A new design strategy for hypothesis testing under response adaptive randomization

Alessandro Baldi Antognini

Department of Statistical Sciences, University of Bologna, Via Belle Arti 41, 40126, Bologna, Italy
e-mail: a.baldi@unibo.it

Alessandro Vagheggini

Unit of Biostatistics and Clinical Trials, Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori (IRST) IRCCS, Via Maroncelli 40, 47014, Meldola (FC), Italy
e-mail: alessandro.vagheggin@irst.emr.it

Maroussa Zagoraiou

Department of Statistical Sciences, University of Bologna, Via Belle Arti 41, 40126, Bologna, Italy
e-mail: maroussa.zagoraiou@unibo.it

and

Marco Novelli

Department of Statistical Sciences, University of Bologna, Via Belle Arti 41, 40126, Bologna, Italy
e-mail: marco.novelli4@unibo.it

Abstract: The aim of this paper is to provide a new design strategy for response adaptive randomization in the case of normal response trials aimed at testing the superiority of one of two available treatments. In particular, we introduce a new test statistic based on the treatment allocation proportion ensuing the adoption of a suitable response adaptive randomization rule that could be more efficient and uniformly more powerful with respect to the classical Wald test. We analyze the conditions under which the suggested strategy, derived by matching an asymptotically best response adaptive procedure and a suitably chosen target allocation, could induce a monotonically increasing power that discriminates with high precision the chosen alternatives. Moreover, we introduce and analyze new classes of targets aimed at maximizing the power of the new statistical test, showing both analytically and via simulations i) how the power function of the suggested test increases as the ethical skew of the chosen target grows, namely overcoming the usual trade-off between ethics and inference, and ii) the substantial gain of inferential precision ensured by the proposed approach.

MSC 2010 subject classifications: Primary 62L05, 62K05; secondary 62G20.

1. Introduction

The demands of individual care and experimental information often come into conflict and the ensuing ethical problem, usually referred to as *individual-versus-collective ethics*, is how to balance the welfare of the patients in the trial against a possible knowledge gain that will improve the care of future patients. In the context of clinical trials, especially in phase-III trials for treatment comparisons, it is widely accepted that Response Adaptive (RA) randomization is a possible answer and this is the reason for which over the past two decades there has been a growing stream of statistical papers on this topic [for a recent review see 2, 5].

Often the conflicting goals related to the ethical demand of maximizing the subjects care and to the statistical aim of drawing correct inferential conclusions with high precision can be formalized into suitable (constrained/combined) optimization problems; in this setting, several authors derived target allocations of the treatments that could be regarded as a valid trade-off among ethics and inference [see e.g. 19, 22, 4]. Generally, these targets depend on the unknown model parameters and they can be approached asymptotically by using suitable RA randomization procedures, namely sequential allocation rules that change at each step the probabilities of treatment assignments to the patients on the basis of earlier responses and past allocations, in order to converge to the chosen target. Classical examples are the well-known doubly-adaptive biased coin design [13] and the efficient randomized adaptive design (ERADE) [14].

Although U.S. government agencies and Health Authorities encourage the adoption of RA procedures [9, 10], their use remains controversial due to some inferential problems that could arise [20, 21]. Indeed, due to the adaptation process, RA designs induce a complex dependence structure between the outcomes; moreover, the allocations are themselves informative on the model parameters and therefore the ensuing statistical inference must be unconditional on the design, so it should take into account the randomness of the design as well [see 12, 5]. In this context, several authors provide the conditions under which the maximum likelihood estimators (MLEs) retain the strong consistency and asymptotic normality properties, that allow one to apply the usual asymptotic inference like, e.g., the classical Wald test. However, the large majority of the design literature is focused on the problem of optimizing the estimation of the treatment effects, while little attention is devoted to hypotheses testing, which has been approached almost exclusively for binary response trials. In particular, Hu and Rosenberger [11, 12] showed that the asymptotic power of the Wald test for testing the equality of the probabilities of success of the treatments is

negatively affected by the variability of the design; then, they introduced the concept of *asymptotically best design*, namely RA procedures which, by satisfying a CLT property with asymptotic variance attaining its minimum (i.e., the Rao-Cramér lower bound), guarantee an asymptotically best power (like, for instance, the ERADE). Adopting a large deviations approach, Azriel et al. [3] derived an optimal target allocation in the case of two treatments that maximizes the asymptotic power of Wald test, showing that the ensuing target is quite close to the balanced one (namely the target allocation optimizing the power under homoscedastic outcomes), and so significantly different from the Neyman allocation (i.e., the optimal target under heteroscedasticity). Moreover, Yi and Wang [23] compared the performances of Wald, score, and likelihood ratio tests under RA designs for binary outcomes, showing how these tests are still asymptotically equivalent, with a slight superiority of Wald test. While in the case of homoscedastic normal response trials, Baldi Antognini et al. [6] recently showed how some target allocations may induce an anomalous behaviour of the power of Wald test, that could be locally decreasing or could vanish as the difference between the treatment effects grows, also suggesting a suitable modified version of Wald statistic which avoids some degenerate scenarios.

Since in some circumstances the treatment allocation proportion ensuing the adoption of an RA design behaves as a proper estimator of the chosen target, that could be consistent and asymptotically normal, the aim of this paper is to provide a new statistical test for clinical trials with normal responses which, based on the proportion of treatment assignments, is very simple to implement and allows one to discriminate with high precision the chosen alternatives. By combining this new test with both i) an asymptotically best RA procedure and ii) a suitably chosen target allocation, it is possible to identify a new design strategy that could be more efficient with respect to the classical Wald test, even if matched with the balanced target in the case of homoscedastic outcomes or the Neyman allocation under heteroscedasticity.

More specifically, we derive the conditions on the suggested design strategy under which the power function of the new test monotonically increases as the sample size or the difference between the treatment effects grows. Moreover, we analyze classes of targets aimed at maximizing the power of the new test, also showing how this power function monotonically increases as the ethical skew of the chosen target grows, namely overcoming the usual trade-off between ethics and inference. Throughout the paper, some simulation studies are performed in order to show the consistent gain of inferential precision ensured by the proposed approach, which leads to a substantial gain of power, in some circumstances 10% higher than the classical Wald test's one.

The paper is structured as follows. Starting from some preliminaries in Section 2, Section 3 introduces the new test for normal homoscedastic outcomes, while Section 4 deals with the case of heteroscedasticity. Several examples are provided and some simulations are performed in order to validate the theoretical results, stressing also some practical implications related to the proposed methodology. Finally, Section 5 deals with some conclusions.

2. Some preliminaries

2.1. Normal model and target allocations

Suppose that patients come to the trial sequentially and are assigned to one of two competing treatments, say A and B. At each step $i \geq 1$, let δ_i be the indicator managing the allocation of the ith subject, more specifically $\delta_i = 1$ if he/she is assigned to A and 0 otherwise, and let Y_i be the corresponding response. Conditionally on the allocations, patients' responses are assumed to be independent following (almost approximately) normal distributions with

$$E(Y_i) = \delta_i \, \mu_A + (1 - \delta_i) \, \mu_B, \quad V(Y_i) = \delta_i \, \sigma_A^2 + (1 - \delta_i) \, \sigma_B^2, \qquad i = 1, \ldots, n,$$

where μ_j and σ_j^2 denote the mean and the variance under treatment $j = A, B$. After n assignments, let $\hat{\mu}_{jn}$ be the MLE of the treatment effect μ_j (i.e., the corresponding sample mean), while $\pi_n = n^{-1} \sum_{i=1}^n \delta_i$ and $(1 - \pi_n)$ are the allocation proportions to A and B, respectively.

Under this setting, inference is usually focused on the difference $\mu = \mu_A - \mu_B$ between the treatment effects, so that from now on suppose that the interest lies in testing the superiority of a treatment (say A) in a "the-larger-the-better scenario", namely testing the null hypothesis $H_0 : \mu = 0$ against the right-tailed alternative $H_1 : \mu > 0$ (the two-tailed alternative can straightforwardly be derived). Notice that, under this scenario, μ_B is usually regarded as a nuisance parameter; moreover, when $\sigma_A^2 = \sigma_B^2 = \sigma^2$ this corresponds to test the stochastic dominance of treatment A over B with respect to the usual stochastic order; while in the case of heteroscedastic responses, when $\sigma_A^2 > \sigma_B^2$ it corresponds to check if A is greater than B in monotone convex order.

The ethical concern consisting in assigning more patients to the better treatment collides with the inferential goal of deriving correct statistical conclusions with high precision. This duality could be overcome by means of the adoption of a suitable target allocation ρ to A (conversely, $1 - \rho$ to B) combining these goals. Generally, these targets should depend on both i) the difference $\mu \in \mathbb{R}$ between the treatment effects, in order to skew the assignments to the better performing treatment, and ii) a tuning parameter $T > 0$ which manages the randomization component of the target, namely low values of T accentuate the ethical skew to the better treatment, while as T grows the ethical component vanishes and the target tends to the balanced one. Thus, from now on we take into account target allocations $\rho(\mu) = \rho(\mu, T)$ satisfying for any fixed $T > 0$ the following properties:

A1: $\rho : \mathbb{R} \to (0, 1)$ with $\rho(x) = 1 - \rho(-x)$;
A2: $\rho(x)$ is increasing in x with $\lim_{x \to \infty} \rho(x) = k \in (1/2, 1]$ (usually $k = 1$);
A3: $\rho(x)$ is twice continuously differentiable in x.

These are natural conditions widely satisfied by the target allocations suggested in the literature. In particular, A1 ensures that both treatments are treated symmetrically, namely the target does not change if the treatment labels are

switched. The ethical requirement A2 ensures that the superior treatment should be favored, while A3 is a technical condition.

As discussed in Baldi Antognini et al. [6], for any chosen T the behaviour of the target ρ could be represented by the cdf of a continuous r.v. centered at 0 and having support on $\mathbb{R}$, like the Normal target suggested in Bandyopadhyay and Biswas [8] and Atkinson and Biswas [1]:

$$\rho_N(x) = \Phi(x/T), \quad x \in \mathbb{R}, \tag{2.1}$$

where Φ is the cdf of the standard normal. This target also coincides with the one proposed by Ivanova et al. [15] in the homoscedastic case with $T = 2\sigma\sqrt{2}$.

From A1, ρ could also be characterized by re-scaling the cdf of a continuous random variable defined in $\mathbb{R}^+ \cup \{0\}$, such as the exponential one

$$\rho_E(x) = 1 - \frac{\exp(-x/T)}{2}, \quad x \geq 0 \tag{2.2}$$

(i.e., the Laplace cdf). Letting $d\rho/dx = \rho'_x$, then $\rho'_x(\cdot)$ can be regarded as the connected pdf: from A2 $\rho'_x(x) \geq 0$ for any $x \in \mathbb{R}$ and, accordingly to A3, $\lim_{x\to\infty} \rho'_x(x) = 0$ and $\lim_{x\to\infty} x\rho'_x(x) = 0$ to ensure integrability. Finally, observe that any given target ρ satisfying A2 with $k = 1$ can be univocally re-scaled to any desired threshold $\tilde{k} \in (1/2, 1)$ by letting $\tilde{\rho} = (2\tilde{k} - 1)\rho + 1 - \tilde{k}$.

Clearly, the target function could not only depend on the difference μ between the treatment effects, but also on some nuisance parameters. In this case, condition A1 could be replaced by

A1': $\rho(x) : \mathbb{R} \to (0,1)$ is invariant under label permutation of the treatments with $\rho(0) = 1/2$.

For instance, to encompass "the-larger-the-better" framework, Zhang and Rosenberger [24] (page 564) briefly discussed the target

$$\rho(x) = \frac{\sigma_A\sqrt{\mu_A}}{\sigma_A\sqrt{\mu_A} + \sigma_B\sqrt{\mu_B}} = \frac{\sigma_A\sqrt{x + \mu_B}}{\sigma_A\sqrt{x + \mu_B} + \sigma_B\sqrt{\mu_B}}, \quad \mu_A, \mu_B > 0.$$

As correctly stated by the authors, this target does not have any natural interpretation. Clearly, it satisfies A1' only in the homoscedastic case; whereas, for heteroscedastic outcomes

$$\rho(0) = \frac{\sigma_A}{\sigma_A + \sigma_B} = \rho^* \neq \frac{1}{2},$$

where ρ^* is the well-known Neyman allocation, one of the most commonly cited targets. However, ρ^* does not have any ethical appeal, since it could assign the majority of patients to the worst treatment, and it does not satisfy assumptions A1'-A2.

2.2. Wald test for RA designs

Given a desired target $\rho(\mu)$, RA randomized procedures could be employed to converge to it. After an initial sample size of n_0 allocations to both treatments

in order to derive non-trivial estimates, at each step $n > 2n_0$, μ is estimated by means of $\hat{\mu}_n = \hat{\mu}_{An} - \hat{\mu}_{Bn}$ and the target is estimated accordingly by $\hat{\rho}_n = \rho(\hat{\mu}_n)$; then, the next assignment is forced to the better performing treatment and the allocation proportion π_n progressively approaches ρ as n grows. Given a target satisfying A1-A3, several authors [see e.g. 17, 7] provided the conditions under which the allocation proportion is a strong consistent estimator of the target, namely $\lim_{n\to\infty} \pi_n = \rho(\mu)$ *a.s.* These conditions, usually satisfied by the RA rules suggested in the literature (like, e.g., the sequential ML design [17], the doubly adaptive biased coin design and the ERADE; for a general discussion see [12, 5]), also guarantee the strong consistency and the asymptotic normality of the MLEs, i.e., $\lim_{n\to\infty}(\hat{\mu}_{An}, \hat{\mu}_{Bn}) = (\mu_A, \mu_B)$ *a.s.* and, as $n \to \infty$,

$$\sqrt{n}\left[\begin{pmatrix}\hat{\mu}_{An}\\\hat{\mu}_{Bn}\end{pmatrix} - \begin{pmatrix}\mu_A\\\mu_B\end{pmatrix}\right] \hookrightarrow \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\dfrac{\sigma_A^2}{\rho(\mu)} & 0\\ 0 & \dfrac{\sigma_B^2}{1-\rho(\mu)}\end{pmatrix}\right), \tag{2.3}$$

that implies

$$\sqrt{n}(\hat{\mu}_n - \mu) \hookrightarrow \mathcal{N}\left(0, \frac{\sigma_A^2[1 - \rho(\mu)] + \sigma_B^2\rho(\mu)}{\rho(\mu)[1 - \rho(\mu)]}\right). \tag{2.4}$$

Taking now into account the case of homoscedastic outcomes with unknown common variance $\sigma^2 = \sigma_A^2 = \sigma_B^2$ (for the heteroscedastic case see Section 4), Wald test statistic is

$$W_n = \sqrt{\frac{n}{\hat{\sigma}_n^2}}\,\hat{\mu}_n\sqrt{\rho(\hat{\mu}_n)[1 - \rho(\hat{\mu}_n)]}, \tag{2.5}$$

where $\hat{\sigma}_n^2$ is the usual pooled sample variance. Since $\lim_{n\to\infty} \pi_n = \rho(\mu)$ *a.s.*, then $\lim_{n\to\infty} \hat{\sigma}_n^2 = \sigma^2$ and $\lim_{n\to\infty} \rho(\hat{\mu}_n) = \rho(\mu)$ *a.s.* Under H_0, W_n converges in distribution to a standard normal r.v. so that, letting z_α be the α-percentile of Φ, the power of the right-sided test W_n of level α can be approximated by

$$\Phi\left(\frac{\sqrt{n}}{\sigma}\mu\sqrt{\rho(\mu)[1 - \rho(\mu)]} - z_{1-\alpha}\right), \quad \mu > 0, \tag{2.6}$$

since in a large sample set-up $\hat{\sigma}_n^2 \approx \sigma^2$ and $\rho(\hat{\mu}_n) \approx \rho(\mu)$. When σ^2 is a-priori known, Wald test (2.5) has the same form with $\hat{\sigma}_n^2$ replaced by σ^2 and therefore the corresponding power function can still be approximated by (2.6). As is well-known, power (2.6) is maximized when the allocations are balanced, namely if $\rho(\mu) = 1/2$ for any μ, which conversely does not satisfy any ethical issue.

Within this framework, a first fundamental requirement of Wald test W_n is that the corresponding power function should tend to one as n grows, that is clearly satisfied for every fixed $\mu > 0$. Moreover, for any fixed n (sufficiently large for the CLT approximation), other additional properties are that the power

should i) converge to one as μ grows and ii) be monotonically increasing in μ. As showed in Baldi Antognini et al. [6], these requirements induce the following restrictions on the adopted target ρ:

$$\lim_{x \to \infty} x^2[1 - \rho(x)] = \infty, \tag{2.7}$$

namely i) is fulfilled for targets with a low ethical improvement (or targets with $k < 1$ in A2), while

$$\rho(x)[1 - \rho(x)] \geq x\rho_x'(x)[\rho(x) - 1/2], \qquad \text{for every } x > 0, \tag{2.8}$$

in order to satisfy ii). Note that these conditions are not trivial and involve the entire functional form of the chosen target. For instance, they fail for ρ_N and ρ_E, while both of them are satisfied by the targets

$$\rho_R(x) = \frac{1}{2} + \frac{x}{2(T + x)}, \quad \text{for } x \geq 0, \tag{2.9}$$

and

$$\rho_S(x) = \frac{1}{2} + \frac{\sqrt{x}}{2(T + \sqrt{x})}, \quad \text{for } x \geq 0. \tag{2.10}$$

3. The new design-based test for homoscedastic normal outcomes

When the desired target satisfies A1-A3, then $\rho(\cdot)$ is monotone in μ and so invertible: in this setting, testing the equality of the treatment effects $H_0 : \mu = 0$ is equivalent to testing $H_0 : \rho(\mu) = 1/2$ (alternative one- or two-tailed hypotheses can be derived accordingly). At the same time, if the chosen RA procedure converges almost surely to the desired target, $\hat{\rho}_n$ and the allocation proportion π_n are competing estimators of ρ. Moreover, when the RA rule satisfies also a CLT property, then the allocation proportion itself could be regarded as a strong consistent and asymptotically normal estimator of ρ and if the design is asymptotically best, then the asymptotic variance of the allocation proportion achieves its minimum.

Thus, in the following we will focus on asymptotically best RA rules, namely RA procedures such that, as $n \to \infty$,

$$\sqrt{n}(\pi_n - \rho(\mu)) \hookrightarrow \mathcal{N}(0, \lambda^2) \tag{3.1}$$

with asymptotic variance attaining the Rao-Cramér lower bound

$$\lambda^2 = \frac{\sigma^2[\rho_\mu'(\mu)]^2}{\rho(\mu)[1 - \rho(\mu)]} \tag{3.2}$$

(recalling that $\mu = \mu_A - \mu_B$, $\partial\rho(\mu_A - \mu_B)/\partial\mu_A = -\partial\rho(\mu_A - \mu_B)/\partial\mu_B$). Thus, letting $\hat{\lambda}_n^2 = \hat{\sigma}_n^2[\rho_\mu'(\hat{\mu}_n)]^2[\pi_n(1 - \pi_n)]^{-1}$ be a suitable consistent estimator of the asymptotic variance λ^2, which employs the current allocation proportion π_n to

estimate $\rho(\mu)$ in order to prevent degenerate scenarios [see 6, for a thorough discussion], we propose the following test statistic

$$Z_n = \hat{\lambda}_n^{-1} \sqrt{n} \left(\pi_n - \frac{1}{2} \right). \tag{3.3}$$

Under H_0, Z_n converges to a standard normal distribution and therefore the rejection region is

$$\pi_n > \frac{\hat{\lambda}_n z_{1-\alpha}}{\sqrt{n}} + \frac{1}{2}.$$

Thus, given the consistency of $\hat{\lambda}_n$, the power of the right-tailed test Z_n of level α can be approximated by

$$\Phi \left(\frac{\sqrt{n}}{\sigma} \left[\frac{\rho(\mu) - 1/2}{\rho'_\mu(\mu)} \right] \sqrt{\rho(\mu)[1 - \rho(\mu)]} - z_{1-\alpha} \right), \quad \mu > 0. \tag{3.4}$$

Remark 3.1. Due to the complex dependence structure induced by RA randomization, the exact distribution of the MLEs as well as that of a given test statistic is not available for a fixed sample size. Therefore, the exact power of a test cannot be computed. Within this framework, the approximated power functions in (2.6) and (3.4) are based on the CLT for the corresponding test statistics. These normal-type approximations have been suggested and applied by several authors (see, e.g., Chapter 3 of Lehmann [16] for a general discussion and [18, 12, 22] for their application to the clinical context) and they are extremely accurate and particularly effective in a moderate-large sample setting, namely the most representative framework in the context of phase-III clinical trials, instead of an asymptotic one (note that, from an asymptotic viewpoint, under the "local alternative" approach the alternative hypothesis takes the form $H_1 : \mu = \delta/\sqrt{n}$, so the power functions of W_n and Z_n converge to the same limit $\Phi \left(\frac{\delta}{2\sigma} - z_{1-\alpha} \right)$, namely these tests perform equally since their asymptotic local powers coincide).

Thus, in this paper we compare tests W_n and Z_n from a methodological viewpoint on the basis of the approximated power functions in (2.6) and (3.4), henceforth referred to as power or approximated power, likewise. The ensuing theoretical properties will also be explored through simulations, where the behaviour of the simulated power confirms our analytical results.

3.1. Properties of the power of test Z_n

To analyze the behaviour of the power of test Z_n it is suitable to rewrite (3.4) as $\Phi \left(\sqrt{n} \sigma^{-1} {}_\rho G(\mu) - z_{1-\alpha} \right)$, since the power depends on μ only through the adopted target by the function

$$_\rho G = {}_\rho G(x) = \frac{[\rho(x) - 1/2]}{\rho'_x(x)} \sqrt{\rho(x)[1 - \rho(x)]}, \qquad x > 0. \tag{3.5}$$

Firstly, observe that $_\rho G(x) > 0$ for any $x > 0$, so that the power of test Z_n tends to one as n grows regardless of the adopted target. Whereas, function (3.4) tends to one as μ grows if and only if the target is chosen such that $\lim_{x\to\infty} {_\rho G(x)} = \infty$, namely if ρ satisfies the following condition:

$$\lim_{x\to\infty} \frac{1 - \rho(x)}{[\rho'_x(x)]^2} = \infty. \tag{3.6}$$

This is trivially fulfilled for targets with $k \in (1/2, 1)$ in A2; whereas when $k = 1$, it is easy to check that condition (3.6) is also satisfied by any previously introduced target.

The following Lemma 1 and Corollary 3.1 provide the conditions on the chosen target under which the ensuing approximated power of test Z_n is monotonically increasing in μ.

Lemma 1. *Approximated power function (3.4) is monotonically increasing in μ if and only if the target is chosen such that*

$$1 - \frac{[\rho(x) - 1/2]^2}{\rho(x)[1 - \rho(x)]} \geq \frac{\rho''_{xx}(x)}{[\rho'_x(x)]^2}[\rho(x) - 1/2], \quad \text{for any } x > 0. \tag{3.7}$$

Proof. From (3.5),

$$_\rho G'_x(x) = \frac{[\rho'_x(x)]^2 - \rho''_{xx}(x)[\rho(x) - 1/2]}{[\rho'_x(x)]^2} \sqrt{\rho(x)[1 - \rho(x)]} - \frac{[\rho(x) - 1/2]^2}{\sqrt{\rho(x)[1 - \rho(x)]}},$$

so that power function (3.4) is increasing in μ if and only if inequality (3.7) is verified. □

In order to identify suitable classes of targets satisfying (3.7), for any given ρ let us introduce the corresponding hazard function $_\rho h(x) = \rho'_x(x)/[1 - \rho(x)]$ for $x > 0$. The following Corollary provides some sufficient conditions on the target allocation guaranteeing that the approximated power of test Z_n is monotonically increasing in μ.

Corollary 3.1. *Given a target allocation ρ, the approximated power function of test Z_n is monotonically increasing in μ if*

$$\frac{1/2 - \rho(x)}{\rho(x)[1 - \rho(x)]} - \frac{\rho''_{xx}(x)}{[\rho'_x(x)]^2} \geq -2, \quad \text{for every } x > 0. \tag{3.8}$$

Moreover, additional conditions guaranteeing the non-negativity of the LHS of (3.8), and therefore the monotonicity in μ, are:

$$_\rho h'_x(x) \leq 0, \quad \text{for every } x > 0, \tag{3.9}$$

or

$$_\rho h'_x(x) > 0 \quad \text{and} \quad {_\rho h^2(x)} \geq 2\,_\rho h'_x(x)\rho(x), \quad \text{for every } x > 0. \tag{3.10}$$

Proof. See Appendix A.1 □

Example 3.1. Taking into account ρ_E in (2.2), the corresponding hazard $_{\rho_E}h(x) = T^{-1}$ is constant in x and therefore the approximated power of test Z_n is monotonically increasing in μ. The same conclusion still holds for ρ_S in (2.10), since $_{\rho_S}h(x) = [2(T\sqrt{x} + x)]^{-1}$ is decreasing in x for every $T > 0$.

Inspired by the Logistic cdf, assuming now the following target

$$\rho_L(x) = [1 + \exp(-x/T)]^{-1}, \quad x \in \mathbb{R}. \tag{3.11}$$

The resulting hazard $_{\rho_L}h(x) = \exp(x/T)[T\exp(x/T) + T]^{-1}$ satisfies condition (3.10), since it is increasing in x and $[\exp(x/T) - 1][\exp(x/T) + 1] > 0$ for every $x, T > 0$; therefore, the power is increasing in μ.

3.2. Gain of power

Now we compare the performances of test Z_n to those of the Wald's one given the same desired target allocation $\rho(\mu)$. The following Theorem shows how Z_n could induce a gain of power with respect to Wald test.

Theorem 3.1. *Adopting a given target ρ, test Z_n is uniformly more powerful than Wald test W_n (i.e., the approximated power (3.4) is uniformly greater than the one in (2.6)) if and only if*

$$\rho(x) - 1/2 \geq x\rho'_x(x), \qquad \text{for any } x > 0, \tag{3.12}$$

with a strict inequality for certain values of $x > 0$. Whereas, Wald test cannot be uniformly more powerful than test Z_n.

Proof. The first statement follows directly from (2.6) and (3.4). Moreover, the non-superiority of Wald test could be easily argued by observing that, due to the property of the target function ρ, the RHS of (3.12) goes to 0 as x grows, while the LHS tends to $k - 1/2 > 0$. Therefore, it does not exists a target such that $\rho(x) - 1/2 \leq x\rho'_x(x)$ for every $x > 0$. □

Example 3.2. By adopting ρ_L in (3.11), test Z_n is uniformly more powerful than W_n. Indeed, from (3.12),

$$\frac{\exp(x/T) - 1}{2[\exp(x/T) + 1]} \geq \frac{x\exp(x/T)}{T[\exp(x/T) + 1]^2},$$

i.e., $T[\exp(2x/T) - 1] \geq 2x\exp(x/T)$ for any $x, T > 0$. The same conclusion still holds for ρ_R in (2.9) and ρ_S in (2.10), via simple algebra. Moreover, condition (3.12) is also fulfilled by ρ_E in (2.2), since $T[1 - \exp(-x/T)] \geq x\exp(-x/T)$ for every $x, T > 0$, and for ρ_N in (2.1) because

$$\Phi(x/T) \geq \frac{x\exp\left(-\dfrac{x^2}{2T^2}\right)}{T\sqrt{2\pi}}$$

is satisfied for any $x, T > 0$ (indeed, for $x > 0$ the LHS is greater than $1/2$, while at $x = T$ the RHS attains its maximum given by $(2\pi e)^{-1/2} \cong 0.242$).

Taking now into account the following target

$$\rho_B(x) = \frac{1}{2} + \frac{1}{2}\left(\frac{x}{1+x}\right)^T, \qquad x \geq 0, \tag{3.13}$$

condition (3.12) is satisfied only for $T \leq 1$; indeed,

$$\frac{1}{2}\left(\frac{x}{x+1}\right)^T \geq \frac{1}{2}\frac{T}{x+1}\left(\frac{x}{x+1}\right)^T$$

only when $x \geq T - 1$ and therefore test Z_n is more powerful than W_n only for $T \leq 1$.

In order to validate these theoretical results and provide some practical implications of the suggested methodology, we have performed a simulation study by adopting ERADE (with randomization parameter 0.5) and starting sample size $n_0 = 1$. The results come from 5000 simulations with $n = 250$, where the responses are assumed homoscedastic normal with $\sigma^2 = 1$ and $\mu_B = 1$. Figure 1 compares the simulated power of Wald test W_n (dashed line) with those of Z_n (solid line) under targets ρ_E and ρ_R for $T = 1$ and 5.

The simulations show the gain of power induced by the adoption of the proposed design strategy. Notice that, even if in this Figure μ varies between 0 and 0.8 – to highlight the differences between the power functions – ρ_E does not satisfy (2.7) and (2.8) and therefore under this target the power of Wald test is not monotonically increasing and vanishes as μ grows (see Table 1 and [6]).

Note that, when the target has a high ethical component (in particular, when ρ'_x grows extremely fast locally around zero), slightly inflated type-I errors of test Z_n could be present (in any case, not greater than 6.5% and not higher than 1.5% with respect to the one of Wald test). This is due to the functional form of the chosen target which, around zero, could induce a slightly unstable behaviour of π_n as an estimator of $\rho(\mu)$. However, this inflated type-I error for test Z_n is not present for every target. For instance, Table 1 compares the simulated power of W_n and Z_n adopting ρ_N and ρ_L with $T = 1$ (we stress the degenerate behaviour of the power of Wald test for large values of μ due to the fact that these targets do not satisfy conditions (2.7) and (2.8)).

We now compare the approximated power of test Z_n ensuing from the adoption of an asymptotically best RA design converging to a given target $\rho(\mu)$ with the one of the classical Wald test under its optimal scenario, namely when the allocations are balanced, showing the conditions on the chosen target guaranteeing the superiority of the proposed design strategy also in this situation.

Theorem 3.2. *Adopting an asymptotically best RA design converging to a target ρ chosen such that*

$$\frac{[\rho(x) - 1/2]\sqrt{\rho(x)[1 - \rho(x)]}}{\rho'_x(x)} \geq \frac{x}{2}, \quad \text{for any } x > 0, \tag{3.14}$$

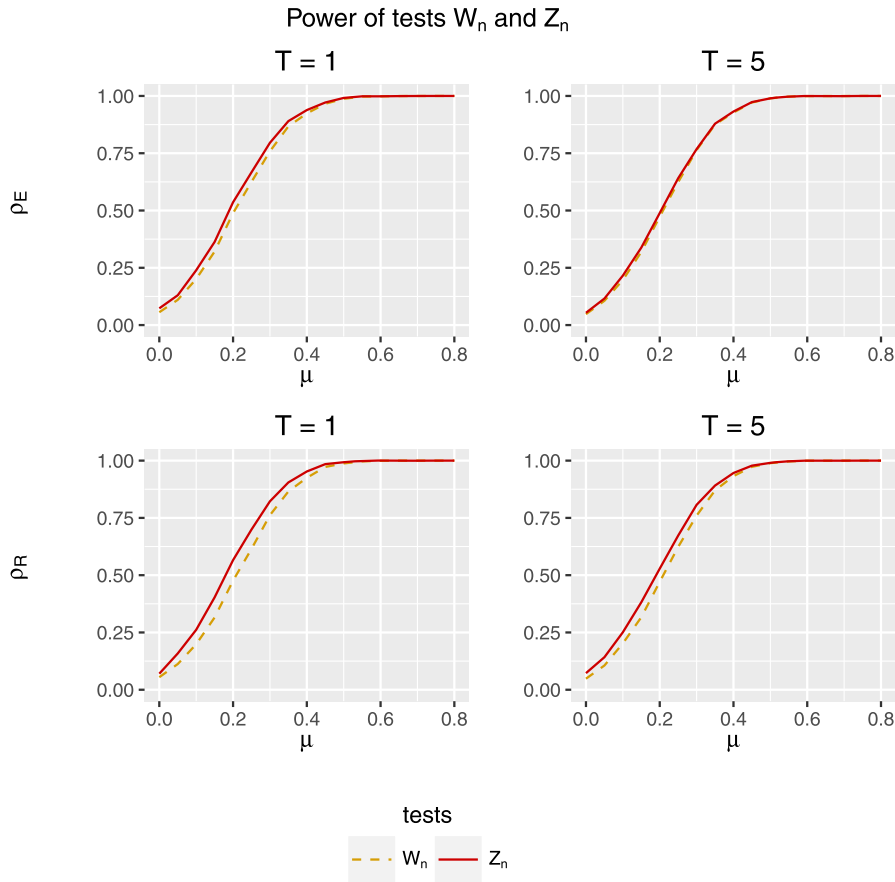*then test Z_n is uniformly more powerful than W_n equipped by balance.*

Power of tests W$_n$ and Z$_n$

FIG 1. *Simulated power of tests W_n (dashed line) and Z_n (solid line) under the ERADE with targets ρ_E and ρ_R.*

Proof. In the case of balance, the approximated power function of Wald test simply becomes $\Phi\left((2\sigma)^{-1}\sqrt{n}\mu - z_{1-\alpha}\right)$, for $\mu > 0$; therefore, condition (3.14) follows directly from (3.4). □

Example 3.3. Adopting an asymptotically best RA design converging to ρ_L, test Z_n is uniformly more powerful than W_n under the balanced design. Indeed, condition (3.14) becomes

$$\frac{T[\exp(x/T) - 1]}{\sqrt{\exp(x/T)}} \geq x,$$

which is satisfied for any $x, T > 0$. This still holds for ρ_E, since

$$T\exp(-x/T)[\exp(x/T) - 1]\sqrt{2\exp(x/T) - 1} \geq x, \qquad \text{for any } x, T > 0,$$

TABLE 1
Simulated power of tests W_n and Z_n adopting ρ_N and ρ_L with $T = 1$.

	ρ_N		ρ_L	
μ	W_n	Z_n	W_n	Z_n
0.0	0.05	0.05	0.05	0.05
0.1	0.20	0.21	0.20	0.21
0.2	0.46	0.47	0.46	0.51
0.3	0.76	0.76	0.76	0.78
0.4	0.93	0.93	0.93	0.95
0.5	0.99	0.99	0.99	0.99
0.6	1.00	1.00	1.00	1.00
1.0	1.00	1.00	1.00	1.00
2.0	0.93	1.00	1.00	1.00
3.0	0.50	1.00	1.00	1.00
4.0	0.08	1.00	1.00	1.00
5.0	0.00	1.00	1.00	1.00
8.0	0.00	1.00	0.90	1.00
10.0	0.00	1.00	0.05	1.00

for ρ_R (because $\sqrt{T(T + 2x)} \geq T$ for any $x, T > 0$) and for ρ_S (since $T/2 \leq \sqrt{T(T + 2\sqrt{x})}$ for any $x, T > 0$). Whereas, adopting ρ_N or ρ_B condition (3.14) does not hold.

Under the same simulation scenario of Example 3.2, Figure 2 shows how the proposed design strategy with targets ρ_E and ρ_R outperforms in terms of power the classical Wald test equipped by balance (note that, in our setting, when $\rho(\mu) = 1/2$ for every μ, then the ERADE becomes an Efron's coin with bias parameter 0.75). As regards test Z_n, the gain of power is quite evident and verifies the previously obtained theoretical results. Clearly, as T increases the ethical component of the targets vanishes and, asymptotically, the allocations tend to be balanced. Therefore, in this case the simulated power of tests Z_n and W_n tend to coincide. As previously showed, test Z_n presents slightly inflated type-I errors, while W_n is quite conservative (due to the balanced allocation) with a type-I error of 0.046. As discussed in Example 3.2, the inflated type-I error of test Z_n can be overcome by taking into account target allocations with a low slope around $\mu = 0$, like ρ_L. Table 2 quantifies numerically the gain in terms of power induced by the proposed strategy, stressing also the ethical impact related to the chosen target. In particular, we compare the values of the power of tests W_n (equipped by balance) and Z_n adopting ρ_R and ρ_L (with $T = 1$) as μ varies, showing a substantial improvement in terms of inferential precision and ethics as well. For instance, taking into account ρ_R, the gain of power is about 11% at $\mu = 0.2$ (also assigning 8% more subjects to the best treatment), while the inflation of type-I error is around 1%. Whereas, adopting ρ_L the type-I error is preserved and the gain in terms of power wrt Wald test is rather high even if lower with respect to ρ_R (around 5% at $\mu = 0.2$ with an ethical improvement of 5%). We stress that, adopting ethical targets $\rho = \rho(\mu)$, although Wald test preserves the nominal type-I error, its power could be locally decreasing or could even vanish as μ grows (see, e.g., [6] and Table 1).
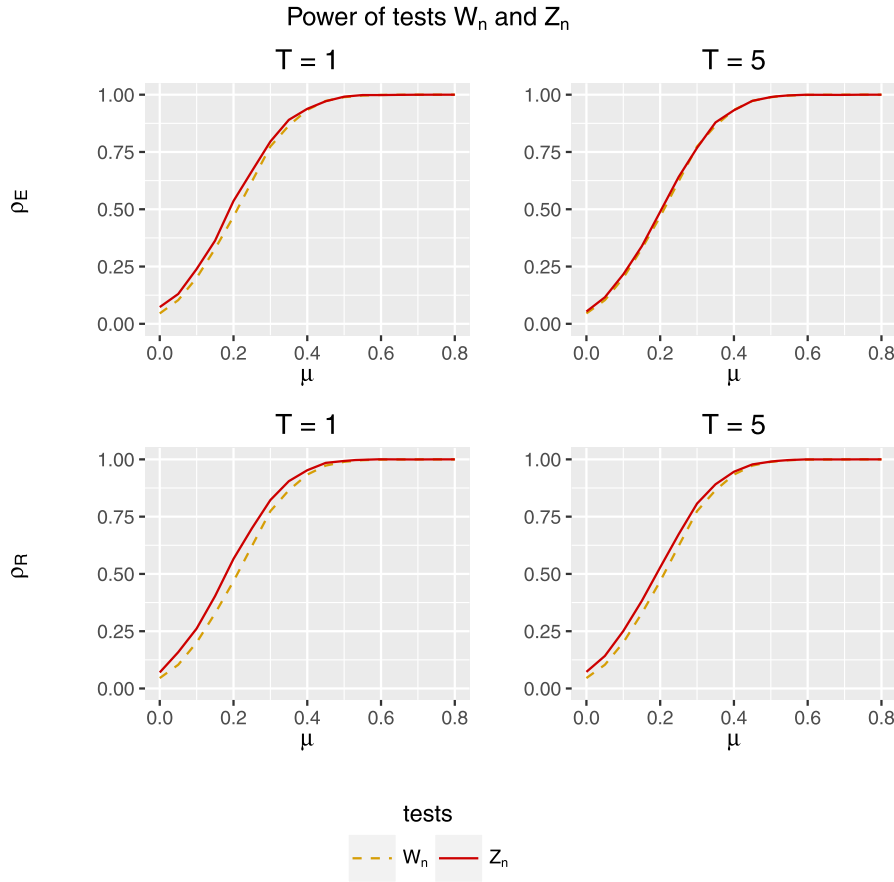
FIG 2. *Simulated power of tests W_n (dashed line) under the balanced allocation and Z_n (solid line) adopting targets ρ_E and ρ_R for $T = 1$ and 5.*

3.3. Ethics improves inference

Taking into account the suggested design strategy, this subsection deals with the improvement in terms of power that could be induced by increasing the ethical skew of suitably chosen targets. In order to stress the dependence on T, in what follows we denote the target function by $\rho = \rho(x, T)$ and, analogously, $_\rho h = {}_\rho h(x, T)$ and $_\rho G = {}_\rho G(x, T)$. Since T manages the randomization component increasingly, while the ethical skew grows as $T \to 0$, from now on we assume that $\rho(x, T)$ is differentiable in T with $\rho'_T(x, T) \leq 0$ for any x, T (analogously to the previous notation, for any function $\varphi(s, t)$ we let $\varphi'_s = \partial\varphi/\partial s$, $\varphi'_t = \partial\varphi/\partial t$ and $\varphi''_{st} = \partial^2\varphi/\partial s\partial t$).

The following results show how, by adopting the new test Z_n, it is possible to overcome the usual trade-off between ethical demands and inferential precision.

TABLE 2

Simulated power of tests W_n under the balanced allocation and Z_n adopting ρ_R and ρ_L with $T = 1$ (within brackets the percentage of subjects allocated to the best treatment).

	W_n	Z_n	
μ	$\rho = 1/2$	ρ_R	ρ_L
0.0	0.05	0.06 (50%)	0.05 (50%)
0.1	0.20	0.26 (55%)	0.21 (53%)
0.2	0.46	0.57 (58%)	0.51 (55%)
0.3	0.77	0.82 (62%)	0.78 (57%)
0.4	0.93	0.95 (64%)	0.95 (60%)
0.5	0.99	0.99 (67%)	0.99 (62%)
0.6	1.00	1.00 (69%)	1.00 (65%)

Theorem 3.3. *If the target $\rho(x, T)$ is chosen such that, for any $x, T > 0$,*

$$\frac{1}{\rho(x,T) - 1/2} + \frac{1}{2\rho(x,T)[1 - \rho(x,T)]} \geq \frac{\rho h'_T(x,T)}{\rho h(x,T)\rho'_T(x,T)}, \qquad (3.15)$$

then the approximated power function of test Z_n is decreasing in T.

Proof. See Appendix A.2. □

Corollary 3.2. *Any given target $\rho(x, T)$ under which the function*

$$\frac{\rho h(x,T)}{\rho(x,T) - 1/2} \qquad (3.16)$$

is increasing in T satisfies (3.15).

Proof. From (3.15), a sufficient condition for the decreasingness in T is that

$$\frac{1}{\rho(x,T) - 1/2} \geq \frac{\rho h'_T(x,T)}{\rho h(x,T)\rho'_T(x,T)}, \quad \text{for any } x, T > 0,$$

i.e.,

$$\frac{\rho h'_T(x,T)}{\rho h(x,T)} - \frac{\rho'_T(x,T)}{\rho(x,T) - 1/2} \geq 0, \quad \text{for any } x, T > 0,$$

which is fulfilled if the function $\ln\{\rho h(x,T)/[\rho(x,T) - 1/2]\}$ is increasing in T. Given the positiveness of $\rho h(x,T)/[\rho(x,T) - 1/2]$, the thesis follows directly. □

Example 3.4. Target ρ_R in (2.9) satisfies Corollary 3.2, since $\rho_R h(x,T) = (T + x)^{-1}$ and therefore

$$\frac{\rho_R h(x,T)}{\rho_R(x,T) - 1/2} = \frac{2}{x}.$$

The same conclusion still holds for ρ_B in (3.13); indeed, the function

$$\frac{\rho_B h(x,T)}{\rho_B(x,T) - 1/2} = -\frac{2T}{x(x+1)\left[\left(\dfrac{x}{x+1}\right)^T - 1\right]}$$

is increasing in T because its partial derivative with respect to T is non-negative for any $x, T > 0$, since

$$1 + \left(\frac{x}{x+1}\right)^T \left[T \ln\left(\frac{x}{x+1}\right) - 1\right] \geq 0.$$

Whereas ρ_E does not satisfy Corollary 3.2 but fulfills condition (3.15), since

$$\frac{1}{\exp(x/T) - 1} + \frac{\exp(x/T)}{2\exp(x/T) - 1} \geq \frac{T}{x} \qquad \text{for every } x, T > 0.$$

The same holds for ρ_L and ρ_S after simple algebra.

Figure 3 shows the behaviour of the simulated power of test Z_n under targets ρ_L, ρ_E and ρ_R for $T = 1$ (solid line) and $T = 3$ (dashed line); results come from the same simulation scenario as before (see Example 3.2). This Figure confirms that, for any considered target, a high ethical component (i.e., low values of T) induces a gain of power.

Thus, adopting test Z_n with a target satisfying (3.15) (or, analogously, (3.16)), small values of T guarantee that more subjects will be assigned to the better treatment and, at the same time, a gain in terms of power to discriminate with higher precision the chosen alternative.

Remark 3.2. It is worth noticing that the classical Wald test does not share this peculiarity. Indeed, from (2.6), the approximated power function of W_n could be expressed as $\Phi\left(\sqrt{n}\sigma^{-1}{}_\rho \tilde{G}(x, T) - z_{1-\alpha}\right)$, where now ${}_\rho \tilde{G}(x, T) = x\sqrt{\rho(x, T)[1 - \rho(x, T)]}$ for $x, T > 0$. Recalling that $\rho'_T(x, T) \leq 0$ for any $x, T > 0$, then

$$_\rho \tilde{G}'_T(x, T) = -\frac{x\rho'_T(x, T)[\rho(x, T) - 1/2]}{\sqrt{\rho(x, T)[1 - \rho(x, T)]}} \geq 0, \qquad \text{for every } x, T > 0,$$

and therefore the approximated power is increasing in T.

The adoption of test Z_n induces a partial order between targets on the basis of ${}_\rho G$, namely given two targets ρ_1 and ρ_2 then $\rho_1 \geq_G \rho_2$ if and only if ${}_{\rho_1}G(x, T) \geq {}_{\rho_2}G(x, T)$ for every $x, T > 0$ (i.e., ρ_1 guarantees a uniformly more powerful test than ρ_2). Within this framework, if two targets have the same functional form with different values of the randomization parameter (namely, $\rho_l = \rho(x, T_l)$), then Theorem 3.3 and Corollary 3.2 characterize the classes of targets guaranteeing a simultaneous improvement of the ethical skew and the approximated power of test Z_n.

Furthermore, for any fixed T, it is possible to find classes of targets having better performances in terms of power with respect to others, as the following Remark shows.

Remark 3.3. Taking into account targets ρ_E in (2.2) and ρ_L in (3.11), then $\rho_E \geq_G \rho_L$, since for any fixed $T > 0$
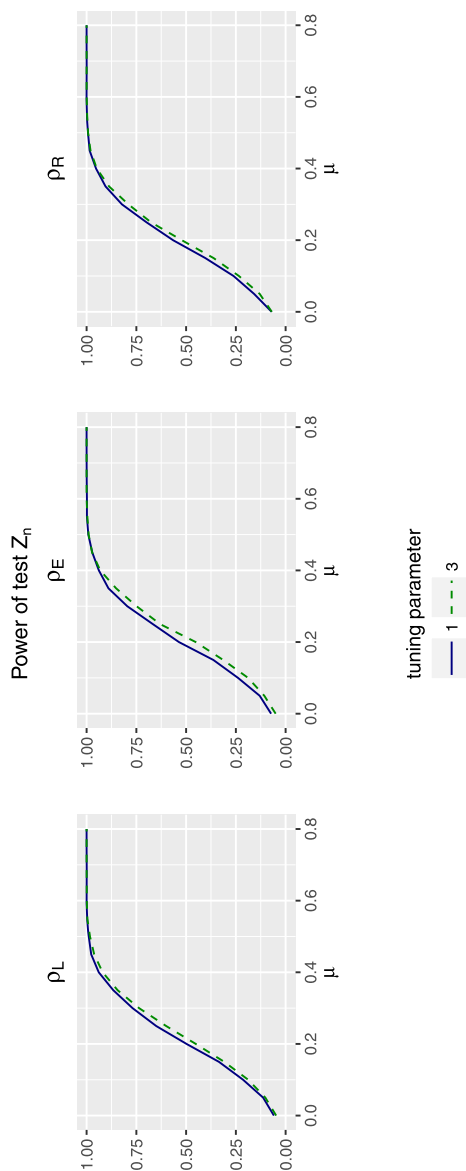
FIG 3. *Simulated power of test Z_n adopting targets ρ_L, ρ_E and ρ_R with $T = 1$ (solid line) and $T = 3$ (dashed line).*

$$\rho_E G(x, T) = \frac{T}{2} \exp(-x/T)[\exp(x/T) - 1]\sqrt{2\exp(x/T) - 1}$$

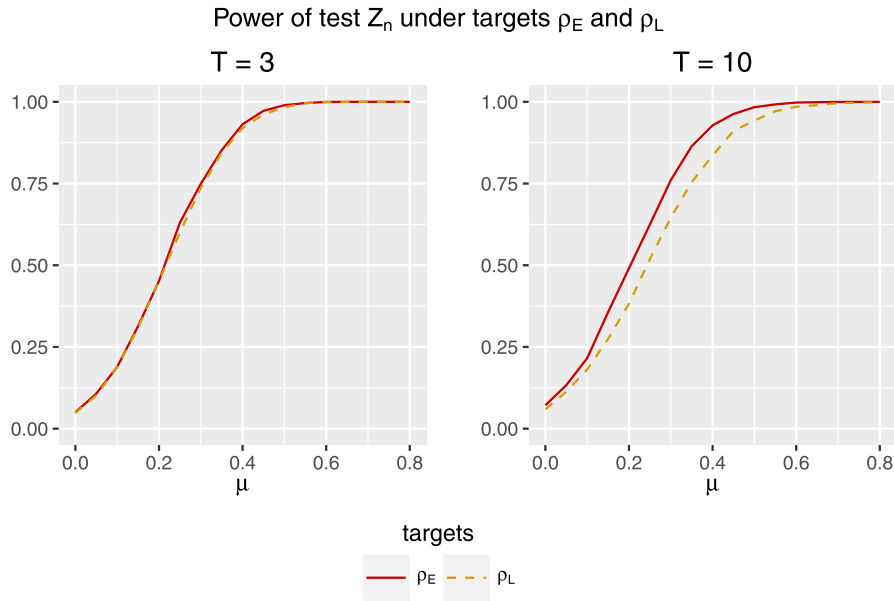$$\geq \frac{T[\exp(x/T) - 1]}{2\sqrt{\exp(x/T)}} = \rho_L G(x, T)$$

for every $x > 0$.

Fɪɢ 4. *Simulated power of test Z_n adopting ρ_E (solid line) and ρ_L (dashed line) for $T = 3$ and 10.*

Under the previous simulation scenario, Figure 4 shows the gain of power induced by ρ_E with respect to ρ_L when $T = 3$ and 10. This gain is quite substantial for high values of T, while, even if it is still present when $T \leq 2$, this gain is not graphically appreciable.

4. Extension to the case of heteroscedastic outcomes

4.1. Wald test

Under heteroscedasticity, from (2.4) the Wald statistic becomes

$$W_n = \sqrt{n}\hat{\mu}_n \sqrt{\frac{\rho(\hat{\mu}_n, T)[1 - \rho(\hat{\mu}_n, T)]}{\hat{\sigma}_{An}^2[1 - \rho(\hat{\mu}_n, T)] + \hat{\sigma}_{Bn}^2 \rho(\hat{\mu}_n, T)}}, \qquad (4.1)$$

where $\hat{\sigma}_{jn}^2$ denotes the usual sample variance of treatment $j = A, B$. Given the consistency of $\rho(\hat{\mu}_n, T)$ and $\hat{\sigma}_{jn}^2$, under H_0 statistics W_n converges in distribution to a standard normal random variable; therefore, the power of the right-sided test of level α can be approximated by

$$\Phi\left(\frac{\sqrt{n}}{\sigma_A}\mu\sqrt{\frac{\rho(\mu, T)[1 - \rho(\mu, T)]}{1 - \rho(\mu, T) + \nu\rho(\mu, T)}} - z_{1-\alpha}\right), \quad \mu > 0, \qquad (4.2)$$

where $\nu = \sigma_B^2/\sigma_A^2$. As is well-known, function (4.2) is maximized by assuming the Neyman target $\rho^* = (1 + \sqrt{\nu})^{-1}$ (discussed in Section 2).

Letting now for any $T > 0$,

$$_\rho\tilde{G}(x,T) = x\sqrt{\frac{\rho(x,T)[1 - \rho(x,T)]}{1 - \rho(x,T) + \nu\rho(x,T)}}, \qquad \text{for } x > 0, \qquad (4.3)$$

it is straightforward to check that $_\rho\tilde{G}(x,T) > 0$ for every $x, T, \nu > 0$, and therefore power function (4.2) tends to one as $n \to \infty$. Moreover, (2.7) is still valid to ensure that the power goes to one as μ grows, while the following Theorem provides the condition under which the power of Wald test is monotonically increasing in μ.

Theorem 4.1. *A target ρ induces a monotonically increasing approximated power function of Wald test W_n if and only if, for $T > 0$,*

$$1 - \rho(x,T) \geq \frac{x\rho_x'(x,T)}{2}, \qquad \text{for any } x > 0. \qquad (4.4)$$

Proof. See Appendix A.3.   □

Example 4.1. Target ρ_R in (2.9) satisfies condition (4.4), since for any $T > 0$

$$\frac{T}{T + x} \geq \frac{Tx}{2(T + x)^2}, \qquad \text{for every } x > 0,$$

and therefore the power of test W_n is monotonically increasing in μ. The same holds for ρ_S after simple algebra, since $4T + 3\sqrt{x} \geq 0$ for every $T, x > 0$.

On the contrary, condition (4.4) is not met by ρ_L (since $2T[1 + \exp(x/T)] \geq x\exp(x/T)$ does not hold for any $x > 0$) and by ρ_E (because (4.4) is satisfied only for $x \leq 2T$).

4.2. The new test under heteroscedasticity

In the case of heteroscedastic outcomes, the asymptotic variance of the treatment allocation proportion under an asymptotically best RA procedures is

$$\lambda^2 = \left[\rho_\mu'(\mu,T)\right]^2 \frac{\sigma_A^2[1 - \rho(\mu,T)] + \sigma_B^2\rho(\mu,T)}{\rho(\mu,T)[1 - \rho(\mu,T)]}, \qquad (4.5)$$

which can be consistently estimated at each step n by

$$\hat{\lambda}_n^2 = [\rho_\mu'(\hat{\mu}_n,T)]^2 \frac{\hat{\sigma}_{An}^2[1 - \pi_n] + \hat{\sigma}_{Bn}^2\pi_n}{\pi_n(1 - \pi_n)}.$$

Under H_0, test $Z_n = \hat{\lambda}_n^{-1}\sqrt{n}\,(\pi_n - 1/2)$ asymptotically converges to a standard normal distribution and therefore the power of the right-tailed test of level α can be approximated by

$$\Phi\left(\frac{\sqrt{n}}{\sigma_A}\left[\frac{\rho(\mu,T) - 1/2}{\rho_\mu'(\mu,T)}\right]\sqrt{\frac{\rho(\mu,T)[1 - \rho(\mu,T)]}{1 - \rho(\mu,T) + \nu\rho(\mu,T)}} - z_{1-\alpha}\right), \quad \mu > 0. \quad (4.6)$$

Clearly, for every $x, T, \nu > 0$, approximated power (4.6) tends to one as n grows. Moreover, condition (3.6) still guarantees that function (4.6) tends to one as μ grows.

The following Theorem provides the condition for the adopted target to preserve the monotonicity in μ of the approximated power function of test Z_n for heteroscedastic responses.

Theorem 4.2. *A target ρ induces a monotonically increasing power of test Z_n if and only if, for every $x, T > 0$,*

$$1 - \frac{[\rho(x,T) - 1/2]^2}{\rho(x,T)[1 - \rho(x,T)]} - \frac{\rho''_{xx}(x,T)[\rho(x,T) - 1/2]}{[\rho'_x(x,T)]^2}$$
$$\geq \frac{[\rho(x,T) - 1/2](\nu - 1)}{2[1 - \rho(x,T) + \nu\rho(x,T)]}. \tag{4.7}$$

Proof. See Appendix A.4. □

For instance, targets ρ_R and ρ_S satisfy condition (4.7) via straightforward algebra.

The following result provides sufficient conditions on a target for inducing a monotonically increasing approximated power of test Z_n.

Corollary 4.1. *Given a target ρ, the approximated power of test Z_n is increasing in μ if, for $\nu \leq 1$, condition (3.8) of Corollary 3.1 holds; whereas, when $\nu > 1$, a sufficient condition is*

$$\frac{1/2 - \rho(x,T)}{\rho(x,T)[1 - \rho(x,T)]} - \frac{\rho''_{xx}(x,T)}{[\rho'_x(x,T)]^2} \geq -3/2, \quad \text{for every } x > 0. \tag{4.8}$$

Proof. See Appendix A.5. □

Thus, regardless of ν, (4.8) guarantees that the approximated power of test Z_n is monotonically increasing in μ and therefore conditions (3.9) and (3.10) of Corollary 3.1 are still sufficient. Thus, from Example 3.1, adopting ρ_E, ρ_L and ρ_S the power of test Z_n is increasing in μ.

Remark 4.1. Given a target ρ, from (4.2) and (4.6) it is straightforward to show that test Z_n is uniformly more powerful than Wald test W_n if $\rho(x,T) - 1/2 \geq x\rho'_x(x,T)$ for any $x, T > 0$, i.e., the same condition in the case of homoscedastic outcomes (see Theorem 3.1).

The next result compares the power of test Z_n ensuing from the adoption of an asymptotically best RA design converging to a given target with those of the classical Wald test under the Neyman allocation.

Corollary 4.2. *Test Z_n ensuing from the adoption of an asymptotically best RA design converging to a target ρ chosen such that, for every $x, T > 0$,*

$$\frac{[\rho(x,T) - 1/2]\sqrt{\rho(x,T)[1 - \rho(x,T)]}}{\rho'_x(x,T)} \geq x\frac{\sqrt{1 - \rho(x,T) + \nu\rho(x,T)}}{1 + \sqrt{\nu}}, \quad (4.9)$$

is uniformly more powerful than W_n under the Neyman allocation. Thus, regardless of ν, a sufficient condition for (4.9) is

$$[\rho(x,T) - 1/2]\sqrt{1 - \rho(x,T)} \geq x\rho'_x(x,T), \quad \forall x, T > 0. \quad (4.10)$$

Proof. Adopting the Neyman allocation, from (4.2) the approximated power function of Wald test becomes $\Phi\left(\sigma_A^{-1}\sqrt{n}\mu(1 + \sqrt{\nu})^{-1} - z_{1-\alpha}\right)$, for $\mu > 0$. Therefore, condition (4.9) follows directly from (4.6), while (4.10) can be easily derived by observing that $\sqrt{\rho(x,T)}(1 + \sqrt{\nu}) \geq \sqrt{1 - \rho(x,T) + \nu\rho(x,T)}$ for every $x, T, \nu > 0$. □

Conditions (4.9) and (4.10) are quite restrictive: within the class of targets previously defined, only ρ_S in (2.10) satisfies (4.10) and therefore guarantees the superiority of test Z_n with respect to the Wald's one combined with Neyman allocation, regardless of the values of σ_A^2 and σ_B^2.

Finally, the next result extends Theorem 3.3 to the case of heteroscedastic responses, showing how test Z_n could overcome the usual trade-off between ethical demands and inferential precision.

Corollary 4.3. *If the target $\rho(x,T)$ is chosen such that, for any $x, T, \nu > 0$,*

$$2[1 - \rho(x,T) + \nu\rho(x,T)]_\rho G'_T(x,T) + {}_\rho G(x,T)\rho'_T(x,T)(1 - \nu) \leq 0, \quad (4.11)$$

then the approximated power of test Z_n is decreasing in T. Moreover, for any $\nu > 0$, the increasingness in T of function (3.16) is sufficient to guarantee the decreasingness in T of the power.

Proof. See Appendix A.6. □

Therefore, from Example 3.4, adopting ρ_R, ρ_S and ρ_B the power of test Z_n is still decreasing in T even under heteroscedasticity.

5. Discussion

In general, the ethical issue of assigning more subjects to the better treatment and the aim of improving the care of future patients by maximizing the inferential precision are two conflicting demands. In this paper, we provided a novel statistical test that allows one – under mild conditions on the adopted target as well as the chosen RA rule – to have a simultaneous improvement regarding both the ethical demands and the power of the test. Under this design strategy, that combines an asymptotically best RA procedure, the proposed test and a suitably chosen target allocation, the compromise among ethics and inference translates into a trade-off between the ethical skew of the target and a possible slightly inflated type-I error. This is particularly true when the sample size is small and the chosen target $\rho(\mu)$ is characterized by a strong ethical im-

provement, since in this case the treatment allocation proportion π_n tends to be slightly unstable as an estimator of $\rho(\mu)$. However, this does not hold for targets ρ_L and ρ_N coupled with $T = 1$ which, under our strategy, are able to control type-I errors.

Moreover, as further simulations omitted here for brevity have showed, for sample sizes lower than 200 combined with a target having a high ethical component and/or small values of T, an inflation of type-I errors could be present for both the suggested test and the classical Wald's one. In order to overcome this drawback, in the case of low sample sizes the ethical component of the target should be quite low, while targets having a strong ethical skew – under which Wald test is not suitable, since its power is not monotonically increasing [6] – require adequately large samples.

In particular, we showed both analytically and via simulations that the choice of target ρ_L with T close to 1 not only preserves nominal type-I errors, but also guarantees a gain in terms of both power and ethics. For instance, under the homoscedastic normal model, if compared with the classical Wald test, our design strategy with ρ_L ($T = 1$) guarantees a gain of power of Z_n around 5% and assigns 5% more patients to the better treatment when $\mu = 0.2$.

Whereas taking into account ρ_E (with $T = 1$), a slightly inflated type-I error is matched with a gain of power around 8% for $\mu = 0.2$ and 9% more patients assigned to the better treatment. In contrast, despite Wald test preserves in general nominal type-I errors, adopting the exponential target it induces an anomalous behaviour of the power that is locally decreasing and vanishes as the difference between the treatment effects grows [6].

Further research beyond the work in the present paper is needed. For instance, a possible way to overcome the instability induced by the discontinuous allocation function of ERADE could be the employment of the Doubly-Adaptive Biased Coin Design [13] with extremely high values of the randomization parameter (indeed, as the deterministic component in the allocation increases, this design tends to be asymptotically best). Moreover, another open topic is how to generalize our design strategy to multi-treatment clinical trials with more complex statistical models. Although the case of several treatments could seem a straightforward extension, the definition and the properties of the RA designs as well as those of the vectorial target functions are not yet available. For instance, ERADE has still not been generalized to the case of several treatments and therefore, up to now, there is not an RA procedure able to converge to any desired target allocation attaining its minimum asymptotic variance.

Acknowledgements

Appendix A: Proofs

A.1. Proof of Corollary 3.1

Given a target ρ, let

$$_\rho H(x) = \frac{1/2 - \rho(x)}{\rho(x)[1 - \rho(x)]} - \frac{\rho''_{xx}(x)}{[\rho'_x(x)]^2}, \quad \text{for } x > 0, \tag{A.1}$$

then

$$_\rho H(x) = \frac{1}{\rho'_x(x)} \left[_\rho h(x) \frac{[1/2 - \rho(x)]}{\rho(x)} - \frac{\rho''_{xx}(x)}{\rho'_x(x)} \right], \quad \text{for } x > 0. \tag{A.2}$$

From equation (3.7), easy algebraic manipulations lead to

$$_\rho H(x) \geq -\frac{1}{\rho(x) - 1/2}, \tag{A.3}$$

where the right-hand side in (A.3) is monotonically increasing in x, tends to negative infinity for $x \to 0$ and reaches -2 as $x \to \infty$. Then, $_\rho H(x) \geq -2$ is sufficient to induce an increasing power function for test Z_n.

Moreover, if $_\rho h'_x(x) \leq 0$ for every $x > 0$, then $[\rho'_x(x)]^2 \leq -[1 - \rho(x)]\rho''_{xx}(x)$. Thus, from equation (A.2), for every $x > 0$

$$_\rho h(x) \frac{[1/2 - \rho(x)]}{\rho(x)} - \frac{\rho''_{xx}(x)}{\rho'_x(x)} \geq {}_\rho h(x) \frac{[1/2 - \rho(x)]}{\rho(x)} + {}_\rho h(x) \geq 0$$

and therefore $_\rho H(x) \geq 0$ for every $x > 0$.

Whereas, if the hazard is increasing, then $\rho'_x(x) \geq 0$ and $[\rho'_x(x)]^2 \geq -[1 - \rho(x)]\rho''_{xx}(x)$ for every $x > 0$. From the first derivative of the hazard we can write $\rho''_{xx}(x) = {}_\rho h'_x(x)[1 - \rho(x)] - {}_\rho h(x)\rho'_x(x)$ and therefore, from equation (A.2), $_\rho h^2(x) \geq 2 {}_\rho h'_x(x)\rho(x)$ for every $x > 0$ implies that $_\rho H(x) \geq 0$ for every $x > 0$.

A.2. Proof of Theorem 3.3

Proof. Starting from equation (3.5), note that

$$_\rho G'_T(x, T) = \frac{\rho'_T(x, T)\sqrt{\rho(x, T)[1 - \rho(x, T)]}}{\rho'_x(x, T)} - \frac{\rho'_T(x, T)[\rho(x, T) - 1/2]^2}{\rho'_x(x, T)\sqrt{\rho(x, T)[1 - \rho(x, T)]}}$$
$$- \frac{\rho''_{xT}[\rho(x, T) - 1/2]\sqrt{\rho(x, T)[1 - \rho(x, T)]}}{[\rho'_x(x, T)]^2}.$$

and thus the power of test Z_n is decreasing in T if and only if, for any $x, T > 0$,

$$\frac{\rho'_x(x, T)\rho'_T(x, T)}{1 - \rho(x, T)} \left[\frac{1 - \rho(x, T)}{\rho(x, T) - 1/2} - \frac{\rho(x, T) - 1/2}{\rho(x, T)} \right] \leq \rho''_{xT}(x, T). \tag{A.4}$$

Since

$$\rho h'_T(x,T) = \frac{\rho''_{xT}(x,T)}{1 - \rho(x,T)} + \frac{\rho'_x(x,T)\rho'_T(x,T)}{[1 - \rho(x,T)]^2},$$

then (A.4) can be written as

$$\frac{1 - \rho(x,T)}{\rho(x,T) - 1/2} + \frac{1}{2\rho(x,T)} \geq \frac{1 - \rho(x,T)}{\rho'_T(x,T)} \frac{\rho h'_T(x,T)}{\rho h(x,T)}$$

and therefore (3.15) follows directly. $\qquad\square$

A.3. Proof of Theorem 4.1

Proof. Starting from equation (4.3), after straightforward calculation it can be shown that $\rho\tilde{G}'_x(x,T) \geq 0$ if and only if, for $T > 0$,

$$\rho(x,T)[1 - \rho(x,T)][1 - \rho(x,T) + \nu\rho(x,T)] \geq$$
$$x\rho'_x(x,T) \left[\frac{[\rho(x,T)]^2}{2}(\nu - 1) + \rho(x,T) - \frac{1}{2} \right], \quad \forall x > 0, \tag{A.5}$$

which directly coincides with (2.8) in the case of homoscedasticity (i.e., for $\nu = 1$). After some manipulations, condition (A.5) can be rewritten as

$$\rho(x,T)[1 - \rho(x,T)]^2 + \frac{x\rho'_x(x,T)[1 - \rho(x,T)]^2}{2} \geq$$
$$\nu[\rho(x,T)]^2 \left\{ \frac{x\rho'_x(x,T)}{2} - [1 - \rho(x,T)] \right\}, \quad \forall x > 0, \tag{A.6}$$

which is always satisfied if its RHS is non-positive, namely when

$$1 - \rho(x,T) \geq \frac{x\rho'_x(x,T)}{2} \qquad \forall x > 0.$$

On the other hand, observe that if $2[1 - \rho(x,T)] < x\rho'_x(x,T)$ for some $x \in \mathbb{R}^+$, then there always exists a sufficiently large value of ν that makes (A.6) false. $\quad\square$

A.4. Proof of Theorem 4.2

Proof. For any given target ρ, let now $\rho A(x,T) = \sqrt{\rho(x,T)[1 - \rho(x,T)]}$, $\rho B(x, T,\nu) = \sqrt{1 - \rho(x,T) + \nu\rho(x,T)}$ and

$$\rho\mathcal{G}(x,T) = \frac{[\rho(x,T) - 1/2]\sqrt{\rho(x,T)[1 - \rho(x)]}}{\rho'_x(x,T)\sqrt{1 - \rho(x,T) + \nu\rho(x,T)}}, \qquad x,T > 0, \tag{A.7}$$

then,

$$\rho\mathcal{G}(x,T) = \frac{\rho G(x,T)}{\rho B(x,T,\nu)}, \qquad x,T > 0. \tag{A.8}$$

Therefore,

$$_\rho\mathcal{G}'_x(x,T) = [\rho'_x(x,T)]^2\,_\rho B(x,T,\nu)\frac{_\rho A^2(x,T) - [\rho(x,T) - 1/2]^2}{_\rho A(x,T)}$$
$$- [\rho(x,T) - 1/2]\,_\rho A(x,T)\frac{2\rho''_{xx}(x,T)\,_\rho B^2(x,T,\nu) + [\rho'_x(x,T)]^2(\nu - 1)}{2\,_\rho B(x,T,\nu)},$$

so that the power function is monotonically increasing if and only if, for any $x > 0$,

$$[\rho'_x(x,T)]^2 - \frac{[\rho'_x(x,T)]^2[\rho(x,T) - 1/2]^2}{_\rho A^2(x,T)} \geq$$
$$\frac{[\rho'_x(x,T)]^2[\rho(x,T) - 1/2](\nu - 1)}{2\,_\rho B^2(x,T,\nu)} + \rho''_{xx}(x)[\rho(x,T) - 1/2].$$

Thus, the thesis follows by means of simple algebraic manipulations. □

A.5. Proof of Corollary 4.1

Proof. Recalling that, from (A.1),

$$_\rho H(x,T) = -\frac{\rho(x,T) - 1/2}{\rho(x,T)[1 - \rho(x,T)]} - \frac{\rho''_{xx}(x,T)}{[\rho'_x(x,T)]^2},$$

straightforward algebraic manipulations on condition (4.7) lead to

$$_\rho H(x,T) \geq -\frac{1}{\rho(x,T) - 1/2} + \frac{\nu - 1}{2[1 - \rho(x,T) + \nu\rho(x,T)]}, \quad \forall x > 0. \quad (A.9)$$

Notice that the RHS of (A.9) is non-positive for $\nu \leq 1$ and therefore (A.9) is satisfied if $H(x,T) \geq -2$. Whereas when $\nu > 1$, if we let

$$f(\nu) = \frac{\nu - 1}{2[1 - \rho(x,T) + \nu\rho(x,T)]},$$

then $f : \mathbb{R}^+ \to \mathbb{R}^+$ is increasing in ν with $\lim_{\nu \to \infty} f(\nu) = [2\rho(x,T)]^{-1}$; then the RHS of (A.9) is negative since

$$-\frac{1}{\rho(x,T) - 1/2} + \frac{\nu - 1}{2[1 - \rho(x,T) + \nu\rho(x,T))]} \leq \frac{-1 - 2\rho(x,T)}{2\rho(x,T)[2\rho(x,T) - 1]} \leq -\frac{3}{2}$$

because $5\rho(x,T)[1 - \rho(x,T)] \geq 0 \geq [\rho(x,T) + 1][\rho(x,T) - 1]$. Thus, for $\nu > 1$, condition (A.9) is satisfied if $H(x,T) \geq -3/2$. □

A.6. Proof of Corollary 4.3

Proof. Starting from (A.8),

$$_\rho\mathcal{G}'_T(x,T) = \frac{_\rho G'_T(x,T)}{\sqrt{1-\rho(x,T)+\nu\rho(x,T)}} + \frac{_\rho G(x,T)\rho'_T(x,T)(1-\nu)}{2[1-\rho(x,T)+\nu\rho(x,T)]^{3/2}}$$

and therefore the power is decreasing in T if and only if (4.11) holds. Moreover, when $\nu \leq 1$ condition (3.15) is sufficient because $_\rho G'_T$ and ρ'_T are both non-positive. Whereas for $\nu > 1$, (4.11) could be rewritten as

$$\frac{_\rho G'_T(x,T)}{\rho'_T(x,T)} \geq {}_\rho G(x,T)\frac{\nu-1}{2[1-\rho(x,T)+\nu\rho(x,T)]}, \tag{A.10}$$

where $(\nu-1)/[1-\rho(x,T)+\nu\rho(x,T)] \leq \rho(x,T)^{-1}$ for any $\nu > 0$. Therefore,

$$\frac{_\rho G'_T(x,T)}{\rho'_T(x,T)} \geq \frac{_\rho G(x,T)}{2\rho(x,T)}, \quad \text{for every } x,T,\nu > 0, \tag{A.11}$$

i.e.,

$$2\frac{_\rho G'_T(x,T)}{_\rho G(x,T)} \leq \frac{\rho'_T(x,T)}{\rho(x,T)}, \quad \text{for every } x,T,\nu > 0, \tag{A.12}$$

is sufficient for (A.10). Thus, due to the positiveness of $_\rho G$ and ρ, condition (A.12) is satisfied if the function $\rho(x,T)/_\rho G^2(x,T)$ is increasing in T, where

$$\frac{\rho(x,T)}{_\rho G^2(x,T)} = \frac{_\rho h(x,T)\rho'_x(x,T)}{[\rho(x,T)-1/2]^2} = \left(\frac{_\rho h(x,T)}{\rho(x,T)-1/2}\right)^2[1-\rho(x,T)],$$

that is clearly satisfied when function (3.16) is increasing in T. $\qquad\square$

References

[1] Atkinson, A. C. and A. Biswas (2005). Bayesian adaptive biased-coin designs for clinical trials with normal responses. *Biometrics 61*, 118–125. MR2135851

[2] Atkinson, A. C. and A. Biswas (2014). *Randomised Response-Adaptive Designs in Clinical Trials*. Boca Raton: Chapman & Hall/CRC Press.

[3] Azriel, D., M. Mandel, and Y. Rinott (2012). Optimal allocation to maximize power of two-sample tests for binary response. *Biometrika 99*, 101–113. MR2899666

[4] Baldi Antognini, A. and A. Giovagnoli (2010). Compound optimal allocation for individual and collective ethics in binary clinical trials. *Biometrika 97*, 935–946. MR2746162

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

[6] Baldi Antognini, A., A. Vagheggini, and M. Zagoraiou (2016). Is the classical wald test always suitable under response-adaptive randomization? *Statistical Methods in Medical Research*, available online, DOI: 10.1177/0962280216680241. MR3825908

[7] Baldi Antognini, A. and M. Zagoraiou (2015). On the almost sure convergence of adaptive allocation procedures. *Bernoulli 21*, 881–908. MR3338650

[8] Bandyopadhyay, U. and A. Biswas (2001). Adaptive designs for normal responses with prognostic factors. *Biometrika 88*, 409–419. MR1844841

[9] CHMP (2007). Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design. Available on line.

[10] FDA (2010). Guidance for industry. Adaptive design clinical trials for drugs and biologics (draft document). Available on line.

[11] Hu, F. and W. F. Rosenberger (2003). Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparisons. *Journal of the American Statistical Association 98*, 671–678. MR2011680

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

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

[14] Hu, F., L.-X. Zhang, and X. He (2009). Efficient randomized adaptive designs. *The Annals of Statistics 37*, 2543–2560. MR2543702

[15] Ivanova, A., A. Biswas, and A. Lurie (2006). Response-adaptive designs for continuous outcomes. *Journal of Statistical Planning and Inference 136*, 1845–1852. MR2255600

[16] Lehmann, E. L. (1999). *Elements of large-sample theory*. Springer Verlag, New York. MR1663158

[17] Melfi, V. and C. Page (2000). Estimation after adaptive allocation. *Journal of Statistical Planning and Inference 29*, 353–363. MR1771124

[18] Rosenberger, W. F. and J. L. Lachin (2002). *Randomization in Clinical Trials: Theory and Practice*. John Wiley & Sons, New York. MR1914364

[19] Rosenberger, W. F., N. Stallard, A. Ivanova, C. N. Harper, and M. L. Ricks (2001). Optimal adaptive designs for binary response trials. *Biometrics 57*, 909–913. MR1863454

[20] Thall, P. F., P. Fox, and J. Wathen (2015). Some caveats for outcome adaptive randomization in clinical trials. In O. Sverdlov (Ed.), *Modern adaptive randomized clinical trials: statistical and practical aspects*, pp. 287–305. Oxford: Chapman & Hall/CRC Biostatistics. MR3676361

[21] Thall, P. F., P. Fox, and J. Wathen (2015). Statistical controversies in clinical research: scientific and ethical problems with adaptive randomization in comparative clinical trials. *Annals of Oncology 26*, 1621–1628.

[22] Tymofyeyev, Y., W. F. Rosenberger, and F. Hu (2007). Implementing optimal allocation in sequential binary response experiments. *Journal of the American Statistical Association 102*, 224–234. MR2345540

[23] Yi, Y. and X. Wang (2011). Comparison of Wald, score, and likelihood ratio tests for response adaptive designs. *Journal of Statistical Theory and Applications 10*, 553–569. MR2907394

[24] Zhang, L.-X. and W. F. Rosenberger (2006). Response-adaptive randomization for clinical trials with continuous outcomes. *Biometrics 62*, 562–569. MR2236838