Optimal and ethical designs for hypothesis testing in multi-arm exponential trials

(Article begins on next page)

09 July 2024

Optimal and ethical designs for hypothesis testing in multi-arm exponential trials

Rosamarie Frieri | Maroussa Zagoraiou

[1]Department of Statistical Sciences,
University of Bologna, Bologna, Italy

Correspondence
*Maroussa Zagoraiou,
Department of Statistical Sciences,
University of Bologna,
Via Belle Arti 41, 40126
Bologna, Italy
Email: maroussa.zagoraiou@unibo.it

Present Address
Department of Statistical Sciences
University of Bologna
Via Belle Arti 41, 40126
Bologna, Italy

Summary

Multi-arm clinical trials are complex experiments which involve several objectives. The demand for unequal allocations in a multi-treatment context is growing and adaptive designs are being increasingly used in several areas of medical research. For uncensored and censored exponential responses, we propose a constrained optimization approach in order to derive the design maximizing the power of the multivariate test of homogeneity, under a suitable ethical constraint. In the absence of censoring, we obtain a very simple closed-form solution that dominates the balanced design in terms of power and ethics. Our suggestion can also accommodate delayed responses and staggered entries, and can be implemented via response adaptive rules. While other targets proposed in the literature could present an unethical behaviour, the suggested optimal allocation is frequently unbalanced by assigning more patients to the best treatment, both in the absence and presence of censoring. We evaluate the operating characteristics of our proposal both theoretically and by simulations, also redesigning a real lung cancer trial, showing that the constrained optimal target guarantees very good performances in terms of ethical demands, power and estimation precision. Therefore, it is a valid and useful tool in designing clinical trials, especially oncological trials and clinical experiments for grave and novel infectious diseases, where the ethical concern is of primary importance.

KEYWORDS:
Multiple treatments, response adaptive randomization, survival trials, unequal allocations

1 | INTRODUCTION

In this paper we deal with the design of randomized multi-arm clinical trials for treatment comparisons to achieve a suitable trade-off between inferential and ethical demands. Most of the randomized clinical trials have been designed to achieve balanced allocation among the treatment groups. Equal allocation frequently maximizes the inferential precision in the estimation of the treatment effects and reflects the condition of *equipoise*, that has been widely recognized as an ethically necessary condition that should hold at the beginning of each trial.[1] However, the balanced allocation may not be efficient and could be strongly inappropriate for clinical trials, in which the ethical concern of individual care could be of crucial importance. Indeed, it is becoming increasingly common the use of unequal allocations not only for ethical reasons,[2,3] and the absolute need of true *equipoise* is object of debate.[4] Moreover, for heterogeneous treatment groups, unequal randomization often outperforms the

balanced design in terms of statistical efficiency. Advantages of unequal randomization can be ramped up in multi-arm trials with the promise of shortening drug development processes.[5]

Many clinical studies for severe/fatal diseases or oncological trials have time-to-event outcomes that can be modelled with the exponential distribution; this model can be used in trials for diseases with a very fast progression to death[6,7] or in combination with a censoring scheme.[8,9] In this context, the ethical demand of maximizing patient's care becomes prominent and the choice of the design should compromise between the conflicting goals of assigning more patients to the best performing treatment(s), while preserving power. These objectives can be formalized in a constrained/combined optimization problem,[10,11,12] whose solution is the so-called optimal *compromise* target. This framework has been adopted for example by Tymofyeyev et al[13] for the binary model, by Biswas et al[14] for both binary and continuous outcomes and by Baldi Antognini et al[15] for the linear homoscedastic model. In general, these optimal targets depend on the unknown model parameters and, under suitable conditions, response-adaptive randomization (RAR) procedures can be implemented to approach the desired target.[16,17,18] Principles and a variety of advantages of adaptive designs have been recently listed by FDA.[19] However, the application of RAR procedures in survival trials presents several complexities since: *(i)* the responses cannot be observed immediately but are naturally delayed, *(ii)* censored observations may be present and *(iii)* patients enrolment is often staggered in time.

Indeed, literature on RAR procedures for survival outcomes is quite scarce. To the best of our knowledge, Zhang and Rosenberger[20] were the first authors suggesting to design a survival trial on the basis of optimality criteria. They derived targets for two-arm trials with exponential and Weibull distribution, by minimizing an approximation of the total expected hazard, subject to power constraints. For several treatments and in absence of censoring, Zhu and Hu[21] derived analytically the optimal allocation that maximizes power for fixed weighted sample size. On the other hand, Sverdlov et al[22] introduced two optimal allocations for censored exponentially distributed outcomes, $NP1$ and $NP2$, based on non linear programming; analytical solution is available only for $NP1$, while $NP2$ can be addressed numerically. However, both these works[21,22] are based on the same constrained optimization framework proposed by Tymofyeyev et al,[13] which requires the choice of two user-selected thresholds: one of them related to a minimum percentage of allocations to each treatment (to avoid degenerate scenarios) and the other regarding the chosen efficacy measure (which is, however, a priori unknown since it depends on the model parameters). A further downside is related to the structure of the ensuing targets, since they do not always send more patients to the best treatment. To overcome these drawbacks, Baldi Antognini et al[15] proposed a new multi-purpose design strategy for the normal homoscedastic model.

This work deals with the problem of how to allocate subjects to $K \geq 2$ treatments for exponential trials. After introducing notation in Section 2, Section 3 discusses the simple set-up without censored observations. Firstly, we derive analytically the design maximizing the power of the multivariate Wald test. In absence of treatments with the same efficacy, the optimal target is a Neyman allocation involving just the clinically best and the worst treatments. Clearly, this allocation presents undesirable properties for both inference and ethics and, on this purpose, we discuss the complex issue of how to take into account patients' health in the design of a trial for more than two treatments. Therefore, we formalize a constrained optimization problem in which the power function is maximized subject to an ethical constraint on the allocation proportions, reflecting the effectiveness of the treatments. We compare the ensuing optimal target to several targets proposed in the literature - including a Bayesian procedure[23] - and we demonstrate that it is superior to the balanced design in terms of power and ethics. Then, we generalize the results by taking into account censored observations, also including delayed responses and staggered entries (Section 4). We implement our proposals with the Doubly adaptive Biased Coin Design[24] (DBCD) to discuss their operating characteristics in several experimental settings. We also perform robustness studies to model misspecifications and we redesign the three-arm KEYNOTE-010[8] clinical trial. We conclude the paper with a discussion and future developments (Section 5), while mathematical details are reported in Appendix A.

2 | FRAMEWORK AND NOTATION

Consider a clinical trial in which patients are allocated sequentially to $K \geq 2$ treatments and let Y_{ij} be the response of the j-th patient assigned to the i-th treatment where Y_{ij} follows an exponential distribution with mean $\theta_i \in \mathbb{R}^+$, for $i = 1, \ldots, K$. Let $\mathbf{N}_n = (N_{1n}, \ldots, N_{Kn})^\top$ be the random allocation vector, whose i-th component is the number of patients assigned to treatment i up to step n, where $n = \mathbf{N}_n^\top \mathbf{1}_K$ and $\mathbf{1}_K$ is the K-dimensional vector of ones. Let $\pi_{in} = N_{in}/n$, then $\boldsymbol{\pi}_n = (\pi_{1n}, \ldots, \pi_{Kn})^\top$ is the vector of allocation proportions such that $\boldsymbol{\pi}_n^\top \mathbf{1}_K = 1$.

After n steps, letting $\hat{\boldsymbol{\theta}}_n = (\hat{\theta}_{1n}, \ldots, \hat{\theta}_{Kn})^\top$ be the vector of the MLEs of the treatment effects, by a well-known result $\hat{\boldsymbol{\theta}}_n \xrightarrow{a.s.} \boldsymbol{\theta}$ and $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} \mathrm{N}(\mathbf{0}_K, \mathbf{M}^{-1})$, where $\mathbf{M} = \mathrm{diag}\left(\pi_{in}\theta_i^{-2}\right)_{i=1,\ldots,K}$ is the Fisher information matrix. In multi-arm trials, the

inferential interest is usually focused on the contrasts, so let us define $\gamma = \mathbf{A}\theta$ where $\mathbf{A} = [\mathbf{1}_{K-1}| -\mathbf{I}_{K-1}]$ and $\mathbf{I}_{K-1}$ is the $(K-1)$-dim identity matrix so that $\gamma = (\theta_1 - \theta_2, \dots, \theta_1 - \theta_K)^\top$. By denoting with $\hat{\gamma}_n = \mathbf{A}\hat{\theta}_n$ the corresponding MLE, it is known that $\hat{\gamma}_n$ is strongly consistent and asymptotically normal with $\sqrt{n}(\hat{\gamma}_n - \gamma) \xrightarrow{d} N(\mathbf{0}_{K-1}, \Sigma)$, where $\Sigma = \mathbf{A}\mathbf{M}^{-1}\mathbf{A}^\top$.

We define the target as the desirable treatment allocation proportion $\rho = (\rho_1, \dots, \rho_K)^\top$, where $\rho^\top \mathbf{1}_K = 1$ and $\rho_i \geq 0$, for $i = 1, \dots, K$. If the latter inequality is strict, namely if $\rho_i > 0$ for all $i = 1, \dots, K$, ρ is called non degenerate. The target can be seen, in a finite set-up, as the actual desired proportion of treatment assignments. Otherwise, it can be found as a limit, for increasing n, to which the allocation proportion should ideally converge. In general, the targets could depend on the unknown parameters and, under suitable conditions, RAR procedures can be carried out to sequentially estimate the model parameters, and then force the assignments to asymptotically approach the chosen target. [17,18]

Without loss of generality, in this work we assume that higher values for the response are more desirable and we let $\theta_1 \geq \theta_2 \geq \dots \geq \theta_K$, i.e. the best treatment is indicated with label 1 and the worst one with label K, admitting also groups of treatments with the same efficacy. This assumption is not restrictive since it is simply a label-coding choice. Indeed, the ordering of treatments is a priori unknown, but for RAR rules the treatment effects are estimated step by step and then their ranking is sequentially updated. Clearly, the contrasts can be defined with respect to any treatment (not necessary the best one) by re-defining the matrix $\mathbf{A}$. For instance, if we set $\mathbf{A} = [\mathbf{I}_{K-1}| -\mathbf{1}_{K-1}]$, then $\gamma = (\theta_1 - \theta_K, \dots, \theta_{K-1} - \theta_K)^\top$ so that the contrasts are defined with respect to the worst performing treatment.

3 | OPTIMAL ALLOCATIONS FOR HYPOTHESIS TESTING FOR THE EXPONENTIAL MODEL

In this section we derive optimal targets for testing the hypothesis of homogeneity among treatment effects. Such overall null-hypothesis is a milestone in the statistical literature and it is the first stage of multiple comparison techniques for many stepwise procedures [25,26] (see also Section 5). We then compare the ensuing target with the optimal allocations for the exponential model proposed in the literature both theoretically and numerically.

3.1 | Single-objective optimal allocation for hypothesis testing

Let us consider the problem of testing the null-hypothesis of equality among the treatment effects,

$$\begin{cases} H_0 : \gamma = \mathbf{0}_{K-1} \\ H_1 : \gamma \neq \mathbf{0}_{K-1}, \end{cases}$$

where $\mathbf{0}_{K-1}$ is the $K-1$ dimensional vector of zeros. After n steps, let $\hat{\mathbf{M}}_n$ and $\hat{\Sigma}_n = \mathbf{A}\hat{\mathbf{M}}_n^{-1}\mathbf{A}^\top$ be consistent estimators of $\mathbf{M}$ and Σ, respectively. Under H_0, the Wald's statistic $W_n = n \cdot \hat{\gamma}_n^\top \hat{\Sigma}_n^{-1} \hat{\gamma}_n \xrightarrow{d} \chi_{K-1}^2(0)$, while under H_1, $W_n \xrightarrow{d} \chi_{K-1}^2(n\phi)$, where $\chi_{K-1}^2(n\phi)$ is a chi-squared r.v. with $K - 1$ degrees of freedom and non centrality parameter (NCP) $n\phi$, where $\phi = \phi(\pi) = \gamma^\top \Sigma^{-1} \gamma$ is given by (see for example Zhu and Hu [21])

$$\phi(\pi) = \sum_{i=1}^{K} \left(\frac{\theta_1 - \theta_i}{\theta_i} \right)^2 \pi_i - \frac{1}{\sum_{i=1}^{K} \frac{\pi_i}{\theta_i^2}} \left(\sum_{i=1}^{K} \frac{\theta_1 - \theta_i}{\theta_i^2} \pi_i \right)^2. \tag{1}$$

For every sample size, the power of the Wald homogeneity test monotonically increases as ϕ grows; as a shorthand we shall often refer to ϕ as the NCP. In the next Theorem, following a similar set-up to Tymofyeyev et al, [13] we derive the allocation proportion maximizing the NPC.

Theorem 1. The target allocation $\tilde{\rho} = (\tilde{\rho}_1, \dots, \tilde{\rho}_K)^\top$ maximizing the power of Wald's test is such that $\phi(\tilde{\rho}) = \left(\frac{\theta_1 - \theta_K}{\theta_1 + \theta_K} \right)^2$. Given $\theta_1 = \dots = \theta_j > \theta_{j+1} \geq \dots \geq \theta_h = \dots = \theta_K$, with $1 \leq j < h \leq K$,

  (i) if $\theta_{j+1} > \theta_h$, then every allocation $\tilde{\rho}$ such that $\sum_{i=1}^{j} \tilde{\rho}_i = \frac{\theta_1}{\theta_1 + \theta_K}$, $\tilde{\rho}_{j+1}, \dots, \tilde{\rho}_{h-1} = 0$ and $\sum_{i=h}^{K} \tilde{\rho}_i = \frac{\theta_K}{\theta_1 + \theta_K}$ is optimal;

  (ii) if $\theta_{j+1} = \theta_h$ i.e. in presence of only two clusters of treatments, every allocation $\tilde{\rho}$ such that $\sum_{i=1}^{j} \tilde{\rho}_i = \frac{\theta_1}{\theta_1 + \theta_K} = 1 - \sum_{i=j+1}^{K} \tilde{\rho}_i$ is optimal.

Proof. See Appendix A.1. □

The targets in *(i)* are degenerate, since every $\tilde{\rho}$ provides at least an empty treatment arm. In particular, in presence of a single superior and inferior treatments ($j = 1$ and $h = K$),

$$\tilde{\rho} = \left(\frac{\theta_1}{\theta_1 + \theta_K}, 0, \ldots, 0, \frac{\theta_K}{\theta_1 + \theta_K} \right)^{\top} \tag{2}$$

i.e. it is a generalization of Neyman allocation involving just the best and the worst treatments, not collecting informations on the intermediates. Target in (2) is always a possible solution for both *(i)* and *(ii)*, and for $K = 2$ we retrieve the usual Neyman allocation. Notice that the only non degenerate optimal targets $\tilde{\rho}$ are those obtained under scenario *(ii)* of Theorem 1.

Example 1. If $\theta = (4, 4, 4, 1)^{\top}$, then $\tilde{\rho} = \left(\frac{4}{5}, 0, 0, \frac{1}{5} \right)^{\top}$ is optimal with $\phi(\tilde{\rho}) = \frac{9}{25}$. Moreover, every combination of $\tilde{\rho}_1, \tilde{\rho}_2$ and $\tilde{\rho}_3$ such that $\sum_{i=1}^{3} \tilde{\rho}_i = \frac{4}{5}$ is optimal, like e.g., $\left(\frac{2}{5}, \frac{2}{5}, 0, \frac{1}{5} \right)^{\top}$ or $\left(\frac{2}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5} \right)^{\top}$.

Remark 1. For trials comparing $K > 2$ treatments, the definition of the ethical issue is highly debated and controversial. For example, the trial may be designed with the requirement of maximizing patients' benefit by a) maximizing the number of subjects receiving the superior treatment(s) or b) minimizing the number of patients treated with the inferior arm(s). While for $K = 2$ these two ethical paradigms are equivalent, in the case of multi-arm trials the implementation of a) does not necessarily satisfy b) and vice versa; moreover, even if a) and b) hold simultaneously, the conclusions may be questionable. For instance, under the-larger-the-better scenario and assuming $\theta_1 > \theta_2 > \theta_3 > \theta_4$, the target $\left(\frac{7}{16}, \frac{2}{16}, \frac{6}{16}, \frac{1}{16} \right)^{\top}$ complies with a) and b), but how it can be considered as *ethical*? Wouldn't it be the target $\left(\frac{7}{16}, \frac{6}{16}, \frac{2}{16}, \frac{1}{16} \right)^{\top}$ more desirable for patients' health? In this paper, a target will be considered as ethical if its components are ordered according to the magnitude of the treatment effects (this definition was mentioned in Theorem 1 of Sverdlov et al[22]). In this way not only a) and b) hold, but the ranking of the ρ_i's reflects the efficacy of θ, i.e. $\rho_i \geq \rho_{i+1} \iff \theta_i \geq \theta_{i+1}$ for all $i = 1, \ldots, K - 1$. Clearly, for $K = 2$ the Neyman allocation for the exponential model is ethical since $\rho_1 \geq \rho_2 \iff \theta_1 \geq \theta_2$.

Despite the optimal unconstrained design $\tilde{\rho}$ maximizes power, it presents undesirable characteristics from both ethical and inferential perspective. In general, this target does not assign patients to the intermediate treatments, giving unreliable variance of the estimate of model parameters. Moreover, $\tilde{\rho}$ is not attractive from an ethical point of view, since it always allocates a fraction of patients to the worst treatment arm.

3.2 | Multi-objective optimal allocation for hypothesis testing

In this section we introduce the multi-purpose optimal target ρ^C maximizing the NCP subject to an ethical constraint reflecting the order among treatments. Specifically, in the following Theorem we derive the closed form solution of the constrained optimization problem

$$\begin{cases} \max \phi(\rho) \\ \text{s.t.} \quad \rho_i \geq \rho_{i+1} \text{ for } i = 1, \ldots, K - 1 \quad \text{and} \quad \sum_{i=1}^{K} \rho_i = 1. \end{cases} \tag{3}$$

Theorem 2. Given $\theta_1 = \cdots = \theta_j > \theta_{j+1} \geq \cdots \geq \theta_h = \cdots = \theta_K$ ($1 \leq j < h \leq K$), let us define

$$x = \frac{\frac{1}{\theta_1} \sum_{i=1}^{K} \left(\frac{1}{\theta_i} - \frac{1}{\theta_1} \right)^2}{\left[\sum_{i=1}^{K} \left(\frac{1}{\theta_i} - \frac{1}{\theta_1} \right) \right] \left[\sum_{i=1}^{K} \left(\frac{1}{\theta_i^2} - \frac{1}{\theta_1^2} \right) \right]}. \tag{4}$$

If $x > K^{-1}$, then the balanced design $\rho^B = \left(\frac{1}{K}, \ldots, \frac{1}{K} \right)^{\top}$ is optimal. Otherwise, when $x \leq K^{-1}$, the solution of (3) is $\rho^C = (\rho_1^C, \ldots, \rho_K^C)^{\top}$ (with $\rho_i^C \geq \rho_{i+1}^C$ for $i = 1, \ldots, K - 1$ and $\mathbf{1}_K^{\top} \rho^C = 1$), where

(i) if $\theta_{j+1} > \theta_h$, then $\rho_{j+1}^C = \ldots = \rho_K^C = x$ (and, clearly, $\sum_{i=1}^{j} \rho_i^C = 1 - (K - j)x$);

(ii) in the case of just two clusters of treatments, namely when $\theta_{j+1} = \theta_h$, then $\sum_{i=1}^{j} \rho_i^C = 1 - (K - j)x = \theta_1/(\theta_1 + \theta_K) = 1 - \sum_{i=j+1}^{K} \rho_i^C$.

Proof. See Appendix A.2. □

The constrained optimal target presents very appealing properties: ρ^C has a very simple form and it is non degenerate, so that there is no need to fix a lower bound for the treatment allocation proportion to the worst treatment(s). Furthermore, it assigns the same proportion of patients to all the treatment arms or it skews the allocations in favour of the best performing treatment(s).

The behaviour of ρ^C for $K = 3$ and several experimental settings is displayed in Table 1. The optimal constrained target allocates a higher proportion of subjects to the superior treatment and this proportion increases as the difference between θ_1 and θ_2 grows (Table 1a). Moreover, ρ_1^C increases as the magnitude of the superior treatment increases (Table 1b). Note that, in the last scenario of Table 1a $x = 1/6$, so that all the targets with $\rho_2^C + \rho_3^C = 1/3$ are optimal.

TABLE 1 Behaviour of the optimal constrained target.

(a) Fixed θ_1 and θ_3 and decreasing θ_2.				(b) Fixed θ_2 and θ_3 and increasing θ_1.			
θ	$\rho_1{}^C$	$\rho_2{}^C$	$\rho_3{}^C$	θ	$\rho_1{}^C$	$\rho_2{}^C$	$\rho_3{}^C$
$(10, 9, 5)^\top$	0.436	0.282	0.282	$(10, 8, 4)^\top$	0.546	0.227	0.227
$(10, 7, 5)^\top$	0.590	0.205	0.205	$(15, 8, 4)^\top$	0.706	0.147	0.147
$(10, 5, 5)^\top$	0.667	0.1665	0.1665	$(20, 8, 4)^\top$	0.774	0.113	0.113

3.3 | Comparison of optimal targets for the exponential model

The aim of this section is to compare, both theoretically and with numerical examples, the constrained optimal target with those proposed in the literature for exponential outcomes. By taking into account the well known A-optimal criterion, Sverdlov and Rosenberg[27] derived the optimal allocation ρ^A minimizing the trace of Σ,

$$\rho_1^A = \frac{\theta_1\sqrt{K-1}}{\theta_1\sqrt{K-1} + \sum_{k=2}^K \theta_k} \quad \text{and} \quad \rho_i^A = \frac{\theta_i}{\theta_1\sqrt{K-1} + \sum_{k=2}^K \theta_k} \quad \text{for } i = 2, \ldots, K,$$

while Wong and Zhu[28] found the D-optimal allocation ρ^D minimizing the determinant of Σ (this target is not available in closed form but is derived as the unique root of a non linear system of equations). Although for linear contrasts these criteria are usually called A_A and D_A, for the sake of notation, we simply refer to them as ρ^A and ρ^D.

For exponential outcomes without censoring, Zhu and Hu[21] derived the optimal allocation ρ^Z maximizing the NCP for fixed n, subject to the constraint that $\rho_i^Z \geq T$ for all $i = 1, \ldots, K$ where $T \in [0, K^{-1}]$ is a user-selected threshold. Nevertheless, no guidelines are provided by the authors to help the choice of T; moreover, Theorem 1 of Zhu and Hu[21] does not include trials in which the treatments are grouped into two clusters (like, e.g., $K = 3$ with $\theta_1 = \theta_2 > \theta_3$).

To measure the effectiveness of such targets in terms of both statistical and ethical performances, we evaluate:

1. the efficiency in terms of power of a given target ρ as $E_\phi(\rho) = \frac{\phi(\rho)}{\phi(\tilde\rho)}$ where $\tilde\rho$ is the unconstrained optimal target defined in Theorem 1;

2. D_A and A_A efficiency, defined by $E_{D_A}(\rho) = \left[\frac{|\mathbf{A}\mathbf{M}^{-1}(\rho^D)\mathbf{A}^\top|}{|\mathbf{A}\mathbf{M}^{-1}(\rho)\mathbf{A}^\top|}\right]^{\frac{1}{K-1}}$ and $E_{A_A}(\rho) = \frac{tr(\mathbf{A}\mathbf{M}^{-1}(\rho^A)\mathbf{A}^\top)}{tr(\mathbf{A}\mathbf{M}^{-1}(\rho)\mathbf{A}^\top)}$, respectively. Accordingly, values of $E_{D_A}(\rho)$ or $E_{A_A}(\rho)$ close to 1 point out that the design ρ has similar performance in terms of estimation precision of the optimal design ρ^D or ρ^A;

3. besides the ethical measures provided by ρ_1 and ρ_K representing the assignments to the best and worst treatments, to assess a global ethical performance we compute $E_e(\rho) = \frac{\sum_{i=1}^K \theta_i\rho_i}{\theta_1}$, that is the ratio between the total expected responses under a given target ρ and the total expected outcomes obtained by assigning all the subjects to the best treatment.

In the comparisons we also take into account the balanced design as a benchmark. However, equal allocation may be suboptimal in trials with heterogeneous variance.[3] In the following Theorem we demonstrate that the constrained optimal target proposed in this paper has always better performance in terms of power and ethical efficiency wrt the balanced design.

Theorem 3. The optimal constrained target ρ^C dominates the balanced allocation ρ^B, namely $E_\phi(\rho^C) \geq E_\phi(\rho^B)$ and $E_e(\rho^C) \geq E_e(\rho^B)$, simultaneously.

Proof. See Appendix A.3. ☐

Remark 2. While ρ^D satisfies the property of Remark 1 (shown by Sverdlov et al[22]), i.e. $\theta_1 \geq ... \geq \theta_K \Leftrightarrow \rho_1^D \geq ... \geq \rho_K^D$, targets ρ^Z and ρ^A could present a controversial behaviour in terms of ethics. In general, ρ^A assigns more subjects to the reference treatment (that does not always coincide with the best one, as in our framework). For instance, if $\theta = (25, 29, 30)^\top$ then $\rho^A = (0.375, 0.307, 0.318)^\top$, so that more subjects receive the worst treatment. Moreover, also ρ^Z is unethical since it does not always satisfy a) and/or b) of Remark 1. Indeed, in the same scenario, by setting $T = 0.2$, we obtain $\rho^Z = (0.425, 0.2, 0.375)^\top$, so it assigns the higher proportion of patients to the less effective treatment.

Further numerical comparisons are reported in Table 2, in which we consider $K = 3$ and 5 treatments and several values of θ. We have included $\tilde{\rho}$ for completeness, even if it is strongly inadequate, as widely discussed in Section 3.1. Let us first consider the case of $K = 3$. As far as power is concerned, the proposed constrained target ρ^C always presents higher power than the competitors except in the scenario $\theta = (8, 5, 4)^\top$, with a loss of 0.7% wrt ρ^Z. In the remaining cases, the gain of ρ^C in terms of power efficiency on the second best is up to 3.4%. With regard to the ethical concern, ρ^C presents values closest to one, except in the first scenario, in which ρ^A has essentially equivalent ethical efficiency. This latter target represents the second best choice in terms of ethical demand. In general, the D_A efficiency of ρ^C is higher for narrower values of the relative differences among θ_1 and θ_2 and it is higher than 77% in all these settings. Very good performances in terms of A_A efficiency are achieved by the constrained target, with values always greater than 90.6%. Note that ρ^Z is nearly constant as the vector of treatment effects changes. Indeed, the allocation proportion to the worst treatments is equal, or very close to T. The results for $K = 5$ treatments enhance the value of our proposal and similar considerations to the case of $K = 3$ treatments still hold. Note that the gain in terms of power efficiency reaches the 19% wrt other designs (see the last scenario). The undesirable behaviour in terms of patients' benefit of ρ^Z, discussed in Remark 2, holds for both $K = 3$ and $K = 5$ treatments.

In Figure 1 we show the behaviour of the above-mentioned efficiency measures for $K = 3$ treatments where θ_1 varies between 10 and 30, $\theta_2 = 9$ and $\theta_3 = 8$. The constrained target ρ^C dominates the other allocations in terms of power and ethics for values of $\theta_1 \geq 15$, whereas for smaller values of θ_1 the maximum loss wrt ρ^Z is 8.8% for $\theta_1 = 10$. The ethical efficiency of ρ^C is increasing in θ_1 for $\theta_1 > 20$ (the same behaviour can be observed only for ρ^A when $\theta_1 > 22$, whereas ρ^D, ρ^B and ρ^Z are always decreasing in θ_1), and the average gain of ρ^C on the second best (ρ^A) is 3.5%. The best performance in terms of D_A efficiency is given by ρ^B, while ρ^Z is the second best for high values of θ_1. On the other hand, $E_{D_A}(\rho^C)$ and $E_{D_A}(\rho^A)$ decrease as θ_1 increases, having lower values with respect to the other targets. Both ρ^Z and ρ^C present A_A efficiency close to 1 (in particular highest values are achieved by ρ^Z for $\theta_1 \in [13, 26]$ while in the remaining configurations of parameters the highest values are reached by ρ^C). It is worth notice that $E_{A_A}(\rho^C)$ is nearly constant w.r.t. θ_1 while A_A efficiency tends to decrease for the remaining targets.

The theoretical properties as well as the comparisons in Table 2 and Figure 1 show that the optimal constrained target ρ^C guarantees very good performance in terms of power, estimation precision and ethical concerns.

3.4 | Implementing the optimal target with RAR procedures

In this section we apply the DBCD[24] to implement the constrained target, under the assumption that responses are available immediately (see e.g. Shulz et al[7]). This procedure starts with a sample of n_0 patients allocated to the K treatments (usually by restricted randomization) to obtain initial estimates of model parameters. After $j > n_0$ subjects are assigned and the responses are observed, the unknown parameters are estimated by $\hat{\theta}_j = (\hat{\theta}_{1j}, \ldots, \hat{\theta}_{Kj})^\top$ and they are used to compute the estimated target $\hat{\rho}_j = (\hat{\rho}_{1j}, \ldots, \hat{\rho}_{Kj})^\top$. Then, the $(j+1)$th patient is randomized to treatment i with probability $\Psi_{j+1,i} = \hat{\rho}_{ij} \left(\frac{\hat{\rho}_{ij}}{\pi_{ij}}\right)^\kappa \left[\sum_{i=1}^{K} \hat{\rho}_{ij} \left(\frac{\hat{\rho}_{ij}}{\pi_{ij}}\right)^\kappa\right]^{-1}$, where $\kappa \in [0, +\infty)$. We apply the above-mentioned procedure to target the constrained optimal design by setting $\kappa = 2$, that represents a good trade-off between randomness and optimality.[29] We take into account the scenarios of Tables 1a and 1b for different sample sizes, where each trial has been replicated 10000 times. The first $\frac{1}{10}$ of the total sample size is assigned to the treatments with restricted randomization, then allocations become response adaptive. The results are provided in Tables 3 and 4, in which we report for every scenario the estimates $\hat{\theta}_n$, the theoretical target, the simulated allocation proportions π_n (with their standard deviations in square brackets) and the simulated average power (W) of Wald's test.

The sequential procedure implementing ρ^C assigns a higher proportion of subjects to the more effective treatment. Clearly, the convergence to the optimal target improves for increasing n, even though, in many experimental scenarios, good convergence is also achieved for $n = 100$ and the simulated allocation proportions are equal or really close to their target values for $n = 250$. Moreover, the convergence improves as the efficacy of the best treatment grows (see Table 4) and the average power exhibits values over 72% in all the considered settings, while for $n \geq 150$ it is above 88%.

TABLE 2 Comparison of optimal targets for $K = 3$ and 5 treatments in different experimental scenarios wrt power efficiency $E_\phi(\rho)$, ethical efficiency $E_e(\rho)$, D_A efficiency $E_{D_A}(\rho)$, and A_A efficiency $E_{A_A}(\rho)$.

$\theta^\top$	ρ	$E_\phi(\rho)$	$E_e(\rho)$	$E_{D_A}(\rho)$	$E_{A_A}(\rho)$
(30, 20, 8)	$\rho^A = (0.602, 0.284, 0.114)^\top$	0.761	0.822	0.933	1
	$\rho^D = (0.441, 0.385, 0.174)^\top$	0.765	0.744	1	0.905
	$\tilde{\rho} = (0.789, 0, 0.211)^\top$	1	0.845	$\to 0$	$\to 0$
	$\rho^C = (0.664, 0.168, 0.168)^\top$	0.889	0.821	0.836	0.906
	ρ^B	0.740	0.644	0.903	0.730
	$\rho^Z = (0.591, 0.200, 0.209)^\top$	0.881	0.780	0.888	0.927
(30, 10, 8)	$\rho^A = (0.702, 0.165, 0.133)^\top$	0.868	0.792	0.864	1
	$\rho^D = (0.464, 0.295, 0.241)^\top$	0.657	0.627	1	0.815
	$\tilde{\rho} = (0.789, 0, 0.211)^\top$	1	0.845	$\to 0$	$\to 0$
	$\rho^C = (0.768, 0.116, 0.116)^\top$	0.900	0.839	0.770	0.973
	ρ^B	0.501	0.533	0.954	0.620
	$\rho^Z = (0.600, 0.200, 0.200)^\top$	0.807	0.720	0.953	0.956
(12, 5, 4)	$\rho^A = (0.653, 0.193, 0.154)^\top$	0.838	0.785	0.899	1
	$\rho^D = (0.449, 0.303, 0.248)^\top$	0.668	0.658	1	0.856
	$\tilde{\rho} = (0.750, 0, 0.250)^\top$	1	0.833	$\to 0$	$\to 0$
	$\rho^C = (0.726, 0.137, 0.137)^\top$	0.872	0.828	0.805	0.968
	ρ^B	0.535	0.583	0.962	0.683
	$\rho^Z = (0.600, 0.200, 0.200)^\top$	0.822	0.750	0.942	0.985
(8, 5, 4)	$\rho^A = (0.557, 0.246, 0.197)^\top$	0.760	0.809	0.944	1
	$\rho^D = (0.411, 0.321, 0.268)^\top$	0.669	0.746	1	0.919
	$\tilde{\rho} = (0.667, 0, 0.333)^\top$	1	0.834	$\to 0$	$\to 0$
	$\rho^C = (0.628, 0.186, 0.186)^\top$	0.801	0.837	0.876	0.973
	ρ^B	0.603	0.708	0.979	0.814
	$\rho^Z = (0.555, 0.200, 0.245)^\top$	0.808	0.802	0.931	0.980
(12, 11, 10, 5, 3)	$\rho^A = (0.453, 0.208, 0.189, 0.094, 0.057)^\top$	0.660	0.854	0.858	1
	$\rho^D = (0.235, 0.232, 0.229, 0.182, 0.123)^\top$	0.719	0.745	1	0.775
	$\tilde{\rho} = (0.800, 0, 0, 0, 0.200)^\top$	1	0.850	$\to 0$	$\to 0$
	$\rho^C = (0.540, 0.115, 0.115, 0.115, 0.115)^\top$	0.810	0.818	0.791	0.856
	ρ^B	0.723	0.683	0.966	0.676
	$\rho^Z = (0.373, 0.150, 0.150, 0.150, 0.177)^\top$	0.796	0.742	0.912	0.869
(12, 10, 8, 6, 4)	$\rho^A = (0.462, 0.192, 0.154, 0.115, 0.077)^\top$	0.574	0.808	0.865	1
	$\rho^D = (0.231, 0.224, 0.211, 0.189, 0.144)^\top$	0.562	0.701	1	0.763
	$\tilde{\rho} = (0.750, 0, 0, 0, 0.250)^\top$	1	0.833	$\to 0$	$\to 0$
	$\rho^C = (0.548, 0.113, 0.113, 0.113, 0.113)^\top$	0.695	0.812	0.783	0.912
	ρ^B	0.577	0.667	0.983	0.683
	$\rho^Z = (0.348, 0.150, 0.150, 0.150, 0.202)^\top$	0.683	0.716	0.927	0.882
(12, 8, 7, 6, 3)	$\rho^A = (0.500, 0.167, 0.146, 0.125, 0.062)^\top$	0.548	0.774	0.840	1
	$\rho^D = (0.236, 0.221, 0.213, 0.202, 0.128)^\top$	0.526	0.640	1	0.718
	$\tilde{\rho} = (0.800, 0, 0, 0, 0.200)^\top$	1	0.850	$\to 0$	$\to 0$
	$\rho^C = (0.612, 0.097, 0.097, 0.097, 0.097)^\top$	0.716	0.805	0.716	0.898
	ρ^B	0.565	0.600	0.973	0.628
	$\rho^Z = (0.363, 0.150, 0.150, 0.150, 0.187)^\top$	0.674	0.672	0.918	0.877

ρ^A, A_A optimal design; ρ^D, D_A optimal design; $\tilde{\rho}$, optimal unconstrained target; ρ^C, optimal constrained target; ρ^B, balanced design; ρ^Z, Zhu and Hu[21] optimal target (with minimum allocation proportion for each arm T equal to 0.2 for $K = 3$ and 0.15 for $K = 5$).
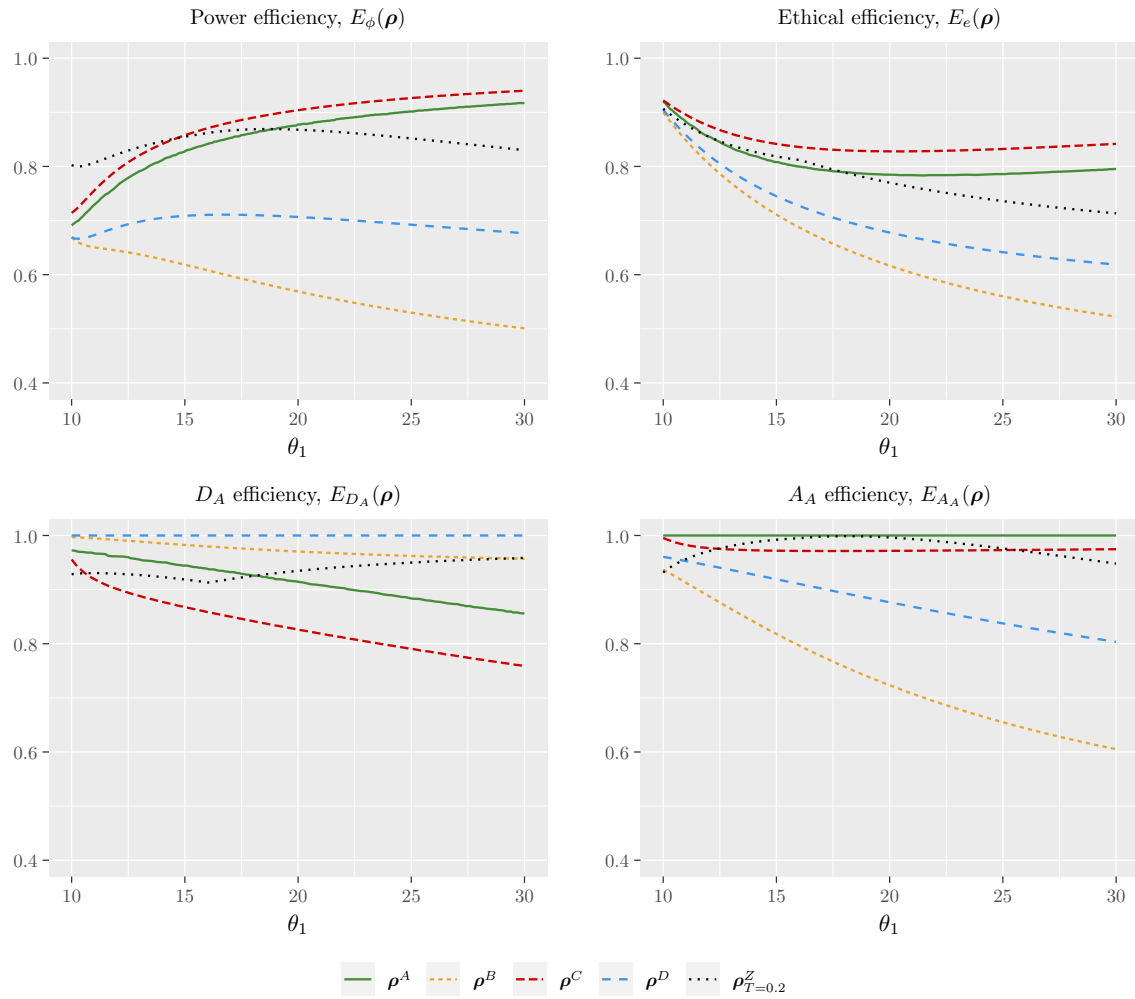
FIGURE 1 Comparisons of optimal allocations wrt several measures of efficiency for $\theta = (\theta_1, 9, 8)^\top$ as θ_1 varies from 10 to 30. ρ^A, A_A optimal design; ρ^B, balanced design; ρ^C, optimal constrained target; ρ^D, D_A optimal design; ρ^Z, Zhu and Hu[21] optimal target (with minimum allocation proportion for each arm T equal to 0.2).

TABLE 3 Simulation results (Scenarios Table 1a). Estimates of model parameters $\hat{\theta}_n$, simulated allocation proportions π_n (their standard deviation in square brackets) and simulated average power W of Wald's test; 10000 iterations.

n	$\theta^\top$	$\hat{\theta}_n^\top$	ρ^C	$\pi_n^\top$		W
100		$(9.6, 8.6, 4.9)$		$(0.44, 0.32, 0.24)$	$[.164, .140, .072]$	0.721
150	$(10, 9, 5)$	$(9.7, 8.7, 5.0)$	$(0.44, 0.28, 0.28)^\top$	$(0.44, 0.32, 0.24)$	$[.143, .122, .050]$	0.885
200		$(9.8, 8.8, 5.0)$		$(0.44, 0.31, 0.25)$	$[.129, .108, .044]$	0.965
250		$(9.8, 8.8, 5.0)$		$(0.44, 0.30, 0.26)$	$[.119, .098, .041]$	0.990
100		$(9.8, 6.8, 4.9)$		$(0.55, 0.24, 0.21)$	$[.135, .093, .062]$	0.731
150	$(10, 7, 5)$	$(9.9, 6.8, 5.0)$	$(0.60, 0.20, 0.20)^\top$	$(0.56, 0.23, 0.21)$	$[.110, .072, .049]$	0.894
200		$(9.9, 6.9, 5.0)$		$(0.57, 0.22, 0.21)$	$[.093, .058, .041]$	0.960
250		$(9.9, 6.9, 5.0)$		$(0.57, 0.22, 0.21)$	$[.083, .049, .038]$	0.987
100		$(9.9, 4.9, 4.9)$		$(0.64, 0.18, 0.18)$	$[.096, .056, .049]$	0.880
150	$(10, 5, 5)$	$(10.0, 4.9, 4.9)$	$(0.66, 0.17, 0.17)^\top$	$(0.65, 0.18, 0.17)$	$[.067, .038, .034]$	0.973
200		$(10.0, 4.9, 4.9)$		$(0.66, 0.17, 0.17)$	$[.053, .030, .027]$	0.996
250		$(10.0, 5.0, 5.0)$		$(0.66, 0.17, 0.17)$	$[.044, .024, .023]$	0.999

TABLE 4 Simulation results (Scenarios Table 1b). Estimates of model parameters $\hat{\theta}_n$, simulated allocation proportions π_n (their standard deviation in square brackets) and simulated average power W of Wald's test; 10000 iterations.

n	$\theta^\top$	$\hat{\theta}_n^\top$	ρ^C	$\pi_n^\top$		W
100		(9.7, 7.6, 3.9)		(0.51, 0.28, 0.21)	[.166, .136, .062]	0.918
150	(10, 8, 4)	(9.8, 7.8, 3.9)	$(0.54, 0.23, 0.23)^\top$	(0.52, 0.27, 0.21)	[.142, .113, .048]	0.988
200		(9.8, 7.8, 4.0)		(0.53, 0.25, 0.22)	[.123, .095, .042]	0.998
250		(9.9, 7.9, 4.0)		(0.53, 0.25, 0.22)	[.110, .081, .039]	1.000
100		(14.9, 7.6, 3.9)		(0.70, 0.16, 0.14)	[.106, .075, .044]	0.998
150	(15, 8, 4)	(14.9, 7.8, 3.9)	$(0.70, 0.15, 0.15)^\top$	(0.70, 0.15, 0.15)	[.076, .049, .034]	1.000
200		(14.9, 7.8, 3.9)		(0.70, 0.15, 0.15)	[.058, .035, .028]	1.000
250		(15.0, 7.9, 4.0)		(0.70, 0.15, 0.15)	[.049, .028, .025]	1.000
100		(19.9, 7.6, 3.8)		(0.77, 0.12, 0.11)	[.073, .047, .034]	1.000
150	(20, 8, 4)	(20.0, 7.7, 3.9)	$(0.78, 0.11, 0.11)^\top$	(0.78, 0.11, 0.11)	[.053, .031, .027]	1.000
200		(20.0, 7.9, 3.9)		(0.78, 0.11, 0.11)	[.043, .024, .022]	1.000
250		(20.0, 7.9, 3.9)		(0.78, 0.11, 0.11)	[.037, .020, .019]	1.000

In Table 5 we summarize the results of the simulated type I error (W_α) of Wald's test, where the nominal value was set to 0.05. Clearly, under this setting $\rho^C = \rho^B$ and DBCD gives simulated allocation proportions that converge to the balanced design: even for $n = 100$, the type I error rate is really close to the nominal value.

TABLE 5 Simulation results of type I error W_α, estimates of model parameters $\hat{\theta}_n$ and simulated allocation proportions π_n (their standard deviation in square brackets); 10000 iterations.

n	$\theta^\top$	$\hat{\theta}_n^\top$	$\pi_n^\top$		W_α
100		(11.6, 11.9, 11.8)	(0.32, 0.33, 0.35)	[.114, .093, .108]	0.052
150	(12, 12, 12)	(11.7, 11.9, 11.9)	(0.33, 0.33, 0.34)	[.104, .083, .093]	0.048
200		(11.8, 11.9, 11.9)	(0.33, 0.33, 0.34)	[.099, .078, .087]	0.048
250		(11.8, 12.0, 12.0)	(0.33, 0.33, 0.34)	[.094, .075, .083]	0.046
100		(3.9, 4.0, 3.9)	(0.32, 0.33, 0.35)	[.114, .093, .108]	0.052
150	(4, 4, 4)	(3.9, 4.0, 4.0)	(0.33, 0.33, 0.34)	[.104, .083, .093]	0.048
200		(3.9, 4.0, 4.0)	(0.33, 0.33, 0.34)	[.099, .078, .087]	0.048
250		(3.9, 4.0, 4.0)	(0.33, 0.33, 0.34)	[.094, .075, .083]	0.046

3.5 | Comparisons of our proposal with the Bayesian adaptive randomization strategy proposed by Trippa et al[23]

Although our approach is frequentist, Bayesian adaptive designs are quite used in practice. Under this framework, the unknown parameters are random quantities and their a priori uncertainty is expressed through the prior distribution which - after observing the outcomes - is updated to the posterior one to make inference. Trippa et al[23] have recently proposed a Bayesian adaptive randomization (BAR) strategy for testing multiple treatments in a trial for recurrent glioblastoma. The aim of BAR is to assign more patients to the best performing treatment, while allocating a similar number of subjects to the control arm. In particular, at each step $j + 1$, the probabilities of enrolling a patient to each experimental arm are based on the posterior probabilities that the effect of the given treatment at step j is higher than that of the control. The probability of assigning a subject to the control is based, instead, on the difference between the number of subjects assigned to the experimental group with the highest sample size and the sample size of the control arm at step j (see Trippa et al[23] for details).

Even if in our setting we do not fix a control arm (for a thorough discussion see Section 5), it is still possible to compare the DBCD procedure targeting ρ^C to BAR by selecting one of the treatments as the control. Moreover, as a benchmark we also consider the completely randomized design (CRD) intending to target ρ^B.

In the simulation study we consider the case of $K = 3$ treatments by taking into account a control arm with mean effect equal to 5 and two different scenarios corresponding to i) two active experimental arms $\theta = (10, 7, 5)^\top$ and ii) a single one $\theta = (10, 5, 5)^\top$. As regards BAR, we set a conjugate inverse-gamma prior on each model parameter (with shape and scale equal to 2 and θ_i, respectively) and the probabilities of assigning the patients to the treatments are computed via Monte Carlo approximation. Following the authors' suggestion, the tuning parametric functions for the experimental arms and the control (i.e. $\gamma(j)$ and $\eta(j)$ in their notation) are set equal to $(j/n)^{1.5}$ and $0.25 \cdot j/n$, respectively.

Besides the importance of testing the global null hypothesis (see Section 5), in many clinical trials the interest may be also focused on alternative power metrics. The pairwise comparisons (experimental vs. control) may be of interest and so the assessment of the design performance on a marginal power metric may be relevant to the experimenter. For instance, in many contexts, it may be important to identify a single treatment to recommend to patients instead of detecting any treatment with a potential therapeutic effect. On this purpose, we investigate the performance of our design, also wrt those of the BAR procedure, in terms of power of the test for the pairwise comparison between the experimental arm with the longest survival time and the control treatment.

In Table 6 we report the average power of the Wald test, the average size of the treatment group for which the mean survival time is the highest/lowest (nSup and nInf) and the total observed survival time (TS) for several sample sizes. We also show the values of the simulated average marginal power W_m, computed as the probability to reject the null of equality among the superior experimental arm and the control arm.

Under these scenarios, both the adaptive procedures are superior to the balanced design in terms of average power of the Wald test. However, the optimal constrained allocation rule presents higher power than BAR with a gap that tends to decrease as the sample size increases, going from 4.6% and 8.6% for scenario i) and ii), respectively, to 0.2%. Both the adaptive randomizations outperform the CRD in terms of marginal power as well, with a gain up to 17.6% in adopting ρ^C for $n = 100$. In scenario i) the RAR targeting ρ^C guarantees the highest values of W_m; in scenario ii) when $n \geq 150$ instead, the BAR leads to slightly higher values of the marginal power.

TABLE 6 Simulated average power of the Wald test (W) and marginal power (W_m), number of patients treated with the superior/inferior arm (nSup and nInf) and total survival time (TS) for the DBCD targeting ρ^C, the Bayesian adaptive procedure (BAR) and the CRD intending to target ρ^B for several sample sizes and 10000 iterations.

		$\theta = (10, 7, 5)^\top$					$\theta = (10, 5, 5)^\top$				
n	Design	W	W_m	nSup	nInf	TS	W	W_m	nSup	nInf	TS
	ρ^C	0.731	0.685	55	21	820	0.880	0.751	64	18	812
100	BAR	0.685	0.617	34	36	730	0.794	0.700	37	37	685
	ρ^B	0.654	0.575	34	33	734	0.741	0.575	34	33	667
	ρ^C	0.894	0.861	84	32	1234	0.973	0.899	97	26	1225
150	BAR	0.873	0.840	50	53	1097	0.941	0.901	55	55	1028
	ρ^B	0.862	0.820	50	50	1102	0.928	0.820	50	50	1002
	ρ^C	0.960	0.946	114	42	1648	0.996	0.965	132	34	1639
200	BAR	0.955	0.940	67	70	1461	0.988	0.973	74	74	1370
	ρ^B	0.947	0.928	66	66	1469	0.977	0.928	67	66	1335
	ρ^C	0.987	0.981	142	53	2067	0.999	0.990	165	43	2054
250	BAR	0.985	0.980	83	87	1824	0.997	0.993	92	92	1712
	ρ^B	0.983	0.977	83	83	1832	0.996	0.977	83	83	1666

The performance of the BAR procedure strongly depends on the choice of the tuning parametric functions. Indeed the BAR procedure may become more competitive in terms of marginal power as $\gamma(j)$ grows. We run further simulations in which we consider $\gamma(j) = 5 \cdot (j/n)^{1.5}$ and, for example, when $\theta = (10, 7, 5)^\top$ and $n = 100$, W_m increases to 0.712. However, as noted by Lin and Bunn,[30] since the control arm is usually the worst treatment, the allocation given by BAR is often not attractive to

clinicians due to a large proportion of patients randomized to the poorest performing arm. In this situation, by assigning more subjects to the best and the worst treatments, BAR presents good performance in terms of power, especially for large values of $\gamma(j)$. In these extreme cases, randomization probabilities to the intermediate arms become approximately null,[23] making BAR procedure affected by similar drawbacks to those of the optimal unconstrained target $\tilde{\rho}$.

The results of Table 6 also show how the implementation of ρ^C leads to a considerable benefit for patients when compared to the others procedures. In the presence of two effective experimental arms the optimal constrained target assigns around 23% more patients to the best performing treatment than BAR, while the proportion grows to 28% in the case of a single active experimental arm. At the same time our procedure allocates, on average, -12% and -15% subjects to the treatment arm exhibiting the lowest mean survival time wrt BAR (for scenario i) and ii), respectively). Moreover, our proposal has consistently higher average observed survival time compared with both BAR and CRD.

We also compare the procedures in terms of the minimum sample size necessary to ensure the average power of the Wald test greater than 80% for the experimental set-up $\theta = (10, 7, 5)^\top$. As regards ρ^C, 117 patients are required while for BAR at least 130 subjects should be enrolled. By using CRD, the minimum sample size reaches 133 units. These differences tend to decrease as the minimum power requested increases: the smallest sample size such that W$\geq$ 90% is 154 for ρ^C, 160 for BAR and 167 for ρ^B.

For completeness we also investigated the hypothetical scenario in which only one of the experimental treatments is better than the control. For instance, if the control arm has mean equal to 7 and the experimental arms have mean effects 10 and 5, an average loss of 1.5% of power is observed compared to the performance of BAR in setting i) of Table 6, that is matched with an improvement of the survival time (for $n = 100$, TS= 779 and for $n = 250$, TS = 1972).

These results provide further confirmation to the arguments in favour of the optimal constrained design as a compromise between statistical power and patients' benefit. Even in the case of designs with comparable performance in terms of power, the ethical gain induced by our procedure makes it more attractive from patients' point of view.

4 | OPTIMAL ALLOCATIONS FOR SURVIVAL TRIALS WITH RIGHT-CENSORING

Consider now a trial in which each patient has exponentially distributed survival time subjected to a random censoring time C. Assuming that his/her censoring is independent of the outcomes and it is the same for each treatment group, let ϵ_i be the probability that a patient belonging to the ith treatment group experience the event of interest (death/failure). Clearly, the probability ϵ depends on the censoring scheme adopted in the study and one of the most popular[20,22] has been introduced by Latta[31] and further studied by Rosenberger and Seshaiyer.[32] This scheme can be summarized as follows: let R be the total recruitment period and D the duration of the trial, subjects arrival times are assumed to be independent and uniformly distributed in $[0, R]$, while each patient is subjected to an independent censoring time over $(0, D)$. In this setting, for any group $i = 1, \ldots, K$

$$\epsilon_i = \epsilon(\theta_i) = 1 - \frac{\theta_i}{D} - \frac{2\theta_i^2}{RD} \exp\left\{-\frac{D}{\theta_i}\right\} - \frac{\theta_i}{D}\left(1 - 2\frac{\theta_i}{R}\right) \exp\left\{-\frac{D-R}{\theta_i}\right\} \tag{5}$$

is a monotonically non increasing function of θ_i, following the idea that the longer the expected survival time is, the smaller the probability for a patient to fail before censoring. Clearly, the likelihood for θ is modified (see e.g. Lawless[33]) and, after n assignments, the Fisher information matrix becomes $\mathbf{M}_\epsilon = \text{diag}\left(\pi_{in}\epsilon_i\theta_i^{-2}\right)_{i=1,\ldots,K}$ and, thus, $\sqrt{n}(\hat{\gamma}_n - \gamma) \xrightarrow{d} \text{N}(\mathbf{0}_{K-1}, \mathbf{\Sigma}_\epsilon)$, where $\mathbf{\Sigma}_\epsilon = \mathbf{A}\mathbf{M}_\epsilon^{-1}\mathbf{A}^\top$. Accordingly, the NCP becomes $\phi_\epsilon(\pi) = \gamma^\top\mathbf{\Sigma}_\epsilon^{-1}\gamma$ and, in general, it is hard to obtain a closed-form expression of both the constrained and unconstrained optimal targets. Nevertheless, solutions can be found numerically with standard optimization software (R, Matlab).

Under the censoring scheme in (5), Table 7 shows the behaviour of the unconstrained optimal target, denoted by $\tilde{\rho}^\epsilon$. As in the uncensored set-up, this target is degenerate and presents similar drawbacks to the one of Theorem 1, both from inferential and ethical viewpoints.

Therefore, we apply the same constrained optimization framework in (3), in which NCP is maximized under the ethical constraint reflecting the effectiveness of the treatments. Since the closed-form solution of this target, denoted by ρ^{C_ϵ}, is not available, in what follows we take into account a smoothing transformation of it in order to obtain a continuous target function implementable via DBCD. In particular, we consider the convolution of ρ^{C_ϵ} with a Gaussian kernel (see e.g. Tymofyeyev et al[13]) with $\sigma^2 = 1$.

Table 8 presents the constrained optimal target ρ^{C_ϵ} in the same setting, i.e. $K = 3$, $R = 55$ and $D = 96$. We display both the theoretical and the smoothed version (in italics). The target ρ^{C_ϵ} skews the assignments to the best performing treatment arm and $\rho_1^{C_\epsilon}$ is increasing in θ_1 (see Table 8b), whereas it is decreasing in θ_2 (Table 8a). Furthermore, ρ^{C_ϵ} and its smoothed

version substantially coincide: only small differences, of order 10^{-3}, are present. This behaviour was also confirmed by further computations, not reported here for brevity, and it suggests that the smoothed target should have similar performance to ρ^{C_ϵ}. Hence, from now on we will refer to the smoothed version of the constrained target with ρ^{C_ϵ}.

TABLE 7 Unconstrained optimal target $\tilde{\rho}^\epsilon$, under the right censoring scheme in (5) with $R = 55$ and $D = 96$.

θ	$\tilde{\rho}_1^\epsilon$	$\tilde{\rho}_2^\epsilon$	$\tilde{\rho}_3^\epsilon$
$(30, 10, 5)^\top$	0.876	0	0.124
$(20, 10, 5)^\top$	0.815	0	0.185
$(10, 10, 5)^\top$	0.336	0.336	0.328
$(10, 7, 5)^\top$	0.673	0	0.327
$(10, 5, 5)^\top$	0.674	0.163	0.163

TABLE 8 Theoretical and smoothed (in italics) optimal constrained target for $R = 55$, $D = 96$ and $\sigma = 1$.

(a) Fixed θ_1, θ_3 and decreasing θ_2.				(b) Fixed θ_2, θ_3 and increasing θ_1.			
θ	$\rho_1^{C_\epsilon}$	$\rho_2^{C_\epsilon}$	$\rho_3^{C_\epsilon}$	θ	$\rho_1^{C_\epsilon}$	$\rho_2^{C_\epsilon}$	$\rho_3^{C_\epsilon}$
$(10, 9, 5)^\top$	0.444	0.278	0.278	$(10, 8, 4)^\top$	0.552	0.224	0.224
	0.444	*0.278*	*0.278*		*0.550*	*0.225*	*0.225*
$(10, 7, 5)^\top$	0.594	0.203	0.203	$(15, 8, 4)^\top$	0.714	0.143	0.143
	0.594	*0.203*	*0.203*		*0.714*	*0.143*	*0.143*
$(10, 5, 5)^\top$	0.672	0.164	0.164	$(20, 8, 4)^\top$	0.786	0.107	0.107
	0.672	*0.164*	*0.164*		*0.782*	*0.109*	*0.109*

4.1 | Comparisons of optimal targets under an independent right censoring scheme

In the presence of censoring, the following optimal allocations have been derived in the literature. The allocation minimizing the trace of Σ_ϵ can be easily derived from the general result for heteroscedastic models,[27] from which we obtain

$$\rho_1^{A_\epsilon} = \frac{\frac{\theta_1}{\sqrt{\epsilon_1}}\sqrt{K-1}}{\frac{\theta_1}{\sqrt{\epsilon_1}}\sqrt{K-1} + \sum_{k=2}^{K}\frac{\theta_k}{\sqrt{\epsilon_k}}} \quad \text{and} \quad \rho_i^{A_\epsilon} = \frac{\frac{\theta_i}{\sqrt{\epsilon_i}}}{\frac{\theta_1}{\sqrt{\epsilon_1}}\sqrt{K-1} + \sum_{k=2}^{K}\frac{\theta_k}{\sqrt{\epsilon_k}}} \quad \text{for } i = 2, \dots, K.$$

The D-optimal design (ρ^{D_ϵ}) minimizing the determinant of Σ_ϵ can be obtained as a solution of a non linear system of equations and it can be found numerically.[22] In addition, based on non linear programming, Sverdlov et al[22] proposed two optimal allocations, ρ^{NP1} and ρ^{NP2}, adopting the same constrained optimization framework of Tymofyeyev et al.[13] In particular ρ^{NP1} and ρ^{NP2} were derived by minimizing the total sample size and the total expected hazard, respectively, under the constraints of a minimum desired (user-selected) proportion $T \in [0, K^{-1}]$ of subjects for each treatment group, and $\phi_\epsilon \geq U$, where U is a positive constant. Only ρ^{NP1} admits a closed-form expression, while ρ^{NP2} can be found numerically (the code was kindly provided by the corresponding author of Sverdlov et al[22]). Since both targets are discontinuous functions of the unknown model parameters, the authors apply a smoothing transformation to the targets using the above-mentioned multivariate Gaussian Kernel.

Remark 3. While ρ^{D_ϵ} and ρ^{C_ϵ} have their components ordered according to the treatment efficacies, ρ^{NP1} and ρ^{NP2} do not always fulfil this characteristic. Indeed, for $\theta = (34, 34, 24)^\top$ with $T = 0.2$, $R = 55$, $D = 96$ (namely under scenario IIc of Table I of Sverdlov et al[22]), the authors obtain $\rho^{NP1} = (0.3, 0.3, 0.4)^\top$ and $\rho^{NP2} = (0.32, 0.32, 0.36)^\top$, where the higher proportion of subjects is treated with the worst treatment. For ρ^{A_ϵ} similar considerations of Remark 2 hold: such allocation is ethical as

long as the reference treatment is also the superior one; whereas if e.g., $\theta = (25, 29, 30)^\top$ (with $R = 55$ and $D = 96$), then $\rho^{A_\epsilon} = (0.367, 0.310, 0.323)^\top$.

We now compare the constrained target in presence of censoring with those mentioned previously by considering $K = 3$, $R = 55$ and $D = 96$. Table 9 reports the theoretical values of ρ^{D_ϵ} and ρ^{A_ϵ}, the smoothed constrained optimal target ρ^{C_ϵ}, ρ^{NP1} and ρ^{NP2} with $T = 0.2$. The efficiency criteria described in Section 3.3 have been considered to measure the performance of the competing targets with respect to statistical and ethical considerations. In particular, the target whose power efficiency's values are closest to 1 is ρ^{C_ϵ}, except for $\theta = (7, 5, 4)^\top$, in which the loss wrt ρ^{NP1} and ρ^{NP2} is 3.1% and 1.5%, respectively. The constrained target outperforms the other allocations in terms of ethical efficiency (except in the first scenario in which ρ^{A_ϵ} gives nearly the same value). As far as estimation efficiency is concerned, ρ^{C_ϵ} shows very good performance in terms of A_A efficiency, always higher than 90.2%, while D_A efficiency is slightly lower. Note that, in all the experimental settings of Table 9 except the last one, ρ^{NP1} and ρ^{NP2} coincide by allocating the same proportion T of patients to the intermediate and the worst treatments. Finally, the proposed ρ^{C_ϵ} is superior to the balanced design in terms of power and ethical efficiency.

TABLE 9 Comparisons of optimal targets in presence of censoring for $K = 3$ treatments in different experimental scenarios wrt power efficiency $E_\phi(\rho)$, ethical efficiency $E_e(\rho)$, D_A efficiency $E_{D_A}(\rho)$, and A_A efficiency $E_{A_A}(\rho)$. Censoring scheme in (5) with $R = 55$ and $D = 96$.

θ	ρ	$E_\phi(\rho)$	$E_e(\rho)$	$E_{D_A}(\rho)$	$E_{A_A}(\rho)$
	$\rho^{A_\epsilon} = (0.625, 0.274, 0.101)^\top$	0.787	0.834	0.922	1
	$\rho^{D_\epsilon} = (0.450, 0.389, 0.161)^\top$	0.798	0.752	1	0.891
$(30, 20, 8)^\top$	$\rho^{C_\epsilon} = (0.684, 0.158, 0.158)^\top$	0.915	0.832	0.818	0.902
	ρ^B	0.762	0.644	0.888	0.702
	$\rho^{NP1} = \rho^{NP2} = (0.6, 0.2, 0.2)^\top$	0.900	0.787	0.884	0.929
	$\rho^{A_\epsilon} = (0.731, 0.150, 0.119)^\top$	0.875	0.813	0.841	1
	$\rho^{D_\epsilon} = (0.472, 0.293, 0.235)^\top$	0.646	0.632	1	0.788
$(30, 10, 8)^\top$	$\rho^{C_\epsilon} = (0.792, 0.104, 0.104)^\top$	0.904	0.854	0.744	0.973
	ρ^B	0.485	0.533	0.950	0.584
	$\rho^{NP1} = \rho^{NP2} = (0.6, 0.2, 0.2)^\top$	0.788	0.720	0.958	0.931
	$\rho^{A_\epsilon} = (0.663, 0.188, 0.149)^\top$	0.840	0.791	0.892	1
	$\rho^{D_\epsilon} = (0.452, 0.302, 0.246)^\top$	0.665	0.661	1	0.849
$(12, 5, 4)^\top$	$\rho^{C_\epsilon} = (0.734, 0.133, 0.133)^\top$	0.873	0.833	0.799	0.969
	ρ^B	0.528	0.583	0.959	0.670
	$\rho^{NP1} = \rho^{NP2} = (0.6, 0.2, 0.2)^\top$	0.817	0.750	0.943	0.980
	$\rho^{A_\epsilon} = (0.527, 0.263, 0.210)^\top$	0.727	0.834	0.953	1
	$\rho^{D_\epsilon} = (0.397, 0.329, 0.274)^\top$	0.665	0.789	1	0.935
$(7, 5, 4)^\top$	$\rho^{C_\epsilon} = (0.586, 0.207, 0.207)^\top$	0.770	0.853	0.903	0.981
	ρ^B	0.628	0.762	0.984	0.853
	$\rho^{NP1} = (0.519, 0.200, 0.281)^\top$	0.801	0.823	0.931	0.964
	$\rho^{NP2} = (0.576, 0.200, 0.224)^\top$	0.785	0.847	0.907	0.976

ρ^{A_ϵ}, A_A optimal design; ρ^{D_ϵ}, D_A optimal design; ρ^{C_ϵ}, smoothed version of the optimal constrained target; ρ^B, balanced design; ρ^{NP1} and ρ^{NP2}, Sverdlov et al[22] optimal targets (with minimum allocation proportion for each arm T equal to 0.2).

4.2 | RAR implementation under censoring

As in Section 3.4, we wish to implement DBCD to target the optimal constrained allocation by taking into account delayed responses and staggered entries. Indeed, in trials with time-to-event outcomes the response of patients to a given treatment is often unavailable before the randomization of the next subject. However, it has been shown[34] that DBCD is quite insensitive to delayed responses under widely satisfied conditions that hold for exponential responses.[20]

We run 10000 trials to evaluate the operating characteristics of the optimal constrained design for $K = 3$ in several illustrative examples. In this case, initial data for the RAR procedure were collected by allocating subjects with equal probabilities until two events were observed in each treatment arm. After that, at each step, subject j is randomized to treatment i with probability Ψ_{ji} (as in Section 3.4). We studied the convergence of the allocation proportion to the desired target and we estimated the power of the Wald and log-rank (LR) tests. The proportion of responses observed during the recruitment, i.e. the observations used in RAR ($\%obs$) and the percentage of patients allocated with RAR ($\%RAR$) will be reported in the Tables.

Table 10 shows the results under the experimental scenarios of Table 8a for different choices of the sample size. In these scenarios $\%obs=86\%$ and the proportion of patients allocated with RAR varies from 88% for $n = 150$ to 92-93% for $n = 300$. As n grows, the convergence to ρ^{C_ϵ} improves and the standard deviations become smaller. For $n \geq 150$ at least 83.7% of power is guaranteed for both tests in all the scenarios, while for $n \geq 250$ the simulated average power is always higher than 0.977.

TABLE 10 Simulation results for different sample sizes and scenarios of Table 8a. Estimates of model parameters $\hat{\theta}_n$, simulated allocation proportions π_n (their standard deviation in square brackets) and average power of Wald's test (W) and log-rank (LR) test, $R = 55$, $D = 96$ and 10000 iterations.

n	$\%RAR$	$\theta^\top$	$\hat{\theta}_n^\top$	ρ^{C_ϵ}	$\pi_n^\top$		W	LR
150	88%		(9.8, 8.9, 5.0)		(0.40, 0.33, 0.27)	[.110, .088, .051]	0.885	0.875
200	90%	(10, 9, 5)	(9.9, 8.9, 5.0)	$(0.444, 0.278, 0.278)^\top$	(0.40, 0.32, 0.28)	[.100, .079, .041]	0.962	0.875
250	91%		(9.9, 8.9, 5.0)		(0.40, 0.32, 0.28)	[.092, .073, .036]	0.987	0.986
300	92%		(9, 9, 8.9, 5.0)		(0.40, 0.32, 0.28)	[.088, .069, .034]	0.995	0.996
150	88%		(9.9, 6.9, 5.0)		(0.48, 0.27, 0.25)	[.107, .069, .052]	0.837	0.842
200	90%	(10, 7, 5)	(9.9, 6.9, 5.0)	$(0.594, 0.203, 0.203)^\top$	(0.48, 0.27, 0.25)	[.095, .0.6, .044]	0.938	0.931
250	91%		(9.9, 6.9, 5.0)		(0.49, 0.26, 0.25)	[.088, .054, .041]	0.977	0.975
300	92%		(9.9, 6.9, 5.0)		(0.49, 0.26, 0.25)	[.084, .050, .039]	0.992	0.991
150	89%		(9.9, 5.0, 5.0)		(0.55, 0.23, 0.22)	[.089, .054, .046]	0.944	0.947
200	90%	(10, 5, 5)	(10.0, 5.0, 5.0)	$(0.672, 0.164, 0.164)^\top$	(0.56, 0.22, 0.22)	[.075, .044, .037]	0.985	0.985
250	92%		(10.0, 5.0, 5.0)		(0.57, 0.22, 0.21)	[.065, .038, .032]	0.996	0.996
300	93%		(10.0, 5.0, 5.0)		(0.58, 0.21, 0.21)	[.058, .033, .029]	0.999	0.999

The skewness in favour of the best performing treatment is achieved by DBCD, but, as is well-known (see Hu et al[34] and reference therein), the convergence to the desired target is affected by delayed responses. To further investigate this aspect, we implement DBCD under experimental scenarios of Table 8b, for $n = 100$ and $n = 250$. We fixed $\theta_2 = 8$ and $\theta_3 = 4$ and the duration of the trial (namely $D = 96$) and we choose two different lengths of the recruitment period R; results are summarized in Table 11. Since ϵ_1 is decreasing in θ_1 and in R (i.e. the censoring in the data is higher for longer survival times and clearly longer recruitment period), it stands out how the effect on the convergence of an increasing sample size is a minor matter wrt the length of the recruitment, whose effect is as important as θ_1 increases. If the recruitment is longer, it is more likely that the number of responses observed, and used to allocate the next patient, is higher, resulting in a better convergence rate. From $R = 55$ to $R = 75$, the gain in terms of $\%obs$ is between 3% and 5%. An additional issue is related to $\%RAR$, which, is smaller for longer survival times. Because of the combination of these factors, better convergence is achieved when $\theta_1 = 10$, whereas, since for increasing θ_1 the target is more unbalanced in favour of treatment 1, DBCD procedure is slower in approaching ρ^{C_ϵ}. On the other hand, standard deviations are smaller for higher n and shorter R. With regard to power of the tests, all these scenarios present values $> 90\%$. Furthermore, note that, even if the convergence to the desired target is affected by delayed responses, when the superiority of the best treatment is pronounced (e.g., $\theta_1 = 15$ or 20), the simulated allocation proportion of subjects assigned to it is always more than twice wrt that of the other treatments.

TABLE 11 Simulation results for $n = 100, 250$ and 10000 iterations. Estimates of model parameters $\hat{\theta}_n$, simulated allocation proportions π_n (their standard deviation in square brackets) and average power of Wald's (W) and log-rank (LR) tests, different R and $D = 96$ for $\theta_2 = 8$ and $\theta_3 = 4$.

R	n	%obs	%RAR	θ_1	$\hat{\theta}_n^\top$	ρ^{C_ϵ}	$\pi_n^\top$	W	LR
55	100	86%	85%		(9.8, 7.8, 3.9)		(0.45, 0.30, 0.25) [.131, .102, .065]	0.910	0.904
	250	86%	91%		(9.9, 7.9, 4.0)	$(0.550, 0.225, 0.225)^\top$	(0.45, 0.29, 0.26) [.096, .070, .038]	1.000	1.000
75	100	89%	87%	10	(9.8, 7.7, 3.9)		(0.47, 0.30, 0.24) [.142, .111, .068]	0.907	0.896
	250	90%	92%		(9.9, 7.9, 4.0)	$(0.552, 0.224, 0.224)^\top$	(0.47, 0.28, 0.25) [.102, .074, .039]	0.999	1.000
55	100	81%	84%		(14.9, 7.8, 3.9)		(0.55, 0.24, 0.21) [.117, .081, .059]	0.996	0.995
	250	81%	90%		(15.0, 7.9, 4.0)	$(0.714, 0.143, 0.143)^\top$	(0.57, 0.22, 0.21) [.076, .046, .036]	1.000	1.000
75	100	85%	85%	15	(14.9, 7.7, 3.9)		(0.59, 0.21, 0.20) [.119, .081, .059]	0.996	0.994
	250	86%	92%		(15.0, 7.9, 4.0)	$(0.716, 0.142, 0.142)^\top$	(0.61, 0.20, 0.19) [.073, .044, .034]	1.000	1.000
55	100	76%	82%		(20.1, 7.8, 3.9)		(0.60, 0.20, 0.20) [.106, .068, .060]	0.999	0.999
	250	77%	90%		(20.0, 7.9, 4.0)	$(0.782, 0.109, 0.109)^\top$	(0.62, 0.19, 0.19) [.065, .038, .034]	1.000	0.999
75	100	81%	84%	20	(20.1, 7.7, 3.9)		(0.64, 0.19, 0.17) [.100, .067, .055]	1.000	0.999
	250	82%	91%		(20.0, 7.9, 4.0)	$(0.786, 0.107, 0.107)^\top$	(0.66, 0.17, 0.17) [.060, .035, .031]	1.000	0.999

The impact of the magnitude of the treatment effects on the convergence is additionally supported by the results in Table 12 which correspond to a set-up of a very rapid fatal disease. In this case the estimated allocation proportions are really close to the target, ensuring that 60% of patients or more receive the best treatment. These results show that, for small values of θ, our proposal has better performance even for small-moderate sample sizes, particularly common in clinical trials for rare diseases.

TABLE 12 Simulation results for different sample sizes, 10000 iterations and $\theta = (3, 2, 1)^\top$. Estimates of model parameters $\hat{\theta}_n$, simulated allocation proportions π_n (their standard deviation in square brackets) and average power of the Wald (W) and log-rank (LR) tests, $R = 55$, $D = 96$.

n	%obs	%RAR	$\hat{\theta}_n^\top$	ρ^{C_ϵ}	$\pi_n^\top$		W	LR
150	96%	93%	(3.0, 1.9, 1.0)		(0.61, 0.20, 0.19)	[.114, .074, .053]	0.998	0.999
200	96%	94%	(3.0, 2.0, 1.0)	$(0.634, 0.183, 0.183)^\top$	(0.61, 0.20, 0.19)	[.093, .058, .042]	1.000	1.000
250	96%	95%	(3.0, 2.0, 1.0)		(0.62, 0.19, 0.19)	[.080, .048, .037]	1.000	1.000
300	96%	95%	(3.0, 2.0, 1.0)		(0.62, 0.19, 0.19)	[.072, .042, .034]	1.000	1.000

Finally Table 13 shows the simulate average type I errors, for the Wald and log-rank (LR$_\alpha$) tests. Both of them are very close to the significance level, that was set to 0.05. Better convergence rate is achieved by the second scenario $\theta = (4, 4, 4)^\top$ as in this case the proportion of data used in the RAR procedure is greater than 90%, while in the first scenario is not higher than 81%.

TABLE 13 Simulation results of type I error for the Wald (W_α) and log-rank (LR_α) tests ($\rho^{C_\epsilon} = \rho^B$). Estimates of model parameters $\hat{\theta}_n$ and simulated allocation proportions π_n (their standard deviation in square brackets), with $R = 55$, $D = 96$; 10000 iterations.

n	%obs	%RAR	$\theta^\top$	$\hat{\theta}_n^\top$	$\pi_n^\top$		W_α	LR_α
150	80%	86%		(11.8, 12.0, 12.0)	(0.32, 0.35, 0.33)	[.085, .075, .079]	0.048	0.061
200	81%	88%		(11.9, 12.0, 12.0)	(0.32, 0.35, 0.33)	[.080, .070, .071]	0.051	0.057
250	81%	89%	(12, 12, 12)	(11.9, 12.0, 12.0)	(0.32, 0.35, 0.33)	[.075, .066, .066]	0.047	0.054
300	81%	90%		(11.9, 12.0, 12.0)	(0.32, 0.35, 0.33)	[.072, .064, .064]	0.047	0.053
150	92%	91%		(3.9, 4.0, 4.0)	(0.33, 0.34, 0.33)	[.095, .077, .083]	0.048	0.056
200	92%	92%		(3.9, 4.0, 4.0)	(0.33, 0.34, 0.33)	[.089, .071, .072]	0.048	0.055
250	93%	93%	(4, 4, 4)	(3.9, 4.0, 4.0)	(0.33, 0.34, 0.33)	[.084, .068, .068]	0.047	0.055
300	93%	94%		(3.9, 4.0, 4.0)	(0.33, 0.33, 0.34)	[.081, .066, .066]	0.046	0.054

4.3 | Robustness of our methodology to model misspecifications and sensitivity analysis

In this section we investigate the performance of the optimal constrained target when the distribution of the survival times is Weibull (Weib), log-logistic (LL) or log-normal (LN) to take into account different shapes of the hazard function. In particular we consider Weib with parameters $s = 0.8$ and 1.25 (monotone increasing and decreasing hazard, respectively), LL distribution with $s = 0.5$ and the LN distribution with $s = 0.8$ (non-monotone hazards).

As it is well-known, in both conventional and adaptive designs, wrong parametric assumptions often lead to power loss and type I error inflation of parametric tests.[35] In our robustness studies, besides the log-rank test, we consider the performance of the Wald test based on the correctly specified distribution of the survival times (as also suggested by Sverdlov et al[35]). More specifically, we compute the test statistic by analysing the final dataset with the distribution according to which the survival times were generated and we report the average power/type I error of Wald's test ($\tilde{W}$ and $\tilde{W}_\alpha$). We summarize in Table 14 the simulation results under different models for the survival times for the DBCD targeting ρ^{C_ϵ} and the CRD intending to target ρ^B, taking into account $n = 250$ patients. These results show that our proposal is - in general - slightly more powerful and it results in a fewer average number of deaths than the balanced design even when the event time distribution is non exponential. The maximum loss of power wrt the one under the exponential model for ρ^{C_ϵ} occurs under a monotone increasing hazard function (-12% and -10% for $\tilde{W}$ and LR respectively) which is, at the same time, associated to a smaller average number of events in the study (-3 deaths). Conversely, when survival times follow a Weib with $s = 1.25$, higher values of power are present. In the case of non-monotone hazard functions an improvement in terms of power is combined with a substantial reduction in the number of deaths in the trial. The type I error is around the nominal level in most of the scenarios considered; a slight inflation is observed in the case of non-monotone hazards. We conclude that our procedure is quite robust to model misspecification in all the considered experimental set-ups for both the log-rank test and the parametric test $\tilde{W}$.

Another form of model misspecification may occur when the patients' accrual rate is non-uniform; to take into account several recruitment patterns we consider the Beta distribution.[35] The cases Beta $(1, 5)$ and Beta $(5, 1)$ (right/left skewed) encompass situations in which the accrual rate decreases/increases over time respectively. On the other hand, Beta $(1/5, 1/5)$ represents scenarios with accelerated recruitment at the beginning and at the end of the trial and Beta $(5, 5)$ refers to studies in which the highest recruitment rate occurs in the middle of the trial. Table 15 shows the simulating operating characteristics of the DBCD targeting ρ^{C_ϵ} and the CRD targeting ρ^B. We report the average proportion of patients whose outcomes are observed during the recruitment, the simulated average power of the Wald test/type I error and the total average number of events for $n = 250$ patients. The power of the Wald test is similar across the five recruitment patterns with a maximum loss of 1.2% corresponding to the left-skewed Beta distribution. On the other hand, under Beta$(1/5, 1/5)$ a slight increase in power is observed. The type I error rate is close to 0.05 showing that our proposal is also robust to misspecification of the recruitment pattern.

4.4 | Redesign of KEYNOTE-010 clinical trial

In this section we illustrate the application of the constrained optimal target by redesigning the KEYNOTE-010 clinical trial[8] (registered at ClinicalTrials.gov, number NCT01905657). The aim of this Phase II/III study was to compare two doses of pembrolizumab (MK-3475) versus docetaxel in patients with non-small cell lung cancer and whose tumors were assessed as being

PD-L1 positive. Between Aug 2013 and Feb 2015, 1034 participants were enrolled and were randomly allocate with a 1:1:1 ratio, with a central interactive voice-response system, to receive treatment A (pembrolizumab 2 mg/kg), treatment B (pembrolizumab 10 mg/kg) or treatment C (docetaxel). Among primary endpoints (overall survival and progression-free survival) we were interested in the overall survival in the intent-to-treat population. At the cut-off date, after 23 months, median overall survival was 10.4 months in group A, 12.7 months in group B, and 8.5 months in group C. In designing this study, the authors assumed exponential distribution for overall survival, so that we run 10000 trials with $n = 1034$ patients with $\boldsymbol{\theta} = (\theta_A, \theta_B, \theta_C)^\top = (15, 18, 12)^\top$, adopting the above-mentioned censoring scheme with $R = 18$ and $D = 23$. The DBCD procedure has been implemented to target the optimal allocations ρ^{C_ϵ}, ρ^{A_ϵ}, ρ^{D_ϵ} and ρ^{NP1} (with $T = 0.2$) and we included the CRD to target ρ^B, as sequential analogue to the equal allocation adopted the original trial. Results are presented in Table 16, in which, for sake of comparison, we have also reported the simulated average efficiencies (see Section 3.3), number of deaths, total observed survival time, number of patients assigned to the inferior and to the superior treatment and the total expected hazard in the study ($H = \sum_{i=1}^{K} N_{in}/\theta_i$). In all the procedures, %obs and %RAR were 60% and 90-91% respectively. The optimal constrained design presents the highest degree of skewness in favour of the best performing treatment arm (B in this case) wrt to the other designs and the ethical property of Remark 1 does not hold for $NP1$ and the A_A allocations (see also Remark 3). All the procedures present similar statistical properties of power and estimation efficiency. Nevertheless, our proposal is superior in terms of ethical characteristics, resulting in 4-8 fewer deaths, longer total survival time, lower average hazard and higher ethical efficiency compared to the other designs. In addition, nInf and nSup highlight the potential advantages in adopting the optimal constrained design in which from 20 up to 74 fewer patients are assigned to the inferior treatment, whereas 428 patients receive the superior treatment, so that 24-83 more patients are injected with treatment B with the maximum gain achieved wrt the balanced design.

Note that this result provide further significance to the ethical definition of Remark 1. Given the gravity of the outcome, a response-adaptive procedure targeting ρ^{C_ϵ} would provide a better trade-off between statistical power and patients' benefits.

TABLE 14 Simulated average power/type I error of the Wald test ($\tilde{W}$ and $\tilde{W}_\alpha$) based on the correctly specified model and the log-rank test (LR and LR$_\alpha$), average number of deaths (Deaths; with their standard deviations in brackets) for the DBCD targeting ρ^{C_ϵ} and the CRD (intended to target ρ^B) for $n = 250$ and 10000 iterations.

	$\boldsymbol{\theta} = (10, 7, 5)^\top$						$\boldsymbol{\theta} = (12, 12, 12)^\top$					
Model	ρ^{C_ϵ}			ρ^B			ρ^{C_ϵ}			ρ^B		
	$\tilde{W}$	LR	Deaths	$\tilde{W}$	LR	Deaths	$\tilde{W}_\alpha$	LR$_\alpha$	Deaths	$\tilde{W}_\alpha$	LR$_\alpha$	Deaths
Exp	0.977	0.975	229 (4)	0.970	0.972	231 (4)	0.047	0.054	218 (5)	0.044	0.051	218 (5)
Weib, $s = 0.8$	0.859	0.870	226 (5)	0.842	0.870	228 (4)	0.050	0.052	213 (5)	0.044	0.050	213 (5)
Weib, $s = 1.25$	0.999	0.998	230 (4)	0.999	0.998	232 (4)	0.049	0.054	221 (5)	0.048	0.053	221 (4)
LL, $s = 0.5$	0.996	0.985	218 (5)	0.996	0.983	221 (5)	0.061	0.055	203 (6)	0.060	0.051	203 (6)
LN, $s = 0.8$	0.999	0.996	221 (5)	0.999	0.995	223 (5)	0.062	0.057	206 (6)	0.067	0.055	206 (6)

TABLE 15 Simulated average power/type I error of the Wald test (W and W$_\alpha$), average percentage of observation during the recruitment period (%obs; only for the DBCD procedure), average number of deaths (Deaths; with their standard deviations in brackets) for the DBCD targeting ρ^{C_ϵ} and the CRD (intended to target ρ^B) under different recruitment patterns, $n = 250$ patients and 10000 iterations.

	$\boldsymbol{\theta} = (10, 7, 5)^\top$					$\boldsymbol{\theta} = (12, 12, 12)^\top$				
Recruitment	ρ^{C_ϵ}			ρ^B		ρ^{C_ϵ}			ρ^B	
	%obs	W	Deaths	W	Deaths	%obs	W$_\alpha$	Deaths	W$_\alpha$	Deaths
Uniform	86%	0.977	229 (5)	0.970	231 (4)	81%	0.047	218 (5)	0.045	218 (5)
Beta $(1, 5)$	95%	0.974	230 (4)	0.974	231 (4)	91%	0.048	219 (5)	0.045	219 (5)
Beta $(5, 1)$	59%	0.965	230 (5)	0.972	231 (4)	48%	0.052	217 (5)	0.045	217 (5)
Beta $(1/5, 1/5)$	67%	0.980	229 (5)	0.975	231 (4)	64%	0.054	218 (5)	0.048	218 (5)
Beta $(5, 5)$	89%	0.972	230 (4)	0.972	231 (4)	81%	0.052	219 (5)	0.045	219 (5)

TABLE 16 Redesign of KEYNOTE-010[8] clinical trial

	Statistical properties					Ethical characteristics					
	W	LR	E_ϕ	E_{D_A}	E_{A_A}	Deaths	TS	nInf	nSup	H	E_e
$\boldsymbol{\pi}_n^{A_\epsilon} = (0.30, 0.34, 0.36)^\top$ [.028, .031, .029]	0.89	0.90	0.74	0.97	0.98	445	6533	365	354	70	0.83
$\boldsymbol{\pi}_n^{D_\epsilon} = (0.34, 0.36, 0.30)^\top$ [.020, .019, .019]	0.88	0.89	0.73	0.99	0.91	441	6569	311	376	71	0.84
$\boldsymbol{\pi}_n^{C_\epsilon} = (0.31, 0.41, 0.28)^\top$ [.053, .079, .035]	0.88	0.88	0.76	0.98	0.94	437	6595	291	428	69	0.86
$\boldsymbol{\pi}_n^{B} = (0.33, 0.34, 0.33)^\top$ [.007, .007, .007]	0.88	0.89	0.72	0.98	0.87	444	6546	345	345	71	0.83
$\boldsymbol{\pi}_n^{NP1} = (0.29, 0.39, 0.32)^\top$ [.057, .065, .047]	0.89	0.89	0.77	0.96	0.91	441	6570	333	404	70	0.85

5 | CONCLUSIONS AND FUTURE RESEARCH

The design of multi-arm clinical experiments is complex, especially when different objectives are involved. The proposed optimal constrained target guarantees a valid trade-off between the inferential goal of maximizing the power of Wald's test of homogeneity and the ethical demand of preserving subjects' care, whereas some targets for time-to-event outcomes could lead to undesirable allocation proportions, as we pointed out in this paper.

In addition, we implement our proposal via DBCD methodology. To assess the simulated operating characteristics of the design, we considered several experimental settings. In the uncensored model, we assumed that responses are available immediately and the convergence rate to the target is excellent. In the presence of censored observations we have also included delayed responses and staggered entries, which could seriously slowed down the convergence of the design. To assess such an impact, we run simulation studies directed to illustrate how the convergence is related to the complex interplay between sample size, survival times, length of recruitment and duration of the trial. Overall, the optimal constrained design reaches a good rate of convergence, provided that a sufficient amount of responses are observed throughout the recruitment phase to let the adaptive procedure work. Moreover, even in experimental scenarios with slower convergence, a fair degree of skewness in favour of the most promising treatment is achieved. The practical applicability of our proposal has been also highlighted by performing robustness studies to model misspecification and by redesigning a real lung cancer trial. Authors are working on developing a user-friendly interface (R-based Shiny Web application) to implement the procedure, which will be available in the future.

The problem of testing the null-hypothesis of equality among treatment effects considered in this paper is useful in many applications. In multi-arm trials a global test comparing all treatments can be carried out prior to making individual pairwise comparisons.[13,21,22] Indeed the overall null-hypothesis is the first stage of multiple comparison methodologies for several stepwise procedures.[25,26] One of the most powerful is the Fisher's least significant difference method which is a two-step test for pairwise comparisons; in the first stage the overall null-hypothesis of homogeneity is tested at level α and then, in case of rejection, all the pairs of interest are tested for equality at the same level of significance. As it is well-known, the design should be tailored on trial objectives: a single global test may be of interest, for instance, in Phase II trials.[36] Especially in anticancer research, due to the dramatic increase of new potential drugs under development, one of the primary Phase II objectives is to evaluate the effect of new treatments and to identify the one(s) that most warrants additional evaluation in a larger Phase III study.[37,38] Others interesting set-ups concern trials in which the choice of the control treatment is not unique: for instance, studies in which the new drug has to be compared to placebo, current commercial products, competitors products (see the comments by Owen and by Bechhofer and Tamhane in the discussion of Hedayat et al[39]). Finally there are situations such that no treatments with demonstrated efficacy exist and so no standard of care is available to be set as an appropriate/fixed control. This is particularly related to rapidly emerging novel infectious diseases such as Ebola[40] or COVID-19. In these cases clinical studies must start quickly and the overall null-hypothesis, as a first step, allows to evaluate several candidate treatments at once.

However, in many other experimental set-ups, the interest may not be solely focused on the power to reject the global null. For instance in the presence of a control treatment, as the pairwise comparisons become relevant, it is important to assess the design performance in terms of marginal power metrics. By taking into account the null hypothesis of equality among the treatment

arm with the longest survival time and that of the control arm, the marginal power induced by our proposal is clearly lower wrt the global one. The lack of marginal power is not generally severe: it may be mild even for moderate sample sizes and tends to vanish as the sample size grows. This loss has an impact, whilst modest, when determining trial sample size.

A further significant aspect of this work is to provide support to the arguments in favour of unbalanced allocation designs.[2,3] Unequal randomization is often more appropriate for patients' health and more powerful than the balanced design, especially for heteroscedastic treatment groups, and we showed that the proposed constrained target shares this property. Hence, the indiscriminate use of the popular equal allocation design in clinical trials should be reconsidered.

The promising performance of this optimization approach lead to further methodological developments to extend its applicability. One of the main directions of future research is to adopt this framework to derive optimal constrained targets for widely used heteroscedastic models like, e.g., for binary trials. Another interesting issue consists in including covariates/prognostic factors, to take also into account patients' heterogeneity.

ACKNOWLEDGMENTS

References

1. Freedman B. Equipoise and the ethics of clinical research. *N Engl J Med* 1987; 317(3): 141–145.

2. Peckham E, Brabyn S, Cook L, et al. The use of unequal randomisation in clinical trials: An update. *Contemp Clin Trials* 2015; 45: 113–122. 10th Anniversary Special Issue.

3. Sverdlov O, Ryeznik Y. Implementing unequal randomization in clinical trials with heterogeneous treatment costs. *Stat Med* 2019; 38(16): 2905–2927.

4. Avins AL. Can unequal be more fair? Ethics, subject allocation, and randomised clinical trials. *J Med Ethics* 1998; 24(6): 401–408.

5. Berry DA. Adaptive Clinical Trials: The Promise and the Caution. *J Clin Oncol* 2011; 29(6): 606–609.

6. Zelen M. Application of exponential models to problems in cancer research. *J R Stat Soc Series A Stat Soc* 1966; 129(3): 368–398.

7. Shulz D, Chernichovsky D, Allweis C. A novel method for quantifying passive-avoidance behavior based on the exponential distribution of step-through latencies. *Pharmacol Biochem Behav* 1986; 25(5): 979–983.

8. Herbst RS, Baas P, Kim D, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *Lancet* 2015; 387(10027): 1540–1550.

9. Fountzilas G, Ciuleanu E, Dafni U, et al. Concomitant radiochemotherapy vs radiotherapy alone in patients with head and neck cancer. *Med Oncol* 2004; 21: 95–107.

10. Cook RD, Wong WK. On the equivalence of constrained and compound optimal designs. *J Am Stat Assoc* 1994; 89(426): 687–692.

11. Jennison C, Turnbull BW. *Group Sequential Methods with Applications to Clinical Trials.* Boca Raton, UK: CRC Press; 1999.

12. Baldi Antognini A, Rosenberger WF, Wang Y, et al. Exact optimum coin bias in Efron's randomization procedure. *Stat Med* 2015; 34(28): 3760–3768.

13. Tymofyeyev Y, Rosenberger WF, Hu F. Implementing optimal allocation in sequential binary response experiments. *J Am Stat Assoc* 2007; 102(477): 224–234.

14. Biswas A, Mandal S, Bhattacharya R. Multi-treatment optimal response-adaptive designs for phase III clinical trials. *J Korean Stat Soc* 2011; 40(1): 33–44.

15. Baldi Antognini A, Novelli M, Zagoraiou M. Optimal designs for testing hypothesis in multiarm clinical trials. *Stat Methods Med Res* 2019; 28(10-11): 3242–3259.

16. Sverdlov O, Ryeznik Y, Wong WK. On optimal designs for clinical trials: An updated review. *J Stat Theory Pract* 2020; 14(10).

17. Rosenberger WF, Lachin JM. *Randomization in clinical trials: theory and practice*. New York, NY: Wiley; 2002.

18. Baldi Antognini A, Giovagnoli A. *Adaptive Designs for Sequential Treatment Allocation*. Boca Raton, UK: CRC Press; 2015.

19. US Food and Drug Administration. Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics. *Available online*. 2019.

20. Zhang L, Rosenberger WF. Response-adaptive randomization for survival trials: the parametric approach. *J R Stat Soc Series C Appl Stat* 2007; 56(2): 153–165.

21. Zhu H, Hu F. Implementing optimal allocation for sequential continuous responses with multiple treatments. *J Stat Plan Inference* 2009; 139(7): 2420–2430.

22. Sverdlov O, Tymofyeyev Y, Wong WK. Optimal response-adaptive randomized designs for multi-armed survival trials. *Stat Med* 2011; 30(24): 2890–2910.

23. Trippa L, Lee EQ, Wen PY, et al. Bayesian adaptive randomized trial design for patients with recurrent glioblastoma. *J Clin Oncol* 2012; 30(26): 3258–3263.

24. Hu F, Zhang LX. Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials. *Ann Stat* 2004; 32(1): 268–301.

25. Hochberg Y, Tamhane A. *Multiple Comparison Procedures*. New York, NY: Wiley; 1987.

26. Christensen R. *Plane Answers to Complex Questions*. New York, NY: Springer-Verlag; 2002.

27. Sverdlov O, Rosenberger WF. On recent advances in optimal allocation designs in clinical trials. *J Stat Theory Pract* 2013; 7(4): 753–773.

28. Wong WK, Zhu W. Optimum treatment allocation rules under a variance heterogeneity model. *Stat Med* 2008; 27(22): 4581–4595.

29. Rosenberger WF, Hu F. Maximizing power and minimizing treatment failures in clinical trials. *Clin Trials* 2004; 1(2): 141–147.

30. Lin J, Bunn V. Comparison of multi-arm multi-stage design and adaptive randomization in platform clinical trials. *Contemp Clin Trials* 2017; 54: 48–59.

31. Latta RB. A Monte Carlo study of some two-sample rank tests with censored data. *J Am Stat Assoc* 1981; 76(375): 713–719.

32. Rosenberger WF, Seshaiyer P. Adaptive survival trials. *J Biopharm Stat* 1997; 7(4): 617–624.

33. Lawless J. *Statistical Models and Methods for Lifetime Data*. New York, NY: Wiley; 2011.

34. Hu F, Zhang LX, Cheung SH, et al. Doubly adaptive biased coin designs with delayed responses. *Can J Stat* 2008; 36(4): 541–559.

35. Sverdlov O, Rosenberger WF, Ryeznik Y. Utility of Covariate-Adjusted Response-Adaptive Randomization in survival trials. *Stat Biopharm Res* 2013; 5(1): 38–53.

36. Jung S, George S. Between-arm comparisons in randomized Phase II trials. *J Biopharm Stat* 2009; 19(3): 456–468.

37. Simon R, Wittes R, Ellenberg S. Randomized Phase II clinical trial. *Cancer Treat Rep* 1986; 69(12): 1375–1381.

38. Lee J, Feng L. Randomized phase II designs in cancer clinical trials: current status and future directions. *J Clin Oncol* 2005; 23(19): 4450–4457.

39. Hedayat AS, Jacroux M, Majumdar D. Optimal designs for comparing test treatments with controls (with discussion). *Statist Sci* 1988; 3(4): 462–476.

40. Magaret A, Angus D, Adhikari N, et al. Design of a multi-arm randomized clinical trial with no control arm. *Contemp Clin Trials* 2016; 46: 12–17.

APPENDIX

A PROOFS

A.1 Proof of Theorem 1

The following Lemma is preliminary to the proof.

Lemma 4. $\phi(\cdot)$ is a concave function of the vector of allocation proportions.

Proof of Lemma 4. Let $\omega_i = (\rho_i \theta_i^{-2})/(\sum_{k=1}^{K} \rho_k \theta_k^{-2})$, then $\omega_i \geq 0$ for $i = 1, \ldots, K$ and $\sum_{i=1}^{K} \omega_i = 1$, namely $\omega = (\omega_1, \ldots, \omega_K)^\top$ could be regarded as a pdf of a (non-negative) discrete r.v. θ with K support points $\theta_1 \geq \ldots \geq \theta_K > 0$. So, letting

$$\bar{\theta}_\omega = \sum_{i=1}^{K} \theta_i \omega_i = \frac{\sum_{i=1}^{K} \rho_i \theta_i^{-1}}{\sum_{i=1}^{K} \rho_i \theta_i^{-2}} \tag{A1}$$

and $M_\omega(\theta^2) = \sum_{i=1}^{K} \theta_i^2 \omega_i = \left(\sum_{i=1}^{K} \rho_i \theta_i^{-2} \right)^{-1}$, then by (1), it is possible to show that

$$\phi(\rho) = \frac{M_\omega(\theta^2) - \bar{\theta}_\omega^2}{M_\omega(\theta^2)} = \frac{V_\omega(\theta)}{M_\omega(\theta^2)}, \tag{A2}$$

where $V_\omega(\theta)$ is the variance of θ evaluated w.r.t. ω. Moreover,

$$\phi(\rho) = \frac{1}{M_\omega(\theta^2)} \sum_{i=1}^{K} (\theta_i - \bar{\theta}_\omega)^2 \omega_i = \sum_{i=1}^{K} a_i \rho_i, \tag{A3}$$

where $a_i = \left(1 - \bar{\theta}_\omega/\theta_i \right)^2$, for $i = 1, \ldots, K$. Note that, for $i = 1, \ldots, K$, $\frac{\partial \phi(\rho)}{\partial \rho_i} = \sum_{j=1}^{K} \frac{\partial a_j}{\partial \rho_i} \rho_j + a_i$ where

$$\frac{\partial a_j}{\partial \rho_i} = 2 \left(1 - \frac{\bar{\theta}_\omega}{\theta_j} \right) \left(-\frac{1}{\theta_j} \right) \frac{\partial \bar{\theta}_\omega}{\partial \rho_i} \quad \text{and} \quad \frac{\partial \bar{\theta}_\omega}{\partial \rho_i} = \frac{\theta_i^{-1}}{\sum_{i=1}^{K} \rho_i \theta_i^{-2}} \left(1 - \frac{\bar{\theta}_\omega}{\theta_i} \right).$$

Thus,

$$\frac{\partial \phi(\rho)}{\partial \rho_i} = a_i - \frac{2}{\theta_i} \left(1 - \frac{\bar{\theta}_\omega}{\theta_i} \right) \sum_{j=1}^{K} \left\{ \frac{\rho_j \theta_j^{-1}}{\sum_{i=1}^{K} \rho_i \theta_i^{-2}} \left(1 - \frac{\bar{\theta}_\omega}{\theta_j} \right) \right\} = a_i,$$

and

$$\frac{\partial^2 \phi(\rho)}{\partial \rho_i^2} = -\frac{2 M_\omega(\theta^2)}{\theta_i^2} \left(1 - \frac{\bar{\theta}_\omega}{\theta_i} \right)^2, \quad \frac{\partial^2 \phi(\rho)}{\partial \rho_i \partial \rho_j} = -\frac{2 M_\omega(\theta^2)}{\theta_i \theta_j} \left(1 - \frac{\bar{\theta}_\omega}{\theta_i} \right) \left(1 - \frac{\bar{\theta}_\omega}{\theta_j} \right).$$

Thus, the Hessian matrix is given by

$$H_\phi(\rho) = -2 M_\omega(\theta^2) \left[\left(1 - \frac{\bar{\theta}_\omega}{\theta_1} \right), \ldots, \left(1 - \frac{\bar{\theta}_\omega}{\theta_K} \right) \right]^\top \left[\left(1 - \frac{\bar{\theta}_\omega}{\theta_1} \right), \ldots, \left(1 - \frac{\bar{\theta}_\omega}{\theta_K} \right) \right], \tag{A4}$$

having one eigenvalue equal to 0 (with multiplicity $K - 1$) and a non-null eigenvalue given by its trace, i.e. $tr\{H_\phi(\rho)\} = -2 M_\omega(\theta^2) \sum_{i=1}^{K} \left(1 - \frac{\bar{\theta}_\omega}{\theta_i} \right)^2 < 0$, which implies the concavity of NCP.

Proof of Theorem 1

Let $CV_\omega(\theta)$ be the coefficient of variation of θ evaluated with respect to ω, by (A2) the NCP can be rewritten as $\phi(\rho) = V_\omega(\theta)\left(V_\omega(\theta) + \bar{\theta}_\omega^2\right)^{-1} = \left(1 + \frac{1}{CV_\omega^2(\theta)}\right)^{-1}$.

Within the class of pdfs with a given mean $\bar{\theta}_\omega$, the variance $V_\omega(\theta)$ is maximized by the one that assigns all the mass of probability to the extremes, in this case θ_1 and θ_K. So let $P(\theta = \theta_K) = \omega^* = 1 - P(\theta = \theta_1)$, then $\bar{\theta}_\omega = \theta_1(1 - \omega^*) + \theta_K \omega^*$ and $V_\omega(\theta) = (\theta_1 - \theta_K)^2 \omega^*(1 - \omega^*)$. Thus, $CV_\omega^2(\theta) = \left[\theta_1(1 - \omega^*) + \theta_K \omega^*\right]^{-2}(\theta_1 - \theta_K)^2 \omega^*(1 - \omega^*)$, where $\frac{\partial CV_\omega}{\partial \omega^*} = 0 \Leftrightarrow \omega^* = \theta_1/(\theta_1 + \theta_K)$. In such a case, $\omega^* = \frac{\rho_K}{\theta_K^2}\left(\frac{\rho_1}{\theta_1^2} + \frac{\rho_K}{\theta_K^2}\right)^{-1}$, thus $\rho_K = \theta_K/(\theta_1 + \theta_K) = 1 - \rho_1$ and, from (A2), $\phi(\rho) = \left(\frac{\theta_1 - \theta_K}{\theta_1 + \theta_K}\right)^2$.

As regards statement i), every allocation $\tilde{\rho}$ such that $\sum_{i=1}^{j} \tilde{\rho}_i = \frac{\theta_1}{\theta_1 + \theta_K} = 1 - \sum_{i=h}^{K} \tilde{\rho}_i$ is optimal. Indeed, in this case $\bar{\theta}_\omega = \left(\sum_{i=1}^{K} \frac{\tilde{\rho}_i}{\theta_i^2}\right)^{-1} \sum_{i=1}^{K} \frac{\tilde{\rho}_i}{\theta_i} = \frac{2\theta_1 \theta_K}{\theta_1 + \theta_K}$ and thus, from (A3), $\phi(\tilde{\rho}) = \left(1 - \frac{\bar{\theta}_\omega}{\theta_1}\right)^2 \frac{\theta_1}{\theta_1 + \theta_K} + \left(1 - \frac{\bar{\theta}_\omega}{\theta_K}\right)^2 \frac{\theta_K}{\theta_1 + \theta_K} = \left(\frac{\theta_1 - \theta_K}{\theta_1 + \theta_K}\right)^2$.

The proof of (ii) is straightforward. Moreover, since the Hessian matrix in (A4) is negative semi-definite we have to check the stationary points. By setting the partial derivatives of the Lagrangian $L(\rho, \lambda) = \phi(\rho) - \lambda\left(\sum_{i=1}^{K} \rho_i - 1\right)$ equal to zero, we obtain a system of K equations $a_i = \lambda$ for $i = 1, \ldots, K$, which admits solutions if and only if $a_1 = \cdots = a_K$. Notice that, for every $i = 1, \ldots, K - 1$,

$$a_i \geq a_{i+1} \Leftrightarrow \left(\frac{\bar{\theta}_\omega}{\theta_i} + \frac{\bar{\theta}_\omega}{\theta_{i+1}}\right) \leq 2 \quad \text{or} \quad \theta_i = \theta_{i+1} \tag{A5}$$

and, since $\theta_K < \bar{\theta}_\omega < \theta_1$, from $a_1 = a_K$ we obtain

$$\left(1 - \frac{\bar{\theta}_\omega}{\theta_1}\right)^2 = \left(1 - \frac{\bar{\theta}_\omega}{\theta_K}\right)^2 \Leftrightarrow \bar{\theta}_\omega = 2\left(\frac{1}{\theta_K} + \frac{1}{\theta_1}\right)^{-1} = \frac{2\theta_1 \theta_K}{\theta_1 + \theta_K}. \tag{A6}$$

Clearly, from (A5), if $\theta_i = \theta_{i+1}$ then $a_i = a_{i+1}$. Furthermore, if $\exists j \in \{1, \ldots, K - 1\}$ such that $\theta_1 = \ldots = \theta_j > \theta_{j+1} \geq \ldots \geq \theta_K$ then $a_1 = a_2 = \ldots = a_j$; since $\theta_i \neq \theta_1$ for $i = j + 1, \ldots, K$, from (A5) follows that $a_i = a_1 \Leftrightarrow \bar{\theta}_\omega\left(\frac{1}{\theta_1} + \frac{1}{\theta_i}\right) = 2 \ \forall i = j + 1, \ldots, K$. By substituting $\bar{\theta}_\omega$ in (A6), then $\theta_i = \theta_K$, for every $i = j + 1, \ldots, K$.

On the other hand, if $\exists h \in \{2, \ldots, K\}$ such that $\theta_1 \geq \ldots \geq \theta_{h-1} > \theta_h = \ldots = \theta_K$, then $a_h = \ldots = a_K$. Clearly, $\theta_i \neq \theta_K$ for $i = 1, \ldots, h - 1$ and therefore, from (A5), $a_i = a_K \Leftrightarrow \bar{\theta}_\omega\left(\frac{1}{\theta_i} + \frac{1}{\theta_K}\right) = 2$. Thus, by substituting $\bar{\theta}_\omega$, it follows that $\theta_1 = \theta_i$ for every $i = 1, \ldots, h - 1$.

A.2 Proof of Theorem 2

The following Lemma is preliminary to the proof.

Lemma 5. Given a non-degenerate target ρ (i.e. such that $\rho_j > 0$ for every j), we have

(i) If $\bar{\theta}_\omega \in \left[\frac{2\theta_K \theta_{K-1}}{\theta_K + \theta_{K-1}}, \frac{2\theta_1 \theta_2}{\theta_1 + \theta_2}\right]$, then $\exists \tilde{i} \in \{2, \ldots, K - 1\}$ such that

$$a_1 \geq a_2 \geq \cdots \geq a_{\tilde{i}} \leq a_{\tilde{i}+1} \leq \cdots \leq a_K \tag{A7}$$

In particular, if $\bar{\theta}_\omega \in \left[\theta_2, \frac{2\theta_1 \theta_2}{\theta_1 + \theta_2}\right]$, then $\tilde{i} = 2$; whereas, if $\bar{\theta}_\omega \in \left[\frac{2\theta_K \theta_{K-1}}{\theta_K + \theta_{K-1}}, \theta_{K-1}\right]$, then $\tilde{i} = K - 1$.

(ii) If $\bar{\theta}_\omega \in \left[\frac{2\theta_1 \theta_2}{\theta_1 + \theta_2}, \theta_1\right)$ then $\tilde{i} = 1$, i.e.

$$a_1 \leq a_2 \leq \cdots \leq a_K. \tag{A8}$$

(iii) If $\bar{\theta}_\omega \in \left(\theta_K, \frac{2\theta_K \theta_{K-1}}{\theta_K + \theta_{K-1}}\right]$, then $\tilde{i} = K$, i.e

$$a_1 \geq a_2 \geq \cdots \geq a_K. \tag{A9}$$

Notice that $\frac{2\theta_K \theta_{K-1}}{\theta_K + \theta_{K-1}} \leq \theta_{K-1}$ and $\frac{2\theta_1 \theta_2}{\theta_1 + \theta_2} \geq \theta_2$.
Moreover,

(iv) if $\bar{a} = a_1$, then (A8) and (A9) are impossible and, therefore, (A7) holds with $a_K > a_1$;

(v) if $\bar{a} > a_1$ then (A9) is impossible;

(vi) if $\bar{a} > a_1 \geq a_2$, then (A8) and (A9) are impossible and, therefore, (A7) holds with $a_K > a_1$.

Proof of Lemma 5. Clearly from (A5), $a_i \geq a_{i+1} \Longleftrightarrow \left(\frac{\bar{\theta}_\omega}{\theta_i} + \frac{\bar{\theta}_\omega}{\theta_{i+1}}\right) \leq 2$. Thus, if $\bar{\theta}_\omega \in [\theta_{K-1}, \theta_2]$, then $\exists \, \tilde{i} \in \{2, \dots, K-1\}$ such that $a_1 \geq a_2 \geq \cdots \geq a_{\tilde{i}} \leq a_{\tilde{i}+1} \leq \cdots \leq a_K$.

Moreover, if $\bar{\theta}_\omega \in (\theta_2, \theta_1)$ and $\left(\frac{\bar{\theta}_\omega}{\theta_1} + \frac{\bar{\theta}_\omega}{\theta_2}\right) \leq 2$ (i.e. $\theta_2 < \bar{\theta}_\omega \leq \frac{2\theta_1\theta_2}{\theta_1+\theta_2}$), then $a_1 \geq a_2 \leq a_3 \leq \dots \leq a_K$, (namely (A7) holds with $\tilde{i} = 2$), while if $\bar{\theta}_\omega \in (\theta_2, \theta_1)$ and $\left(\frac{\bar{\theta}_\omega}{\theta_1} + \frac{\bar{\theta}_\omega}{\theta_2}\right) > 2$ (i.e. $\frac{2\theta_1\theta_2}{\theta_1+\theta_2} < \bar{\theta}_\omega < \theta_1$), then $a_1 < a_2 \leq a_3 \leq \dots \leq a_K$, namely (A7) holds with $\tilde{i} = 1$. Furthermore, if $\bar{\theta}_\omega \in (\theta_K, \theta_{K-1})$ and $\left(\frac{\bar{\theta}_\omega}{\theta_K} + \frac{\bar{\theta}_\omega}{\theta_{K-1}}\right) \leq 2$ (i.e. $\theta_K < \bar{\theta}_\omega \leq \frac{2\theta_K\theta_{K-1}}{\theta_K+\theta_{K-1}}$), then $a_1 \geq a_2 \geq \dots \geq a_{K-1} \geq a_K$, (namely (A7) holds with $\tilde{i} = K$), whereas if if $\bar{\theta}_\omega \in (\theta_K, \theta_{K-1})$ and $\left(\frac{\bar{\theta}_\omega}{\theta_K} + \frac{\bar{\theta}_\omega}{\theta_{K-1}}\right) > 2$ (i.e. $\frac{2\theta_K\theta_{K-1}}{\theta_K+\theta_{K-1}} < \bar{\theta}_\omega < \theta_{K-1}$), then $a_1 \geq a_2 \geq \dots \geq a_{K-1} < a_K$, namely (A7) holds with $\tilde{i} = K-1$. The rest of the proof is straightforward.

The behaviour of the sequence $\{a_i, i = 1, \dots, K\}$ will be used for the proof of Theorem 2.

Proof of Theorem 2

Maximization problem (3) can be address via Lagrange multipliers with $L(\rho, \lambda_1, \dots, \lambda_K) = \phi(\rho) - \sum_{i=1}^{K-1} \lambda_i(\rho_{i+1} - \rho_i) - \lambda_K\left(\sum_{i=1}^{K} \rho_i - 1\right)$. By setting $\partial L(\rho, \lambda_1, \dots, \lambda_K)/\partial \rho_i = 0$ for $i = 1, \dots, K$ we obtain

$$\begin{cases} a_1 + \lambda_1 = \lambda_K \\ a_i - \lambda_{i-1} + \lambda_i = \lambda_K, \quad i = 2, \dots, K-1, \\ a_K - \lambda_{K-1} = \lambda_K \end{cases}$$

namely, by summing all the equations, $\lambda_i = \sum_{j=1}^{i}(\bar{a} - a_j)$ $(i = 1, \dots, K-1)$ and $\lambda_K = \bar{a} = \frac{1}{K}\sum_{i=1}^{K}\left(1 - \frac{\bar{\theta}_\omega}{\theta_i}\right)^2 > 0$.

Case 1 $\lambda_i > 0 \quad \forall i = 1, \dots, K-1$.

In this case $\rho_1 = \dots = \rho_K$, namely the corresponding target is ρ^B, so that

$$\bar{\theta}_\omega = \left(\sum_{i=1}^{K} \frac{1}{\theta_i^2}\right)^{-1} \sum_{i=1}^{K} \frac{1}{\theta_i} \tag{A10}$$

and, from (A5), $a_1 \geq a_2$ since $\left(\frac{1}{\theta_2} - \frac{1}{\theta_1}\right)^2 + \sum_{i=3}^{K} \frac{1}{\theta_i}\left\{\left(\frac{1}{\theta_i} - \frac{1}{\theta_1}\right) + \left(\frac{1}{\theta_i} - \frac{1}{\theta_2}\right)\right\} \geq 0$.

Condition $\lambda_1 > 0$, i.e., $\bar{a} > a_1$, corresponds to $\left(\sum_{i=1}^{K} \frac{1}{\theta_i^2}\right)^{-1} \sum_{i=1}^{K} \frac{1}{\theta_i} > 2\left(\sum_{i=1}^{K} \frac{1}{\theta_i} - \frac{K}{\theta_1}\right)\left(\sum_{i=1}^{K} \frac{1}{\theta_i^2} - \frac{K}{\theta_1^2}\right)^{-1} \Leftrightarrow x > K^{-1}$. Then, from ($vi$) of Lemma 5, the sequence $\{a_i, i = 1, \dots, K\}$ behaves as in (A7), with at least $a_K > a_1$. Therefore $i\bar{a} > \sum_{j=1}^{i} a_j$ for $i = 2, \dots, \tilde{i}$, while for $i > \tilde{i}$, the sequence becomes increasing. Since $K - 1 > (a_1 + \dots + a_{K-1})/a_K$ (we recall that $a_i < a_K$ for $i = 1, \dots, K-1$), then $(K-1)\bar{a} > \sum_{j=1}^{K-1} a_j$ and therefore $\lambda_i > 0$ for $i = \tilde{i} + 1, \dots, K-1$.

Case 2 $\lambda_1 = 0$ and $\lambda_i > 0 \quad \forall i = 2, \dots, K-1$.

In this case $\rho_2 = \dots = \rho_K = \zeta$, so the ensuing optimal target is $(1 - [K-1]\zeta, \zeta, \dots, \zeta)$ (where, clearly, $\zeta \in [0, K^{-1}]$), which is admissible iff $\bar{a} = a_1$ and

$$i\bar{a} > \sum_{j=1}^{i} a_j \qquad i = 2, \dots, K-1. \tag{A11}$$

Now, $\bar{a} = a_1$ corresponds to

$$\bar{\theta}_\omega = \frac{2\left(\sum_{i=1}^{K} \frac{1}{\theta_i} - \frac{K}{\theta_1}\right)}{\sum_{i=1}^{K} \frac{1}{\theta_i^2} - \frac{K}{\theta_1^2}}. \tag{A12}$$

Moreover, $\bar{\theta}_\omega = \dfrac{\frac{1}{\theta_1} + \zeta\left(\sum_{i=1}^{K} \frac{1}{\theta_i} - \frac{K}{\theta_1}\right)}{\frac{1}{\theta_1^2} + \zeta\left(\sum_{i=1}^{K} \frac{1}{\theta_i^2} - \frac{K}{\theta_1^2}\right)}$, so that $\zeta = x$ and the ensuing target ρ^C is admissible provided that $x \leq K^{-1}$. From (iv) in Lemma 5, the sequence of $\{a_i, i = 1, \dots, K\}$ behaves as in (A7), with at least $a_K > a_1$, and we need to check that $\lambda_i > 0$ for every $i = 2, \dots, K-1$.

- If $\theta_1 = \theta_2$, then $a_1 = a_2$ which contradicts (A11) (clearly, the same reasoning holds in the more general case of a cluster of several superior treatments).

- If $\theta_1 > \theta_2 = \dots = \theta_K$, then $a_2 = \dots = a_K$ and, together with $\bar{a} = a_1$, implies that $a_1 = \dots = a_K$ which contradicts (A11).

- If $\theta_1 > \theta_2 \geq \cdots \geq \theta_K$ with $\theta_2 > \theta_K$, from (A12) after tedious algebra it follows that $\left(\frac{\bar{\theta}_\omega}{\theta_1} + \frac{\bar{\theta}_\omega}{\theta_2}\right) < 2$, i.e., $a_1 > a_2$ and, combined with $\bar{a} = a_1$, it ensures that $\lambda_i > 0$ for $i = 2, \ldots, \tilde{i}$. Moreover for $i > \tilde{i}$, the sequence $\{a_i\}$ becomes increasing and, therefore, if $(K-1)\bar{a} > \sum_{i=1}^{K-1} a_i$ then $\lambda_i > 0$, for $i = \tilde{i} + 1, \ldots, K-1$. This condition is trivially satisfied, since $\bar{a} = a_1$ and

$$(K-1)\bar{a} + a_K > \sum_{i=1}^{K-1} a_i + a_K \iff (K-1)\bar{a} + a_K > K\bar{a} \iff a_K > \bar{a} = a_1. \tag{A13}$$

Case 3 $\lambda_i = 0 \quad \forall i = 1, \ldots, K-1$

In such a case, $\bar{a} = a_1 = \cdots = a_{K-1}$, which clearly implies that $a_K = \bar{a}$. As shown in A.1, this implies that $\theta_1 = \cdots = \theta_j > \theta_{j+1} = \cdots = \theta_K$ (i.e., there are two clusters of treatments). Thus, every $\rho^{\tilde{C}} = (\rho_1^{\tilde{C}}, \ldots, \rho_K^{\tilde{C}})^\top$ such that $\sum_{i=1}^j \rho_i^{\tilde{C}} = \frac{\theta_1}{\theta_1 + \theta_K}$ and $\sum_{i=j+1}^K \rho_i^{\tilde{C}} = \frac{\theta_K}{\theta_1 + \theta_K}$, is optimal. Indeed, in such a case $\bar{\theta}_\omega = 2\theta_1 \theta_K / (\theta_1 + \theta_K)$ and $\phi(\rho^{\tilde{C}}) = \bar{a} = (\theta_1 - \theta_K)^2 / (\theta_1 + \theta_K)^2$, that coincides with the maximum of ϕ in the unconstrained optimization. Moreover, adopting $\rho^{\tilde{C}}$, $x = \frac{\theta_K}{(K-j)(\theta_1 + \theta_K)}$ and therefore, $\sum_{i=1}^j \rho_i^{\tilde{C}} = 1 - x(K-j) = 1 - \sum_{i=j+1}^K \rho_i^{\tilde{C}}$. Since the components of $\rho^{\tilde{C}}$ are ordered according to the magnitude of the treatment effects, then $\rho_j^{\tilde{C}} \leq \rho_k^{\tilde{C}}$ (for $k = 1, \ldots, j-1$) and $\rho_{j+1}^{\tilde{C}} \geq \rho_k^{\tilde{C}}$ (for $k = j+1, \ldots, K$). Thus, $\rho_j^{\tilde{C}} \leq \frac{1-x(K-j)}{j}$, $\rho_{j+1}^{\tilde{C}} \geq \frac{x(K-j)}{K-j} = x$ and, clearly, $j^{-1}[1 - x(K-j)] \geq x$, i.e., $x \leq K^{-1}$.

Case 4 $\lambda_1 > 0$ and at least one $\lambda_i = 0$ for $i \in [2; K-1]$.

Under this scenario, $\bar{a} > a_1$ and thus (A9) of Lemma 5 is impossible. Therefore, the sequence $\{a_i, i = 1, \ldots, K\}$ behaves as in (A7) or (A8). Moreover, $\lambda_i = 0 \iff i = \sum_{j=1}^i a_j / \bar{a}$ (with $2 \leq i \leq K-1$). Since $\frac{a_1}{\bar{a}} < 1$, then it exists at least one a_l (with $2 \leq l \leq i$) such that $\frac{a_i}{\bar{a}} > 1$, i.e. $\bar{a} < a_l$, and clearly $\bar{a} < a_l \leq a_i \leq a_{i+1}$. However, this is impossible since

- if $\lambda_{i+1} = 0$, then $(i+1)\bar{a} = \sum_{j=1}^i a_j + a_{i+1} \iff \bar{a} = a_{i+1}$ which contradicts $\bar{a} < a_{i+1}$;

- if $\lambda_{i+1} > 0$, then $(i+1)\bar{a} > \sum_{j=1}^i a_j + a_{i+1} \iff \bar{a} > a_{i+1}$ but this is impossible since $\bar{a} < a_{i+1}$.

Case 5 $\lambda_1 = 0$ and at least one $\lambda_i > 0$ and $\lambda_{i+1} = 0$ with $i \in \{2, \ldots, K-2\}$

Under this setting, $\bar{a} = a_1$ and from Lemma 5 the sequence behaves as in (A7). Since $\lambda_{i+1} = 0$, then

$$i = \frac{a_2}{\bar{a}} + \cdots + \frac{a_{i+1}}{\bar{a}}. \tag{A14}$$

From (A14) and $\lambda_i > 0$ (i.e. $i > \frac{a_1}{\bar{a}} + \frac{a_2}{\bar{a}} + \cdots + \frac{a_i}{\bar{a}}$) it follows that $\bar{a} = a_1 < a_{i+1}$ and, given the behaviour of $\{a_i, i = 1, \ldots, K\}$, also $a_{i+1} \leq a_{i+2}$. This scenario is impossible since

- if $\lambda_{i+2} = 0$, then $(i+2)\bar{a} = a_1 + \cdots + a_{i+1} + a_{i+2}$, which combined with (A14), gives $\bar{a} = a_{i+2}$ contradicting that $\bar{a} < a_{i+1} \leq a_{i+2}$;

- if $\lambda_{i+2} > 0$, then $(i+2)\bar{a} > a_1 + \cdots + a_{i+1} + a_{i+2}$, which combined with (A14), gives $\bar{a} > a_{i+2}$ but this is impossible.

Case 6 $\lambda_1 = \lambda_2 = \cdots = \lambda_j = 0$ and $\lambda_i > 0 \ \forall i = j+1, \ldots, K-1$

Under this setting, the ensuing optimal target is $\rho^{\check{C}} = (\rho_1^{\check{C}}, \ldots, \rho_j^{\check{C}}, v, \ldots, v)^\top$ with $\rho_i^{\check{C}} \geq \rho_{i+1}^{\check{C}} \geq v$ for $i = 1, \ldots, j-1$ and $\sum_{i=1}^j \rho_i^{\check{C}} = 1 - (K-j)v$ (where, clearly, $v \leq K^{-1}$). This target is admissible iff i) $\bar{a} = a_1 = \ldots = a_j$ and ii) $i\bar{a} > \sum_{k=1}^i a_i$ for $i = j+1, \ldots K-1$.

Condition $\lambda_1 = 0$ implies $\bar{a} = a_1$, so that (A12) holds and, from (iv) of Lemma 5, the sequence $\{a_i, i = 1, \ldots, K\}$ behaves as in (A7) with $a_K > a_1$.

- If $\theta_1 > \theta_2$ then by (A5), $a_1 = a_2 \iff \left(\frac{\bar{\theta}_\omega}{\theta_1} + \frac{\bar{\theta}_\omega}{\theta_2}\right) = 2$. Since $\theta_1 > \bar{\theta}_\omega$ it has to be $\theta_2 < \bar{\theta}_\omega$, namely $a_1 = a_2 \leq a_3 \leq \ldots \leq a_K$, which is impossible given that $\bar{a} = a_1$. Similar reasoning applies for all the pairs $\theta_i > \theta_{i+1}$ for $i = 2, \ldots, j-1$.

- If $\theta_1 = \cdots = \theta_j > \theta_{j+1} \geq \ldots \geq \theta_K$, with $j \in \{2, \ldots, K-1\}$, then $a_1 = \cdots = a_j$ and

$$\bar{\theta}_\omega = \frac{[1 - (K-j)v]\theta_1^{-1} + v \sum_{i=j+1}^K \theta_i^{-1}}{[1 - (K-j)v]\theta_1^{-2} + v \sum_{i=j+1}^K \theta_i^{-2}}$$

so that, from (A12), after tedious algebra it follows that $v = x$ (clearly, $x \leq K^{-1}$ in order to be admissible). Finally, we need to verify ii). These conditions are trivially satisfied following the same arguments of Case 2 (see (A13)).

Since the components of $\rho^{\check{C}}$ are ordered according to the magnitude of the treatment effects, then $\rho_j^{\check{C}} \le \rho_k^{\check{C}}$ (for $k = 1, \dots, j-1$) and, clearly, $\rho_j^{\check{C}} \le \frac{1-x(K-j)}{j}$ (recalling that $\sum_{i=1}^{j} \rho_i^{\check{C}} = 1 - (K-j)x$). Moreover, given the above-mentioned ordering, $j^{-1}[1 - x(K-j)] \ge x$, i.e., $x \le K^{-1}$. When $\theta_{j+1} = \dots = \theta_K$, then $x = \frac{\theta_K}{(K-j)(\theta_1 + \theta_K)}$ and this is a special case of Case 3. In addition, note that under $\rho^{\check{C}}$,

$$\phi(\rho^{\check{C}}) = a_1[1 - (K-j)x] + x \sum_{j=i+1}^{K} a_i = \bar{a} - K\bar{a}x + K\bar{a}x = \bar{a}, \tag{A15}$$

A.3 Proof of Theorem 3

Clearly, when $x > K^{-1}$, $\rho^C = \rho^B$ and thus, we will consider only the case $x \le K^{-1}$. Denote by $\bar{\theta}_\omega^B = \sum_{i=1}^{K} \frac{1}{\theta_i} / \sum_{i=1}^{K} \frac{1}{\theta_i^2}$ the expression in (A10). Then, from (A3), $\phi(\rho^B) = \frac{1}{K} \sum_{i=1}^{K} \left(1 - \frac{\bar{\theta}_\omega^B}{\theta_i}\right)^2 = 1 - \bar{\theta}_\omega^B \frac{1}{K} \sum_{i=1}^{K} \frac{1}{\theta_i}$.

- Under statement (i) of Theorem 2, $\rho^C = (\rho_1^C, \dots, \rho_j^C, x, \dots, x)^\top$, and letting $\bar{\theta}_\omega^C$ be (A1) under ρ^C, then from (A15) follows that $\phi(\rho^C) = \frac{1}{K} \sum_{i=1}^{K} \left(1 - \frac{\bar{\theta}_\omega^C}{\theta_i}\right)^2 = 1 - 2\bar{\theta}_\omega^C \frac{1}{K} \sum_{i=1}^{K} \frac{1}{\theta_i} + \left(\bar{\theta}_\omega^C\right)^2 \frac{1}{K} \sum_{i=1}^{K} \frac{1}{\theta_i^2}$.

  Thus, $\phi(\rho^B) \le \phi(\rho^C)$ since $-\bar{\theta}_\omega^B \frac{1}{K} \sum_{i=1}^{K} \frac{1}{\theta_i} \le -2\bar{\theta}_\omega^C \frac{1}{K} \sum_{i=1}^{K} \frac{1}{\theta_i} + \left(\bar{\theta}_\omega^C\right)^2 \frac{1}{K} \sum_{i=1}^{K} \frac{1}{\theta_i^2} \Leftrightarrow (\bar{\theta}_\omega^B - \bar{\theta}_\omega^C)^2 \ge 0$.

- Under statement (ii) of Theorem 2, then $\phi(\rho^C) = \phi(\tilde{\rho}) = \left(\frac{\theta_1 - \theta_K}{\theta_1 + \theta_K}\right)^2$ and, clearly, $E_\phi(\rho^C) = 1$.

Taking now into account the ethical efficiency, denoting by $\bar{\theta} = K^{-1} \sum_{i=1}^{K} \theta_i$, then $E_e(\rho^B) = \frac{\bar{\theta}}{\theta_1}$. Under statement (i) of Theorem 2 we obtain $E_e(\rho^C) = \theta_1^{-1} \left\{\theta_1[1 - (K-j)]x + x \sum_{i=j+1}^{K} \theta_i\right\} = \theta_1^{-1} \left[\theta_1(1 - Kx) + Kx\bar{\theta}\right]$. Thus, $E_e(\rho^C) \ge E_e(\rho^B)$ since $\theta_1(1 - Kx) \ge \bar{\theta}(1 - Kx)$ (recalling that $\theta_1 > \bar{\theta}$). The case under statement (ii) follows easily by similar arguments.