Stat. Meth. & Appl. (2007) 16:321–346
DOI 10.1007/s10260-006-0042-4

ORIGINAL ARTICLE

# A numerical study for comparing two response-adaptive designs for continuous treatment effects

**Anna Maria Paganoni · Piercesare Secchi**

Accepted: 15 December 2006 / Published online: 24 January 2007
© Springer-Verlag 2007

**Abstract** We study two sequential, response-adaptive randomized designs for clinical trials; one has been proposed in Bandyopadhyay and Biswas (Biometrika 88: 409–419, 2001) and in Biswas and Basu (Sankhya Ser B 63:27–42, 2001), the other stems from the randomly reinforced urn introduced and studied in Muliere et al. (J Stat Plan Inference 136:1853–1874, 2006a). Both designs can be used in clinical trials where the response from each patient is a continuous variable. Comparison is conducted through numerical studies and along a new guideline for the evaluation of a response-adaptive design.

**Keywords** Response adaptive designs · Clinical trials · Urn schemes

## 1 Introduction

In this paper we compare two sequential, response-adaptive randomized designs suitable for clinical trials where the observed response from each patient is a continuous variable.

Suppose patients enter the trial sequentially; a sequential, response-adaptive randomized design assigns each new patient in the trial to a treatment with a probability that changes along the trial according to the data that have already accrued about treatment effects. Aiming, for ethical reasons, at maximizing the patient's personal experience while treated in the trial, sequential response-adaptive designs incline to assign more patients to the better treatment, while seeking to keep randomness as a basis for statistical inference; for a

A. M. Paganoni · P. Secchi (✉)
Dipartimento di Matematica "F. Brioschi", Politecnico di Milano,
Piazza Leonardo da Vinci, 32, 20123 Milano, Italy
e-mail: piercesare.secchi@polimi.it

foundational description of this approach to sequential design see Flournoy and Rosenberger (1995), Rosenberger (1996, 2002) and the book by Rosenberger and Lachin (2002).

Most of the literature on sequential, response-adaptive designs deals with the case where responses are binary variables. Among the few exceptions, are the design proposed by Bandyopadhyay and Biswas (2001), further explored in Biswas and Basu (2001), and the design based on the randomly reinforced urn introduced in Muliere et al. (2006a). In this paper we will compare these two adaptive designs for clinical trials with responses on the continuous scale: details on the designs and their definitions are given in the next section after setting the probabilistic stage for the analysis.

The comparison is carried along a guideline that we introduce in Sect. 3. We believe that the analysis generated by this procedure is deeply rooted in two foundational motivations, moving most research on response-adaptive designs:

1. to generate experimental data adequate for statistical inference on treatment effects;
2. to randomly allocate patients to treatments, using the information about treatments generated along the trial for biasing the allocation probabilities toward the better treatment.

This point of view is very much in the spirit of Hu and Rosenberger (2003). The quantities we propose in Sect. 3 for the evaluation of a response-adaptive design, are difficult to compute analytically. Hence in Sect. 4, for both designs object of this paper, we conduct a numerical study for eliciting them in different illustrative instances; this put us in the position to construct a comparative analysis for the two designs. A discussion on the two designs merits concludes the paper.

## 2 Two response-adaptive designs for continuous responses

We indicate with $M_i$ and $N_i$ the response for patient $i = 1, 2, \ldots$, depending on whether the patient has been allocated to treatment $A$ or treatment $B$ respectively. We assume that the random variables $M_1, M_2, \ldots, M_i, \ldots$ are i.i.d. with probability distribution function $\mu$, that the random variables $N_1, N_2, \ldots, N_i, \ldots$ are i.i.d. with probability distribution function $\nu$ and that the two sequences are independent. Furthermore we assume that the expected values of the response distributions,

$$m_\mu = \int x\mu(\mathrm{d}x) \quad \text{and} \quad m_\nu = \int x\nu(\mathrm{d}x),$$

respectively, exist and are finite.

**Definition 1** A sequential design $\rho$ is a sequence $(\rho_0, \rho_1, \ldots, \rho_i, \ldots)$ such that $\rho_0 \in [0, 1]$ while, for $i \geq 1$, $\rho_i$ is a measurable function from the space $(\{0, 1\} \times \mathbf{R})^i$ to the space $[0, 1]$.

If patients enter the trial sequentially, the design $\rho$ describes the strategy followed by the experimenter for allocating patients to treatments. In fact, for $i = 1, 2, \ldots$, let us indicate with $X_i$ the indicator function which takes value 1 if patient $i$ is allocated to treatment $A$ and 0 if patient $i$ is allocated to treatment $B$; moreover, let $\mathcal{F}_i$ be the sigma-field generated by

$$X_1, X_1M_1 + (1 - X_1)N_1, \ldots, X_i, X_iM_i + (1 - X_i)N_i,$$

i.e. the sigma-field representing the information available to the experimenter before the allocation of the $(i+1)$th patient to a treatment. We assume that $X_1$ has Bernoulli($\rho_0$) distribution and that, for $i \geq 1$, the conditional distribution of $X_{i+1}$, given $\mathcal{F}_i$, is

$$\text{Bernoulli}(\rho_i(X_1, X_1M_1 + (1 - X_1)N_1, \ldots, X_i, X_iM_i + (1 - X_i)N_i)).$$

When the trial sample size is fixed *a priori* and the target number of patients that are going to be treated is $n$, specification of the elements $\rho_n, \rho_{n+1}, \ldots$ of $\rho$ becomes irrelevant.

In this paper we will consider and compare two different sequential designs. Both designs are randomized and response-adaptive; that is, both modify, along the trial, the probabilities of allocation to treatments according to the information provided by past data. Furthermore, both designs can be used in clinical trials where responses are continuous variables and the ultimate goal is to decide the treatment generating the response with highest mean.

### 2.1 A response-adaptive design generated by a randomly reinforced urn

Fix $0 < \lambda \leq \upsilon < \infty$, a transformation function $\phi : \mathbf{R} \to [\lambda, \upsilon]$ and a couple of positive real numbers $(b_0, w_0)$. Define the design $\sigma$ by setting $\sigma_0 = b_0/(b_0 + w_0)$ and, for $i \geq 1$ and $(x_1, e_1, \ldots, x_i, e_i) \in (\{0, 1\} \times \mathbf{R})^i$, let

$$\sigma_i(x_1, e_1, \ldots, x_i, e_i) = \frac{b_0 + \sum_{j=1}^{i} x_j \phi(e_j)}{b_0 + w_0 + \sum_{j=1}^{i} \phi(e_j)}.$$

That is, according to $\sigma$, given the past allocations $X_1, \ldots, X_i$ and the observed responses $X_1M_1 + (1 - X_1)N_1, \ldots, X_iM_i + (1 - X_i)N_i$, the experimenter allocates the next patient $i + 1$ to treatment $A$ with probability

$$\begin{aligned}&\sigma_i(X_1, X_1M_1 + (1 - X_1)N_1, \ldots, X_i, X_iM_i + (1 - X_i)N_i) \\ &= \frac{b_0 + \sum_{j=1}^{i} X_j \phi(M_j)}{b_0 + w_0 + \sum_{j=1}^{i} X_j \phi(M_j) + \sum_{j=1}^{i} (1 - X_j)\phi(N_j)}.\end{aligned}$$

The sequential design $\sigma$ is randomized and response-adaptive; it is implemented by an urn containing initially $b_0$ balls of color $A$ and $w_0$ balls of color $B$. When

patient $i = 1, 2, \ldots$ enters the trial, the experimenter allocates him to a treatment by sampling a ball from the urn. After the treatment effect $X_i M_i + (1 - X_i)N_i$ is observed, and before allocating the $(i + 1)$th patient, the urn composition is reinforced with a number $\phi(X_i M_i + (1 - X_i)N_i)$ of balls of the same color as that of the last one sampled.

In Muliere et al. (2006a) it is proved that, if

$$\int \phi(x)\mu(\mathrm{d}x) > \int \phi(x)\nu(\mathrm{d}x), \tag{1}$$

then, whatever the initial composition $(b_0, w_0)$,

$$P[\lim_{i\to\infty} \sigma_i(X_1, X_1 M_1 + (1 - X_1)N_1, \ldots, X_i, X_i M_i + (1 - X_i)N_i) = 1] = 1; \tag{2}$$

that is, asymptotically, with probability one the design $\sigma$ assigns patients to treatment with $\phi$-transformed response with highest mean. For instance, if treatment effects are positive and bounded real quantities with different mean values and we let $\phi$ to be the identity function, the design $\sigma$ asymptotically allocates patients to the treatment with highest mean value.

The design $\sigma$ crucially depends on the transformation $\phi$ : this can be taken to be a utility function as argumented in Muliere et al. (2006b). In fact, assume that the response distributions $\mu$ and $\nu$ belong to a class $\mathcal{P}$ of probability distributions on $\mathfrak{R}$. The experimenter is able to express preferences among the elements of $\mathcal{P}$; the preference pattern is represented by a bounded utility function $\phi : \mathfrak{R} \to \mathfrak{R}$ such that if $P_1$ and $P_2$ are in $\mathcal{P}$ then $P_1$ is preferred to $P_2$ if and only if

$$\int \phi(x)P_1(\mathrm{d}x) > \int \phi(x)P_2(\mathrm{d}x)$$

while $P_1$ and $P_2$ are equivalent if and only if $\int \phi(x)\, P_1(\mathrm{d}x) = \int \phi(x)\, P_2(\mathrm{d}x)$. Conditions which guarantee the existence of a bounded utility function $\phi$ such that expected utilities of the elements of $\mathcal{P}$ are ordered in the same way as the true preferences among the $P \in \mathcal{P}$ can be found, for instance, in DeGroot (1970) and Fishburn (1981). It is to be noted that if $\phi$ is such a utility, then $\phi_1 = b\phi + c$, with $b > 0$, is also a utility function that represents the same preference pattern among the elements of $\mathcal{P}$ : hence, without loss of generality, we assume that $0 < \lambda \le \phi(x) \le \upsilon < \infty$ for all $x \in \mathfrak{R}$. Moreover, by letting $\phi$ to be a bounded utility function defined on $\mathfrak{R}^k$, $k \ge 1$, we may easily implement the design $\sigma$ even in situations, more general than those considered in this paper, where responses are multivariate and $\mu$ and $\nu$ are probability distributions on $\mathfrak{R}^k$.

If the trial sample size is fixed to be a very large $n$, (2) supports the claim that $\sigma$ will bias the allocation probabilities of the $n$ patients toward the treatment associated with the $\phi$-transformed response with highest mean—i.e. with highest utility—whatever the initial urn composition $(b_0, w_0)$. However, since convergence in (2) is typically slow, for small $n$ the effect of the initial composition

$(b_0, w_0)$ on the allocation probabilities will be relevant. In this paper we follow this simple rule for setting it; first we allocate an equal number $k$ of patients to treatment $A$ and to treatment $B$ and we observe the responses $M_1, \ldots, M_k$ and $N_{k+1}, \ldots, N_{2k}$ respectively. Then we set

$$b_0 = \sum_{i=1}^{k} \phi(M_i), \quad w_0 = \sum_{i=1}^{k} \phi(N_{k+i})$$

and we proceed to allocate patient $2k+1, 2k+2, \ldots$ to treatments following the design $\sigma$ with initial composition $(b_0, w_0)$. The integer number $k$ is called the *initialization parameter*: in the numerical studies illustrated in this paper it will be set equal to a small fraction of the trial sample size $n$.

*Remark 1* Rate of convergence in (2) is a very interesting and still open problem: a stimulating challenge for the authors.

*Remark 2* In a Bayesian analysis, different choices of $b_0$ and $w_0$ would incorporate different prior believes about the better treatment.

*Remark 3* In the special case where the $\phi$-transformed responses after treatment, $\phi(M_1)$ and $\phi(N_1)$, are equally distributed and the common reinforcement distribution is degenerate at a point mass $m > 0$, the randomly reinforced urn implementing the design $\sigma$ is in fact a Polya urn and the infinite sequence of allocations $X_1, X_2, \ldots$ is exchangeable with de Finetti measure equal to a Beta. For general transformation function $\phi$ and response distributions $\mu$ and $\nu$ with bounded support, the allocation sequence $\{X_n\}$ is only asymptotically exchangeable with associated de Finetti measure equal to the distribution of the almost sure limit of the sequence of allocation probabilities $\{\sigma_n(X_1, X_1M_1 + (1-X_1)N_1, \ldots, X_n, X_nM_n + (1-X_n)N_n)\}$; this limit distribution is generally unknown, but it is characterized in Aletti et al. (2005) as the unique solution of a functional equation involving unknown distributions on [0,1]. For a more detailed study of the theoretical properties of the design $\sigma$, see also May et al. (2005) and Muliere et al. (2006a).

*Remark 4* The design $\sigma$ is a conceptual extension of a response adaptive design proposed by Durham et al. (1998) for clinical trials with dichotomous responses; its definition has been influenced by the popular randomized-play-the-winner (RPW) design by Wei and Durham (1978). When responses are dicothomous, the main difference between RPW and $\sigma$ is that with RPW a successful response to the currently allocated treatment increases the probability of allocating the same treatment to the next patient in the study, whereas a failure increases the probability of allocating the next patient to the alternative treatment. With $\sigma$ allocation probabilities are reinforced when successes to treatment are observed, but they remain unchanged after failures. As a consequence, the target allocation for $\sigma$, i.e. the limit probability of allocating a patient to a treatment, is degenerate on the treatment with highest probability of success

whereas RPW targets a proportion in (0,1) function of both treatment success probabilities.

### 2.2 The triple-B response-adaptive design

In Bandyopadhyay and Biswas (2001) and in Biswas and Basu (2001) a sequential design $\tau$ for comparing mean responses after treatments is proposed which accommodates for robust estimates of the means of response distributions.

Let $G$ be a continuous cumulative distribution function, symmetric about 0, like, for instance, the cumulative distribution $\Phi$ of a standard normal. Set $\tau_0 = 0, \tau_1(x_1, e_1) = 1$ for all $(x_1, e_1) \in \{0,1\} \times \mathbf{R}$ and, for $i = 2, 3, \ldots$, and $(x_1, e_1, \ldots, x_i, e_i) \in (\{0,1\} \times \mathbf{R})^i$ define

$$\tau_i(x_1, e_1, \ldots, x_i, e_i) = G\left(\frac{\overline{M}_i - \overline{N}_i}{c}\right),$$

where

$$\overline{M}_i = \frac{\sum_{j=1}^{i} x_j e_j}{\sum_{j=1}^{i} x_j}, \quad \overline{N}_i = \frac{\sum_{j=1}^{i} (1 - x_j) e_j}{\sum_{j=1}^{i} (1 - x_j)}$$

and $c$ is an appropriate scaling constant. Hence, according to $\tau$, the first patient in the trial is allocated to treatment $A$, the second to treatment $B$ and then, from the third patient on, having observed past allocations $X_1 = 0, X_2 = 1, X_3, \ldots, X_i$ and responses $M_1, N_2, X_3 M_3 + (1 - X_3) N_3, \ldots, X_i M_i + (1 - X_i) N_i$, the $(i+1)$th patient is allocated to treatment $A$ with probability $G((\overline{M}_i - \overline{N}_i)/c)$ where

$$\overline{M}_i = \frac{\sum_{j=1}^{i} X_j M_j}{\sum_{j=1}^{i} X_j} \quad \text{and} \quad \overline{N}_i = \frac{\sum_{j=1}^{i} (1 - X_j) N_j}{\sum_{j=1}^{i} (1 - X_j)} \tag{3}$$

represent the current estimates of the mean of the distributions $\mu$ and $\nu$ respectively. The sequential design $\tau$ is randomized and response-adaptive.

*Remark 5* Modifications in the definition of $\tau$ when one wants robust estimates of the means, instead of sample means as in (3), are easily implemented and have been considered in Biswas and Basu (2001). Analogously, a suitable choice of the $\phi$ transformation in the design $\sigma$ would meet robustness concerns. We do not elaborate further on this theme; it will be the topic of future research.

Clearly, the design $\tau$ favors the treatment which has led to larger responses on average in the past. In point of fact, Biswas and Basu (2001) claim that

$$P\left[\lim_{i \to \infty} \tau_i(X_1, M_1, \ldots, X_i, X_i M_i + (1 - X_i) N_i) = G\left(\frac{m_\mu - m_\nu}{c}\right)\right] = 1; \tag{4}$$

see also Bandyopadhyay and Biswas (2001). Parallel to the choice of the initialization parameter $k$ for the sequential design $\sigma$ described in the previous subsection, an important question for the determination of $\tau$ is the choice of the scaling constant $c$. Small values of $c$ make the design sensitive to outliers particularly during the early stages of the experiment. Larger values of $c$ will cause the design to be less and less adaptive to the information provided by the experiment while it is carried on, eventually pulling the allocation ratio toward the 50:50 pattern. For $\sigma$, this happens when the initialization parameter $k$ is set equal to half the trial sample size $n$.

*Remark 6* Rate of convergence in (4) is also unknown.

*Remark 7* The design $\tau$ could be extended to handle situations where responses to treatments are observed along with covariates playing the rôle of prognostic factors: see Bandyopadhyay and Biswas (2001) and Atkinson and Biswas (2005). Similar modifications to $\sigma$ have yet to be studied.

## 3 A guideline for evaluating and comparing different response-adaptive designs

The goal of a randomized, response-adaptive sequential design is twofold:

1. to generate experimental data adequate for statistical inference on treatment effects;
2. to randomly allocate patients to treatments, using the information about treatments generated along the trial for biasing the allocation probabilities toward the better treatment.

When considering a response-adaptive design for a particular clinical trial, a standard, non-adaptive alternative is usually available; generally speaking, this is the *default design* that the experimenter would implement for carrying on the trial in the absence of a response-adaptive competitor, to be followed by a *default inferential analysis* applied to the data generated by the design. Default design and default inferential analysis represent the experimenter's *default plan*. We believe that the experimenter might be persuaded to use a response-adaptive design, instead of the default design, if we show that:

(a) the response-adaptive design makes it possible to perform an inferential analysis with the same optimality characteristics as those guaranteed by the default plan; and
(b) the number of patients allocated in the trial to the worse treatment by the response-adaptive design is less than that provided by the default design.

*Remark 8* Condition (b) could be modified by focusing on a different optimality criterion. For instance, when responses are survival times, the sum of the survival times of the patients involved in the trial could generate more concern than the number of patients allocated to the worse treatment.

Informed by (a) and (b), we propose a guideline for the evaluation and comparison of a response-adaptive design: the idea is to focus on a benchmark, a basic inferential problem often encountered in practice, and to base the design's evaluation on the conditions for which requirements (a) and (b) are simultaneously satisfied. As a seminal example, in this paper we will fix our benchmark to be the problem of comparing the means of two normal distributions, with same known variance, in order to test if one mean is greater than the other: the method is however easily extendable to more complex benchmarks.

Hence, assume that the goal of the experiment is to test whether the mean response generated by treatment $A$ is equal to the mean response generated by treatment $B$ against the alternative hypothesis that $A$ generates a response with greater mean. In symbols:

$$H_0 : m_\mu = m_\nu \quad vs. \quad H_1 : m_\mu > m_\nu. \tag{5}$$

Moreover, assume that the response distributions $\mu$ and $\nu$ are Normal with common known variance $v_0^2$.

A standard design and inferential analysis for the problem (5) are to randomly allocate $n_A$ patients to treatment $A$, $n_B$ patients to treatment $B$ and, at the end of the experiment, to perform a one-sided $z$-test. For any given level $\alpha \in (0, 1)$ and any even trial sample size $n \geq 2$, the power of the test is maximized if $n_A = n_B = n/2$; we refer to this design and inference with the expression *a balanced, one-sided z-test*. In the following pages, it is going to play the role of default plan.

Suppose that the level $\alpha$ of the test is assigned and that $n$ has been chosen so that the balanced, one-sided z-test has a given power $1 - \beta$ when

$$\delta = m_\mu - m_\nu$$

is greater than or equal to a specific, clinically relevant difference $\delta_0 > 0$. In order to convince the experimenter to switch from the default design to a competitor response-adaptive design, we need:

(a) to elicit an $\alpha$-level test, function of the experimental data generated by the response-adaptive design, for proving $H_0$ versus $H_1$. For a trial sample size $n^*$, the power of the test must be at least $1 - \beta$ when $\delta \geq \delta_0$;
(b) to show that, when the trial sample size is $n^*$ and $\delta \geq \delta_0$, the random number $N_B$ of patients allocated to treatment $B$ is less than or equal to $n/2$, with high probability.

If (a) and (b) hold simultaneously, the experimenter adopting the response-adaptive design knows that the probability that a significant result will be obtained if a clinically relevant difference between the two treatments exists (i.e. the power of the test) is not less than the power of the test in the default plan at the smallest clinically relevant difference; this might happen at the cost of a number $n^*$ of patients in the trial greater than or equal to $n$, but with the assurance that, if there is a clinically relevant difference between the treatments,

with high probability less patients than those allocated by the default design will in fact experience the worse treatment.

The numerical studies illustrated in this paper evaluate and compare the two response-adaptive designs described in the previous section in the light of the benchmark inferential problem described above. To explore conditions for which requirements (a) and (b) are simultaneously met, we will in fact consider two different analysis.

### 3.1 First analysis

We fix the trial sample size $n$ of the default plan and a level $\alpha \in (0, 1)$ and we compute, as a function of the difference $\delta = m_\mu - m_\nu$, the power $1 - \beta(\delta)$ of the balanced, one-sided z-test of level $\alpha$. Next, for the competitor response-adaptive design and for different values of $\delta$, we compute through simulation the smallest trial sample size $n^* = n^*(\delta)$ such that a given $\alpha$-level test (to be specified in the next section) has power greater than or equal to the power $1 - \beta(\delta)$ of the $z$-test. We also compute through simulation the distribution of the random number $N_B$ of patients allocated to treatment $B$ by the response-adaptive design when the trial sample size is $n^* = n^*(\delta)$. The typical situation is summarized in Fig. 1.

Inspection of Fig. 1 shows the existence of three different regions for the values of $\delta$. For small values of $\delta$, $n^*$ is larger than $n$ but $N_B$ is not smaller than $n/2$ with high probability: this is the "red zone" where condition (a) above is met, but not (b). That is: in order to get the same power as that of the default plan with sample size $n$, the response-adaptive design needs a sample size $n^* > n$, but the higher cost due to a larger sample size is not compensated by a gain in terms of less patients allocated to the worse treatment. Moderate values of $\delta$ fall in the "yellow zone": (a) and (b) are met at the cost of a larger sample

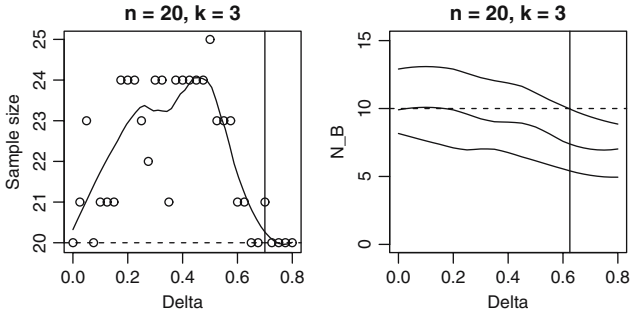


**Fig. 1** *First analysis* the picture on the left represents the function $n^* = n^*(\delta)$ by means of the values determined by simulation and a loess curve interpolating them; the *horizontal dotted line* corresponds to the sample size $n$, while the *vertical line* indicates the value $\delta_1^G$. On the right picture, the first quartile, the median and the third quartile of the distribution of $N_B$ are represented as a function of $\delta$; the *horizontal dotted line* corresponds to the value $n/2$ while the *vertical line* indicates the value $\delta_1^Y$

size $n^*$ for the response-adaptive design. Finally, large values of $\delta$ belong to the "green zone": (a) and (b) are satisfied and $n^*$ is less than or equal to $n+1$.

If the experimenter believes that values of $\delta$ in the "red zone" are clinically relevant, he shouldn't exchange the default design for the response-adaptive design. If the smallest clinically relevant value $\delta_0$ for $\delta$ falls in the "yellow zone", the experimenter might switch from the default design to the response-adaptive design, at a cost of a larger trial sample size. Finally, when $\delta_0$ falls in the "green zone" it seems unreasonable not to use the response-adaptive design.

To make these ideas more precise, we define two crucial $\delta$-values. The first is

$$\delta_1^Y = \inf\{\delta > 0 : q_3(N_B(\delta)) \leq n/2\},$$

where $q_3(N_B(\delta))$ represents the third quartile of the distribution of the random number of patients $N_B$ allocated to the worse treatment $B$ by the response-adaptive design when the trial sample size is $n^*(\delta)$. The second crucial $\delta$-value is

$$\delta_1^G = \inf\{\delta \geq \delta_1^Y : n^*(\delta) \leq n\}.$$

Assuming that $q_3(N_B(\delta))$ is a decreasing function of $\delta$, if $\delta_0$ is the smallest clinically relevant difference for the means of $\mu$ and $\nu$ and $\delta_0 < \delta_1^Y$, it is not advisable to switch from the default design with sample size $n$ to the response-adaptive design; if $\delta_0 \geq \delta_1^Y$ switching to the response-adaptive design will put less patients on the worse treatment with high probability, while preserving the same power as the balanced z-test. This however happens at the cost of a larger trial sample size if $\delta_1^Y \leq \delta_0 < \delta_1^G$.

*Remark 9* When defining $\delta_1^Y$, a less restrictive analysis would consider a smaller quantile than $q_3$ of the distribution of the random number of patients $N_B$ allocated to the worse treatment $B$ by the response-adaptive design when the trial sample size is $n^*(\delta)$. For instance,

$$\delta_1^Y = \inf\{\delta > 0 : q_2(N_B(\delta)) \leq n/2\},$$

where $q_2$ is for the median.

### 3.2 Second analysis

We fix a level $\alpha \in (0,1)$ and a large power $1-\beta$. As a function of $\delta = m_\mu - m_\nu$, we compute the smallest sample size $n$ such that a balanced, one-sided z-test of level $\alpha$ for the hypothesis in (5) has power greater than or equal to $1-\beta$. Next, through simulation, we compute the smallest sample size $n^* = n^*(\delta)$ for which a test (to be specified in the next section) based on the data generated by the competitor response-adaptive design, has power greater than or equal to $1-\beta$ when $m_\mu - m_\nu = \delta$. We then compute, again through simulation, the distribution of the random number $N_B$ of patients allocated to treatment $B$, when

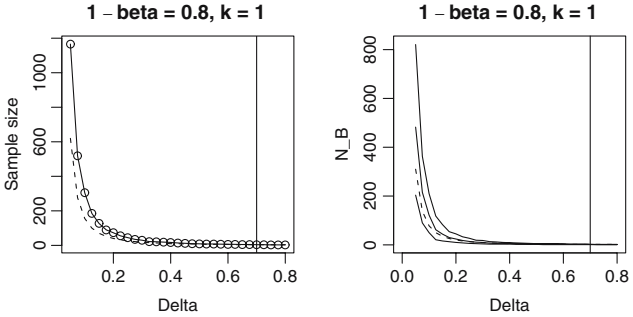


**Fig. 2** *Second analysis* for the picture on the left, the black dotted line represents $n = n(\delta)$ while the function $n^* = n^*(\delta)$ is represented by the *continuous black line*.The *vertical line* corresponds to the value $\delta_2^G$. On the right picture, the *black dotted line* represents $n/2$ for different values of $\delta$, while the three *continuous red lines* represent, respectively, the third quartile, the median and the first quartile of the distribution of $N_B$. The vertical value corresponds to the value $\delta_2^Y$

the response-adaptive design is adopted and the trial sample size is $n^* = n^*(\delta)$. Figure 2 illustrates the typical situation.

Define

$$\delta_2^Y = \inf\left\{\delta > 0 : q_3(N_B(\delta)) \leq \frac{n(\delta)}{2}\right\}$$

where $q_3(N_B(\delta)$ represents the third quartile of the distribution of the random number $N_B$ of patients allocated by the response-adaptive design to the worse treatment $B$ and the trial sample size is $n^* = n^*(\delta)$. Moreover set

$$\delta_2^G = \inf\left\{\delta \geq \delta_2^Y : n^*(\delta) \leq n(\delta)\right\}.$$

As in our first analysis, we identify three different regions for the values of $\delta$; as before they are called the "red zone", the "yellow zone" and the "green zone" and they are represented by the, possibly void, intervals $(0, \delta_2^Y), [\delta_2^Y, \delta_2^G)$ and $[\delta_2^G, \infty)$. To illustrate, assume that $\delta_0$ is the smallest clinically relevant value for the difference $\delta = m_\mu - m_\nu$. If $\delta_0 < \delta_2^Y$, the experimenter using the response-adaptive design will obtain a significant result for $\delta = \delta_0$ with probability at least $1 - \beta$ by enrolling $n^* = n^*(\delta_0)$ patients in the trial. However with such a sample size, there is a probability greater than 0.25 that the number of patients in the trial allocated to the worse treatment $B$ is larger than $n/2$, the number of patients allocated to $B$ by the default design. Hence, values of $\delta_0$ less than the value $\delta_2^Y$ falls in the "red zone" where the experimenter would be wise not to abandon the default design in favor of the response-adaptive competitor. For values of $\delta_0 \geq \delta_2^G$ the smallest trial sample size for the default plan with power greater than or equal to $1-\beta$ is negligibly different from the smallest sample size for the response-adaptive design guaranteeing a power greater than or equal to $1 - \beta$ : these values of $\delta_0$ belong to the "green zone". When $\delta_2^Y \leq \delta_2^G$, a value

of $\delta_0$ in the green zone denounces as unreasonable the adoption of the default design since, at no higher cost in terms of sample size, the response-adaptive design would put the experimenter in the position to perform an $\alpha$-level test with the same power $1-\beta$ for $\delta = \delta_0$ as the test belonging to the default plan, but allocating a smaller number of patients to the worse treatment with probability greater than 0.75. Finally, for values of $\delta_0$ in the "yellow zone" $[\delta_2^Y, \delta_2^G)$, switching from the default design to the response-adaptive competitor will allocate less patients to the worse treatment with high probability but at the cost of a larger trial sample size.

## 4 Numerical studies

We are now ready to evaluate and compare the response-adaptive designs $\sigma$ and $\tau$ defined in Sect. 2, along the guideline illustrated in the previous pages and using as benchmark inferential problem the balanced, one-sided $z$-test for testing the hypothesis (5).

Before proceeding, we need to specify a test statistic and a rejection region for the testing problem (5) when data are acquired according to one of the two response-adaptive designs under scrutiny. For a given trial sample size $n$, an obvious candidate for the test function is:

$$Z_0 = \frac{\bar{M} - \bar{N}}{v_0 \sqrt{\frac{1}{N_A} + \frac{1}{N_B}}},$$

where $N_A = \sum_{i=1}^{n} X_i$ and $N_B = n - N_A$ are the random number of patients allocated by the design to treatment $A$ and $B$, respectively, and

$$\bar{M} = \frac{1}{N_A} \sum_{i=1}^{n} X_i M_i, \quad \bar{N} = \frac{1}{N_B} \sum_{i=1}^{n} (1 - X_i) N_i$$

are the observed sample means for the responses of patients allocated to treatment $A$ and treatment $B$ respectively. (Recall that $v_0^2$ is the known common variance of the Normal distributions $\mu$ and $\nu$.)

**Proposition 1** *For both response-adaptive designs $\sigma$ and $\tau$, when $H_0$ is true and the means of $\mu$ and $\nu$ are the same, the test statistic $Z_0$ is asymptotically standard normal, as the trial sample size n goes to infinity.*

*Proof* For the asymptotic normality of $Z_0$ when the design $\tau$ is adopted, equation (4) in Bandyopadhyay and Biswas (2001) presents a result without proof conditional on the treatment assignments. Formally, the unconditional result follows from Hu and Zhang (2004).

In order to prove that $Z_0$ is asymptotically normal when the design $\sigma$ is adopted and $H_0$ is true, we need two facts that are proved in May et al. (2005):

(i) As $n$ goes to infinity, $N_A/n$ converges almost surely to a random variable $Z_\infty \in [0,1]$. The distribution of $Z_\infty$ has no point masses.
(ii) The sequence of allocation variables $\{X_n\}$ is asymptotically exchangeable with de Finetti measure $Z_\infty$. Hence: for all $j \geq 0$, conditionally on $Z_\infty$, the random variables $(X_n, \ldots, X_{n+j})$ are asymptotically i.i.d. with distribution Bernoulli$(Z_\infty)$, as $n$ grows to infinity.

From (i) and (ii) it follows that, given $Z_\infty$, the asymptotic conditional distribution of $Z_0$ is standard normal. Hence, the unconditional asymptotic distribution of $Z_0$ is also standard normal. □

For $\alpha \in (0,1)$ and a large trial sample size $n$, the previous result supports the rejection region

$$R_\alpha = \{X_1, X_1M_1 + (1-X_1)N_1, \ldots, X_n, X_nM_n + (1-X_n)N_n : Z_0 > z_{1-\alpha}\} \qquad (6)$$

for testing the hypothesis (5) at a level $\alpha$; $z_{1-\alpha}$ is the $(1-\alpha)$th quantile of a standard normal distribution. The simulations conducted for our first analysis will show that, even for small and moderate sample sizes ($n = 20, 40, 100$), $P(R_\alpha)$ is close to $\alpha$ when $H_0$ is true.

For $\delta = m_\mu - m_\nu \geq 0$ and a trial sample size $n$, let

$$\pi_\alpha(n, \delta) = P[Z_0 > z_{1-\alpha}]$$

be the power function of the test with rejection region $R_\alpha$. Recall that the power function for a balanced, one-sided z-test for the hypothesis (5) is equal to

$$1 - \Phi\left(z_{1-\alpha} - \frac{\delta}{2v_0}\sqrt{n}\right).$$

The simulations for the first and the second analysis, described in their generalities in the previous section, are carried out for the designs $\sigma$ and $\tau$ with the following common settings:

- $\mu$ and $\nu$ are normal distribution with means $m_\mu$ and $m_\nu$ respectively and common known variance $v_0^2$;
- $v_0 = 0.25$;
- $m_\mu \in [1, 1.8]$, $m_\nu = 1$;
- the level of the test is set to be $\alpha = 0.05$;
- for the design $\sigma$, the transformation function

$$\phi(x) = \begin{cases} 0.1 & \text{if } x \leq 0.1, \\ x \quad \text{if } 0.1 & < x < 10, \\ 10 & \text{if } x \geq 10; \end{cases}$$

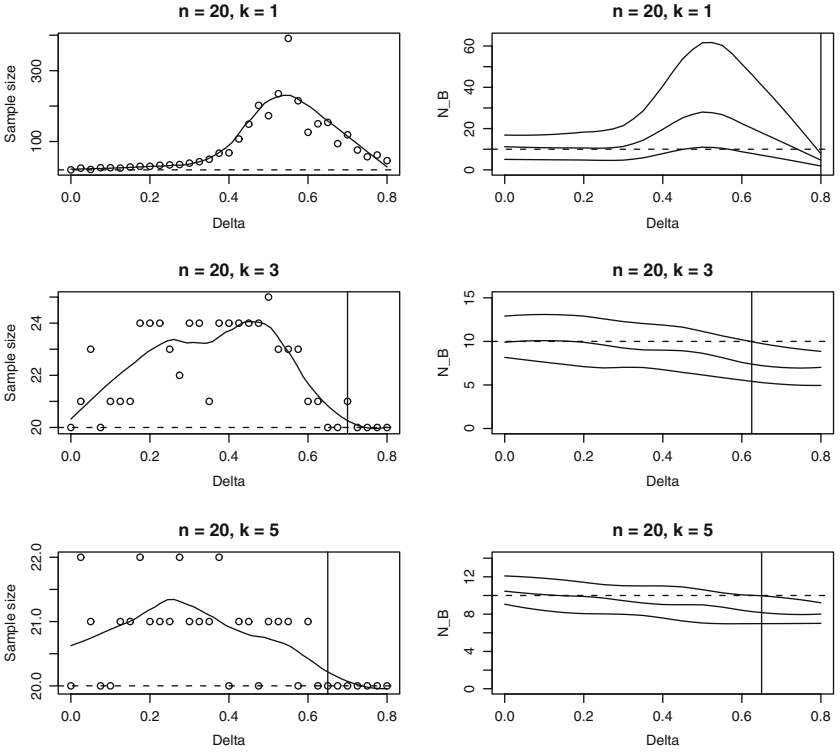


**Fig. 3** *Design* $\sigma$, *first analysis* $n = 20, k = 1, 3, 5$. Pictures on the left side represent the function $n^* = n^*(\delta)$ by means of the values determined by simulation and a loess curve interpolating them; the *horizontal dotted line* corresponds to the sample size $n$, while the *vertical line* indicates the value $\delta_1^G$. On the right side, the first quartile, the median and the third quartile of the distribution of $N_B$ are represented as a function of $\delta$; the *horizontal dotted line* corresponds to the value $n/2$ while the *vertical line* indicates the value $\delta_1^Y$

hence $\phi$ is morally the identity, given that, for all admissible choices of their means, both $\mu$ and $\nu$ assign probability close to 1 to the interval $[0.1, 10]$. The initialization parameter $k$ is set equal to $1, 3, 5$;

- for the design $\tau$, $G$ is set equal to the cumulative distribution function $\Phi$ of a standard normal and the normalizing constant $c = 1, 5, 10$.

## 4.1 Simulations for the first analysis

We fix $n = 20, 40, 100$. For both response-adaptive designs $\sigma$ and $\tau$, $\delta \in [0, 0.8]$, first we find the smallest $n^* = n^*(\delta)$ such that

$$\pi(n^*, \delta) \geq 1 - \Phi\left(z_{1-\alpha} - \frac{\delta}{2\nu_0}\sqrt{n}\right) \tag{7}$$

with $\delta \in [0, 0.8]$.

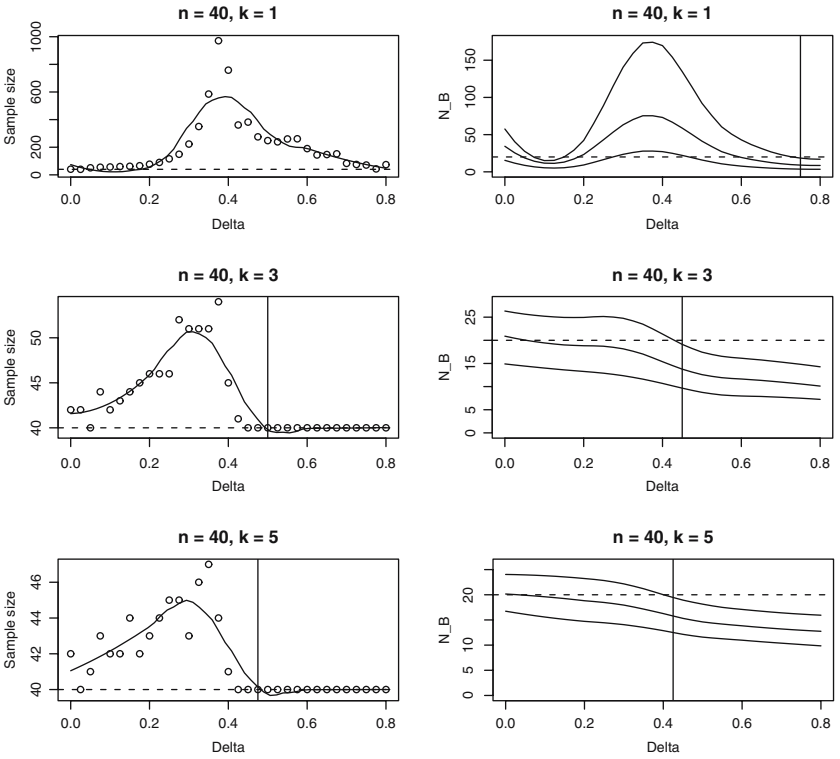


**Fig. 4** *Design* $\sigma$, *first analysis* $n = 40, k = 1, 3, 5$. Pictures on the left side represent the function $n^* = n^*(\delta)$ by means of the values determined by simulation and a loess curve interpolating them; the *horizontal dotted line* corresponds to the sample size $n$, while the *vertical line* indicates the value $\delta_1^G$. On the right side, the first quartile, the median and the third quartile of the distribution of $N_B$ are represented as a function of $\delta$; the *horizontal dotted line* corresponds to the value $n/2$ while the *vertical line* indicates the value $\delta_1^Y$

**Table 1** *First analysis*: values of $\delta_1^G$ and $\delta_1^Y$

| Design $\sigma$ | $\delta_1^G$ | $\delta_1^Y$ | Design $\tau$ | $\delta_1^G$ | $\delta_1^Y$ |
|---|---|---|---|---|---|
| $n = 20, k = 1$ | >0.8 | 0.8 | $n = 20, c = 1$ | 0.725 | 0.325 |
| $n = 20, k = 3$ | 0.7 | 0.625 | $n = 20, c = 5$ | 0.8 | 0.8 |
| $n = 20, k = 5$ | 0.65 | 0.65 | $n = 20, c = 10$ | >0.8 | >0.8 |
| $n = 40, k = 1$ | >0.8 | 0.75 | $n = 40, c = 1$ | 0.350 | 0.225 |
| $n = 40, k = 3$ | 0.5 | 0.45 | $n = 40, c = 5$ | 0.625 | 0.625 |
| $n = 40, k = 5$ | 0.475 | 0.425 | $n = 40, c = 10$ | >0.8 | >0.8 |
| $n = 100, k = 1$ | >0.8 | 0.6 | $n = 100, c = 1$ | 0.275 | 0.125 |
| $n = 100, k = 3$ | 0.375 | 0.325 | $n = 100, c = 5$ | 0.45 | 0.45 |
| $n = 100, k = 5$ | 0.325 | 0.25 | $n = 100, c = 10$ | >0.8 | >0.8 |

The integer $n^*$ is found via simulation: for $\delta \in \{0, 0.025, 0.05, \ldots, 0.775, 0.8\}$, we begin the simulation by setting $n^* = n$ and we estimate the power $\pi(n^*, \delta)$ by iteratively running an experiment where $n^*$ patients are sequentially allocated

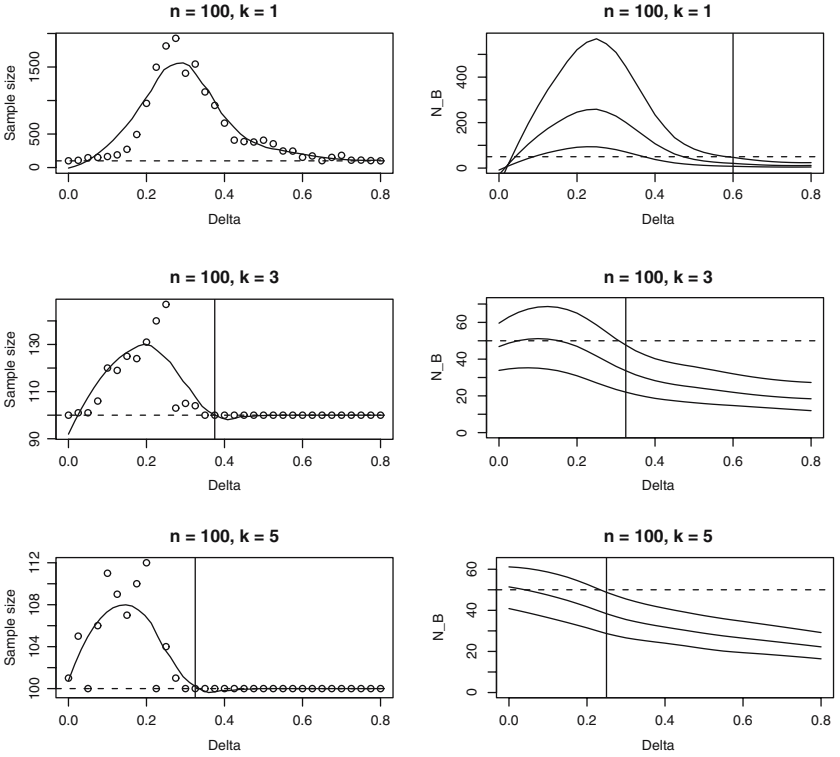


**Fig. 5** *Design* $\sigma$, *first analysis* $n = 40, k = 1, 3, 5$. Pictures on the left side represent the function $n^* = n^*(\delta)$ by means of the values determined by simulation and a loess curve interpolating them; the *horizontal dotted line* corresponds to the sample size $n$, while the *vertical line* indicates the value $\delta_1^G$. On the right side, the first quartile, the median and the third quartile of the distribution of $N_B$ are represented as a function of $\delta$; the *horizontal dotted line* corresponds to the value $n/2$ while the *vertical line* indicates the value $\delta_1^Y$

**Table 2** *Second analysis*: values of $\delta_2^G$ and $\delta_2^Y$

| Design $\sigma$ | $\delta_2^G$ | $\delta_2^Y$ | Design $\tau$ | $\delta_2^G$ | $\delta_2^Y$ |
|---|---|---|---|---|---|
| $1-\beta = 0.8, k = 1$ | >0.5 | >0.5 | $1-\beta = 0.8, c = 1$ | 0.2 | 0.2 |
| $1-\beta = 0.8, k = 3$ | 0.41 | 0.41 | $1-\beta = 0.8, c = 5$ | 0.57 | 0.57 |
| $1-\beta = 0.8, k = 5$ | 0.32 | 0.32 | $1-\beta = 0.8, c = 10$ | 0.67 | 0.67 |
| $1-\beta = 0.9, k = 1$ | >0.5 | >0.5 | $1-\beta = 0.9, c = 1$ | 0.22 | 0.2 |
| $1-\beta = 0.9, k = 3$ | 0.42 | 0.42 | $1-\beta = 0.9, c = 5$ | 0.64 | 0.64 |
| $1-\beta = 0.9, k = 5$ | 0.36 | 0.36 | $1-\beta = 0.9, c = 10$ | 0.79 | 0.79 |
| $1-\beta = 0.95, k = 1$ | >0.5 | >0.5 | $1-\beta = 0.95, c = 1$ | 0.29 | 0.2 |
| $1-\beta = 0.95, k = 3$ | 0.46 | 0.46 | $1-\beta = 0.95, c = 5$ | 0.74 | 0.74 |
| $1-\beta = 0.95, k = 5$ | 0.38 | 0.38 | $1-\beta = 0.95, c = 10$ | >0.8 | >0.8 |

to treatment $A$ or treatment $B$ according to the response-adaptive design while, at the end of the experiment, the hypothesis $H_0$ is rejected or accepted according to the test statistic $Z_0$ and the rejection region $R_\alpha$ in (6). The experiment is

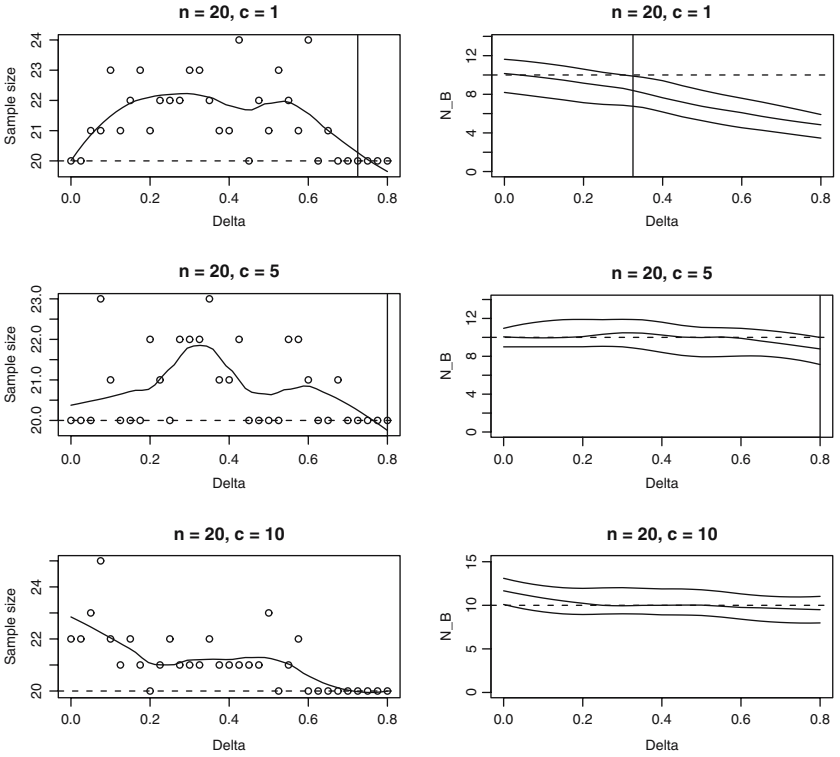


**Fig. 6** *Design* $\tau$, *first analysis* $n = 20, c = 1, 5, 10$. Pictures on the left side represent the function $n^* = n^*(\delta)$ by means of the values determined by simulation and a loess curve interpolating them; the *horizontal dotted line* corresponds to the sample size $n$, while the *vertical line* indicates the value $\delta_1^G$. On the right side, the first quartile, the median and the third quartile of the distribution of $N_B$ are represented as a function of $\delta$; the *horizontal dotted line* corresponds to the value $n/2$ while the *vertical line* indicates the value $\delta_1^Y$

iterated 1,000 times and the frequency of rejection is considered as an estimate of $\pi(n^*, \delta)$; if this frequency is strictly smaller than $1 - \Phi(z_{1-\alpha} - \frac{\delta}{2v_0}\sqrt{n})$, the integer $n^*$ is increased by one and a new experiment, with the new value for $n^*$, is iteratively run. And so on until an $n^*$ satisfying (7) is found. Next, for $\delta \in \{0, 0.025, 0.05, \ldots, 0.775, 0.8\}$, the distribution of $N_B$ when the trial sample size is $n^* = n^*(\delta)$ is approximated by the empirical distribution generated by running for 1,000 times the experiment where $n^*$ patients are allocated to treatment $A$ or $B$ according to the response-adaptive design; the first quartile, the median and the third quartile of the empirical distribution of the 1,000 deviates thus obtained are considered to be estimates of the corresponding quantiles for the distribution of $N_B$. Results of these simulations for the design $\sigma$ are illustrated in Figs. 3, 4, 5, those for $\tau$ in Figs. 6, 7, 8; all the empirical points have been interpolated by means of a polynomial local regression using a loess of degree 2: fitting is by weighted least square, with span equal to 0.55 and tricubic

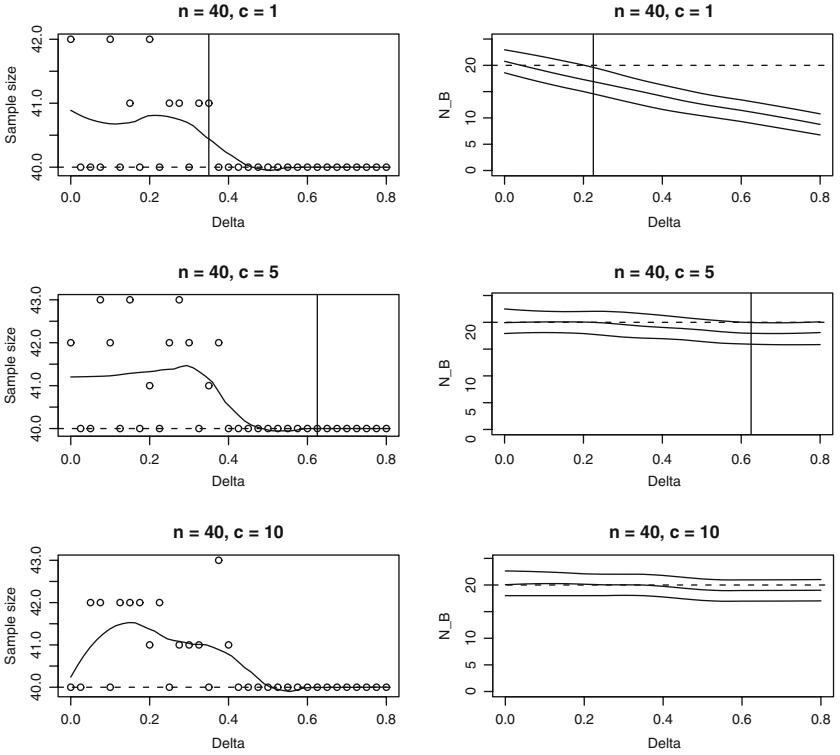


**Fig. 7** *Design* $\tau$, *first analysis* $n = 40, c = 1, 5, 10$. Pictures on the left side represent the function $n^* = n^*(\delta)$ by means of the values determined by simulation and a loess curve interpolating them; the *horizontal dotted line* corresponds to the sample size $n$, while the *vertical line* indicates the value $\delta_1^G$. On the right side, the first quartile, the median and the third quartile of the distribution of $N_B$ are represented as a function of $\delta$; the *horizontal dotted line* corresponds to the value $n/2$ while the *vertical line* indicates the value $\delta_1^Y$

weight. The quantities $\delta_1^Y$ and $\delta_1^G$ have been computed on the loess curves; their values appear in Table 1.

## 4.2 Simulations for the second analysis

We fix $1 - \beta = 0.8, 0.9, 0.95$. For $\delta \in [0.025, 0.8]$ we find the smallest value of $n = n(\delta)$ such that

$$1 - \Phi\left(z_{1-\alpha} - \frac{\delta}{2v_0}\sqrt{n}\right) \geq 1 - \beta$$

and the smallest $n^* = n^*(\delta)$ such that

$$\pi(n^*, \delta) \geq 1 - \beta. \tag{8}$$

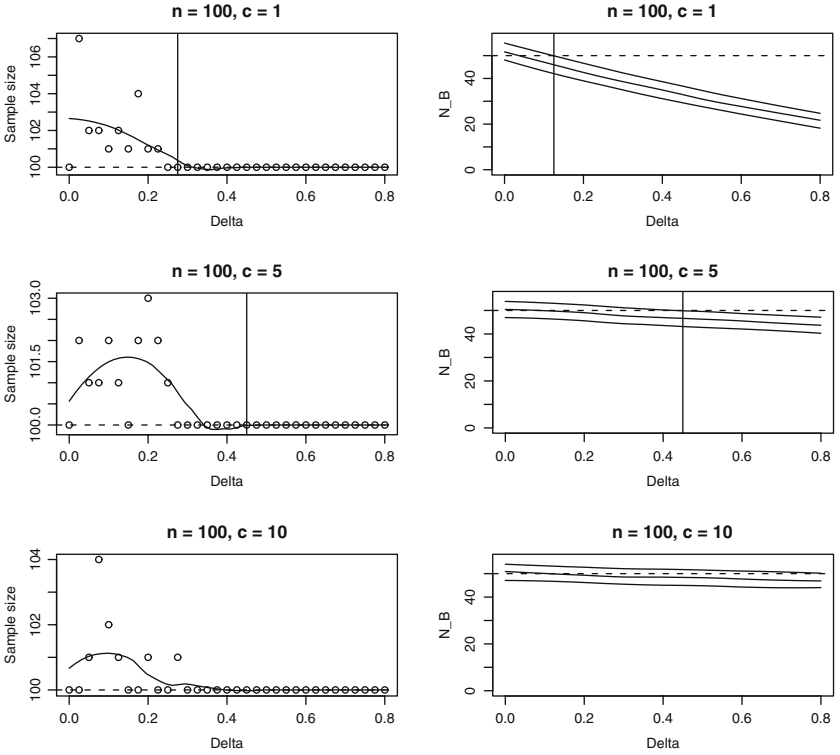


**Fig. 8** *Design* $\tau$, *first analysis* $n = 100, c = 1, 5, 10$. Pictures on the left side represent the function $n^* = n^*(\delta)$ by means of the values determined by simulation and a loess curve interpolating them; the *horizontal dotted line* corresponds to the sample size $n$, while the *vertical line* indicates the value $\delta_1^G$. On the right side, the first quartile, the median and the third quartile of the distribution of $N_B$ are represented as a function of $\delta$; the *horizontal dotted line* corresponds to the value $n/2$ while the *vertical line* indicates the value $\delta_1^Y$

The integer $n^*$ is found via simulation. For $\delta \in \{0.025, 0.05 \ldots, 0.775, 0.8\}$, we begin by setting $n^* = n_0$ : for $\sigma$ the initial value $n_0 = \max(n, 2k)$ while for $\tau$ we let $n_0 = \max(n, 2)$. We then estimate the power $\pi(n^*, \delta)$ by iteratively running an experiment where $n^*$ patients are sequentially allocated to treatment $A$ or treatment $B$ according to the response-adaptive design and the hypothesis $H_0$ is rejected or accepted according to the test statistic $Z_0$ and the rejection region $R_\alpha$ in (6). The experiment is iterated 5,000 times and the frequency of rejection is considered as an estimate of $\pi(n^*, \delta)$; if this frequency is strictly smaller than $1 - \beta$, the integer $n^*$ is increased by one and a new experiment is iteratively run. And so on until an $n^*$ satisfying (8) is found. Simulations than proceed as before: the third quartile, the median and the first quartile of the distribution of $N_B = N_B(\delta)$, when the trial sample size is $n^*(\delta)$, are estimated by means of the corresponding empirical quantiles based on 5,000 deviates generated by the distribution of $N_B(\delta)$, for $\delta \in \{0.025, 0.05 \ldots, 0.775, 0.8\}$. Results of these

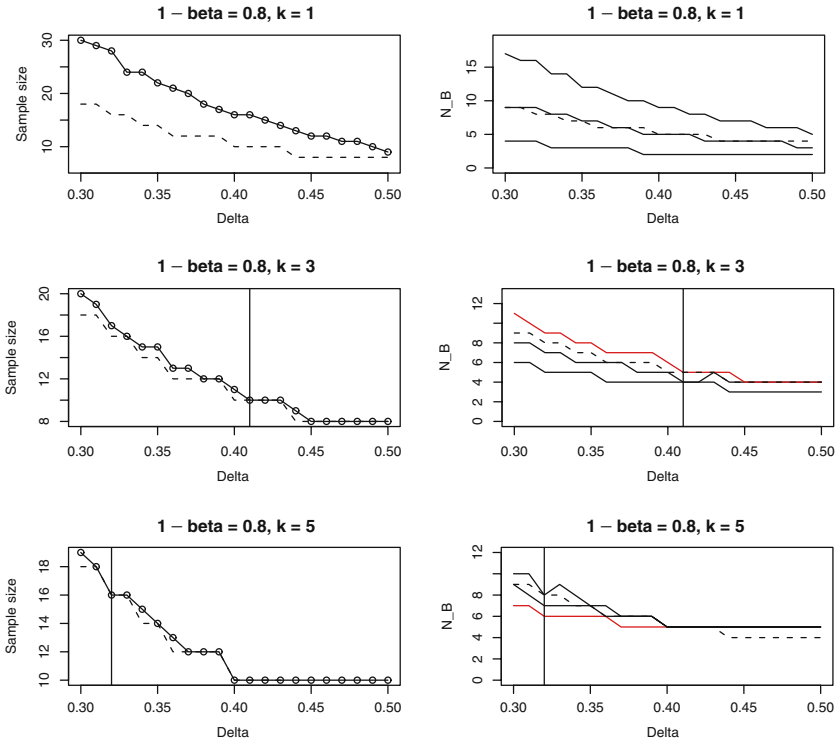


**Fig. 9** *Design* $\sigma$, *second analysis* $1-\beta = 0.8,\ k = 1,3,5$. Pictures on the left: the *black dotted line* represents $n = n(\delta)$ while the function $n^* = n^*(\delta)$ is represented by the *continuous black line*. *Vertical lines* correspond to the value $\delta_2^G$. Pictures on the right: the *black dotted line* represents $n/2$ for different values of $\delta$, while the three *continuous red lines* represent, respectively, the third quartile, the median and the first quartile of the distribution of $N_B$. *Vertical lines* correspond to the value $\delta_2^Y$

simulations, for the design $\sigma$ are illustrated in Figs. 9, 10, 11 and for $\tau$ in Figs. 12, 13, 14; for graphical convenience we focus the figures on suitable ranges for $\delta$. These results suggest that both functions $n^*(\delta)$ and $q_3(N_B(\delta))$ are monotonically decreasing with $\delta$; we interpolated the empirical points by joining them with linear segments. The quantities $\delta_2^Y$ and $\delta_2^G$ have been computed on the interpolated curves; their values appear in Table 2. For the design $\sigma$, notice that $n^*(\delta) \geq 2k$ for all values of $\delta$, where the integer $k$ is the initialization parameter.

## 5 Discussion

The simulations studies described in the previous section give important information about the two designs $\sigma$ and $\tau$ and the parameters $k$ and $c$ determining them.

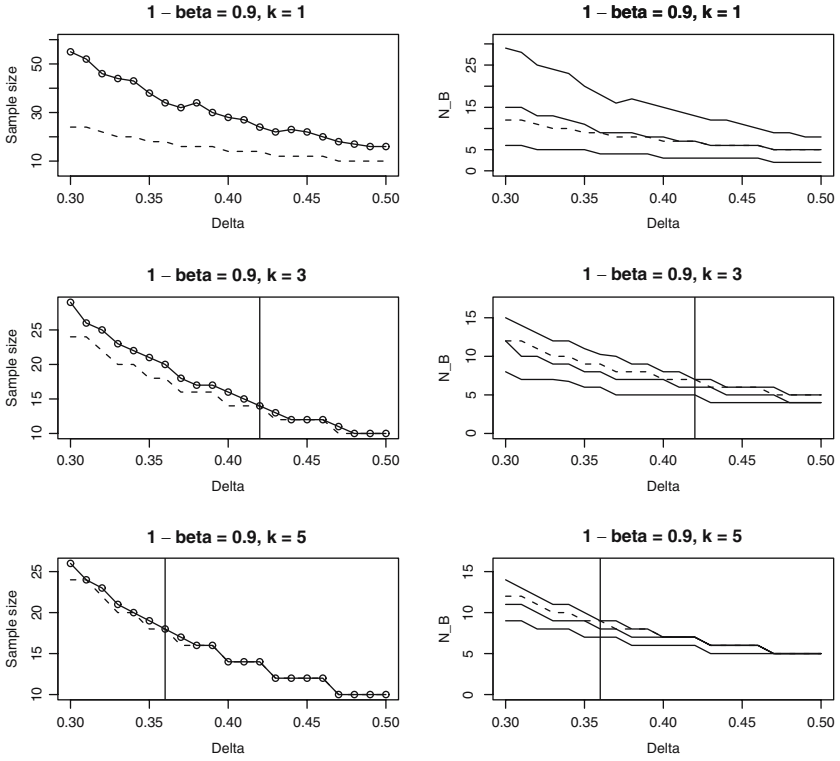


**Fig. 10** *Design* $\sigma$, *second analysis* $1-\beta = 0.9,\ k = 1,3,5$. Pictures on the left: the *black dotted line* represents $n = n(\delta)$ while the function $n^* = n^*(\delta)$ is represented by the *continuous black line*. *Vertical lines* correspond to the value $\delta_2^G$. Pictures on the right: the *black dotted line* represents $n/2$ for different values of $\delta$, while the three *continuous red lines* represent, respectively, the third quartile, the median and the first quartile of the distribution of $N_B$. *Vertical lines* correspond to the value $\delta_2^Y$

## 5.1 First analysis

Larger values of the initialization parameter $k$ for $\sigma$, or of the scaling constant $c$ for $\tau$, move the corresponding design toward a balanced allocation; as a consequence we might expect that both $\delta_1^G(\sigma)$ and $\delta_1^G(\tau)$ are decreasing functions of $k$ and $c$, respectively, for all sample sizes $n$ of the reference default plan. This fact is confirmed by the simulations as long as $\delta_1^G > \delta_1^Y$. However, the analogies between the roles of the two parameters stop here.

In fact higher values of $c$ for the design $\tau$ imply more robustness to outliers, as has been pointed out in Biswas and Basu (2001), but also larger values for $\delta_1^Y(\tau)$; in other words, since $\delta_1^Y(\tau)$ is an increasing function of $c$, for a large $c$ the number $N_B$ of patients allocated to the inferior treatment is smaller than $n/2$ with high probability only when the difference $\delta$ between the means of the two treatments is large, maybe larger than the smallest clinically relevant difference

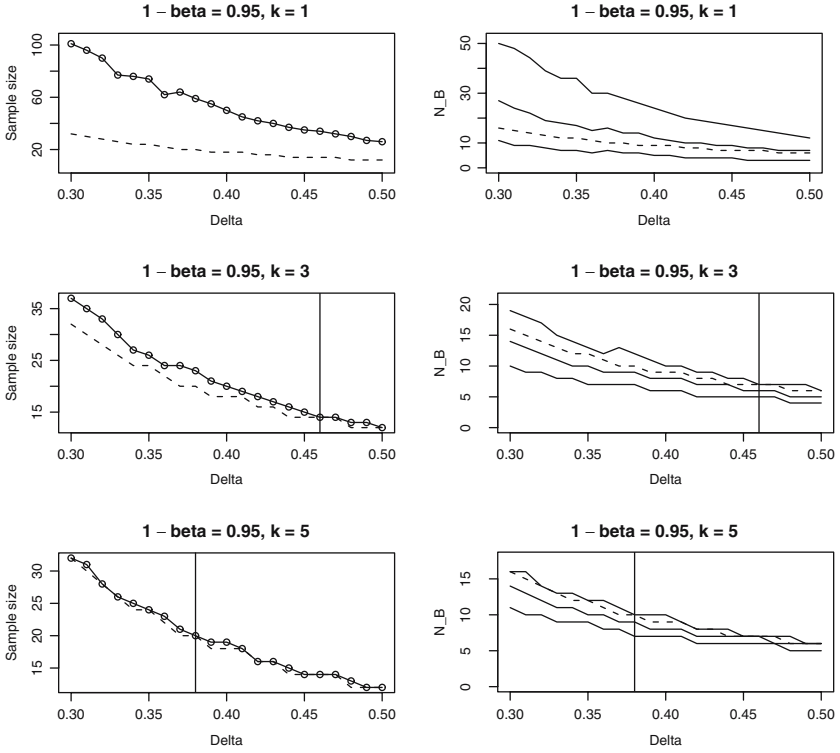


**Fig. 11** *Design* $\sigma$, *second analysis* $1-\beta = 0.95,\ k = 1,3,5$. Pictures on the left: the *black dotted line* represents $n = n(\delta)$ while the function $n^* = n^*(\delta)$ is represented by the *continuous black line*. *Vertical lines* correspond to the value $\delta_2^G$. Pictures on the right: the *black dotted line* represents $n/2$ for different values of $\delta$, while the three *continuous red lines* represent, respectively, the third quartile, the median and the first quartile of the distribution of $N_B$. *Vertical lines* correspond to the value $\delta_2^Y$

$\delta_0$. The variability of the distribution of $N_B$, measured by its interquartile range, is also affected by changes in the values of $c$, larger values of $c$ implying a smaller IQR for the distribution of $N_B$.

The role that the initialization parameter $k$ plays in the determination of the critical value $\delta_1^Y(\sigma)$ is more complicated. The simulations show that for small values of $k$, $\delta_1^Y(\sigma)$ is a decreasing function of $k$. However for larger values of $k$, $\delta_1^Y(\sigma)$ must be an increasing function of $k$ and $\delta_1^Y(\sigma) = \infty$ in the limit case when $k = n/2$. In fact, when the reference sample size of the default plan is $n = 20$, the simulation supports the conjecture that $\delta_1^Y(\sigma)$, as a function of $k$, reaches its minimum for $k$ close to 3. For $n = 40$ and $n = 100$, the value of $k$ for which $\delta_1^Y(\sigma)$ is minimum is conjectured to be larger than 5, hence in both cases we observe a decreasing sequence of values for $\delta_1^Y(\sigma)$, for $k = 1,3,5$. Moreover; the IQR of the distribution of $N_B$ seems to be a decreasing function of $k$. Both facts discourage the use of very small values, like $k = 1$, for the initialization

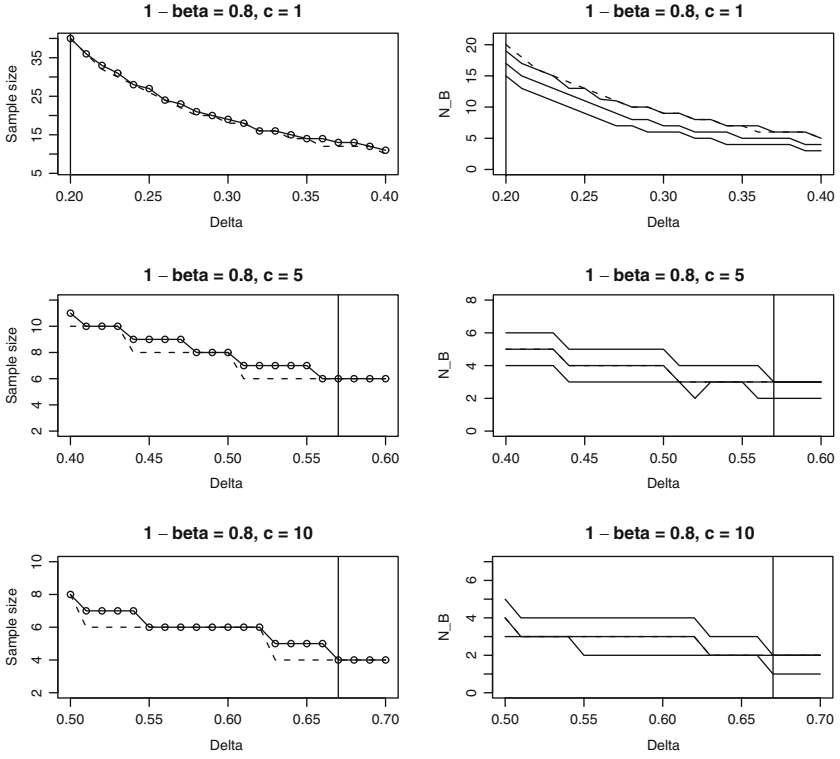


**Fig. 12** *Design* $\tau$, *second analysis* $1-\beta = 0.8,\ c = 1, 5, 10$. Pictures on the left: the *black dotted line* represents $n = n(\delta)$ while the function $n^* = n^*(\delta)$ is represented by the *continuous black line*. Vertical lines correspond to the value $\delta_2^G$. Pictures on the right: the *black dotted line* represents $n/2$ for different values of $\delta$, while the three *continuous red lines* represent, respectively, the third quartile, the median and the first quartile of the distribution of $N_B$. *Vertical lines* correspond to the value $\delta_2^Y$

parameter of the design $\sigma$ and suggest the existence of an optimal value for $k$ depending on the sample size $n$ of the default plan.

For both designs $\sigma$ and $\tau$, once the respective parameters $k$ or $c$ are fixed, the values $\delta_1^G$ and $\delta_1^Y$ are decreasing functions of $n$, the reference sample size of the default plan. Hence for clinical trials where the planned sample size $n$ of the default plan is large, it is to be expected that both response-adaptive designs $\sigma$ and $\tau$ could be effective alternatives, even in cases where the smallest clinically relevant difference $\delta_0$ is small. By looking at the values of $\delta_1^Y$ when the sample size of the default plan is $n = 100$, one can conclude that, at the possible cost of a larger sample size for the trial, the design $\sigma$ becomes a viable alternative to a balanced design for values of the smallest clinically relevant difference $\delta_0$ greater than 0.25; if one is willing to choose a small value for the scaling constant $c$, the design $\tau$ is better than a balanced design even for values of $\delta_0$ greater than 0.12.

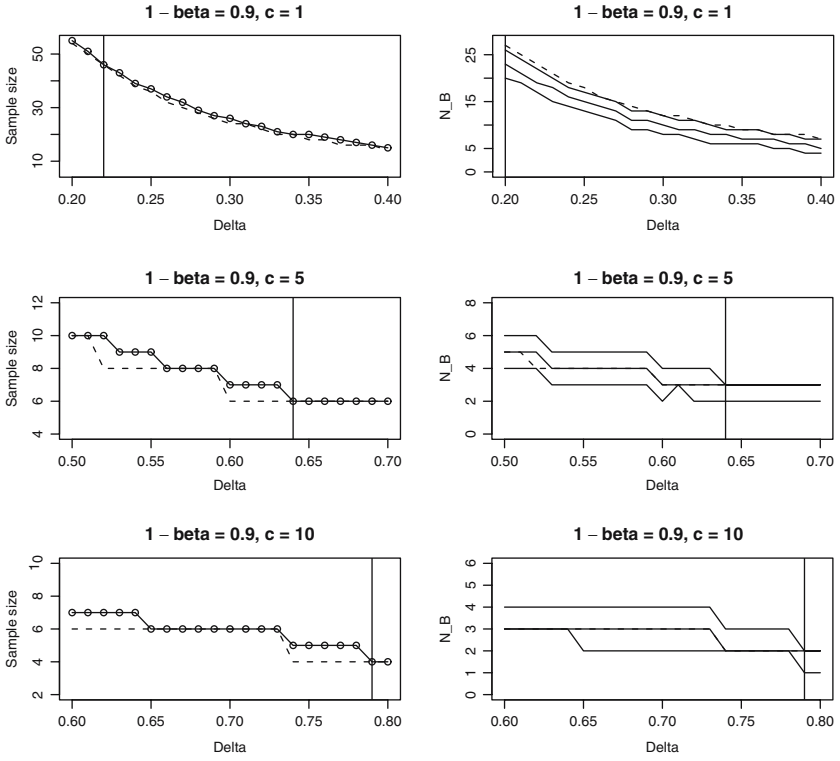


**Fig. 13** *Design* $\tau$, *second analysis* $1-\beta = 0.9,\ c = 1, 5, 10$. Pictures on the left: the *black dotted line* represents $n = n(\delta)$ while the function $n^* = n^*(\delta)$ is represented by the *continuous black line*. *Vertical lines* correspond to the value $\delta_2^G$. Pictures on the right: the *black dotted line* represents $n/2$ for different values of $\delta$, while the three *continuous red lines* represent, respectively, the third quartile, the median and the first quartile of the distribution of $N_B$. *Vertical lines* correspond to the value $\delta_2^Y$

## 5.2 Second analysis

The results of the second analysis confirm the findings of the first.

As it was to be expected, all the simulations show that for both designs $\sigma$ and $\tau$, the critical value $\delta_2^G$ is an increasing function of the power $1-\beta$. For a fixed power $1-\beta$, both $\delta_2^G(\sigma)$ and $\delta_2^G(\tau)$ are non increasing functions of the initialization parameter $k$ and the scaling constant $c$, respectively, as long as $\delta_2^G > \delta_2^Y$.

The critical value $\delta_2^Y(\tau)$ increases with $c$ for any value of the power $1-\beta$; hence a stronger protection against outliers is paid in terms of larger values for the smallest difference $\delta$ for which the response-adaptive design becomes a viable alternative to the default design. For a fixed value of $c$, the simulations support the conjecture that $\delta_2^Y(\tau)$ is an increasing function of $1-\beta$.

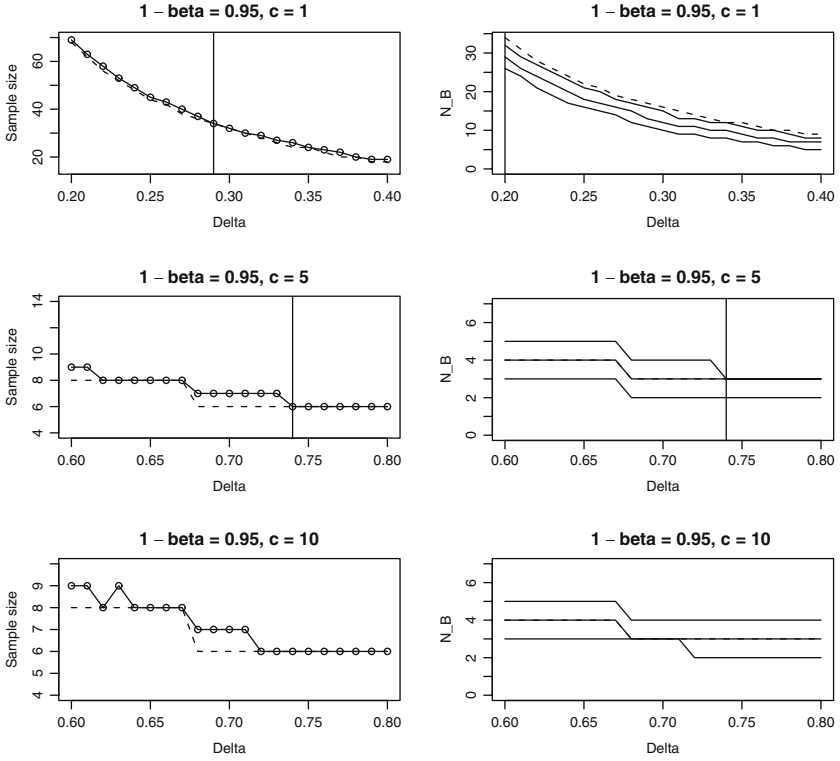


**Fig. 14** *Design* $\tau$, *second analysis* $1-\beta = 0.95,\ c = 1, 5, 10$. Pictures on the left: the *black dotted line* represents $n = n(\delta)$ while the function $n^* = n^*(\delta)$ is represented by the *continuous black line*. *Vertical lines* correspond to the value $\delta_2^G$. Pictures on the right: the *black dotted line* represents $n/2$ for different values of $\delta$, while the three *continuous red lines* represent, respectively, the third quartile, the median and the first quartile of the distribution of $N_B$. *Vertical lines* correspond to the value $\delta_2^Y$

For the design $\sigma$ it is confirmed that very small values of the initialization parameter $k$ are to be discouraged: in fact, for $1-\beta = 0.8$, the value $\delta_2^Y(\sigma)$ seems to reach a minimum for values of $k$ close to 5, while this minimum is attained for values of $k$ larger than 5 when $1-\beta > 0.9$: this findings will stimulate future research on the optimal choice for the parameter $k$.

This second analysis seems slightly in favor to the design $\tau$: for moderate values of the scaling constant $c$, the values of $\delta_2^Y(\tau)$ are generally smaller than those obtained with the design $\sigma$, even when we believe we are dealing with a value for the initialization parameter $k$ close to the optimum ($1-\beta = 0.8, k = 5$).

## 6 Conclusion

Research on response-adaptive designs for clinical trials with continuous responses is gaining momentum. In this study we compared two different

response-adaptive designs for continuous responses by introducing a criterion that evaluates them by considering their performances when competing with a benchmark, non-adaptive standard alternative.

Over all the two designs seem to have similar merits; when confronted with a balanced, one-sided $z$-test, both look like viable alternatives when the smallest clinically relevant difference $\delta_0$ between the means of the two responses assumes moderate to large values. In fact, when power is the main concern, our findings are slightly in favor of the design $\tau$ proposed by Bandyopadhyay and Biswas (2001) and Biswas and Basu (2001), implemented with a small or moderate scaling constant $c$. In contrast, the appeal of the design $\sigma$ generated by the randomly reinforced urn of Muliere et al. (2006a) stems from the fact that it targets an asymptotic allocation where all the patients are given the better treatment, a property not shared by the design $\tau$.

## References

Aletti G, May C, Secchi P (2005) On the distribution of the limit proportion for a two-color, randomly reinforced urn with equal reinforcement distributions. MOX report no. 74, MOX-Dipartimento di Matematica, Politecnico di Milano

Atkinson AC, Biswas A (2005) Bayesian adaptive biased-coin design for clinical trials with normal repsonses. Biometrics 61:118–125

Bandyopadhyay U, Biswas A (2001) Adaptive designs for normal responses with prognostic factors. Biometrika 88:409–419

Biswas A, Basu A (2001) Robust adaptive designs in clinical trials for continuous responses. Sankhya Ser B 63:27–42

DeGroot MH (1970) Optimal statistical decisions. McGraw-Hill, New York

Durham SC, Flournoy N, Li W (1998) A sequential design for maximizing the probability of a response. Can J Stat 26(3):479–495

Fishburn PC (1981) Subjective expected utility: a review of normative theories. Theory Decis 13:139–199

Flournoy N, Rosenberger WF (eds) (1995) Adaptive designs. IMS, Hayward

Hu F, Rosenberger WF (2003) Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparisons. J Am Stat Assoc 98:671–678

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

May C, Paganoni AM, Secchi P (2005) On a two-color generalized Polya urn. Metron, LXIII: 115–134

Muliere P, Paganoni AM, Secchi P (2006a) A randomly reinforced urn. J Stat Plann Inference 136:1853–1874

Muliere P, Paganoni AM, Secchi P (2006b) Randomly reinforced urns for clinical trials with continuous responses. SIS—Proceedings of the XLIII Scientific Meeting, Invited Session 9, CLEUP Padova, pp 403–414

Rosenberger WF (1996) New directions in adaptive designs. Stat Sci 11:137–149

Rosenberger WF (2002) Randomized urn models and sequential design. Seq Anal 21:1–28

Rosenberger WF, Lachin JM (2005) Randomization in clinical trials. Wiley, New York

Wei LJ, Durham S (1978) The randomized play-the-winner rule in medical trials. J Am Stat Assoc 73:840–843