Appl. Math. J. Chinese Univ.
2009, 24(1): 1-13

A new family of covariate-adjusted response adaptive designs and their properties

ZHANG Li-Xin[1] HU Fei-fang[2]

Abstract. It is often important to incorporate covariate information in the design of clinical trials. In literature there are many designs of using stratification and covariate-adaptive randomization to balance certain known covariate. Recently, some covariate-adjusted response-adaptive (CARA) designs have been proposed and their asymptotic properties have been studied (Ann. Statist. 2007). However, these CARA designs usually have high variabilities. In this paper, a new family of covariate-adjusted response-adaptive (CARA) designs is presented. It is shown that the new designs have less variables and therefore are more efficient.

§1 Introduction

Response-adaptive designs for clinical trials incorporate sequentially accruing response data into future allocation probabilities. A major objective of response-adaptive designs in clinical trials is to minimize the number of patients that is assigned to the inferior treatment to a degree that still generates useful statistical inferences. The preliminary idea of response adaptive randomization can be traced back to Thompson[17] and Robbins[11]. A lot of response-adaptive designs has already been proposed in literature (e.g., Rosenberger and Lachin[12], Hu and Rosenberger[5]). Much recent work has focused on proposing better randomized adaptive designs. The three main components for evaluating a response-adaptive design are allocation proportion, efficiency (power), and variability. The issue of efficiency or power was discussed by Hu and Rosenberger[4], who showed that the efficiency is a decreasing function of the variability induced by the randomization procedure for any given allocation proportion. Hu, Rosenberger and Zhang[6] showed that there is an asymptotic lower bound on the variability of response-adaptive designs. A response-adaptive design that attains this lower bound will be said to be first order efficient. More recently, Hu, Zhang and He[9] proposed a new family of efficient randomized adaptive designs that can adapt to any desired allocation proportion. But all these studies are limited to the designs that do not incorporate covariates.

Received: 2008-12-20
MR Subject Classification: 62L05,60F05,60F15
Keywords: adaptive design, covariate, efficiency, asymptotic variability
Digital Object Identifier(DOI): 10.1007/s11766-009-0001-6
Partially supported by the National Natural Science Foundation of China (10771192)[1]; NSF Awards DMS-0349048 of USA[2]

In many clinical trials (Pocock and Simon[10], Taves[16]), covariate information is available and has a strong influence on the responses of patients. For instance, the efficacy of a hypertensive drug is related to a patient's initial blood pressure and cholesterol level, whereas the effectiveness of a cancer treatment may depend on whether the patient is a smoker or a nonsmoker. Covariate-adaptive designs have been proposed to balance covariates among treatment groups (see Pocock and Simon[10], Taves[16] and Zelen[19]). Hu and Rosenberger[5] defined a covariate-adjusted response-adaptive (CARA) design as a design that incorporates sequentially the history information of accruing response data and covariate as well as the observed covariate information of the incoming patient into future allocation probabilities.

In a CARA design, the assignment of a treatment depends on the history information and the covariate of the incoming patient. This generates a certain level of technical complexity for studying the properties of the design. Zhang, et al[21] got a limit success on CARA designs by proposing a class of CARA designs that allow a wide spectrum of applications to very general statistical models and obtaining the asymptotic properties to provide a statistical basis for inferences after using this kind of designs. However, the CARA designs in Zhang, et al[21] often have high variabilities and therefore are not efficient (Hu and Rosenberger[4]). The major purpose of this paper is to study the variability and efficiency of CARA designs and to propose a new family of CARA designs with small variabilities.

The paper is organized as follows. In §2, the Fisher information and the best asymptotic variability are derived for a CARA design with any given target allocation proportion. We will find that the Fisher information and the variability depend on the distribution of each individual response, the target function and the distribution of the covariate. In §3, we propose a new CARA design that can be adapted to target any allocation function and in which a parameter can be tuned such that the asymptotic variability approaches to the best one. The design proposed by Zhang, et al[21] is a special case of this new design and has the largest variability in all this kind of designs. The new design is also an extension of the doubly adaptive biased coin design (DBCD) proposed by Eisele and Woodroofe[2] and Hu and Zhang[7]. The technical proofs are put in the Appendix.

§2 Variability and efficiency of CARA designs

2.1 General framework of CARA designs

Given a clinical trial with K treatments. Supposing that a patient with a covariate vector $\boldsymbol{\xi}$ is assigned to treatment k, $k = 1, \cdots, K$, and the observed response is Y_k, we assume that the response Y_k has a conditional distribution $f_k(y_k|\boldsymbol{\theta}_k, \boldsymbol{\xi})$ for the given covariate $\boldsymbol{\xi}$. Here $\boldsymbol{\theta}_k$, $k = 1, \cdots, K$, are unknown parameters, and $\boldsymbol{\Theta}_k \subset \mathbf{R}^d$ is the parameter space of $\boldsymbol{\theta}_k$.

In an adaptive design, we let $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots$ be the sequence of random treatment assignments. For the m-th subject, $\boldsymbol{X}_m = (X_{m,1}, \cdots, X_{m,K})$ represents the assignment of treatment such that if the m-th subject is allocated to treatment k, then all elements in $\boldsymbol{X}_m$ are 0 except for the k-th component, $X_{m,k}$, which is 1. Suppose that $\{Y_{m,k},\ k = 1, \cdots, K,\ m = 1, 2 \cdots\}$ denotes the responses such that $Y_{m,k}$ is the response of the m-th subject to treatment k, $k = 1, \cdots, K$. In practical situations, only $Y_{m,k}$ with $X_{m,k} = 1$ is observed. Denote $\boldsymbol{Y}_m = (Y_{m,1}, \cdots, Y_{m,K})$.

Also, assume that covariate information is available in the clinical study. Let $\boldsymbol{\xi}_m$ be the covariate of the m-th subject. We assume that $\{(Y_{m,1},\cdots,Y_{m,K},\boldsymbol{\xi}_m),\ m=1,2,\cdots\}$ is a sequence of i.i.d. random vectors, the distributions of which are the same as that of $(Y_1,\cdots,Y_K,\boldsymbol{\xi})$. Further, let $\mathscr{X}_m=\sigma(\boldsymbol{X}_1,\cdots,\boldsymbol{X}_m)$, $\mathscr{Y}_m=\sigma(\boldsymbol{Y}_1,\cdots,\boldsymbol{Y}_m)$ and $\mathscr{Z}_m=\sigma(\boldsymbol{\xi}_1,\cdots,\boldsymbol{\xi}_m)$ be the sigma fields corresponding to the responses, assignments and covariates respectively, and let $\mathscr{F}_m=\sigma(\mathscr{X}_m,\mathscr{Y}_m,\mathscr{Z}_m)$ be the sigma field of the history. A general covariate-adjusted response-adaptive (CARA) design is defined by

$$\psi_{m+1,k}=\mathsf{P}(X_{m+1,k}=1|\mathscr{F}_m,\boldsymbol{\xi}_{m+1})=\mathsf{P}(X_{m+1,k}=1|\mathscr{X}_m,\mathscr{Y}_m,\mathscr{Z}_{m+1}),\ k=1,\cdots,K,$$

the conditional probabilities of assigning treatments $1,\cdots,K$ to the mth patient, conditioned on the entire history including the information of all previous m assignments, responses, and covariate vectors, plus the information of the current patient's covariate vector.

2.2 CARA designs with a target

Let $N_{m,k}$ be the number of subjects assigned to treatment k in the first m assignments and write $\boldsymbol{N}_m=(N_{m,1},\cdots,N_{m,K})$. Then $\boldsymbol{N}_m=\sum_{i=1}^m\boldsymbol{X}_i$. Further, let

$$N_{n,k|\boldsymbol{x}}=\sum_{m=1}^n X_{m,k}I\{\boldsymbol{\xi}_m=\boldsymbol{x}\}$$

be the number of subjects with covariate $\boldsymbol{x}$ that is randomized to treatment k, $k=1,\cdots,K$, in the n trials, and $N_n(\boldsymbol{x})=\sum_{m=1}^n I\{\boldsymbol{\xi}_m=\boldsymbol{x}\}$ be the total number of subjects with covariate $\boldsymbol{x}$. Write $\boldsymbol{\theta}=(\boldsymbol{\theta}_1,\cdots,\boldsymbol{\theta}_K)$. Because the value of $\boldsymbol{\theta}$ and the covariate determine the distributions of the outcomes, and accordingly, the effects of each treatment, in many cases one would like to define a CARA design such that the "conditional" allocation proportion for a given covariate $\boldsymbol{x}$ converges to a pre-specified proportion which is a function of $\boldsymbol{\theta}$ and $\boldsymbol{x}$. That is,

$$\frac{N_{n,k|x}}{N_n(x)}\to\pi_k(\boldsymbol{\theta},\boldsymbol{x}),\ k=1,\cdots,K, \tag{2.1}$$

where $\pi_1(\boldsymbol{\theta},\boldsymbol{x})$, $\cdots$, $\pi_K(\boldsymbol{\theta},\boldsymbol{x})$ are K known functions. We call them target allocation functions. Examples for the choice of target functions are discussed by Zhang, et al[21], Rosenberger, et al[13], Rosenberger, Vidyashankar and Agarwal[14] and Hu and Rosenberger[5]. Recently, Tymofyeyev, Rosenberger and Hu[18] developed a general framework to obtain optimal allocation proportion for K-treatment clinical trials. However, when $\mathsf{P}(\boldsymbol{\xi}=\boldsymbol{x})=0$, for example, in the continuous covariate case, the "conditional" allocation proportion $N_{n,k|x}/N_n(x)$ is not well-defined because both the numerator and denominator are zeros almost surely. As compared with (2.1), it is more meaningful to allocate each individual patient to treatment k with a probability close to $\pi_k(\boldsymbol{\theta},\boldsymbol{x})$ for a given covariate $\boldsymbol{x}$. So we consider a class of CARA designs with a property that

$$\mathsf{P}(X_{m+1,k}=1|\mathscr{F}_m,\boldsymbol{\xi}_{m+1}=\boldsymbol{x})\to\pi_k(\boldsymbol{\theta},\boldsymbol{x})\ \text{ a.s.} \tag{2.2}$$

The next theorem tells us that (2.2) implies (2.1). Write $\rho_k(\boldsymbol{\theta})=\mathsf{E}\pi_k(\boldsymbol{\theta},\boldsymbol{\xi})$, $k=1,\cdots,K$, $\boldsymbol{\rho}(\boldsymbol{\theta})=(\rho_1(\boldsymbol{\theta}),\cdots,\rho_K(\boldsymbol{\theta}))$ and $\boldsymbol{\pi}(\boldsymbol{\theta},\boldsymbol{x})=(\pi_1(\boldsymbol{\theta},\boldsymbol{x}),\cdots,\pi_K(\boldsymbol{\theta},\boldsymbol{x}))$.

Theorem 2.1. If (2.2) is satisfied, then

$$\frac{N_{n,k|\boldsymbol{x}}}{N_n(\boldsymbol{x})}\to\pi_k(\boldsymbol{\theta},\boldsymbol{x})\ \text{a.s. on the event}\ \{N_n(x)\to\infty\} \tag{2.3}$$

and

$$\frac{N_{n,k}}{n}\to\rho_k(\boldsymbol{\theta})\ \text{a.s.} \tag{2.4}$$

Here, "A a.s. on B" means that $\mathsf{P}(B \setminus A) = 0$ for two events A and B. Further, if the density of the covariate is positive at $\boldsymbol{x}$, then

$$\lim_{r\searrow 0}\lim_{n\to\infty}\frac{N_{n,k|B(\boldsymbol{x},r)}}{N_n(B(\boldsymbol{x},r))} = \pi_k(\boldsymbol{\theta},\boldsymbol{x}) \ \text{ a.s.}, \tag{2.5}$$

where $N_{n,k|B(\boldsymbol{x},r)} = \sum_{m=1}^n X_{m,k} I\{\boldsymbol{\xi}_m \in B(\boldsymbol{x},r)\}$, $N_n(B(\boldsymbol{x},r)) = \sum_{m=1}^n I\{\boldsymbol{\xi}_m \in B(\boldsymbol{x},r)\}$, $B(\boldsymbol{x},r)$ is a ball with the center $\boldsymbol{x}$ and the radius r.

Notice, when $\mathsf{P}(\boldsymbol{\xi} = \boldsymbol{x}) = 0$, though the allocation proportion $N_{n,k|\boldsymbol{x}}/N_n(\boldsymbol{x})$ is not well-defined, (2.3) is trivial because $\mathsf{P}(N_n(\boldsymbol{x}) \to \infty) = 0$. Accurately, (2.3) makes sense only in the discrete covariate case and (2.5) is a version of (2.3) for continuous covariates.

2.3 Variability and efficiency

For response-adaptive designs which do not incorporate covariates, Hu, Rosenberger and Zhang[6] found the lower bound of the asymptotic variability of a design, i.e., of the allocation proportions of the design. A design is called asymptotically efficient if its asymptotic variability attains the lower bound. Next, we study the variability and efficiency of CARA designs. Suppose, given $\boldsymbol{\xi}$, that the response Y_k of a trial of treatment k has a distribution in the exponential family, and takes the form

$$f_k(y_k|\boldsymbol{\xi},\boldsymbol{\theta}_k) = \exp\left\{(y_k\mu_k - a_k(\mu_k))/\phi_k + b_k(y_k,\phi_k)\right\} \tag{2.6}$$

with link function $\mu_k = h_k(\boldsymbol{\xi}\boldsymbol{\theta}_k^T)$, where $\boldsymbol{\theta}_k = (\theta_{k1},\cdots,\theta_{kd})$, $k = 1,\cdots,K$, are coefficients. Assume that the scale parameter ϕ_k is fixed. It is easily checked that $\mathsf{E}[Y_k|\boldsymbol{\xi}] = a_k'(\mu_k)$, $\mathsf{Var}(Y_k|\boldsymbol{\xi}) = a_k''(\mu_k)\phi_k$,

$$\frac{\partial \log f_k(y_k|\boldsymbol{\xi},\boldsymbol{\theta}_k)}{\partial \boldsymbol{\theta}_k} = \frac{1}{\phi_k}\{y_k - a_k'(\mu_k)\}h_k'(\boldsymbol{\xi}\boldsymbol{\theta}_k^T)\boldsymbol{\xi},$$

$$\frac{\partial^2 \log f_k(y_k|\boldsymbol{\xi},\boldsymbol{\theta}_k)}{\partial \boldsymbol{\theta}_k^2} = \frac{1}{\phi_k}\Big\{ - a_k''(\mu_k)[h_k'(\boldsymbol{\xi}\boldsymbol{\theta}_k^T)]^2 + [y_k - a_k'(\mu_k)]h_k''(\boldsymbol{\xi}\boldsymbol{\theta}_k^T)\Big\}\boldsymbol{\xi}^T\boldsymbol{\xi}$$

and, given $\boldsymbol{\xi}$, the conditional Fisher information matrix is

$$\boldsymbol{I}_k(\boldsymbol{\theta}_k|\boldsymbol{\xi}) = -\mathsf{E}\Big[\frac{\partial^2 \log f_k(Y_k|\boldsymbol{\xi},\boldsymbol{\theta}_k)}{\partial \boldsymbol{\theta}_k^2}\Big|\boldsymbol{\xi}\Big] = \frac{1}{\phi_k}a_k''(\mu_k)[h_k'(\boldsymbol{\xi}\boldsymbol{\theta}_k^T)]^2\boldsymbol{\xi}^T\boldsymbol{\xi}.$$

For the observations up to stage n, the likelihood function is

$$L(\boldsymbol{\theta}) = \prod_{j=1}^n\prod_{k=1}^K [f_k(Y_{j,k}|\boldsymbol{\xi}_j,\boldsymbol{\theta}_k)]^{X_{j,k}} = \prod_{k=1}^K\prod_{j=1}^n [f_k(Y_{j,k}|\boldsymbol{\xi}_j,\boldsymbol{\theta}_k)]^{X_{j,k}} := \prod_{k=1}^K L_k(\boldsymbol{\theta}_k) \tag{2.7}$$

with $\log L_k(\boldsymbol{\theta}_k) \propto \sum_{j=1}^n X_{j,k}\{Y_{j,k} - a_k(\mu_{j,k})\}$, $\mu_{j,k} = h_k(\boldsymbol{\theta}_k^T\boldsymbol{\xi}_j)$, $k = 1,2,\cdots,K$. Write

$$\boldsymbol{I}_k = \mathsf{E}[\pi_k(\boldsymbol{\theta},\boldsymbol{\xi})\boldsymbol{I}_k(\boldsymbol{\theta}_k|\boldsymbol{\xi})], \quad k = 1,\cdots,K. \tag{2.8}$$

Then $-\mathsf{E}_{\boldsymbol{\theta}}\left[\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_k^2}\right] = \sum_{j=1}^n \mathsf{E}_{\boldsymbol{\theta}}\left[X_{j,k}\boldsymbol{I}_k(\boldsymbol{\theta}_k|\boldsymbol{\xi}_j)\right] = n\boldsymbol{I}_k + o(n)$. It follows that the entire Firsher information matrix is

$$\boldsymbol{I}_n(\boldsymbol{\theta}) = -\mathsf{E}_{\boldsymbol{\theta}}\left[\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2}\right] = n\mathrm{diag}(\boldsymbol{I}_1,\cdots,\boldsymbol{I}_K) + o(n).$$

Thus we obtain the following theorem.

Theorem 2.2. Suppose the responses follow the generalized linear model (2.6) and the design satisfies (2.2). Let $\boldsymbol{I}(\boldsymbol{\theta}) = \mathrm{diag}(\boldsymbol{I}_1,\cdots,\boldsymbol{I}_K)$. Then the Firsher information matrix satisfies

$$\boldsymbol{I}_n(\boldsymbol{\theta}) = n\boldsymbol{I}(\boldsymbol{\theta}) + o(n),$$

and the asymptotic variance-covariance matrix of an asymptotically efficient estimator of $\boldsymbol{\theta}$ is $\boldsymbol{I}^{-1}(\boldsymbol{\theta})/n$.

The limit proportion $\boldsymbol{\rho}(\boldsymbol{\theta}) = (\rho_1(\boldsymbol{\theta}), \cdots, \rho_K(\boldsymbol{\theta}))$ depends on both the parameter $\boldsymbol{\theta}$ and the distribution of $\boldsymbol{\xi}$. When the distribution of $\boldsymbol{\xi}$ is known, according to Theorem 2.2, the asymptotic variance-covariance matrix of an asymptotically efficient estimator of $\boldsymbol{\rho}(\boldsymbol{\theta})$ is

$$\frac{1}{n}\frac{\partial \boldsymbol{\rho}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\boldsymbol{I}^{-1}(\boldsymbol{\theta})\left(\frac{\partial \boldsymbol{\rho}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T.$$

While, if the parameter $\boldsymbol{\theta}$ is known, then the non-parameter maximal likelihood estimator (MLE) of $\boldsymbol{\rho}(\boldsymbol{\theta}) = \mathsf{E}[\boldsymbol{\pi}(\boldsymbol{\theta}, \boldsymbol{\xi})]$ is $\frac{1}{n}\sum_{m=1}^{n}\boldsymbol{\pi}(\boldsymbol{\theta}, \boldsymbol{\xi}_m)$ and its variance-covariance matrix is $\frac{\mathsf{Var}\{\pi(\boldsymbol{\theta},\boldsymbol{\xi})\}}{n}$. So, in the general case that the parameter $\boldsymbol{\theta}$ and the distribution of $\boldsymbol{\xi}$ are both unknown, the asymptotic variance-covariance matrix of an asymptotic efficient estimator of $\boldsymbol{\rho}(\boldsymbol{\theta})$ is $\boldsymbol{B}(\boldsymbol{\theta})/n$, where

$$\boldsymbol{B}(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{\rho}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\boldsymbol{I}^{-1}(\boldsymbol{\theta})\left(\frac{\partial \boldsymbol{\rho}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T + \mathsf{Var}\{\pi(\boldsymbol{\theta}, \boldsymbol{\xi})\}.$$

The allocation proportion $\boldsymbol{N}_n/n$ in an adaptive design with property (2.2) will converge to $\boldsymbol{\rho}(\boldsymbol{\theta})$ according to Theorem 2.1. So we can now define an asymptotically efficient CARA design as follows.

Definition 2.1. A covaraite-adjusted response-adaptive design with target function $\boldsymbol{\pi}(\boldsymbol{\theta}, \boldsymbol{x})$ is called asymptotically efficient if it satisfies (2.2) and

$$n^{1/2}\big(\boldsymbol{N}_n/n - \boldsymbol{\rho}(\boldsymbol{\theta})\big) \xrightarrow{\mathscr{D}} N\big(\mathbf{0}, \boldsymbol{B}(\boldsymbol{\theta})\big), \tag{2.9}$$

where $\boldsymbol{B}(\boldsymbol{\theta})$ is called the best asymptotic variability.

Zhang, Hu, Cheung and Chan[21] proposed a CARA design (we refer it as ZHCC's design) by defining $\mathsf{P}(X_{m+1,k} = 1|\mathscr{F}_m, \boldsymbol{\xi}_m) = \pi_k(\widehat{\boldsymbol{\theta}}_m, \boldsymbol{\xi}_{m+1})$, where $\widehat{\boldsymbol{\theta}}_m$ is the MLE of $\boldsymbol{\theta}$ based on the observations up to stage m. It has been shown that ZHCC's design satisfies (2.2) and $n^{1/2}\big(\boldsymbol{N}_n/n - \boldsymbol{\rho}(\boldsymbol{\theta})\big) \xrightarrow{\mathscr{D}} N\big(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta})\big)$, where

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = 2\frac{\partial \boldsymbol{\rho}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\boldsymbol{I}^{-1}(\boldsymbol{\theta})\left(\frac{\partial \boldsymbol{\rho}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T + \mathrm{diag}(\boldsymbol{\rho}(\boldsymbol{\theta})) - \big(\boldsymbol{\rho}(\boldsymbol{\theta})\big)^T\boldsymbol{\rho}(\boldsymbol{\theta}).$$

It is easily seen that

$$\begin{aligned}&\mathrm{diag}(\boldsymbol{\rho}(\boldsymbol{\theta})) - \big(\boldsymbol{\rho}(\boldsymbol{\theta})\big)^T\boldsymbol{\rho}(\boldsymbol{\theta})\\ =&\mathsf{Var}\{\pi(\boldsymbol{\theta}, \boldsymbol{\xi})\} + \mathsf{E}\left[\mathrm{diag}(\pi(\boldsymbol{\theta}, \boldsymbol{\xi})) - (\pi(\boldsymbol{\theta}, \boldsymbol{\xi}))^T\,\pi(\boldsymbol{\theta}, \boldsymbol{\xi})\right] \geq \mathsf{Var}\{\pi(\boldsymbol{\theta}, \boldsymbol{\xi})\},\end{aligned}$$

where $\boldsymbol{A} \geq \boldsymbol{B}$ means that $\boldsymbol{A} - \boldsymbol{B}$ is non-negative definite. Hence, ZHCC's design is not asymptotically efficient.

It is of significance to find an asymptotically efficient CARA design for any given target function $\boldsymbol{\pi}(\boldsymbol{\theta}, \boldsymbol{x})$. In the next section, we will propose a new class of CARA designs with an asymptotic variability being able to approach the best one.

§3 Covariate-adjusted DBCD

Our new design is based on the idea of the doubly adaptive biased coin design (DBCD) proposed by Eisele and Woodroofe[2], and extended by Hu and Zhang[7]. In the scenario without covariates, the Hu and Zhang's extension can target any desired allocation and can approach the lower bound of the asymptotic variability by tuning a parameter. In this section, we modify the DBCD to incorporate covariates. For simplicity, we only consider the two-treatment case ($K = 2$).

Covariate-adjusted DBCD (CADBCD): To start, we let $\boldsymbol{\theta}_0$ be an initial estimate of $\boldsymbol{\theta}$, and assign m_0 subjects to each treatment by using a restricted randomization. Assume that m ($m \geq 2m_0$) subjects have been assigned to treatments. Their responses $\{\boldsymbol{Y}_j,\ j = 1, \cdots, m\}$ and the corresponding covariates $\{\boldsymbol{\xi}_j,\ j = 1, \cdots, m\}$ are observed. We let $\widehat{\boldsymbol{\theta}}_m = (\widehat{\boldsymbol{\theta}}_{m,1}, \widehat{\boldsymbol{\theta}}_{m,2})$ be an estimate of $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. Here, for each $k = 1, 2$, $\widehat{\boldsymbol{\theta}}_{m,k} = \widehat{\boldsymbol{\theta}}_{m,k}(Y_{j,k}, \boldsymbol{\xi}_j : X_{j,k} = 1, j = 1, \cdots, m)$ is the estimator of $\boldsymbol{\theta}_k$ that is based on the observed $N_{m,k}$-size sample $\{(Y_{j,k}, \boldsymbol{\xi}_j)$: for which $X_{j,k} = 1, j = 1 \cdots, m\}$. Write $\widehat{\rho}_m = \frac{1}{m}\sum_{i=1}^{m} \pi_1(\widehat{\boldsymbol{\theta}}_m, \boldsymbol{\xi}_i)$ and $\widehat{\pi}_m = \pi_1(\widehat{\boldsymbol{\theta}}_m, \boldsymbol{\xi}_{m+1})$. Next, when the $(m+1)$-th subject is ready for randomization and the corresponding covariate $\boldsymbol{\xi}_{m+1}$ is recorded, we assign the patient to treatment 1 with a probability of

$$\psi_{m+1,1} = \frac{\widehat{\pi}_m \left(\frac{\widehat{\rho}_m}{N_{m,1}/m}\right)^{\gamma}}{\widehat{\pi}_m \left(\frac{\widehat{\rho}_m}{N_{m,1}/m}\right)^{\gamma} + (1-\widehat{\pi}_m)\left(\frac{1-\widehat{\rho}_m}{1-N_{m,1}/m}\right)^{\gamma}} \tag{3.1}$$

and to treatment 2 with a probability of $\psi_{m+1,2} = 1 - \psi_{m+1,1}$, where $\gamma \geq 0$ is a constant that controls the degree of randomness of the procedure, from the most random when $\gamma = 0$ to the deterministic when $\gamma = \infty$. ZHCC's design is a special case of CADBCD with $\gamma = 0$.

Asymptotic properties. For studying the asymptotic properties, we assume the target allocation function $\pi_1(\boldsymbol{\theta}^*, \boldsymbol{x})$ satisfies the following condition.

Condition 3.1. We assume that the parameter space $\boldsymbol{\Theta}_k$ is a bounded domain in $\mathbf{R}^d$, and that the true value $\boldsymbol{\theta}_k$ is an interior point of $\boldsymbol{\Theta}_k$, $k = 1, 2$.

1. For each fixed $\boldsymbol{x}$, $0 < \pi_1(\boldsymbol{\theta}^*, \boldsymbol{x}) < 1$ is a continuous function of $\boldsymbol{\theta}^*$ in the closure of $\boldsymbol{\Theta}_1 \times \boldsymbol{\Theta}_2$.
2. $\pi_1(\boldsymbol{\theta}^*, \boldsymbol{\xi})$ is twice differentiable with respect to $\boldsymbol{\theta}^*$, and the expectations of $\|\partial\pi_1(\boldsymbol{\theta}, \boldsymbol{\xi})/\partial\boldsymbol{\theta}\|^2$ and $\sup_{\|\boldsymbol{\theta}^* - \boldsymbol{\theta}\| \leq \delta} \|\partial^2\pi_1(\boldsymbol{\theta}^*, \boldsymbol{\xi})/\partial\boldsymbol{\theta}^2\|$ are finite for some $\delta > 0$.

Write $v = \mathsf{E}[\pi_1(\boldsymbol{\theta}, \boldsymbol{\xi})]$, then $0 < v < 1$ due to Condition 3.1.

Theorem 3.1. Suppose that for $k = 1, 2$,

$$\widehat{\boldsymbol{\theta}}_{nk} - \boldsymbol{\theta}_k = \frac{1}{n}\sum_{m=1}^{n} X_{m,k}\boldsymbol{h}_k(Y_{m,k}, \boldsymbol{\xi}_m)\big(1 + o(1)\big) + o(n^{-1/2}) \quad \text{a.s.}, \tag{3.2}$$

where $\boldsymbol{h}_k$ are functions with $\mathsf{E}[\boldsymbol{h}_k(Y_k, \boldsymbol{\xi})|\boldsymbol{\xi}] = \mathbf{0}$. We also assume that $\mathsf{E}\|\boldsymbol{h}_k(Y_k, \boldsymbol{\xi})\|^2 < \infty$, $k = 1, 2$. Then under Condition 3.1 we have

$$\mathsf{P}\big(X_{n,1} = 1\big) \to v; \quad \mathsf{P}\big(X_{n,1} = 1|\mathscr{F}_{n-1}, \boldsymbol{\xi}_n = \boldsymbol{x}\big) \to \pi_1(\boldsymbol{\theta}, \boldsymbol{x}) \text{ a.s.} \tag{3.3}$$

and

$$\frac{N_{n,1}}{n} - v = O\Big(\sqrt{\frac{\log\log n}{n}}\Big) \text{ a.s.}; \quad \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} = O\Big(\sqrt{\frac{\log\log n}{n}}\Big) \text{ a.s.} \tag{3.4}$$

Further, let $\boldsymbol{V}_k = \mathsf{E}\{\pi_k(\boldsymbol{\theta}, \boldsymbol{\xi})(\boldsymbol{h}_k(Y_k, \boldsymbol{\xi}))^T\boldsymbol{h}_k(Y_k, \boldsymbol{\xi})\}, k = 1, 2, \boldsymbol{V} = \mathrm{diag}\big(\boldsymbol{V}_1, \boldsymbol{V}_2\big)$, $\sigma_1^2 = \mathsf{E}[\pi_1(\boldsymbol{\theta}, \boldsymbol{\xi})(1 - \pi_1(\boldsymbol{\theta}, \boldsymbol{\xi}))]$, $\sigma_2^2 = \mathsf{Var}\{\pi_1(\boldsymbol{\theta}, \boldsymbol{\xi})\}$, $\sigma_3^2 = \mathsf{E}\frac{\partial\pi_1(\boldsymbol{\theta}, \boldsymbol{\xi})}{\partial\boldsymbol{\theta}}\boldsymbol{V}\big(\mathsf{E}\frac{\partial\pi_1(\boldsymbol{\theta}, \boldsymbol{\xi})}{\partial\boldsymbol{\theta}}\big)^T$, $\lambda = \gamma\frac{\sigma_1^2}{v(1-v)}$ and $\sigma^2 = \frac{\sigma_1^2 + \sigma_3^2}{1+2\lambda} + \sigma_2^2 + \sigma_3^2$. Then,

$$\sqrt{n}(N_{n,1}/n - v) \xrightarrow{D} N(0, \sigma^2) \ \text{ and } \ \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{V}). \tag{3.5}$$

The proof of this Theorem is a little complicated and will be stated in the Appendix. According to (3.3), CADBCD satisfies (2.2). The asymptotic variability σ^2 of the design takes the values from the maximum $2\sigma_3^2 + v(1-v)$ when $\gamma = 0$ to the minimum $\sigma_2^2 + \sigma_3^2$ when $\gamma = \infty$.

The next result for the generalized linear model is a corollary of Theorem 3.1. The proof is given in the Appendix through the verification of Condition (3.2).

Corollary 3.1. Suppose the distributions of the responses follow the generalized linear model (2.6) and satisfy the following regular condition

$$H(\delta) =: \mathsf{E}_{\boldsymbol{\theta}}\Big[\sup_{\|\boldsymbol{z}\|\le\delta}\Big\|\frac{\partial^2 \log f_k(Y_k|\boldsymbol{\xi},\boldsymbol{\theta}_k)}{\partial\boldsymbol{\theta}_k^2}\Big|_{\boldsymbol{\theta}_k}^{\boldsymbol{\theta}_k+\boldsymbol{z}}\Big\|\Big] \to 0 \text{ as } \delta \to 0, \tag{3.6}$$

where $f(x)|_a^b = f(b) - f(a)$. Under Condition 3.1, if the matrices $\boldsymbol{I}_1$ and $\boldsymbol{I}_2$ defined as in (2.8) are nonsingular and the MLE $\widehat{\boldsymbol{\theta}}_m$, which maximizes the likelihood function (2.7), is unique, then we have (3.3), (3.4), and (3.5) with $\boldsymbol{V} = \boldsymbol{I}^{-1}(\boldsymbol{\theta})$ and $\boldsymbol{I}(\boldsymbol{\theta}) = \mathrm{diag}(\boldsymbol{I}_1, \boldsymbol{I}_2)$.

It is obvious that $B(\boldsymbol{\theta}) = \sigma_2^2 + \sigma_3^2$ is the best asymptotic variability of CARA designs with two treatments according to Definition 2.1. For the CADBCD,

$$\sigma^2 = \frac{\sigma_1^2 + \sigma_3^2}{1 + 2\gamma\frac{\sigma_1^2}{v(1-v)}} + B(\boldsymbol{\theta}) > B(\boldsymbol{\theta}) \text{ but } \sigma^2 \searrow B(\boldsymbol{\theta}) \text{ as } \gamma \nearrow \infty.$$

This means that CADBCD is not asymptotically efficient but it can approach an asymptotically efficient CARA design if γ is chosen large. ZHCC's design is a special case of CADBCD which has the largest variability.

§4 Conclusion Remarks

We have proposed a family of covariate-adjusted response-adaptive designs that are fully randomized and asymptotically efficient. CADBCD can be viewed as a generalization of Hu and Zhang's doubly adaptive biased coin design[7] for incorporating covariate information. The asymptotic properties derived here provide the theoretical foundation for inference based on CADBCD.

In this paper, we have assumed that the responses in each treatment group are available without delay. In practice, there is no logistical difficulty in incorporating delayed responses into CADBCD, provided that some responses become available during the course of the allocation in the experiment, and thus we can always update the estimates whenever new data become available. For clinical trials with uniform (or exponential) patient entry and exponential response times (see Bai, Hu and Rosenberger[1], Hu and Zhang[8] and Zhang, et al[20] for example), it is easy to verify the theoretical results in §2 and §3.

§5 Appendix: Proofs

Proof of Theorem 2.1. Notice $\mathsf{E}[X_{m+1,k}|\mathscr{F}_m] = \mathsf{E}[\psi_{m+1,k}|\mathscr{F}_m] \to \rho_k(\boldsymbol{\theta})$ by (2.2) and

$$\{\sum_{m=1}^{n}(X_{m,k} - \mathsf{E}[X_{m,k}|\mathscr{F}_{m-1}]), \mathscr{F}_n\}$$

is a martingale. (2.4) follows immediately. For (2.3), let $\mathscr{G}_m = \sigma(\mathscr{F}_m, \boldsymbol{\xi}_{m+1})$. Then

$$\{\sum_{m=1}^{n}(X_{m,k} - \mathsf{E}[X_{m,k}|\mathscr{G}_{m-1}])I\{\boldsymbol{\xi}_m = x\}, \mathscr{G}_m\}$$

is a martingale with $\sum_{m=1}^{n}\mathsf{E}\big[\{(X_{m,k}-\mathsf{E}[X_{m,k}|\mathscr{G}_{m-1}])I\{\boldsymbol{\xi}_m=x\}\}^2|\mathscr{G}_{m-1}\big]\le N_n(x)$. It follows that $\frac{\sum_{m=1}^{n}(X_{m,k}-\mathsf{E}[X_{m,k}|\mathscr{G}_{m-1}])I\{\boldsymbol{\xi}_m=x\}}{N_n(x)}\to 0$ a.s. on $\{N_n(x)\to\infty\}$ by Theorem 3.3.10 of Stout [15]. On the other hand,

$$\frac{\sum_{m=1}^{n}(\mathsf{E}[X_{m,k}|\mathscr{G}_{m-1}]-\pi_k(\boldsymbol{\theta},x))I\{\boldsymbol{\xi}_m=x\}}{N_n(x)}\to 0 \text{ a.s. on } \{N_n(x)\to\infty\}$$

by (2.2). So, (2.3) is proved. For (2.5), notice $\frac{N_n(B(\boldsymbol{x},r))}{n}\to\mathsf{P}\{\boldsymbol{\xi}\in B(\boldsymbol{x},r)\}>0$ a.s. With a similar argument we have

$$\lim_{n\to\infty}\frac{N_{n,k|B(\boldsymbol{x},r)}}{N_n(B(\boldsymbol{x},r))}=\lim_{n\to\infty}\frac{\sum_{m=1}^{n}\pi_k(\boldsymbol{\theta},\boldsymbol{\xi}_m)I\{\boldsymbol{\xi}_m\in B(\boldsymbol{x},r)\}}{N_n(B(\boldsymbol{x},r))}=\frac{\mathsf{E}[\pi_k(\boldsymbol{\theta},\boldsymbol{\xi})I\{\boldsymbol{\xi}\in B(\boldsymbol{x},r)\}]}{\mathsf{P}\{\boldsymbol{\xi}\in B(\boldsymbol{x},r)\}} \text{ a.s.}$$

Letting $r\searrow 0$ yields (2.5).

Proof of Theorem 3.1. The proof is a little complicate and long. We will complete via four steps.

Step 1. We show that (3.4) and

$$\widehat{\rho}_m=v+O(\sqrt{\log\log m/m}) \text{ a.s.} \tag{5.1}$$

Write $\pi_1=\pi_1(\boldsymbol{\theta},\boldsymbol{\xi})$ for short. Let

$$M_{n,1}=\sum_{m=1}^{n}\left(X_{m,1}-\mathsf{E}[X_{m,1}|\mathscr{F}_{m-1},\boldsymbol{\xi}_m]\right),\quad M_{n,2}=\sum_{m=1}^{n}\left(\pi_1(\boldsymbol{\theta},\boldsymbol{\xi}_m)-\mathsf{E}\pi_1\right),$$

$$\boldsymbol{Q}_{n,k}=\sum_{m=1}^{n}X_{m,k}\boldsymbol{h}_k(Y_{m,k},\boldsymbol{\xi}_m) \text{ for } k=1,2,$$

$\boldsymbol{Q}_n=(\boldsymbol{Q}_{n,1},\boldsymbol{Q}_{n,2})$ and $M_{n,3}=\boldsymbol{Q}_n\left(\mathsf{E}\frac{\partial\pi_1}{\partial\boldsymbol{\theta}}\right)^T$. Then $\boldsymbol{Q}_n$ and $M_{n,j}$, $j=1,2,3$, are martingales. According to the law of iterated logarithm (LIL) for martingales, we have

$$\boldsymbol{Q}_n=O(\sqrt{\log\log n/n}) \text{ and } M_{n,j}=O(\sqrt{\log\log n/n}) \text{ a.s.} j=1,2,3. \tag{5.2}$$

Hence, by (3.2) it is easily shown that

$$\widehat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}=O(\sqrt{\log\log m/m})\text{a.s.} \tag{5.3}$$

It follows that

$$\begin{aligned}\widehat{\pi}_m=&\pi_1(\widehat{\boldsymbol{\theta}}_m,\boldsymbol{\xi}_{m+1})=\pi_1(\boldsymbol{\theta},\boldsymbol{\xi}_{m+1})+(\widehat{\boldsymbol{\theta}}_m-\boldsymbol{\theta})\left(\frac{\partial\pi_1(\boldsymbol{\theta},\boldsymbol{\xi}_{m+1})}{\partial\boldsymbol{\theta}}\right)^T\\&+O(1)\|\widehat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}\|^2\sup_{\|\boldsymbol{\theta}^*-\boldsymbol{\theta}\|\le\delta}\left\|\frac{\partial^2\pi_1(\boldsymbol{\theta}^*,\boldsymbol{\xi}_{m+1})}{\partial\boldsymbol{\theta}^2}\right\| \qquad (5.4)\\=&\pi_1(\boldsymbol{\theta},\boldsymbol{\xi}_{m+1})+(\widehat{\boldsymbol{\theta}}_m-\boldsymbol{\theta})\mathsf{E}\frac{\partial\pi}{\partial\boldsymbol{\theta}}+(\widehat{\boldsymbol{\theta}}_m-\boldsymbol{\theta})\left[\frac{\partial\pi_1(\boldsymbol{\theta},\boldsymbol{\xi}_{m+1})}{\partial\boldsymbol{\theta}}-\mathsf{E}\frac{\partial\pi_1}{\partial\boldsymbol{\theta}}\right]^T\\&+O(1)\frac{\log\log m}{m}\sup_{\|\boldsymbol{\theta}^*-\boldsymbol{\theta}\|\le\delta}\left\|\frac{\partial^2\pi_1(\boldsymbol{\theta}^*,\boldsymbol{\xi}_{m+1})}{\partial\boldsymbol{\theta}^2}\right\| \text{ a.s.}\end{aligned}$$

It is easily shown that $\sum_{m=1}^{n}(\widehat{\boldsymbol{\theta}}_m-\boldsymbol{\theta})\left[\frac{\partial\pi_1(\boldsymbol{\theta},\boldsymbol{\xi}_{m+1})}{\partial\boldsymbol{\theta}}-\mathsf{E}\frac{\partial\pi_1}{\partial\boldsymbol{\theta}}\right]^T=o((\log n)^2)$ a.s. and

$$\sum_{m=1}^{n}\frac{\log\log m}{m}\sup_{\|\boldsymbol{\theta}^*-\boldsymbol{\theta}\|\le\delta}\left\|\frac{\partial^2\pi_1(\boldsymbol{\theta}^*,\boldsymbol{\xi}_{m+1})}{\partial\boldsymbol{\theta}^2}\right\|=o((\log n)^2) \text{ a.s.}$$

It follows that

$$\sum_{m=1}^{n}\widehat{\pi}_m=\sum_{m=1}^{n}\pi_1(\boldsymbol{\theta},\boldsymbol{\xi}_{m+1})+\sum_{m=1}^{n}(\widehat{\boldsymbol{\theta}}_m-\boldsymbol{\theta})\left(\mathsf{E}\frac{\partial\pi_1}{\partial\boldsymbol{\theta}}\right)^T+o((\log n)^2)\text{ a.s.}\tag{5.5}$$

Similarly,

$$\begin{aligned}\widehat{\rho}_m&=\frac{1}{m}\sum_{i=1}^{m}\pi_1(\widehat{\boldsymbol{\theta}}_m,\boldsymbol{\xi}_i)\\ &=\frac{1}{m}\sum_{i=1}^{m}\pi_1(\boldsymbol{\theta},\boldsymbol{\xi}_i)+(\widehat{\boldsymbol{\theta}}_m-\boldsymbol{\theta})\left(\mathsf{E}\frac{\partial\pi_1(\boldsymbol{\theta},\boldsymbol{\xi})}{\partial\boldsymbol{\theta}}\right)^T\\ &\quad+(\widehat{\boldsymbol{\theta}}_m-\boldsymbol{\theta})\frac{1}{m}\sum_{i=1}^{m}\left[\frac{\partial\pi_1(\boldsymbol{\theta},\boldsymbol{\xi}_i)}{\partial\boldsymbol{\theta}}-\mathsf{E}\frac{\partial\pi_1(\boldsymbol{\theta},\boldsymbol{\xi})}{\partial\boldsymbol{\theta}}\right]^T\\ &\quad+O(1)\|\widehat{\boldsymbol{\theta}}_m-\boldsymbol{\theta}\|^2\frac{1}{m}\sum_{i=1}^{m}\sup_{\|\boldsymbol{\theta}^*-\boldsymbol{\theta}\|\le\delta}\left\|\frac{\partial^2\pi_1(\boldsymbol{\theta}^*,\boldsymbol{\xi}_i)}{\partial\boldsymbol{\theta}^2}\right\|\qquad(5.6)\\ &=\frac{1}{m}\sum_{i=1}^{m}\pi_1(\boldsymbol{\theta},\boldsymbol{\xi}_i)+(\widehat{\boldsymbol{\theta}}_m-\boldsymbol{\theta})\left(\mathsf{E}\frac{\partial\pi_1(\boldsymbol{\theta},\boldsymbol{\xi})}{\partial\boldsymbol{\theta}}\right)^T+O(\frac{\log\log m}{m}).\qquad(5.7)\end{aligned}$$

It follows that

$$\widehat{\rho}_m=v+\frac{1}{m}\sum_{i=1}^{m}[\pi_1(\boldsymbol{\theta},\boldsymbol{\xi}_i)-\mathsf{E}\pi_1]+O(\sqrt{\log\log m/m})=v+O(\sqrt{\log\log m/m})\text{ a.s.}$$

and $\sum_{m=1}^{n}\widehat{\pi}_m=nv+O(\sqrt{n\log\log n})$ a.s. Now, write

$$g(\pi,a,b)=\frac{\pi(b/a)^\gamma}{\pi(b/a)^\gamma+(1-\pi)((1-b)/(1-a))^\gamma}.\tag{5.8}$$

Then $\psi_{m+1,1}=g\left(\widehat{\pi}_m,N_{m,1}/m,\widehat{\rho}_m\right)$. It is easily seen that $g(\pi,a,b)$ is a non-decreasing function of b, and so $g(\pi,a,b)\le g(\pi,a,a)=\pi$ if $a\ge b$. Let $l_n=\max\{m\le n:N_{m,1}/m\le\widehat{\rho}_m\}$, then $\psi_{m+1,1}\le\widehat{\pi}_m$ when $m\ge l_n+1$. Hence

$$\begin{aligned}N_{n,1}=&N_{l_n+1,1}+M_{n,1}-M_{l_n+1,1}+\sum_{m=l_n+1}^{n-1}\psi_{m+1,1}\\ \le&1+N_{l_n,1}+M_{n,1}-M_{l_n+1,1}+\sum_{m=l_n+1}^{n-1}\widehat{\pi}_m\\ \le&1+l_n\widehat{\rho}_{l_n}+M_{n,1}-M_{l_n+1,1}+\sum_{m=1}^{n-1}\widehat{\pi}_m-\sum_{m=1}^{l_n}\widehat{\pi}_m\qquad(5.9)\\ \le&nv+O(\sqrt{n\log\log n})\text{ a.s.}\end{aligned}$$

Similarly, $n-N_{n,1}\le n(1-v)+O(\sqrt{n\log\log n})$ a.s. (3.4) and (5.1) are now proved.

Step 2. We show (3.3) and the asymptotic normality of $\widehat{\boldsymbol{\theta}}_n$.

By (3.4) and (5.1), $\widehat{\rho}_n/(N_{n,1}/n)\to1$ a.s., and hence (3.3) is proved. Further, $\psi_{m,1}-$

$\pi_1(\widehat{\boldsymbol{\theta}}_{m-1}, \boldsymbol{\xi}_m) \to 0$ a.s., then it is easily checked that $\boldsymbol{Q}_n$ is a martingale with

$$\begin{aligned}&\frac{1}{n}\sum_{m=1}^{n}\mathsf{E}\left[(\Delta\boldsymbol{Q}_n)^T\boldsymbol{\Delta}Q_n\right]\\ =&\frac{1}{n}\sum_{m=1}^{n}\operatorname{diag}\Big(\mathsf{E}\left[\psi_{m,1}\boldsymbol{h}_1(Y_{m,1},\boldsymbol{\xi}_m)^T\boldsymbol{h}_1(Y_{m,1},\boldsymbol{\xi}_m)\right],\\ &\qquad\mathsf{E}\left[\psi_{m,2}\boldsymbol{h}_2(Y_{m,2},\boldsymbol{\xi}_m)^T\boldsymbol{h}_2(Y_{m,2},\boldsymbol{\xi}_m)\right]\Big)\to\boldsymbol{V}.\end{aligned}$$

So, applying the central limit theorem for martingales yields $n^{1/2}(\widehat{\boldsymbol{\theta}}_n-\boldsymbol{\theta})\xrightarrow{\mathscr{D}}N(\boldsymbol{0},\boldsymbol{V})$. The proof of Step 2 is completed.

Step 3. We show that

$$\psi_{m+1,1}=\widehat{\pi}_m-\gamma\frac{\widehat{\pi}_m(1-\widehat{\pi}_m)}{v(1-v)}\left(\frac{N_{m,1}}{m}-\widehat{\rho}_m\right)+O(\frac{\log\log m}{m})\ \text{ a.s.}\tag{5.10}$$

Let $g(\pi,a,b)$ be defined as in (5.8). By some elementary argument, it can be showed that

$$\sup_{0\le\pi\le1}\left|g(\pi,a,b)-\pi+\gamma\frac{\pi(1-\pi)}{v(1-v)}(a-b)\right|=O((a-v)^2+(b-v)^2),\tag{5.11}$$

as $(a,b)\to(v,v)$. By (3.4) and (5.1), it follows that

$$\sup_{0\le\pi\le1}\left|g(\pi,N_{m,1}/m,\widehat{\rho}_m)-\pi+\gamma\frac{\pi(1-\pi)}{v(1-v)}\left(\frac{N_{m,1}}{m}-\widehat{\rho}_m\right)\right|=O(\frac{\log\log m}{m})\ \text{ a.s.}$$

(5.10) is now proved.

Step 4. At last, we show the asymptotic normality of N_n.

Notice $N_{m,1}/m-\widehat{\rho}_m=O(\sqrt{\log\log m/m})$ a.s. With the same argument as deriving (5.5), we can show that

$$\sum_{m=1}^{n}\frac{\widehat{\pi}_m(1-\widehat{\pi}_m)}{v(1-v)}\left(\frac{N_{m,1}}{m}-\widehat{\rho}_m\right)=\sum_{m=1}^{n}\frac{\mathsf{E}[\pi_1(1-\pi_1)]}{v(1-v)}\left(\frac{N_{m,1}}{m}-\widehat{\rho}_m\right)+o((\log n)^2)\ \text{a.s.}$$

By (5.10) it follows that

$$\begin{aligned}\sum_{m=1}^{n}\psi_{m-1,1}=&\sum_{m=1}^{n}\pi_1(\boldsymbol{\theta},\boldsymbol{\xi}_m)+\sum_{m=0}^{n-1}(\widehat{\boldsymbol{\theta}}_m-\boldsymbol{\theta})\left(\mathsf{E}\frac{\partial\pi_1}{\partial\boldsymbol{\theta}}\right)^T\\ &-\lambda\sum_{m=1}^{n-1}\left(\frac{N_{m,1}}{m}-\widehat{\rho}_m\right)+o((\log n)^2)\ \text{a.s.}\end{aligned}$$

Then

$$\begin{aligned}&N_{n,1}-nv=M_{n,1}+\sum_{m=1}^{n}\psi_{m-1,1}-nv\\ =&M_{n,1}+M_{n,2}+\sum_{m=0}^{n-1}(\widehat{\boldsymbol{\theta}}_m-\boldsymbol{\theta})\left(\mathsf{E}\frac{\partial\pi_1}{\partial\boldsymbol{\theta}}\right)^T-\lambda\sum_{m=1}^{n-1}\left(\frac{N_{m,1}}{m}-\widehat{\rho}_m\right)+o((\log n)^2)\\ =&M_{n,1}+M_{n,2}+\lambda\sum_{m=1}^{n-1}\frac{M_{m,2}}{m}+(\lambda+1)\sum_{m=0}^{n-1}(\widehat{\boldsymbol{\theta}}_m-\boldsymbol{\theta})\left(\mathsf{E}\frac{\partial\pi_1}{\partial\boldsymbol{\theta}}\right)^T\\ &-\lambda\sum_{m=1}^{n-1}\left(\frac{N_{m,1}}{m}-v\right)+o((\log n)^2)\ \text{a.s.}\end{aligned}$$

$$
\begin{aligned}
=&M_{n,1}+\left(M_{n,2}+\lambda\sum_{m=1}^{n-1}\frac{M_{m,2}}{m}\right)+(1+o(1))\left((\lambda+1)\sum_{m=1}^{n-1}\frac{M_{m,3}}{m}\right)\\
&-\lambda\sum_{m=1}^{n-1}\left(\frac{N_{m,1}}{m}-v\right)+o(n^{1/2})\ \text{a.s.}
\end{aligned}
$$

On the other hand,

$$
\begin{gathered}
\mathsf{E}[\Delta M_{m,i}\Delta M_{m,j}|\mathscr{F}_{m-1}]=0,\ \ i\neq j,\\
\mathsf{E}[(\Delta M_{m,1})^2|\mathscr{F}_{m-1}]=\mathsf{E}[\psi_{m,1}(1-\psi_{m,1})|\mathscr{F}_{m-1}]\to\sigma_1^2\ \ \text{a.s.},\\
\mathsf{E}[(\Delta M_{m,2})^2|\mathscr{F}_{m-1}]=\mathsf{Var}[\pi_1(\boldsymbol{\theta},\boldsymbol{\xi}_m)]=\sigma_2^2,
\end{gathered}
$$

and

$$
\mathsf{E}[(\Delta M_{m,3})^2|\mathscr{F}_{m-1}]=\mathsf{E}\frac{\partial\pi}{\partial\boldsymbol{\theta}}\mathsf{E}[(\Delta\boldsymbol{Q}_m)^T\Delta\boldsymbol{Q}_m|\mathscr{F}_{m-1}]\left(\mathsf{E}\frac{\partial\pi}{\partial\boldsymbol{\theta}}\right)^T\to\sigma_3^2\ \ \text{a.s.}
$$

By applying the function central limit theorem (c.f., Corollary 3.1 of Hall and Heyde[3]), we have $n^{-1/2}\left(M_{[nt],1},M_{[nt],2},M_{[nt],3}\right)\xrightarrow{\mathscr{D}}\left(\sigma_1B_t^{(1)},\sigma_2B_t^{(2)},\sigma_3B_t^{(3)}\right)$, where $B_t^{(i)}$, $i=1,2,3$, are three independent standard Brownian motions. Then with the same argument as in Hu and Zhang[7], one can show that $n^{-1/2}(N_{[nt],1}-[nt]v)\xrightarrow{\mathscr{D}}G_t$, where

$$
G_t=\sigma_1t^{-\lambda}\int_0^t x^\lambda dB_x^{(1)}+\sigma_2B_t^{(2)}+(\lambda+1)\sigma_3t^{-\lambda}\int_0^t x^{\lambda-1}B_x^{(3)}dx
$$

is a solution of the equation

$$
G_t=\sigma_1B_t^{(1)}+\sigma_2\left(B_t^{(2)}+\lambda\int_0^t\frac{B_x^{(2)}}{x}dx\right)+(\lambda+1)\sigma_3\int_0^t\frac{B_x^{(3)}}{x}dx-\lambda\int_0^t\frac{G_x}{x}dx
$$

with $G_0=0$. It is easily checked that

$$
\mathsf{Var}(G_t)=t\left[\frac{\sigma_1^2}{1+2\lambda}+\sigma_2^2+\frac{2(\lambda+1)}{1+2\lambda}\sigma_3^2\right]=t\left[\frac{\sigma_1^2+\sigma_3^2}{1+2\lambda}+\sigma_2^2+\sigma_3^2\right].
$$

Hence $n^{1/2}(N_{n,1}/n-v)\xrightarrow{D}N(0,\sigma^2)$.

Proof of Corollary 3.1. It is sufficient to show the strong continency of MLE $\widehat{\boldsymbol{\theta}}_m$:

$$
\widehat{\boldsymbol{\theta}}_n\to\boldsymbol{\theta}.\tag{5.12}
$$

In fact, if (5.12) is proved, then by (5.4) and (5.6) we have $\widehat{\rho}_n\to v$ a.s. and $\frac{1}{n}\sum_{m=1}^n\widehat{\pi}_m\to v$ a.s. By (5.9) we will have $N_n/n\to v$ a.s. It follows that $\psi_{m,k}-\pi_k(\widehat{\boldsymbol{\theta}}_{m-1},\boldsymbol{\xi}_m)\to0$ a.s. by (5.11). The rest of proof is similar to that of Corollary 3.1 by Zhang, et al[21].

For (5.12), it suffices to show that, for any $\delta>0$ small enough, with probability one for m large enough, we have

$$
\log L_k(\boldsymbol{\theta}_k^*)<\log L_k(\boldsymbol{\theta}_k),\ \text{if}\ \|\boldsymbol{\theta}_k^*-\boldsymbol{\theta}_k\|=\delta.\tag{5.13}
$$

We consider the case $k=1$ only. The application of Taylor's theorem yields

$$
\begin{aligned}
&\frac{1}{m}\log L_1(\boldsymbol{\theta}_1^*)-\frac{1}{m}\log L_1(\boldsymbol{\theta}_1)\\
=&(\boldsymbol{\theta}_1^*-\boldsymbol{\theta}_1)\frac{1}{m}\frac{\partial\log L_1}{\partial\boldsymbol{\theta}_1}\Big|_{\boldsymbol{\theta}_1}+(\boldsymbol{\theta}_1^*-\boldsymbol{\theta}_1)\frac{1}{m}\frac{\partial^2\log L_1}{\partial\boldsymbol{\theta}_1^2}\Big|_{\boldsymbol{\theta}_1}(\boldsymbol{\theta}_1^*-\boldsymbol{\theta}_1)^T\\
&+(\boldsymbol{\theta}_1^*-\boldsymbol{\theta}_1)\Big\{\frac{1}{m}\int_0^1\Big[\frac{\partial^2\log L_1}{\partial\boldsymbol{\theta}_1^2}\Big|_{\boldsymbol{\theta}_1}^{\boldsymbol{\theta}_1+t(\boldsymbol{\theta}_1^*-\boldsymbol{\theta}_1)}\Big]dt\Big\}(\boldsymbol{\theta}_1^*-\boldsymbol{\theta}_1)^T.
\end{aligned}
$$

Write

$$f(a,b,\boldsymbol{z},\boldsymbol{\xi})=\frac{\pi_1(\boldsymbol{z},\boldsymbol{\xi})\left(\frac{b}{a}\right)^{\gamma}}{\pi_1(\boldsymbol{z},\boldsymbol{\xi})\left(\frac{b}{a}\right)^{\gamma}+(1-\pi_1(\boldsymbol{z},\boldsymbol{\xi}))\left(\frac{1-b}{1-a}\right)^{\gamma}}.$$

It is obvious that f is a continuous function of a, b and $\boldsymbol{z}$ for each given $\boldsymbol{\xi}$. By applying the law of large numbers for martingales, one can show that $\frac{1}{m}\frac{\partial \log L_1}{\partial \boldsymbol{\theta}_1}\Big|_{\boldsymbol{\theta}_1}\to \mathbf{0}$ a.s. and

$$\frac{\partial^2 \log L_1}{\partial \boldsymbol{\theta}_1^2}\Big|_{\boldsymbol{\theta}_1}=\sum_{j=2}^{m}\left(\mathsf{E}\big[f(a,b,\boldsymbol{z},\boldsymbol{\xi})\boldsymbol{I}_1(\boldsymbol{\theta}_1|\boldsymbol{\xi})\big]\right)\Big|_{a=\frac{N_{j-1}}{j-1},b=\widehat{\rho}_{j-1},\boldsymbol{z}=\widehat{\boldsymbol{\theta}}_{j-1}}+o(m)\ \text{ a.s.}$$

For details of the proof one can refer to Zhang, et al[21]. Further, it is obvious that

$$\limsup \widehat{\rho}_m\le \lim\frac{1}{m}\sum_{j=1}^{m}\sup_{\boldsymbol{\theta}}\pi_1(\boldsymbol{\theta},\boldsymbol{\xi}_j)=\mathsf{E}[\sup_{\boldsymbol{\theta}}\pi_1(\boldsymbol{\theta},\boldsymbol{\xi})]<1\ \text{ a.s.},$$

where the superior is taken over the parameter space. And similarly

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{m=1}^{n}\widehat{\pi}_m\le \mathsf{E}[\sup_{\boldsymbol{\theta}}\pi_1(\boldsymbol{\theta},\boldsymbol{\xi})]<1\ \text{ a.s.}$$

By (5.9),

$$\limsup N_{n,1}/n\le \mathsf{E}[\sup_{\boldsymbol{\theta}}\pi_1(\boldsymbol{\theta},\boldsymbol{\xi})]<1\ \text{ a.s.}$$

By considering $1-\widehat{\rho}_m$ and $n-N_{n,1}$ instead of $\widehat{\rho}_m$ and $N_{n,1}$ respectively, we have

$$\liminf \widehat{\rho}_m\ge \mathsf{E}[\inf_{\boldsymbol{\theta}}\pi_1(\boldsymbol{\theta},\boldsymbol{\xi})]>0\ \text{ and }\ \liminf N_{n,1}/n\ge \mathsf{E}[\inf_{\boldsymbol{\theta}}\pi_1(\boldsymbol{\theta},\boldsymbol{\xi})]>0\ \text{a.s.}$$

So we may assume that $\widehat{\rho}_m, N_{n,1}/n\in[\delta_0,1-\delta_0]$ for some $0<\delta_0<1$. On the other hand, it is obvious that $\boldsymbol{y}\mathsf{E}\big[f(a,b,\boldsymbol{z},\boldsymbol{\xi})\boldsymbol{I}_1(\boldsymbol{\theta}_1|\boldsymbol{\xi})\big]\boldsymbol{y}^T$ is a continuous function of $a,b,\boldsymbol{y},\boldsymbol{z}$, and is positive for all $0<a,b<1$, $\boldsymbol{y}\neq\mathbf{0}$ and all $\boldsymbol{z}$. It follows that there is a constant $c_0>0$ for which

$$\liminf_{j\to\infty}\min_{\boldsymbol{y}:\|\boldsymbol{y}\|=1}\left(\boldsymbol{y}\mathsf{E}\big[f(a,b,\boldsymbol{z},\boldsymbol{\xi})\boldsymbol{I}_1(\boldsymbol{\theta}_1|\boldsymbol{\xi})\big]\boldsymbol{y}^T\right)_{a=\frac{N_{j-1}}{j-1},b=\widehat{\rho}_{j-1},\boldsymbol{z}=\widehat{\boldsymbol{\theta}}_{j-1}}>c_0\ \text{a.s.}$$

So with probability one for m large enough it holds that

$$\begin{aligned}&\frac{1}{m}\log L_1(\boldsymbol{\theta}_1^*)-\frac{1}{m}\log L_1(\boldsymbol{\theta}_1)\\ \le&-\|\boldsymbol{\theta}_1^*-\boldsymbol{\theta}_1\|^2\Big\{\frac{1}{m}\sum_{j=2}^{m}\min_{\boldsymbol{y}:\|\boldsymbol{y}\|=1}\left(\boldsymbol{y}\mathsf{E}\big[f(a,b,\boldsymbol{z},\boldsymbol{\xi})\boldsymbol{I}_1(\boldsymbol{\theta}_1|\boldsymbol{\xi})\big]\boldsymbol{y}^T\right)_{a=\frac{N_{j-1}}{j-1},b=\widehat{\rho}_{j-1},\boldsymbol{z}=\widehat{\boldsymbol{\theta}}_{j-1}}\Big\}\\ &+\|\boldsymbol{\theta}_1^*-\boldsymbol{\theta}_1\|^2H(\|\boldsymbol{\theta}_1^*-\boldsymbol{\theta}_1\|)+o(1)\\ \le&-c_0\delta^2+\delta^2H(\delta)+o(1)<0\ \text{uniformly in}\ \boldsymbol{\theta}_1^*\ \text{with}\ \|\boldsymbol{\theta}_1^*-\boldsymbol{\theta}_1\|=\delta\end{aligned}$$

when δ is small enough. (5.13) is thus proved.

References

1 Bai Z D, Hu F, Rosenberger W F. Asymptotic properties of adaptive designs with delayed response, Ann Statist, 2002, 30: 122-139.

2 Eisele J, Woodroofe M. Central limit theorems for doubly adaptive biased coin designs, Ann Statist, 1995, 23: 234-254.

3 Hall P, Heyde C C. Martingale Limit Theory and its Applications, London: Academic Press, 1980.

4 Hu F, Rosenberger W F. Evaluating response-adaptive randomization procedures for treatment comparisons, J Amer Statist Assoc, 2003, 98: 671-678.

5 Hu F, Rosenberger W F. The Theory of Response-Adaptive Randomization in Clinical Trials, New York: John Wiley and Sons, 2006.

6 Hu F, Rosenberger W F, Zhang, L X. Asymptotically best response-adaptive randomization procedures, J Statist Plan Inference, 2006, 136: 1911-1922.

7 Hu F, Zhang L X. Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials, Ann Statist, 2004, 32: 268-301.

8 Hu F, and Zhang L X. Asymptotic normality of urn models for clinical trials with delayed response, Bernoulli, 2004, 10(3): 447-463.

9 Hu F, Zhang L X, He X. Efficient randomized adaptive designs, Ann Statist, To appear.

10 Pocock S J, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial, Biometrics, 1975, 31: 103-115.

11 Robbins H. Some aspects of the sequential design of experiments, Bull Amer Math Soc, 1952, 58: 527-535.

12 Rosenberger W F, Lachin J M. Randomization in Clinical Trials Theory and Practice, New York: John Wiley and Sons, 2002.

13 Rosenberger W F, Stallard N, Ivanova A, et al. Optimal adaptive designs for binary response trials, Biometrics, 2001, 57: 909-913.

14 Rosenberger W F, Vidyashankar A N, Agarwal D K. Covariate-adjusted response-adaptive designs for binary response, J Biopharm Statist, 2001, 11: 227-236.

15 Stout W F. Almost Sure Convergence, New York: Academic Press, 1974.

16 Taves D R. Minimization: a new method of assigning patients to treatment and control groups, Clin Pharm Ther, 1974, 15: 443-453.

17 Thompson W R. On the likelihood that one unknown probability exceeds another in the view of the evidence of the two samples, Biometrika, 1933, 25: 275-294.

18 Tymofyeyev Y, Rosenberger W F, Hu F. Implementing optimal allocation in sequential binary response experiments, J Amer Statist Assoc, 2007, 102: 224-234.

19 Zelen M. The randomization and stratification of patients to clinical trials, Journal of Chronic Diseases, 1974, 27: 365-375.

20 Zhang L X, Chan W S, Cheung S H, et al. A generalized urn model for clinical trials with delayed responses, Statistica Sinica, 2006, 17: 387-409.

21 Zhang L X, Hu F, Cheung S H, et al. Asymptotic properties of covariate-adjusted adaptive designs, Ann Statist, 2007, 35: 1166-1182.

1 Institute of Statist. and Dept. of Math., Zhejiang Univ., Hongzhou 310027, China

2 Dept. of Statist., Univ. of Virginia, Charlottesville, VA 22904-4135