Is the classical Wald test always suitable under response-adaptive randomization?

Page Proof Instructions and Queries

Journal Title: Statistical Methods in Medical Research (SMM)

Article Number: 680241

Greetings, and thank you for publishing with SAGE. We have prepared this page for your review. Please respond to each of the below queries by digitally marking this PDF using Adobe Reader (free at https://get.adobe.com/reader).

Please use *only* the circled tools to indicate your requests and responses, as edits via other tools/methods are not compatible with our software. To ask a question or request a formatting change (such as italics), please click the tool and then choose "Text Callout." To access the necessary tools, choose "Comment" from the right-side menu.

No.	Query
	Please confirm that all author information, including names, affiliations, sequence, and contact details, is correct.
	Please review the entire document for typographical errors, mathematical errors, and any other necessary corrections; check headings, tables, and figures.
	Please confirm that the Funding and Conflict of Interest statements are accurate.
	Please ensure that you have obtained and enclosed all necessary permissions for the reproduction of artistic works, (e.g. illustrations, photographs, charts, maps, other visual material, etc.) not owned by yourself. Please refer to your publishing agreement for further information.
	Please note that this proof represents your final opportunity to review your article prior to publication, so please do send all of your changes now.

Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
0(0) 1–18

DOI: 10.1177/0962280216680241
smm.sagepub.com

SAGE

Is the classical Wald test always suitable under response-adaptive randomization?

Alessandro Baldi Antognini,[1] Alessandro Vagheggini[1] and Maroussa Zagoraiou[2]

Abstract

The aim of this paper is to analyze the impact of response-adaptive randomization rules for normal response trials intended to test the superiority of one of two available treatments. Taking into account the classical Wald test, we show how response-adaptive methodology could induce a consistent loss of inferential precision. Then, we suggest a modified version of the Wald test which, by using the current allocation proportion to the treatments as a consistent estimator of the target, avoids some degenerate scenarios and so it should be preferable to the classical test. Furthermore, we show both analytically and via simulations how some target allocations may induce a locally decreasing power function. Thus, we derive the conditions on the target guaranteeing its monotonicity and we show how a correct choice of the initial sample size allows one to overcome this drawback regardless of the adopted target.

Keywords

Adaptive experiments, asymptotic tests, comparative clinical trials, ethics, power

1 Introduction

Adaptive experiments are sequential procedures where the decision about how to proceed next is made according to a pre-established rule that makes use of the information accrued along the way. Even if their use remains controversial due to some inferential problems that could arise,[1,2] adaptive designs are widely used in different experimental fields and they are nowadays considered as a panacea for ethical issues posed by randomized clinical trials. This is especially true for phase III trials, where patients are enrolled step-by-step and are assigned to one of two or more available treatments to be compared. In this context, randomization is regarded as a must and, when is combined with the adaptive nature of the experiment, it means that the treatments are assigned to the next unit by allocation probabilities that make use of the past information. However, the updating process cannot take place in a haphazard manner, which could undermine the validity and integrity of the ensuing statistical analysis. Thus, the design of these experiments requires special care and it is not surprising that statistical research on this topic has become very popular over the past two decades, also due to the strong encouragement from US Government agencies and health authorities.[3,4]

Due to the peculiarity of clinical context, often there are several competing goals related to the ethical demand of maximizing the subjects care and to the statistical aim of drawing correct inferential conclusions with high precision. By formalizing these goals into suitable optimization problems, several authors provided target allocations of the treatments that could represent a valid trade-off among ethics and inference.[5–12] In general, these targets depend on the unknown model parameters and they can be approached asymptotically by using suitable response-adaptive (RA) randomization procedures, such as the doubly adaptive biased coin design[13] and the efficient randomized-adaptive design (ERADE),[14] converging to them.

RA designs are a class of sequential allocation rules where the probabilities of treatment assignments change at each step on the basis of earlier responses and past allocations. Starting from an initial sample of observations on

[1]Department of Statistical Sciences, University of Bologna, Italy
[2]Department of Business Administration and Law, University of Calabria, Italy

Corresponding author:
Alessandro Baldi Antognini, Department of Statistical Sciences, University of Bologna, Via delle Belle Arti 41, 40126, Bologna, Italy.
Email: a.baldi@unibo.it

each treatment (usually based on restricted randomization) to derive a non-trivial estimation, at each step these designs estimate the unknown parameters as well as the target and then force the next allocation to converge to the target.

Under these procedures the resulting statistical analysis requires refined tools able to allow for the complex dependence structure, since (i) the assignments are a stochastic process, making the resulting responses dependent, and (ii) inference must be unconditional on the design, because the allocations are themselves informative on the parameters of the model.[15,16] Although the asymptotic properties of both (i) the usual maximum likelihood estimators (MLEs) and (ii) the allocation process are well-established, the large majority of the literature[9,13,17–25] is focused on the implications of the RA methodology in terms of estimation of the treatment effects, while little attention is devoted to hypotheses testing,[7,12,26–28] almost exclusively for binary data.

The aim of this paper is to analyze the impact of RA designs for hypothesis testing in the case of normally response trials for checking the superiority of one of two available treatments. Taking into account the classical Wald test, we first show how the RA methodology could induce an anomalous behavior of the power function. Then, we suggest a modified version of Wald test which, by using the current allocation proportion to the treatments as a consistent estimator of the target, avoids some degenerate scenarios and so it should be preferable than the classical test. Furthermore, we show both analytically and via simulations how some target allocations may induce an additional anomalous behavior of the power function, which could be locally decreasing. Thus, we derive the conditions on the target guaranteeing the monotonicity of the ensuing power, showing also how a correct choice of the initial sample size allows one to overcome this drawback regardless of the adopted target.

The paper is structured as follows. Starting from the notation and some preliminaries in Section 2, Sections 3 and 4 deal with the asymptotic power of the Wald-type Z-tests under RA randomization procedures, highlighting their drawbacks. Section 5 describes some practical implications via a simulation study, while Section 6 deals with some general conclusions about the applicability of RA randomization procedures for hypothesis testing.

2 Preliminaries

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 denote the allocation of the ith subject, with $\delta_i = 1$ if they is assigned to A and 0 otherwise, and let Y_i be the corresponding outcome that is assumed to be normally distributed with

$$E(Y_i) = \delta_i \mu_A + (1 - \delta_i)\mu_B \quad \text{and} \quad V(Y_i) = \sigma^2, \quad i \geq 1 \tag{1}$$

where μ_A and μ_B are the treatment effects and $\sigma^2 \in \mathbb{R}^+$ denotes the common variance. Assuming that the outcomes are conditionally independent given the treatment assignments, the usual goal consists of identifying the superiority of a given treatment, so that the inferential interest lies in estimating or testing $\mu = \mu_A - \mu_B$, while it is customary to regard μ_B (or the sum of the treatment effects) as a nuisance. Thus, from now on we take into account the problem of testing the hypothesis H_0: $\mu = 0$ versus H_1: $\mu > 0$ (the case H_1: $\mu \neq 0$ can be easily derived in an analogous way).

Several proposals have been made in the literature in order to derive suitable target allocations (ρ; $1 - \rho$) to A and B, respectively (either as finite sample allocations or as asymptotic proportions to be approximated in a large sample set-up) that achieve a good trade-off between ethical concerns and inferential precision. One of the main proposals consists in formalizing these objectives into a combined/constrained optimization problem and find the targets that are optimal with respect to the chosen approach (see Chapter 5 of Baldi Antognini and Giovagnoli[16] and the paper by Biswas and Bhattacharya[11] for a recent review). In general, the ensuing target depends on the unknown model parameters, i.e. $\rho = \rho(\mu)$, and assuming without loss of generality "the-larger-the-better" scenario (namely, treatment A is better than B if and only if $\mu_A > \mu_B$) it should satisfy the following conditions:

- T1 ρ: $\mathbb{R} \to (0; 1)$ is a symmetric function with $\rho(-x) = 1 - \rho(x)$, ensuring that both treatments are treated likewise;
- T2 $\rho(x)$ is increasing in x, meaning that any gain in terms of the relative superiority of a given treatment should skew the assignments by increasing its desirability;
- T3 $\rho(\cdot)$ is twice continuously differentiable with bounded derivatives.

Note the following.

- From T1, $\rho(x) \neq \{0; 1\}$ guarantees that the comparative experiments do not collapse into the observation of just one treatment; moreover, $\rho(0) = 1/2$ and therefore, due to the symmetric structure of $\rho(\cdot)$ around the point $(0; 1/2)$, we could simply model the target function for $x > 0$.
- Ethical requirement T2 ensures that the superior treatment should be favored and, combined with T1, guarantees that the desirability of either treatment is the same if and only if the two treatment arms equally perform.
- the target $\rho(\cdot)$ may depend on the nuisance parameter too and in this case conditions T1–T3 should be satisfied for any given value of the nuisance. For example, for μ_A, $\mu_B > 0$, Zhang and Rosenberger[29] suggested the target $\sqrt{\mu_A}/(\sqrt{\mu_A} + \sqrt{\mu_B})$, while Baldi Antognini and Giovagnoli[16] analyzed the target $\mu_A/(\mu_A + \mu_B)$. Clearly, these choices correspond to

$$\rho_Z(x) = \frac{1}{2} + \frac{\sqrt{x + \mu_B} - \sqrt{\mu_B}}{2(\sqrt{x + \mu_B} + \sqrt{\mu_B})}, \quad \text{for } x > 0 \tag{2}$$

and

$$\rho_R(x) = \frac{1}{2} + \frac{x}{2(2\mu_B + x)}, \quad \text{for } x > 0 \tag{3}$$

respectively, both satisfying T1–T3 for any $\mu_B \in \mathbb{R}^+$.

Although non-necessary from a mathematical perspective, an additional ethical requirement that is almost always satisfied by the targets suggested in the literature[9,11,12,23,29,30] is

T4 $\lim_{x \to \infty} \rho(x) = 1$, namely the target function has to approach 1 as A performs infinitely better than B (analogously, from T1, $\lim_{x \to -\infty} \rho(x) = 0$).

In such a case, the behavior of the target could be represented by the cumulative distribution function (cdf) of a continuous symmetric random variable centered at 0 with support $\mathbb{R}$, like e.g. the normal target[11,23,30]

$$\rho_N(x) = \Phi(x/T), \quad x \in \mathbb{R} \tag{4}$$

where Φ denotes the cdf of the standard normal, the Cauchy target

$$\rho_C(x) = \frac{1}{2} + \frac{\arctan(x/T)}{\pi}, \quad x \in \mathbb{R} \tag{5}$$

or the logistic one

$$\rho_L(x) = \left(1 + e^{-x/T}\right)^{-1}, \quad x \in \mathbb{R} \tag{6}$$

The tuning parameter $T > 0$ manages the ethical component of the target: small values of T tend to skew the subjects' assignment to the superior treatment, while as T grows the ethical component vanishes and the target tends to the balanced one. In this setting, it is natural to regard $\rho'(\cdot)$ as the connected pdf, where, from T3, this density should be uniformly continuous, so that $\lim_{x \to \infty} \rho'(x) = 0$ and also $\lim_{x \to \infty} x\rho'(x) = 0$ to ensure integrability.

Alternatively, by using the symmetric property T1, $\rho(x)$ could be modeled for $x \in \mathbb{R}^+$ as a suitably re-scaled cdf of a positive random variable, like e.g. the exponential target

$$\rho_E(x) = \begin{cases} 1 - \dfrac{e^{-x/T}}{2}, & \text{if } x \geq 0, \\ \dfrac{e^{x/T}}{2}, & \text{if } x < 0. \end{cases} \tag{7}$$

Figure 1 shows the different behavior of all of the above-mentioned targets (where, for simplicity, we set $T = \mu_B = 1$).

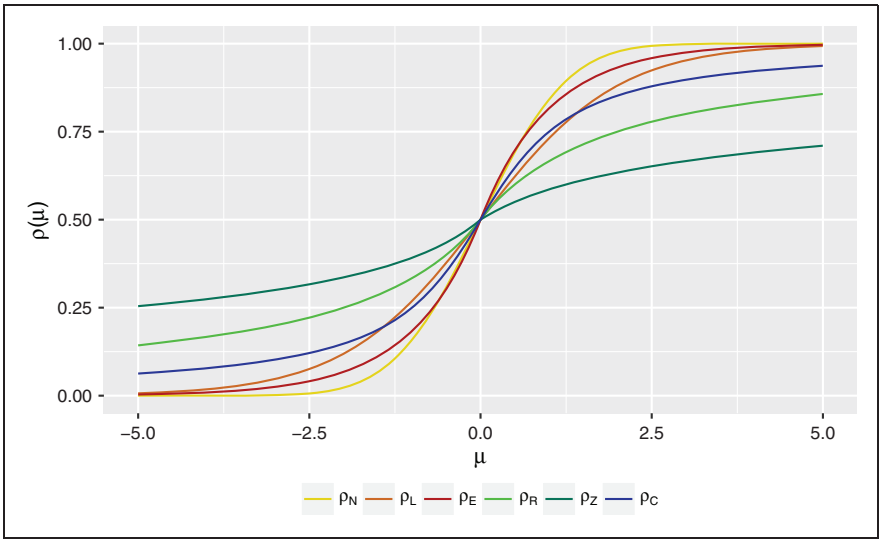

Figure 1. Target functions ρ_N, ρ_C, ρ_L and ρ_E (with $T = 1$), ρ_R and ρ_Z (with $\mu_B = 1$).

Remark 1. *Every above-mentioned target (namely, $\rho_N(\cdot)$, $\rho_C(\cdot)$, $\rho_L(\cdot)$, $\rho_Z(\cdot)$, $\rho_R(\cdot)$ and $\rho_E(\cdot)$) satisfies T4. However, this condition could be relaxed by assuming a re-scaled target function $\tilde{\rho}(\cdot)$ such that $\lim_{x\to\infty}\tilde{\rho}(x) = c \in (1/2; 1]$; clearly, it does not induce substantial implications in practice, since $\tilde{\rho}(\cdot)$ is univocally associated with a cdf $\rho(\cdot)$ satisfying T1–T4, given by $\rho(x) = [\tilde{\rho}(x) - (1-c)]/(2c-1)$.*

3 The Wald-type *Z*-tests under RA randomization procedures

3.1 RA designs and asymptotic inference

Several RA designs have been suggested in the literature with the aim of converging to a desired target $\rho(\mu)$ depending on the unknown model parameters. After the starting sample of n_0 observations assigned to each treatment, at each step $n > 2n_0$ these designs estimate the difference μ between the treatment effects by $\hat{\mu}_n = \hat{\mu}_{An} - \hat{\mu}_{Bn}$, where $(\hat{\mu}_{An}, \hat{\mu}_{Bn})$ are the MLEs of (μ_A, μ_B), i.e. the sample means. Thus, the target is estimated by $\hat{\rho}_n = \rho(\hat{\mu}_n)$ and then the next allocation is forced to gradually approach the target.

An example is the ERADE[14] defined by

$$\Pr(\delta_{n+1} = 1 \mid \delta_1, \ldots, \delta_n; Y_1, \ldots, Y_n) = \begin{cases} \gamma\rho(\hat{\mu}_n), & \text{if } \pi_n > \rho(\hat{\mu}_n) \\ \rho(\hat{\mu}_n), & \text{if } \pi_n = \rho(\hat{\mu}_n), \\ 1 - \gamma[1 - \rho(\hat{\mu}_n)], & \text{if } \pi_n < \rho(\hat{\mu}_n) \end{cases}$$

where $\pi_n = n^{-1}\sum_{i=1}^{n}\delta_i$ and $1-\pi_n$ denote the allocation proportions to A and B, respectively, and $\gamma \in [0;1)$ is a randomization parameter. As shown by Hu et al.,[14] the ERADE asymptotically approaches any chosen target $\rho(\mu)$ satisfying T1–T3, namely it guarantees that $\lim_{n\to\infty}\pi_n = \rho(\mu)$ almost surely.

In general, even if the MLEs coincide with those of the non-sequential setting, their distribution under RA designs is not the same as when the observations are independent and identically distributed (i.i.d.), due to the dependence structure induced by the adaptation process. However, given a target $\rho(\mu)$ satisfying T1–T3, consistency and asymptotic normality of the MLEs are ensured provided that the RA design is chosen such that $\lim_{n\to\infty}\pi_n = \rho(\mu)$ almost surely.[16] Indeed, as n tends to infinity $(\hat{\mu}_{An}, \hat{\mu}_{Bn}) \to (\mu_A, \mu_B)$ almost surely and

$$\sqrt{n}\left[\begin{pmatrix}\hat{\mu}_{An}\\ \hat{\mu}_{Bn}\end{pmatrix} - \begin{pmatrix}\mu_A\\ \mu_B\end{pmatrix}\right] \hookrightarrow N\left(\begin{pmatrix}0\\0\end{pmatrix}; \begin{pmatrix}\dfrac{\sigma^2}{\rho(\mu)} & 0\\ 0 & \dfrac{\sigma^2}{1-\rho(\mu)}\end{pmatrix}\right).$$

Thus, assuming that the common variance σ^2 is a priori known,

$$\sqrt{n}(\hat{\mu}_n - \mu) \hookrightarrow N\left(0; \frac{\sigma^2}{\rho(\mu)[1-\rho(\mu)]}\right) \tag{8}$$

and, recalling that $\rho(\hat{\mu}_n)$ is a consistent estimator of $\rho(\mu)$ due to the continuity of the target function, the classical Wald test statistic is (see, for instance, Yi and Wang)[27]

$$W_n = \sqrt{\frac{n\rho(\hat{\mu}_n)[1-\rho(\hat{\mu}_n)]}{\sigma^2}}\hat{\mu}_n \tag{9}$$

As is well-known, under H_0 the statistic W_n converges asymptotically to a standard normal distribution (or, alternatively, W_n^2 follows a chi-squared distribution with one degree of freedom (dof)) and thus the asymptotic test rejects H_0 if

$$\hat{\mu} > z_{1-\alpha}\sqrt{\frac{\sigma^2}{n\rho(\hat{\mu}_n)[1-\rho(\hat{\mu}_n)]}} \tag{10}$$

where z_α is the α-percentile of Φ. Therefore, the (asymptotic) power of the right-sided Z-test of level α is

$$\Phi\left(\mu\sqrt{\frac{n\rho(\mu)[1-\rho(\mu)]}{\sigma^2}} - z_{1-\alpha}\sqrt{\frac{\rho(\mu)[1-\rho(\mu)]}{\rho(\hat{\mu}_n)[1-\rho(\hat{\mu}_n)]}}\right), \quad \mu > 0$$

which can be approximated by

$$\Phi\left(\mu\sqrt{\frac{n\rho(\mu)[1-\rho(\mu)]}{\sigma^2}} - z_{1-\alpha}\right), \quad \mu > 0 \tag{11}$$

since asymptotically $\rho(\hat{\mu}_n) \approx \rho(\mu)$.

Under the same hypotheses, we suggest an alternative version of the Wald test which can be constructed by using π_n, instead of $\rho(\hat{\mu}_n)$, as a consistent estimator of $\rho(\mu)$, namely by replacing W_n in (9) with

$$\tilde{W}_n = \sqrt{\frac{n\pi_n(1-\pi_n)}{\sigma^2}}\hat{\mu}_n \tag{12}$$

and rejecting H_0 if

$$\hat{\mu} > z_{1-\alpha}\sqrt{\frac{\sigma^2}{n\pi_n(1-\pi_n)}} \tag{13}$$

A formal derivation of this test can be obtained from the following version of the central limit theorem (CLT)[20]

$$\begin{pmatrix} \sqrt{N_{An}}(\hat{\mu}_{An} - \mu_A) \\ \sqrt{N_{Bn}}(\hat{\mu}_{Bn} - \mu_B) \end{pmatrix} \hookrightarrow N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}\right), \quad \text{as } n \to \infty,$$

where $N_{An} = n\pi_n$ and $N_{Bn} = n(1-\pi_n)$ denote the number of allocations to the two treatments after n steps. Indeed, as n grows,

$$\sqrt{n\pi_n(1-\pi_n)}(\hat{\mu}_n - \mu) \hookrightarrow N\left(0; \sigma^2\right) \tag{14}$$

and therefore the rejection region (13), as well as power function (11), follows directly recalling that $\pi_n \to \rho(\mu)$ almost surely as n tends to infinity.

Remark 2. *When the common variance σ^2 is unknown, it can be estimated at each step n by the usual pooled sample variance s^2_{pn}, which is a consistent estimator for σ^2 under RA designs satisfying $\lim_{n\to\infty}\pi_n = \rho(\mu)$ almost surely. Thus, from the CLTs in (8) and (14), as n tends to infinity,*

$$\sqrt{\frac{n\rho(\hat{\mu}_n)[1-\rho(\hat{\mu}_n)]}{s^2_{pn}}}(\hat{\mu}_n - \mu) \quad \text{and} \quad \sqrt{\frac{n\pi_n(1-\pi_n)}{s^2_{pn}}}(\hat{\mu}_n - \mu)$$

converge in distribution to a standard normal random variable. Then, both tests $\boldsymbol{W_n}$ and $\tilde{W}_n$ can be adjusted by substituting σ^2 with s^2_{pn} and therefore the corresponding power functions can still be approximated by (11), since in a large sample set-up $s^2_{pn} \approx \sigma^2$.

3.2 Asymptotic approximation, target function and starting sample size

The performances of the Wald tests described above, as well as the quality of the CLT approximation of the power function in (11), are strictly related to the chosen target. Test $\boldsymbol{W_n}$ is based on the asymptotic approximation $\rho(\hat{\mu}_n) \approx \rho(\mu)$, which, from T4, could admit degenerate situations where $\rho(\hat{\mu}_n) \approx 0$ or 1. Indeed, if the desired target $\rho(\cdot)$ has a strong ethical component (namely it satisfies T4 and rapidly increases as μ grows), when the difference between the treatment effects is remarkable, so is $\hat{\mu}_n$ and thus $\rho(\hat{\mu}_n)$ tends to one; consequently, the right-hand side of (10) grows to infinity and therefore H_0 tends to be always accepted (i.e. the power goes to zero).

As an example, Figure 2 shows the behavior of the Wald test $\boldsymbol{W_n}$ in a simulated trial where the chosen target functions are ρ_N in (4), ρ_L in (6) and ρ_E in (7) (colors are yellow, orange and red for $T = 0.5$, 1 and 2, respectively) and ρ_R in (3) (with $\mu_B = 1$), where the ERADE is employed with $\gamma = 0.5$. The results come from 5000 simulations with sample sizes $n = 75$, 150, 250 and starting sample of $n_0 = 2$ observations on each treatment, where the responses are generated following a Gaussian distribution with $\sigma^2 = 1$, $\mu_B = 1$ and $\mu_A = \mu_B + k$, where $k \geq 0$. Taking into account targets ρ_N, ρ_L and ρ_E, the ensuing power tends to be quite poor, it is not monotonically increasing in μ and tends to zero as μ grows (although under ρ_L and ρ_E the non-monotonicity is not noticeable in the plots for $T = 2$, the power function is still decreasing, but this behavior is present for larger values of μ). Moreover, this anomalous behavior is accentuated as the ethical component of the target grows (i.e. small values of T), since in such a case the ethical skew tends to assign all subjects to the superior treatment also for small values of μ. Therefore, the consistency and asymptotic normality of the MLEs are strongly compromised, as well as the quality of the approximation of power (11). This is particularly true for small sample sizes, where also the type I errors become slightly inflated, as shown in Table 1. This is due to the fact that, given the RA nature of the procedure, when the sample size is small and the chosen target is characterized by a strong ethical impact, then π_n tends to be slightly more unstable, as an estimator of $\rho(\mu)$, than $\rho(\hat{\mu}_n)$ as μ tends to zero and therefore the type I errors for $\tilde{W}_n$ are more inflated than those for W_n. This becomes more evident for ρ_N, since it has the highest ethical impact if compared with the other considered targets (see Figure 1), especially for $T = 0.5$.

While the choice of ρ_R (which satisfies T4, but with a lower ethical improvement with respect to ρ_N, ρ_L and ρ_E) always guarantees a suitable behavior of the power of the test, that goes to one as μ grows, also preserving a correct type I error (see Table 1).

As it can be easily seen from the power function in (11), a crucial condition for the applicability of Wald test W_n is that the chosen target ρ should satisfy $\lim_{x\to\infty} x^2[1-\rho(x)] = \infty$. This condition characterizes the ethical improvement of the target and prescribes that $1-\rho$ should tend to zero more slowly than x^{-2}, in order to avoid the degenerate scenarios discussed previously. For instance, as also shown in Figure 2, adopting ρ_R then $\lim_{x\to\infty} x^2[1-\rho_R(x)] = \infty$ (which holds for ρ_C and ρ_Z too), while under ρ_N (and also for ρ_E and ρ_L) this limit goes to zero.

Now taking into account $\tilde{W}_n$, it is based on the asymptotic approximation $\pi_n \approx \rho(\mu)$ which could also admit the extreme scenarios $\pi_n \approx 0$ or 1 due to T4. However, these two cases could be verified only for $n_0 = 0$, while in practical applications for any fixed sample size n

$$\tau_n = \frac{n_0}{n} \leq \pi_n \leq \frac{n-n_0}{n} = 1 - \tau_n \tag{15}$$

where

$$\tau_n \in \left[\frac{1}{n}; \frac{1}{2} - \frac{1}{n}\right], \quad \text{for } n \text{ even}$$

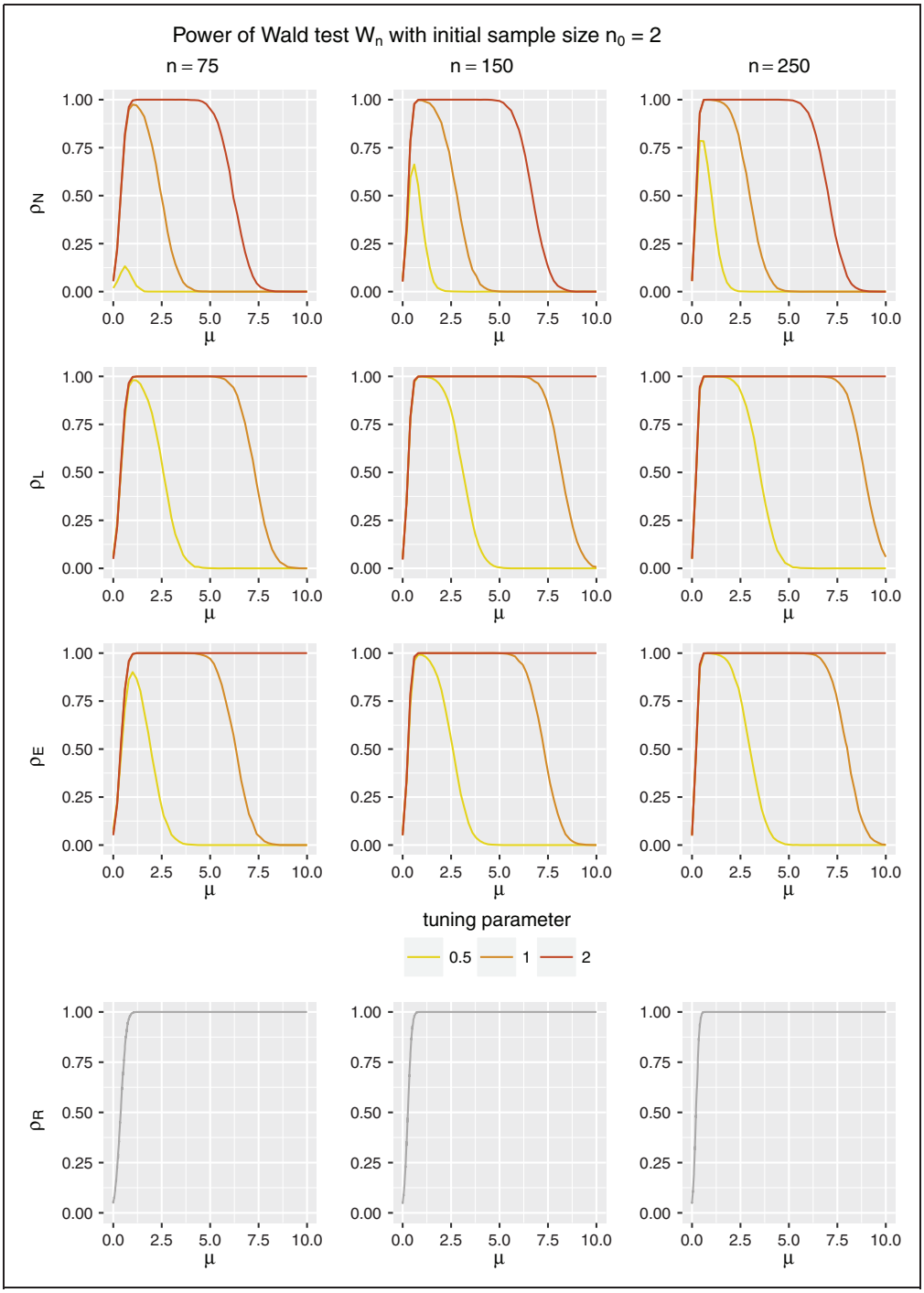

Figure 2. Power of the Wald test W_n under ρ_N, ρ_L and ρ_E ($T = 0.5$, 1 and 2) and ρ_R ($\mu_B = 1$) with $n = 75$, 150, 250 and starting sample size $n_0 = 2$.

due to the starting sample of $n_0 \geq 1$ allocations made on each treatment (for n odd, $\tau_n \leq (n-1)/2$). Note that the cases $\tau_n = 0$ or $1/2$ should be excluded. Indeed, $\tau_n = 1/2$ corresponds to assume no adaptivity of the design (i.e. every allocation is made by restricted randomization). Whereas $\tau_n = 0$ implies that no starting sample has been taken into account, namely the RA procedure starts with an initial guess (μ_{A0}, μ_{B0}) possibly derived from earlier trials, which ensures the applicability of the RA methodology even when no patients are assigned to the treatments (potentially also modifying the usual sample means accordingly, as discussed by Hu et al.).[14]

Table 1. Type I errors of the tests W_n and $\tilde{W}_n$ for targets ρ_N, ρ_L and ρ_E (with $T = 0.5$, 1 and 2), and ρ_R (with $\mu_B = 1$).

		$n = 75$		$n = 150$		$n = 250$	
ρ	T	W_n	$\tilde{W}_n$	W_n	$\tilde{W}_n$	W_n	$\tilde{W}_n$
ρ_N	0.5	0.02	0.12	0.07	0.11	0.06	0.10
	1	0.06	0.06	0.05	0.05	0.05	0.05
	2	0.05	0.05	0.05	0.05	0.06	0.05
ρ_L	0.5	0.06	0.06	0.06	0.06	0.05	0.05
	1	0.06	0.06	0.05	0.05	0.05	0.05
	2	0.05	0.05	0.05	0.05	0.05	0.05
ρ_E	0.5	0.08	0.09	0.07	0.07	0.06	0.06
	1	0.06	0.06	0.05	0.05	0.05	0.05
	2	0.05	0.05	0.05	0.05	0.05	0.05
ρ_R		0.05	0.05	0.05	0.05	0.05	0.05

Thus, to take into account the initial samples, a more suitable asymptotic approximation of the allocation proportion satisfying condition (15) is $\pi_n \approx \rho(\mu)(1-2\tau_n)+\tau_n$, then the resulting power function of $\tilde{W}_n$ becomes

$$\Phi\left(\mu\sqrt{n}\sqrt{\frac{\tau_n(1-\tau_n)+(1-2\tau_n)^2\rho(\mu)[1-\rho(\mu)]}{\sigma^2}}-z_{1-\alpha}\right), \quad \mu>0 \tag{16}$$

Note that this simply corresponds to assume a re-scaled target $\tilde{\rho}(\mu) = \rho(\mu)(1-2\tau_n)+\tau_n$ satisfying T1–T3 where, instead of T4, for any fixed sample size n

$$\lim_{x\to\infty}\tilde{\rho}(x) = 1-\tau_n \quad \text{and} \quad \lim_{x\to-\infty}\tilde{\rho}(x) = \tau_n$$

(i.e. from Remark 1, $c = 1-\tau_n$).

Figure 3 shows the performance of $\tilde{W}_n$ under the same simulation scenarios adopted in Figure 2.

If compared with the classical Wald test, $\tilde{W}_n$ is much more robust with respect to degenerate situations induced by the choice of a given target, due to the effect of the starting sample size. Indeed, for any considered target the power function tends to one as μ grows. As shown previously, under ρ_N, ρ_L and ρ_E the ethical component of the target grows as T decreases and therefore the type I errors tend to be inflated, especially for small sample sizes (see Table 1).

Clearly, from power functions (11) and (16), the performance of the classical Wald test W_n (for any chosen starting sample size n_0) is substantially the same of $\tilde{W}_n$ with $n_0 = 0$. Therefore, an application of RA procedures with $n_0 \geq 1$ combined with test $\tilde{W}_n$ is preferable from a practical viewpoint.

However, does every choice of the target guarantee suitable properties of the power function? And how strong should be the ethical skew in order to avoid an anomalous behavior of the power?

4 Properties of the power function

Assuming without loss of generality $\sigma = 1$, if we let for any $x > 0$

$$g_n(x) = x\sqrt{n}\sqrt{\tau_n(1-\tau_n)+(1-2\tau_n)^2\rho(x)[1-\rho(x)]}-z_{1-\alpha} \tag{17}$$

the power of test $\tilde{W}_n$ is $\Phi(g_n(\mu))$ and a first requirement is that the power function should tend to one as the sample size grows. Due to the properties of $\Phi(\cdot)$, this condition is always satisfied since, for every fixed $x > 0$, $\lim_{n\to\infty} g_n(x) = \infty$.

Moreover, for any fixed n (sufficiently large for the CLT approximation), additional fundamental requirements are:

C1: the power should reach 1 as μ tends to infinity;

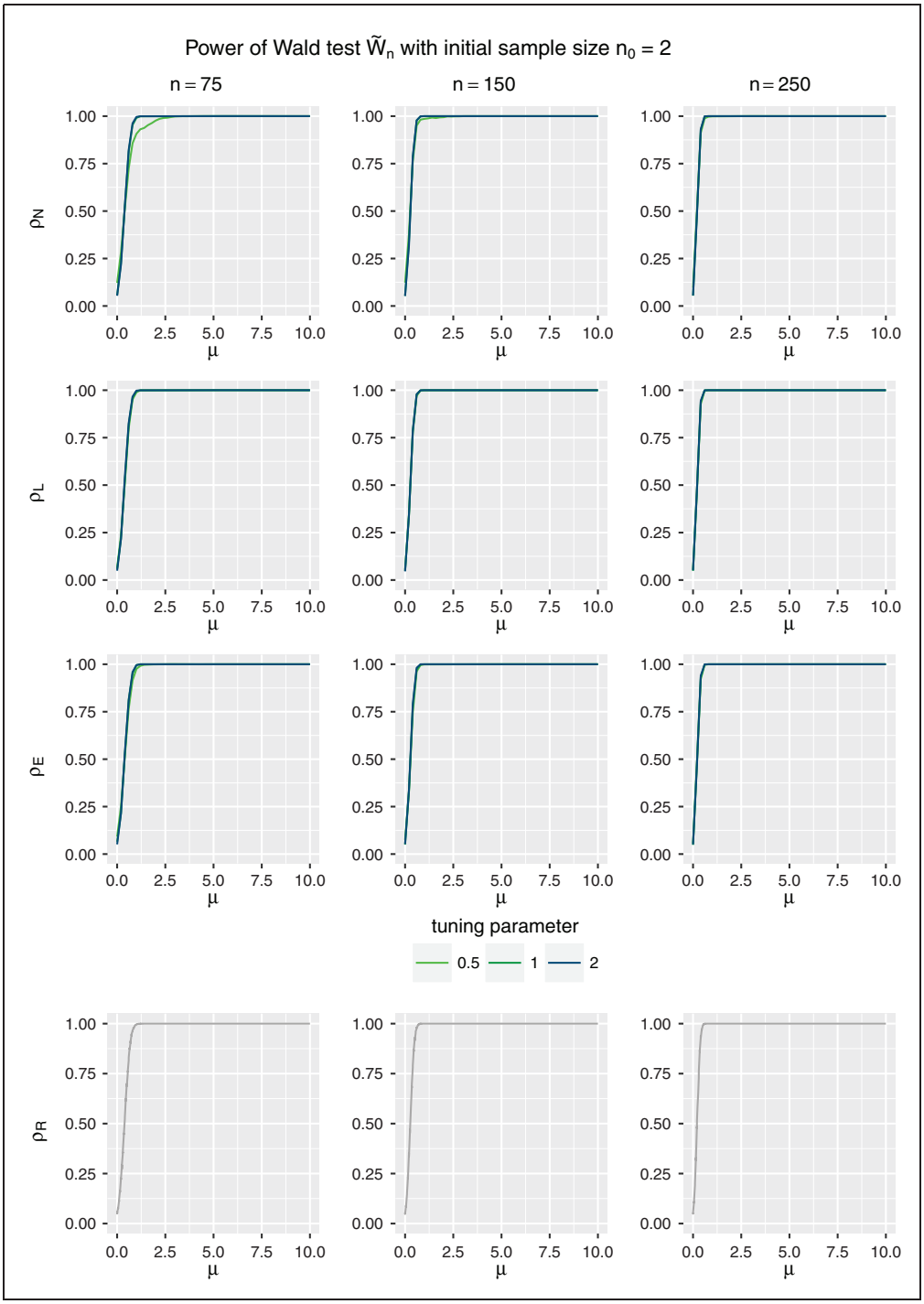

Figure 3. Power of the Wald test $\tilde{W}_n$ under ρ_N, ρ_L and ρ_E ($T = 0.5$, 1 and 2) and ρ_R ($\mu_B = 1$) with $n = 75$, 150, 250 and starting sample size $n_0 = 2$.

C2: the power should be increasing in μ.

Provided that $\tau_n \neq 0$, condition C1 is always satisfied since $\lim_{x \to \infty} g_n(x) = \infty$. Whereas if $\tau_n = 0$ (i.e. $n_0 = 0$), C1 is fulfilled only when $\lim_{x \to \infty} x^2[1 - \rho(x)] = \infty$, namely for targets with a low ethical improvement (as discussed in Section 3 for W_n).

Condition C2 means that, for n sufficiently large (and for any fixed n_0), $g_n(x)$ should be increasing in x. This crucial property is not generally guaranteed for any chosen target allocation (even if $n_0 \neq 0$), as the following theorem shows.

Theorem 1. *A target ρ induces a monotonically increasing power function of Wald test $\tilde{W}_n$ if and only if*

$$\rho(x)[1-\rho(x)] > x\rho'(x)\left[\rho(x)-\frac{1}{2}\right], \quad \text{for all } x>0 \tag{18}$$

Proof. See the Appendix. □

Example 1. *Taking into account ρ_Z in (2), condition (18) becomes*

$$\frac{\sqrt{\mu_B}\sqrt{x+\mu_B}}{(\sqrt{x+\mu_B}+\sqrt{\mu_B})^2} > \frac{x\sqrt{\mu_B}(\sqrt{x+\mu_B}-\sqrt{\mu_B})}{4\sqrt{x+\mu_B}(\sqrt{x+\mu_B}+\sqrt{\mu_B})^3}, \quad \text{for } x>0$$

namely

$$(3x+4\mu_B)(\sqrt{x+\mu_B}+\sqrt{\mu_B})+2x\sqrt{\mu_B}>0, \quad \text{for } x>0$$

that is clearly satisfied since $\mu_B>0$; therefore, adopting ρ_Z the power is monotonically increasing in μ.

To obtain suitable classes of targets satisfying Theorem 1, it could be useful to take into account the hazard function (widely used in the survivor analysis literature) associated with a given target ρ, by letting

$$h_\rho(x) := \frac{\rho'(x)}{1-\rho(x)}, \quad \text{for any } x>0$$

Note that, from T2 and T3, the hazard $h_\rho:\mathbb{R}^+ \to \mathbb{R}^+$ is a continuous function with $\lim_{x\to 0} h_\rho(x) = 2\rho'(0)$.

Corollary 1. *Given a target allocation ρ, if the corresponding hazard satisfies the condition $\lim_{x\to\infty} xh_\rho(x) = K > 2$, then the power of test $\tilde{W}_n$ is locally decreasing. Whereas if $xh_\rho(x) \leq 2$ for every $x>0$, then the power is* monotonically increasing.

Proof. See the Appendix. □

Example 2. *From Corollary 1, every target with constant or monotonically increasing hazard leads to an asymptotic power which is not monotonically increasing. For instance, the exponential target ρ_E has a constant hazard $h_{\rho_E}(x) = T^{-1}$ for every $x>0$, which induces a locally decreasing power since $\lim_{x\to\infty} xh_{\rho_E}(x) = \lim_{x\to\infty} xT^{-1} = \infty$. Also for the Normal and the Logistic targets the power is not monotonically increasing, since*

$$xh_{\rho_N}(x) = \frac{xe^{-x^2/(2T^2)}}{T\sqrt{2\pi}[1-\Phi(x/T)]} \to \infty, \quad \text{as } x\to\infty$$

and

$$xh_{\rho_L}(x) = \frac{x\rho_L(x)}{T} \to \infty, \quad \text{as } x\to\infty$$

Whereas assuming ρ_R in (3) the corresponding hazard is $h_{\rho_R}(x) = (2\mu_B + x)^{-1}$, thus $xh_{\rho_R}(x) \leq 2$ for any $x>0$ and therefore the power is monotonically increasing. The same conclusion still holds for the Cauchy target ρ_C in (5), since

$$xh_{\rho_C}(x) = \frac{2Tx}{(T^2+x^2)[\pi - 2\arctan(x/T)]} \leq 2$$

An additional characterization of target functions inducing a locally decreasing power can be derived through differential inequalities.

Corollary 2. *Given a target ρ, if there exists $\eta > 0$ such that*

$$x^2\rho(x)[1-\rho(x)] \leq \eta^2\rho(\eta)[1-\rho(\eta)], \quad \text{for all } x \geq \eta$$

then the power of test $\tilde{W}_n$ is not monotonically increasing.

Proof. See the Appendix. □

Example 3. *Taking into account the Logistic target ρ_L with $T=1$, then for any $x \geq 3$*

$$x^2\rho_L(x)[1-\rho_L(x)] = \frac{x^2e^x}{(1+e^x)^2} \leq \frac{3^2e^3}{(1+e^3)^2} = 0.40659$$

and therefore the ensuing power is locally decreasing.

Since the power function of W_n in (11) could be regarded as a special case of (16) with $\tau_n = 0$, all of the previous results about monotonicity also hold for the classical Wald test W_n. However, even if some target allocations do not guarantee that the ensuing power is monotonically increasing, the following result shows how $\tilde{W}_n$ combined with a suitable choice of both (i) the sample size and (ii) the starting samples allow one to avoid this drawback.

Theorem 2. *For any chosen target ρ, letting*

$$\beta_\rho = \max_{x\in\mathbb{R}^+}\left\{x\rho'(x)\left[\rho(x)-\frac{1}{2}\right]-\rho(x)[1-\rho(x)]\right\}$$

then there exists a couple $(n^*; \tau_{n^*})$ with

$$n^* = 2\sqrt{4\beta_\rho+1} \quad \text{and} \quad \tau_{n^*} = \frac{1}{2}-\frac{1}{n^*}$$

such that, for all $n > n^*$ and for all $\tau_n > \tau_{n^*}$ the power function of the Wald test $\tilde{W}_n$ is monotonically increasing.

Proof. See the Appendix. □

Remark 3. *Theorem 2 derives the minimum ratio τ_{n^*} between the starting sample and the sample sizes ensuring the monotonicity of the power function. However, for any given target ρ, the power of the test $\tilde{W}_n$ is monotonically increasing in τ_n for any μ. Indeed, as τ_n grows then the RA component of the design vanishes and therefore the power reaches the one of the balanced design. This is clearly another view for the already discussed trade-off between ethics and inferential precision.*

Taking into account the previously defined targets inducing a locally decreasing power, namely ρ_N, ρ_L and ρ_E, Table 2 shows how the sample size and the starting sample can be chosen in order to obtain a monotonically increasing power function.

In general, for every target the (minimum) sample size requested is $n \geq 3$, that is always satisfied in practice. While as regards the choice of the starting samples, the ensuing condition is not-trivially fulfilled, especially for the

Table 2. Computations of n^* and τ_{n^*} for the targets ρ_N, ρ_L and ρ_E.

ρ	T	β_ρ	n^*	τ_{n^*}
ρ_N	For all $T>0$	0.031	2.12	3%
ρ_L	For all $T>0$	0.018	2.07	2%
ρ_E	For all $T>0$	0.011	2.04	1%

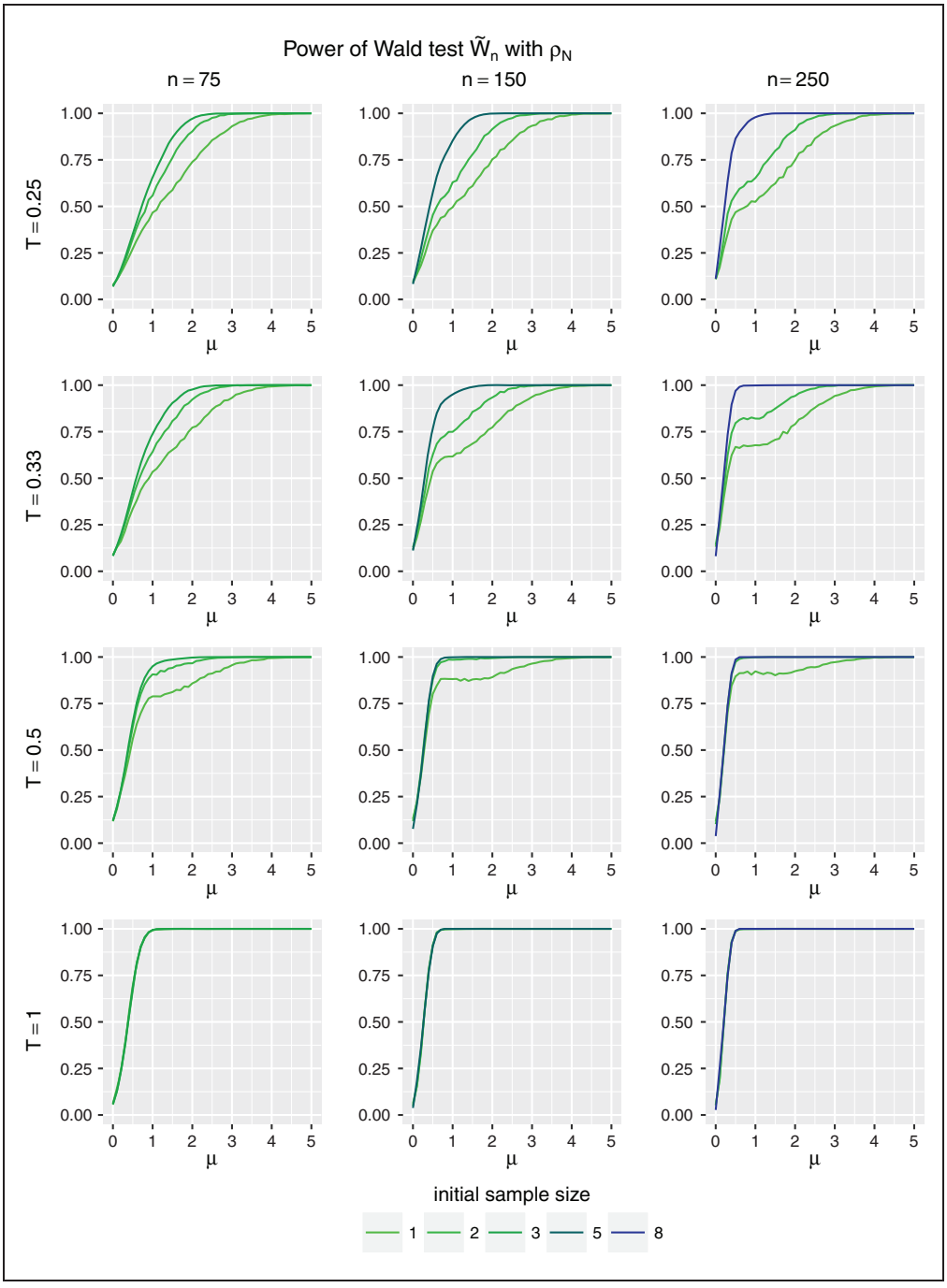

Figure 4. Power of the Wald test $\tilde{W}_n$ under ρ_N with $T = 0.25$, 0.33, 0.5 and 1: colors go from green to blue for $n_0 = 1$, 2 and $\lceil \tau_{n^*} \times n \rceil$, with $n = 75$, 150 and 250.

large sample framework of asymptotic inference. Indeed, for a sample size of $n = 250$, if we choose ρ_N then $n_0 = 8$ starting allocations on each treatment guarantee a monotonically increasing power (i.e. only the remaining 234 assignments will be allocated in the RA way).

5 A simulation study

Section 3 collects the theoretical results allowing the applicability of Wald-type Z-tests under RA designs, while in Section 4 we analyze the corresponding power function from a theoretical point of view. In particular, we show

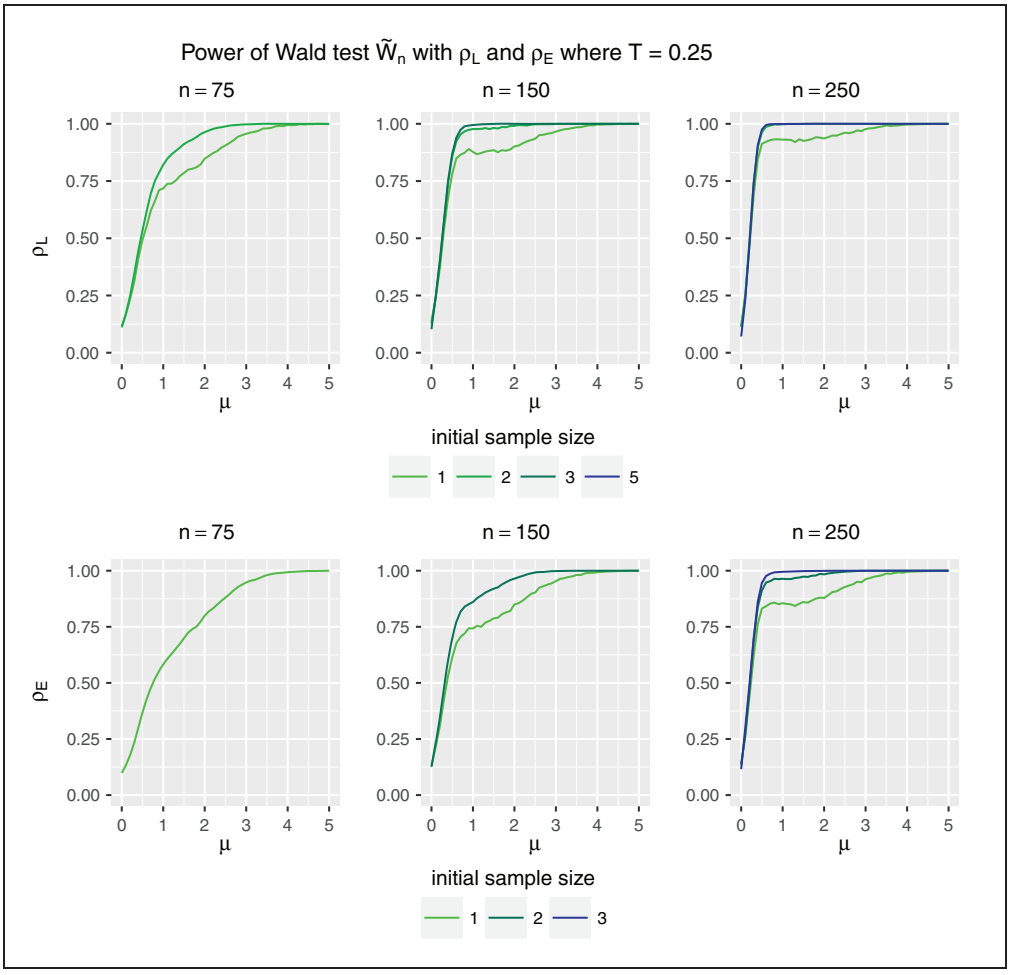

Figure 5. Power of the Wald test $\tilde{W}_n$ under ρ_L and ρ_E with $T = 0.25$: colors go from green to blue for $n_0 = 1$, 2 and $\lceil \tau_{n^*} \times n \rceil$, with $n = 75$, 150 and 250.

that, for certain classes of targets, the ensuing power is locally decreasing in the difference between the treatment effects and could not tend to one as μ grows, stressing also how a suitable choice of the starting sample overcomes this drawback.

In this section we focus on the practical implications in terms of loss of power by means of a simulation study, where the chosen RA procedure is the ERADE with $\gamma = 0.5$. The results come from 5000 simulations with sample sizes $n = 75$, 150 and 250, where the responses are generated following a Gaussian distribution with $\sigma^2 = 1$, $\mu_B = 1$ and $\mu_A = \mu_B + k$, with $k \geq 0$. The considered targets are ρ_N, ρ_L and ρ_E with several values of the tuning parameter T; for each scenario, the initial sample size n_0 spans from 1 to $\lceil \tau_{n^*} \times n \rceil$, i.e. the smallest integer greater than or equal to $\tau_{n^*} \times n$ (colors go from green to blue for $\tilde{W}_n$ and from yellow to red for W_n, as the starting sample grows). For the sake of readability, Figures 4–6 show the simulated power functions of $\tilde{W}_n$ and W_n only for an initial sample size $n_0 = 1$, 2 and $\lceil \tau_{n^*} \times n \rceil$; for other potential values of n_0 (that could be present depending on the chosen sample size and the considered target) the local non-monotonicity is still present but not easily noticeable, and it tends to be negligible as n_0 approaches $\tau_{n^*} \times n$.

Figure 4 shows the behavior of the power function of Wald test $\tilde{W}_n$ adopting ρ_N. In general, the power tends to become increasing as the tuning parameter T grows; this is clearly coherent since $\rho \to 1/2$ as the randomization component of the target grows and therefore the power tends to that of the balanced design, namely $\Phi\left(2^{-1}\sqrt{n}\mu - z_{1-\alpha}\right)$, that is monotonically increasing in μ. While for small values of T, where the ethical skew is stronger, also the CLT approximation is partially compromised, as it can be shown from the inflated type I errors that occur for $T = 0.25$ and 0.33 or for small sample size $n = 75$ (see also Figure 5). Taking into account ρ_N in Figure 4, even if for $T = 1$ the non-monotonicity of the power is not noticeable in the plots, the power function is still slightly decreasing, but this behavior is present for large values of μ, hidden from the property of $\Phi(\cdot)$. Indeed,

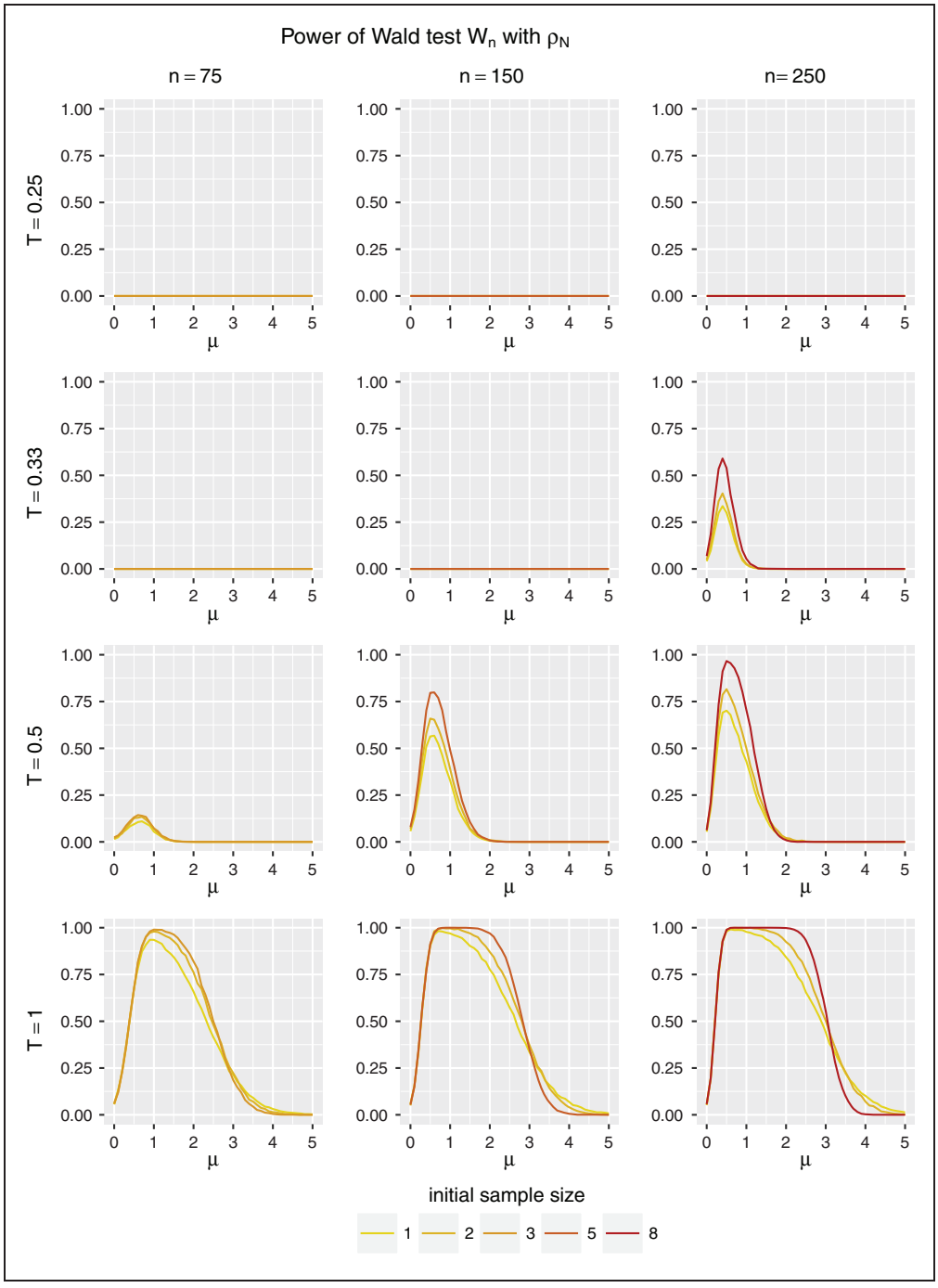

Figure 6. Power of the Wald test W_n under ρ_N: colors go from yellow to red for $n_0 = 1$, 2 and $\lceil \tau_{n^*} \times n \rceil$, with $n = 75$, 150 and 250.

for $n = 75$ and $n_0 = 2$ the ensuing power is slightly decreasing for values of μ between 1.75 and 1.9. The same behavior holds for ρ_E and ρ_L as it can be seen in Figure 5 (as also confirmed by further simulations, omitted here for brevity).

As regards the starting sample size, the plots (see Figures 4 and 5) are very similar for all of the considered targets: for low values of n_0 the power function is locally decreasing, while as the sample dimension increases, so should n_0 in order to ensure a monotonically increasing power. Finally, note that, for small values of T, the gain in terms of power highly increases even for small increments of n_0 as discussed in Remark 3.

In order to explain the inadequacy of the classical Wald test, Figures 6 shows the power of W_n when target ρ_N is employed under the same simulation scenarios described previously. As it can be seen, the anomalous behavior of

the power function is strongly accentuated for small sample sizes and small values of T; moreover, every choice of the starting sample size does not allow to avoid degenerate situations in which the power vanishes as μ grows (the same clearly holds for ρ_E and ρ_L).

6 Discussion

The choice of the target function plays a crucial role in RA methodology, since it incorporates ethical requirements with inferential goals. In general, targets should skew the assignments towards the best treatment and a fundamental question is how strong should be the ethical improvement to obtain a suitable trade-off between ethical aims and inferential precision.

Our paper is focused on this problem by taking into account hypothesis testing instead of the classical estimation. Even though we do not suggest a specific target, we show the inadequacy of classes of target functions for hypothesis testing in comparative clinical trials, stressing also the crucial role of the initial sample size.

Under RA randomization procedures, the classical Wald test W_n could be applied only when the desired target ρ has a low ethical skew, namely when $\lim_{\mu\to\infty}\mu^2[1-\rho(\mu)]\to\infty$ (e.g. under re-scaled targets not satisfying T4), while $\tilde{W}_n$ should be generally preferable.

Moreover, for testing hypotheses, RA randomization procedures should not be applied without a starting sample. Indeed, for certain targets the choice $n_0=0$ may induce an accentuated anomalous behavior of the power, that becomes strongly decreasing even for small differences between the treatment effects, while for high values of μ the RA rule tends to allocate every subject to the best treatment inducing a null power. On the other hand, any choice of the starting sample size $n_0\geq 1$ could be suitable if the chosen target satisfies condition (18): in this case, small values of n_0 improve the ethical goals of the RA design. Whereas when the desired ρ induces a locally decreasing power, then the starting sample size should be chosen in an accurate manner, as shown in Theorem 2, that clearly conflicts with the general suggestion $n_0=2$ given by Hu et al.[14]

Finally, we wish to stress that our results still hold even for an alternative hypothesis H_1: $\mu\neq 0$, where all of the previous conclusions about the monotonicity of the power function could be interpreted in terms of monotonicity of the non-centrality parameter ϕ of a non-central chi-square distribution with one dof. Indeed, taking into account the classical Wald test, from the CLT in (8) under H_0 the statistic W_n^2 is asymptotically distributed as a central chi-square with one dof, so that H_0 is rejected when $W_n^2>\chi^2_{1;1-\alpha}=z^2_{1-\alpha/2}$, where $\chi^2_{1;1-\alpha}$ is the $(1-\alpha)$-percentile of a central chi-square with one dof. While under the alternative hypothesis, W_n^2 converges to a non-central chi-square with one dof and non-centrality parameter $\phi=\mu^2 n\rho(\mu)[1-\rho(\mu)]/\sigma^2$ and therefore the previously obtained results still hold (taking now into account $|\mu|$ instead of μ), since the power is a monotonically increasing function of ϕ (as in Hu and Rosenberger[31] and Tymofyeyev et al.).[7]

Acknowledgments

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

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

References

1. Thall PF, Fox P and Wathen J. Statistical controversies in clinical research: scientific and ethical problems with adaptive randomization in comparative clinical trials. *Ann Oncol* 2015; **26**: 1621–1628.
2. Thall PF, Fox P and Wathen J. Some caveats for outcome adaptive randomization in clinical trials. In: Sverdlov O (ed.) *Modern adaptive randomized clinical trials: statistical and practical aspects*. Oxford: Chapman & Hall/CRC Biostatistics, 2015, pp.287–305.
3. CHMP. Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design, 2007. Available at: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003616.pdf
4. FDA. Guidance for industry. Adaptive design clinical trials for drugs and biologics (draft document), 2010. Available at: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM201790.pdf
5. Rosenberger WF, Stallard N, Ivanova A, et al. Optimal adaptive designs for binary response trials. *Biometrics* 2001; **57**: 909–913.

6. Biswas A, Bhattacharya R and Zhang L. Optimal response-adaptive designs for continuous responses in phase III trials. *Biometrical J* 2007; **49**: 928–940.
7. Tymofyeyev Y, Rosenberger WF and Hu F. Implementing optimal allocation in sequential binary response experiments. *J Am Statist Assoc* 2007; **102**: 224–234.
8. Biswas A and Bhattacharya R. Optimal response-adaptive designs for normal responses. *Biometrical J* 2009; **51**: 193–202.
9. Baldi Antognini A and Giovagnoli A. Compound optimal allocation for individual and collective ethics in binary clinical trials. *Biometrika* 2010; **97**: 935–946.
10. Baldi Antognini A and Zagoraiou M. Multi-objective optimal designs in comparative clinical trials with covariates: the reinforced doubly-adaptive biased coin design. *Ann Statist* 2012; **40**: 1315–1345.
11. Biswas A and Bhattacharya R. Response-adaptive designs for continuous treatment responses in phase III clinical trials: A review. *Statist Meth Med Res* 2016; **25**(1): 81–100.
12. Biswas A and Bhattacharya R. Near efficient target allocations in response-adaptive randomization. *Statist Meth Med Res* 2016; **25**(2): 807–820.
13. Hu F and Zhang LX. Asymptotic properties of doubly adaptive biased coin designs for multi-treatment clinical trials. *Ann Statist* 2004; **32**: 268–301.
14. Hu F, Zhang LX and He X. Efficient randomized adaptive designs. *Ann Statist* 2009; **37**: 2543–2560.
15. Hu F and Rosenberger WF. *The Theory of Response-Adaptive Randomization in Clinical Trials*. New York: John Wiley & Sons, 2006.
16. Baldi Antognini A and Giovagnoli A. *Adaptive Designs for Sequential Treatment Allocation*. Boca Raton, FL: Chapman & Hall/CRC Biostatistics, 2015.
17. Rosenberger WF and Sriram TN. Estimation for an adaptive allocation design. *J Statist Planning Inference* 1996; **59**: 309–319.
18. Durham SD, Flournoy N and Rosenberger WF. Asymptotic normality of maximum likelihood estimators from multiparameter response-driven designs. *J Statist Planning Inference* 1997; **60**: 69–76.
19. Rosenberger WF, Flournoy N and Durham SD. Asymptotic normality of maximum likelihood estimators from multiparameter response-driven designs. *J Statist Planning Inference* 1997; **60**: 69–76.
20. Melfi V and Page C. Estimation after adaptive allocation. *J Statist Planning Inference* 2000; **29**: 353–363.
21. Melfi V, Page C and Geraldes M. An adaptive randomized design with application to estimation. *Can J Statist* 2001; **29**: 107–116.
22. Atkinson AC and Biswas A. Adaptive biased-coin designs for skewing the allocation proportion in clinical trials with normal responses. *Statist Med* 2005; **24**: 2477–2492.
23. Atkinson AC and Biswas A. Bayesian adaptive biased-coin designs for clinical trials with normal responses. *Biometrics* 2005; **61**: 118–125.
24. Baldi Antognini A and Giovagnoli A. On the large sample optimality of sequential designs for comparing two or more treatments. *Sequential Anal* 2005; **24**: 205–217.
25. Geraldes M, Melfi V, Page C, et al. The doubly adaptive weighted difference design. *J Statist Planning Inference* 2006; **136**: 1923–1939.
26. Zhu H and Hu F. Sequential monitoring of response-adaptive randomized clinical trials. *Ann Statist* 2010; **38**(4): 2218–2241.
27. Yi Y and Wang X. Comparison of Wald, score, and likelihood ratio tests for response adaptive designs. *J Statist Theory Applicat* 2011; **10**: 553–569.
28. Azriel D, Mandel M and Rinott Y. Optimal allocation to maximize power of two-sample tests for binary response. *Biometrika* 2012; **99**: 101–113.
29. Zhang LX and Rosenberger WF. Response-adaptive randomization for clinical trials with continuous outcomes. *Biometrics* 2006; **62**: 562–569.
30. Bandyopadhyay U and Biswas A. Adaptive designs for normal responses with prognostic factors. *Biometrika* 2001; **88**: 409–419.
31. Hu F and Rosenberger WF. Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparisons. *J Am Statist Assoc* 2003; **98**: 671–678.

Appendix

Proof of Theorem 1

From equation (17),

$$g_n'(x) = \sqrt{n}\sqrt{\tau_n(1-\tau_n) + (1-2\tau_n)^2\rho(x)[1-\rho(x)]} + + \frac{x\sqrt{n}(1-2\tau_n)^2\left[\frac{1}{2} - \rho(x)\right]\rho'(x)}{\sqrt{\tau_n(1-\tau_n) + (1-2\tau_n)^2\rho(x)[1-\rho(x)]}}$$

and thus the power function is monotonically increasing if and only if

$$\rho(x)[1-\rho(x)]+\frac{n_0(n-n_0)}{(n-2n_0)^2} > x\rho'(x)\left[\rho(x)-\frac{1}{2}\right], \quad \text{for all } x>0 \tag{19}$$

From T1–T3, $\rho(x)>1/2$ and $\rho'(x)>0$ for every $x>0$; therefore, condition (18) implies (19) since $n_0(n-n_0)/(n-2n_0)^2$ is positive. On the other hand, suppose that there exists $I_{\mathbb{R} \text{ such that}}$

$$\rho(x)[1-\rho(x)] \le x\rho'(x)\left[\rho(x)-\frac{1}{2}\right], \quad \text{forevery} x \in I$$

In this case the power is monotonically increasing if, for every $x \in I$,

$$\frac{n_0(n-n_0)}{(n-2n_0)^2} > x\rho'(x)\left[\rho(x)-\frac{1}{2}\right]-\rho(x)[1-\rho(x)] \tag{20}$$

However, for any fixed n_0, the left-hand side of (20) is monotonically decreasing in n and tends to zero as n grows; therefore, condition (20) does not hold for sufficiently large n.

Proof of Corollary 1

From condition (18), a target ρ induces a monotonically increasing power if and only if

$$\frac{\rho(x)}{\rho(x)-\frac{1}{2}} > xh_\rho(x), \quad \text{for all } x>0 \tag{21}$$

Note that the left-hand side in (21) is monotonically decreasing in x, tends to infinity for $x \to 0$ and goes to two as x grows. Thus, if $\lim_{x\to\infty} xh_\rho(x)=K>2$, then (21) does not hold for sufficiently large x. Whereas if $xh_\rho(x) \le 2$ for every $x>0$, then (21) is satisfied.

Proof of Corollary 2

Letting $q_\rho(x) := \rho(x)[1-\rho(x)]$, then it is straightforward to see that $q'_\rho(x)=\rho'(x)[1-2\rho(x)]$ and

$$\rho'(x)\left[\rho(x)-\frac{1}{2}\right]=-\frac{q'_\rho(x)}{2}$$

Hence, condition (18) could be rewritten as follows

$$-\frac{xq'_\rho(x)}{2} < q_\rho(x), \quad \text{forevery} x>0$$

namely $q'_\rho(x) > -2x^{-1}q_\rho(x)$ for every $x>0$. Thus, if there exists $\eta>0$ such that $q'_\rho(x) \le -2x^{-1}q_\rho(x)$ for $x>\eta$, by applying Grönwall's inequality, we obtain

$$q_\rho(x) \le q_\rho(\eta)\exp\left\{\int_\eta^x -\frac{2dt}{t}\right\}, \quad \text{for } x>\eta$$

that is,

$$q_\rho(x) \le q_\rho(\eta)\left(\frac{\eta}{x}\right)^2, \quad \text{for } x>\eta$$

and therefore the power function in (16) is not monotonically increasing.

Proof of Theorem 2

Letting for any $x > 0$

$$B_\rho(x) := x\rho'(x)\left[\rho(x) - \frac{1}{2}\right] - \rho(x)[1 - \rho(x)]$$

then $B_\rho(x)$ is a continuous and bounded function with $\lim_{x\to 0} \mathrm{B}_\rho(x) = -1/4$ and $\lim_{x\to\infty} B_\rho(x) = 0$. From (19), $g_n'(x) > 0$ if and only if for every $x > 0$

$$G(\tau_n) := \frac{\tau_n(1-\tau_n)}{(1-2\tau_n)^2} > B_\rho(x) \tag{22}$$

where the function $G(\cdot)$ is monotonically increasing and reaches its maximum $\tilde{G}$ given by

$$\tilde{G} = \begin{cases} \frac{n^2-4}{16}, & \text{for} n \text{even}, \text{at} \tau_n = \frac{1}{2} - \frac{1}{n}, \\ \frac{n^2-1}{4}, & \text{for} n \text{odd}, \text{at} \tau_n = \frac{1}{2} - \frac{1}{2n}. \end{cases}$$

Thus, $\tilde{G} > \beta_\rho$ for any $n > n^* = 2\sqrt{4\beta_\rho + 1}$ and, from the monotonicity of $G(\cdot)$, there exists a unique τ_{n^*} given by

$$\tau_{n^*} = \frac{1}{2} - \frac{1}{2\sqrt{4\beta_\rho + 1}}$$

such that, for every $\tau_n > \tau_{n^*}$, $G(\tau_n) > \beta_\rho$. Therefore, condition (22) is satisfied.