Ann Inst Stat Math (2018) 70:353–371
https://doi.org/10.1007/s10463-016-0596-3

Fixed-width confidence interval for covariate-adjusted response-adaptive designs

Uttam Bandyopadhyay[1] · **Atanu Biswas**[2]

Received: 12 January 2016 / Revised: 29 July 2016 / Published online: 21 January 2017

Abstract In this paper, we obtain fixed-width confidence interval for covariate-adjusted response-adaptive designs. Specifically, we consider logistic regression model and the normal regression model for binary and continuous responses, respectively, both in the situations for presence and absence of treatment–covariate interactions. Simulation study and real-data analysis are carried out.

Keywords Coverage probability · Fixed-width confidence interval · Logistic regression model · Normal regression model · Stopping time · Treatment–covariate interaction

1 Introduction

Fixed-width confidence interval estimation is an important problem in sequential analysis literature. See the booklength discussion by Ghosh et al. (1997) in this context. However, the problem becomes complicated if the data come from some dependent process. Response-adaptive design (see Atkinson and Biswas 2014) is such a set up where the allocation among two (or more) competing treatments depends on the sigma field generated by all the previous allocation-and-response history. Towards this direction, Bandyopadhyay and Biswas (2015) worked on fixed-width confidence interval

✉ Atanu Biswas
atanu@isical.ac.in

Uttam Bandyopadhyay
ubandyopadhyay08@gmail.com

[1] Department of Statistics, University of Calcutta, 35 Ballygunge Circular Road, Kolkata 700 019, India

[2] Applied Statistics Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata 700 108, India

estimation of a real-valued function measuring treatment difference. However, so far our knowledge goes, not much work on fixed-width confidence interval estimation in the covariate-adjusted response-adaptive (CARA) set up has been done; possibly the only exception is the work of Chang and Park (2013). The present paper aims to fulfil that gap.

In this paper, we consider a typical response-adaptive set up in phase III clinical trials with two treatments (indexed by A and B), where the objective is to skew the allocation in favor of the better treatment (see, for example, Atkinson and Biswas 2014) so that eventually a larger number of allocation is done by the better treatment in a sequential allocation procedure. So the subjects under consideration are patients of the trial. Each subject receives one of the two treatments, A and B, using some randomized allocation technique. More the treatment difference, the allocation should be more skewed in favor of the better treatment. Early studies on response-adaptive designs focused on finding the allocation probabilities based on allocation and response history of the previously allocated patients (see, e.g., the randomized play-the-winner rule of Wei and Durham 1978). However, a success from a patient with favorable treatment condition should have different weight than a success from a patient with adverse condition. Later, the covariate history of the previously allocated patients was also used in finding the allocation probabilities (see, e.g., the covariate-adjusted randomized play-the-winner rule of Bandyopadhyay and Biswas 1999), the normal cumulative distribution function-type link function-based design for continuous responses of Bandyopadhyay and Biswas (2001). Then, the CARA designs came into play with the provision of using the covariate information of the current patient in finding the allocation probabilities (see, e.g., Biswas and Coad 2005). Here, for example, the allocation probability of an old patient differs from that of a young patient.

Let δ_k be the allocation indicator of the kth treatment, $k = A, B$, where $\delta_A (= 1 - \delta_B) = 1$ or 0 according as a subject receives treatment A or treatment B. Let X_k be the potential response to treatment k, $k = A, B$. Clearly, for subject, either X_A or X_B is observed depending on the treatment given to this subject. Then, the observed response of a subject is $Y = \delta_A X_A + \delta_B X_B$.

The response is usually influenced by some covariate Z, which is real or vector valued. We assume that, given Z, X_k follows the distribution according to the density or mass function (cf. McCullagh and Nelder 1989)

$$f_k(x|\theta_k, Z) = \exp\left[\frac{1}{\phi_k}(x\mu_k(Z) - a_k(\mu_k(Z))) + b(x, \phi_k)\right],$$

where

(i) $\mu_k(Z) = h_k(\theta_k^\top Z)$ is the link function,
(ii) $\theta_k = (\alpha_k, \beta_k^\top)^\top$, in which α_k (real valued) is the effect due to treatment k and β_k (may be vector valued) is the corresponding regression coefficient, and $Z = (1, Z_1^\top)^\top$,
(iii) $\phi_k, a_k(\cdot)$ and $b_k(\cdot, \cdot)$ are some specific functions.

We assume that $\theta_k \in \Omega_k$, a bounded subset of $\mathbb{R}^r$ $(r \geq 2)$, $k = A, B$. We further assume that Z is discrete valued and is uniformly bounded. This is justified as, in real clinical trials, covariates are the measurements on some clinical characteristics of human beings (e.g., blood pressure, blood sugar, age, etc.), and thus they vary within certain finite limits. In other words, we can find $M \in (0, \infty)$ such that $0 < ||Z|| < M$. We consider $\phi_k = 1$, $k = A, B$. Specifically, in this paper we consider

(a) logistic regression model in which X_k, given Z, follows Bernoulli $(p_k(Z))$ distribution, where we assume that

$$\text{logit}(p_k(Z)) = \alpha_k + \beta_k^\top Z_1, \quad k = A, B,$$

and

(b) normal regression model in which X_k, given Z, follows $N(\mu_k(Z), 1)$ distribution, where we assume that

$$\mu_k(Z) = \alpha_k + \beta_k^\top Z_1, \quad k = A, B.$$

In this paper, we provide fixed-width confidence interval (CI) of $\alpha_A - \alpha_B$ using the data obtained under covariate-adjusted response-adaptive (CARA) design. We carry out the analysis in the situations for the presence and absence of treatment–covariate interactions. Note that, $\alpha_A - \alpha_B$ can be interpreted as the *treatment difference* at $Z = 0$ under no treatment–covariate interaction (i.e., under $\beta_A = \beta_B$). We can as well carry out our fixed-width confidence interval estimation for other parametric functions in a similar way.

The rest of the paper is organized as follows. In Sect. 2, we discuss some preliminary results regarding the CARA design. The fixed-width confidence-interval and the related results are obtained in Sect. 3. Numerical illustration through simulations and redesigning real data is done in Sect. 4. Section 5 concludes, followed by technical details in "Appendix".

2 CARA design and related results

2.1 CARA design

CARA design has been studied by different authors. There has been growing interest in this context since the work of Rosenberger et al. (2001) was published in the context of response-adaptive designs for binary responses. Several references are available in Atkinson and Biswas (2014) and Biswas and Bhattacharya (2016). Zhang et al. (2007) studied the analytical properties of CARA designs under a general linear model set up.

The CARA design consists of setting the allocation function $\pi_A(\theta, Z)$ $(= 1 - \pi_B(\theta, Z))$ such that $\pi_A(\theta, Z) = P(\delta_A = 1|Z)$, where $0 < \pi_A(\theta, Z) < 1$ and $\theta = (\theta_A^\top, \theta_B^\top)^\top$. We assume that

(I) $\pi_A(\theta, Z)$ is continuous in θ for each covariate value of Z.

(II) $\rho_A = \rho_A(\theta) = E\,[\pi_A(\theta, Z)] \in (0, 1)$.

For example, the allocation function

$$\pi_A(\theta, Z) = P_\theta(X_B < X_A|Z) + \frac{1}{2}P_\theta(X_A = X_B|Z),$$

where X_k is distributed according to the probability model (a) or (b), satisfies (I) and (II) and is used in the present study.

As θ is unknown, CARA design starts with n_0 subjects to each treatment so as to get a reasonable initial estimate of the parameter. Let n be the target number of subjects to be examined in the trial. The corresponding observations on (Z, δ_A, Y) are $\{(Z_i, \delta_{Ai}, Y_i), i \le n\}$, in which the allocation indicators $\{\delta_{Ai}, 2n_0 \le i \le n\}$ are distributed according to the probability model

$$P\left[\delta_{A,i+1}=1|\mathscr{F}_i, Z_{i+1}\right]=1-P\left[\delta_{B,i+1}=1|\mathscr{F}_i, Z_{i+1}\right]=\pi_A(\widehat{\theta}_i, Z_{i+1}), \quad i \ge 2n_0,$$

for allocating the $(i + 1)$st subject to treatment A based on $\mathscr{F}_i$, the σ-field of the history $\{(Z_j, \delta_{Aj}, Y_j), j \le i\}$ and the current covariate Z_{i+1}, where $\widehat{\theta}_i = (\widehat{\theta}_{Ai}^\top, \widehat{\theta}_{Bi}^\top)^\top$ is a consistent estimator of θ based on $\mathscr{F}_i$.

2.2 Estimation

For estimation of θ, we use the maximum likelihood (ML) method where the log-likelihood function

$$L_n(\theta) = \sum_{i=1}^{n} \sum_{k=A,B} \delta_{ki}\,[(Y_i\mu_k(Z_i) - a_k(\mu_k(Z_i))) + b_k(Y_i)]$$

is maximized. In the presence of treatment–covariate interactions, the ML estimator, $\widehat{\theta}_n = (\widehat{\theta}_{An}^\top, \widehat{\theta}_{Bn}^\top)^\top$, of θ is the solution of the likelihood equations

$$\sum_{i=1}^{n} \delta_{ki}(Y_i - a_k'(\mu_k(Z_i)))Z_i^\top = 0, \tag{1}$$

for $k = A, B$. The solution is unique and the likelihood function attains absolute maximum at the respective unique solution. For model (a), the above equations reduce to

$$\sum_{i=1}^{n} \delta_{ki} Y_i Z_i = \sum_{i=1}^{n} \delta_{ki} Z_i Z_i^\top p_k(Z_i), \quad k = A, B, \tag{2}$$

whereas, for model (b), the equations become

$$\sum_{i=1}^{n} \delta_{ki} Y_i Z_i = \sum_{i=1}^{n} \delta_{ki} Z_i Z_i^\top \theta_k, \qquad k = A, B. \tag{3}$$

In the absence of treatment–covariate interactions, we have $\mu_k(Z) = \mu_{kc}(Z) = h_k(\alpha_k + \beta^\top Z_1), k = A, B$, so that $\theta = \theta_c = (\alpha_A, \alpha_B, \beta^\top)^\top \in \Omega_c$, a bounded subset of $\mathbb{R}^{r+1}$. Here, β represents the common regression coefficient. As in (1), we have on setting $[a'_k(\mu_{kc}(Z_i))]^\top = (a'_{1k}(\mu_{kc}(Z_i)), [a'_{2k}(\mu_{kc}(Z_i))]^\top)$,

$$\begin{aligned} &\sum_{i=1}^{n} \delta_{ki}(Y_i - a'_{1k}(\mu_{kc}(Z_i))) = 0, \qquad k = A, B, \\ &\sum_{i=1}^{n} \sum_{k=A,B} \delta_{ki}(Y_i - a'_{2k}(\mu_{kc}(Z_i)))Z_{1i}^\top = 0, \end{aligned} \tag{4}$$

which, for model (a), becomes

$$\sum_{i=1}^{n} \delta_{ki} Y_i = \sum_{i=1}^{n} \delta_{ki} p_{kc}(Z_i), \qquad k = A, B, \tag{5}$$

$$\sum_{i=1}^{n} \sum_{k=A,B} \delta_{ki} Y_i Z_{1i} = \sum_{i=1}^{n} \sum_{k=A,B} \delta_{ki} Z_{1i} Z_{1i}^\top p_{kc}(Z_i), \tag{6}$$

and, for model (b), becomes

$$\sum_{i=1}^{n} \delta_{ki} Y_i = \sum_{i=1}^{n} \delta_{ki}(\alpha_k + \beta^\top Z_{1i}), \qquad k = A, B, \tag{7}$$

$$\sum_{i=1}^{n} \sum_{k=A,B} \delta_{ki} Y_i Z_i = \sum_{i=1}^{n} \sum_{k=A,B} \delta_{ki}(\alpha_k + \beta^\top Z_{1i}) Z_{1i}^\top, \tag{8}$$

where $p_{kc}(Z) = \exp(\alpha_k + \beta^\top Z_1)/[1 + \exp(\alpha_k + \beta^\top Z_1)]$. The solution of the above equations is denoted by $\theta^*_{cn} = (\alpha^*_{An}, \alpha^*_{Bn}, \beta^{*\top}_n)^\top$.

2.3 Related asymptotic results

Let $n_k = \sum_{i=1}^{n} \delta_{ki}$ be the number of subjects treated by treatment k, and $n_k(Z)$ be that for a given covariate value Z ($k = A, B$) when n responses are obtained. Then, in the presence of treatment–covariate interactions, it follows that (cf. Zhang et al. 2007), as $n \to \infty$,

$$\frac{n_k}{n} \to \rho_k \qquad \text{and} \qquad \frac{n_k(Z)}{n} \to \pi_k(\theta, Z)$$

almost surely, $k = A, B$. Moreover, $\widehat{\theta}_n$ is strongly consistent for θ. Next, consider the following matrix:

$$I(\theta|Z) = -E\left[\sum_{k=A,B} \pi_k(\theta, Z)\frac{\mathrm{d}^2 f_k(X_k|\theta_k, Z)}{\mathrm{d}\theta^2}\bigg| Z\right],$$

which, by a routine computation, can be simplified to

$$Diag\,(\pi_A(\theta, Z)\tau_A(Z), \pi_B(\theta, Z)\tau_B(Z)) \otimes (ZZ^\top)$$

with $\tau_k(\theta_k, Z) = \tau_k(Z) = p_k(Z)(1-p_k(Z))$ and 1 for models (a) and (b), respectively, $k = A, B$. Now, as in Theorem 2.1 of Zhang et al. (2007), we state the following result.

Result 1 *As* $n \to \infty$,

$$\sqrt{n}(\widehat{\theta}_n - \theta) \to N_{2r}(0, \Sigma)$$

in distribution, where

$$\Sigma = \Sigma(\theta) = \{E\,[I(\theta|Z)]\}^{-1} = \mathrm{Diag}(\Sigma_A, \Sigma_B)$$

is positive-definite with $\Sigma_k = [E(\pi_k(\theta, Z)\tau_k(Z)ZZ^\top)]^{-1}$, $k = A, B$.

Let $\sigma^2 = \sigma^2(\theta)$ (> 0) be the sum of the first principal diagonal elements from the matrices Σ_k, $k = A, B$. Then, given d (> 0) and $\alpha \in (0, 1)$, we can find the positive integer

$$\nu = \nu(\theta) = \min\left\{m \geq 2n_0 : m \geq \frac{u^2\sigma^2(\theta)}{d^2}\right\},$$

where $P(|N(0, 1)| \leq u) = 1 - \alpha$ and n_0 $(\geq r)$ is a pre-fixed positive integer. Now, writing $\widehat{\theta}_{k\nu} = (\widehat{\alpha}_{k\nu}, \widehat{\beta}_{k\nu}^\top)^\top$, Result 1 implies that, as $d \downarrow 0$,

$$\sqrt{\nu}\,(\widehat{\alpha}_{A\nu} - \widehat{\alpha}_{B\nu} - \alpha_A + \alpha_B) \to N(0, \sigma^2(\theta))$$

in distribution. Hence, we get

$$\lim_{d\downarrow 0} P_\theta\,\{|\widehat{\alpha}_{A\nu} - \widehat{\alpha}_{B\nu} - \alpha_A + \alpha_B| \leq d\} = 1 - \alpha \qquad \text{for all } \theta \in \Omega_A \times \Omega_B. \tag{9}$$

In the absence of treatment–covariate interactions, the parameter θ becomes $\theta_c \in \Omega_c$, a bounded subset of $\mathbb{R}^{r+1}$ and, thus, using θ_{cn}^* in place of $\widehat{\theta}_n$, we can obtain similar

results to those which are found under the presence of treatment–covariate interactions. However, for getting the asymptotic distribution of θ^*_{cn}, we need to consider the matrix

$$I_c(\theta_c|Z) = -E\left[\sum_{k=A,B} \pi_k(\theta_c, Z)\frac{\mathrm{d}^2 f_k(X_k|\theta_{kc}, Z)}{\mathrm{d}\theta_c^2}\middle| Z\right],$$

where $\theta_{kc} = (\alpha_k, \beta^\top)^\top$, $k = A, B$. Hence, unlike the previous one, we get

$$I_c(\theta_c|Z) = \begin{pmatrix} \pi_A(\theta_c, Z)\tau_{Ac}(Z) & 0 & \pi_A(\theta_c, Z)\tau_{Ac}(Z)Z_1^\top \\ 0 & \pi_B(\theta_c, Z)\tau_{Bc}(Z) & \pi_B(\theta_c, Z)\tau_{Bc}(Z)Z_1^\top \\ \pi_A(\theta_c, Z)\tau_{Ac}(Z)Z_1 & \pi_B(\theta_c, Z)\tau_{Bc}(Z)Z_1 & \sum_{k=A,B}\pi_k(\theta_c, Z)\tau_{kc}(Z)Z_1Z_1^\top \end{pmatrix}$$

with $\tau_{kc}(\theta_{kc}, Z) = \tau_{kc}(Z) = p_{kc}(Z)(1 - p_{kc}(Z))$ and 1 for models (a) and (b), respectively, $k = A, B$. Hence, as in Result 1, we get the following result.

Result 2 *As* $n \to \infty$,

$$\sqrt{n}(\theta^*_{cn} - \theta_c) \to N_{r+1}(0, \Sigma_c)$$

in distribution, where $\Sigma_c = \Sigma(\theta_c) = \{E[I_c(\theta_c|Z)]\}^{-1}$ *is positive definite.*

Let $\sigma_c^2 = \sigma_c^2(\theta_c) > 0$ be the sum of the first two principal diagonal elements of Σ_c. Then, setting

$$\nu_c = \nu_c(\theta_c) = \min\left\{m \geq 2n_0 : m \geq \frac{u^2\sigma_c^2}{d^2}\right\}$$

with $2n_0 \geq r + 1$, we have $\theta^*_{c\nu_c} = (\alpha^*_{A\nu_c}, \alpha^*_{B\nu_c}, \beta^{*\top}_{\nu_c})^\top$, and hence, by Result 2,

$$\sqrt{\nu_c}\left(\alpha^*_{A\nu_c} - \alpha^*_{B\nu_c} - \alpha_A + \alpha_B\right) \to N(0, \sigma_c^2)$$

in distribution as $d \downarrow 0$. This implies, as in (9),

$$\lim_{d\downarrow 0} P_\theta\left\{\left|\widehat{\alpha}^*_{A\nu_c} - \widehat{\alpha}^*_{B\nu_c} - \alpha_A + \alpha_B\right| \leq d\right\} = 1 - \alpha \qquad \text{for all } \theta \in \Omega_c.$$

3 Fixed-width confidence interval

In practice, σ^2 and σ_c^2 are unknown as θ and θ_c are unknown. Estimating θ sequentially from the likelihood equations (see (2) and (3)), a strongly consistent estimator of

$E[I(\theta|Z)]$ (see Note A.1 of "Appendix") can be obtained by the following sequences of matrices:

$$\widehat{I}_m = \text{Diag}\left\{\frac{1}{m}\sum_{i=1}^{m}\delta_{ki}\tau_k\left(\widehat{\theta}_{km}, Z_i\right)Z_i Z_i^{\top},\ k = A, B\right\}, \qquad m \geq 2n_0.$$

Similarly, estimating θ_c from the Eqs. (5)–(8), a strongly consistent estimator of $E[I_c(\theta_c|Z)]$ is obtained from

$$I^*_{cm} = \begin{pmatrix} \frac{1}{m}\sum_{i=1}^{m}\delta_{Ai}\tau_{Ac}\left(\theta^*_{Acm}, Z_i\right) & 0 & \frac{1}{m}\sum_{i=1}^{m}\delta_{Ai}\tau_{Ac}\left(\theta^*_{Acm}, Z_i\right)Z_{1i}^{\top} \\ 0 & \frac{1}{m}\sum_{i=1}^{m}\delta_{Bi}\tau_{Bc}\left(\theta^*_{Bcm}, Z_i\right) & \frac{1}{m}\sum_{i=1}^{m}\delta_{Bi}\tau_{Bc}\left(\theta^*_{Bcm}, Z_i\right)Z_{1i}^{\top} \\ \frac{1}{m}\sum_{i=1}^{m}\delta_{Ai}\tau_{Ac}\left(\theta^*_{Acm}, Z_i\right)Z_{1i} & \frac{1}{m}\sum_{i=1}^{m}\delta_{Bi}\tau_{Bc}\left(\theta^*_{Bcm}, Z_i\right)Z_{1i} & \frac{1}{m}\sum_{i=1}^{m}\sum_{k=A,B}\delta_{ki}\tau_{kc}\left(\theta^*_{kcm}, Z_i\right)Z_{1i}Z_{1i}^{\top} \end{pmatrix},$$

for $m \geq 2n_0$.

Now, to estimate σ^2 and σ_c^2, we consider two continuous functions, g and g_c, such that $g(\Sigma) = \sigma^2$ and $g_c(\Sigma_c) = \sigma_c^2$. Then, $\widehat{\sigma}_m^2 = g(\widehat{\Sigma}_m)$ and $\sigma^{*2}_{cm} = g_c(\Sigma^*_{cm})$, where $\widehat{\Sigma}_m = \widehat{I}_m^{-1}$ and $\widehat{\Sigma}_{cm} = I_m^{*-1}$, will provide strongly consistent estimators of σ^2 and σ_c^2, respectively. Now, we define the stopping rules:

$$N = N(d) = \min\left\{m \geq 2n_0 : m \geq \frac{u^2\widehat{\sigma}_m^2}{d^2}\right\} \tag{10}$$

and

$$N_c = N_c(d) = \min\left\{m \geq 2n_0 : m \geq \frac{u^2\sigma^{*2}_{cm}}{d^2}\right\}.$$

Then, we prove the following results.

Result 3 *For every $d > 0$, $P_\theta(N < \infty) = 1$ for all $\theta \in \Omega_A \times \Omega_B$.*

Proof As $\widehat{\sigma}_m^2$ is strongly consistent for σ^2 (> 0), we have

$$P_\theta(N > m) \leq P_\theta\left(m < \frac{u^2}{d^2}\widehat{\sigma}_m^2\right) \to P(\phi) = 0 \qquad \text{for all } \theta \in \Omega_A \times \Omega_B$$

as $m \to \infty$. □

Note By the standard argument as above, we have for each $d > 0$,

$$P_\theta(N < \infty) = 1 \qquad \text{for all } \theta \in \Omega_c.$$

Result 4 *As $d \downarrow 0$, we have*

$$\frac{N}{v} \to 1 \quad \textit{almost surely}.$$

Proof Note that, by definition (10), $N = N(d) \uparrow \infty$ as $d \downarrow 0$. Hence, as $d \downarrow 0$, $N \to \infty$ almost surely. This gives, as $d \downarrow 0$,

$$N_k = \sum_{i=1}^{N} \delta_{ki} \to \infty \quad \text{almost surely, } k = A, B.$$

Further, let $\widehat{\theta}_N$ be that value of θ for which, as in (1), we have

$$\sum_{i=1}^{N} \delta_{ki}(Y_i - a_k'(\mu_k(Z_i)))Z_i^{\top} = 0, \qquad k = A, B. \tag{11}$$

Then, as $d \downarrow 0$, almost sure convergence of $\widehat{\theta}_v$ to θ implies that of $\widehat{\theta}_N$ to θ. Moreover, martingale convergence theorem, together with the convergence of $\widehat{\theta}_N$ to θ, implies that

$$\frac{N_k}{N} \to \rho_k \quad \text{almost surely as } d \downarrow 0, \qquad k = A, B,$$

which implies $\widehat{\Sigma}_N$ converges almost surely to Σ.

Now, combining all these, we ultimately get, as $d \downarrow 0$,

$$g(\widehat{\Sigma}_N) = \widehat{\sigma}_N^2 \to \sigma^2 \quad \text{almost surely.}$$

Hence, using the definition of N, we get

$$\frac{Nd^2}{u^2} \to \sigma^2 \quad \text{almost surely as } d \downarrow 0,$$

which using the fact that

$$\frac{vd^2}{u^2} \to \sigma^2 \quad \text{as } d \downarrow 0$$

implies the required result. □

Note 1 *Proceeding along the same line as above, we can show that*

$$\frac{N_c}{v_c} \to 1 \quad \textit{almost surely as } d \downarrow 0.$$

Here, as in (11)*, we find* $\theta^*_{N_c}$ *from the equations under* (4) *setting* $n = N_c$*. Then, as above,* $\theta^*_{N_c}$ *is strongly consistent for* θ_c*.*

Note 2 *As* $d \downarrow 0$*,*

$$\frac{N_k(Z)}{N} \to \pi_k(\theta, Z) \quad \textit{and} \quad \frac{N_{ck}(Z)}{N_c} \to \pi_k(\theta_c, Z) \quad \textit{almost surely}, \qquad k = A, B,$$

where $N_k(Z)$ ($N_{ck}(Z)$) is the number of subjects receiving treatment k for a given covariate value Z when treatment–covariate interactions are present (absent).

Next, as in Zhang et al. (2007), it is possible to set r-component vector $U_k(Y, Z)$ such that

$$\widehat{\theta}_{kv} = \theta_k + \frac{1}{v}\sum_{i=1}^{v} \delta_{ki} U_k(Y_i, Z_i) + \epsilon_{kd}, \tag{12}$$

where

$$E\,[U_k(X_k, Z)|Z] = 0, \qquad E[||U_k(X_k, Z)||^2] < \infty,$$

and, as $d \downarrow 0$,

$$\frac{1}{\sqrt{v}}||\epsilon_{kd}|| \to 0 \quad \textit{almost surely}, \qquad k = A, B.$$

In particular, we assume that for the vector Z, $E||Z||^2 < \infty$, which holds since Z is uniformly bounded, and the matrix $E[ZZ^\top]$ is positive definite. Then, for model (a), (12) holds when

$$\begin{aligned} U_k(X_k, Z) &= I_{Zk}^{-1} Z(X_k - p_k(Z)) \quad \textit{with} \quad I_{Zk} \\ &= E[\pi_k(\theta, Z) p_k(Z)(1 - p_k(Z)) ZZ^\top], \qquad k = A, B, \end{aligned}$$

whereas, for model (b), (12) holds when

$$U_k(X_k, Z) = I_{Zk}^{-1} Z(X_k - Z^\top \theta_k) \quad \textit{with} \quad I_{Zk} = E[\pi_k(\theta, Z) ZZ^\top], \qquad k = A, B.$$

Here, we note that the expression of $Diag(I_{ZA}, I_{ZB})$ is same as that of $E[I(\theta|Z)]$. Let us rewrite $\widehat{\theta}_N = (\theta_{AN}^\top, \widehat{\theta}_{BN}^\top)^\top$ with $\widehat{\theta}_{kN} = (\widehat{\alpha}_{kN}, \widehat{\beta}_{kN}^\top)^\top$, $k = A, B$. Then, we get the following result.

Result 5 *As $d \downarrow 0$,*

$$P_\theta\,[|\widehat{\alpha}_{AN} - \widehat{\alpha}_{BN} - \alpha_A + \alpha_B| \le d] \to 1 - \alpha \qquad \textit{for all} \quad \theta \in \Omega_A \times \Omega_B.$$

Proof As in (9), the result will be proved if, for $d \downarrow 0$,

$$\sqrt{N}(\widehat{\theta}_{kN} - \theta_k) \to N_r\,(0, \Sigma_k)$$

in distribution, for $k = A, B$.

Writing

$$S_{kn} = \sum_{i=1}^{n} \delta_{ki} U_k(Y_i, Z_i), \qquad k = A, B,$$

we observe that $\{S_{kn},\ \mathscr{F}_n\}$ is a martingale and, hence, $\{||S_{kn} - S_{kv}||;\ \mathscr{F}_n,\ n \geq v\}$ is submartingale, $k = A, B$. By Result A.2, it follows that

$$\frac{1}{\sqrt{v}} S_{kv} \to N_r\,(0, \Sigma_k) \tag{13}$$

in distribution as $d \downarrow 0$, $k = A, B$. Moreover, using Kolmogorov maximal inequality for non-negative submartingales (see, for example, Sen 1981, p. 13), it follows by some standard arguments that for each $\epsilon > 0$ and $\eta > 0$, there are $d^* = d(\epsilon, \eta)$ with $n^* = n(d^*)$ and $\delta = \delta(\epsilon, \eta) > 0$ such that

$$P_\theta \left\{ \max_{|n-v| \leq \delta v} ||S_{kn} - S_{kv}|| < \epsilon \sqrt{v} \right\} < \eta \tag{14}$$

for all $v \geq n^*$. Hence, combining (12) with (14), we get

$$P_\theta \left[\max_{|n-v| \leq \delta v} ||\widehat{\theta}_{kn} - \widehat{\theta}_{kv}|| < \epsilon / \sqrt{v} \right] < \eta \tag{15}$$

for all $v \geq n^*$, $k = A, B$. Thus, Anscombe (1952) condition holds for such ML estimators under CARA design. Now, setting

$$\widehat{\theta}_{kN} = \theta_k + (\widehat{\theta}_{kv} - \theta_k) + (\widehat{\theta}_{kN} - \widehat{\theta}_{kv})$$

and using (12), (13) and (15), the required result follows. □

Next, we consider the case of no interaction between treatment and covariates in which, as in (12), it is also possible to find $(r + 1)$-component vectors $U_{kc}(Y, Z)$, $k = A, B$, such that

$$\theta^*_{v_c} = \theta_c + \frac{1}{v_c} \sum_{i=1}^{v_c} \sum_{k=A,B} \delta_{ki} U_{kc}(Y_i, Z_i) + \epsilon_{cd},$$

where

$$E\,[U_{kc}(X_k, Z)|Z] = 0, \qquad E[||U_{kc}(X_k, Z)||^2] < \infty$$

and, as $d \downarrow 0$,

$$\frac{1}{\sqrt{v_c}} ||\epsilon_{cd}|| \to 0 \quad \text{almost surely}, \qquad k = A, B.$$

In particular, for model (a), we have (setting $\phi_k = 1, k = A, B$)

$$U_{Ac}(X_A, Z) = I_{Zc}^{-1} \begin{pmatrix} X_A - p_{Ac}(Z) \\ 0 \\ Z_1\,(X_A - p_{Ac}(Z)) \end{pmatrix} \qquad \text{and}$$

$$U_{Bc}(X_B, Z) = I_{Zc}^{-1} \begin{pmatrix} 0 \\ X_B - p_{Bc}(Z) \\ Z_1 \left(X_B - p_{Bc}(Z)\right) \end{pmatrix},$$

where

$$I_{Zc} = E \begin{pmatrix} \pi_A(\theta_c, Z) V_{Ac}(Z) & 0 & \pi_A(\theta_c, Z) V_{Ac}(Z) Z_1^\top \\ 0 & \pi_B(\theta_c, Z) V_{Bc}(Z) & \pi_B(\theta_c, Z) V_{Bc}(Z) Z_1^\top \\ \pi_A(\theta_c, Z) V_{Ac}(Z) Z_1 & \pi_B(\theta_c, Z) V_{Bc}(Z) Z_1 & (Z_1 Z_1^\top) \sum_{k=A,B} \pi_k(\theta_c, Z) V_{kc}(Z) \end{pmatrix},$$

with $V_{kc}(Z) = p_{kc}(Z)\,(1 - p_{kc}(Z))$, $k = A, B$. Similarly, for model (b), we have

$$U_{Ac}(X_A, Z) = I_{Zc}^{-1} \begin{pmatrix} X_A - Z^\top \theta_{Ac} \\ 0 \\ Z_1 \left(X_A - Z^\top \theta_{Ac}\right) \end{pmatrix} \quad \text{and}$$

$$U_{Bc}(X_B, Z) = I_{Zc}^{-1} \begin{pmatrix} 0 \\ X_B - Z^\top \theta_{Bc} \\ Z_1 \left(X_B - Z^\top \theta_{Bc}\right) \end{pmatrix}$$

with

$$I_{Zc} = E \begin{pmatrix} \pi_A(\theta_c, Z) & 0 & \pi_A(\theta_c, Z) Z_1^\top \\ 0 & \pi_B(\theta_c, Z) & \pi_B(\theta_c, Z) Z_1^\top \\ \pi_A(\theta_c, Z) Z_1 & \pi_B(\theta_c, Z) Z_1 & Z_1 Z_1^\top \end{pmatrix}.$$

Note that the expression of I_{Zc} is same as that of $E\,[I_c(\theta_c|Z)]$.

Now, as in Result 5, we get

Result 6 *As* $d \downarrow 0$,

$$P_\theta \left[|\alpha^*_{AN_c} - \alpha^*_{BN_c} - \alpha_A + \alpha_B| \le d\right] \to 1 - \alpha \quad \textit{for all } \theta \in \Omega_c.$$

4 Evaluation

We have done detailed computation through simulations under different setup, but here, for the sake of brevity, we present only a brief illustrative snapshop only. Moreover, we illustrate the proposed procedure through one data example which has been used by many authors for the last two decades for illustrative purposes in the context of response-adaptive designs.

4.1 Simulation

For illustration, in Tables 1 and 2, we report the results obtained from the computations from 10,000 iterations on normally distributed responses for $\alpha = 0.05$. We assumed that $\mu_A = \alpha_A + \beta_A Z_1$ and $\mu_B = \alpha_B + \beta_B Z_1$, where Z_1 can take only two possible

Table 1 The values of $E(N)$, CP for the CARA design with normally distributed responses for $d = 0.30, 0.50$ and different values of α_A, when $\alpha = 0.05$ and $\sigma = 1$; and Prop_A, $\text{Prop}_A^{(1)}$ and $\text{Prop}_A^{(0)}$ for the CARA design

α_A	$E(N)$	CP	$Prop_A$	$Prop_A^{(1)}$	$Prop_A^{(0)}$
$d = 0.50$; here $2n_0 = 10$					
0.0	136.3835 (0.4415)	0.9532 (0.0021)	0.5003 (0.0011)	0.5014 (0.0017)	0.4992 (0.0014)
0.2	140.0700 (0.5022)	0.9511 (0.0022)	0.5713 (0.0011)	0.5746 (0.0016)	0.5777 (0.0014)
0.4	152.0789 (0.7118)	0.9510 (0.0022)	0.6424 (0.0010)	0.6493 (0.0015)	0.6356 (0.0013)
0.6	176.0866 (1.0588)	0.9504 (0.0022)	0.7090 (0.0009)	0.7172 (0.0013)	0.7009 (0.0012)
0.8	211.0685 (1.4334)	0.9516 (0.0021)	0.7679 (0.0008)	0.7805 (0.0011)	0.7555 (0.0011)
1.0	265.2023 (1.9929)	0.9459 (0.0023)	0.8183 (0.0007)	0.8303 (0.0009)	0.8063 (0.0010)

Standard errors (SE's) are given in the parentheses. Here $\alpha_B = 0$, $\beta_A = \beta_B = 1$

Table 2 The values of $E(N)$, CP for the CARA design with normally distributed responses for $d = 0.30, 0.50$ and different values of α_A, when $\alpha = 0.05$ and $\sigma = 1$; and Prop_A, $\text{Prop}_A^{(1)}$ and $\text{Prop}_A^{(0)}$ for the CARA design

α_A	$E(N)$	CP	Prop_A	$\text{Prop}_A^{(1)}$	$\text{Prop}_A^{(0)}$
$d = 0.50$; here $2n_0 = 10$					
0.0	138.0170 (0.4086)	0.9493 (0.0022)	0.3432 (0.0009)	0.1939 (0.0012)	0.4913 (0.0014)
0.2	140.6379 (0.4780)	0.9511 (0.0022)	0.4039 (0.0010)	0.2464 (0.0014)	0.5590 (0.0014)
0.4	151.4740 (0.6108)	0.9461 (0.0023)	0.4648 (0.0010)	0.2997 (0.0016)	0.6279 (0.0013)
0.6	170.5558 (0.7955)	0.9483 (0.0022)	0.5273 (0.0010)	0.3647 (0.0016)	0.6880 (0.0012)
0.8	202.4080 (1.0722)	0.9500 (0.0022)	0.5925 (0.0010)	0.4370 (0.0016)	0.7467 (0.0011)
1.0	260.3217 (1.9763)	0.9452 (0.0023)	0.6580 (0.0009)	0.5153 (0.0015)	0.7995 (0.0010)
$d = 0.30$; here $2n_0 = 10$					
0.0	356.3226 (0.4450)	0.9500 (0.0022)	0.3320 (0.0007)	0.1653 (0.0009)	0.4977 (0.0009)
0.2	365.6799 (0.5195)	0.9509 (0.0022)	0.3947 (0.0007)	0.2173 (0.0010)	0.5710 (0.0009)
0.4	394.0609 (0.6988)	0.9497 (0.0022)	0.4619 (0.0007)	0.2791 (0.0011)	0.6438 (0.0009)
0.6	447.2048 (1.0112)	0.9461 (0.0023)	0.5316 (0.0007)	0.3503 (0.0011)	0.7120 (0.0008)
0.8	532.1538 (1.3932)	0.9424 (0.0023)	0.6004 (0.0006)	0.4262 (0.0010)	0.7740 (0.0007)
1.0	650.4065 (1.7324)	0.9459 (0.0026)	0.6675 (0.0006)	0.5068 (0.0010)	0.8278 (0.0006)

Standard errors (SE's) are given in the parentheses. Here $\alpha_B = 0$, $\beta_A = 1$ and $\beta_B = 2$

values, 1 and 0; that is Z_1 is a Bernoulli(0.5) variable. We considered fixed the values of α_B, β_A and β_B. In Table 1, we considered $\alpha_B = 0$ and $\beta_A = \beta_B = 1$, whereas in Table 2, $\alpha_B = 0$, $\beta_A = 1$ and $\beta_B = 2$ are considered. In Table 1, $d = 0.50$ is taken, but in Table 2, the two possible values of d, namely $d = 0.5$ and 0.3, are considered. We have considered $\sigma = 1$ throughout. Also an initial sample size of $2n_0 = 10$ is taken, which is equally distributed among the two competing treatments. These are used to get initial estimates of the parameters. The response-adaptive allocation procedure is carried out from the $(2n_0 + 1)$st patient onwards. For the allocation

purpose, the estimates of the parameters are obtained for every patient starting from the patient number $(2n_0 + 1)$. We then varied α_A (≥ 0) to compute the average value of sample size N, i.e., $E(N)$, and the corresponding coverage probability (CP) for obtaining a fixed-width confidence interval of $\alpha_A - \alpha_B$ by our given procedure. For response-adaptive allocation, we considered CARA version of the link-function-based allocation function,

$$\pi_A(\theta_i, Z_{i+1}) = \Phi\left(\frac{\widehat{\alpha}_{Ai} - \widehat{\alpha}_{Bi} + (\widehat{\beta}_{Ai} - \widehat{\beta}_{Bi})Z_{1,i+1}}{\widehat{\sigma}_i}\right) \tag{16}$$

of Bandyopadhyay and Biswas (2001), where $\Phi(\cdot)$ is the cdf of a standard normal distribution. Here, as Z_1 is binary, we also report the proportion of allocation to treatment A for the two strata separately (given by $\text{Prop}_A^{(1)}$ and $\text{Prop}_A^{(0)}$). The proportions of allocation to the better treatment A $(= \text{Prop}_A = 1 - \text{Prop}_B)$ are also provided. The standard errors of N, CP, Prop_A, $\text{Prop}_A^{(1)}$ and $\text{Prop}_A^{(0)}$ are also given in parentheses.

Clearly, $E(N)$ is smaller for larger value of d when all other parameters are same. The values of CP are almost 0.95. The Prop_A-values for the adaptive allocation are more than 50% when $\mu_A > \mu_B$, which is an important advantage of using adaptive allocation rule. In fact, these Prop_A-values increase with the increase in $\mu_A - \mu_B$ (for some fixed μ_A).

The initial sample size is an important issue for any such adaptive allocation. The initial samples are used to get initial estimates of the underlying parameters to get the allocation procedure started. So, the initial sample size should be large enough to get estimates of all the parameters. Moreover, it should be large enough so that the initial estimates become reliable. Usually, without any prior information about the superiority of one treatment over the other, the initial samples are equally taken on the two competing treatments. It is expected that a large initial sample size will produce good initial estimates and, hence, the first few adaptive allocations for the patients $2n_0 + 1, 2n_0 + 2, \ldots$ will be reliable and less variable. Thus, the adaptive mechanism will yield good and reliable allocation with small variation. On the other hand, a large initial sample size will allocate a large number of patients equally, and as a result the final allocation will be less skewed in favor of the better performing treatment.

To understand this effect, we carried out a simulation study where we find the values of $E(N)$, CP for the CARA design with normally distributed responses for $\alpha_A = 0.4$, $\alpha_B = 0$, $\beta_A = 1$, $\sigma = 1$ when $\alpha = 0.05$. Two values of β_B, namely 1 and 2, are considered. We also obtain Prop_A, $\text{Prop}_A^{(1)}$ and $\text{Prop}_A^{(0)}$ values. Standard errors (SE's) are given in the parentheses. These are reported in Table 3 for $d = 0.5$. Here, we consider four values of initial sample size, namely $2n_0 = 6, 10, 20, 30$. From Table 3, we observe that the expected sample size decreases as the initial sample size increases. Also the allocation probability to the better treatment decreases with the increase in $2n_0$-value. Moreover, the standard errors decrease with $2n_0$. However, the coverage probabilities are almost 0.95 for all the cases. We carried out computations with some other values of α_B and d. The results are in the same direction and, hence, we did not report them for the sake of brevity. Thus, we need to set the initial sample size carefully; not too small so that the initial adaptive allocations become unstable and

Table 3 The values of $E(N)$, CP for the CARA design with normally distributed responses for $d = 0.50$, $\alpha = 0.05$ and $\sigma = 1$; and Prop_A, $\text{Prop}_A^{(1)}$ and $\text{Prop}_A^{(0)}$ for the CARA design

$2n_0$	$E(N)$	CP	Prop_A	$\text{Prop}_A^{(1)}$	$\text{Prop}_A^{(0)}$
$\beta_B = 1$					
6	159.8554 (0.9921)	0.9492 (0.0022)	0.6458 (0.0012)	0.6543 (0.0018)	0.6376 (0.0014)
10	151.4740 (0.6108)	0.9461 (0.0023)	0.4648 (0.0010)	0.2997 (0.0016)	0.6279 (0.0013)
20	142.8337 (0.4065)	0.9520 (0.0021)	0.6285 (0.0008)	0.6320 (0.0012)	0.6248 (0.0011)
30	138.9402 (0.3105)	0.9492 (0.0022)	0.6165 (0.0007)	0.6187 (0.0010)	0.6142 (0.0010)
$\beta_B = 2$					
6	160.6003 (1.1300)	0.9475 (0.0022)	0.4654 (0.0012)	0.3022 (0.0019)	0.6267 (0.0014)
10	151.4740 (0.6108)	0.9461 (0.0023)	0.4648 (0.0010)	0.2997 (0.0016)	0.6279 (0.0013)
20	142.9273 (0.3839)	0.9489 (0.0022)	0.4672 (0.0008)	0.3154 (0.0011)	0.6173 (0.0012)
30	138.8313 (0.2974)	0.9490 (0.0022)	0.4713 (0.0007)	0.3327 (0.0010)	0.6086 (0.0010)

Standard errors (SE's) are given in the parentheses. Here $\alpha_A = 0.4$, $\alpha_B = 0$, $\beta_A = 1$ and $\beta_B = 1, 2$

highly variable, and not too large to spoil the essence of skewing allocations in the response-adaptive design. In the data example of Sect. 4.2, we set $2n_0 = 10$.

4.2 Data example

In this section, we illustrate the proposed methodology using the data from the fluoxetine trial originally reported by Tamura et al. (1994). It was a trial of fluoxetine hydrochloride to treat outpatients of depressive disorder. The response variable is the reduction of HAMD_{17} score which is originally measured in a 17-point scale. These data are considered to be continuous as in Atkinson and Biswas (2005). The covariate considered here is the indicator of shortened rapid eye movement latency. Out of 88 samples, the number of patients treated by Fluoxetine (treatment A) and placebo (treatment B) was 45 and 43, almost 50:50 allocation. From the data, the proportion of shortened rapid eye movement latency (REML) was exactly 50% (44 in number). Using our model, the estimates of the parameters came out to be $\widehat{\alpha}_A = 0.3606$, $\widehat{\alpha}_B = 2.9100$, $\widehat{\beta}_A = 1.4071$, $\widehat{\beta}_B = -3.1190$ and $\widehat{\sigma}^2 = 6.9372^2$. Pretending them to be true values, we carry out our proposed procedure of fixed-width confidence interval estimation with the link-function based CARA design (16). We summarize the results for obtaining the fixed-width confidence interval estimation for $\alpha_A - \alpha_B$ in Table 4 (upper half) with $\alpha = 0.05$, $d = 0.8, 0.6, 0.4$. This corresponds to the case when $Z_1 = 0$. Then, in the bottom half of Table 4, we provide the results corresponding to the fixed-width confidence interval estimation of $\alpha_A - \alpha_B + (\beta_A - \beta_B)$, which corresponds to the case when $Z_1 = 1$. We observe that we can simultaneously achieve skewed allocation in favor of treatment A in the shortened REML group and skewed allocation in favor of treatment B in the other group (keeping the overall allocation nearly 50%), and fixed-width confidence interval estimation can be carried out at the same time.

Table 4 Results from data analyses of fluoxetine trial: $\alpha = 0.05$, $d = 0.40, 0.60, 0.80$

d	$E(N)$	CP	Prop_A	$\text{Prop}_A^{(1)}$	$\text{Prop}_A^{(0)}$
Case of $\alpha_A - \alpha_B$					
$d = 0.80$	70.6296 (0.2175)	0.9455 (0.0023)	0.4780 (0.0007)	0.6799 (0.0009)	0.2801 (0.0009)
$d = 0.60$	127.2192 (0.3156)	0.9499 (0.0022)	0.4770 (0.0005)	0.7054 (0.0007)	0.2510 (0.0007)
$d = 0.40$	293.0890 (0.5114)	0.9447 (0.0023)	0.4737 (0.0003)	0.7250 (0.0005)	0.2237 (0.0005)
Case of $\alpha_A - \alpha_B + (\beta_A - \beta_B)$					
$d = 0.80$	66.1572 (0.1844)	0.9344 (0.0025)	0.4957 (0.0007)	0.6917 (0.0009)	0.2954 (0.0009)
$d = 0.60$	117.1493 (0.2704)	0.9383 (0.0024)	0.4947 (0.0005)	0.7185 (0.0007)	0.2681 (0.0007)
$d = 0.40$	266.8701 (0.4254)	0.9467 (0.0022)	0.4949 (0.0004)	0.7440 (0.0005)	0.2450 (0.0005)

Standard errors are given in parentheses. Here $2n_0 = 20$

5 Concluding remarks

In this work, we have examined the impact of adaptive allocation rules by determining a positive integer N, which is the stopping variable, where d (> 0) and $\alpha \in (0, 1)$ are pre-specified. The theoretical results provide a basis for the application of this technique. It is observed that if we could set some d and α at the outset, we can have a very clear-cut stopping rule. The numerical simulations and the illustration with real data show the practical applicability of the approach for stopping an experiment. It is clear that we have gain in sample size and proportion of patients to the better treatment by the adaptive procedure. The two-stage procedure has similar application with the feature that we need to keep on estimating the parameters of the second stage.

Thus, the proposed procedure uses the particular CARA design under consideration, but aims at finding the sample size appropriately to carry out fixed-width confidence interval estimation of the particular treatment difference of interest. Our theoretical development followed by simulation exercise and data illustration indicate that the proposed procedure succeeds to achieve the targeted width while carrying out the ethical CARA design.

Acknowledgements The authors wish to thank the Editor and one anonymous referee for their careful reading and constructive suggestions which led to some improvement over an earlier version of the manuscript.

Appendix A

Result 7 *As* $m \to \infty$,

$$\frac{1}{m}\sum_{i=1}^{m} \delta_{ki}\,\tau_k(\theta_k, Z_i) Z_i Z_i^{\top} \to I_{Zk}$$

almost surely, $k = A, B$.

Proof It is not difficult to observe that

$$\sum_{i=1}^{m}[\delta_{ki}\tau_k(\theta_k, Z_i)Z_iZ_i^\top - E(\pi_k(\widehat{\theta}_{i-1}, Z_i)\tau_k(\theta_k, Z_i)Z_iZ_i^\top)],$$

where $\widehat{\theta_i}$ is the ML estimator based on $\{\delta_{kj}, Y_j, Z_j;\ j = 1, 2, \ldots, i; k = A, B\}$, is a martingale sequence. Hence, applying the same technique as in (A.14) under Corollary 3.1 of Zhang et al. (2007), the required result follows. □

Note 3 *Note A.1. If* $\tau_k(\theta_k, Z) = 1$, *the above result implies*

$$\widehat{I}_{mk} = \frac{1}{m}\sum_{i=1}^{m}\delta_{ki}Z_iZ_i^\top \to E[\pi_k(\theta, Z)ZZ^\top] \tag{17}$$

almost surely as $m \to \infty$.

If $\tau_k(\theta_k, Z) = \frac{e^{\theta_k^\top Z}}{(1+e^{\theta_k^\top Z})^2}$, then the mean-value theorem yields

$$|\tau_k(\widehat{\theta}_{km}, Z) - \tau_k(\theta_k, Z)| < \frac{M}{2}||\widehat{\theta}_{km} - \theta_k||, \tag{18}$$

where M is described in Sect. 2. Consequently, (17) and (18) imply

$$\begin{aligned}\widehat{I}_{mk} &= \frac{1}{m}\sum_{i=1}^{m}\delta_{ki}\left[\tau_k(\theta_k, Z_i) + O\left(||\widehat{\theta}_{km} - \theta_k||\right)\right]Z_iZ_i^\top \\ &= \frac{1}{m}\sum_{i=1}^{m}\delta_{ki}\tau_k(\theta_k, Z_i)Z_iZ_i^\top + O(||\widehat{\theta}_{km} - \theta_k||).O(1)\end{aligned}$$

almost surely. Hence, by the above result, it follows that

$$\widehat{I}_{mk} \to E[\pi_k(\theta, Z)p_k(Z)(1 - p_k(Z))ZZ^\top]$$

almost surely as $m \to \infty$, $k = A, B$.

Result 8 *As* $\nu \to \infty$,

$$\frac{1}{\sqrt{\nu}}S_{k\nu} \to N_r(0, \Sigma_k)$$

in distribution with Σ_k *given in Result 1,* $k = A, B$.

Proof Let ℓ_k, $k = A, B$, be two r-component fixed vectors. Then, setting $a_i = \ell_A^\top I_{ZA}^{-1}Z_i$ and $b_i = \ell_B^\top I_{ZB}^{-1}Z_i$, consider

$$V_{\nu i} = \frac{1}{\sqrt{\nu}}\left[\delta_{Ai}a_i(X_{Ai} - \mu_A(Z_i)) + \delta_{Bi}b_i(X_{Bi} - \mu_B(Z_i))\right],$$

where $\mu_k(Z_i) = p_k(Z_i)$ and $\theta_k^\top Z_i$ for models (a) and (b), $i = 1, 2, \ldots, \nu$, $k = A, B$. The sequences $\{V_{\nu i},\ 1 \le i \le \nu,\ \nu \ge 1\}$ represent differences corresponding to the martingale sequence $T_\nu = \sum_{i=1}^{\nu} V_{\nu i}$, $\nu \ge 1$. Then, we get

$$E(T_\nu^2) = \sum_{i=1}^{\nu} E(V_{\nu i}^2) \le \sum_{k=A,B} \ell^\top I_{Zk}^{-1} E[\tau_k(\theta_k, Z) Z Z^\top] I_{Zk}^{-1} \ell_k < \infty.$$

Hence, by martingale central limit theorem (see, for example. Theorem 3.2 of Hall and Heyde 1980, p. 58), it will follow that

$$T_\nu \to N(0, \eta^2)$$

in distribution as $\nu \to \infty$, where $\eta^2 = \sum_{k=A,B} (\ell_k^\top \Sigma_k \ell_k)$, provided all the conditions of Theorem 3.2 in Hall and Heyde (1980) are satisfied. Here, all the conditions, except (3.19) of this theorem, are trivially satisfied. Now, to prove this non-trivial condition, write

$$\sum_{i=1}^{\nu} W_i = \sum_{i=1}^{\nu} [\nu V_{\nu i}^2 - a_i^2 \tau_A(\theta, Z_i) \delta_{Ai} - b_i^2 \tau_B(\theta, Z_i) \delta_{Bi}],$$

where

$$W_i = \delta_{Ai} a_i^2 [(X_{Ai} - \mu_A(Z_i))^2 - \tau_A(\theta, Z_i)] + \delta_{Bi} b_i^2 [(X_{Bi} - \mu_B(Z_i))^2 - \tau_B(\theta, Z_i)].$$

Using Theorem 2.13 (iii) of Hall and Heyde (1980), we get, as $\nu \to \infty$,

$$\frac{1}{\nu} \sum_{i=1}^{\nu} W_i \to 0$$

in probability, which implies

$$\sum_{i=1}^{\nu} V_{\nu i}^2 \to \eta^2$$

in probability as $\nu \to \infty$. Hence, by the Cramer–Wold device, we get the required result. □

References

Anscombe, F. J. (1952). Large sample-theory of sequential estimation. *Proceedings of the Cambridge Philosophical Society*, *48*, 600–607.

Atkinson, A. C., Biswas, A. (2005). Bayesian adaptive biased-coin designs for clinical trials with normal responses. *Biometrics*, *61*, 118–125.

Atkinson, A. C., Biswas, A. (2014). *Randomized response-adaptive designs in clinical trials*. Boca Raton: Chapman & Hall/CRC.

Bandyopadhyay, U., Biswas, A. (1999). Allocation by randomized play-the-winner rule in the presence of prognostic factors. *Sankhya Series B*, *61*, 397–412.

Bandyopadhyay, U., Biswas, A. (2001). Adaptive designs for normal responses with prognostic factors. *Biometrika*, *88*, 409–419.

Bandyopadhyay, U., Biswas, A. (2015). Sequential and two-stage fixed-width confidence interval estimation in response-adaptive designs. *Sequential Analysis* (Special issue: Celebrating seventy years of Charles Stein's 1945 seminal paper on two-stage sampling) **34**, 350–363.

Biswas, A., Bhattacharya, R. (2016). Response-adaptive designs for continuous treatment responses in phase III clinical trials: A review. *Statistical Methods in Medical Research*. (In press).

Biswas, A., Coad, D. S. (2005). A general multi-treatment adaptive design for multivariate responses. *Sequential Analysis*, *24*, 139–158.

Chang, Y.-C. I., Park, E. (2013). Sequential estimation for covariate-adjusted response-adaptive designs. *Journal of the Korean Statistical Society*, *42*, 105–116.

Ghosh, M., Mukhopadhyay, N., Sen, P. K. (1997). *Sequential estimation*. New York: Wiley.

Hall, P., Heyde, C. C. (1980). *Martingale limit theory and its application*. New York: Academic Press.

McCullagh, P., Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman and Hall/CRC.

Rosenberger, W. F., Stallard, N., Ivanova, A., Harper, C., Ricks, M. (2001). Optimal adaptive designs for binary response trials. *Biometrics*, *57*, 909–913.

Sen, P. K. (1981). *Sequential nonparametrics*. New York: Wiley.

Tamura, R., Faries, D., Andersen, J., Heiligenstein, J. (1994). A case study of an adaptive clinical trial in the treatment of out-patients with depressive disorder. *Journal of the American Statistical Association*, *89*, 768–776.

Wei, L. J., Durham, S. (1978). The randomized play-the-winner rule in medical trials. *Journal of the American Statistical Association*, *73*, 838–843.

Zhang, L. X., Hu, F., Cheung, S. H., Chan, W. S. (2007). Asymptotic properties of covariate-adjusted adaptive designs. *Annals of Statistics*, *35*, 1166–1182.