Robustness of the Sequential Efficient Design for Identifying a Target Subpopulation

Robustness of the Sequential Efficient Design for identifying a target subpopulation

Sara Cecconi[1], Rosamarie Frieri[1], Maroussa Zagoraiou[1], and Alessandro Baldi Antognini[1]

Department of Statistical Sciences, University of Bologna; sara.cecconi4@unibo.it, rosamarie.frieri2@unibo.it, maroussa.zagoraiou@unibo.it, a.baldi@unibo.it

Abstract. Precision medicine is an innovative approach for tailoring treatments based on individual characteristics or biomarkers. Enrichment is a main strategy in the development of clinical trials for precision medicine as it "is the prospective use of any patient characteristic to select a study population in which detection of a drug effect (if one is, in fact, present) is more likely than it would be in an unselected population." (FDA, 2019) [1]. Although a binary biomarker can be used to stratify the subjects, the situation is more complex if a continuous biomarker is associated with patient responses. Recently Baldi Antognini et al. [2] provided the optimal allocations for inference on the threshold of a continuous biomarker and proposed a new Covariate-Adaptive procedure, called the Sequential Efficient Design (SED), to implement these designs sequentially. In this paper we push forward the results by exploring the robustness of the SED under possible deviation from the linearity assumption of the response-biomarker relationship, taking also into account comparisons with the permuted block design.

Keywords: threshold, continuous biomarker, covariate-adaptive randomization, unbalanced allocation

1 Introduction

Adapting to genetic and biomarker-induced heterogeneity has become imperative in contemporary clinical trials, shifting their traditional focus from determining average treatment effects across all patients. The identification of a beneficial subpopulation is now a critical aspect of precision medicine. Strategies for handling continuous biomarkers include discretization ([3, 4]), setting cutoffs equal to a quantile of the biomarker distribution [5], and direct threshold estimation ([2, 6]), with our focus on the latter. In enrichment designs, originated in oncology [7], an initial full-population phase is conducted to determine a biomarker threshold and, in a subsequent phase, the benefitting subgroup is enrolled. Despite potential benefits, the complex experimental framework raises challenges, questioning the worth of enrichment [8]. Optimal decision-making for updating enrollment criteria, especially when estimating continuous biomarker thresholds, is a major concern. Understanding the biomarker-treatment relationship is crucial before driving enrichment study designs. A classical approach is to adopt

the linear treatment-by-covariate interaction with heteroscedastic outcomes. The aim of this paper is to study the robustness of SED in terms of threshold estimation, also in comparison with the well-known Permuted Block Design (PBD) under possible non-linear response-biomarker relationships.

2 Model and biomarker threshold

Consider a clinical trial where n statistical units are sequentially assigned to one of two competing treatments A or B. As the i-th patient is ready to be randomized, her/his biomarker value $X_i = x_i$ is recorded and she/he is assigned to treatment A or B based on a given randomization rule, with $\delta_i = 1$ or $\delta_i = 0$ if the patient is assigned to A or B, respectively. Then, her/his response Y_i is recorded. We assume that the biomarker X is quantitative, not under the experimenters' control and $X_1, X_2, \dots$ are i.i.d. random variables with finite variance. After n assignments, let $\boldsymbol{Y}_n = (Y_1, \dots, Y_n)^t$, $\mathbf{x}_n = (x_1, \dots, x_n)^t$, $\boldsymbol{\delta}_n = (\delta_1, \dots, \delta_n)^t$, $\mathbf{1}_n$ the n-dim vector of ones, and $\mathbf{G}_n = [\boldsymbol{\delta}_n : \mathbf{1}_n - \boldsymbol{\delta}_n : \text{diag}(\boldsymbol{\delta}_n)\mathbf{x}_n : \text{diag}(\mathbf{1}_n - \boldsymbol{\delta}_n)\mathbf{x}_n]$. Conditionally on the treatments and the biomarker, subjects' responses are assumed to be independent following

$$E(\mathbf{Y}_n|\boldsymbol{\delta}_n, \mathbf{x}_n) = \mathbf{G}_n\boldsymbol{\zeta}, \quad \text{var}(\mathbf{Y}_n|\boldsymbol{\delta}_n, \mathbf{x}_n) = \boldsymbol{\Sigma}_n, \tag{1}$$

where $\boldsymbol{\zeta}^t = (\mu_A, \mu_B, \beta_A, \beta_B)$ denotes the parameter of interest, μ_A and μ_B the baseline treatment effects and β_A and β_B the possibly different regression parameters related to A and B, respectively, and $\boldsymbol{\Sigma}_n = \text{diag}\left(\delta_i\sigma_A^2 + (1-\delta_i)\sigma_B^2\right)_{i=1,\dots,n}$, with σ_A^2 and σ_B^2 being the variances of the responses in the two groups. Under this setting, if $(\mathbf{G}_n^t\boldsymbol{\Sigma}_n^{-1}\mathbf{G}_n)^{-1}$ exists, then the estimator of $\boldsymbol{\zeta}$ is $\hat{\boldsymbol{\zeta}}_n = (\mathbf{G}_n^t\boldsymbol{\Sigma}_n^{-1}\mathbf{G}_n)^{-1}\mathbf{G}_n^t\boldsymbol{\Sigma}_n^{-1}\mathbf{Y}_n$ and $\text{var}(\hat{\boldsymbol{\zeta}}_n) = (\mathbf{G}_n^t\boldsymbol{\Sigma}_n^{-1}\mathbf{G}_n)^{-1} = n^{-1}\mathbf{M}^{-1}$, where $\mathbf{M}$ is the average information matrix.

Clearly, for each value of x, the superiority/inferiority of A with respect to B depends on the sign of $\mu_A + x\beta_A - (\mu_B + x\beta_B)$: assuming that highest responses are preferable, A is better than B if the sign is positive. Under this setting, the true biomarker threshold - that delineates the benefitting subpopulation - is given by $x^* = -(\mu_A - \mu_B)/(\beta_A - \beta_B)$. Note that in the absence of treatment-by-biomarker interaction, for every x, $\beta_A = \beta_B$; thus, the relative performance of the treatments is the same and it is equal to the baseline treatment difference $\mu_A - \mu_B$.

Focusing on inference on the threshold, Baldi Antognini et al. [2] derived the optimal design that minimizes the variance of the estimated threshold $\hat{x}^*$, also maximizing the power of the Wald test for testing $H_0 : \beta_A = \beta_B$ versus $H_1 : \beta_A \neq \beta_B$. This optimal design combines the well-known Neyman allocation $\pi_{\text{Neyman}} = \sigma_A/(\sigma_A + \sigma_B)$ with the following conditions

$$m_A(x^i) = m_B(x^i) \quad \text{for} \quad i = 1, 2, \tag{2}$$

where $m_A(x) = \boldsymbol{\delta}_n^t\mathbf{x}_n/\boldsymbol{\delta}_n^t\mathbf{1}_n$ and $m_B(x) = (\mathbf{1}_n - \boldsymbol{\delta}_n)^t\mathbf{x}_n/(n - \boldsymbol{\delta}_n^t\mathbf{1}_n)$ denote the biomarker's empirical means in the two groups, while $m_A(x^2) = \mathbf{x}_n^t\text{diag}(\boldsymbol{\delta}_n)\mathbf{x}_n/\boldsymbol{\delta}_n^t\mathbf{1}_n$ and $m_B(x^2) = \mathbf{x}_n^t\text{diag}(\mathbf{1}_n - \boldsymbol{\delta}_n)\mathbf{x}_n/(n - \boldsymbol{\delta}_n^t\mathbf{1}_n)$ represent the second moments.

3 SED and simulation study

Since the patients' biomarker values are unknown at the beginning of the trial, the optimal design can be implemented in a sequential fashion by forcing, at each step, the treatment allocation to approach the optimum asymptotically. When the treatment variances σ_j^2 $(j = A, B)$ are a-priori known (and thus π_{Neyman}), Baldi Antognini et al. [2] introduced the SED, namely a new Covariate-Adaptive randomization procedure based on the sequential minimization of the Euclidean norm of the following imbalance vector

$$\boldsymbol{b}_n^t = n[\pi_n - \pi_{\text{Neyman}}, \pi_n m_{A_n}(x) - \pi_{\text{Neyman}} m_n(x), \pi_n m_{A_n}(x^2) - \pi_{\text{Neyman}} m_n(x^2)],$$

where the subscript n denotes that the corresponding quantities are evaluated at step n, $\pi_n = n^{-1}\boldsymbol{\delta}_n^t \mathbf{1}_n$ is the proportion of patients allocated to A and $m_n(x) = n^{-1}\mathbf{x}_n^t \mathbf{1}_n$, $m_n(x^2) = n^{-1}\mathbf{x}_n^t \mathbf{x}_n$ are fixed quantities depending on the enrolled patients. When the $(n+1)$th patient is enrolled and her/his biomarker value x_{n+1} is observed, we can calculate the potential imbalance achieved by the procedure by assigning such subject to A or B, i.e., $\boldsymbol{b}_{n+1|A}$ or $\boldsymbol{b}_{n+1|B}$, respectively. Adopting SED, the allocation probability of $(n+1)$th patient to A is

$$Pr(\delta_{n+1} = 1 | \boldsymbol{\delta}_n, \mathbf{x}_n, x_{n+1}) = \begin{cases} \pi_{\text{Neyman}} + \varepsilon, & \text{if } \|\boldsymbol{b}_{n+1|A}\|^2 < \|\boldsymbol{b}_{n+1|B}\|^2 \\ \pi_{\text{Neyman}}, & \text{if } \|\boldsymbol{b}_{n+1|A}\|^2 = \|\boldsymbol{b}_{n+1|B}\|^2 \\ \pi_{\text{Neyman}} - \varepsilon, & \text{if } \|\boldsymbol{b}_{n+1|A}\|^2 > \|\boldsymbol{b}_{n+1|B}\|^2 \end{cases}$$

where $0 < \varepsilon < \min\{\pi_{\text{Neyman}}; 1 - \pi_{\text{Neyman}}\}$ is a skewing parameter controlling the degree of randomness. We run a simulation study to investigate the effect of possible misspecifications of the response-biomarker relationship in treatment group A, while the expected outcome for treatment B is considered to be linear. Assuming standard normal biomarker distribution, we consider two cases of non-linear response-biomarker relationship, as displayed in Figure 1. The true data-generating model for responses is given by a blue line for treatment A and a red line for B and the true threshold (blue curve and red line intersection point) is denoted by x^*. In particular:
Case 1. $E[Y|\delta = 1, X = x] = 0.5 + x + 0.7e^x$; $E[Y|\delta = 0, X = x] = 1.7 + 0.3x$
Case 2. $E[Y|\delta = 1, X = x] = 1.5 + x + 0.3x^3$; $E[Y|\delta = 0, X = x] = 2.2 + 0.3x$.
The simulation study takes into account two types of model misspecification. In **Scenario 1** the trial is conducted believing (erroneously) that the response-biomarker relationship is linear. Thus, at the end, a linear model is fitted to the data (black line in Figure 1) and the intersection point with the red line x_ℓ^* is found by sampling 10000 draws from the standard normal distribution to fit the blue curve. In **Scenario 2** the trial is conducted believing that the response-biomarker relationship is not linear (with an exponential term in case 1 and a cubic term in case 2). Thus the correct model is fitted to the data. The metrics of interest, x_ℓ^* and x^*, are reported in Table 1 with the corresponding probability of belonging to the target population (i.e., with biomarker value above the threshold), denoted by ξ_ℓ and ξ.

Fig. 1: Relationship between treatment responses and biomarker.

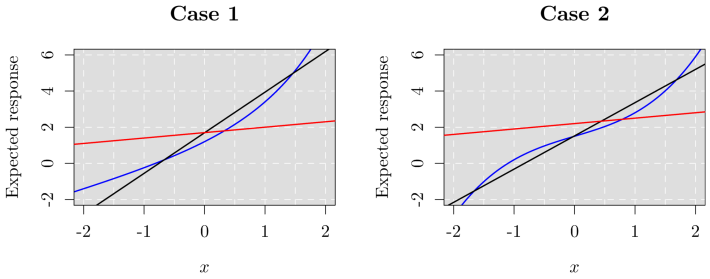

Table 1: Threshold and probability of belonging to the target population for the two cases considered.

	x_ℓ^*	x^*	ξ_ℓ	ξ	$x^* - x_\ell^*$
Case 1	0.028	0.327	0.489	0.372	0.299
Case 2	0.437	0.789	0.331	0.215	0.352

All trials have been simulated 20000 times assuming standard normal biomarker distribution and $n = 30, 50, 70, 100$. We set $\sigma_A = 3$ and $\sigma_B = 1$, so that $\pi_{\text{Neyman}} = 3/4$, and $\epsilon = 0.15$. We compare the performance of the SED with those of the PBD with block size 4 targeting π_{Neyman} in scenario 1 (Table 2) and scenario 2 (Table 3). In Table 2 we report the estimated threshold based on the linear model $\hat{x}_\ell^*$, the estimated ξ_ℓ, i.e., $\hat{\xi}_\ell$, the simulated average variance in estimating the threshold and $E(\hat{x}_\ell^* - x^*)^2$. The SED presents lower values of the variance and $E(\hat{x}_\ell^* - x^*)^2$ with respect to the PBD with differences that tend to diminish for higher values of the sample size. The results for scenario 2 are in Table 3, where $\hat{x}_m^*$ is an estimate of x^* and $\hat{\xi}_m$ is the corresponding proportion of patients in the target population. Accordingly, we report the empirical variance of $\hat{x}_m^*$ and $E(\hat{x}_m^* - x^*)^2$. When fitting the correct model, the estimated threshold is much closer to the true value x^*, with the SED having lower values of the variance and $E(\hat{x}_m^* - x^*)^2$ with respect to the PBD. Table 2 shows that when fitting the data with the linear model, the values of $\hat{x}_\ell^*$ do not present a monotonic behavior for n up to 100. Whereas, when the correct model is fitted, results in Table 3 show that $\hat{x}_m^*$ monotonically increases towards the true x^*, as n grows. Note that with a larger sample size (as shown by further simulations omitted for brevity), $\hat{x}_\ell^*$ tends to x_ℓ^*, which is still different from x^* (see Table 1).

In both Tables 2 and 3, we report γ, i.e., the proportion of times in which $\hat{\xi}_\ell$ or $\hat{\xi}_m = 0$ or 1: in these cases the threshold cannot be identified based on the simulated data (so all the other metrics in Table 2 and 3 have been calculated removing those trials). Clearly, the role of γ becomes more relevant as the threshold tends to assume more "extreme" values. Our results show that, as expected, γ decreases as n increases, with lower values for the SED with respect to those achieved under the PBD.

Table 2: Simulation results in scenario 1. Estimated threshold assuming linear model ($\hat{x}^*_\ell$), estimated proportion of patients in the target population ($\hat{\xi}_\ell$).

		n	$\hat{x}^*_\ell$	$\hat{\xi}_\ell$	var($\hat{x}^*_\ell$)	$E(\hat{x}^*_\ell - x^*)^2$	γ
Case 1	SED	30	0.035	0.487	0.270	0.356	0.053
	PBD	30	0.042	0.484	0.288	0.370	0.059
	SED	50	0.050	0.481	0.164	0.241	0.012
	PBD	50	0.057	0.478	0.177	0.250	0.015
	SED	70	0.042	0.484	0.106	0.187	0.003
	PBD	70	0.054	0.480	0.118	0.193	0.004
	SED	100	0.039	0.484	0.068	0.152	0.000
	PBD	100	0.044	0.482	0.073	0.153	0.001
Case 2	SED	30	0.405	0.361	0.310	0.458	0.106
	PBD	30	0.393	0.366	0.337	0.494	0.118
	SED	50	0.466	0.338	0.219	0.323	0.044
	PBD	50	0.470	0.337	0.234	0.336	0.052
	SED	70	0.482	0.328	0.162	0.256	0.019
	PBD	70	0.488	0.326	0.174	0.265	0.023
	SED	100	0.474	0.327	0.113	0.212	0.006
	PBD	100	0.482	0.324	0.120	0.214	0.007

4 Discussion

The relationship between treatment response and covariates/biomarkers is generally deduced from previous studies but, in practice, the true behavior is not generally known. In this paper, we assess the performance of the SED when some deviation from the linearity assumption could occur, also in comparison with the PBD. When the data are fitted by using the linear model, even a light model misspecification may largely affect the estimated threshold and its variance (scenario 1). Instead, finding the threshold by estimating the correct model (scenario 2) guarantees values of $\hat{x}^*_m$ closer to the true threshold x^*. In both cases, the SED shows better performance than the permuted block design. Note that we implemented the PBD 3:1 targeting the Neyman allocation instead of the classical permuted block design with ratio 2:2, according to the optimal design derived in [2]. Finally, it is worth recalling that the PBD randomization could be affected by high predictability, especially adopting an unbalanced allocation.

Acknowledgement

A. Baldi Antognini, R. Frieri and M. Zagoraiou 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

Table 3: Simulation results in scenario 2. Estimated threshold assuming the correct model ($\hat{x}^*_m$), estimated proportion of patients in the target population ($\hat{\xi}_m$).

		n	$\hat{x}^*_m$	$\hat{\xi}_m$	$var(\hat{x}^*_m)$	$E(\hat{x}^*_m - x^*)^2$	γ
Case 1	SED	30	0.272	0.404	0.282	0.286	0.070
	PBD	30	0.258	0.409	0.290	0.295	0.079
	SED	50	0.308	0.388	0.186	0.186	0.024
	PBD	50	0.303	0.389	0.190	0.191	0.028
	SED	70	0.317	0.383	0.130	0.130	0.009
	PBD	70	0.319	0.382	0.138	0.138	0.009
	SED	100	0.324	0.377	0.090	0.090	0.002
	PBD	100	0.321	0.379	0.094	0.094	0.003
Case 2	SED	30	0.509	0.339	0.543	0.621	0.065
	PBD	30	0.475	0.350	0.558	0.657	0.065
	SED	50	0.660	0.287	0.402	0.418	0.024
	PBD	50	0.647	0.291	0.422	0.442	0.026
	SED	70	0.729	0.258	0.286	0.289	0.009
	PBD	70	0.712	0.264	0.309	0.315	0.010
	SED	100	0.759	0.243	0.205	0.206	0.002
	PBD	100	0.750	0.246	0.216	0.217	0.003

code [2022TRB44L] - CUP [J53D23003270006]. A. Baldi Antognini and S. Cecconi were supported by EU funding within the NextGenerationEU-MUR PNRR Extended Partnership initiative on Emerging Infectious Diseases (Project no. PE00000007, INF-ACT).

References

1. FDA Enrichment strategies for clinical trials to support determination of effectiveness of human drugs and biological products guidance for industry *Available online* (2019).
2. Baldi Antognini A, Frieri R, Rosenberger W, et al. Optimal design for inference on the threshold of a biomarker. *Stat Methods Med Res* 2024; 33: 321 - 343.
3. Luty SE, Carter JD, McKenzie JM, et al. Randomised controlled trial of interpersonal psychotherapy and cognitive-behavioural therapy for depression. *Br J Psychiatry* 2007; 190: 496–502.
4. Steingrimsson JA, Betz J, Qian T, et al. Optimized adaptive enrichment designs for three-arm trials: learning which subpopulations benefit from different treatments. *Biostatistics* 2021; 22: 283–297.
5. Stallard N. Adaptive enrichment designs with a continuous biomarker. *Biometrics* 2023; 79: 9–19.
6. 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.
7. Follmann D. Adaptively changing subgroup proportions in clinical trials. *Stat Sin* 1997; 7: 1085–1102.
8. Baldi Antognini A, Frieri R and Zagoraiou M. New insights into adaptive enrichment designs. *Stat Papers* 2023; 64: 1305-1328.