

Biomarker-driven optimal designs for patient enrollment restriction

Alessandro Baldi Antognini , Sara Cecconi ,
Rosamarie Frieri  and Maroussa Zagoraiou 

Statistical Methods in Medical Research

2025, Vol. 34(9) 1821–1837

© The Author(s) 2025



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/09622802251327690

journals.sagepub.com/home/smm

Abstract

The rapidly developing field of personalized medicine is giving the opportunity to treat patients with a specific regimen according to their individual demographic, biological, or genomic characteristics, known also as biomarkers. While binary biomarkers simplify subgroup selection, challenges arise in the presence of continuous ones, which are often categorized based on data-driven quantiles. In the context of binary response trials for treatment comparisons, this paper proposes a method for determining the optimal cutoff of a continuous predictive biomarker to discriminate between sensitive and insensitive patients, based on their relative risk. We derived the optimal design to estimate such a cutoff, which requires a set of equality constraints that involve the unknown model parameters and the patients' biomarker values and are not directly attainable. To implement the optimal design, a novel covariate-adjusted response-adaptive randomization is introduced, aimed at sequentially minimizing the Euclidean distance between the current allocation and the optimum. An extensive simulation study shows the performance of the proposed approach in terms of estimation efficiency and variance of the estimated cutoff. Finally, we show the potential severe ethical impact of adopting the data-dependent median to identify the subpopulations.

Keywords

binary responses, continuous biomarker, covariate-adjusted response-adaptive randomization, personalized medicine, threshold identification

1 Introduction and model

Nowadays almost all branches of medicine are moving toward personalized medicine based on the belief that all patients cannot be successfully treated with the same therapy. In the presence of some evidence that the effect of a treatment may differ in certain subpopulations, personalized (also known as precision medicine) may be extremely useful. More precisely, personalized medicine is the tailoring of medical treatments to the individual characteristics, or biomarkers, of each patient. The process of selection of the subpopulations based on one or more biomarkers is called enrichment: subjects are screened for their biomarker profile, and then only those with or without certain characteristics are included in the trial and could be suitably randomized to the competing treatments.

While enrichment trials are specifically applied in the design phase, in terms of both enrollment restrictions and treatment allocation process, traditional subgroup identification methods are ex-post procedures for analyzing data already accrued in randomized clinical trials. The vast majority of them are classification algorithms of heuristic nature, without model specifications; they are based on classification trees combined with machine learning techniques, aimed at identifying a set of predictive biomarkers (among many available covariates, some of which are of prognostic nature) as well as suitable subsets of their support (see, e.g. Foster et al.¹). Other proposals have been introduced for continuous endpoints under the

Department of Statistical Sciences, University of Bologna, Bologna, Italy

Corresponding author:

Maroussa Zagoraiou, Department of Statistical Sciences, University of Bologna, Via Belle Arti, 41, 40126, Bologna, Italy.

Email: maroussa.zagoraiou@unibo.it

classical linear model, often in the causal inference framework or inspired by latent variables approaches (see for a recent review by Loh et al.²).

In several cases, the effects of biomarkers on treatments are not explicit. Therefore, adaptive enrichment design methodologies, using the accrued information on previous subjects' responses to find the benefitting population, can be applied. In the presence of a binary (or categorical) biomarker, the patients' subgroups are well-defined. Nevertheless, when the biomarker of interest is defined on a continuous scale (e.g. age, cholesterol, and blood pressure), a widely used approach consists of discretizing the biomarker via a cutoff given by a data-driven quantile, usually the median, in order to define the subgroups. However, as stated by many authors, "discretization of a continuous biomarker using sample percentiles results in significant information loss and should be avoided" (see Polley and Dignam³ and Zhang and Molinaro⁴). Moreover, often a single candidate predictive biomarker is identified via preliminary information, but a suitable cutpoint has not been established.⁵ Under this framework, patient enrollment restriction based on the biomarker would be inappropriate. Indeed, an erroneous identification of the threshold for guiding future treatment decisions for individual patients could lead to potentially severe consequences.

Regrettably, there is still a lack of studies that estimate the threshold directly on a continuous scale to correctly discriminate among the so-called sensitive (biomarker-positive) and insensitive (biomarker-negative) patients.⁶ Recently, for a single-arm trial, Spencer et al.⁷ presented a biomarker-adaptive threshold design to determine if a subpopulation with a clinically relevant response rate exists. Some proposals have been developed in the context of survival trials (see Trippa et al.,⁸ Renfro et al.,⁹ and Diao et al.¹⁰). On the other hand, Lin et al.¹¹ and Frieri et al.¹² adopted a bivariate-normal model to estimate the biomarker threshold, using the correlation biomarker-response as an explicit gauge of the predictive nature of the biomarker. Baldi Antognini et al.¹³ used the linear treatment-by-covariate interaction approach to find the cutoff and provide the optimal allocations for both parameter estimation and cutoff identification.

Inspired by numerous real clinical trials documented in the literature, consider as a motivating example the NSABP B-35 trial described by Margolese et al.¹⁴ By enrolling 3104 patients, this study compares anastrozole versus tamoxifen in postmenopausal women with hormone receptor-positive ductal carcinoma in situ undergoing lumpectomy plus radiotherapy, showing that anastrozole provides a significant improvement in breast cancer-free interval in women younger than 60 years.

Since a binary model for the biomarker-treatment relationship is frequently required for therapeutic success/failure cases (see, e.g. Vinnat and Chevret¹⁵), in this paper, we approach the challenging issue of identifying the optimal cutpoint of a continuous predictive biomarker based on the relative risk for binary trials for treatment comparisons. Our aim is to provide optimal allocations for inference on the threshold and then describe a suitable covariate-adjusted response-adaptive (CARA) procedure to implement this optimum. After introducing the statistical model in Section 2, Section 3 deals with classical D and A optimality criteria, also providing the variance of the estimated threshold. We derive the optimal design that simultaneously optimizes all the above-mentioned criteria, which consists of a set of conditions, involving the unknown patients' biomarker values and the unknown model parameters. Thus, Section 4 is dedicated to the implementation of the optimal design through a suitably defined CARA rule. Finally, in Section 5, we perform a simulation study adopting normal and log-normal distribution for the biomarker, in order to compare the performance of the new CARA procedure with that of conventional trial designs. To stress the practical impact of our procedure, we redesign a clinical trial for severe sepsis disease, taking also into account the Virtual Twins method¹ for subgroup identification. Additionally, the ethical impact resulting from the use of a median-based biomarker cutoff has been explored, comparing it with the benefit gained from using the relative risk-based cutoff.

2 Binary responses

We consider an oncological trial where patients are sequentially assigned to one of two competing anticancer agents T or C . Let us denote by X the chosen quantitative predictive biomarker, which is not under the experimenters' control. When the i th subject is ready to be randomized, the biomarker value $X_i = x_i$ is recorded and she/he receives one of the agents based on a given randomization rule and a treatment indicator variable records the assignment: if the patient is assigned to T , $\delta_i = 1$ while $\delta_i = 0$ otherwise. Thus, $n^{-1} \sum_{i=1}^n \delta_i = \pi$ is the subjects' proportion allocated to T and $n^{-1} \sum_{i=1}^n (1 - \delta_i) = 1 - \pi$ to C . We assume X_1, X_2, \dots independent and identically distributed random variables having common density function with finite expected value and variance. A binary tumor status Y_i is examined after the treatment assignment to directly measure any anticancer activity (e.g. the ctDNA clearance in Spreafico et al.¹⁶) and we take $Y_i = 1$ in the case of successful response of the i th patient (clearly, this setting applies to any dichotomous outcomes of interest, such as mortality of patients with hypoxemic acute respiratory failure after different oxygenation therapies in the HIGH clinical trial, see Azoulay et al.¹⁷).

Conditionally on the treatments and the biomarker, subjects' responses are assumed to be independent following a logistic regression model:

$$\begin{aligned}
 P(Y_i = 1 | \delta_i = 1, X_i = x_i) &= p_i^T = \frac{1}{1 + e^{-(\alpha^T + \beta^T x_i)}} \\
 P(Y_i = 1 | \delta_i = 0, X_i = x_i) &= p_i^C = \frac{1}{1 + e^{-(\alpha^C + \beta^C x_i)}}
 \end{aligned}
 \tag{1}$$

where, since X is a predictive biomarker, $\beta^T \neq \beta^C$.

The primary goal of the paper is to assess whether an anticancer agent is better than the other for a subgroup of subjects based on their biomarker. By taking into account the relative risk,

$$R(x) = \frac{P(Y_i = 1 | \delta_i = 1, X_i = x)}{P(Y_i = 1 | \delta_i = 0, X_i = x)}$$

it is easy to show that, for every subject with biomarker value x greater than the unique cutoff

$$x^* = \frac{\alpha^T - \alpha^C}{\beta^C - \beta^T}$$

agent T is better than C (namely $R(x) > 1$) provided that $\beta^T > \beta^C$ (or for $x < x^*$ if $\beta^T < \beta^C$). This subpopulation of patients (often referred to as target or benefitting subpopulation) is generally the focus of the investigator assuming that T is regarded as the new/experimental treatment. Moreover, after the identification of the cutoff, a secondary objective is to evaluate the impact of the discretization of the continuous biomarker using a sample percentile instead of x^* , in terms of ethical loss (namely the percentage of subjects assigned to the inferior treatment).

As regards the notation, throughout the paper let $\mathbf{x}_n^t = (x_1, \dots, x_n)$, $\boldsymbol{\delta}_n^t = (\delta_1, \dots, \delta_n)$, and $\mathbf{y}_n^t = (y_1, \dots, y_n)$ be the vector of biomarker values, assignments, and outcomes observed after n steps and $\mathbf{0}_k$ is the k -dimensional vector of zeros.

3 Trial designs

Optimal design theory is recognized as an important tool to achieve allocations that yield the best performance out of a range of potential designs. The literature is rich with optimal design methodologies that can be adopted by clinical researchers to improve the efficiency of the drug development process. The key idea is to select a criterion that measures the loss as a function of the inverse of the Fisher information matrix since it is an asymptotic approximation of the variance–covariance matrix of the maximum likelihood estimators (MLEs). Denoting by $\boldsymbol{\theta} = (\alpha^T, \alpha^C, \beta^T, \beta^C)^t$ the vector of the unknown parameters, let $\hat{\boldsymbol{\theta}}_n = (\hat{\alpha}_n^T, \hat{\alpha}_n^C, \hat{\beta}_n^T, \hat{\beta}_n^C)^t$ be the MLE of $\boldsymbol{\theta}$ after n steps, then the Fisher information is (see Appendix 1)

$$\mathbf{I}_n = \sum_{i=1}^n \begin{pmatrix} \delta_i p_i^T (1 - p_i^T) & 0 & \delta_i p_i^T (1 - p_i^T) x_i & 0 \\ 0 & (1 - \delta_i) p_i^C (1 - p_i^C) & 0 & (1 - \delta_i) p_i^C (1 - p_i^C) x_i \\ \delta_i p_i^T (1 - p_i^T) x_i & 0 & \delta_i p_i^T (1 - p_i^T) x_i^2 & 0 \\ 0 & (1 - \delta_i) p_i^C (1 - p_i^C) x_i & 0 & (1 - \delta_i) p_i^C (1 - p_i^C) x_i^2 \end{pmatrix}$$

To simplify the notation, let us define the variance of the response of patient i with $v_i = \delta_i p_i^T (1 - p_i^T) + (1 - \delta_i) p_i^C (1 - p_i^C)$; clearly, if $\delta_i = 1$ then $v_i = p_i^T (1 - p_i^T)$, whereas $v_i = p_i^C (1 - p_i^C)$ if $\delta_i = 0$. Thus, $\delta_i v_i = \delta_i p_i^T (1 - p_i^T)$ and $(1 - \delta_i) v_i = (1 - \delta_i) p_i^C (1 - p_i^C)$, so that

$$M^T(v) = \frac{\sum_{i=1}^n \delta_i v_i}{\sum_{i=1}^n \delta_i} \quad \text{and} \quad M^C(v) = \frac{\sum_{i=1}^n (1 - \delta_i) v_i}{\sum_{i=1}^n (1 - \delta_i)}$$

simply denote the means of the observed variances in the two groups. Moreover, since

$$v^T = \left(\frac{\delta_i v_i}{\sum_{i=1}^n \delta_i} : i = 1, \dots, n \right) \quad \text{and} \quad v^C = \left(\frac{(1 - \delta_i) v_i}{\sum_{i=1}^n (1 - \delta_i)} : i = 1, \dots, n \right)$$

represent two discrete probability distributions, we denote by

$$M_v^T(x) = \frac{\sum_{i=1}^n \delta_i v_i x_i}{\sum_{i=1}^n \delta_i v_i} \quad \text{and} \quad M_v^C(x) = \frac{\sum_{i=1}^n (1 - \delta_i) v_i x_i}{\sum_{i=1}^n (1 - \delta_i) v_i}$$

the mean of the biomarker with respect to v^T and v^C , respectively, with $V_v^j(x) = M_v^j(x^2) - [M_v^j(x)]^2$ ($j = T, C$) the corresponding variances. Thus, the average information matrix $\mathbf{M}_n = \mathbf{I}_n/n$ is

$$\mathbf{M}_n = \begin{pmatrix} \pi M^T(v) & 0 & \pi M^T(v) M_v^T(x) & 0 \\ 0 & (1 - \pi) M^C(v) & 0 & (1 - \pi) M^C(v) M_v^C(x) \\ \pi M^T(v) M_v^T(x) & 0 & \pi M^T(v) M_v^T(x^2) & 0 \\ 0 & (1 - \pi) M^C(v) M_v^C(x) & 0 & (1 - \pi) M^C(v) M_v^C(x^2) \end{pmatrix} \quad (2)$$

and its inverse is given by (see Appendix 2)

$$\mathbf{M}_n^{-1} = \begin{pmatrix} \frac{M_v^T(x^2)}{\pi M^T(v) V_v^T(x)} & 0 & -\frac{M_v^T(x)}{\pi M^T(v) V_v^T(x)} & 0 \\ 0 & \frac{M_v^C(x^2)}{(1 - \pi) M^C(v) V_v^C(x)} & 0 & -\frac{M_v^C(x)}{(1 - \pi) M^C(v) V_v^C(x)} \\ -\frac{M_v^T(x)}{\pi M^T(v) V_v^T(x)} & 0 & \frac{1}{\pi M^T(v) V_v^T(x)} & 0 \\ 0 & -\frac{M_v^C(x)}{(1 - \pi) M^C(v) V_v^C(x)} & 0 & \frac{1}{(1 - \pi) M^C(v) V_v^C(x)} \end{pmatrix}$$

3.1 Optimality criteria

When the interest is in the estimation of the whole parameter vector θ , D and A optimality criteria are typically used. The former criterion minimizes the volume of the confidence ellipsoid for θ , so that the D optimal design minimizes the determinant of \mathbf{M}_n^{-1} . Whereas, the A optimality minimizes the mean variance of the estimators, so the trace of \mathbf{M}_n^{-1} should be minimized. Concerning the estimation of x^* , the MLE of the threshold after n steps is $\hat{x}_n^* = (\hat{\alpha}_n^T - \hat{\alpha}_n^C)/(\hat{\beta}_n^C - \hat{\beta}_n^T)$ and the next theorem provides an approximated closed-form for its variance, as well as the expression of D and A optimality criteria.

Theorem 3.1. *After n assignments,*

$$\det \text{var}(\hat{\theta}_n) = \frac{1}{\pi^2 (1 - \pi)^2 [M^T(v)]^2 [M^C(v)]^2 V_v^T(x) V_v^C(x)} \quad (3)$$

and

$$\text{tr} \text{var}(\hat{\theta}_n) = \frac{1 + M_v^T(x^2)}{\pi M^T(v) V_v^T(x)} + \frac{1 + M_v^C(x^2)}{(1 - \pi) M^C(v) V_v^C(x)} \quad (4)$$

As regards the threshold estimation, since $\beta^T \neq \beta^C$, through a first-order approximation,

$$\text{var}(\hat{x}_n^*) \approx \frac{1}{(\beta^T - \beta^C)^2} \left\{ \frac{\pi M^T(v) + (1 - \pi) M^C(v)}{\pi(1 - \pi) M^T(v) M^C(v)} + \frac{[M_v^T(x) - x^*]^2}{\pi M^T(v) V_v^T(x)} + \frac{[M_v^C(x) - x^*]^2}{(1 - \pi) M^C(v) V_v^C(x)} \right\} \quad (5)$$

Proof. See Appendix 3. □

3.2 Optimal designs

Assuming model (1), the following theorem provides the optimal design for the estimation of the cutoff, which is also D and A optimal.

Theorem 3.2. An allocation δ_n^* satisfying

$$\begin{cases} \pi M^T(v) = (1 - \pi)M^C(v) \\ M_v^T(x) = M_v^C(x) \\ V_v^T(x) = V_v^C(x) \end{cases} \quad (6)$$

is D and A optimal and is optimal for estimating the threshold x^* as well.

Proof. See Appendix 4. □

Conditions (6) define a vast class of optimal designs through equality constraints, involving both the observed biomarker values and the unknown model parameters. In practice, optimality requires the equality of the biomarker weighted means and variances with respect to v^T and v^C , as well as the equality of $\pi M^T(v)$ to $(1 - \pi)M^C(v)$.

Since in our setting incoming patients' biomarker cannot be controlled by the experimenter, given $M(v) = n^{-1} \sum_{i=1}^n v_i$ and the marginal quantities $M_v(x) = \sum_{i=1}^n v_i x_i / \sum_{i=1}^n v_i$ and $V_v(x) = M_v(x^2) - [M_v(x)]^2$ (i.e. mean and variance evaluated w.r.t. $v = (v_i / \sum_{i=1}^n v_i : i = 1, \dots, n)$), the selection of $\pi, M^T(v), M_v^T(x)$ and $V_v^T(x)$ uniquely determines $(1 - \pi), M^C(v), M_v^C(x)$ and $V_v^C(x)$:

$$\begin{aligned} M(v) &= \pi M^T(v) + (1 - \pi)M^C(v), \\ M_v(x) &= [\pi M^T(v)M_v^T(x) + (1 - \pi)M^C(v)M_v^C(x)] / M(v) \quad \text{and} \\ V_v(x) &= [\pi M^T(v)V_v^T(x) + (1 - \pi)M^C(v)V_v^C(x)] / M(v) \end{aligned} \quad (7)$$

Thus, conditions (6) can be restated as $2\pi M^T(v) = M(v)$, $M_v^T(x) = M_v(x)$, $V_v^T(x) = V_v(x)$. If, in addition, a balanced allocation $\pi = 1/2$ is assumed, the optimum requires $M(v) = M^T(v) = M^C(v)$ and clearly $M_v^T(x) = M_v^C(x) = M_v(x)$ and $V_v^T(x) = V_v^C(x) = V_v(x)$.

The dependence of the optimal designs (derived in Theorem 3.2) on the unknown model parameters is highlighted in the next remark.

Remark 3.1. Unlike the results obtained in Baldi Antognini et al.,¹³ conditions (6) cannot be achieved in practice, since they depend on all the observed allocations and biomarker values, as well as on the unknown model parameter θ . Thus, even in the unrealistic scenario in which the experimenter could be able to choose the biomarker values of the subjects, the closeness to the optimum cannot be checked since it depends on θ . Indeed, consider the following toy example with $n = 4$, where we assume $\mathbf{x}_4 = (0, 0, 0.8, 0.8)^t$ and $\delta_4 = (0, 1, 0, 1)^t$. In such a case $\pi = 1/2$, $\sum_{i=1}^4 \delta_i x_i = \sum_{i=1}^4 (1 - \delta_i) x_i = 0.8$ and $\sum_{i=1}^4 \delta_i x_i^2 = \sum_{i=1}^4 (1 - \delta_i) x_i^2 = 0.64$, so the equality between the first and second empirical moments of the biomarker is achieved. When $\theta^t = (-0.5, 0.2, 1.2, 0.2)$ (i.e. $x^* = 0.7$), conditions (6) are satisfied with $M^T(v) = M^C(v) = 0.24$, $M_v^T(x) = M_v^C(x) = 0.40$, and $V_v^T(x) = V_v^C(x) = 0.16$. However, if $\theta^t = (1.8, 2.5, 1.8, 0.8)$ (inducing the same threshold $x^* = 0.7$), then the considered allocation δ_4 is not optimal since $M^T(v) = 0.08 \neq M^C(v) = 0.05$, $M_v^T(x) = 0.18 \neq M_v^C(x) = 0.29$, and $V_v^T(x) = 0.11 \neq V_v^C(x) = 0.15$. Moreover, under this choice of θ does not exist an allocation satisfying conditions (6).

4 CARA procedure to implement the optimum

As pointed out previously, the optimal designs derived in Theorem 3.2 depend on the patients' biomarker values and the unknown model parameters. In these settings, CARA randomization could provide a possible solution to asymptotically approach the optimum. However, standard CARA procedures proposed in the literature are based on a target, namely on the complete specification of the functional form of the optimal design (see, e.g. Baldi Antognini and Giovagnoli¹⁸) and they are not applicable in this setting, where optimality is characterized only by equality constraints. Thus, we now propose a new CARA randomization procedure to achieve the optimum in a sequential manner. From (6), the optimal design requires that

$$\lambda^t = (\pi M^T(v) - (1 - \pi)M^C(v), M_v^T(x) - M_v^C(x), M_v^T(x^2) - M_v^C(x^2)) = \mathbf{0}_3$$

so we suggest the biomarker-adjusted response-adaptive (BiomARA) to minimize the Euclidean norm $\|\lambda\| = \sqrt{\lambda^t \lambda}$ sequentially. It goes without saying that λ depends, at each step n , on the set of collected allocations and patients' biomarker values, as well as on the unknown θ , namely $\lambda_n = \lambda(x_n; \delta_n; \theta)$, so it needs to be sequentially estimated taking also into account patients' responses. To stress this dependence, from now on we let for every $i \geq 1$ $p_i^j = p^j(\theta, x_i)$ and

$v^j(\theta, x_i) = p^j(\theta, x_i)[1 - p^j(\theta, x_i)]$, with $j = T, C$. After n assignments, we could evaluate the current MLE $\hat{\theta}_n$ of θ and so estimate all the previous individual success probabilities and response variances, respectively, by

$$(p^j(\hat{\theta}_n, x_i))_{i=1, \dots, n} \quad \text{and} \quad (v^j(\hat{\theta}_n, x_i))_{i=1, \dots, n}, \quad \text{for } j = T, C$$

Thus, λ can be estimated via

$$\hat{\lambda}_n = \begin{pmatrix} \frac{\sum_{i=1}^n \delta_i v^T(\hat{\theta}_n, x_i)}{n} - \frac{\sum_{i=1}^n (1 - \delta_i) v^C(\hat{\theta}_n, x_i)}{n} \\ \frac{\sum_{i=1}^n \delta_i v^T(\hat{\theta}_n, x_i) x_i}{\sum_{i=1}^n \delta_i v^T(\hat{\theta}_n, x_i)} - \frac{\sum_{i=1}^n (1 - \delta_i) v^C(\hat{\theta}_n, x_i) x_i}{\sum_{i=1}^n (1 - \delta_i) v^C(\hat{\theta}_n, x_i)} \\ \frac{\sum_{i=1}^n \delta_i v^T(\hat{\theta}_n, x_i) x_i^2}{\sum_{i=1}^n \delta_i v^T(\hat{\theta}_n, x_i)} - \frac{\sum_{i=1}^n (1 - \delta_i) v^C(\hat{\theta}_n, x_i) x_i^2}{\sum_{i=1}^n (1 - \delta_i) v^C(\hat{\theta}_n, x_i)} \end{pmatrix}$$

and the BiomARA procedure can be implemented according to the following algorithm:

Step 0: assign n_0 patients with restricted randomization to T and C and set $j = n_0$;

Step 1: based on x_j, δ_j and y_j , estimate θ via $\hat{\theta}_j$;

Step 2: when the $(j+1)$ th patient arrives with biomarker value x_{j+1} , calculate $v^T(\hat{\theta}_j, x_{j+1})$ and $v^C(\hat{\theta}_j, x_{j+1})$ in order to evaluate the potential distances $\|\hat{\lambda}_{j+1}^{(T)}\|$ and $\|\hat{\lambda}_{j+1}^{(C)}\|$ from the optimum that will occur by assigning T or C to this subject, where
 - if $\delta_{j+1} = 1$

$$\hat{\lambda}_{j+1}^{(T)} = \begin{pmatrix} \frac{\sum_{i=1}^j \delta_i v^T(\hat{\theta}_j, x_i) + v^T(\hat{\theta}_j, x_{j+1})}{j+1} - \frac{\sum_{i=1}^j (1 - \delta_i) v^C(\hat{\theta}_j, x_i)}{j+1} \\ \frac{\sum_{i=1}^j \delta_i v^T(\hat{\theta}_j, x_i) x_i + v^T(\hat{\theta}_j, x_{j+1}) x_{j+1}}{\sum_{i=1}^j \delta_i v^T(\hat{\theta}_j, x_i) + v^T(\hat{\theta}_j, x_{j+1})} - \frac{\sum_{i=1}^j (1 - \delta_i) v^C(\hat{\theta}_j, x_i) x_i}{\sum_{i=1}^j (1 - \delta_i) v^C(\hat{\theta}_j, x_i)} \\ \frac{\sum_{i=1}^j \delta_i v^T(\hat{\theta}_j, x_i) x_i^2 + v^T(\hat{\theta}_j, x_{j+1}) x_{j+1}^2}{\sum_{i=1}^j \delta_i v^T(\hat{\theta}_j, x_i) + v^T(\hat{\theta}_j, x_{j+1})} - \frac{\sum_{i=1}^j (1 - \delta_i) v^C(\hat{\theta}_j, x_i) x_i^2}{\sum_{i=1}^j (1 - \delta_i) v^C(\hat{\theta}_j, x_i)} \end{pmatrix}$$

- if $\delta_{j+1} = 0$

$$\hat{\lambda}_{j+1}^{(C)} = \begin{pmatrix} \frac{\sum_{i=1}^j \delta_i v^T(\hat{\theta}_j, x_i)}{j+1} - \frac{\sum_{i=1}^j (1 - \delta_i) v^C(\hat{\theta}_j, x_i) + v^C(\hat{\theta}_j, x_{j+1})}{j+1} \\ \frac{\sum_{i=1}^j \delta_i v^T(\hat{\theta}_j, x_i) x_i}{\sum_{i=1}^j \delta_i v^T(\hat{\theta}_j, x_i)} - \frac{\sum_{i=1}^j (1 - \delta_i) v^C(\hat{\theta}_j, x_i) x_i + v^C(\hat{\theta}_j, x_{j+1}) x_{j+1}}{\sum_{i=1}^j (1 - \delta_i) v^C(\hat{\theta}_j, x_i) + v^C(\hat{\theta}_j, x_{j+1})} \\ \frac{\sum_{i=1}^j \delta_i v^T(\hat{\theta}_j, x_i) x_i^2}{\sum_{i=1}^j \delta_i v^T(\hat{\theta}_j, x_i)} - \frac{\sum_{i=1}^j (1 - \delta_i) v^C(\hat{\theta}_j, x_i) x_i^2 + v^C(\hat{\theta}_j, x_{j+1}) x_{j+1}^2}{\sum_{i=1}^j (1 - \delta_i) v^C(\hat{\theta}_j, x_i) + v^C(\hat{\theta}_j, x_{j+1})} \end{pmatrix}$$

Step 3: assign the $(j+1)$ th patient to T with probability

$$P(\delta_{j+1} = 1 \mid \delta_j, y_j, x_{j+1}) = \begin{cases} 0.5 + \epsilon, & \text{if } \|\hat{\lambda}_{j+1}^{(T)}\| < \|\hat{\lambda}_{j+1}^{(C)}\| \\ 0.5, & \text{if } \|\hat{\lambda}_{j+1}^{(T)}\| = \|\hat{\lambda}_{j+1}^{(C)}\|, \\ 0.5 - \epsilon, & \text{if } \|\hat{\lambda}_{j+1}^{(T)}\| > \|\hat{\lambda}_{j+1}^{(C)}\| \end{cases} \tag{8}$$

where $\epsilon \in (0, 0.5)$, and then observe y_{j+1} .

Set $j = j + 1$ and repeat Steps 1–3 until the enrollment process is completed.

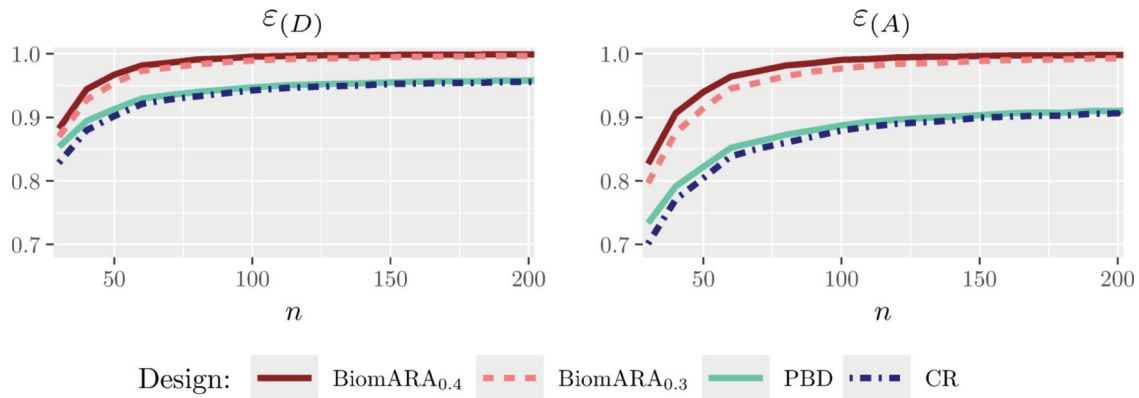


Figure 1. *D* and *A* efficiency for standard normal biomarker distribution and $\theta = (0, 0.3, 1.2, 0.1)^t$.

It is important to notice that, at Step 2, the response of the $(j+1)$ th patient is yet to be observed, so $v^T(\hat{\theta}_j, x_{j+1})$ and $v^C(\hat{\theta}_j, x_{j+1})$ depend on the biomarker value of patient $j + 1$ but are calculated via $\hat{\theta}_j$.

Given the allocation function in (8), the experimenter is able to control the degree of randomness of the procedure through ϵ . At the extremes, the BiomARA with $\epsilon = 0$ coincides with the completely randomized design, while for $\epsilon = 0.5$ the procedure becomes deterministic.

5 Simulation results

To show the performances of the BiomARA, the present section is dedicated to its operating characteristics via an extensive simulation study. The suggested CARA procedure is implemented under several experimental scenarios corresponding to various hypothetical clinical settings, also in comparison with the permuted block design (PBD) having block size 4 and the complete randomization (CR), as standard benchmarks commonly used in practice. The patients’ biomarker X has been simulated under two different distributions: a standard normal $N(0, 1)$ and, to take into account a nonsymmetrical distribution on \mathbb{R}^+ , a log-normal (LN) distribution such that $\ln(X) \sim N(0, 1)$. For each scenario, we simulate 10,000 trials and we explore the behavior of BiomARA with $\epsilon = 0.3$ and 0.4 , denoted by BiomARA_{0.3} and BiomARA_{0.4}, respectively.

5.1 Efficiencies

To evaluate the relative efficiencies of a specific design, we consider *D* and *A* efficiency, denoted by $\epsilon_{(D)}$ and $\epsilon_{(A)}$, whose analytical expressions are reported in Appendix B. These measures are used to compare a design with the optimum in terms of the corresponding criteria, with values equal to 1 in the case of maximum efficiency (see e.g. Atkinson et al.¹⁹). Analogously, a measure of efficiency in terms of threshold estimation $\epsilon_{(x^*)}$ could be defined as the ratio of the variance in (5) calculated in the optimum (6) and its value in a generic design, that is,

$$\epsilon_{(x^*)} = \frac{4 \left\{ 1 + \frac{[M_v(x) - x^*]^2}{V_v(x)} \right\}}{\frac{M(v)}{\pi(1-\pi)M^T(v)M^C(v)} + \frac{[M_v^T(x) - x^*]^2}{\pi M^T(v)V_v^T(x)} + \frac{[M_v^C(x) - x^*]^2}{(1-\pi)M^C(v)V_v^C(x)}}$$

Adopting the optimal designs in (6) implemented via the BiomARA, we simulate the behavior of the above-mentioned efficiency measures, also including CR and PBD.

Figure 1 and the left panel of Figure 2 show the results for the standard normal distribution with $\theta^t = (0, 0.3, 1.2, 0.1)$ as n varies between 30 and 200. Whereas, Figure 3 and the right panel of Figure 2 illustrate the results under the LN distribution with $\theta^t = (-2, 0, 1.2, 0.2)$ for n between 50 and 300. In both cases, BiomARA always shows a significant gain in terms of efficiency with respect to PBD and CR. Contrary to the behavior of PBD and CR, the efficiency under BiomARA increases as n grows and, for the normal biomarker, reaches its maximum value even for very small sample sizes.

Since PBD and CR have similar performance in terms of all the measures of efficiency and in accordance with the recent literature concerning the selection of the biasing probability in the presence of covariates (see e.g. Ma and Hu²⁰), in the following simulations we omit both CR and BiomARA_{0.3}.

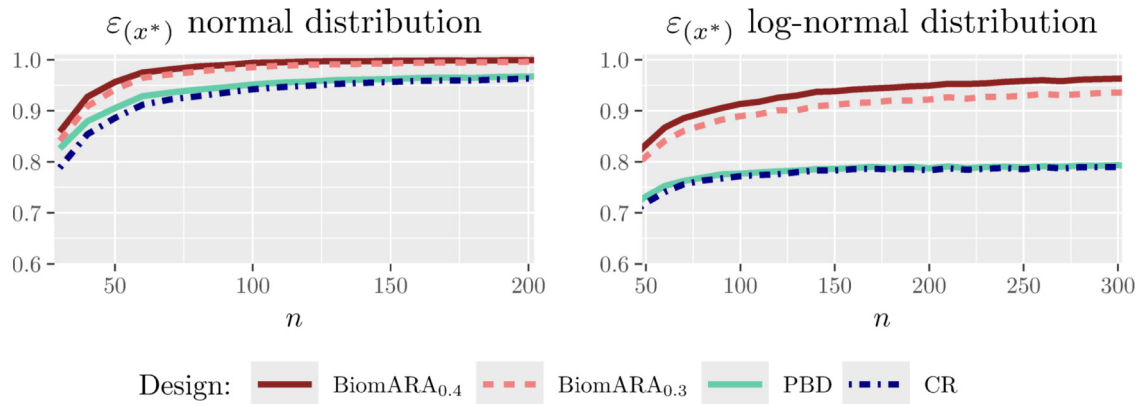


Figure 2. Efficiency of threshold estimation $\varepsilon_{(x^*)}$ when the biomarker has a standard normal distribution and $\theta = (0, 0.3, 1.2, 0.1)^t$ (left panel) and when the biomarker has a log-normal distribution and $\theta = (-2, 0, 1.2, 0.2)^t$ (right panel).

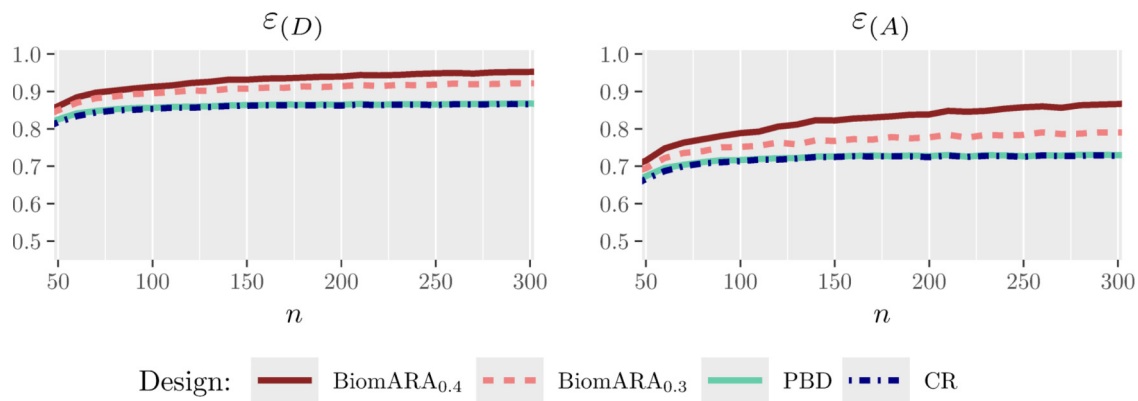


Figure 3. D and A efficiency for log-normal biomarker distribution and $\theta = (-2, 0, 1.2, 0.2)^t$.

5.2 Variance of the estimated cutoff

Focusing on inference on the threshold, we now consider five experimental scenarios (N I–N V), taking a standard normal biomarker with $n = 100$ and 200 . For each scenario (represented graphically in the left panel of Figure 4), in Table 1, we report θ , the true threshold x^* and $1 - \Phi(x^*)$, namely the probability for a subject to have biomarker value greater than x^* and so to benefit from T (since $\beta^T > \beta^C$). From N I to N V, the value of x^* decreases and thus the reported results refer to an increasing proportion of patients benefitting from T , to mimic a range of possible real clinical settings. BiomARA_{0.4} and PBD are compared in terms of the empirical variance of the estimated threshold and the corresponding efficiency. In general, $\text{var}(\hat{x}^*)$ takes smaller values under our procedure with respect to those under PBD and BiomARA shows always higher values of $\varepsilon_{(x^*)}$ that reach the maximum for $n = 200$.

Under the LN distribution for the biomarker, we take into account five different scenarios (LN I–LN V), represented graphically in the right panel of Figure 4: the results are reported in Table 2 for $n = 200$ and 300 . With regard to $\text{var}(\hat{x}^*)$, similar considerations to the normal distribution case hold, while $\varepsilon_{(x^*)}$ seems to grow slower with n , as already observed in the right panel of Figure 2.

In general, the introduced BiomARA procedure converges to the suggested optimal designs, due to the independent and identically distributed nature of the covariate process that guarantees a stabilizing behavior for large samples (by the strong law of large numbers); however, the speed of convergence seems to be strongly related to the symmetry of the covariate distribution: Figure 5 shows the decreasing behavior of $\|\hat{\lambda}_n\|$ as n grows for N II and LN II. The same graphical representation has been observed in all the other settings, omitted here for brevity, especially due to the slower convergence of the estimators in the asymmetric scenarios.

Note that, due to the variability in the estimation of θ (especially for small sample sizes or when x^* is closer to the extremes of the support of the biomarker distribution), some of the simulated trials may end with the erroneous conclusion that no threshold exists (namely the estimated threshold is outside the range of variation of the biomarker). In these cases,

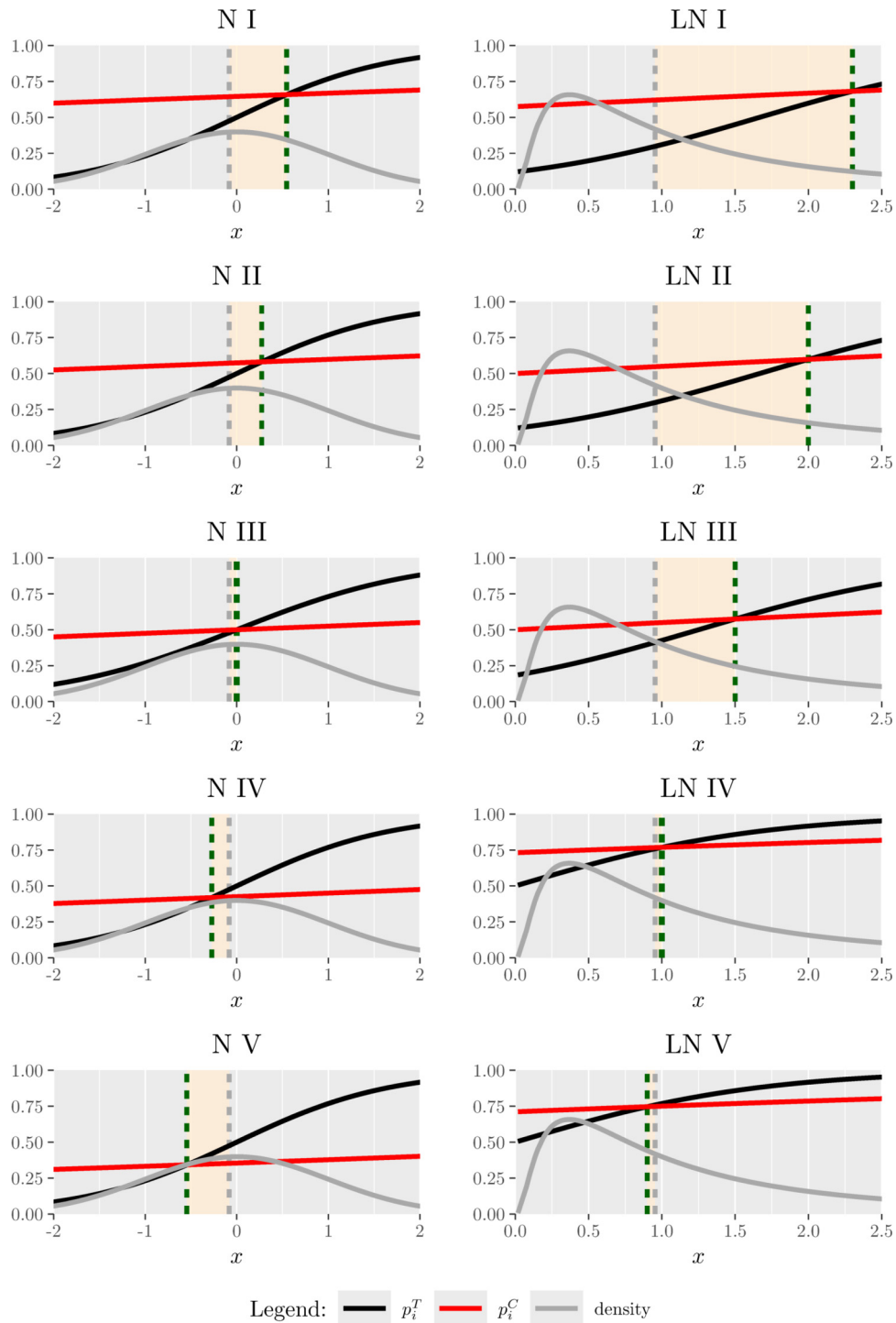


Figure 4. Biomarker density function and success probabilities of T (black curve) and C (red curve) as x varies. The displayed vertical lines correspond to the estimated median (grey dashed) and \hat{x}^* (green dashed).

the estimated proportion of patients with biomarkers greater than the threshold is one (the whole population benefits from T) or zero (nobody benefits from T). The results of Tables 1 and 2 are achieved by omitting those runs. In practice, the proportion of simulated trials leading to an erroneous conclusion depends on the number of subjects enrolled: clearly, the more patients are enrolled, the closer the estimated threshold will be to the true value, leading to a more accurate identification of the patient benefitting subpopulation. Indeed, preliminary clinical studies for identifying subpopulations

Table 1. Simulation results for standard normal biomarker with BiomARA_{0.4}.

	α^T	α^C	β^T	β^C	x^*	$1 - \Phi(x^*)$	n		$\text{var}(\hat{x}^*)$	$\epsilon_{(x^*)}$
N I	0	0.6	1.2	0.1	0.55	29%	100	BiomARA	0.27	0.99
								PBD	0.29	0.94
							200	BiomARA	0.15	1.00
								PBD	0.16	0.96
N II	0	0.3	1.2	0.1	0.27	40%	100	BiomARA	0.25	0.99
								PBD	0.27	0.95
							200	BiomARA	0.11	1.00
								PBD	0.12	0.97
N III	0	0	1	0.1	0	50%	100	BiomARA	0.35	1.00
								PBD	0.38	0.97
							200	BiomARA	0.17	1.00
								PBD	0.17	0.98
N IV	0	-0.3	1.2	0.1	-0.27	60%	100	BiomARA	0.27	0.99
								PBD	0.28	0.95
							200	BiomARA	0.12	1.00
								PBD	0.13	0.97
N V	0	-0.6	1.2	0.1	-0.55	71%	100	BiomARA	0.28	0.99
								PBD	0.29	0.93
							200	BiomARA	0.15	1.00
								PBD	0.16	0.95

N: normal; BiomARA: biomarker-adjusted response-adaptive; PBD: permuted block design.

require larger sample sizes than traditional nonpersonalized medicine trials (Mackey and Bengtsson²¹) and a metric based on this proportion has been used as a criterion to select the sample size (see Frieri et al.¹²). In the presented simulation study, we investigated only settings in which the percentage of reaching a wrong conclusion was lower than 5%. Based on this, higher sample sizes have been set for the LN distribution.

5.3 The ethical impact of discretizing a continuous biomarker

In the presence of a continuous predictive biomarker, common practice has been to take its median to discriminate among the two subpopulations of biomarker-positive (sensitive) and biomarker-negative (insensitive) patients. Nevertheless, a clinical study to identify a threshold could be a preliminary stage for later phases and should determine some guidance for future use of T and C .⁶ In practice, if the true threshold would be a priori known, we would assign the treatment only to patients belonging to the sensitive subpopulation. We perform a simulation study to emulate such a situation in order to show the ethical impact of taking the data-driven median as a threshold instead of the one suggested in this paper, calculated via the relative risk.

To this end, we take into account the results obtained in Tables 1 and 2 (for $n = 200$) as if they represent a preliminary phase through which the researcher had the opportunity to estimate both the median and the cutoff. Figure 4 describes the experimental scenarios N I–N V and LN I–LN V: the black (red) curve corresponds to the success probability of T (C)—namely p_i^T (p_i^C)—while the gray curve represents the standard normal density (left panel) and the LN density (right panel). Finally, the gray dashed line is the corresponding estimated median, while the green dashed line is \hat{x}^* (calculated by using the estimates of model parameters found in the first phase). We simulate patients' biomarkers and we assign them to T or C by using (i) the estimated median and (ii) \hat{x}^* . Then, we compute the percentage of patients receiving the inferior treatment (% IT) when using the estimated median to discriminate the subpopulations, instead of \hat{x}^* . The results are reported in Table 3. Clearly, as also suggested by Figure 4, this proportion is very small for N III and LN IV–V, since \hat{x}^* is closer to the median. However, this percentage could grow up to 20% for N and 32% for LN, stressing that discretizing a continuous biomarker could lead to very serious ethical consequences and should be avoided.

5.4 Redesign a clinical trial for severe sepsis disease

The following simulation study is intended to mimic a clinical trial for patients with severe sepsis disease. The primary outcome is status (deceased or not) at 28 days after treatment. Based on this dataset, since age is found to be a predictive biomarker, the goal is to find the age above which the success probability of the treatment is higher than the success probability of the control. The values of the age of patients, ranging from 33 to 93 years, were sampled with replacement

Table 2. Simulation results for the BiomARA_{0.4} in the case of LN distribution for the biomarker, where $F(\cdot)$ denotes its cdf.

	α^T	α^C	β^T	β^C	x^*	$1 - F(x^*)$	n		$\text{var}(\hat{x}^*)$	$\epsilon_{(x^*)}$
LN I	-2	0.5	1.2	0.2	2.5	18%	200	BiomARA	0.77	0.93
								PBD	0.91	0.78
							300	BiomARA	0.37	0.95
								PBD	0.41	0.78
LN II	-2	0	1.2	0.2	2	24%	200	BiomARA	0.29	0.95
								PBD	0.31	0.79
							300	BiomARA	0.16	0.96
								PBD	0.16	0.79
LN III	-1.5	0	1.2	0.2	1.5	34%	200	BiomARA	0.17	0.97
								PBD	0.21	0.85
							300	BiomARA	0.10	0.98
								PBD	0.11	0.86
LN IV	0	1	1.2	0.2	1	50%	200	BiomARA	0.36	0.94
								PBD	0.42	0.87
							300	BiomARA	0.24	0.95
								PBD	0.24	0.88
LN V	0	0.9	1.2	0.2	0.9	63%	200	BiomARA	0.28	0.93
								PBD	0.43	0.91
							300	BiomARA	0.18	0.94
								PBD	0.20	0.93

cdf: cumulative distribution function; LN: log-normal; BiomARA: biomarker-adjusted response-adaptive; PBD: permuted block design.

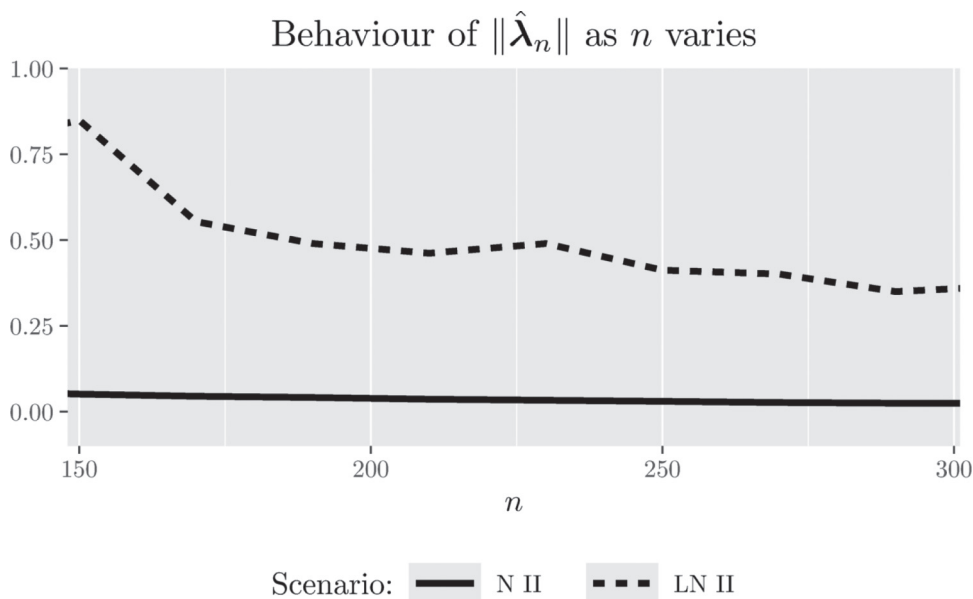


Figure 5. Behavior of $\|\hat{\lambda}_n\|$ as n increases for scenarios N II and LN II. N: normal; LN: log-normal.

from the dataset available at <https://biopharmnet.com/subgroup-analysis-software/> while patients' responses were simulated across different experimental scenarios corresponding to different underlying truths about the predictive strength of the biomarker and, consequently, the threshold. Scenario A refers to $\alpha^T = -3.74$, $\alpha^C = -1.71$, $\beta^T = 0.055$, and $\beta^C = 0.017$, obtained by fitting a logistic model to the data: in this case, the true threshold is around 53 years. Scenario B takes into account $\beta_T = 0.052$ (the corresponding threshold is 58 years). In this study, the BiomARA (with $\epsilon = 0.3$) is compared to a design strategy (Permuted Block Design and Virtual Twins method) which employs the PBD of size 4 for randomizing subjects and the Virtual Twins method¹ as a subgroup identification method (implemented by the R package `aVirtualTwins` available on CRAN). For each scenario, Table 4 reports the average number of patients in the treatment group n_T , the average number of patients with age greater than the estimated threshold $n_{\hat{x}^*}$, the average $|\hat{x}^* - x|$ as a measure

Table 3. Proportion of patients assigned to the inferior treatment if the estimated median is used instead of the estimated threshold based on the relative risk.

	% IT
N I	20%
N II	10%
N III	3%
N IV	10%
N V	20%
	% IT
LN I	32%
LN II	25%
LN III	15%
LN IV	3%
LN V	4%

% IT: percentage of patients receiving the inferior treatment; N: normal; LN: log-normal.

Table 4. Simulation results from the sepsis trial for different values of β^T and x^* with $\alpha^T = -3.74$, $\alpha^C = -1.71$, and $\beta^C = 0.017$.

	β^T	x^*	n		n_T	$n_{\hat{x}^*}$	$\text{var}(\hat{x}^*)$	$ \hat{x}^* - x $	<i>err</i>
A	0.055	53	500	BiomARA	254	314	39.2	4.8	3%
			500	PBDVT	250	194	40.1	11.8	≈ 0
			750	BiomARA	382	472	27.1	4.0	1%
			750	PBDVT	375	291	28.1	11.8	≈ 0
			1000	BiomARA	509	630	20.6	3.5	≈ 0
			1000	PBDVT	500	383	23.6	11.8	≈ 0
B	0.052	58	500	BiomARA	254	406	22.8	5.1	2%
			500	PBDVT	250	270	26.8	10.5	≈ 0
			750	BiomARA	381	613	17.3	4.1	≈ 0
			750	PBDVT	375	401	19.6	10.4	≈ 0
			1000	BiomARA	507	825	15.2	3.5	≈ 0
			1000	PBDVT	500	532	17.3	10.6	≈ 0

BiomARA: biomarker-adjusted response-adaptive; PBDVT: permuted block design and virtual twins method.

of bias and the percentage *err* of simulations in which the trial ends up with a wrong conclusion. BiomARA shows smaller values of both the variance and the bias of the threshold, while the Virtual Twins method seems to suffer from a severe bias in this context. Finally, as previously discussed, *err* is strongly sensitive to the sample size and to the closeness of the estimated threshold to the extremes of the support of the biomarker; however, due to the strong consistency of the MLEs, *err* vanishes as n increases regardless of the true threshold.

6 Discussion

In this paper, we address the complex issue of patient enrollment restriction in the presence of a continuous predictive biomarker. Despite the advances in understanding disease mechanisms and the increasing discovery of biomarkers that affect patients' responses to treatments, methodological results regarding continuous biomarker cutpoint identification and evaluation remain relatively few.

For binary response trials, in this work, we show that optimal designs for the estimation of the model parameters and the cutoff of a continuous predictive biomarker require multiple conditions to be satisfied, involving the patients' biomarker values and the unknown model parameters. Such an optimal design can be implemented sequentially by adopting the new CARA procedure suggested in Section 4. An extensive simulation study, including a redesign of a clinical trial, highlights the advantages of the proposed approach.

Our research includes an analysis aimed at pointing out the pitfalls of the common and problematic approach of adopting (empirical) median-based biomarker cutoffs. From an ethical perspective, a fundamental requirement of a clinical trial is to ensure an overall benefit for the entire sample of enrolled patients and a widely used ethical criterion focuses on

maximizing the percentage of patients who receive the best treatment. Given a prognostic biomarker (i.e. in the absence of treatment/covariate interactions), the relative performance of the treatments is the same for every subject's profile; whereas, for predictive biomarkers (namely in the presence of treatment/covariate interactions), the superiority/inferiority of a given treatment, as well as their discrepancy, depends on the subject's profiles. In this setting, discretizing continuous biomarkers is widely discouraged due to inferential and ethical repercussions (Polley and Dignam,³ Zhang and Molinaro,⁴ Bennette and Vickers,²² and Royston et al.²³). In contrast, our approach provides a cutoff identification methodology directly on a continuous scale which guarantees that more patients are assigned to the superior treatment (based on their biomarker), in line with the International Code of Medical Ethics, "a physician shall act in the patient's best interest when providing medical care." We strongly believe that the discretization of continuous biomarkers should be avoided and we encourage the researchers to abandon it.

The suggested procedure is suitable for early phase studies as it is aimed at guiding future treatment decisions for individual patients. It is worth mentioning that the approach proposed in this paper can be applied to any continuous biomarker distribution, provided that widely satisfied assumptions hold.

Some potential directions for future research have been identified. For instance, extending our approach by considering the generalized linear model family would certainly expand its applicability to a broader spectrum of clinical trial scenarios. In addition, we intend to extend our procedure to accommodate more than two treatments. Lastly, yet importantly, incorporating multiple biomarkers could make the identification of the target population more accurate.

Acknowledgments

The authors of this paper wish to thank the Editor and the referees, who made substantial comments that improved the paper.

Declaration of conflicting interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.


Funding


The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: A. Baldi Antognini, R. Frieri and M. Zagoraïou were supported by the project Funded by the European Union—NextGenerationEU through the Italian Ministry of University and Research under the National Recovery and Resilience Plan (PNRR)—Mission 4 Education and research—Component 2 From research to business—Investment 1.1 Notice Prin 2022—DD N. 104 del 2/2/2022, title [Optimal and adaptive designs for modern medical experimentation], proposal code [2022TRB44L]—CUP [J53D23003270006]. A. Baldi Antognini and S. Cecconi were supported by European Union funding within the NextGenerationEU-MUR PNRR Extended Partnership initiative on Emerging Infectious Diseases (Project No. PE00000007, INF-ACT).

ORCID iDs

Alessandro Baldi Antognini  <https://orcid.org/0000-0002-7426-8208>

Sara Cecconi  <https://orcid.org/0009-0000-1187-7252>

Rosamarie Frieri  <https://orcid.org/0000-0003-2707-183X>

Maroussa Zagoraïou  <https://orcid.org/0000-0002-8461-8240>

References

1. Foster J, Taylor J and Ruberg S. Subgroup identification from randomized clinical trial data. *Stat Med* 2011; **30**: 2867–2880.
2. Loh W, Cao L and Zhou P. Subgroup identification for precision medicine: A comparative review of 13 methods. *Wiley Interdiscip Rev: WIREs Data Min Knowl Discov* 2019; **9**: e1326.
3. Polley MYC and Dignam JJ. Statistical considerations in the evaluation of continuous biomarkers. *J Nucl Med* 2021; **62**: 605–611.
4. Zhang Y and Molinaro AM. Categorizing continuous biomarkers: More cons than pros. *Neurooncol Pract* 2022; **9**: 81–82.
5. Jiang W, Freidlin B and Simon R. Biomarker adaptive threshold design: a procedure for evaluating treatment with possible biomarker-defined subset effect. *J Natl Cancer Inst* 2007; **99**: 1036–1043.
6. Baldi Antognini A, Frieri R and Zagoraïou M. New insights into adaptive enrichment designs. *Stat Pap* 2023; **64**: 1305–1328.
7. Spencer A, Harbron C, Mander A, et al. An adaptive design for updating the threshold value of a continuous biomarker. *Stat Med* 2016; **35**: 4909–4923.
8. Trippa L, Lee EQ, Wen PY, et al. Bayesian adaptive randomized trial design for patients with recurrent glioblastoma. *J Clin Oncol* 2012; **30**: 3258.
9. Renfro LA, Coughlin CM, Grothey AM, et al. Adaptive randomized phase II design for biomarker threshold selection and independent evaluation. *Chin Clin Oncol* 2014; **3**: 3489.
10. Diao G, Dong J, Zeng D, et al. Biomarker threshold adaptive designs for survival endpoints. *J Biopharm Stat* 2018; **28**: 1038–1054.
11. Lin Z, Flournoy N and Rosenberger W. Inference for a two-stage enrichment design. *Ann Stat* 2021; **49**: 2697–2720.

12. Frieri R, Rosenberger W, Flournoy N, et al. Design considerations for two-stage enrichment trial. *Biometrics* 2022; **79**: 2565–2576.
13. Baldi Antognini A, Frieri R WRF, et al. Optimal design for inference on the threshold of a biomarker. *Stat Methods Med Res* 2024; **33**: 321–343.
14. Margolese R, Cecchini R, Julian T, et al. Anastrozole versus tamoxifen in postmenopausal women with ductal carcinoma in situ undergoing lumpectomy plus radiotherapy (NSABP B-35): a randomised, double-blind, phase 3 clinical trial. *Lancet* 2016; **387**: 849–856.
15. Vinnat V and Chevret S. Enrichment bayesian design for randomized clinical trials using categorical biomarkers and a binary outcome. *BMC Med Res Methodol* 2022; **22**: 54.
16. Spreafico A, Hansen A, Abdul Razak A, et al. The future of clinical trial design in oncology. *Cancer Discov* 2021; **11**: 822–837.
17. Azoulay E, Lemiale V, Mokart D, et al. Effect of high-flow nasal oxygen vs standard oxygen on 28-day mortality in immunocompromised patients with acute respiratory failure: The HIGH randomized clinical trial. *JAMA* 2018; **320**: 2099–2107.
18. Baldi Antognini A and Giovagnoli A. *Adaptive designs for sequential treatment allocation*. Boca Raton, FL: Chapman & Hall/CRC, 2015.
19. Atkinson AC, Donev AN and Tobias R. *Optimum experimental designs, with SAS*. Oxford: Oxford University Press, 2007.
20. Ma Z and Hu F. Balancing continuous covariates based on kernel densities. *Contemp Clin Trials* 2013; **34**: 262–269.
21. Mackey H and Bengtsson T. Sample size and threshold estimation for clinical trials with predictive biomarkers. *Contemp Clin Trials* 2013; **36**: 664–672.
22. Bennette C and Vickers A. Against quantiles: categorization of continuous variables in epidemiologic research, and its discontents. *BMC Med Res Methodol* 2012; **12**: 1–5.
23. Royston P, Altman D and Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med* 2006; **25**: 127–141.

A. Appendix

A.1 Fisher information matrix

As is well-known, the Fisher information matrix \mathbf{I}_n and the average (per observation) information $\mathbf{M}_n = n^{-1}\mathbf{I}_n$ play a crucial role in the theory of design of experiments since, under the commonly assumed models, the inverse \mathbf{M}_n^{-1} asymptotically approximates the variance–covariance matrix of the MLEs of the unknown parameters. Thus, many of the so-called alpha-betical criteria (e.g. A , D , and E optimality) are simply defined as convex functions of \mathbf{M}_n^{-1} , which should be minimized in order to achieve the minimum possible loss of information in an experiment. In this Appendix, by using standard derivation techniques of the log-likelihood function we firstly derive the average (per observation) information matrix \mathbf{M}_n in (2) for the logistic regression model, while the expression of its inverse will be provided in Appendix 2.

Under model (1), the log-likelihood is

$$\begin{aligned} \ell(\boldsymbol{\theta}; y) &= \sum_{i=1}^n \left\{ \delta_i y_i \ln \left(\frac{p_i^T}{1 - p_i^T} \right) + (1 - \delta_i) y_i \ln \left(\frac{p_i^C}{1 - p_i^C} \right) - \ln \left(1 + e^{\delta_i(\alpha^T + \beta^T x_i) + (1 - \delta_i)(\alpha^C + \beta^C x_i)} \right) \right\} \\ &= \sum_{i=1}^n \delta_i y_i (\alpha^T + \beta^T x_i) + \sum_{i=1}^n (1 - \delta_i) y_i (\alpha^C + \beta^C x_i) - \sum_{i=1}^n \ln \left(1 + e^{\delta_i(\alpha^T + \beta^T x_i) + (1 - \delta_i)(\alpha^C + \beta^C x_i)} \right) \end{aligned}$$

Let $p_i = \delta_i p_i^T + (1 - \delta_i) p_i^C$ and $\eta_i = \delta_i \eta_i^T + (1 - \delta_i) \eta_i^C$ then $p_i = e^{\eta_i} / (1 + e^{\eta_i})$, namely $\eta_i = \ln(p_i / (1 - p_i))$, and $p_i(1 - p_i) = e^{\eta_i} / (1 + e^{\eta_i})^2$. Since $\delta_i p_i = \delta_i p_i^T$ and $(1 - \delta_i) p_i = (1 - \delta_i) p_i^C$, then

$$\frac{\partial}{\partial \alpha^T} \ln(1 + e^{\eta_i}) = \frac{\delta_i e^{\eta_i}}{1 + e^{\eta_i}} = \delta_i p_i \quad \text{and} \quad \frac{\partial}{\partial \beta^T} \ln(1 + e^{\eta_i}) = \frac{\delta_i x_i e^{\eta_i}}{1 + e^{\eta_i}} = \delta_i x_i p_i$$

(the derivatives wrt α^C and β^C have the same expressions with δ_i replaced by $1 - \delta_i$), obtaining

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta}; y)}{\partial \alpha^T} &= \sum_{i=1}^n \delta_i (y_i - p_i), & \frac{\partial \ell(\boldsymbol{\theta}; y)}{\partial \alpha^C} &= \sum_{i=1}^n (1 - \delta_i) (y_i - p_i) \\ \frac{\partial \ell(\boldsymbol{\theta}; y)}{\partial \beta^T} &= \sum_{i=1}^n \delta_i (y_i - p_i) x_i, & \frac{\partial \ell(\boldsymbol{\theta}; y)}{\partial \beta^C} &= \sum_{i=1}^n (1 - \delta_i) (y_i - p_i) x_i \end{aligned}$$

Since

$$\frac{\partial p_i}{\partial \alpha^T} = \frac{\delta_i e^{\eta_i}}{(1 + e^{\eta_i})^2} = \delta_i p_i (1 - p_i) \quad \text{and} \quad \frac{\partial p_i}{\partial \beta^T} = \frac{\delta_i e^{\eta_i} x_i}{(1 + e^{\eta_i})^2} = \delta_i p_i (1 - p_i) x_i$$

we have

- $\frac{\partial}{\partial \alpha^T} \left(\frac{\partial \ell(\theta; \mathbf{y})}{\partial \alpha^T} \right) = - \sum_{i=1}^n \delta_i p_i (1 - p_i) = - \sum_{i=1}^n \delta_i p_i^T (1 - p_i^T)$
- $\frac{\partial}{\partial \alpha^C} \left(\frac{\partial \ell(\theta; \mathbf{y})}{\partial \alpha^C} \right) = - \sum_{i=1}^n (1 - \delta_i) p_i (1 - p_i) = - \sum_{i=1}^n (1 - \delta_i) p_i^C (1 - p_i^C)$
- $\frac{\partial}{\partial \beta^T} \left(\frac{\partial \ell(\theta; \mathbf{y})}{\partial \beta^T} \right) = - \sum_{i=1}^n \delta_i p_i (1 - p_i) x_i^2 = - \sum_{i=1}^n \delta_i p_i^T (1 - p_i^T) x_i^2$
- $\frac{\partial}{\partial \beta^C} \left(\frac{\partial \ell(\theta; \mathbf{y})}{\partial \beta^C} \right) = - \sum_{i=1}^n (1 - \delta_i) p_i (1 - p_i) x_i^2 = - \sum_{i=1}^n (1 - \delta_i) p_i^C (1 - p_i^C) x_i^2$
- $\frac{\partial}{\partial \alpha^T} \left(\frac{\partial \ell(\theta; \mathbf{y})}{\partial \beta^T} \right) = - \sum_{i=1}^n \delta_i p_i (1 - p_i) x_i = - \sum_{i=1}^n \delta_i p_i^T (1 - p_i^T) x_i$
- $\frac{\partial}{\partial \alpha^C} \left(\frac{\partial \ell(\theta; \mathbf{y})}{\partial \beta^C} \right) = - \sum_{i=1}^n (1 - \delta_i) p_i (1 - p_i) x_i = - \sum_{i=1}^n (1 - \delta_i) p_i^C (1 - p_i^C) x_i$

and, clearly, all the remaining second-order mixed partial derivatives are equal to zero.

A.2 Inverse of the average Fisher information matrix

In what follows we omit the subscript n for the sake of simplicity. By defining $\mathbf{P} = \text{diag}(\pi, 1 - \pi)$, $\mathbf{H} = \text{diag}(M^T(v), M^C(v))$, $\mathbf{K} = \text{diag}(M_v^T(x), M_v^C(x))$, and $\mathbf{L} = \text{diag}(M_v^T(x^2), M_v^C(x^2))$, then the average Fisher information matrix \mathbf{M} in (2) can be rewritten as

$$\mathbf{M} = \text{diag}(\mathbf{P}, \mathbf{P}) \text{diag}(\mathbf{H}, \mathbf{H}) \left(\begin{array}{c|c} \mathbb{I}_2 & \mathbf{K} \\ \hline \mathbf{K} & \mathbf{L} \end{array} \right)$$

where \mathbb{I}_2 denotes the 2-dimensional identity matrix. Let $\mathbf{V} = \text{diag}(V_v^T(x), V_v^C(x)) = \mathbf{L} - \mathbf{K}^2$, then $(\mathbb{I}_2 - \mathbf{K}^2 \mathbf{L}^{-1})^{-1} = \mathbf{V}^{-1} \mathbf{L}$ and $\mathbf{L}^{-1} + \mathbf{L}^{-1} \mathbf{K}^2 \mathbf{V}^{-1} \mathbf{L} \mathbf{L}^{-1} = \mathbf{V}^{-1}$; therefore,

$$\begin{aligned} \mathbf{M}^{-1} &= \text{diag}(\mathbf{P}^{-1}, \mathbf{P}^{-1}) \text{diag}(\mathbf{H}^{-1}, \mathbf{H}^{-1}) \left(\begin{array}{c|c} \mathbf{V}^{-1} \mathbf{L} & -\mathbf{V}^{-1} \mathbf{K} \\ \hline -\mathbf{V}^{-1} \mathbf{K} & \mathbf{V}^{-1} \end{array} \right) = (\mathbf{P} \mathbf{H} \mathbf{V})^{-1} \otimes \left(\begin{array}{c|c} \mathbf{L} & -\mathbf{K} \\ \hline -\mathbf{K} & \mathbb{I}_2 \end{array} \right) \\ &= \left(\begin{array}{cccc} \frac{M_v^T(x^2)}{\pi M^T(v) V_v^T(x)} & 0 & -\frac{M_v^T(x)}{\pi M^T(v) V_v^T(x)} & 0 \\ 0 & \frac{M_v^C(x^2)}{(1 - \pi) M^C(v) V_v^C(x)} & 0 & -\frac{M_v^C(x)}{(1 - \pi) M^C(v) V_v^C(x)} \\ -\frac{M_v^T(x)}{\pi M^T(v) V_v^T(x)} & 0 & \frac{1}{\pi M^T(v) V_v^T(x)} & 0 \\ 0 & -\frac{M_v^C(x)}{(1 - \pi) M^C(v) V_v^C(x)} & 0 & -\frac{1}{(1 - \pi) M^C(v) V_v^C(x)} \end{array} \right) \end{aligned}$$

A.3 Optimality criteria

By noticing that $\mathbf{M}^{-1} = \mathbf{A} \mathbf{B}$, with

$$\mathbf{A} = \text{diag} \left[\frac{1}{\pi M^T(v) V_v^T(x)}, \frac{1}{(1 - \pi) M^C(v) V_v^C(x)}, \frac{1}{\pi M^T(v) V_v^T(x)}, \frac{1}{(1 - \pi) M^C(v) V_v^C(x)} \right]$$

and

$$\mathbf{B} = \left(\begin{array}{c|c} \mathbf{L} & -\mathbf{K} \\ \hline -\mathbf{K} & \mathbb{I}_2 \end{array} \right)$$

the D optimality criterion in (3) follows easily since $\det \mathbf{B} = \det(\mathbf{L} - \mathbf{K}^2) = V_v^T(x) V_v^C(x)$ and $\det \mathbf{A} = [\pi M^T(v) V_v^T(x)]^{-2} [(1 - \pi) M^C(v) V_v^C(x)]^{-2}$. The proof for the A optimality criterion in (4) is straightforward.

As regards the threshold estimation, let $\alpha = \alpha^T - \alpha^C$ and $\beta = \beta^T - \beta^C$, then $\hat{\alpha} = \hat{\alpha}^T - \hat{\alpha}^C$ and $\hat{\beta} = \hat{\beta}^T - \hat{\beta}^C$. By letting

$$\mathbf{J}^t = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

then

$$\begin{aligned} \text{var}(\hat{\alpha}; \hat{\beta}) &= \frac{1}{n} \mathbf{J}^t \mathbf{M}^{-1} \mathbf{J} \\ &= \frac{1}{n} \begin{pmatrix} \frac{M_v^T(x^2)}{\pi M^T(v) V_v^T(x)} + \frac{M_v^C(x^2)}{(1-\pi) M^C(v) V_v^C(x)} & -\frac{M_v^T(x)}{\pi M^T(v) V_v^T(x)} - \frac{M_v^C(x)}{(1-\pi) M^C(v) V_v^C(x)} \\ -\frac{M_v^T(x)}{\pi M^T(v) V_v^T(x)} - \frac{M_v^C(x)}{(1-\pi) M^C(v) V_v^C(x)} & \frac{1}{\pi M^T(v) V_v^T(x)} + \frac{1}{(1-\pi) M^C(v) V_v^C(x)} \end{pmatrix} \end{aligned}$$

and, via the classical first-order Taylor expansion,

$$\text{var}(\hat{x}^*) \approx \frac{1}{\beta^2} \left[\text{var}(\hat{\alpha}) - \frac{2\alpha \text{Cov}(\hat{\alpha}, \hat{\beta})}{\beta} + \frac{\alpha^2 \text{var}(\hat{\beta})}{\beta^2} \right] \quad (9)$$

Let us define

$$\omega = \frac{\pi M^T(v)}{M(v)} \quad (10)$$

and recalling the expression in (7), then $\omega M(v) = \pi M^T(v)$ and $(1-\omega)M(v) = (1-\pi)M^C(v)$, so that expression (1) can be written as

$$\begin{aligned} \text{var}(\hat{x}^*) &\approx \frac{1}{\beta^2 M(v)} \left\{ \frac{1}{\omega V_v^T(x)} \left[M_v^T(x^2) + \frac{2\alpha}{\beta} M_v^T(x) + \frac{\alpha^2}{\beta^2} \right] + \frac{1}{(1-\omega) V_v^C(x)} \left[M_v^C(x^2) + \frac{2\alpha}{\beta} M_v^C(x) + \frac{\alpha^2}{\beta^2} \right] \right\} \\ &= \frac{1}{\beta^2 M(v)} \left\{ \frac{1}{\omega(1-\omega)} + \frac{\left[M_v^T(x) + \frac{\alpha}{\beta} \right]^2}{\omega V_v^T(x)} + \frac{\left[M_v^C(x) + \frac{\alpha}{\beta} \right]^2}{(1-\omega) V_v^C(x)} \right\} \end{aligned}$$

A.4 Proof of Theorem 3.2

We now show that the allocation in (6) is optimal with respect to any convex criterion Ψ of \mathbf{M} , thus for criteria (3) and (4), as well as for (5). Let us denote by \mathbf{M}^* the average information matrix under the optimal design in (6) as

$$\mathbf{M}^* = \frac{M(v)}{2} \begin{pmatrix} \mathbb{1}_2 & M_v(x) \mathbb{1}_2 \\ M_v(x) \mathbb{1}_2 & M_v(x^2) \mathbb{1}_2 \end{pmatrix}$$

Note that \mathbf{M}^* is invariant with respect to the label switching of T and C . Consider also that for any generic design associated with an information matrix \mathbf{M} in (2), by simultaneously permutating the first two rows and columns and the bottom two rows and the two right-hand columns of \mathbf{M} , the information matrix $\tilde{\mathbf{M}}$ is obtained, which corresponds to the design that switches T and C . We have that $\Psi(\tilde{\mathbf{M}}) = \Psi(\mathbf{M})$ and, from (7), $\mathbf{M} + \tilde{\mathbf{M}} = 2\mathbf{M}^*$, so that by convexity

$$\Psi(\mathbf{M}^*) = \Psi(2^{-1} [\mathbf{M} + \tilde{\mathbf{M}}]) \leq 2^{-1} [\Psi(\mathbf{M}) + \Psi(\tilde{\mathbf{M}})] = \Psi(\mathbf{M})$$

Notice that every allocation δ_n^* satisfying (6) induces the same value of $M(v)$, $M_v(x)$ and $M_v(x^2)$. In addition, by letting $\tilde{\mathbf{b}}^t = (1, -1, -\alpha/\beta, \alpha/\beta)$ then $\beta^{-2} \tilde{\mathbf{b}}^t \mathbf{M}^{-1} \tilde{\mathbf{b}}$ = $\text{var}(\hat{x}^*)$, (i.e. criterion in (5)) is a convex function, invariant for label switching of the two treatments. Thus, the optimal design for estimating the threshold is still the one in (6).

B. Efficiencies

Now we derive the relative efficiency of a generic design δ compared to the optimal one in (6). Since $\det \text{var}(\hat{\theta}|\delta^*) = 16/[M(v)]^4 V_v^2(x)$ the D efficiency is

$$\varepsilon_{(D)} = \left[\frac{\det \text{var}(\hat{\theta}|\delta^*)}{\det \text{var}(\hat{\theta}|\delta)} \right]^{1/4} = 2 \left[\frac{\omega^2(1-\omega)^2 V_v^T(x) V_v^C(x)}{V_v^2(x)} \right]^{1/4}$$

Since $\text{tr} \text{var}(\hat{\theta}|\delta^*) = 4[V_v(x) + M_v(x)^2 + 1]/M(v)V_v(x)$ the A efficiency is

$$\varepsilon_{(A)} = \frac{\text{tr} \text{var}(\hat{\theta}|\delta^*)}{\text{tr} \text{var}(\hat{\theta}|\delta)} = \frac{\frac{4[V_v(x)+M_v(x)^2+1]}{V_v(x)}}{\frac{V_v^T(x)+M_v^T(x)^2+1}{\omega V_v^T(x)} + \frac{V_v^C(x)+M_v^C(x)^2+1}{(1-\omega)V_v^C(x)}}$$

As regards the threshold estimation,

$$\text{var}(\hat{x}^*|\delta^*) = \frac{4}{\beta^2 M(v)} \left\{ 1 + \frac{[M_v(x) + \frac{\alpha}{\beta}]^2}{V_v(x)} \right\}$$

so that the corresponding efficiency is

$$\varepsilon_{(x^*)} = \frac{\text{var}(\hat{x}^*|\delta^*)}{\text{var}(\hat{x}^*|\delta)} = \frac{4 \left\{ 1 + \frac{[M_v(x) + \frac{\alpha}{\beta}]^2}{V_v(x)} \right\}}{\frac{1}{\omega(1-\omega)} + \frac{[M_v^T(x) + \frac{\alpha}{\beta}]^2}{\omega V_v^T(x)} + \frac{[M_v^C(x) + \frac{\alpha}{\beta}]^2}{(1-\omega)V_v^C(x)}}$$