Electronic Journal of Statistics
Vol. 15 (2021) 3217–3248
ISSN: 1935-7524
https://doi.org/10.1214/21-EJS1864

Optimal designs for testing the efficacy of heterogeneous experimental groups*

Alessandro Baldi Antognini,
Rosamarie Frieri,
Marco Novelli and
Maroussa Zagoraiou†

Department of Statistical Sciences, University of Bologna, Via Belle Arti 41, 40126, Bologna, Italy
e-mail: a.baldi@unibo.it; rosamarie.frieri2@unibo.it; marco.novelli4@unibo.it
maroussa.zagoraiou@unibo.it

Abstract: This paper develops a unified framework for deriving optimal designs for hypothesis testing in the presence of several heteroscedastic groups. In particular, the obtained optimal designs are generalized Neyman allocations involving only two experimental groups. In order to account for the ordering among the treatments, particularly relevant in the clinical context for ethical reasons, we provide the optimal design for testing under constraints reflecting their effectiveness. The advantages of the suggested allocations are illustrated both theoretically and through several numerical examples, also compared with other designs proposed in the literature, showing a substantial gain in terms of both power and ethics.

AMS 2000 subject classifications: Primary 62K05, 62L05; secondary 62G20, 60F05.
Keywords and phrases: ANOVA, ethics, power, response-adaptive randomization, Wald test.

Received August 2020.

1. Introduction

This paper addresses the issue of designing experiments for comparing several treatments when the principal inferential aim is testing the homogeneity of the treatment effects. Starting from the classical one-way analysis of variance (ANOVA) of Sir. R.A. Fisher, the problem of comparing the equality of several means has a long history in the statistical literature and covers all the applied fields. Over the past 50's, there has been a growing stream of papers about the design of experiments for treatment comparisons; however, they are almost exclusively focused on estimation precision. In particular, having in mind the linear homoscedastic model set-up, balancing the allocations among treatments is often considered as desirable, since this strategy optimizes the usual alphabetical criteria for the estimation of the treatment effects. However, balance could be highly inefficient in the case of heteroscedasticity, or when inference is

*Rosamarie Frieri was supported by the PhD Program GlaxoSmithKline Biologicals SA.
†Corresponding author.

focused on the treatment contrasts; moreover, it could be strongly inappropriate for clinical trials, since the demand of individual care often induces to skew the allocations to the best performing drugs.

Although the ability of detecting a significant treatment difference is a fundamental issue for statistical inference, in the design of experiment literature very little attention has been devoted to hypothesis testing, also due to the underlined complex mathematical structure. Only recently there has been a growing interest on this topic, in particular in the clinical/pharmaceutical research, also due to the encouragement of Health Authorities [5]. In the context of binary trials, Tymofyeyev et al. [14] were among the first authors to derive the design maximizing the non-centrality parameter (NCP) of the Wald test of homogeneity. In general, this design is a degenerate allocation involving only the best and the worst treatments, with no observations on the intermediate ones: for this reason, a lower bound of each treatment allocation proportion is superimposed and the related constrained optimal design is derived. By applying the same methodology to exponential outcomes, Zhu and Hu [17] derived the corresponding optimal allocations, while Sverdlov et al. [13] extended their results in the presence of censoring. In general, the ensuing designs are discontinuous (non-degenerate) functions of the unknown model parameters (i.e., they are locally optimal) and, by suitable smoothing transformations, they could be implemented in a sequential fashion via response-adaptive randomization procedures, namely sequential rules that change the treatment allocation probabilities to approximate the chosen target (for a review, see [2, 11]).

Recently, Baldi Antognini et al. [3] derived the design maximizing the power of the test of homogeneity for normal homoscedastic data, which is a balanced allocation involving only the best and the worst treatments; moreover, by imposing the ethical constraint that the treatment allocation proportions should reflect the a-priori unknown ordering among their effects, they also derived a non-degenerate optimal target implementable via response-adaptive randomization. Under the same framework, assuming that the treatment ordering is known, Singh and Davidov [12] discussed the optimal designs for restricted and unrestricted statistical inference (which are equivalent for large samples) by adopting a maxi-min approach, in order to overcome local optimality problems.

The aim of the present paper is to present a unified framework for deriving optimal designs for hypothesis testing in the presence of several experimental groups, also encompassing the general ANOVA set-up with heteroscedastic errors. In particular, the optimal designs are generalized Neyman allocations involving only two treatments, not necessarily the best and the worst ones. In order to account for the ordering among treatments (which could be particularly relevant in the clinical context, for ethical reasons), we derive constrained optimal designs, where the allocation proportions are themselves ordered as the treatment efficacies. Since the ordering among the effects is generally a-priori unknown, the ensuing allocations are locally optimal designs that can be approached by response-adaptive randomization procedures after suitable smoothing techniques. Several illustrative examples are provided for normal, binary, Poisson and exponential data (with and without censoring), also amending

some previously obtained results. The properties of these designs are described both theoretically and through numerical examples, showing a substantial gain in terms of power and ethics as well.

The paper is structured as follows. Section 2 deals with optimal designs for hypothesis testing, taking into account both unconstrained and constrained optimization. Section 3 discusses the performance of the proposed allocations both analytically and numerically, also compared with other designs suggested in the literature. Section 4 deals with a general discussion about our results, including also their implementation via response-adaptive randomization methodology, while Section 5 concludes the paper. The mathematical details are available in the Appendix.

2. Main results

2.1. Preliminaries

Suppose we have $K \geq 2$ competing treatments and let $\boldsymbol{\delta}_i = (\delta_{i1}, \ldots, \delta_{iK})^t$ be the indicator managing the allocation of the ith subject, namely $\delta_{ik} = 1$ if he/she is assigned to treatment k ($k = 1, \ldots, K$) and 0 otherwise. Given the assignments, the observations Y_is are assumed to be independent and identically distributed belonging to the exponential family parameterized in such a way that $\theta_k \in \Theta \subseteq \mathbb{R}$ denotes the mean effect of treatment k, while $v_k = v(\theta_k) \in \mathbb{R}^+$ represents the corresponding variance ($k = 1, \ldots, K$) and we set $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)^t$ and $\boldsymbol{v} = (v_1, \ldots, v_K)^t$. Under such a mean-value parameterization, the Maximum Likelihood Estimators (MLEs) of the treatment effects are the sample means: special cases of practical relevance are binary $B(\theta_k)$ ($\theta_k \in (0;1)$, $v(\theta_k) = \theta_k(1-\theta_k)$) and Poisson $P(\theta_k)$ ($\theta_k \in \mathbb{R}^+$, $v(\theta_k) = \theta_k$) trials for dichotomous and count data, respectively, while normal model $N(\theta_k; v_k)$ (with $\theta_k \in \mathbb{R}$ and $v(\theta_k) = v_k$ independent from θ_k) is also encompassed for continuous responses as well as the exponential one $exp(\theta_k)$ ($\theta_k \in \mathbb{R}^+$, $v(\theta_k) = \theta_k^2$) for survival outcomes.

After n allocations, let $\boldsymbol{N}_n = \sum_{i=1}^n \boldsymbol{\delta}_i = (N_{n1}, \ldots, N_{nK})^t$, where $N_{nk} = \sum_{i=1}^n \delta_{ik}$ denotes the number of assignments to treatment k and, clearly, $\boldsymbol{N}_n^t \mathbf{1}_K = n$ (here $\mathbf{1}_K$ is the K-dim vector of ones); while $\boldsymbol{\rho} = n^{-1}\boldsymbol{N}_n$ is the vector of the treatment allocation proportions, where $\rho_k = n^{-1}N_{nk} \geq 0$ for $k = 1, \ldots, K$ and $\boldsymbol{\rho}^t \mathbf{1}_K = 1$ for every n. Let $\hat{\boldsymbol{\theta}}_n = (\hat{\theta}_{n1}, \ldots, \hat{\theta}_{nK})^t$ be the MLE of $\boldsymbol{\theta}$ (i.e., $\hat{\theta}_{nk} = N_{nk}^{-1} \sum_{i=1}^n \delta_{ik} Y_i$), under well-known regularity conditions $\hat{\boldsymbol{\theta}}_n$ is strongly consistent and asymptotically normal with $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} \mathrm{N}(\mathbf{0}_K, \mathbb{M}^{-1})$, where $\mathbf{0}_K$ is the K-dim vector of zeros and $\mathbb{M} = \mathbb{M}(\boldsymbol{\rho}) = \mathrm{diag}\,(\rho_k/v_k)_{k=1,\ldots,K}$ is the Fisher information associated with $\boldsymbol{\theta}$.

In this setting the inferential focus is on the contrasts $\mathbb{A}^t\boldsymbol{\theta}$ where, considering without loss of generality (wlog) the first treatment as the reference one, $\mathbb{A}^t = [\mathbf{1}_{K-1} | -\mathbb{I}_{K-1}]$ and $\mathbb{I}_{K-1}$ is the $(K-1)$-dim identity matrix. The MLE $\mathbb{A}^t\hat{\boldsymbol{\theta}}_n$ is strongly consistent and asymptotically normal with $\sqrt{n}\mathbb{A}^t(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} \mathrm{N}(\mathbf{0}_{K-1}, \mathbb{A}^t\mathbb{M}^{-1}\mathbb{A})$. Let $\hat{v}_{kn}$s be consistent estimators of the treatment variances, then $\hat{\mathbb{M}}_n = \mathrm{diag}\,(\rho_k/\hat{v}_{kn})_{k=1,\ldots,K}$ and $W_n = n\hat{\boldsymbol{\theta}}_n^t \mathbb{A}[\mathbb{A}^t\hat{\mathbb{M}}_n^{-1}\mathbb{A}]^{-1}\mathbb{A}^t\hat{\boldsymbol{\theta}}_n$ is

the Wald statistic usually employed for testing the hypothesis of homogeneity of the effects $\mathrm{H}_0 : \mathbb{A}^t\boldsymbol{\theta} = \mathbf{0}_{K-1}$ vs $\mathrm{H}_1 : \mathbb{A}^t\boldsymbol{\theta} \neq \mathbf{0}_{K-1}$. Under H_0, $W_n \xrightarrow{d} \chi^2_{K-1}$, namely it converges to a (central) χ^2 with $K-1$ degrees of freedom (dof), provided that $\rho_k > 0$ for every $k = 1, \ldots, K$. Thus, the power of the α-level test is $\Pr(W_n > q_{K-1,\alpha})$, where $q_{b,\alpha}$ be the $(1-\alpha)$-percentile of a χ^2_b.

Inspired by the central limit theorem, for large sample W_n can be regarded as a quadratic form of a normal vector; therefore, several authors (see, e.g., [7, 10, 11, 13, 14, 17]) proposed to approximate the power by $\Pr\left(\chi^2_{K-1}(n\phi) > q_{K-1,\alpha}\right)$, where $\chi^2_{K-1}(n\phi)$ is a non-central χ^2 distribution with $K-1$ dof and NCP $n\phi$ with $\phi = \phi(\boldsymbol{\rho}) = \boldsymbol{\theta}^t\mathbb{A}[\mathbb{A}^t\mathbb{M}^{-1}\mathbb{A}]^{-1}\mathbb{A}^t\boldsymbol{\theta}$. As is well-known, for fixed dof the non-central χ^2 distribution is stochastically increasing in the NCP, so the approximate power is an increasing function of $\phi(\boldsymbol{\rho})$.

Remark 2.1. *Albeit for normal data with known variances $W_n \sim \chi^2_{K-1}(n\phi(\boldsymbol{\rho}))$ for any sample size n (by replacing $\hat{\mathbb{M}}_n$ with $\mathbb{M}$ in the Wald statistic), in general the non-central chi-square approximation of the power could present some limitations. Indeed, even if Wald [15] proved that the difference between the cumulative distribution functions of W_n and $\chi^2_{K-1}(n\phi)$ vanishes uniformly as n grows, both W_n and $\chi^2_{K-1}(n\phi)$ diverge, as well as their expectations. Therefore, the quality of this approximation gets worse by moving progressively farther out in the tails. However, this approximation could be effective for moderate-large samples (more appealing in the context of phase-III clinical trials) and for alternatives close to H_0. Indeed, as showed by many authors (see, e.g., [4, 9]), the non-central chi-square distribution is asymptotically correct under the local alternative approach. In this framework, the power is evaluated asymptotically by taking into account a set of contiguous alternatives $\{\mathbb{A}^t\boldsymbol{\theta}_n\}$ converging to $\mathbf{0}_{K-1}$ at the rate $n^{-1/2}$, under which W_n and $\chi^2_{K-1}(\boldsymbol{\theta}_n^t\mathbb{A}[\mathbb{A}^t\mathbb{M}^{-1}\mathbb{A}]^{-1}\mathbb{A}^t\boldsymbol{\theta}_n)$ tend to be asymptotically equivalent (namely, the difference between their cumulative distribution functions vanishes uniformly as n grows).*

Notice that both the likelihood ratio and score tests are asymptotically equivalent to the Wald test and, for normal homoscedastic outcomes, likelihood ratio and Wald's tests coincide. Thus, from now on we take into account the NCP as the inferential criterion to be maximized, since it represents an increasing measure of the approximate power (and of the exact power for normal data).

2.2. Unconstrained optimal design for testing

In this section we derive the optimal design maximizing the NCP, namely the allocation $\tilde{\boldsymbol{\rho}}$, defined on the simplex $\tilde{\rho}_k \geq 0$ for every $k = 1, \ldots, K$ with $\tilde{\boldsymbol{\rho}}^t\mathbf{1}_K = 1$, maximizing $\phi(\cdot)$. As is well-known, for $K = 2$ the optimal design is the so-called Neyman allocation $\tilde{\rho}_1 = (1 + \sqrt{v_2/v_1})^{-1} = 1 - \tilde{\rho}_2$.

Assuming (wlog) that high responses are preferable, the treatment outcomes are ordered on the basis of their effects and, for ease of notation, we assume (wlog) that $\theta_1 \geq \ldots \geq \theta_K$ (i.e., the best treatment will be labelled as the first one, while the Kth treatment as the worst, admitting also clusters with the same

efficacy), with at least one strict inequality. We wish to stress that this is a simple label-coding intended to avoid more complex notation; clearly, the treatment ranking is a-priori unknown, but it can be estimated sequentially, as we will discuss in Section 4. For Bernoulli, Poisson, exponential and normal homoscedastic models, this ordering corresponds to the classical stochastic order, while it does not imply a specific ordering for normal heteroscedastic data. For a given design $\boldsymbol{\rho}$, let $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)^t$ with $\pi_k = \pi_k(\boldsymbol{\rho}) = \rho_k v_k^{-1}/(\sum_{i=1}^K \rho_i v_i^{-1}) \geq 0$ $(k = 1, \ldots, K)$ then $\boldsymbol{\pi}^t \mathbf{1}_K = 1$ and, after straightforward calculation,

$$\phi(\boldsymbol{\rho}) = \left(\sum_{k=1}^K \rho_k v_k^{-1} \right) V_{\boldsymbol{\pi}}(\theta), \tag{2.1}$$

where $V_{\boldsymbol{\pi}}(\theta) = \sum_{k=1}^K (\theta_k - \bar{\theta}_{\boldsymbol{\pi}})^2 \pi_k$ and $\bar{\theta}_{\boldsymbol{\pi}} = \sum_{k=1}^K \theta_k \pi_k$ are the variance and the mean of θ (with K possibly different ordered support points $\theta_1 \geq \ldots \geq \theta_K$), evaluated with respect to (wrt) the pdf $\boldsymbol{\pi}$. Clearly, for homoscedastic treatment groups $v_k = \nu$ $(k = 1, \ldots, K)$ and $\boldsymbol{\pi} = \boldsymbol{\rho}$.

Remark 2.2. *This representation is quite general and covers the case of data following the exponential distribution exp(θ_k) subject to an independent right censoring scheme (which is a common feature of survival trials). Indeed, let $\epsilon_k = \epsilon(\theta_k) : \mathbb{R}^+ \to (0; 1)$ be the probability that a failure/death occurs before censoring in the kth group ($k = 1 \ldots, K$), that are assumed to be constant for every subject in each group, then ϵ is a decreasing function depending on the particular censoring scheme adopted in the trial (one of the most general is described in [16]). In such a case, $\phi(\boldsymbol{\rho})$ in (2.1) should be simply re-parametrized by substituting each treatment variance $v_k = \theta_k^2$ with $\dot{v}_k = \theta_k^2/\epsilon(\theta_k)$ $(k = 1, \ldots, K)$. However, we wish to stress that some models are not encompassed by our framework like, e.g., the normal distribution where the variance v_k is not independent from the mean θ_k.*

The next Lemma shows some general properties of the function $\phi(\boldsymbol{\rho})$.

Lemma 2.1. *Let $\boldsymbol{a} = (a_1, \ldots, a_K)^t$, with $a_k = v_k^{-1} \left(\theta_k - \bar{\theta}_{\boldsymbol{\pi}}\right)^2 \geq 0$ for $k = 1, \ldots, K$. Then, $\phi(\boldsymbol{\rho}) = \boldsymbol{a}^t \boldsymbol{\rho}$ is a concave super-harmonic function with non-negative gradient $\nabla \phi = \boldsymbol{a}$ and Laplacian $\nabla^2 \phi < 0$.*

Proof. See Appendix A.1. □

From now on, we denote by $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_K$ the canonical base of $\mathbb{R}^K$ and we set $\varsigma_{ik} = \left(1 + \sqrt{v_k/v_i}\right)^{-1} = 1 - \varsigma_{ki}$, for any $i, k \in \{1, \ldots, K\}$.

Theorem 2.1. *The unconstrained optimal design $\tilde{\boldsymbol{\rho}}$ maximizing the NCP of Wald's test is such that*

$$\phi(\tilde{\boldsymbol{\rho}}) = \max_{i,k \in \{1,\ldots,K\}} \left(\frac{\theta_i - \theta_k}{\sqrt{v_i} + \sqrt{v_k}} \right)^2. \tag{2.2}$$

Thus, if the pair of treatments $\{\tilde{i}, \tilde{k}\}$ maximizing the right-hand side (RHS) of (2.2) is unique, then

$$\tilde{\boldsymbol{\rho}} = \tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}} = \boldsymbol{e}_{\tilde{i}}\varsigma_{\tilde{i}\tilde{k}} + \boldsymbol{e}_{\tilde{k}}\varsigma_{\tilde{k}\tilde{i}} = \left(0, \ldots, 0, \frac{\sqrt{v_{\tilde{i}}}}{\sqrt{v_{\tilde{i}}} + \sqrt{v_{\tilde{k}}}}, 0, \ldots, 0, \frac{\sqrt{v_{\tilde{k}}}}{\sqrt{v_{\tilde{i}}} + \sqrt{v_{\tilde{k}}}}, 0, \ldots, 0\right)^t,$$

i.e., $\tilde{\boldsymbol{\rho}}$ corresponds to a Neyman allocation involving only this pair of treatments.

Moreover, if the pair $\{\tilde{i}, \tilde{k}\}$ maximizing the RHS of (2.2) is not unique, but there exists one (or more) other pair(s) of indexes $\{\tilde{i}', \tilde{k}'\}$ such that

$$\left(\frac{\theta_{\tilde{i}} - \theta_{\tilde{k}}}{\sqrt{v_{\tilde{i}}} + \sqrt{v_{\tilde{k}}}}\right)^2 = \left(\frac{\theta_{\tilde{i}'} - \theta_{\tilde{k}'}}{\sqrt{v_{\tilde{i}'}} + \sqrt{v_{\tilde{k}'}}}\right)^2 = \max_{i,k \in \{1,\ldots,K\}} \left(\frac{\theta_i - \theta_k}{\sqrt{v_i} + \sqrt{v_k}}\right)^2, \tag{2.3}$$

then $\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}}$ and $\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}'}$ are both optimal as well as every mixture, namely

$$\phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}}) = \phi\left(\omega\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}} + (1-\omega)\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}'}\right) = \phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}'}), \quad \forall \omega \in [0;1].$$

Example 2.1. *Let $K = 3$ be the treatments to be compared, with $\boldsymbol{\theta} = (3, 2, 1)^t$ and $\boldsymbol{v} = (1, 4, 9)^t$. From (2.2), the maximum of ϕ is attained at the pair $\{1, 3\}$, so that the optimal design is $\tilde{\boldsymbol{\rho}}_{13} = (1/4, 0, 3/4)$, with $\phi(\tilde{\boldsymbol{\rho}}_{13}) = 1/4$. Under the same setting, if $v_3 = 25$ (instead of $v_3 = 9$) then $\tilde{\boldsymbol{\rho}}_{12}$ and $\tilde{\boldsymbol{\rho}}_{13}$ are both optimal designs with $\phi(\tilde{\boldsymbol{\rho}}_{12}) = \phi(\tilde{\boldsymbol{\rho}}_{13}) = 1/9$; moreover, every convex combination of $\tilde{\boldsymbol{\rho}}_{12}$ and $\tilde{\boldsymbol{\rho}}_{13}$ is still optimal, i.e., every design $\tilde{\boldsymbol{\rho}} = (\tilde{\rho}_1, 4\tilde{\rho}_1 - 2/3, 5/3 - 5\tilde{\rho}_1)^t$ with $\tilde{\rho}_1 \in [1/3; 1/6]$ is s.t. $\phi(\tilde{\boldsymbol{\rho}}) = 1/9$.*

The presence of clusters of treatments is also accounted for as a special case, as the following Corollary shows.

Corollary 2.1. *If $\{\tilde{i}, \tilde{k}\}$ is the pair of treatments maximizing the RHS of (2.2) and there exists one (or more) other treatment(s) $\tilde{i}'$ such that $\theta_{\tilde{i}} = \theta_{\tilde{i}'}$ and $v_{\tilde{i}} = v_{\tilde{i}'}$ (i.e., there exists a cluster of equal treatments involved in the maximization), then also $\{\tilde{i}', \tilde{k}\}$ maximizes the RHS of (2.2) and*

$$\phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}}) = \phi(\omega\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}} + (1-\omega)\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}}) = \phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}}), \quad \forall \omega \in [0;1],$$

namely every design $\tilde{\boldsymbol{\rho}}$ s.t. $\tilde{\rho}_{\tilde{i}} + \tilde{\rho}_{\tilde{i}'} = \varsigma_{\tilde{i}\tilde{k}} = \varsigma_{\tilde{i}'\tilde{k}}$ and $\tilde{\rho}_{\tilde{k}} = \varsigma_{\tilde{k}\tilde{i}}$ is optimal.

If the treatments are grouped in two clusters, namely $\exists h \in \{1, \ldots, K-1\}$ s.t. $(\theta_1, v_1) = (\theta_i, v_i)$ for $i = 1, \ldots, h$ and $(\theta_K, v_K) = (\theta_i, v_i)$ for $i = h+1, \ldots, K$, with $\theta_1 > \theta_K$, then every design $\tilde{\boldsymbol{\rho}}$ s.t. $\sum_{i=1}^{h} \tilde{\rho}_i = \varsigma_{1K}$ and $\sum_{i=h+1}^{K} \tilde{\rho}_i = \varsigma_{K1}$ is optimal (namely, the Neyman allocation is spanned over the two clusters).

Proof. See Appendix A.3. □

Remark 2.3. *In general, the unconstrained optimal design $\tilde{\boldsymbol{\rho}}$ is a (generalized) Neyman allocation involving just two treatments (or mixture of Neyman targets in some special cases). Hence, adopting $\tilde{\boldsymbol{\rho}}$, the power of Wald's test can be approximated by $\Pr\left(\chi_1^2(n\phi(\tilde{\boldsymbol{\rho}})) > q_{1,\alpha}\right)$. As correctly stated by Singh and Davidov [12], although $\tilde{\boldsymbol{\rho}}$ is a degenerate target, it is still optimal since $\Pr\left(\chi_1^2(n\phi(\tilde{\boldsymbol{\rho}})) > q_{1,\alpha}\right) \geq \Pr\left(\chi_b^2(n\phi(\tilde{\boldsymbol{\rho}})) > q_{b,\alpha}\right)$, for $b \geq 1$ (recalling that, for every fixed dof, χ^2-distributions are stochastically increasing in the NCP).*

As discussed in Remark 2.2, for exponential outcomes with censoring the treatment variance should be re-scaled. Thus, from Theorem 2.1, the optimal design maximizing the NCP is the generalized Neyman allocation $\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}} = \boldsymbol{e}_{\tilde{i}}\varsigma_{\tilde{i}\tilde{k}} + \boldsymbol{e}_{\tilde{k}}\varsigma_{\tilde{k}\tilde{i}}$, where now $\varsigma_{\tilde{i}\tilde{k}} = \left(1 + \theta_{\tilde{k}}\sqrt{\epsilon(\theta_{\tilde{i}})}/\theta_{\tilde{i}}\sqrt{\epsilon(\theta_{\tilde{k}})}\right)^{-1}$, on the pair $\{\tilde{i}, \tilde{k}\}$ s.t.

$$\phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}}) = \max_{i,k\in\{1,\ldots,K\}} \left(\frac{\theta_i - \theta_k}{\frac{\theta_i}{\sqrt{\epsilon(\theta_i)}} + \frac{\theta_k}{\sqrt{\epsilon(\theta_k)}}} \right)^2 . \tag{2.4}$$

The maximization of the RHS in (2.4) depends on the specific form of the adopted censoring through $\epsilon(\cdot)$, and it could involve each pair of treatments, not necessarily $\{1, K\}$ (i.e., the one with the best and the worst treatments). Thus, this result conflicts with the optimal design obtained in [13] where, setting the minimum treatment allocation proportion equal to 0 leads to $\tilde{\rho}_{\{1K\}}$, as the following example shows.

Example 2.2. *As in [13], we take into account the censoring scheme suggested in [16] with duration $D = 96$ and recruitment period $R = 55$, for $K = 3$ treatments with $\boldsymbol{\theta}^t = (150, 5, 1)$ we obtain $\epsilon(\theta_1) = 0.239$, $\epsilon(\theta_2) = 0.948$ and $\epsilon(\theta_3) = 0.990$. Thus, the maximum of the RHS in (2.4) is given by the pair $\{2, 3\}$ and therefore the optimal design is $\tilde{\boldsymbol{\rho}}_{23} = (0, 0.836, 0.164)^t$ with $\phi(\tilde{\boldsymbol{\rho}}_{23}) = 0.424$, instead of $\tilde{\boldsymbol{\rho}}_{13} = (0.997, 0, 0.003)^t$ for which $\phi(\tilde{\boldsymbol{\rho}}_{13}) = 0.234$.*

While the general cases of normal heteroscedastic outcomes and exponential responses with censoring should be analysed by Theorem 2.1 and Corollary 2.1, the next Corollary provides some useful simplifications for the most common models, where the Neyman target $\tilde{\boldsymbol{\rho}}_{1K}$ involving the best and the worst treatments is optimal.

Corollary 2.2. *Let $\theta_1 = \ldots = \theta_h \geq \theta_{h+1} \geq \ldots \geq \theta_{K-s} > \theta_{K-s+1} = \ldots = \theta_K$, where h, s are positive integers with $h + s \leq K$, then*

- *for binary trials, every $\tilde{\boldsymbol{\rho}}$ s.t.*

$$\sum_{i=1}^{h} \tilde{\rho}_i = \frac{\sqrt{\theta_1(1-\theta_1)}}{\sqrt{\theta_1(1-\theta_1)} + \sqrt{\theta_K(1-\theta_K)}} = \varsigma_{1K} = 1 - \sum_{i=K-s+1}^{K} \tilde{\rho}_i$$

  is optimal with

$$\phi(\tilde{\boldsymbol{\rho}}) = \left(\frac{\theta_1 - \theta_K}{\sqrt{\theta_1(1-\theta_1)} + \sqrt{\theta_K(1-\theta_K)}} \right)^2 ;$$

- *under Poisson outcomes, every $\tilde{\boldsymbol{\rho}}$ s.t. $\sum_{i=1}^{h} \tilde{\rho}_i = \sqrt{\theta_1}/(\sqrt{\theta_1} + \sqrt{\theta_K}) = \varsigma_{1K} = 1 - \sum_{i=K-s+1}^{K} \tilde{\rho}_i$ is optimal with $\phi(\tilde{\boldsymbol{\rho}}) = \left\{(\theta_1 - \theta_K)/(\sqrt{\theta_1} + \sqrt{\theta_K})\right\}^2$;*
- *under exponential responses in the absence of censoring, every design $\tilde{\boldsymbol{\rho}}$ s.t. $\sum_{i=1}^{h} \tilde{\rho}_i = \theta_1/(\theta_1 + \theta_K) = \varsigma_{1K} = 1 - \sum_{i=K-s+1}^{K} \tilde{\rho}_i$ is optimal with $\phi(\tilde{\boldsymbol{\rho}}) = \{(\theta_1 - \theta_K)/(\theta_1 + \theta_K)\}^2$;*

- *under* $N(\theta_k; v)$ *for* $k = 1, \ldots, K$, *every* $\tilde{\boldsymbol{\rho}}$ *s.t.* $\sum_{i=1}^{h} \tilde{\rho}_i = 1/2 = \varsigma_{1K} = \sum_{i=K-s+1}^{K} \tilde{\rho}_i$ *is optimal with* $\phi(\tilde{\boldsymbol{\rho}}) = (\theta_1 - \theta_K)^2/4v$.

Moreover, when $h = s = 1$ *(i.e., in the absence of clusters of best and worst treatments) the optimal design* $\tilde{\boldsymbol{\rho}} = \tilde{\boldsymbol{\rho}}_{1K} = \boldsymbol{e}_1 \varsigma_{1K} + \boldsymbol{e}_K (1 - \varsigma_{1K})$ *is unique.*

Proof. See Appendix A.4. □

Remark 2.4. *Theorem 2.1 and Corollary 2.2 complement the results in [14] and [17], by covering every possible scenario of clusters of treatments. Indeed, in both papers* h *and* s *are assumed to be positive integers with* $h + s < K$, *therefore the special case when all treatments are grouped into two clusters is excluded (e.g., for* $K = 3$, $h = 2$ *and* $s = 1$ *or* $h = 1$ *and* $s = 2$*), also showing that every design spanning the Neyman target over the two clusters is optimal (instead of assuming as unique solution* $\tilde{\boldsymbol{\rho}} = h^{-1} \varsigma_{1K} \sum_{i=1}^{h} \boldsymbol{e}_i + s^{-1} \varsigma_{K1} \sum_{i=K-s+1}^{K} \boldsymbol{e}_i$ *as in [13, 14, 17]).*

2.3. Constrained optimal designs for testing

This section deals with the problem of finding the design $\boldsymbol{\rho}^* = (\rho_1^*, \ldots, \rho_K^*)^t$ maximizing the NCP of the Wald test of homogeneity under the (ethical) constraints $\rho_1^* \geq \ldots \geq \rho_K^*$, reflecting the (unknown) treatment ordering $\theta_1 \geq \ldots \geq \theta_K$. Due to the complexity induced by this general framework, we need to introduce the following notation. Let

$$T = \frac{\left[\sum_{k=1}^{K} v_k^{-1}\right] \left[\sum_{k=1}^{K} (\theta_1 - \theta_k)^2 v_k^{-1}\right]}{\left[\sum_{k=1}^{K} (\theta_1 - \theta_k) v_k^{-1}\right]^2} \tag{2.5}$$

and, for $i = 1, \ldots, K-1$,

$$\sigma_i = \frac{\frac{1}{i} \sum_{k=1}^{i} v_k^{-1}}{\frac{1}{K} \sum_{k=1}^{K} v_k^{-1}}, \ \beta_i = \frac{\frac{1}{i} \sum_{k=1}^{i} (\theta_1 - \theta_k) v_k^{-1}}{\frac{1}{K} \sum_{k=1}^{K} (\theta_1 - \theta_k) v_k^{-1}}, \ \gamma_i = \frac{\frac{1}{i} \sum_{k=1}^{i} (\theta_1 - \theta_k)^2 v_k^{-1}}{\frac{1}{K} \sum_{k=1}^{K} (\theta_1 - \theta_k)^2 v_k^{-1}}$$

where, clearly, $\beta_1 = \gamma_1 = 0$.

Lemma 2.2. *For every* $i = 1, \ldots, K-1$, $\sigma_i \geq \beta_i \geq \gamma_i \geq 0$ *and, if* $\gamma_i = 1$, *then* $\beta_i > 1$. *Moreover,* $T > 1$.

Proof. See Appendix A.5. □

Theorem 2.2. *If there exists* $\breve{k} \in \{1, \ldots, K-1\}$ *such that:*

P1a: $\beta_{\breve{k}} \left(2\sigma_{\breve{k}} - \beta_{\breve{k}} - \beta_{\breve{k}} \sigma_{\breve{k}}\right) / \sigma_{\breve{k}}^2 < T(1 - \gamma_{\breve{k}}) \leq 1 + \sigma_{\breve{k}} - 2\beta_{\breve{k}}$,

P1b: $A_{\breve{k}}^2 (1 - \sigma_i) - 2A_{\breve{k}} (1 - \beta_i) + T(1 - \gamma_i) > 0$, *for every* $i \neq \breve{k}$,

where

$$A_{\breve{k}} = \begin{cases} \frac{T(1-\gamma_{\breve{k}})}{1-\beta_{\breve{k}} + \left[(1-\beta_{\breve{k}})^2 - T(1-\sigma_{\breve{k}})(1-\gamma_{\breve{k}})\right]^{1/2}}, & \text{if } \gamma_{\breve{k}} \neq 1 \\ 2(\beta_{\breve{k}} - 1)/(\sigma_{\breve{k}} - 1), & \text{if } \gamma_{\breve{k}} = 1, \end{cases} \tag{2.6}$$

then

$$\boldsymbol{\rho}^* = \boldsymbol{\rho}^*_{\breve{k}} = \left(\frac{1-\tau(K-\breve{k})}{\breve{k}}\right)\sum_{i=1}^{\breve{k}} \boldsymbol{e}_i + \tau \sum_{i=\breve{k}+1}^{K} \boldsymbol{e}_i, \tag{2.7}$$

with

$$\tau = \frac{\sigma_{\breve{k}} A_{\breve{k}} - \beta_{\breve{k}}}{K[1-\beta_{\breve{k}} - A_{\breve{k}}(1-\sigma_{\breve{k}})]} \tag{2.8}$$

and

$$\phi(\boldsymbol{\rho}^*_{\breve{k}}) = \left(\frac{A^2_{\breve{k}} - 2A_{\breve{k}} + T}{TK}\right)\sum_{k=2}^{K}(\theta_1-\theta_k)^2 v_k^{-1}. \tag{2.9}$$

If $\breve{k}$ is not unique but there exists (one or more) $\breve{i}$ satisfying $P1$ (namely $P1a$ and $P1b$, where now $P1b$ should hold for every $i \neq \{\breve{k}, \breve{i}\}$), then $A_{\breve{k}} = A_{\breve{i}}$ and $\boldsymbol{\rho}^*$ can be obtained by any convex combination of the corresponding constrained optimal designs $\boldsymbol{\rho}^*_{\breve{k}}$ and $\boldsymbol{\rho}^*_{\breve{i}}$ in (2.7), so that the NCP in (2.9) still holds.

Whereas, if

P2: $T(1-\gamma_i) > 1 + \sigma_i - 2\beta_i$, for $i = 1, \ldots, K-1$,

the constrained optimal design is balanced, namely $\boldsymbol{\rho}^* = \boldsymbol{\rho}_B = K^{-1}\mathbf{1}_K$, with

$$\phi(\boldsymbol{\rho}_B) = \left(\frac{T-1}{TK}\right)\sum_{k=2}^{K}(\theta_1-\theta_k)^2 v_k^{-1}. \tag{2.10}$$

In all the other scenarios $\boldsymbol{\rho}^{*t} = \left(\boldsymbol{\rho}^{*t}_{[K-\dot{c}]}, 0^t_{\dot{c}}\right)$, where $\dot{c} \in \{1, \ldots, K-2\}$ is the minimum number of inferior treatments that should be omitted in order to satisfy $P1$ or $P2$ and $\boldsymbol{\rho}^*_{[K-\dot{c}]}$ is the previously defined constrained optimal design evaluated by taking into account the remaining $K - \dot{c}$ (superior) treatments.

Proof. See Appendix A.6. □

Remark 2.5. *If $\dot{c} = K-2$, then $P1b$ vanishes, while $P1a$ and $P2$ identify two exhaustive and disjoint sets, namely $T \leq 1 + \sigma_1$ (i.e., $v_1 \geq v_2$) and $T > 1 + \sigma_1$, respectively. In such a case,*

$$\boldsymbol{\rho}^{*t} = \begin{cases} \left(\varsigma_{12}; \varsigma_{21}, \mathbf{0}^t_{K-2}\right), & \text{if } v_1 \geq v_2, \\ \left(2^{-1}\mathbf{1}^t_2, \mathbf{0}^t_{K-2}\right), & \text{if } v_1 < v_2 \end{cases}$$

and therefore the constrained optimal design is the Neyman or a balanced allocation involving just the two superior treatments.

Example 2.3. *Consider now the case of $K = 3$. When $\boldsymbol{\theta} = (23, 22.5, 22)^t$ and $\boldsymbol{v} = (100, 10, 11)^t$, condition $P2$ holds and therefore the constrained optimal design is the balanced one, i.e. $\boldsymbol{\rho}^* = 3^{-1}\mathbf{1}_3$, with $\phi(\boldsymbol{\rho}^*) = 0.0057$. Under the same setting, if $\boldsymbol{v} = (65, 10, 3.1)^t$ then there exists a unique $\breve{k} \in \{1, 2\}$ satisfying $P1a$-$P1b$, namely $\breve{k} = 1$, so that $\boldsymbol{\rho}^* = \boldsymbol{\rho}^*_1 = (0.508, 0.246, 0.246)^t$ with $\phi(\boldsymbol{\rho}^*_1) = 0.0104$, whereas if $\boldsymbol{v} = (80, 10, 3.1)^t$ then $\breve{k} = 2$ and therefore $\boldsymbol{\rho}^* = \boldsymbol{\rho}^*_2 =$*

$(0.361, 0.361, 0.278)^t$ *with* $\phi(\boldsymbol{\rho}_2^*) = 0.0096$. *When* $\boldsymbol{v} = (65.37, 10, 3.1)^t$, *then both* $\breve{k} = 1$ *and* $\breve{k} = 2$ *satisfy* $P1a$-$P1b$ *and* $A_1 = A_2 = 0.965$; *thus, every combination* $\boldsymbol{\rho}^* = \omega\boldsymbol{\rho}_1^* + (1-\omega)\boldsymbol{\rho}_2^*$ – *where* $\omega \in [0;1]$, $\boldsymbol{\rho}_1^* = (0.504, 0.248, 0.248)^t$ *and* $\boldsymbol{\rho}_2^* = (0.360, 0.360, 0.280)^t$ – *is optimal with* $\phi(\boldsymbol{\rho}^*) = \phi(\boldsymbol{\rho}_1^*) = \phi(\boldsymbol{\rho}_2^*) = 0.0103$. *If* $\boldsymbol{v} = (5, 1, 65)^t$ *instead, neither* $P1a$ *nor* $P2$ *are satisfied and* $\dot{c} = 1$; *since* $v_1 > v_2$, *then* $\boldsymbol{\rho}^{*t} = (\varsigma_{12}, \varsigma_{21}, 0) = (0.691, 0.309, 0)$ *with* $\phi(\boldsymbol{\rho}^*) = 0.0239$; *analogously,* $\dot{c} = 1$ *also when* $\boldsymbol{v} = (1, 5, 65)^t$, *but now* $\boldsymbol{\rho}^* = (0.5, 0.5, 0)^t$ *with* $\phi(\boldsymbol{\rho}^*) = 0.0208$, *since* $v_1 \leq v_2$.

Despite the generality of Theorem 2.2, the next Corollary shows that in many practical situations the constrained optimal design $\boldsymbol{\rho}^*$ is a non-degenerate target having a simple functional form.

Corollary 2.3. *Assume that the variance* $v(\cdot)$ *is non-decreasing in* θ *and*

$$\sigma_1(1-\gamma_i) \geq \sigma_i + \gamma_i - 2\beta_i, \qquad \text{for } i = 2, \ldots, K-1. \tag{2.11}$$

Thus, when $\theta_1 > \theta_2$, *the constrained optimal design is*

$$\boldsymbol{\rho}^* = \begin{cases} \boldsymbol{\rho}_1^* & \text{if } T \leq 1 + \sigma_1, \\ \boldsymbol{\rho}_B & \text{if } T > 1 + \sigma_1. \end{cases} \tag{2.12}$$

Whereas, in the presence of a cluster of superior treatments $\theta_1 = \cdots = \theta_j > \theta_{j+1} \geq \cdots \geq \theta_K$ $(j = 2, \ldots, K-1)$, *then* $\boldsymbol{\rho}^* = \boldsymbol{\rho}^B$ *when* $T > 1 + \sigma_1$, *while for* $T \leq 1 + \sigma_1$ *every convex combination of* $\boldsymbol{\rho}_1^*, \ldots, \boldsymbol{\rho}_j^*$ *is optimal.*

In particular, normal homoscedastic, Poisson and exponential models satisfy condition (2.11) and τ *in (2.8) is given by*

$$\tau = \begin{cases} \dfrac{\sum_{k=2}^K (\theta_1-\theta_k)^2}{2\left[\sum_{k=2}^K (\theta_1-\theta_k)\right]^2}, & \text{for } N(\theta_k;\nu); \\[2ex] \dfrac{\theta_1\left[\sum_{k=2}^K\left(\frac{1}{\theta_k}-\frac{1}{\theta_1}\right)\right]^{1/2} - \left[\sum_{k=2}^K(\theta_1-\theta_k)\right]^{1/2}}{\left[\sum_{k=1}^K \frac{\theta_1}{\theta_k} - K\right]\left[\sum_{k=2}^K(\theta_1-\theta_k)\right]^{1/2}}, & \text{for } P(\theta_k); \\[2ex] \dfrac{\frac{1}{\theta_1}\sum_{k=2}^K\left(\frac{1}{\theta_k}-\frac{1}{\theta_1}\right)^2}{\left[\sum_{k=2}^K\left(\frac{1}{\theta_k}-\frac{1}{\theta_1}\right)\right]\left[\sum_{k=2}^K\left(\frac{1}{\theta_k^2}-\frac{1}{\theta_1^2}\right)\right]}, & \text{for } exp(\theta_k). \end{cases}$$

Proof. See Appendix A.7. □

Although the hypothesis of Corollary 2.3 do not hold for binary outcomes, the constrained optimal design has an analogous form, as the following proposition shows.

Proposition 2.1. *Under the binary model, when* $\theta_1 > \theta_2$ *the constrained optimal design is* $\boldsymbol{\rho}^*$ *in (2.12), where*

$$\tau = \frac{\frac{\sum_{k=1}^K \frac{(\theta_1-\theta_k)}{(1-\theta_k)\theta_k}}{R} - 1}{\sum_{k=1}^K \frac{(1-\theta_1)\theta_1}{(1-\theta_k)\theta_k} - K} \text{ and } R = \sqrt{\left(\sum_{k=1}^K \frac{\theta_1-\theta_k}{(1-\theta_1)(1-\theta_k)}\right)\left(\sum_{k=1}^K \frac{\theta_1-\theta_k}{\theta_1\theta_k}\right)}.$$

Whereas, if $\theta_1 = \cdots = \theta_j$ $(j = 2, \ldots, K-1)$*, then* $\boldsymbol{\rho}^* = \boldsymbol{\rho}_B$ *when* $T > 1+\sigma_1$*, while for* $T \leq 1+\sigma_1$ *every convex combination of* $\boldsymbol{\rho}_1^*, \ldots, \boldsymbol{\rho}_j^*$ *is optimal.*

Proof. See Appendix A.8 □

3. Analytical and numerical comparisons

This section is dedicated to the performance assessment of the newly introduced optimal designs. Starting with the normal model, $\boldsymbol{\rho}^*$ and $\tilde{\boldsymbol{\rho}}$ will be compared with the balanced allocation $\boldsymbol{\rho}_B$ and the design $\boldsymbol{\rho}_M = (\boldsymbol{e}_1 + \boldsymbol{e}_K)/2$ proposed by Baldi Antognini et al. [3] and Singh and Davidov [12], which is the optimal design for normal homoscedastic data (i.e., $\tilde{\boldsymbol{\rho}} = \boldsymbol{\rho}_M$) and it is also the target maximizing the minimum power for both restricted and unrestricted likelihood ratio tests under the simple order restriction $\theta_1 \geq \ldots \geq \theta_K$. Moreover, we will also consider the design $\boldsymbol{\rho}_A$ provided in [6], which allocates observations to the experimental groups proportionally to the absolute values of the Abelson and Tukey [1] scores (note that, for $K = 3$, $\boldsymbol{\rho}_M = \boldsymbol{\rho}_A$). For binary and exponential responses, we compare our proposals with $\boldsymbol{\rho}_B$ and the design $\boldsymbol{\rho}_H$ proposed by Tymofyeyev et al. [14] and Zhu and Hu [17], that maximizes the NCP under the constraint of a minimum prefixed threshold of allocations to each treatment.

In what follows, both inferential and ethical criteria will be assessed. For every design $\boldsymbol{\rho}$, besides the approximate power $\mathcal{P}_n(\boldsymbol{\rho}) = \Pr\left(\chi_b^2(n\phi(\boldsymbol{\rho})) > q_{b,0.05}\right)$ (where we set $\alpha = 0.05$ and b is the appropriate dof), an alternative measure of efficiency is simply provided by $\Lambda(\boldsymbol{\rho}) = \phi(\boldsymbol{\rho})/\phi(\tilde{\boldsymbol{\rho}})$, namely the ratio between the NCP induced by $\boldsymbol{\rho}$ over its optimal value (i.e., the one corresponding to the unconstrained optimal design $\tilde{\boldsymbol{\rho}}$). Taking into account $\boldsymbol{\rho}_M$, from (2.1) and (2.2) it follows that

$$\Lambda(\boldsymbol{\rho}_M) = \frac{\frac{(\theta_1-\theta_K)^2}{2(v_1+v_K)}}{\max\limits_{i,k\in\{1,\ldots,K\}}\left(\frac{\theta_i-\theta_k}{\sqrt{v_i}+\sqrt{v_k}}\right)^2}. \tag{3.1}$$

From Theorem 2.1, when $\{1; K\}$ is the pair of treatments maximizing the RHS of (2.2), then $\Lambda(\boldsymbol{\rho}_M) = \left\{2[\varsigma_{1K}^2 + (1-\varsigma_{1K})^2]\right\}^{-1} \geq 1/2$, guaranteeing good performance even in the presence of strong heteroscedasticity between the two extreme treatment groups (namely when $\varsigma_{1K} \to \{0; 1\}$). If $\{\tilde{i}, \tilde{k}\} \neq \{1; K\}$, then $\Lambda(\boldsymbol{\rho}_M)$ tends to vanish; indeed, assuming for example the normal heteroscedastic model with $v_1 = v_K$ and $v_{\tilde{i}} = v_{\tilde{k}}$ (where, clearly, $v_1 > v_{\tilde{i}}$), then $\Lambda(\boldsymbol{\rho}_M) \to 0$ as $v_{\tilde{i}}/v_1 \to 0$ (e.g., by letting $\boldsymbol{\theta} = (15, 14, 13, 10, 9)^t$ and $v = (40, 1, 35, 1, 40)^t$, then $\Lambda(\boldsymbol{\rho}_M) = 0.056$). Adopting $\boldsymbol{\rho}_B$ instead, from (2.1) and (2.10)

$$\Lambda(\boldsymbol{\rho}_B) = \frac{\left(\frac{T-1}{T}\right)\frac{1}{K}\sum_{k=1}^{K}\frac{(\theta_1-\theta_k)^2}{v_k}}{\max\limits_{i,k\in\{1,\ldots,K\}}\left(\frac{\theta_i-\theta_k}{\sqrt{v_i}+\sqrt{v_k}}\right)^2}. \tag{3.2}$$

For normal homoscedastic data, $\Lambda(\boldsymbol{\rho}_B) = 4\,(\theta_1-\theta_K)^{-2}K^{-1}\sum_{i=1}^{K}(\theta_i-\bar{\theta})^2$, where $\bar{\theta} = K^{-1}\sum_{k=1}^{K}\theta_k$. By applying the Von Szokefalvi-Nagy inequality,

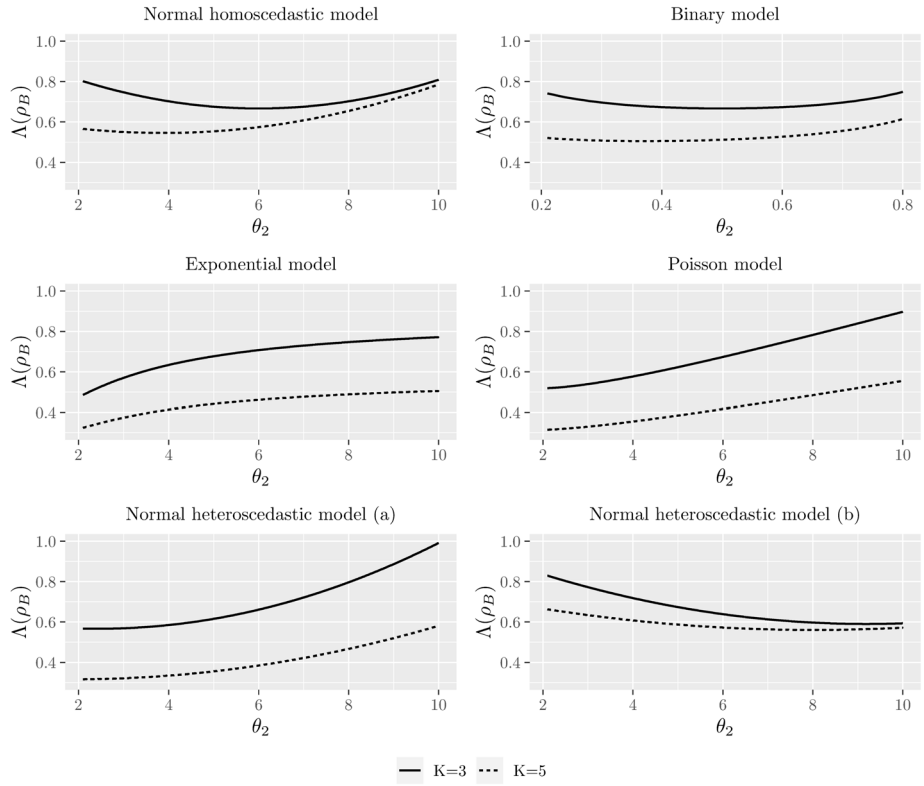

Figure 1: Behaviour of $\Lambda(\boldsymbol{\rho}_B)$ for $K = 3$ and 5 treatments as θ_2 varies. For the normal homoscedastic, exponential and Poisson models $\boldsymbol{\theta}^t = (11, \theta_2, 1)$ and $\boldsymbol{\theta}^t = (11, \theta_2, 2, 1.5, 1)$, while for binary data $\boldsymbol{\theta}^t = (0.9, \theta_2, 0.1)$ and $\boldsymbol{\theta}^t = (0.9, \theta_2, 0.2, 0.15, 0.1)$. For the normal heteroscedastic case, scenario (a) corresponds to $\boldsymbol{v}^t = (12, 9, 2)$ and $\boldsymbol{v}^t = (20, 16, 12, 9, 2)$ and scenario (b) to $\boldsymbol{v}^t = (2, 12, 9)$ and $\boldsymbol{v}^t = (2, 12, 20, 16, 9)$.

$\Lambda(\boldsymbol{\rho}_B) \geq 2K^{-1}$, namely $\boldsymbol{\rho}_B$ tends to exhibit poor performances as the number of treatment groups grows. In this regard, Figure 1 displays the behaviour of $\Lambda(\boldsymbol{\rho}_B)$ for different models with $K = 3$ and 5 treatments as θ_2 varies. In particular, for the normal, exponential and Poisson models we consider $\boldsymbol{\theta}^t = (11, \theta_2, 1)$ for $K = 3$ and $\boldsymbol{\theta}^t = (11, \theta_2, 2, 1.5, 1)$ for $K = 5$, where in both scenarios θ_2 varies between 2 and 10; whereas for the binary model we set $\boldsymbol{\theta}^t = (0.9, \theta_2, 0.1)$ for $K = 3$ and $\boldsymbol{\theta}^t = (0.9, \theta_2, 0.2, 0.15, 0.1)$ for $K = 5$, with θ_2 varying between 0.2 and 0.8. As regards normal responses we take into account both the homoscedastic (with $\nu = 1$) and the heteroscedastic cases. For all the considered models, the performance of $\boldsymbol{\rho}_B$ considerably deteriorates for $K = 5$, especially for binary, exponential, Poisson and normal – scenario (a) – models, whose efficiencies are always lower than 61%. For normal homoscedastic and binary outcomes, $\Lambda(\boldsymbol{\rho}_B)$

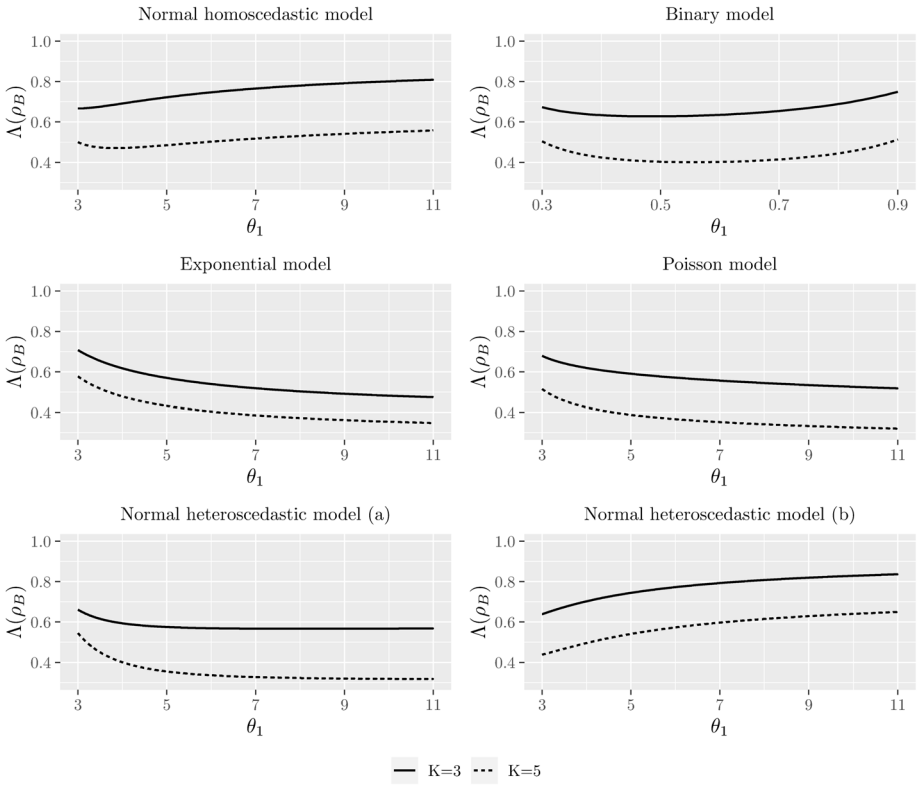

Figure 2: Behaviour of $\Lambda(\boldsymbol{\rho}_B)$ for $K = 3$ and 5 treatments as θ_1 varies. For the normal homoscedastic, exponential and Poisson models $\boldsymbol{\theta}^t = (\theta_1, 2, 1)$ and $\boldsymbol{\theta}^t = (\theta_1, 2.5, 2, 1.5, 1)$, while for binary data $\boldsymbol{\theta}^t = (\theta_1, 0.2, 0.1)$ and $\boldsymbol{\theta}^t = (\theta_1, 0.25, 0.2, 0.15, 0.1)$. For the normal heteroscedastic case, scenario (a) corresponds to $\boldsymbol{v}^t = (12, 9, 2)$ and $\boldsymbol{v}^t = (20, 16, 12, 9, 2)$ and scenario (b) to $\boldsymbol{v}^t = (2, 12, 9)$ and $\boldsymbol{v}^t = (2, 12, 20, 16, 9)$.

tends to grow for values of θ_2 close to θ_3, with a maximum efficiency around 80%. This behaviour is even more pronounced in the normal heteroscedastic scenario (a) for $K = 3$. Whereas, heteroscedastic scenario (b) exhibits decreasing $\Lambda(\boldsymbol{\rho}_B)$ in θ_2 for both $K = 3$ and 5. As regards Poisson and exponential models the performances improve as $\theta_2 \to \theta_1$, whereas for θ_2 close to θ_3 the efficiency becomes even lower than 40%. The behaviour of $\Lambda(\boldsymbol{\rho}_B)$ for increasing θ_1 is reported in Figure 2, where $\boldsymbol{\theta}^t = (\theta_1, 2, 1)$ and $\boldsymbol{\theta}^t = (\theta_1, 2.5, 2, 1.5, 1)$ for normal, exponential and Poisson models with $\theta_1 \in [3, 11]$; while for binary outcomes $\boldsymbol{\theta}^t = (\theta_1, 0.2, 0.1)$ and $\boldsymbol{\theta}^t = (\theta_1, 0.25, 0.2, 0.15, 0.1)$ with $\theta_1 \in [0.3, 0.9]$. Similarly to Figure 1, when $K = 5$ poor performances in terms of $\Lambda(\boldsymbol{\rho}_B)$ are observed. For the binary, exponential and Poisson models, the efficiency is below 75% for $K = 3$ and it is below 51% for $K = 5$. As regards the normal model,

graphical evidence points out that the best performances in terms of efficiency are achieved in the homoscedastic and heteroscedastic scenario (b), while in scenario (a) $\Lambda(\boldsymbol{\rho}_B) < 57\%$ and $\Lambda(\boldsymbol{\rho}_B) < 32\%$ for $K = 3$ and 5, respectively.

As regards ethics, in the multi-treatment context several ethical measures could be adopted, some of them are only model-specific. In our general set-up, as a measure of ethics we take into account the total expected outcome $\mathcal{E}_n(\boldsymbol{\rho}) = n\boldsymbol{\theta}^t\boldsymbol{\rho}$; the corresponding ethical efficiency $(\boldsymbol{\theta}^t\boldsymbol{\rho} - \theta_K)/(\theta_1 - \theta_K) \in [0; 1]$ will be provided within brackets. Since $\mathcal{E}_{2n}(\boldsymbol{\rho}) = 2\mathcal{E}_n(\boldsymbol{\rho})$, in the following tables the ethical criterion will be provided only for $n = 100$.

As previously showed, the unconstrained optimal design $\tilde{\boldsymbol{\rho}}$ maximizes the NCP and the approximate power too; therefore, it does not exists a target with better performances in terms of both ethics and inference wrt $\tilde{\boldsymbol{\rho}}$, simultaneously. Whereas, in some circumstances, $\tilde{\boldsymbol{\rho}}$ dominates $\boldsymbol{\rho}_B$ and $\boldsymbol{\rho}_M$, as discussed in the following proposition.

Proposition 3.1. *Let $\{\tilde{i}, \tilde{k}\}$ be the pair of treatments maximizing the RHS of (2.2) with $\tilde{i} < \tilde{k}$. If $\varsigma_{\tilde{i}\tilde{k}} \geq (\bar{\theta} - \theta_{\tilde{k}})/(\theta_{\tilde{i}} - \theta_{\tilde{k}})$ then $\tilde{\boldsymbol{\rho}}$ dominates $\boldsymbol{\rho}_B$, namely $\mathcal{E}_n(\tilde{\boldsymbol{\rho}}) \geq \mathcal{E}_n(\boldsymbol{\rho}_B)$ and $\mathcal{P}_n(\tilde{\boldsymbol{\rho}}) \geq \mathcal{P}_n(\boldsymbol{\rho}_B)$, simultaneously, for every n. Analogously, if $\varsigma_{\tilde{i}\tilde{k}} \geq (\theta_1 + \theta_K - 2\theta_{\tilde{k}})/[2(\theta_{\tilde{i}} - \theta_{\tilde{k}})]$, then $\tilde{\boldsymbol{\rho}}$ dominates $\boldsymbol{\rho}_M$. Moreover, if the variance $v(\cdot)$ of a given statistical model is non-decreasing and $\theta_{\tilde{i}} + \theta_{\tilde{k}} \geq \max\{2\bar{\theta}; \theta_1 + \theta_K\}$, then $\tilde{\boldsymbol{\rho}}$ dominates both $\boldsymbol{\rho}_M$ and $\boldsymbol{\rho}_B$.*

Proof. The proof follows easily from (2.1) after some algebra, by observing that $\mathcal{E}_n(\boldsymbol{\rho}_B) = n\bar{\theta}$, $\mathcal{E}_n(\boldsymbol{\rho}_M) = n(\theta_1 + \theta_K)/2$, while $\mathcal{E}_n(\tilde{\boldsymbol{\rho}}) = n\left[\theta_{\tilde{i}}\varsigma_{\tilde{i}\tilde{k}} + \theta_{\tilde{k}}(1 - \varsigma_{\tilde{i}\tilde{k}})\right]$. □

Starting from the case of normal responses, Tables 1 and 2 summarize the performances in terms of $\mathcal{P}_n$ with $n = 50$ and 100 (within brackets the corresponding efficiency evaluated wrt $\tilde{\boldsymbol{\rho}}$) and $\mathcal{E}_n$ (with $n = 100$) of the considered allocations for $K = 3$, 4 and 5 treatments, as $\boldsymbol{\theta}$ and $\boldsymbol{v}$ vary. Whereas, Tables 3 and 4 show the results in the case of binary and exponential outcomes, respectively, where the minimum proportion of subjects assigned to each treatment group for $\boldsymbol{\rho}_H$ is set to 0.2 ($K = 3, 4$) and 0.15 ($K = 5$).

Let us first consider the results for normal responses: $\tilde{\boldsymbol{\rho}}$ exhibits the highest approximate power, with an ethical efficiency varying between 42.9% and 72.2%, while $\boldsymbol{\rho}^*$ substantially shows the highest ethical efficiency (between 51.8% and 70.2%), also guaranteeing valid inferential performances (for $n = 100$, its efficiency is always greater than 65.6%). Excluding the homoscedastic scenario, the approximate power of $\boldsymbol{\rho}_M$ is strictly related to the treatment variances: for unordered variances $\boldsymbol{\rho}_M$ exhibits poor performances (showing in some cases an extremely low efficiency, equal to 39.3% for $K = 4$ and $n = 50$), while when the variances are ordered as the treatment effects, then $\mathcal{P}_n$ tends to increase. Moreover, the ethical efficiency of $\boldsymbol{\rho}_M$ is always equal to 0.5 and in several scenarios $\boldsymbol{\rho}_M$ is dominated by $\boldsymbol{\rho}^*$, especially when the number of the treatment groups increases. The balanced design $\boldsymbol{\rho}_B$ is always dominated by $\boldsymbol{\rho}^*$ and it often shows the worst performances in terms of precision and ethics in all the considered scenarios. The approximate power provided by $\tilde{\boldsymbol{\rho}}$ with $n = 50$ observations tends to be quite similar to the one of $\boldsymbol{\rho}_B$ with $n = 100$ subjects.

TABLE 1
$\mathcal{P}_n$ and $\mathcal{E}_n$ (efficiencies within brackets) for normal responses with $K = 3$ and 4 treatments, as $\boldsymbol{\theta}$ and $\boldsymbol{v}$ vary.

$\boldsymbol{\theta} = (1.5, 1.1, 1)^t$	$\boldsymbol{\rho}$	$\mathcal{P}_{50}(\boldsymbol{\rho})$	$\mathcal{P}_{100}(\boldsymbol{\rho})$	$\mathcal{E}_{100}(\boldsymbol{\rho})$
	$\boldsymbol{\rho}^* = (.494, .253, .253)^t$	.283 (.667)	.519 (.736)	127 (.544)
$\boldsymbol{v} = (1, 1, 1)^t$	$\tilde{\boldsymbol{\rho}} = \boldsymbol{\rho}_M = \boldsymbol{\rho}_A$	.424 (1)	.705 (1)	125 (.500)
	$\boldsymbol{\rho}_B$	.257 (.606)	.475 (.674)	120 (.400)
	$\boldsymbol{\rho}^* = (.5, .5, 0)^t$	.211 (.977)	.372 (.976)	130 (.600)
$\boldsymbol{v} = (1, 2, 6)^t$	$\tilde{\boldsymbol{\rho}} = (.414, .586, 0)^t$	.216 (1)	.381 (1)	127 (.531)
	$\boldsymbol{\rho}_M = \boldsymbol{\rho}_A$	.157 (.727)	.267 (.701)	125 (.500)
	$\boldsymbol{\rho}_B$	.153 (.708)	.269 (.706)	120 (.400)
	$\boldsymbol{\rho}^* = (.668, .166, .166)^t$	.121 (.688)	.200 (.656)	135 (.702)
$\boldsymbol{v} = (6, 2, 1)^t$	$\tilde{\boldsymbol{\rho}} = (.71, 0, .29)^t$	.176 (1)	.305 (1)	136 (.710)
	$\boldsymbol{\rho}_M = \boldsymbol{\rho}_A$	.157 (.892)	.267 (.875)	125 (.500)
	$\boldsymbol{\rho}_B$	.098 (.557)	.151 (.495)	120 (.400)
	$\boldsymbol{\rho}^* = \tilde{\boldsymbol{\rho}} = (.586, .414, 0)^t$	.216 (1)	.381 (1)	133 (.669)
$\boldsymbol{v} = (2, 1, 6)^t$	$\boldsymbol{\rho}_M = \boldsymbol{\rho}_A$	.143 (.662)	.240 (.630)	125 (.500)
	$\boldsymbol{\rho}_B$	.135 (.625)	.230 (.604)	120 (.400)
$\boldsymbol{\theta} = (2, 1.8, 1.1, 1)^t$	$\boldsymbol{\rho}$	$\mathcal{P}_{50}(\boldsymbol{\rho})$	$\mathcal{P}_{100}(\boldsymbol{\rho})$	$\mathcal{E}_{100}(\boldsymbol{\rho})$
	$\boldsymbol{\rho}^* = (.37, .21, .21, .21)^t$	.747 (.793)	.971 (.972)	156 (.560)
$\boldsymbol{v} = (1, 1, 1, 1)^t$	$\tilde{\boldsymbol{\rho}} = \boldsymbol{\rho}_M$	.942 (1)	.999 (1)	150 (.500)
	$\boldsymbol{\rho}_B$	.729 (.774)	.965 (.966)	148 (.475)
	$\boldsymbol{\rho}_A = (.433, .067, .067, .433)^t$	.829 (.880)	.989 (.999)	149 (.493)
	$\boldsymbol{\rho}^* = (1/3, 1/3, 1/3, 0)^t$	.467 (.622)	.778 (.810)	163 (.633)
$\boldsymbol{v} = (1, 1.5, 2, 7)^t$	$\tilde{\boldsymbol{\rho}} = (.414, 0, .586, 0)^t$	.751 (1)	.961 (1)	147 (.473)
	$\boldsymbol{\rho}_M$	.424 (.565)	.705 (.734)	150 (.500)
	$\boldsymbol{\rho}_B$	.386 (.514)	.692 (.720)	148 (.475)
	$\boldsymbol{\rho}_A = (.433, .067, .067, .433)^t$	.329 (.438)	.610 (.635)	149 (.493)
	$\boldsymbol{\rho}^* = (.309, .309, .191, .191)^t$	.364 (.561)	.662 (.726)	158 (.575)
$\boldsymbol{v} = (7, 2, 1.5, 1)^t$	$\tilde{\boldsymbol{\rho}} = (0, .586, 0, .414)^t$	.649 (1)	.912 (1)	147 (.469)
	$\boldsymbol{\rho}_M$	.424 (.653)	.705 (.773)	150 (.500)
	$\boldsymbol{\rho}_B$	.347 (.535)	.637 (.698)	148 (.475)
	$\boldsymbol{\rho}_A = (.433, .067, .067, .433)^t$	.311 (.479)	.580 (.636)	149 (.493)
	$\boldsymbol{\rho}^* = (.275, .275, .225, .225)^t$	.340 (.472)	.627 (.661)	152 (.518)
$\boldsymbol{v} = (12, 1.5, 9, 1)^t$	$\tilde{\boldsymbol{\rho}} = (0, .55, 0, .45)^t$	.720 (1)	.949 (1)	144 (.440)
	$\boldsymbol{\rho}_M$	.284 (.393)	.501 (.528)	150 (.500)
	$\boldsymbol{\rho}_B$	.337 (.468)	.622 (.655)	148 (.475)
	$\boldsymbol{\rho}_A = (.433, .067, .067, .433)^t$	.252 (.350)	.477 (.503)	149 (.493)

Even if $\boldsymbol{\rho}_A = \boldsymbol{\rho}_M$ when $K = 3$ regardless of the values of the model parameters, in the experimental scenarios of Table 1 with $K = 4$, $\boldsymbol{\rho}_A$ presents even lower approximate power than $\boldsymbol{\rho}_M$, with efficiencies between 35% and 48% in the case of heteroscedasticity. Excluding the homoscedastic case, for $K = 4$ the approximate power provided by $\boldsymbol{\rho}_A$ is always lower than the one by $\boldsymbol{\rho}^*$. For what concerns ethics, the performances of $\boldsymbol{\rho}_A$ are very similar to those of $\boldsymbol{\rho}_M$ and $\boldsymbol{\rho}_B$.

As regards binary trials in Table 3, the constrained optimal design $\boldsymbol{\rho}^*$ and

TABLE 2
$\mathcal{P}_n$ and $\mathcal{E}_n$ (efficiencies within brackets) for normal responses with $K = 5$ treatments and $\boldsymbol{\theta} = (3, 2.7, 2, 1.2, 1)^t$ as $\boldsymbol{v}$ varies.

$\boldsymbol{v}$	$\boldsymbol{\rho}$	$\mathcal{P}_{50}(\boldsymbol{\rho})$	$\mathcal{P}_{100}(\boldsymbol{\rho})$	$\mathcal{E}_{100}(\boldsymbol{\rho})$
$(1,1,1,1,1)^t$	$\boldsymbol{\rho}^* = (.36, .16, .16, .16, .16)^t$	.999 (.999)	1 (1)	218 (.592)
	$\tilde{\boldsymbol{\rho}} = \boldsymbol{\rho}_M$	1 (1)	1 (1)	200 (.500)
	$\boldsymbol{\rho}_B$	.998 (.998)	1 (1)	198 (.490)
	$\boldsymbol{\rho}_A = (.408, .092, 0, .092, .408)^t$	1 (1)	1 (1)	199 (.495)
$(1, 1.5, 2, 3, 15)^t$	$\boldsymbol{\rho}^* = (.277, .241, .241, .241, 0)^t$	.843 (.846)	.992 (.992)	225 (.626)
	$\tilde{\boldsymbol{\rho}} = (.366, 0, 0, .634, 0)^t$	.997 (1)	1 (1)	186 (.429)
	$\boldsymbol{\rho}_M$	.705 (.707)	.942 (.942)	200 (.500)
	$\boldsymbol{\rho}_B$	.765 (.767)	.978 (.978)	198 (.490)
	$\boldsymbol{\rho}_A = (.408, .092, 0, .092, .408)^t$	.718 (.720)	.961 (.961)	199 (.495)
$(12, 3, 2, 1.5, 1)^t$	$\boldsymbol{\rho}^* = (.287, .287, .142, .142, .142)^t$	.794 (.800)	.985 (.985)	223 (.616)
	$\tilde{\boldsymbol{\rho}} = (0, .634, 0, 0, .366)^t$	.993 (1)	1 (1)	208 (.539)
	$\boldsymbol{\rho}_M$	.792 (.798)	.975 (.975)	200 (.500)
	$\boldsymbol{\rho}_B$	.762 (.767)	.978 (.978)	198 (.490)
	$\boldsymbol{\rho}_A = (.408, .092, 0, .092, .408)^t$	.747 (.725)	.971 (.971)	199 (.495)
$(5, 3, 10, 1, 15)^t$	$\boldsymbol{\rho}^* = (.4, .2, .2, .2, 0)^t$	.836 (.857)	.991 (.991)	238 (.690)
	$\tilde{\boldsymbol{\rho}} = (.691, 0, 0, .309, 0)^t$	.976 (1)	1 (1)	244 (.722)
	$\boldsymbol{\rho}_M$	.609 (.624)	.885 (.885)	200 (.500)
	$\boldsymbol{\rho}_B$	.696 (.713)	.957 (.957)	198 (.490)
	$\boldsymbol{\rho}_A = (.408, .092, 0, .092, .408)^t$	.718 (.736)	.962 (.962)	199 (.495)

$\boldsymbol{\rho}_H$ tend to perform quite similarly in terms of inference, with an efficiency always higher than 68%, whereas $\boldsymbol{\rho}_B$ shows the lowest approximate power with a maximum loss up to 44%. Taking into account the ethical criterion, that in this case corresponds to the total expected successes, $\boldsymbol{\rho}^*$ and $\tilde{\boldsymbol{\rho}}$ guarantee the highest ethical efficiency (with only one exception, where $\mathcal{E}_n(\boldsymbol{\rho}_B)$ is slightly bigger wrt $\mathcal{E}_n(\tilde{\boldsymbol{\rho}})$), with an ethical gain up to 8 and 12 successes wrt $\boldsymbol{\rho}_H$ and $\boldsymbol{\rho}_B$, respectively. Similar considerations still hold for exponential responses reported in Table 4, where the minimum inferential efficiency becomes 73% for $\boldsymbol{\rho}^*$, 67% for $\boldsymbol{\rho}_H$ and 58% for the balanced allocation. The ethical gain (i.e., the additional total expected survival time in this context) induced by $\boldsymbol{\rho}^*$ wrt $\boldsymbol{\rho}_H$ ranges from 36 to 90 (corresponding to a gain in efficiency between $5\% - 18\%$) and it is even more evident for $\boldsymbol{\rho}_B$ (the additional expected survival is up to 250, with an ethical gain up to 33% in terms of efficiency). The unconstrained optimal design $\tilde{\boldsymbol{\rho}}$ exhibits the highest approximate power and, at the same time, the greatest ethical gain, therefore it dominates all the other designs. Similarly to the case of normal responses, the approximate power induced by $\tilde{\boldsymbol{\rho}}$ with $n = 50$ observations tends to be quite similar to those of $\boldsymbol{\rho}_B$ and $\boldsymbol{\rho}_H$ with $n = 100$ subjects. In general, for all the considered models, both $\tilde{\boldsymbol{\rho}}$ and $\boldsymbol{\rho}^*$ present high values of ethical efficiency. Moreover, for normal homoscedastic, binary and exponential responses, it is evident that $\boldsymbol{\rho}_B$ is dominated by $\boldsymbol{\rho}^*$ and also by $\tilde{\boldsymbol{\rho}}$ (with the exception of the second last scenario of Table 3). As further comparisons (omitted here for brevity) showed, the results for Poisson outcomes are substantially the same of those of the exponential model. The case of exponential responses

TABLE 3
$\mathcal{P}_n$ and $\mathcal{E}_n$ (efficiencies within brackets) for binary responses with $K = 3$, 4 and 5 treatments, as $\boldsymbol{\theta}$ varies.

$\boldsymbol{\theta}$	$\boldsymbol{\rho}$	$\mathcal{P}_{50}(\boldsymbol{\rho})$	$\mathcal{P}_{100}(\boldsymbol{\rho})$	$\mathcal{E}_{100}(\boldsymbol{\rho})$
$(.4, .1, .05)^t$	$\boldsymbol{\rho}^* = (.658, .171, .171)^t$	.827 (.882)	.987 (.988)	29 (.682)
	$\tilde{\boldsymbol{\rho}} = (.692, 0, .308)^t$	.938 (1)	.999 (1)	29 (.692)
	$\boldsymbol{\rho}_H = (.593, .2, .207)^t$	.821 (.875)	.986 (.987)	27 (.624)
	$\boldsymbol{\rho}_B$	.663 (.707)	.932 (.933)	18 (.381)
$(.6, .4, .25)^t$	$\boldsymbol{\rho}^* = (.480, .260, .260)^t$	.516 (.675)	.827 (.855)	46 (.591)
	$\tilde{\boldsymbol{\rho}} = (.531, 0, .469)^t$	.765 (1)	.967 (1)	44 (.531)
	$\boldsymbol{\rho}_H = (.432, .2, .368)^t$	.566 (.74)	.869 (.899)	43 (.518)
	$\boldsymbol{\rho}_B$	.485 (.634)	.796 (.823)	42 (.476)
$(.4, .3, .1, .05)^t$	$\boldsymbol{\rho}^* = (.562, .146, .146, .146)^t$	.725 (.773)	.964 (.965)	29 (.688)
	$\tilde{\boldsymbol{\rho}} = (.692, 0, 0, .308)^t$	.938 (1)	.999 (1)	29 (.692)
	$\boldsymbol{\rho}_H = (.4, .2, .2, .2)^t$	.693 (.739)	.952 (.953)	25 (.571)
	$\boldsymbol{\rho}_B$	.611 (.651)	.910 (.911)	21 (.464)
$(.5, .2, .15, .1)^t$	$\boldsymbol{\rho}^* = (.583, .139, .139, .139)^t$	.729 (.774)	.965 (.966)	35 (.635)
	$\tilde{\boldsymbol{\rho}} = (.625, 0, 0, .375)^t$	.942 (1)	.999 (1)	35 (.625)
	$\boldsymbol{\rho}_H = (.4, .2, .2, .2)^t$	.670 (.711)	.942 (.943)	29 (.475)
	$\boldsymbol{\rho}_B$	.525 (.557)	.846 (.847)	24 (.344)
$(.8, .7, .6, .5, .1)^t$	$\boldsymbol{\rho}^* = (.316, .171, .171, .171, .171)^t$	.992 (.992)	1 (1)	58 (.683)
	$\tilde{\boldsymbol{\rho}} = (.571, 0, 0, 0, .429)^t$	1 (1)	1 (1)	50 (.571)
	$\boldsymbol{\rho}_H = (.246, .15, .15, .15, .304)^t$	.998 (.998)	1 (1)	50 (.567)
	$\boldsymbol{\rho}_B$	.990 (.990)	1 (1)	54 (.629)
$(.55, .4, .3, .1, .05)^t$	$\boldsymbol{\rho}^* = (.544, .114, .114, .114, .114)^t$	.924 (.925)	.999 (.999)	40 (.692)
	$\tilde{\boldsymbol{\rho}} = (.695, 0, 0, 0, .305)^t$	.999 (1)	1 (1)	40 (.695)
	$\boldsymbol{\rho}_H = (.378, .15, .15, .15, .172)^t$	.909 (.910)	.998 (.998)	34 (.573)
	$\boldsymbol{\rho}_B$	.817 (.818)	.989 (.989)	28 (.460)

with censoring tends to be similar to that of Normal heteroscedastic data with variances ordered as the treatment effects, and it is strongly affected by the chosen censoring scheme.

Finally, inspired by [12], it is interesting to compare the above-considered designs when the values of the model parameters induce unfavorable configurations. From (2.1), due to the heteroscedasticity, the NCP tends to vanish as the variances of the treatment group grow, regardless of the chosen design; so, for any fixed maximum difference between the treatment effects, there is no unique least favorable configuration, but different scenarios that tend to degenerate as the signal to noise ratio vanishes. In particular, Table 5 shows the performances of the considered targets for normal and exponential responses, where the maximum distance $\theta_1 - \theta_K$ has been set equal to 1. The upper part of the table summarizes the results for normal responses as the variances vary. In the first scenario, $\tilde{\boldsymbol{\rho}}$ shows the highest values of $\mathcal{P}_n$ and $\mathcal{E}_n$, with a gain in terms of approximate power up to 10% wrt $\boldsymbol{\rho}_M = \boldsymbol{\rho}_A$ and up to 9% wrt $\boldsymbol{\rho}^* = \boldsymbol{\rho}_B$. In the second and the third scenarios, the variances are at least ten times the value of the treatment effects; this has the effect of vanishing the differences among the

TABLE 4
$\mathcal{P}_n$ and $\mathcal{E}_n$ (efficiencies within brackets) for exponential data with $K = 3$, 4 and 5 treatments, as $\boldsymbol{\theta}$ varies.

$\boldsymbol{\theta}$	$\boldsymbol{\rho}$	$\mathcal{P}_{50}(\boldsymbol{\rho})$	$\mathcal{P}_{100}(\boldsymbol{\rho})$	$\mathcal{E}_{100}(\boldsymbol{\rho})$
$(4,2,1)^t$	$\boldsymbol{\rho}^* = (.722, .139, .139)^t$	.950 (.961)	.999 (.999)	331 (.769)
	$\tilde{\boldsymbol{\rho}} = (.8, 0, .2)^t$	.989 (1)	1 (1)	340 (.800)
	$\boldsymbol{\rho}_H = (.6, .2, .2)^t$	.941 (.951)	.999 (.999)	300 (.667)
	$\boldsymbol{\rho}_B$	.856 (.866)	.992 (.992)	233 (.444)
$(10,7,3)^t$	$\boldsymbol{\rho}^* = (.634, .183, .183)^t$	.902 (.932)	.997 (.997)	817 (.739)
	$\tilde{\boldsymbol{\rho}} = (.769, 0, .231)^t$	.968 (1)	1 (1)	838 (.769)
	$\boldsymbol{\rho}_H = (.574, .2, .226)^t$	.902 (.932)	.997 (.997)	782 (.688)
	$\boldsymbol{\rho}_B$	.849 (.877)	.991 (.991)	667 (.524)
$(11,9,5,3)^t$	$\boldsymbol{\rho}^* = (.625, .125, .125, .125)^t$	.875 (.892)	.995 (.995)	899(.749)
	$\tilde{\boldsymbol{\rho}} = (.786, 0, 0, .214)^t$	.981 (1)	1 (1)	929 (.786)
	$\boldsymbol{\rho}_H = (.4, .2, .2, .2)^t$	.843 (.859)	.992 (.992)	780 (.600)
	$\boldsymbol{\rho}_B$	.789 (.804)	.982 (.982)	700 (.500)
$(14,10,7,5)^t$	$\boldsymbol{\rho}^* = (.619, .127, .127, .127)^t$	.672 (.732)	.943 (.946)	1145 (.717)
	$\tilde{\boldsymbol{\rho}} = (.737, 0, 0, .263)^t$	.918 (1)	.997 (1)	1163 (.737)
	$\boldsymbol{\rho}_H = (.4, .2, .2, .2)^t$	.619 (.674)	.915 (.918)	1000 (.556)
	$\boldsymbol{\rho}_B$	.534 (.582)	.854 (.857)	900 (.444)
$(7,5,4,3,2)^t$	$\boldsymbol{\rho}^* = (.624, .094, .094, .094, .094)^t$	.759 (.778)	.977 (.977)	569 (.737)
	$\tilde{\boldsymbol{\rho}} = (.778, 0, 0, 0, .222)^t$	.975 (1)	1 (1)	589 (.778)
	$\boldsymbol{\rho}_H = (.378, .15, .15, .15, .172)^t$	.713 (.731)	.963 (.963)	479 (.558)
	$\boldsymbol{\rho}_B$	.603 (.618)	.911 (.911)	420 (.440)
$(14,13,10,5,4)^t$	$\boldsymbol{\rho}^* = (.58, .105, .105, .105, .105)^t$	.825 (.846)	.990 (.990)	1148 (747)
	$\tilde{\boldsymbol{\rho}} = (.778, 0, 0, 0, .222)^t$	.975 (1)	1 (1)	1178 (.778)
	$\boldsymbol{\rho}_H = (.4, .15, .15, .15, .15)^t$	.804 (.825)	.987 (.987)	1040 (.640)
	$\boldsymbol{\rho}_B$	.741 (.760)	.972 (.972)	920 (.520)

treatment means, leading to a drastically reduced approximate power. In such cases, the performances of $\tilde{\boldsymbol{\rho}}$ and $\boldsymbol{\rho}_M$ tend to coincide (also due to the fact that the treatment variances are quite similar), while $\boldsymbol{\rho}^*$ behaves similarly to $\boldsymbol{\rho}_B$. The bottom part of Table 5 deals with the case of exponential responses. In the first scenario, $\tilde{\boldsymbol{\rho}}$ provides a considerable gain in terms of approximate power wrt all the competitors; however, since the difference $\theta_1 - \theta_K$ is held constant, $\mathcal{P}_n$ greatly decreases as the values of $\boldsymbol{\theta}$ increase and all the considered targets tend to perform quite similarly (in particular for $n = 50$).

4. Implementation via response-adaptive randomization and discussion

The unconstrained optimal design $\tilde{\boldsymbol{\rho}}$ in Theorem 2.1 and the constrained one $\boldsymbol{\rho}^*$ in Theorem 2.2 depend on the unknown model parameters and therefore they are a-priori unknown (i.e., locally optimal). The dependence on the model parameters acts in terms of both i) the a-priori unknown treatment ordering and ii) the functional form of the optimal design itself, which is often a degenerate

TABLE 5
$\mathcal{P}_n$ and $\mathcal{E}_n$ (efficiencies within brackets) for normal and exponential responses with $K = 3$ treatments.

$\boldsymbol{\theta} = (2, 1.5, 1)^t$	$\boldsymbol{\rho}$	$\mathcal{P}_{50}(\boldsymbol{\rho})$	$\mathcal{P}_{100}(\boldsymbol{\rho})$	$\mathcal{E}_{100}(\boldsymbol{\rho})$
$\boldsymbol{v} = (1, 4, 25)^t$	$\boldsymbol{\rho}^* = \boldsymbol{\rho}_B$	.165 (.757)	.294 (.764)	150 (.500)
	$\tilde{\boldsymbol{\rho}} = (.333, .667, 0)^t$	.218 (1)	.384 (1)	167 (.667)
	$\boldsymbol{\rho}_M = \boldsymbol{\rho}_A$	.165 (.757)	.283 (.737)	150 (.500)
$\boldsymbol{v} = (20, 15, 10)^t$	$\boldsymbol{\rho}^* = (.5, .25, .25)^t$	.1 (.656)	.155 (.601)	163 (.625)
	$\tilde{\boldsymbol{\rho}} = (.586, 0, .414)^t$	.152 (1)	.258 (1)	159 (.586)
	$\boldsymbol{\rho}_M = \boldsymbol{\rho}_A$	.149 (.980)	.252 (.977)	150 (.500)
	$\boldsymbol{\rho}_B$	.096 (.629)	.146 (.567)	150 (.500)
$\boldsymbol{v} = (25, 22, 20)^t$	$\boldsymbol{\rho}^* = (.462, .269, .269)^t$	.081 (.696)	.114 (.617)	160 (.597)
	$\tilde{\boldsymbol{\rho}} = (.528, 0, .472)^t$	.116 (1)	.184 (1)	153 (.528)
	$\boldsymbol{\rho}_M = \boldsymbol{\rho}_A$	.116 (.998)	.183 (.997)	150 (.500)
	$\boldsymbol{\rho}_B$	.079 (.682)	.110 (.598)	150 (.500)
Exponential responses	$\boldsymbol{\rho}$	$\mathcal{P}_{50}(\boldsymbol{\rho})$	$\mathcal{P}_{100}(\boldsymbol{\rho})$	$\mathcal{E}_{100}(\boldsymbol{\rho})$
$\boldsymbol{\theta} = (2, 1.5, 1)^t$	$\boldsymbol{\rho}^* = (.558, .221, .221)^t$	.445 (.670)	.753 (.823)	167 (.669)
	$\tilde{\boldsymbol{\rho}} = (.667, .0, .333)^t$	.654 (1)	.915 (1)	167 (.667)
	$\boldsymbol{\rho}_H = (.504, .0.2, .296)^t$	.466 (.712)	.777 (.849)	160 (.604)
	$\boldsymbol{\rho}_B$	.399 (.610)	.697 (.761)	150 (.500)
$\boldsymbol{\theta} = (5, 4.5, 4)^t$	$\boldsymbol{\rho}^* = (.48, .26, .26)^t$	.085 (.688)	.122 (.614)	461 (.61)
	$\tilde{\boldsymbol{\rho}} = (.556, .0, .444)^t$	.123 (1)	.199 (1)	456 (.556)
	$\boldsymbol{\rho}_H = (.433, .0.2, .367)^t$	.089 (.724)	.132 (.663)	453 (.533)
	$\boldsymbol{\rho}_B$	.082 (.669)	.117 (.59)	450 (.500)
$\boldsymbol{\theta} = (20, 19.5, 19)^t$	$\boldsymbol{\rho}^* = (.453, .274, .274)^t$	.052 (.962)	.054 (.929)	1959 (.589)
	$\tilde{\boldsymbol{\rho}} = (.513, .0, .487)^t$	.054 (1)	.058 (1)	1951 (.513)
	$\boldsymbol{\rho}_H = (.408, .0.2, .392)^t$	.052(.963)	.054 (.937)	1951 (.508)
	$\boldsymbol{\rho}_B$	.052 (.960)	.053 (.926)	1950 (.500)

allocation with no assignments to some treatment groups.

Unlike other alternative approaches suggested in the literature, which are intended to mediate the design criterion onto the entire parameter space to obtain good overall performances (such as, e.g., Bayesian or maxi-min approaches), in this paper we consider response-adaptive randomization as the natural solution to this local optimality problem. Under this framework, the exact optimal designs are sequentially estimated step-by-step: on the basis of earlier responses and past assignments, the unknown parameters are estimated along with the treatment ordering (which could change as the trial progresses) and thus, the next assignment is randomly forced to progressively approach the optimal target (for instance, by applying the Doubly Adaptive Biased Coin Design of [8]).

Even if response-adaptive randomization methodology seems a natural choice in order to implement the optimal proposed allocations, we wish to stress that the optimal designs $\tilde{\boldsymbol{\rho}}$ and $\boldsymbol{\rho}^*$ cannot be targeted directly, due to the fact that i) their functional forms are locally discontinuous around the subset of Θ^K under which the treatment ordering changes (namely, where θ_i tends to coincide

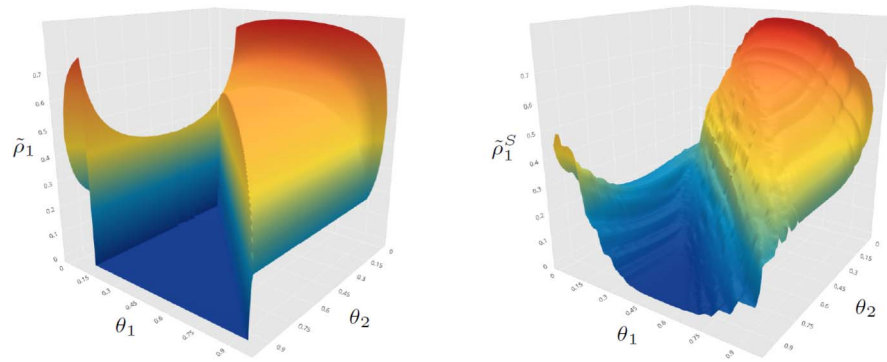

Figure 3: Comparison of $\tilde{\rho}_1(\theta)$ and $\tilde{\rho}_1^S(\theta)$ as $\theta = (\theta_1, \theta_2, 0.15)^t$ vary.

with one or more θ_ks) and ii) in several scenarios these optimal designs lay on the boundary. To overcome these drawbacks, that prevent the applicability of standard response-adaptive randomization methodology, a smoothing transformation (e.g., via a Gaussian kernel) can be applied to obtain a continuous and non-degenerate version of these targets (see, e.g., [14]). In particular, for the mono-parametric exponential family we take into account the convolution of $\tilde{\boldsymbol{\rho}} = \tilde{\boldsymbol{\rho}}(\boldsymbol{\theta}) = (\tilde{\rho}_1(\boldsymbol{\theta}), \ldots, \tilde{\rho}_K(\boldsymbol{\theta}))^t$ (or, analogously, $\boldsymbol{\rho}^*$), with a K-dim Gaussian kernel

$$G(\boldsymbol{\theta}) = (2\pi\sigma^2)^{-K/2} \exp\left(-\frac{\theta_1^2 + \cdots + \theta_K^2}{2\sigma^2}\right),$$

($\sigma^2 > 0$ controls the degree of smoothing), namely we define the smoothed version $\tilde{\boldsymbol{\rho}}^S(\boldsymbol{\theta}) = (\tilde{\rho}_1^S(\boldsymbol{\theta}), \ldots, \tilde{\rho}_K^S(\boldsymbol{\theta}))^t$ of $\tilde{\boldsymbol{\rho}}$, by letting

$$\tilde{\rho}_k^S(\boldsymbol{\theta}) = (\tilde{\rho}_k * G)(\boldsymbol{\theta}) = \int_{\Theta^K} \tilde{\rho}_k(\boldsymbol{x}) G(\boldsymbol{\theta} - \boldsymbol{x}) d\boldsymbol{x}, \quad k = 1, \ldots, K \tag{4.1}$$

(which could be naturally extended to the case of heteroscedastic normal model where $\tilde{\boldsymbol{\rho}} = \tilde{\boldsymbol{\rho}}(\boldsymbol{\theta}; \boldsymbol{v}) : \Theta^K \times \mathbb{R}^{+K} \to [0; 1]$).

This smoothing transformation essentially impacts on the points of discontinuity of $\tilde{\boldsymbol{\rho}}$ and on its boundary, so that the smoothed optimal design $\tilde{\boldsymbol{\rho}}^S$ obey the classical regularity conditions of continuity, non-degeneracy and differentiability (see, e.g., [7, 2]), that allows for the standard asymptotic inference for response-adaptive randomization procedures (therefore, all the asymptotics in Section 2.1 are still valid). For instance, taking into account binary trials with $K = 3$ treatments, Figure 3 illustrates the behaviour of the first component of $\tilde{\boldsymbol{\rho}}$ and the one of its smoothed version $\tilde{\boldsymbol{\rho}}^S$, with $\boldsymbol{\theta} = (\theta_1, \theta_2, 0.15)^t$ as θ_1 and θ_2 vary in $[0; 1]$ (where now the treatment ordering is free to change on the basis of the values of θ_1 and θ_2).

In order to show how the performances of $\tilde{\boldsymbol{\rho}}$ and $\tilde{\boldsymbol{\rho}}^S$ (as well as those of $\boldsymbol{\rho}^*$ and $\boldsymbol{\rho}^{*S}$) are quite similar, Table 6 presents the approximate power $\mathcal{P}_n$ and total expected outcomes $\mathcal{E}_n$ under the same scenarios of Table 3 with $K = 3$ treatments.

TABLE 6
$\mathcal{P}_n$ and $\mathcal{E}_n$ (efficiencies within brackets) for binary responses with $K = 3$ treatments, as $\boldsymbol{\theta}$ varies.

$\boldsymbol{\theta}$	$\boldsymbol{\rho}$	$\mathcal{P}_{50}(\boldsymbol{\rho})$	$\mathcal{P}_{100}(\boldsymbol{\rho})$	$\mathcal{E}_{100}(\boldsymbol{\rho})$
$(.4, .1, .05)^t$	$\boldsymbol{\rho}^* = (.658, .171, .171)^t$	.827 (.882)	.987 (.988)	29 (.682)
	$\boldsymbol{\rho}^{*S} = (.660, .170, .170)^t$	.827 (.882)	.987 (.988)	29 (.684)
	$\tilde{\boldsymbol{\rho}} = (.692, 0, .308)^t$	.938 (1)	.999 (1)	29 (.692)
	$\tilde{\boldsymbol{\rho}}^S = (.742, .031, .227)^t$	.872 (.930)	.994 (.995)	31 (.747)
$(.6, .4, .25)^t$	$\boldsymbol{\rho}^* = (.480, .260, .260)^t$	.516 (.675)	.827 (.855)	46 (.589)
	$\boldsymbol{\rho}^{*S} = (.476, .262, .262)^t$	.516 (.675)	.827 (.855)	46 (.591)
	$\tilde{\boldsymbol{\rho}} = (.531, 0, .469)^t$	.765 (1)	.967 (1)	44 (.531)
	$\tilde{\boldsymbol{\rho}}^S = (.549, .041, .409)^t$	.645 (.843)	.923 (.954)	45 (.567)

5. Conclusions

This paper discusses optimal designs for hypothesis testing in the presence of heterogeneous experimental groups, also encompassing the general one-way ANOVA with heteroscedastic errors. In particular, we derive the allocation maximizing the NCP of the classical Wald test of homogeneity about the treatment contrasts; this optimal design is a generalized Neyman allocation involving only two treatments, not necessarily the best and the worst ones. Moreover, to account for the ordering among treatments, we derive the optimal design maximizing the NCP of the homogeneity test, subject to an ethical constraint reflecting the efficacy of the competing treatments. Due to the dependence on the unknown model parameters, these allocations are locally optimal and therefore a-priori unknown. Moreover, these designs are degenerate allocations with possible local discontinuities. To avoid these drawbacks, a smoothing transformation via Gaussian kernel implemented via response-adaptive randomization has been proposed. The suggested convolution avoids i) degeneracies by assigning a non-null mass on each treatment and ii) potential discontinuities. Thus, these smoothed optimal designs could be approached via standard response-adaptive randomization procedures that, by estimating at each step the unknown parameters as well as the treatment ordering, change sequentially the probabilities of treatment assignments in order to converge to the desired target. To provide a flexible and easy to use tool for implementing the proposed optimal designs and their smoothed versions, a fully documented R code is available in the GitHub repository https://bit.ly/38ER2CD.

As showed in Section 3, the balanced allocation is strongly inappropriate due to heteroscedasticity. Moreover, any deviation from the assumption of homoscedasticity could also affect the inferential performances of $\boldsymbol{\rho}_M$ and $\boldsymbol{\rho}_A$, which tend to exhibit a very low approximate power, especially in the case of unordered variances. In such scenarios, $\tilde{\boldsymbol{\rho}}^S$ provides a remarkable gain in terms of inferential efficiency wrt all the other designs, guaranteeing also high ethical standards. Nevertheless, whenever the ethical dimension plays a crucial role, $\boldsymbol{\rho}^{*S}$ could be preferable since, although it places more emphasis on the ethical aspects, it still guarantees good inferential performances, also wrt $\boldsymbol{\rho}_H$.

Acknowledgments

We are very grateful to the referees for their helpful suggestions and comments, which led to a substantially improved version of the paper.

Appendix A: Proofs

A.1. Proof of Lemma 2.1

From (2.1), $\phi(\boldsymbol{\rho}) = \left(\sum_{k=1}^{K} \rho_k v_k^{-1}\right) V_{\boldsymbol{\pi}}(\theta) = \boldsymbol{a}^t \boldsymbol{\rho}$, where $V_{\boldsymbol{\pi}}(\theta) = \sum_{k=1}^{K}(\theta_k - \bar{\theta}_{\boldsymbol{\pi}})^2 \pi_k$, $\bar{\theta}_{\boldsymbol{\pi}} = \sum_{k=1}^{K} \theta_k \pi_k$ and $\pi_k = \pi_k(\boldsymbol{\rho}) = \rho_k v_k^{-1}/(\sum_{i=1}^{K} \rho_i v_i^{-1})$. Since for every $i = 1, \ldots, K$,

$$\frac{\partial V_{\boldsymbol{\pi}}(\theta)}{\partial \rho_i} = \frac{\partial \sum_{k=1}^{K} \theta_k^2 \pi_k}{\partial \rho_i} - \frac{\partial \bar{\theta}_{\boldsymbol{\pi}}^2}{\partial \rho_i},$$

where

$$\frac{\partial \sum_{k=1}^{K} \theta_k^2 \pi_k}{\partial \rho_i} = \left(v_i \sum_{k=1}^{K} \frac{\rho_k}{v_k}\right)^{-1} \left(\theta_i^2 - \sum_{k=1}^{K} \theta_k^2 \pi_k\right)$$

and

$$\frac{\partial \bar{\theta}_{\boldsymbol{\pi}}^2}{\partial \rho_i} = \left(v_i \sum_{k=1}^{K} \frac{\rho_k}{v_k}\right)^{-1} 2(\theta_i \bar{\theta}_{\boldsymbol{\pi}} - \bar{\theta}_{\boldsymbol{\pi}}^2),$$

then

$$\frac{\partial V_{\boldsymbol{\pi}}(\theta)}{\partial \rho_i} = \left(v_i \sum_{k=1}^{K} \frac{\rho_k}{v_k}\right)^{-1} [(\theta_i - \bar{\theta}_{\boldsymbol{\pi}})^2 - V_{\boldsymbol{\pi}}(\theta)]$$

and therefore

$$\frac{\partial \phi(\boldsymbol{\rho})}{\partial \rho_i} = \frac{V_{\boldsymbol{\pi}}(\theta)}{v_i} - \frac{V_{\boldsymbol{\pi}}(\theta) - (\theta_i - \bar{\theta}_{\boldsymbol{\pi}})^2}{v_i} = a_i.$$

Furthermore,

$$\frac{\partial^2 \phi(\boldsymbol{\rho})}{\partial \rho_i \partial \rho_j} = -2 \left(\sum_{k=1}^{K} \frac{\rho_k}{v_k}\right)^{-1} \frac{(\theta_i - \bar{\theta}_{\boldsymbol{\pi}})}{v_i} \frac{(\theta_j - \bar{\theta}_{\boldsymbol{\pi}})}{v_j},$$

so the Hessian matrix

$$H(\phi) = -2 \left(\sum_{k=1}^{K} \frac{\rho_k}{v_k}\right)^{-1} \begin{bmatrix} \frac{(\theta_1 - \bar{\theta}_{\boldsymbol{\pi}})}{v_1} \\ \vdots \\ \frac{(\theta_K - \bar{\theta}_{\boldsymbol{\pi}})}{v_K} \end{bmatrix} \cdot \begin{bmatrix} \frac{(\theta_1 - \bar{\theta}_{\boldsymbol{\pi}})}{v_1} & \ldots & \frac{(\theta_K - \bar{\theta}_{\boldsymbol{\pi}})}{v_K} \end{bmatrix}$$

is negative semi-definite, since it has $K-1$ null eigenvalues and one eigenvalue (that coincides with the Laplacian)

$$\mathrm{tr} H(\phi) = \nabla^2 \phi = -2 \left(\sum_{k=1}^{K} \frac{\rho_k}{v_k}\right)^{-1} \sum_{k=1}^{K} \frac{a_k}{v_k} < 0,$$

provided that $\mathbb{A}^t \boldsymbol{\theta} \neq \mathbf{0}_{K-1}$.

A.2. Proof of Theorem 2.1

Firstly, we will show that the unconstrained optimal design $\tilde{\boldsymbol{\rho}}$ maximizing the NCP of Wald test is of the form of (2.2). Let $\boldsymbol{\rho}^{(l)}$ denote a generic target with l components different from 0, then $\boldsymbol{\rho}^{(1)}$ corresponds to one of the standard basis $\boldsymbol{e}_k$ $(k = 1, \ldots, K)$ and clearly $\phi(\boldsymbol{\rho}^{(1)}) = 0$, attaining the minimum of the NCP. Taking now into account $\boldsymbol{\rho}^{(2)} = \sum_{k\in\{i,j\}} \boldsymbol{e}_k \rho_k$, where $\rho_j = 1 - \rho_i \in (0; 1)$, from (2.1) we have

$$\phi\left(\boldsymbol{\rho}^{(2)}\right) = \left(\sum_{k\in\{i,j\}} \frac{\rho_k}{v_k}\right)^{-1} (\theta_i - \theta_j)^2 \left\{\frac{\rho_i}{v_i}\frac{1-\rho_i}{v_j}\right\},$$

and

$$\phi\left(\boldsymbol{\rho}^{(2)}\right) \leq \max_{i,j\in\{1,2,\ldots,K\}} (\theta_i - \theta_j)^2 \left\{\max_{\rho_i\in[0,1]} \left(\sum_{k\in\{i,j\}} \frac{\rho_k}{v_k}\right)^{-1} \frac{\rho_i}{v_i}\frac{1-\rho_i}{v_j}\right\}. \quad \text{(A.1)}$$

First, observe that

$$\max_{\rho_i\in[0,1]} \frac{\frac{\rho_i(1-\rho_i)}{v_i v_j}}{\sum_{k\in\{i,j\}} \frac{\rho_k}{v_k}} = \max_{\rho_i\in[0,1]} \left\{\frac{v_i}{\rho_i} + \frac{v_j}{1-\rho_i}\right\}^{-1} = \min_{\rho_i\in[0,1]} \frac{v_i}{\rho_i} + \frac{v_j}{1-\rho_i},$$

which is clearly minimised by the Neyman allocation $\tilde{\rho}_i = \varsigma_{ij}$. Thus, from (A.1),

$$\phi(\tilde{\boldsymbol{\rho}}) = \max_{i,j\in\{1,2,\ldots,K\}} (\theta_i - \theta_j)^2 \left\{\frac{\varsigma_{ij}(1-\varsigma_{ij})}{\varsigma_{ij}(v_j - v_i) + v_i}\right\} = \max_{i,j\in\{1,2,\ldots,K\}} \left(\frac{\theta_j - \theta_i}{\sqrt{v_i} + \sqrt{v_j}}\right)^2,$$

so that if the pair of treatments $\{\tilde{i}, \tilde{k}\}$ maximizing the RHS of the previous equation is unique, then $\tilde{\boldsymbol{\rho}} = \tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}} = \boldsymbol{e}_{\tilde{i}}\varsigma_{\tilde{i}\tilde{k}} + \boldsymbol{e}_{\tilde{k}}\varsigma_{\tilde{k}\tilde{i}}$. When the pairs $\{\tilde{i}, \tilde{k}\}$ and $\{\tilde{i}', \tilde{k}'\}$ satisfy (2.3), then both $\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}}$ and $\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}'}$ are optimal designs, namely

$$\phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}}) = \phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}'}) = \max_{i,k\in\{1,\ldots,K\}} \left(\frac{\theta_i - \theta_k}{\sqrt{v_i} + \sqrt{v_k}}\right)^2. \quad \text{(A.2)}$$

Moreover, every mixture of $\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}}$ and $\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}'}$ is still optimal, since for every $\omega \in [0; 1]$

$$\phi\left(\omega\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}} + (1-\omega)\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}'}\right) \geq \omega\phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}}) + (1-\omega)\phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}'}) = \phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}}) = \phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}'}),$$

due to the concavity of $\phi(\cdot)$ (see Lemma 2.1), but from (A.2), $\phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}}) = \phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}'}) \geq \phi\left(\omega\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}} + (1-\omega)\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}'}\right)$. Thus, $\phi\left(\omega\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}} + (1-\omega)\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}'}\right) = \phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}}) = \phi(\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}'})$.

In order to deal with other scenarios, namely $\phi(\boldsymbol{\rho}^{(l)})$ with $l > 2$, assume wlog $i < j < t$. For the case $\boldsymbol{\rho}^{(3)} = \sum_{k\in\{i,j,t\}} \boldsymbol{e}_k \rho_k$, with $\rho_t = 1 - \rho_i - \rho_j$, note that $\nabla\phi(\boldsymbol{\rho}^{(3)})$ can be expressed as

$$\frac{\partial\phi(\boldsymbol{\rho}^{(3)})}{\partial\rho_k} = \frac{(\theta_k - \bar{\theta}_{\boldsymbol{\pi}})^2}{v_k} - \frac{(\theta_t - \bar{\theta}_{\boldsymbol{\pi}})^2}{v_t} = a_k - a_t, \quad k = i, j.$$

If $\exists\, k$ s.t. $a_k \neq a_t$, then $a_k > a_t$ implies that $\phi(\boldsymbol{\rho}^{(3)})$ is increasing in ρ_k, namely the NCP is maximized for $\tilde{\rho}_t = 0$, and conversely, $a_k < a_t$ implies that $\phi(\boldsymbol{\rho}^{(3)})$ is decreasing in ρ_k that is, it is maximized for $\tilde{\rho}_k = 0$, so that in both cases $\phi(\boldsymbol{\rho}^{(3)}) \leq \phi(\tilde{\boldsymbol{\rho}})$. If instead $\nabla\phi(\boldsymbol{\rho}^{(3)}) = \mathbf{0}$, namely $a_i = a_j = a_t$, from $a_i = a_t$ we get $\bar{\theta}_{\boldsymbol{\pi}} = (\sqrt{v_t}\theta_i + \sqrt{v_i}\theta_t)/(\sqrt{v_t} + \sqrt{v_i})$ and, from (2.1),

$$\phi(\boldsymbol{\rho}^{(3)}) = a_t = v_t^{-1}\left(\theta_t - \frac{\sqrt{v_t}\theta_i + \sqrt{v_i}\theta_t}{\sqrt{v_t} + \sqrt{v_i}}\right)^2 = \left(\frac{\theta_i - \theta_t}{\sqrt{v_t} + \sqrt{v_i}}\right)^2 \leq \phi(\tilde{\boldsymbol{\rho}}).$$

By applying the same reasoning, it follows that $\phi(\boldsymbol{\rho}^{(l)}) \leq \phi(\tilde{\boldsymbol{\rho}})$ for $l = 4, \ldots, K$. Indeed, taking into account $\phi(\boldsymbol{\rho}^{(K)})$, if $\nabla\phi(\boldsymbol{\rho}^{(K)}) = \mathbf{0}$ then $a_1 = \ldots = a_K$; from $a_1 = a_K$ we get $\bar{\theta}_{\boldsymbol{\pi}} = (\sqrt{v_K}\theta_1 + \sqrt{v_1}\theta_K)/(\sqrt{v_K} + \sqrt{v_1})$, so that

$$\phi(\tilde{\boldsymbol{\rho}}^{(K)}) = \left(\frac{\theta_1 - \theta_K}{\sqrt{v_1} + \sqrt{v_K}}\right)^2 \leq \phi(\tilde{\boldsymbol{\rho}});$$

while if $\nabla\phi(\boldsymbol{\rho}^{(K)}) \neq \mathbf{0}$, then $\phi(\boldsymbol{\rho}^{(K)})$ could be maximized by some $\boldsymbol{\rho}^{(K-1)}$ and, recursively, it follows that $\phi(\boldsymbol{\rho}^{(K)}) \leq \phi(\tilde{\boldsymbol{\rho}})$, which concludes the proof.

A.3. Proof of Corollary 2.1

Since $\theta_{\tilde{i}} = \theta_{\tilde{i}'}$ and $v_{\tilde{i}} = v_{\tilde{i}'}$, the pair $\{\tilde{i}, \tilde{k}\}$ maximizes the RHS of (2.2) if and only if $\{\tilde{i}', \tilde{k}\}$ maximizes the RHS of (2.2). Therefore, from Theorem 2.1, both $\tilde{\boldsymbol{\rho}}_{\tilde{i}\tilde{k}}$ and $\tilde{\boldsymbol{\rho}}_{\tilde{i}'\tilde{k}}$ are optimal designs as well as every mixture of them, where clearly $\varsigma_{\tilde{i}\tilde{k}} = \varsigma_{\tilde{i}'\tilde{k}}$.

The case of two clusters of treatments follows easily from the previous result by noticing that

$$\max_{i,k\in\{1,\ldots,K\}}\left(\frac{\theta_i - \theta_k}{\sqrt{v_i} + \sqrt{v_k}}\right)^2 = \left(\frac{\theta_1 - \theta_K}{\sqrt{v_1} + \sqrt{v_K}}\right)^2,$$

namely the pair $\{1, K\}$ surely maximizes the RHS of (2.2).

A.4. Proof of Corollary 2.2

Letting $i < k$, then $\theta_i \geq \theta_k$ and therefore the RHS of (2.2) becomes

$$\max_{i<k}\left(\frac{\theta_i - \theta_k}{\sqrt{v_i} + \sqrt{v_k}}\right)^2 = \left(\max_{i<k}\frac{\theta_i - \theta_k}{\sqrt{v_i} + \sqrt{v_k}}\right)^2. \tag{A.3}$$

For binary trials, by letting

$$\frac{\theta_i - \theta_k}{\sqrt{v_i} + \sqrt{v_k}} = \frac{\theta_i - \theta_k}{\sqrt{\theta_i(1-\theta_i)} + \sqrt{\theta_k(1-\theta_k)}} = F(\theta_i, \theta_k),$$

then $F(\cdot)$ attains its minimum when $\theta_i = \theta_k$; moreover, $F(\theta_i, \theta_k)$ is increasing in $\theta_i \in (\theta_k; \theta_1]$, since

$$\frac{\partial F(\theta_i, \theta_k)}{\partial \theta_i} = \frac{\theta_i(1-\theta_k) + \theta_k(1-\theta_i) + 2\sqrt{\theta_i\theta_k(1-\theta_i)(1-\theta_k)}}{2\sqrt{(1-\theta_i)\theta_i}\left(\sqrt{(1-\theta_i)\theta_i} + \sqrt{(1-\theta_k)\theta_k}\right)^2} > 0,$$

and decreasing in $\theta_k \in [\theta_K; \theta_i)$, because

$$\frac{\partial F(\theta_i, \theta_k)}{\partial \theta_k} = \frac{-\theta_i(1-\theta_k) - \theta_k(1-\theta_i) - 2\sqrt{\theta_i\theta_k(1-\theta_i)(1-\theta_k)}}{2\sqrt{(1-\theta_k)\theta_k}\left(\sqrt{(1-\theta_i)\theta_i} + \sqrt{(1-\theta_k)\theta_k}\right)^2} < 0.$$

Thus,

$$\max_{i<k}\left(\frac{\theta_i - \theta_k}{\sqrt{\theta_i(1-\theta_i)} + \sqrt{\theta_k(1-\theta_k)}}\right) = \frac{\theta_1 - \theta_K}{\sqrt{\theta_1(1-\theta_1)} + \sqrt{\theta_K(1-\theta_K)}},$$

namely, the optimal design is $\tilde{\boldsymbol{\rho}}_{1K}$ with

$$\varsigma_{1K} = \sqrt{\theta_1(1-\theta_1)}/(\sqrt{\theta_1(1-\theta_1)} + \sqrt{\theta_K(1-\theta_K)})$$

and $\phi(\tilde{\boldsymbol{\rho}}_{1K}) = \left\{(\theta_1 - \theta_K)/(\sqrt{\theta_1(1-\theta_1)} + \sqrt{\theta_K(1-\theta_K)})\right\}^2$.

Taking into account Poisson outcomes,

$$\max_{i<k}\left(\frac{\theta_i - \theta_k}{\sqrt{\theta_i} + \sqrt{\theta_k}}\right) = \max_{i<k}\left(\sqrt{\theta_i} - \sqrt{\theta_k}\right) = \sqrt{\theta_1} - \sqrt{\theta_K}$$

and therefore $\tilde{\boldsymbol{\rho}}_{1K}$ is the optimal design with $\varsigma_{1K} = \sqrt{\theta_1}/(\sqrt{\theta_1} + \sqrt{\theta_K})$ and $\phi(\tilde{\boldsymbol{\rho}}_{1K}) = \left\{(\theta_1 - \theta_K)/(\sqrt{\theta_1} + \sqrt{\theta_K})\right\}^2$. For exponential responses,

$$\max_{i<k}\left(\frac{\theta_i - \theta_k}{\theta_i + \theta_k}\right) = \frac{\theta_1 - \theta_K}{\theta_1 + \theta_K} = \max_{x>1}\left(\frac{x-1}{x+1}\right),$$

where $x = \theta_i/\theta_k \in [1; \theta_1/\theta_K]$; since the RHS is increasing in x, the optimal design is $\tilde{\boldsymbol{\rho}}_{1K}$ with $\varsigma_{1K} = \theta_1/(\theta_1 + \theta_K)$ and $\phi(\tilde{\boldsymbol{\rho}}_{1K}) = \{(\theta_1 - \theta_K)/(\theta_1 + \theta_K)\}^2$. For normal homoscedastic outcomes, from (A.3),

$$\max_{i<k}\left(\frac{\theta_i - \theta_k}{2\sqrt{v}}\right)^2 = \frac{(\theta_1 - \theta_K)^2}{4v}$$

is attained at the optimal design $\tilde{\boldsymbol{\rho}}_{1K}$ with $\varsigma_{1K} = 1/2$. In the presence of clusters of best(worst) treatments, ς_{1K} (ς_{K1}) should be spanned over the corresponding group of treatments.

A.5. Proof of Lemma 2.2

From now on let $\vartheta_i = \theta_1 - \theta_i$ for $i = 1, \ldots, K$, then $0 = \vartheta_1 \leq \ldots \leq \vartheta_K$, with at least one strict inequality holds. For every $i = 1, \ldots, K-1$, $\sigma_j \geq \beta_j$; indeed,

$$\left(\sum_{i=1}^{K} \frac{\vartheta_i}{v_i}\right)\left(\sum_{i=1}^{j} \frac{1}{v_i}\right) \geq \left(\sum_{i=1}^{j} \frac{\vartheta_i}{v_i}\right)\left(\sum_{i=1}^{K} \frac{1}{v_i}\right) \iff$$
$$\left(\sum_{i=j+1}^{K} \frac{\vartheta_i}{v_i}\right)\left(\sum_{i=1}^{j} \frac{1}{v_i}\right) \geq \left(\sum_{i=1}^{j} \frac{\vartheta_i}{v_i}\right)\left(\sum_{i=j+1}^{K} \frac{1}{v_i}\right)$$

and therefore, by rearranging the terms in the summations,

$$\sigma_j - \beta_j = \sum_{k=1}^{j} \frac{1}{v_k}\left\{\sum_{i=j+1}^{K} \frac{\vartheta_i - \vartheta_j}{v_i}\right\} \geq 0,$$

since $\vartheta_i \geq \vartheta_j$ for every $i > j$. Moreover, note that $\sigma_j = \beta_j$ if and only if $\vartheta_j = \ldots = \vartheta_K$. Analogously, $\beta_j \geq \gamma_j$, since

$$\beta_j - \gamma_j = \sum_{k=1}^{j} \frac{\vartheta_k}{v_k}\left\{\sum_{i=j+1}^{K} \frac{\vartheta_i(\vartheta_i - \vartheta_j)}{v_i}\right\} \geq 0,$$

and the equality holds if i) $0 = \vartheta_1 = \ldots = \vartheta_j$ or ii) $\vartheta_j = \ldots = \vartheta_K$. Case i) corresponds to $\beta_1 = \ldots = \beta_j = \gamma_1 = \ldots = \gamma_j = 0$, while under ii) $\sigma_j = \beta_j = \gamma_j$. If $\gamma_j = 1$, then $\beta_j = \gamma_j$ if and only if $0 = \vartheta_1 = \ldots = \vartheta_K$; indeed, under ii), $\sigma_j = \beta_j = \gamma_j = 1$, that implies $(K-j)\sum_{i=1}^{j} v_i^{-1} = j\sum_{i=j+1}^{K} v_i^{-1}$ and

$$(K-j)\sum_{i=1}^{j} \frac{\vartheta_i}{v_i} = j\vartheta_j \sum_{i=j+1}^{K} \frac{1}{v_i} = \vartheta_j(K-j)\sum_{i=1}^{j} \frac{1}{v_i},$$

namely $\sum_{i=1}^{j}(\vartheta_i - \vartheta_j)/v_i = 0$, i.e. $0 = \vartheta_1 = \ldots = \vartheta_j$, which, combined with ii), corresponds to $\vartheta_1 = \ldots = \vartheta_K = 0$ (i.e., $\theta_1 = \ldots = \theta_K$).

Furthermore, from (2.5), due to the Cauchy-Schwarz inequality

$$T = \left[\sum_{k=1}^{K}\left(v_k^{-1/2}\right)^2\right]\left[\sum_{k=1}^{K}\left(\vartheta_k v_k^{-1/2}\right)^2\right]\left[\sum_{k=1}^{K}\vartheta_k v_k^{-1}\right]^{-2} \geq 1,$$

where the equality holds if and only if $\vartheta_1 = \ldots = \vartheta_K = 0$.

A.6. Proof of Theorem 2.2

To avoid a cumbersome notation, we let $\bar{\vartheta}_{\boldsymbol{\pi}} = \sum_{k=1}^{K} \vartheta_k \pi_k$ and $\bar{\vartheta}_{\boldsymbol{\pi}_B} = (\sum_{k=1}^{K} \vartheta_k v_k^{-1})/(\sum_{k=1}^{K} v_k^{-1})$ (namely assuming $\boldsymbol{\pi}_B = \boldsymbol{\pi}(\boldsymbol{\rho}_B)$). Recalling that $\phi(\boldsymbol{\rho}) = \boldsymbol{a}^t\boldsymbol{\rho}$, where clearly $a_i = v_i^{-1}(\theta_i - \bar{\theta}_{\boldsymbol{\pi}})^2 = v_i^{-1}(\vartheta_i - \bar{\vartheta}_{\boldsymbol{\pi}})^2$.

Moreover, from now on we let $\bar{a}_k = k^{-1}\sum_{i=1}^{k} a_i$ for $k = 1, \ldots, K-1$, while $\bar{a} = K^{-1}\sum_{i=1}^{K} a_i$.

The constrained maximization problem of Theorem 2.2 can be address via Lagrange multipliers, where $L(\boldsymbol{\rho}, \lambda_1, \ldots, \lambda_K) = \phi(\boldsymbol{\rho}) - \sum_{i=1}^{K-1} \lambda_i(\rho_{i+1} - \rho_i) - \lambda_K \left(\sum_{i=1}^{K} \rho_i - 1\right)$. Setting $\partial L(\boldsymbol{\rho}, \lambda_1, \ldots, \lambda_K)/\partial \rho_i = 0$ for $i = 1, \ldots, K$, then

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

namely, by summing all the equations, $\lambda_K = \bar{a} > 0$ (since $\vartheta_K > 0$) and $\lambda_i = i(\bar{a} - \bar{a}_i)$ for $i = 1, \ldots, K-1$.

Case 1: $\lambda_i = 0$ and $\lambda_j > 0$ for every $j \neq i$. Under this scenario, the corresponding target is $\boldsymbol{\rho}_i^* = \xi \sum_{k=1}^{i} \boldsymbol{e}_k + \tau \sum_{k=i+1}^{K} \boldsymbol{e}_k$, where $\tau \in (0, K^{-1}]$ and $\xi = [1 - \tau(K-i)]i^{-1} \geq \tau$, since $\mathbf{1}_K^t \boldsymbol{\rho}_i^* = 1$. Under $\boldsymbol{\rho}_i^*$,

$$\bar{\vartheta}_{\boldsymbol{\pi}} = \bar{\vartheta}_{\boldsymbol{\pi}_B} \cdot \left\{ \frac{\tau K + (1 - \tau K)\beta_i}{\tau K + (1 - \tau K)\sigma_i} \right\}, \tag{A.4}$$

where $[\tau K + (1-\tau K)\beta_i]/[\tau K + (1-\tau K)\sigma_i] \in (\beta_i/\sigma_i, 1]$. Therefore, $\boldsymbol{\rho}_i^*$ is optimal iff $\bar{a} = \bar{a}_i$ and $\bar{a} > \bar{a}_j$ for every $j \neq i$. Condition $\bar{a} = \bar{a}_i$ can be restated as

$$\bar{\vartheta}_{\boldsymbol{\pi}}^2 \left(\frac{1}{K} \sum_{k=1}^{K} \frac{1}{v_k} \right) [1 - \sigma_i] - 2\bar{\vartheta}_{\boldsymbol{\pi}} \left(\frac{1}{K} \sum_{k=1}^{K} \frac{\vartheta_k}{v_k} \right) [1 - \beta_i] + \left(\frac{1}{K} \sum_{k=1}^{K} \frac{\vartheta_k^2}{v_k} \right) [1 - \gamma_i] = 0.$$

Clearly, if $\sigma_i = 1$ then $\bar{\vartheta}_{\boldsymbol{\pi}} = \bar{\vartheta}_{\boldsymbol{\pi}_B}\{T(1-\gamma_i)/[2(1-\beta_i)]\}$. If $\sigma_i \neq 1$ instead, this is a quadratic function wrt $\bar{\vartheta}_{\boldsymbol{\pi}}$ and the corresponding roots are $\bar{\vartheta}_{\boldsymbol{\pi}} = \bar{\vartheta}_{\boldsymbol{\pi}_B} \cdot A_i^{\pm}$, where $A_i^{\pm} = [1 - \beta_i \pm \sqrt{R_i}]/(1 - \sigma_i)$ and $R_i = (1-\beta_i)^2 - T(1-\sigma_i)(1-\gamma_i)$. Thus, by (A.4), $A_i^{\pm} = [\tau K + (1-\tau K)\beta_i]/[\tau K + (1-\tau K)\sigma_i] \in (\beta_i/\sigma_i, 1]$. However, A_i^+ is not an admissible solution since (unless the degenerate case $\theta_1 = \ldots = \theta_K$, which is clearly excluded):

- if $\sigma_i < 1$, $A_i^+ \leq 1 \iff \sqrt{R_i} \leq \beta_i - \sigma_i < 0$;
- if $\sigma_i > 1$, $A_i^+ > \beta_i/\sigma_i \iff \sigma_i\sqrt{R_i} < \beta_i - \sigma_i < 0$.

Taking into account A_i^-, when $\gamma_i \neq 1$ it can be rewritten as $A_i^- = T(1-\gamma_i)/[1 - \beta_i + \sqrt{R_i}]$ (which also encompasses the case $\sigma_i = 1$), and therefore the definition of A_i in (2.6) follows immediately. Now, $A_i \in (\beta_i/\sigma_i, 1]$ when $P1a$ holds. Indeed,

- when $\gamma_i < 1$, $A_i \leq 1 \iff T(1-\gamma_i) \leq 1 + \sigma_i - 2\beta_i$, which also guarantees that $R_i \geq 0$ (since $(1-\beta_i)^2 \geq T(1-\sigma_i)(1-\gamma_i)$ when $\sigma_i \geq 1$, while for $\sigma_i < 1$ then $(1-\beta_i)^2/(1-\sigma_i) \geq 1 + \sigma_i - 2\beta_i$). In addition, $A_i > \beta_i/\sigma_i$ when $T(1-\gamma_i) > \beta_i(2\sigma_i - \beta_i - \beta_i\sigma_i)/\sigma_i^2$ (which also implies that $T(1-\gamma_i) > (1-\beta_i)\beta_i/\sigma_i$, since $\beta_i(1-\beta_i)/\sigma_i < \beta_i(2\sigma_i - \beta_i - \beta_i\sigma_i)/\sigma_i^2$);
- when $\gamma_i > 1$, $A_i > \beta_i/\sigma_i \iff T(1-\gamma_i) > \beta_i(2\sigma_i - \beta_i - \beta_i\sigma_i)/\sigma_i^2$, also ensuring that $R_i \geq 0$ (since $(1-\beta_i)^2/(1-\sigma_i) \leq \beta_i(2\sigma_i - \beta_i - \beta_i\sigma_i)/\sigma_i^2$). Moreover $A_i \leq 1 \iff 1 - \beta_i \leq T(1-\gamma_i) \leq 1 + \sigma_i - 2\beta_i$ (where $T(1-\gamma_i) \geq 1 - \beta_i$ is trivially satisfied when $A_i > \beta_i/\sigma_i$);

- if $\gamma_i = 1$, condition $P1a$ reduces to $2\sigma_i - \beta_i - \beta_i\sigma_i < 0 \leq 1 + \sigma_i - 2\beta_i$.

Condition $\bar{a} > \bar{a}_j$ for every $j \neq i$ corresponds to $P1b$; indeed, by combining

$$\bar{\vartheta}^2_{\boldsymbol{\pi}} \left(\frac{1}{K} \sum_{k=1}^{K} \frac{1}{v_k} \right) [1-\sigma_j] - 2\bar{\vartheta}_{\boldsymbol{\pi}} \left(\frac{1}{K} \sum_{k=1}^{K} \frac{\vartheta_k}{v_k} \right) [1-\beta_j] + \left(\frac{1}{K} \sum_{k=1}^{K} \frac{\vartheta_k^2}{v_k} \right) [1-\gamma_j] > 0$$

($j = 1, \ldots, K-1$, $j \neq i$) with $\bar{\vartheta}_{\boldsymbol{\pi}} = \bar{\vartheta}_{\boldsymbol{\pi}_B} \cdot A_i$, then

$$\frac{\left(\sum_{k=1}^{K} \vartheta_k v_k^{-1}\right)^2 \left[A_i^2(1-\sigma_j) - 2A_i(1-\beta_j)\right]}{\sum_{k=1}^{K} v_k^{-1}} + \left(\sum_{k=1}^{K} \frac{\vartheta_k^2}{v_k} \right) (1-\gamma_j) > 0,$$

namely $T^{-1} \left(\sum_{k=1}^{K} \vartheta_k^2 v_k^{-1}\right) \left[A_i^2(1-\sigma_j) - 2A_i(1-\beta_j) + T(1-\gamma_j)\right] > 0$, i.e., $f_j(A_i) > 0$, where $f_j(z) = z^2(1-\sigma_j) - 2z(1-\beta_j) + T(1-\gamma_j)$. Thus, if there exists a treatment $i \in \{1, \ldots, K-1\}$ s.t. $P1$ holds, then $\boldsymbol{\rho}_i^*$ is optimal with $\tau = (\sigma_i A_i - \beta_i)/\{K[1-\beta_i - A_i(1-\sigma_i)]\}$. Moreover under $\boldsymbol{\rho}_i^*$, since $\bar{a} = \bar{a}_i$,

$$\phi(\boldsymbol{\rho}_i^*) = \boldsymbol{a}^t \boldsymbol{\rho}_i^* = i\bar{a}[1 - \tau(K-i)]i^{-1} + \tau\bar{a}(K-i) = \bar{a}$$

and therefore (2.9) follows directly.

Case 2: $\lambda_i = \lambda_j = 0$ and $\lambda_k > 0$ for every $k \neq \{i, j\}$. Assuming (wlog) $1 \leq i < j \leq K-1$, the corresponding target has the form

$$\boldsymbol{\rho}^* = \left(\frac{1 - (j-i)\eta - (K-j)\nu}{i} \right) \sum_{k=1}^{i} \boldsymbol{e}_k + \eta \sum_{k=i+1}^{j} \boldsymbol{e}_k + \nu \sum_{k=j+1}^{K} \boldsymbol{e}_k, \tag{A.5}$$

where $\eta \geq \nu \in (0, K^{-1}]$ and $j(\eta - \nu) \leq 1 - K\nu \leq j(\eta - \nu) + i(1-\eta)$. Under this scenario, $\bar{a} = \bar{a}_i = \bar{a}_j$ and $\bar{a} > \bar{a}_k$ for $k \neq \{i, j\}$. Thus, if $P1$ holds for both i and j, then $A_i = A_j$ (namely, $\bar{\vartheta}_{\boldsymbol{\pi}} = \bar{\vartheta}_{\boldsymbol{\pi}_B} \cdot A_i = \bar{\vartheta}_{\boldsymbol{\pi}_B} \cdot A_j$) and $P1b$ should be satisfied for every $k \neq \{i, j\}$. Indeed, if $A_i \neq A_j$, from $P1b$ it follows that $f_j(A_i) > 0$ but, at the same time, $f_i(A_j) > 0$, which is impossible since $f_j(A_i) > 0 \iff f_i(A_j) < 0$. Adopting $\boldsymbol{\rho}^*$ in (A.5), since $\bar{a} = \bar{a}_i = \bar{a}_j$,

$$\begin{aligned}\phi(\boldsymbol{\rho}^*) = \boldsymbol{a}^t \boldsymbol{\rho}^* &= \left(\frac{1 - (j-i)\eta - (K-j)\nu}{i} \right) \sum_{k=1}^{i} a_k + \eta \sum_{k=i+1}^{j} a_k + \nu \sum_{k=j+1}^{K} a_k \\ &= \left(\frac{1 - (j-i)\eta - (K-j)\nu}{i} \right) i\bar{a} + \eta\bar{a}(j-i) + \nu\bar{a}(K-j) = \bar{a},\end{aligned}$$

so that, combined with (2.9), $\phi(\boldsymbol{\rho}^*) = \phi(\boldsymbol{\rho}_i^*) = \phi(\boldsymbol{\rho}_j^*)$. Moreover, every mixture of $\boldsymbol{\rho}_i^*$ and $\boldsymbol{\rho}_j^*$ is optimal since, due to the concavity of $\phi(\cdot)$,

$$\phi(\boldsymbol{\rho}_i^*) = \phi(\boldsymbol{\rho}_j^*) \geq \phi\left(\omega\boldsymbol{\rho}_i^* + (1-\omega)\boldsymbol{\rho}_j^*\right) \geq \omega\phi(\boldsymbol{\rho}_i^*) + (1-\omega)\phi(\boldsymbol{\rho}_j^*) = \phi(\boldsymbol{\rho}_i^*) = \phi(\boldsymbol{\rho}_j^*).$$

Case 3: $\lambda_i > 0$ for every $i = 1, \ldots, K-1$. Under this setting, $\boldsymbol{\rho}^* = \boldsymbol{\rho}_B$ and $\bar{\vartheta}_{\boldsymbol{\pi}} = \bar{\vartheta}_{\boldsymbol{\pi}_B}$. Balance is optimal provided that $\bar{a} > \bar{a}_i$ for $i = 1, \ldots, K-1$, i.e.

$$\bar{\vartheta}^2_{\boldsymbol{\pi}_B} \left(\frac{1}{K} \sum_{k=1}^{K} \frac{1}{v_k} \right) [1-\sigma_i] - 2\bar{\vartheta}_{\boldsymbol{\pi}_B} \left(\frac{1}{K} \sum_{k=1}^{K} \frac{\vartheta_k}{v_k} \right) [1-\beta_i] + \left(\frac{1}{K} \sum_{k=1}^{K} \frac{\vartheta_k^2}{v_k} \right) [1-\gamma_i] > 0,$$

namely $f_i(1) > 0$ for every $i = 1, \ldots, K-1$ (condition $P2$) and

$$\phi(\boldsymbol{\rho}_B) = \bar{a} = \frac{1}{K}\left[\bar{\vartheta}^2_{\boldsymbol{\pi}_B}\sum_{k=1}^{K}\frac{1}{v_k} - 2\bar{\vartheta}_{\boldsymbol{\pi}_B}\sum_{k=1}^{K}\frac{\vartheta_k}{v_k} + \sum_{k=1}^{K}\frac{\vartheta_k^2}{v_k}\right] = \left(\frac{T-1}{TK}\right)\sum_{k=1}^{K}\frac{\vartheta_k^2}{v_k}.$$

Case 4: in all the other scenarios, namely when $P2$ does not hold and $\nexists i$ satisfying $P1$, then the Lagrangian does not admit critical points, so that only boundary solutions could be optimal. Due to the constraints $\rho_1^* \geq \ldots \geq \rho_K^*$, then $\rho_i^* = 0$ implies that $\rho_j^* = 0$ for every $j > i$; moreover, $\rho_2^* = 0$ should be excluded since it corresponds to the minimum of the NCP. Starting from the case with $\rho_K^* = 0$ and using the same notation of Appendix A.2, notice that $\phi\left(\boldsymbol{\rho}^{(K-1)^t}; 0\right) = \phi\left(\boldsymbol{\rho}^{(K-1)}\right)$; thus, if $P1$ or $P2$ are satisfied, then $\dot{c} = 1$ and therefore $\phi\left(\boldsymbol{\rho}^{(K-1)}\right) \leq \phi\left(\boldsymbol{\rho}^*_{[K-1]}\right)$; otherwise $\rho^*_{K-1} = \rho^*_K = 0$ and, iteratively, the same reasoning should be applied at most for $\dot{c} = K-2$, for which results of Remark 2.5 hold.

A.7. Proof of Corollary 2.3

If $v(\theta)$ is non-decreasing in θ, then $\{v_k^{-1}; k = 1, \ldots, K\}$, $\{\vartheta_k v_k^{-1}; k = 1, \ldots, K\}$ and $\{\vartheta_k^2 v_k^{-1}; k = 1, \ldots, K\}$ are non decreasing in k and therefore $\{(\sigma_k, \beta_k, \gamma_k); k = 1, \ldots, K-1\}$ is still non-decreasing with $1 \geq \sigma_k \geq \beta_k \geq \gamma_k \geq 0$ for every $k = 1, \ldots, K-1$. Under condition (2.11),

- $T > 1 + \sigma_1$ implies that $P2$ holds, since $(1+\sigma_1)(1-\gamma_i) > 1 + \sigma_i - 2\beta_i$ for $i = 2, \ldots, K-1$;
- while if $T \leq 1 + \sigma_1$, then $A_1 \in (0; 1]$ and $P1$ is satisfied, since $f_i(A_1) > 0$ for every $i = 2, \ldots, K-1$; indeed, it corresponds to

$$T\left[(1-\sigma_i) - (1-\gamma_i)(1-\sigma_1)\right] + 2(\beta_i - \gamma_i)\left[1 + \sqrt{1 - T(1-\sigma_1)}\right] > 0, \tag{A.6}$$

  which trivially holds if $1 - \sigma_i \geq (1-\gamma_i)(1-\sigma_1)$, while for $1 - \sigma_i < (1-\gamma_i)(1-\sigma_1)$ the left-hand side (LHS) in (A.6) is monotonically decreasing in T and (A.6) is satisfied for $T = 1 + \sigma_1$.

For the normal homoscedastic model the proof is straightforward, since $\sigma_i = 1$ for $i = 1, \ldots, K-1$ and condition (2.11) is trivially satisfied because $\beta_i \geq \gamma_i$ for $i = 2, \ldots, K-1$. Under Poisson data, let $\psi_k = \theta_1/\theta_k$ for $k = 1, \ldots, K$ (where $1 = \psi_1 \leq \ldots \leq \psi_K$) and $\bar{\psi} = K^{-1}\sum_{i=1}^{K}\psi_i$ (while $\bar{\psi}_k = k^{-1}\sum_{i=1}^{k}\psi_i$). Thus, $K^{-1}\sum_{k=1}^{K} v_k^{-1} = \bar{\psi}/\theta_1$ and $K^{-1}\sum_{k=1}^{K}\vartheta_k v_k^{-1} = \bar{\psi} - 1$, so that (2.11) becomes

$$\frac{1-\gamma_j}{\bar{\psi}} \geq \frac{\bar{\psi}_j}{\bar{\psi}} + \gamma_j - 2\left(\frac{\bar{\psi}_j - 1}{\bar{\psi} - 1}\right),$$

namely

$$(\bar{\psi}_j - 1)\left[\frac{2}{\bar{\psi}-1} - \frac{1}{\bar{\psi}}\right] \geq \left(\frac{\bar{\psi}+1}{\bar{\psi}}\right)\gamma_j \iff \frac{\bar{\psi}_j - 1}{\bar{\psi} - 1} = \beta_j \geq \gamma_j.$$

For exponential data, by letting $\bar{\psi^2} = K^{-1}\sum_{i=1}^K \psi_i^2$ and $\bar{\psi_k^2} = k^{-1}\sum_{i=1}^k \psi_i^2$, then $K^{-1}\sum_{k=1}^K v_k^{-1} = \bar{\psi^2}/\theta_1^2$ and $K^{-1}\sum_{k=1}^K \vartheta_k v_k^{-1} = (\bar{\psi^2} - \bar{\psi})/\theta_1$, while $K^{-1}\sum_{k=1}^K \vartheta_k^2 v_k^{-1} = \bar{\psi^2} - 2\bar{\psi} + 1$. Hence, after some algebra, (2.11) becomes

$$\bar{\psi_k^2} - 1 \geq \frac{\bar{\psi_k^2} - 2\bar{\psi_k} + 1}{\bar{\psi^2} - 2\bar{\psi} + 1}(\bar{\psi^2} - 1) \iff \frac{\bar{\psi_k} - 1}{\bar{\psi} - 1} \geq \frac{\bar{\psi_k^2} - 1}{\bar{\psi^2} - 1}$$

and, by rearranging the terms in the summation, we obtain

$$(K-k)\left[\sum_{i=1}^k \psi_i(\psi_i - 1)\right] - \sum_{h=1}^k \sum_{i=k+1}^K \psi_h\psi_i(\psi_h - \psi_i) \geq k \sum_{i=k+1}^K \psi_i(\psi_i - 1) \iff$$
$$\sum_{h=1}^k \sum_{i=k+1}^K \psi_h(\psi_i - 1)(\psi_i - \psi_h + 1) \geq \sum_{h=1}^k \sum_{i=k+1}^K \psi_i(\psi_i - 1) \iff$$
$$\sum_{h=1}^k \sum_{i=k+1}^K (\psi_i - 1)(\psi_h - 1)(\psi_i - \psi_h) \geq 0,$$

which is satisfied since $\psi_i \geq 1$ for every $i = 1, \ldots, K$ and $\psi_i \geq \psi_h$ (since $h < i$).

A.8. Proof of Proposition 2.1

For binary trials, $\exists!\,\tilde{i} \in \{1, \ldots, K\}$ s.t. $a_1 \geq \cdots \geq a_{\tilde{i}} \leq a_{\tilde{i}+1} \leq \cdots \leq a_K$, namely $\{a_i, i = 1, \ldots, K\}$ is decreasing for $k \leq \tilde{i}$ and increasing for $k > \tilde{i}$. Indeed, by equating $a_i = a_{i+1}$ we get

$$\bar{\vartheta}_{\boldsymbol{\pi}}^2(v_{i+1} - v_i) - 2\bar{\vartheta}_{\boldsymbol{\pi}}(\vartheta_i v_{i+1} - \vartheta_{i+1} v_i) + \vartheta_i^2 v_{i+1} - \vartheta_{i+1}^2 v_i = 0. \tag{A.7}$$

Clearly, if $v_i = v_{i+1}$, then $\bar{\vartheta}_{\boldsymbol{\pi}} = (\vartheta_i + \vartheta_{i+1})/2$, while if $v_i \neq v_{i+1}$ eq. (A.7) is quadratic wrt $\bar{\vartheta}_{\boldsymbol{\pi}}$, whose corresponding roots are $\bar{\vartheta}_{\boldsymbol{\pi}} = Q_i^{\pm}$, with $Q_i^{\pm} = [\vartheta_i\sqrt{v_{i+1}} \pm \vartheta_{i+1}\sqrt{v_i}]/[\sqrt{v_{i+1}} \pm \sqrt{v_i}]$. However, since $0 = \vartheta_1 < \bar{\vartheta}_{\boldsymbol{\pi}} < \vartheta_K \leq 1$, then Q_i^- is not a feasible solution; indeed, if $v_{i+1} > v_i$, then $Q_i^- < 0$, while if $v_{i+1} < v_i$ then $Q_i^- > 1$. Thus, the only feasible solution is $Q_i^+ = \varsigma_{i+1,i}\vartheta_i + \varsigma_{i,i+1}\vartheta_{i+1} \in [\vartheta_i; \vartheta_{i+1}]$ (which also encompasses the case $v_i = v_{i+1}$) and

$$a_i \geq a_{i+1} \iff (\bar{\vartheta}_{\boldsymbol{\pi}} \geq Q_i^+) \cup (\vartheta_i = \vartheta_{i+1}). \tag{A.8}$$

Since the sequence $\{Q_i^+, i = 1, \ldots, K-1\}$ is increasing in i, combined with (A.8) it guarantees that $\exists!\,\tilde{i} \in \{1, \ldots, K\}$, such that $a_1 \geq \cdots \geq a_{\tilde{i}} \leq a_{\tilde{i}+1} \leq \cdots \leq a_K$. Moreover, if $\bar{a} \geq a_1$, then $\bar{a} > \bar{a}_{K-1}$ (i.e., $\bar{a} < a_K$), since $(K-1)\bar{a} + a_K > \sum_{k=1}^{K-1} a_k + a_K$ is always satisfied due to the fact that $\bar{a} < \max_{i=1,\ldots,K} a_i$.

Case 1: $\lambda_1 = 0$ and $\lambda_i > 0$ for $i = 2, \ldots, K-1$. Under this setting the optimal target is $\boldsymbol{\rho}_1^* = [1 - (K-1)\tau]\boldsymbol{e}_1 + \tau\sum_{k=2}^K \boldsymbol{e}_k$, where τ follows easily from Theorem 2.2 after simple algebra. Condition $\bar{a} = a_1$, namely $A_1 \in (0, 1]$, holds iff $T \leq 1 + \sigma_1$. In this case also $P1b$ is satisfied, i.e. $\bar{a} > \bar{a}_i$ for $i = 2, \ldots, K-1$, since $\bar{a} > \bar{a}_{K-1}$ due to the behaviour of $\{a_i, i = 1, \ldots, K\}$.

Case 2: $\lambda_1 = \cdots = \lambda_j = 0$ (with $j = 2, \ldots, K-2$) and $\lambda_i > 0$ for $i > j$. Under this setting $\bar{a} = \bar{a}_i$ for $i = 1, \ldots, j$ and $\bar{a} > \bar{a}_i$ for $i = j+1, \ldots, K-1$, then $\theta_1 = \cdots = \theta_j$ (which implies $A_1 = \cdots = A_j$) and $P1a$ holds for $i = 1, \ldots, j$ iff $T \leq 1 + \sigma_1$. Following the same reasoning of Case 1, $P1b$ for $i = j+1, \ldots, K-1$ also holds, so that $\boldsymbol{\rho}_1^*, \boldsymbol{\rho}_2^*, \ldots, \boldsymbol{\rho}_j^*$ are optimal designs. Moreover, due to the concavity of $\phi(\cdot)$, every convex combination of $\boldsymbol{\rho}_1^*, \boldsymbol{\rho}_2^*, \ldots, \boldsymbol{\rho}_j^*$ is still optimal.

Case 3: $\lambda_i > 0$ for every $i = 1, \ldots, K-1$. Under this scenario $\boldsymbol{\rho}^* = \boldsymbol{\rho}_B$ and when $T > 1 + \sigma_1$ then $\bar{a} > a_1$, which also guarantees $\bar{a} > \bar{a}_i \, \forall i = 2, \ldots, K-1$ (namely $P2$ holds), due to the behaviour of $\{a_i, i = 1, \ldots, K\}$.

Every other scenario is impossible since:

- $\lambda_1 > 0$ and at least one $\lambda_i = 0$ with $i \in \{2, \ldots, K-1\}$, namely $\bar{a} > a_1$ and $\bar{a} = \bar{a}_i$ where at least $a_i > \bar{a}$, clearly implies that $\bar{a} < a_i < a_{i+1}$. However, if $\lambda_{i+1} > 0$, then $\bar{a} = \bar{a}_i > \bar{a}_{i+1} \iff \bar{a} > a_{i+1}$, while if $\lambda_{i+1} = 0$, then $\bar{a} = \bar{a}_i = \bar{a}_{i+1} \iff \bar{a} = a_{i+1}$;
- $\lambda_1 = 0$ and at least one $\lambda_i > 0$ and $\lambda_{i+1} = 0$ with $i \in \{2, \ldots, K-2\}$, namely $\bar{a} = a_1$ and $\bar{a} = \bar{a}_{i+1}$, so that $a_2 < \bar{a} < a_{i+1} < a_{i+2}$, but if $\lambda_{i+2} = 0$, then $\bar{a} = a_{i+2}$, while if $\lambda_{i+2} > 0$, then $\bar{a} > a_{i+2}$.

References

[1] Abelson, R. P. and Tukey, J. W. (1963). Efficient Utilization of Non-Numerical Information in Quantitative Analysis General Theory and the Case of Simple Order. *The Annals of Mathematical Statistics* **34(4)** 1347–1369. MR0156411

[2] Baldi Antognini, A. and Giovagnoli, A. (2015). *Adaptive Designs for Sequential Treatment Allocation*. Chapman & Hall/CRC Biostatistics. MR3616325

[3] Baldi Antognini, A., Novelli, M. and Zagoraiou, M. (2019). Optimal designs for testing hypothesis in multiarm clinical trials. *Stat. Methods Med. Res.* **28** 3242–3259. MR4002696

[4] Cox, D. R. and Hinkley, D. W. (1974). *Theoretical Statistics*. Chapman and Hall, London. MR0370837

[5] FDA (2018). Guidance for Industry. Adaptive Design Clinical Trials for Drugs and Biologics (draft document). Available on line.

[6] Hirotsu, C. and Herzberg, A. M. (1987). Optimal Allocation of Observations for Inference on k Ordered Normal Population Means. *Australian Journal of Statistics* **29** 151–165. MR0912551

[7] Hu, F. and Rosenberger, W. F. (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials*. New York: John Wiley & Sons. MR2245329

[8] Hu, F. and Zhang, L. X. (2004). Asymptotic properties of doubly adaptive biased coin designs for multi-treatment clinical trials. *Ann. Stat.* **32** 268–301. MR2051008

[9] Lehmann, E. L. and Romano, J. P. (2006). *Testing statistical hypotheses*. Springer Science & Business Media. MR2135927

[10] Lyles, R. H., Lin, H.-M. and Williamson, J. M. (2007). A practical approach to computing power for generalized linear models with nominal, count, or ordinal responses. *Statistics In Medicine* **26** 1632–1648. MR2359164

[11] Rosenberger, W. F. and Lachin, J. M. (2015). *Randomization in clinical trials: theory and practice*. John Wiley & Sons. MR3443072

[12] Singh, S. P. and Davidov, O. (2019). On the design of experiments with ordered treatments. *J. R. Statist. Soc. Series B* **81** 881–900. MR4025401

[13] Sverdlov, O., Tymofyeyev, Y. and Wong, W. K. (2011). Optimal response-adaptive randomized designs for multi-armed survival trials. *Stat. Med.* **30** 2890–2910. MR2844691

[14] Tymofyeyev, Y., Rosenberger, W. F. and Hu, F. (2007). Implementing optimal allocation in sequential binary response experiments. *J. Am. Stat. Assoc.* **102** 224–234. MR2345540

[15] Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society* **54** 426–482. MR0012401

[16] Zhang, L. X. and Rosenberger, W. F. (2007). Response-adaptive randomization for survival trials: the parametric approach. *Appl. Statist.* **56** 153–165. MR2359239

[17] Zhu, H. and Hu, F. (2009). Implementing optimal allocation for sequential continuous responses with multiple treatments. *J. Stat. Plan. Infer.* **139** 2420–2430. MR2508003