ARCHIVIO ISTITUZIONALE
DELLA RICERCA

Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Compound optimal allocations for survival clinical trials

This is the final peer-reviewed author's accepted manuscript (postprint) of the following publication:

Published Version:
Baldi Antognini, A., Novelli, M., Zagoraiou, M., Vagheggini, A. (2020). Compound optimal allocations for survival clinical trials. BIOMETRICAL JOURNAL, 62(7 (November)), 1730-1746 [10.1002/bimj.201900232].

Availability:
This version is available at: https://hdl.handle.net/11585/765822 since: 2020-12-03

Published:
DOI: http://doi.org/10.1002/bimj.201900232

Terms of use:

Some rights reserved. The terms and conditions for the reuse of this version of the manuscript are specified in the publishing policy. For all terms of use and more information see the publisher's website.

This item was downloaded from IRIS Università di Bologna (https://cris.unibo.it/).
When citing, please refer to the published version.

(Article begins on next page)

24 July 2024

Compound optimal allocations for survival clinical trials

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

[1] Department of Statistical Sciences, University of Bologna, Bologna, Italy
[2] Unit of Biostatistics and Clinical Trials, Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori (IRST) IRCCS, Meldola (FC), Italy

Received zzz, revised zzz, accepted zzz

The aim of the present paper is to provide optimal allocations for comparative clinical trials with survival outcomes. The suggested targets are derived adopting a compound optimization strategy based on a subjective weighting of the relative importance of inferential demands and ethical concerns. The ensuing compound optimal targets are continuous functions of the treatment effects, so we provide the conditions under which they can be approached by standard response-adaptive randomization procedures, also guaranteeing the applicability of the classical asymptotic inference. The operating characteristics of the suggested methodology are verified both theoretically and by simulation, including the robustness to model misspecification. With respect to the other available proposals, our strategy always assigns more patients to the best treatment without compromising inference, taking into account estimation efficiency and power as well. We illustrate our procedure by redesigning two real oncological trials.

Key words: Censoring; Ethics; Hypothesis testing; Oncological trials; Response-Adaptive Randomization.

1 Introduction

In the biomedical and pharmaceutical research for treatment comparisons, randomized trials are commonly considered to be the gold standard. Indeed, a randomization component in the assignments tends to mitigate several types of bias, including the accidental bias due to unknown covariates/confounders and the selection bias induced by the investigators. Moreover, another important issue in clinical trials is the ethical conflict between the care for the well-being of the subjects involved in the trial and the rigorous pursuit of further knowledge (Royall, 1991).

The competing ethical and inferential goals can be analytically characterized into suitable optimization problems allowing to derive target allocations which may act as a legitimate compromise (see, e.g., Rosenberger et al., 2001; Biswas et al., 2007; Biswas and Bhattacharya, 2009; Baldi Antognini and Giovagnoli, 2010). In general, they depend on the unknown model parameters and can be targeted by appropriate Response-Adaptive Randomization (RAR) procedures (see, e.g., Hu and Rosenberger, 2006; Baldi Antognini and Giovagnoli, 2015), namely a class of sequential allocation rules which at each step change the probabilities of the treatment assignments on the basis of the accumulating information, with the aim of skewing allocations toward the superior treatment.

In oncological phase III comparative trials, the primary endpoint is usually to evaluate the efficacy of an experimental treatment against a control one in terms of survival. These studies commonly rely on a 1:1 randomization without particular attention of patients' care, i.e., characterized by a low ethical concern

[*]Correspondence: Alessandro Baldi Antognini, Department of Statistical Sciences, University of Bologna, Via Belle Arti 41, 40126, Bologna, Italy.
e-mail: a.baldi@unibo.it

© 0 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

since any patient has the same probability of being assigned to each treatment despite the accumulating information. Furthermore, they have to span over a long period in order to observe the time-to-event response variable which is by definition delayed (i.e., not instantaneously observable). Early theoretical attempts to apply RAR procedures to time-to-event endpoints date back to the mid 1990s (Yao and Wei, 1996; Hallstrom et al., 1996; Rosenberger and Seshaiyer, 1997). By the use of non-parametric test statistics, they mapped the treatment effects to the unit interval in order to skew the allocation probabilities towards the arm displaying longer survival. Their results highlighted increased ethics without compromising the statistical power. Unfortunately, the inferential properties of these early proposals were far from optimal. More recently, Zhang and Rosenberger (2007) proposed a parametric approach to modelling the survival allowing to analytically derive optimal targets to be attained by means of RAR procedures. Their results highlighted a power increase with respect to the usual 1:1 randomization along with a higher proportion of patients assigned to the better treatment. Biswas and Mandal (2004); Sverdlov et al. (2011) further extended the range of practical applications, also by taking into account multiple arm trials; their extensive simulation study shows results analogous to the two-arm's ones.

The present work aims at deriving an optimal target able to increase the ethical gain in trials with time-to-event responses. In particular, we propose a combined optimization approach that, by using either fixed or flexible weights (governing the relative importance of ethics wrt inference), leads to an optimal target allocation depending on the unknown treatment effects. The properties of the suggested target allocation have been theoretically studied, also providing the conditions for the applicability of RAR methods as well as of the standard asymptotic inference. Simulated results along with a redesign of two real clinical trials have been employed to highlight the ethical gain induced by our proposal, which may help practitioners in the planning phase. R software was used to obtain the numerical results and to carry out the simulations; the source code to reproduce them is available as Supporting Information on the journal's web page (http://onlinelibrary.wiley.com/doi/xxx/suppinfo).

In Section 2 the notation and the model are introduced along with the optimal targets for both ethics and inference. Section 3 contains the theoretical results on the compound optimal target, whereas Section 4 generalizes the previous results by taking into account right censoring. Section 5 illustrates the analytical properties of our procedure, while the operating characteristics of the newly introduced target are investigated by an extensive simulation study in Section 6, including type-I error control and robustness to model misspecifications. In Section 7 our methodology is used to redesign two real oncological trials. Some practical recommendations end the paper.

2 Notation and model

Let A and B be two competing treatments. Suppose that subjects enter the trial sequentially and at each step only one treatment will be assigned according to a given randomization rule. At each step $i \geq 1$, let δ_i be the indicator managing the allocation of the ith subject, with $\delta_i = 1$ if he/she is assigned to A and 0 otherwise. Given the treatment assignments, patient's responses (i.e., survival times) Ys relative to either treatment are assumed to be independent and identically distributed (*iid*) belonging to the exponential distribution with pdf $f(y;\theta_j) = \theta_j^{-1}\exp(-y/\theta_j)$, where $\theta_j \in \mathbb{R}^+$ denotes the effect of treatment j ($j = A, B$). In this setting, the treatment outcomes are stochastically ordered on the basis of their effects and without loss of generality we assume that high responses are preferable for patient's care. As is well-known, the exponential distribution is one of the most commonly used model in the context of survival analysis, since the hazard rates θ_A^{-1} and θ_B^{-1} are constant over time. Now we start our exposition with this simple setup, that will be later generalized by taking into account censoring.

At the end of the trial with n subjects, let $N_{An} = \sum_{i=1}^n \delta_i$ and $N_{Bn} = n - N_{An}$ be the assignments to both treatments, so that $\rho = n^{-1}N_{An}$ is the allocation proportion to A (respectively, $1-\rho$ to B). Then, the MLEs of the treatment effects coincide with the sample means, namely $\hat{\theta}_{An} = N_{An}^{-1}\sum_{i=1}^n \delta_i Y_i$ and $\hat{\theta}_{Bn} = N_{Bn}^{-1}\sum_{i=1}^n (1-\delta_i)Y_i$.

© 0 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

2.1 Optimal allocation for inference

Assuming that the inferential interest consists in estimating/testing the superiority of a given treatment (say A) with respect to the gold standard (B), the parameter of interest is the difference $\theta_A - \theta_B$ between the treatment effects or, alternatively, their ratio $\gamma = \theta_A/\theta_B$, while θ_B is usually regarded as a nuisance. When the aim consists in optimal estimation, typically the goal is to minimize the variance of the estimated treatment difference

$$V(\hat{\theta}_{An} - \hat{\theta}_{Bn} \mid \rho) = n^{-1} \left\{ \frac{\theta_A^2}{\rho} + \frac{\theta_B^2}{1-\rho} \right\}. \tag{1}$$

As is well-known, for every sample size n the variance in (1) is minimized by the Neyman allocation $\rho_N = \theta_A/(\theta_A + \theta_B)$.

Remark 2.1 Notice that $\rho_N = \rho_N(\gamma) = \gamma/(\gamma+1)$ and thus $\rho_N(\gamma) = 1 - \rho_N(\gamma^{-1})$ for every $\gamma \in \mathbb{R}^+$. Namely, by expressing ρ_N in terms of the ratio γ between the treatment effects, it induces a symmetric structure of the target around the point $\rho_N(1) = 1/2$. Therefore, contrary to the binary case, under this setting the Neyman allocation is ethical, since it always assigns the majority of patients to the best treatment.

Moreover, by minimizing (1), ρ_N is also optimal for hypothesis testing, since it maximizes the power of the usual t-test employed for testing $H_0 : \theta_A = \theta_B$ against $H_1 : \theta_A > \theta_B$. Indeed, due to the asymptotic normality of the classical Wald statistic

$$W_n = \frac{\sqrt{n}(\hat{\theta}_{An} - \hat{\theta}_{Bn})}{\sqrt{\frac{\hat{\theta}_{An}^2}{\rho} + \frac{\hat{\theta}_{Bn}^2}{1-\rho}}},$$

the power of the test of level α can be approximated by $\Phi\left(\frac{\sqrt{n}(\theta_A - \theta_B)}{\sqrt{\frac{\theta_A^2}{\rho} + \frac{\theta_B^2}{1-\rho}}} - z_{1-\alpha} \right)$ for $\theta_A - \theta_B \geq 0$, i.e.,

$$\Phi\left(\frac{\sqrt{n}(\gamma-1)\sqrt{\rho(1-\rho)}}{\sqrt{\rho(1-\gamma^2)+\gamma^2}} - z_{1-\alpha} \right), \quad \gamma \geq 1, \tag{2}$$

where Φ denotes the cdf of the standard normal r.v. and z_α is the α-percentile of Φ. For every n, ρ_N maximizes power (2) and the same conclusion still holds even for the two-sided alternative $H_1 : \theta_A \neq \theta_B$, where the power is an increasing function of the non-centrality parameter $n(\theta_A - \theta_B)^2/[\theta_A^2\rho^{-1} + \theta_B^2(1-\rho)^{-1}]$ of a non-central chi-square distribution with 1 degree of freedom.

Since $V(\hat{\theta}_{An} - \hat{\theta}_{Bn} \mid \rho_N) = n^{-1}(\theta_A + \theta_B)^2 \leq V(\hat{\theta}_{An} - \hat{\theta}_{Bn} \mid \rho)$, from now on we can use

$$\mathcal{C}_I(\rho) = \frac{V(\hat{\theta}_{An} - \hat{\theta}_{Bn} \mid \rho_N)}{V(\hat{\theta}_{An} - \hat{\theta}_{Bn} \mid \rho)} = \frac{(\theta_A + \theta_B)^2}{\frac{\theta_A^2}{\rho} + \frac{\theta_B^2}{1-\rho}} = \frac{(\gamma+1)^2\rho(1-\rho)}{\rho(1-\gamma^2)+\gamma^2} \in [0;1] \tag{3}$$

as a criterion of inferential efficiency that should be maximized.

2.2 Ethically optimal target

Dealing with design criteria aimed at measuring the ethical cost of the clinical trial, one of the most popular is the proportion of patients receiving the best treatment, namely

$$\mathcal{C}_E(\rho) = \rho \mathbb{I}_{\{\gamma>1\}} + (1-\rho)\mathbb{I}_{\{\gamma<1\}} = \begin{cases} \rho, & \text{if } \gamma > 1, \\ 1-\rho, & \text{if } \gamma < 1, \end{cases} \tag{4}$$

© 0 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

where $\mathbb{I}_Q$ denotes the indicator function of the event Q. The ethical efficiency $\mathcal{C}_E(\rho) \in [0;1]$ is a linear function of the allocation proportion ρ (increasing or decreasing on the basis of the superiority/inferiority of A wrt B) and is clearly maximized by assigning all subjects to the best treatment, namely by $\rho_E = \mathbb{I}_{\{\gamma>1\}}$, under which $\mathcal{C}_E$ achieves its optimum value equal to 1. Contrary to $\mathcal{C}_I(\rho)$ in (3), criterion $\mathcal{C}_E(\rho)$ depends only on the sign of $\gamma - 1$, but it does not depend on the magnitude of γ. Moreover, if $\gamma = 1$ there is no longer a best treatment and therefore every allocation could be considered as ethically equivalent.

Remark 2.2 Within this setting, another possible ethical measure to be maximized is the total expected outcomes, i.e., $\tilde{\mathcal{C}}_E(\rho) = \rho\theta_A + (1-\rho)\theta_B$. However, provided that $\gamma \neq 1$, $\tilde{\mathcal{C}}_E(\rho)$ is simply a re-scaled version of $\mathcal{C}_E(\rho)$, namely it provides the same information in the scale $[\min\{\theta_A, \theta_B\}; \max\{\theta_A, \theta_B\}] = \theta_B[\min\{\gamma, 1\}; \max\{\gamma, 1\}]$ instead of $[0;1]$. For these reasons, from now on $\mathcal{C}_E(\rho)$ in (4) will be used as the ethical criterion to be maximized.

3 Compound Optimality

Ideally, we would simultaneously maximize both the inferential efficiency and the ethical one. However, $\mathcal{C}_I$ competes directly with $\mathcal{C}_E$ and the problem is how to achieve a good trade-off between inferential goals and ethical demands. In this setting a natural solution consists in assuming a compound optimization approach, namely by formalizing a compound criterion that combines the competing objectives with suitable weights. Since $\mathcal{C}_I$ and $\mathcal{C}_E$ are efficiency measures (i.e., these criteria are non-negative functions, lying in [0;1], that should be maximized), a natural compromise consists in taking the following combination

$$\mathcal{C}_\omega(\rho) = \omega\mathcal{C}_E(\rho) + (1-\omega)\,\mathcal{C}_I(\rho), \tag{5}$$

as the compromise criterion to be maximized, where $\omega \in [0;1]$ represents the relative importance of ethics wrt inference. Clearly, the ethical weight can be selected by the investigator according to the relative importance of $\mathcal{C}_E$ and $\mathcal{C}_I$ in the actual clinical trial. The choice $\omega = 0$ indicates complete concern about inference (so the resulting optimal target coincides with ρ_N), while a weight of $\omega = 1$ indicates that ethics is the only goal (leading to ρ_E). Besides a fixed constant, ω could also be chosen according to the values of γ, due to the fact that more attention must be paid to ethics when the treatment effects differ significantly, while inferential care could be crucial when θ_A is close to θ_B, since it is harder to discriminate between the two treatments (in such a case the ethical cost is quite low). So, we shall assume $\omega = \omega(\gamma) : \mathbb{R}^+ \to [0;1]$ to be a continuous function such that:

i) $\omega(\gamma) = \omega(\gamma^{-1})$ for every $\gamma \in \mathbb{R}^+$, in order to deal with the treatments symmetrically;

ii) $\omega(1) = 0$ and ω increasing in γ for $\gamma > 1$ for ethical reasons.

For instance, by modifying the cdf of the standard log-normal distribution, we could assume

$$\omega_a(\gamma) = \begin{cases} 2\Phi(\ln(\gamma^{-a})) - 1, & \text{if } 0 < \gamma < 1, \\ 2\Phi(\ln(\gamma^{a})) - 1, & \text{if } \gamma \geq 1, \end{cases} \tag{6}$$

where $a \geq 1$ is a tuning parameter managing the ethical skew. Weighting functions ω_a in (6) are plotted in Figure 1 for $a = 1, 2, 3$ and 4.

Remark 3.1 Clearly, additional weighting functions can be constructed by opportunely modifying the cdf of any log-symmetric distribution (Seshadri, 1965; Jones, 2008). Moreover, since in the clinical practice small differences between the treatment effects are often assumed to be negligible up to a given threshold, usually called the minimal clinically relevant effect, weight (6) could be extended accordingly.

Since $\mathcal{C}_\omega(\rho)$ is differentiable wrt ρ, the compound optimal target $\rho^*_\omega = \arg\max \mathcal{C}_\omega(\rho)$ can be easily found by (5) via simple algebra.

© 0 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

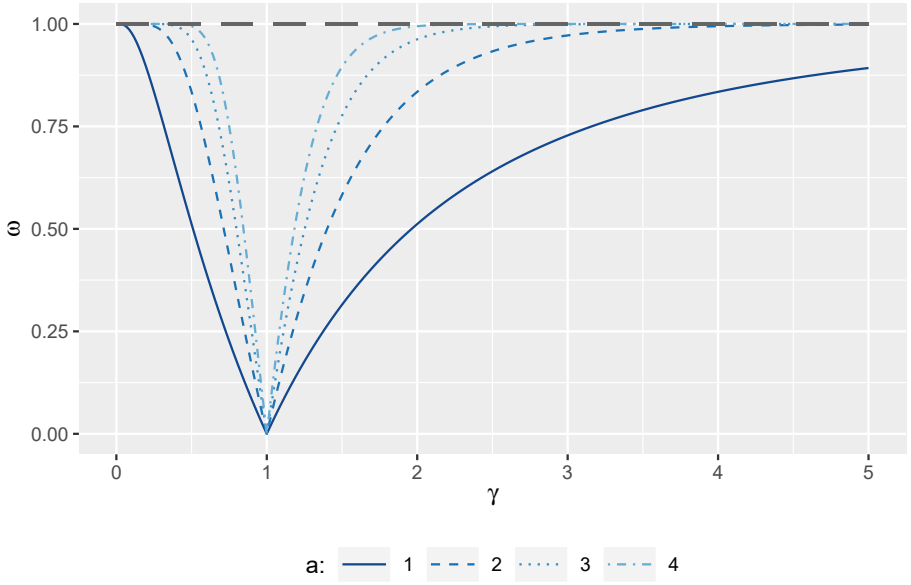

Figure 1 Log-Normal weighting function $\omega_a(\gamma)$ as a varies.

Theorem 3.2 *For any choice of the weight ω, the optimal compound target $\rho^*_\omega \in [0;1]$ is unique. If $\omega \geq [1+(1-\max\{\rho_N; 1-\rho_N\})^2]^{-1}$, then $\rho^*_\omega = \rho_E$; otherwise, if $\omega < [1+(1-\max\{\rho_N; 1-\rho_N\})^2]^{-1}$,*

$$\rho^*_\omega = \frac{\rho_N^2\beta + \gamma\sqrt{1+\beta(2\rho_N-1)}}{1+\beta(2\rho_N-1)+\gamma\sqrt{1+\beta(2\rho_N-1)}}, \tag{7}$$

where $\beta = [\omega/(1-\omega)]\mathrm{sgn}(\gamma-1)$. Moreover,

P1: *ρ^*_ω always assigns more subjects to the best treatment wrt ρ_N,*

P2: *by choosing $\omega < 4/5$, then $\rho^*_\omega \in (0;1)$,*

P3: *ρ^*_ω is monotonically increasing in ω,*

P4: *by choosing $\omega/(1-\omega) < 4/\sqrt{3}$ (i.e., $\omega < 0.698$), then ρ^*_ω is monotonically increasing in γ.*

P r o o f. See the Appendix. □

The suggested compound optimal target is a continuous function of the unknown model parameters and therefore it is a priori unknown. However, from property P2, this allocation can be approached asymptotically by using suitable RAR procedures - like, e.g., the Doubly-Adaptive Biased Coin Design (DBCD) (see Hu and Zhang, 2004) - namely sequential allocation rules that change at each step the probabilities of treatment assignments on the basis of the accrued information (i.e., the available responses and past allocations), in order to converge to the chosen target. After a starting sample of subjects assigned to either treatment, at each step the treatment effects are estimated by MLEs and therefore γ and ρ^*_ω are estimated accordingly. Thus, the next assignment is forced to approach the target increasingly as the distance between the current allocation proportion and the estimated target grows.

© 0 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Remark 3.3 Although the optimal target in (7) is not twice-differentiable, which implies that the treatment allocation proportion of the DBCD is not asymptotically normally distributed, ρ^*_ω is a continuous function of the unknown parameters, which guarantees - together with P2 - the applicability of standard RAR methodology, as well as the classical likelihood-based asymptotic inference. Indeed, the treatment allocation proportion converges almost surely to ρ^*_ω and the consistency of the MLEs is guaranteed along with their asymptotic normality (Baldi Antognini and Zagoraiou, 2015; Baldi Antognini and Giovagnoli, 2015). However, to preserve the asymptotic normality of the design, a smoothing transformation of the optimal target could be applied, as described in Tymofyeyev et al. (2007).

4 Survival trials with right censoring

Since a common feature of survival data is the presence of censoring, now we extend the exponential model introduced in Section 2 by assuming that the patients are subject to an independent right censoring scheme. In particular, for the ith individual let (Y_i, D_i) be the pair of outcomes that will be observed, where D_i is the indicator of not-censoring (namely the indicator of the event of interest, like death/failure) that is assumed to be independent of Y_i, so that $(Y_i, D_i) = (y, 1)$ when the ith patient is not censored (namely an event is observed) and y represents its survival time, while for $(Y_i, D_i) = (y, 0)$ the patient is censored and y is the censoring time. Within this setting, the MLE of the treatment effect θ_j $(j = A, B)$ is $\hat{\theta}_{jn} = S^Y_{jn}/S^D_{jn}$, where S^Y_{jn} and S^D_{jn} denote the total observed survival time and the total number of failures for group j, respectively, i.e., $S^Y_{An} = \sum_{i=1}^n \delta_i Y_i$, $S^Y_{Bn} = \sum_{i=1}^n (1-\delta_i) Y_i$, $S^D_{An} = \sum_{i=1}^n \delta_i D_i$ and $S^D_{Bn} = \sum_{i=1}^n (1-\delta_i) D_i$.

Instead of (1), the variance of the estimated treatment difference becomes

$$V(\hat{\theta}_{An} - \hat{\theta}_{Bn} \mid \rho) = n^{-1} \left\{ \frac{\theta_A^2}{\rho p_A} + \frac{\theta_B^2}{(1-\rho) p_B} \right\}, \tag{8}$$

where, for any $i = 1, \ldots, n$, $p_A = E[D_i \mid \delta_i = 1] = \Pr(D_i = 1 \mid \delta_i = 1)$ and $p_B = E[D_i \mid \delta_i = 0] = \Pr(D_i = 1 \mid \delta_i = 0)$ are the probabilities that a failure occurs before censoring in the two groups, that are assumed to be constant for each subject assigned to the same treatment (Wald statistic W_n as well as its power (2) should be modified accordingly).

Clearly, p_j depends on θ_j and the particular censoring scheme adopted in the trial (that involves the duration of the experiment, the recruitment period, the chosen follow-up as well as some model assumptions about the patient entries and the censoring time). For instance, one of the most general censoring schemes is described in Lawless (2011) and Zhang and Rosenberger (2007). More specifically, suppose patients entry times are staggered and *iid* uniformly distributed over the recruitment period R, $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{U}[0, R]$. Since for practical reasons each patient may be observed for a maximum follow-up time $S - R$, where S is the total study duration, individuals are subjected to independent censoring times *iid* uniformly distributed over the duration of the trial, namely $C_1, \ldots, C_n \overset{iid}{\sim} \mathcal{U}[0, S]$, independently of $X_1, \ldots, X_n$. Thus,

$$p(\theta_j) = 1 - \frac{\theta_j}{S} + \exp\left\{-\frac{S}{\theta_j}\right\} \frac{\theta_j}{SR} \left(\exp\left\{\frac{R}{\theta_j}\right\} (2\theta_j - R) - 2\theta_j \right), \qquad j = A, B. \tag{9}$$

In general, since the censoring scheme is common to the two treatments, from now on we assume that $p_j = p(\theta_j)$, where $p(\cdot) : \mathbb{R}^+ \to [0; 1]$ is a non-increasing function such that $\lim_{x \to 0} p(x) = 1$ and $\lim_{x \to \infty} p(x) = 0$ (stressing that the probability of observing a failure should be decreasing in the expected lifetime, regardless of the chosen censoring).

Remark 4.1 Notice that this is a simple model re-parameterization of (1) with $\tilde{\theta}_j = \theta_j / \sqrt{p(\theta_j)}$ and therefore $\tilde{\gamma} = \tilde{\theta}_A / \tilde{\theta}_B$. It is straightforward to show that, for every sample size n, the variance in (8) as well as the power are optimized by the Neyman allocation $\tilde{\rho}_N = \tilde{\theta}_A / (\tilde{\theta}_A + \tilde{\theta}_B)$, which is still ethical,

© 0 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

since it assigns the majority of patients to the best treatment (indeed, if $\theta_A > \theta_B$ then $\tilde{\theta}_A > \tilde{\theta}_B$, due to the decreasingness of $p(\cdot)$, and therefore $\tilde{\rho}_N > 1/2$) (see also Sverdlov et al., 2011). Obviously, in the absence of censoring $D_i = 1$ for any i (i.e., $p(x) = 1\ \forall x$), so that $\tilde{\theta}_j = \theta_j$ ($j = A, B$) and therefore $\tilde{\rho}_N = \rho_N$. Finally note that, in the presence of censoring, the form of the inferential criterion $\mathcal{C}_I(\rho)$ is still the same as in (3) with $\tilde{\gamma}$ instead of γ.

5 Numerical comparison among target allocations

This section is dedicated to a performance comparison between the new proposal, ρ^*_ω, and other targets proposed in the literature for survival trials. In particular, we compare the behavior of the Neyman allocation $\tilde{\rho}_N$, the target proposed by Zhang and Rosenberger (2007),

$$\tilde{\rho}_{ZR} = \frac{\sqrt{\theta_A^3\, p(\theta_B)}}{\sqrt{\theta_A^3\, p(\theta_B)} + \sqrt{\theta_B^3\, p(\theta_A)}}$$

and the one suggested by Biswas and Mandal (2004)

$$\tilde{\rho}_{BM} = \frac{\theta_A \sqrt{p(\theta_B)\left[1 - \exp\left(-c/\theta_B\right)\right]}}{\theta_A \sqrt{p(\theta_B)\left[1 - \exp\left(-c/\theta_B\right)\right]} + \theta_B \sqrt{p(\theta_A)\left[1 - \exp\left(-c/\theta_A\right)\right]}},$$

where the survival time is dichotomized and c represents the threshold above which the outcome is considered to be a success.

To assess the performance of the considered targets several criteria will be taken into account. More specifically, from a statistical point of view we will consider the power of Wald test (the significance level is always set equal to 5%) and the estimation efficiency $\mathcal{C}_I(\rho)$, while $\mathcal{C}_E(\rho)$ in (4) will be adopted as a measure of ethics. In this section we study the analytical properties of the suggested optimal target by assuming the parameter values as known. The operating characteristics of our target implemented via DBCD will be discussed in Section 6 also showing its practical implications by redesigning the trial on metastatic breast cancer in Jones et al. (2005) in Section 7.

Firstly, we consider the scenario without censoring, that is $p(\theta_j) = 1$ for $j = A, B$. Table 1 summarizes the ethical behavior of the considered targets as γ varies, with $\theta_A \in [10; 20]$ and $\theta_B = 10$ months, respectively. For the optimal compound target, we take into account fixed ethical weights $\omega \in \{0.3, 0.4, 0.5, 0.6, 0.69\}$, as well as the weighting function ω_a in (6) with $a = 1,\ 1.5$ and 2, while for ρ_{BM} we adopt the thresholds $c = 9, 12$. Clearly, from now on the weighting function ω_a will be always re-scaled via $4/(4 + \sqrt{3})$, in order to satisfy property P4.

Table 2 shows the power of Wald test adopting the considered targets as γ varies with a sample size $n = 250$, while Table 3 summarizes the results in terms of estimation efficiency $\mathcal{C}_I(\rho)$ in (3).

The newly introduced target shows good performance from both ethical and inferential viewpoints. As an example, with $\omega = 0.4$, ρ^*_ω induces a consistent ethical improvement if compared to the other targets (up to 9% more patients are assigned to the best treatment wrt ρ_N, and up to 7% and 8% wrt ρ_{ZR} and ρ_{BM}, respectively), guaranteeing substantially the same performances in terms of power and with a maximum loss in estimation efficiency of 3%. Clearly, the ethical gain of ρ^*_ω grows as ω increases; for instance, with $\omega \in [0.5, 0.6]$, ρ^*_ω assigns up to 18% more patients to the best treatment wrt the Neyman target and up to 16% and 17% for ρ_{ZR} and ρ_{BM}, respectively. Note that this is particularly evident for low values of γ making ρ^*_ω a valid choice in oncological trials where ethics might play a crucial role.

While the behaviour of ρ^*_ω changes considerably as ω varies, moving from an allocation close to ρ_N for low values of ω to a much stronger ethical one for high values of ω, ρ_{BM} appears to be only slightly dependent on the choice of the threshold c (this is also confirmed by other results, not showed here, obtained with different values of c). Targets ρ_{ZR} and ρ_{BM} show comparable performances in both statistical power and estimation precision, with a more ethical assignment wrt ρ_N especially for high values of γ. Finally,

© 0 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 1 Proportion of patients assigned to the best treatment as γ varies.

			ρ_{BM}		ρ^*_ω as ω varies					$\rho^*_{\omega_a}$		
γ	ρ_N	ρ_{ZR}	$c=9$	$c=12$	0.3	0.4	0.5	0.6	0.69	$a=1$	$a=1.5$	$a=2$
1	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
1.1	0.52	0.54	0.53	0.53	0.58	0.61	0.64	0.70	0.78	0.53	0.53	0.54
1.2	0.55	0.57	0.56	0.56	0.60	0.62	0.66	0.71	0.79	0.56	0.57	0.58
1.3	0.57	0.60	0.59	0.58	0.62	0.64	0.68	0.73	0.79	0.59	0.60	0.61
1.4	0.58	0.62	0.61	0.61	0.63	0.66	0.69	0.74	0.80	0.61	0.63	0.64
1.5	0.60	0.65	0.63	0.63	0.65	0.67	0.70	0.75	0.80	0.63	0.65	0.67
1.6	0.62	0.67	0.65	0.65	0.66	0.69	0.72	0.76	0.81	0.65	0.68	0.70
1.7	0.63	0.69	0.67	0.67	0.68	0.70	0.73	0.77	0.81	0.67	0.70	0.73
1.8	0.64	0.71	0.69	0.68	0.69	0.71	0.74	0.77	0.82	0.69	0.72	0.75
1.9	0.66	0.72	0.70	0.70	0.70	0.72	0.75	0.78	0.82	0.71	0.74	0.77
2	0.67	0.74	0.72	0.71	0.71	0.74	0.76	0.79	0.83	0.72	0.75	0.78

Table 2 Approximated power of Wald test as γ varies.

			ρ_{BM}		ρ^*_ω as ω varies					$\rho^*_{\omega_a}$		
γ	ρ_N	ρ_{ZR}	$c=9$	$c=12$	0.3	0.4	0.5	0.6	0.69	$a=1$	$a=1.5$	$a=2$
1	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
1.1	0.19	0.19	0.19	0.19	0.19	0.19	0.18	0.17	0.16	0.19	0.19	0.19
1.2	0.42	0.42	0.42	0.42	0.41	0.41	0.40	0.38	0.34	0.42	0.42	0.42
1.3	0.66	0.66	0.66	0.66	0.66	0.66	0.64	0.62	0.56	0.66	0.66	0.66
1.4	0.84	0.84	0.84	0.84	0.84	0.83	0.82	0.80	0.75	0.84	0.84	0.83
1.5	0.94	0.93	0.93	0.93	0.94	0.93	0.92	0.91	0.88	0.93	0.93	0.93
1.6	0.98	0.98	0.98	0.98	0.98	0.98	0.97	0.97	0.95	0.98	0.98	0.97
1.7	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98	0.99	0.99	0.99
1.8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00
1.9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 3 Estimation efficiency as γ varies.

		ρ_{BM}		ρ^*_ω as ω varies					$\rho^*_{\omega_a}$		
γ	ρ_{ZR}	$c=9$	$c=12$	0.3	0.4	0.5	0.6	0.69	$a=1$	$a=1.5$	$a=2$
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.2	1.00	1.00	1.00	0.99	0.97	0.94	0.88	0.75	1.00	1.00	1.00
1.4	0.99	1.00	1.00	0.99	0.98	0.95	0.89	0.78	1.00	0.99	0.98
1.6	0.99	0.99	1.00	0.99	0.98	0.95	0.90	0.81	0.99	0.98	0.97
1.8	0.98	0.99	0.99	0.99	0.98	0.96	0.91	0.83	0.99	0.97	0.94
2	0.97	0.99	0.99	0.99	0.98	0.96	0.92	0.84	0.99	0.96	0.93

as far as the estimation efficiency is concerned, all targets perform well; for $\omega \leq 0.4$ the maximum loss in efficiency is about 3%, with a substantial equivalence to ρ_N for $\omega \leq 0.3$. Although for $\omega \geq 0.6$ the estimation efficiency induced by ρ^*_ω is slightly lower wrt ρ_N, this choice of ω corresponds to the highest ethical improvement.

As regards the weighting function ω_a, the ethical component of the target increases as a grows. In particular, for low values of γ (i.e., $\gamma \leq 1.2$) $\rho^*_{\omega_a}$ tends to be only slightly dependent on the choice of a, whereas for $\gamma \geq 1.3$ the ethical improvement of the target strongly increases as a varies. The behaviour

© 0 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

of $\rho^*_{\omega_1}$ is quite similar to the one of ρ_{BM}, while for $a > 1$ the ethical gain is greater wrt the other targets. For instance, $\rho^*_{\omega_{1.5}}$ increases the number of patients assigned to the superior treatment up to 8%, 4% and 2% wrt ρ_N, ρ_{BM} and ρ_{ZR}, respectively. While the power is substantially equivalent to that of the other competitors, only a small loss in estimation precision is observed (lower than 4% wrt Neyman allocation, 3% wrt ρ_{BM} and 1% wrt ρ_{ZR}).

Taking now into account the censoring scheme in (9), we consider the same previous set up with the recruiting period $R = 48$ months and the total duration $S = 120$ months. Tables 4-6 show the proportion of patients assigned to treatment A, the power of Wald test and the behaviour of the estimation efficiency for the considered targets as $\tilde{\gamma}$ varies.

Table 4 Proportion of patients assigned to the best treatment as $\tilde{\gamma}$ varies.

			$\tilde{\rho}_{BM}$		$\tilde{\rho}^*_\omega$ as ω varies					$\tilde{\rho}^*_{\omega_a}$		
$\tilde{\gamma}$	$\tilde{\rho}_N$	$\tilde{\rho}_{ZR}$	$c=9$	$c=12$	0.3	0.4	0.5	0.6	0.69	$a=1$	$a=1.5$	$a=2$
1	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
1.1	0.52	0.54	0.53	0.53	0.58	0.61	0.65	0.70	0.78	0.53	0.54	0.54
1.2	0.55	0.57	0.56	0.56	0.60	0.63	0.66	0.72	0.79	0.56	0.57	0.58
1.3	0.57	0.60	0.59	0.59	0.62	0.65	0.68	0.73	0.79	0.59	0.60	0.61
1.4	0.59	0.63	0.61	0.61	0.64	0.66	0.70	0.74	0.80	0.61	0.63	0.65
1.5	0.61	0.65	0.64	0.63	0.65	0.68	0.71	0.75	0.80	0.64	0.66	0.68
1.6	0.62	0.68	0.66	0.65	0.67	0.69	0.72	0.76	0.81	0.66	0.68	0.71
1.8	0.64	0.70	0.68	0.67	0.68	0.71	0.73	0.77	0.82	0.68	0.71	0.73
1.9	0.65	0.72	0.70	0.69	0.70	0.72	0.74	0.78	0.82	0.70	0.73	0.75
2	0.66	0.73	0.71	0.71	0.71	0.73	0.75	0.79	0.83	0.71	0.74	0.77
2.1	0.68	0.75	0.73	0.72	0.72	0.74	0.76	0.80	0.83	0.73	0.76	0.79

Table 5 Approximated power of Wald test as $\tilde{\gamma}$ varies.

			$\tilde{\rho}_{BM}$		$\tilde{\rho}^*_\omega$ as ω varies					$\tilde{\rho}^*_{\omega_a}$		
$\tilde{\gamma}$	$\tilde{\rho}_N$	$\tilde{\rho}_{ZR}$	$c=9$	$c=12$	0.3	0.4	0.5	0.6	0.69	$a=1$	$a=1.5$	$a=2$
1	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
1.1	0.18	0.18	0.18	0.18	0.18	0.17	0.17	0.16	0.15	0.18	0.18	0.18
1.2	0.39	0.39	0.39	0.39	0.39	0.38	0.38	0.36	0.32	0.39	0.39	0.39
1.3	0.62	0.62	0.62	0.62	0.62	0.61	0.60	0.58	0.53	0.62	0.62	0.62
1.4	0.80	0.80	0.80	0.80	0.80	0.79	0.78	0.76	0.71	0.80	0.80	0.80
1.5	0.91	0.91	0.91	0.91	0.91	0.90	0.90	0.88	0.85	0.91	0.91	0.90
1.6	0.96	0.96	0.96	0.96	0.96	0.96	0.96	0.95	0.93	0.96	0.96	0.96
1.8	0.99	0.98	0.99	0.99	0.99	0.98	0.98	0.98	0.97	0.99	0.98	0.98
1.9	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00
2.1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

© 0 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 6 Estimation efficiency as $\tilde{\gamma}$ varies.

		$\tilde{\rho}_{BM}$		$\tilde{\rho}^*_\omega$ as ω varies					$\tilde{\rho}^*_{\omega_a}$		
$\tilde{\gamma}$	$\tilde{\rho}_{ZR}$	$c=9$	$c=12$	0.3	0.4	0.5	0.6	0.69	$a=1$	$a=1.5$	$a=2$
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.1	1.00	1.00	1.00	0.99	0.97	0.94	0.87	0.72	1.00	1.00	1.00
1.2	1.00	1.00	1.00	0.99	0.97	0.94	0.88	0.75	1.00	1.00	1.00
1.3	1.00	1.00	1.00	0.99	0.98	0.95	0.89	0.77	1.00	1.00	0.99
1.4	0.99	1.00	1.00	0.99	0.98	0.95	0.89	0.79	1.00	0.99	0.98
1.5	0.99	1.00	1.00	0.99	0.98	0.95	0.90	0.80	1.00	0.99	0.98
1.6	0.99	0.99	1.00	0.99	0.98	0.95	0.90	0.81	0.99	0.98	0.97
1.8	0.98	0.99	0.99	0.99	0.98	0.95	0.91	0.82	0.99	0.98	0.96
1.9	0.98	0.99	0.99	0.99	0.98	0.96	0.91	0.83	0.99	0.97	0.95
2	0.98	0.99	0.99	0.99	0.98	0.96	0.92	0.84	0.99	0.97	0.94
2.1	0.97	0.99	0.99	0.99	0.98	0.96	0.92	0.85	0.99	0.96	0.93

The results obtained considering the right censoring scheme substantially reflect those just described, confirming the ethical gain of the newly introduced target wrt the other competitors. In this more realistic scenario, adopting $0.3 \leq \omega \leq 0.4$, $\tilde{\rho}^*_\omega$ always exhibits the highest assignment proportion to the best treatment, with a maximum loss in power less than 1% wrt the other targets. The same considerations apply to the estimation precision in Table 6: $\tilde{\rho}^*_\omega$ shows good inferential performances with only a slight loss wrt ρ_N for high values of ω. Note that, from the previous tables, the weighting function ω_a with $a \in (1;2)$ guarantees a good compromise between ethical and inferential concerns, as the induced ethical gains are not obtained at the expense of poor inferential performance.

6 Sensitivity analysis and robustness of our methodology

In this section we investigate the operating characteristics of our proposal in terms of type-I error control/power of the test, convergence of the allocation proportion to the target, and robustness of the suggested methodology to model misspecification. In what follows, we take into account the censoring scheme previously described with $R = 48$ and $S = 120$ months; each trial was replicated 30000 times and the DBCD (with randomization parameter 2) has been adopted with a starting block randomization until at least one event is observed in each arm.

6.1 Type-I error control

Table 7 summarizes the results of the sensitivity analysis for type-I error control of both Wald and LR tests for several values of n as the common treatment effect $\theta_A = \theta_B = \theta$ varies.

In general, an inflation of the type-I error up to $1 - 1.5\%$ could be present, especially for lower sample sizes and for fixed weights $\omega = 0.4$ and 0.5, which provide a strong ethical skew even for small differences in the treatment effects. On the other hand, log-normal weights seems to preserve the type-I errors in particular for $n \geq 400$ and for $a \leq 1.5$. Moreover, as n increases, Wald and LR tests tend to preserve the type-I error with log-normal weights, whereas they do not share this property in the case of fixed weights with the exception of $\omega = 0.3$. For these reasons, in what follows we will focus on $\omega = 0.3$ and the adaptive weights ω_a.

Other simulation results not shown here for brevity highlight that for sample sizes $n < 150$ the convergence of the allocation proportion to the target consistently deteriorates and its standard deviation grows, which tend to induce inflated type-I errors. We wish to stress that, for survival trials, the evolution of the RAR procedure, its convergence as well as the inferential performances are strictly related to the number of events observed during the experiment (which clearly depends on a complex combination of the magnitude

© 0 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 7 Simulated type-I errors of Wald and Log-Rank (LR) tests with DBCD adopting ρ^*_ω and $\rho^*_{\omega_a}$ as $\theta_A = \theta_B = \theta$ and n vary.

		ρ^*_ω						$\rho^*_{\omega_a}$					
		$\omega = 0.3$		$\omega = 0.4$		$\omega = 0.5$		$a = 1$		$a = 1.5$		$a = 2$	
θ	n	Wald	LR	Wald	LR	Wald	LR	Wald	LR	Wald	LR	Wald	LR
1	200	.058	.060	.061	.061	.062	.062	.055	.060	.059	.062	.059	.064
	300	.058	.057	.060	.059	.060	.061	.055	.057	.055	.055	.060	.058
	400	.055	.058	.055	.058	.059	.060	.053	.054	.054	.056	.054	.057
	500	.055	.055	.056	.057	.059	.059	.054	.054	.055	.054	.051	.054
5	200	.058	.058	.060	.059	.062	.063	.056	.057	.057	.061	.061	.063
	300	.057	.058	.060	.060	.062	.061	.055	.055	.056	.058	.058	.059
	400	.053	.056	.056	.059	.058	.060	.053	.054	.055	.055	.053	.055
	500	.054	.057	.057	.058	.058	.059	.054	.055	.052	.052	.053	.055
10	200	.057	.060	.061	.059	.062	.061	.057	.059	.058	.061	.062	.065
	300	.055	.057	.059	.061	.061	.063	.054	.055	.055	.058	.058	.059
	400	.055	.058	.057	.059	.058	.059	.052	.053	.054	.055	.055	.056
	500	.055	.056	.056	.057	.060	.059	.053	.055	.055	.055	.054	.055

of the treatment effects, the total number of enrolled patients, the recruitment period and the duration of the trial).

6.2 Comparisons among competitors

Table 8 collects the simulated operating characteristics of targets $\tilde{\rho}^*_\omega$ ($\omega = 0.3$), $\tilde{\rho}^*_{\omega_a}$ ($a = 1$, 1.5 and 2), $\tilde{\rho}_{BM}$ (with $c = 11$ in the upper part of the Table and $c = 13$ in the lower one), $\tilde{\rho}_N$ and $\tilde{\rho}_{ZR}$, implemented via DBCD (with randomization parameter 2) under different alternative treatment effects, i.e., $(\theta_A, \theta_B) = (12, 10)$ and $(15, 10)$, in which the right censoring scheme of Section 4 has been considered. As a benchmark, we also take into account the Completely Randomized (CR) design aimed at targeting the balanced allocation $\rho^*_{CR} = 1/2$; clearly, CR is not affected by the censoring scheme. The considered operating characteristics are i) the observed allocation proportion ρ to A and its standard deviation (sd) within brackets, ii) the power of Wald and LR tests and iii) the total observed ST. The sample size n is set to be equal to 300, 400 and 500.

From the upper part of Table 8 - namely for $(\theta_A, \theta_B) = (12, 10)$ - it can be seen that all the considered targets show similar power performances with a slight gain wrt CR (i.e., 1%-2%): this is more apparent for lower sample sizes and when Wald test is adopted. Furthermore, the ethical gain induced by the adoption of RAR procedures is quite substantial. Indeed, between 6% to 10% more patients are assigned to the better treatment when the compound target is used; in general, $\tilde{\rho}_{BM}$ and $\tilde{\rho}_N$ skew slightly less the assignments wrt our proposal, while $\tilde{\rho}_{ZR}$ lies between the cases $\tilde{\rho}^*_{\omega_1}$ and $\tilde{\rho}^*_{\omega_{1.5}}$. Longer STs are especially appreciable for the compound targets and further highlight the increased ethics induced by our proposal; clearly, as the sample size increases, STs increase as well. Lastly, the decreasing standard deviation (sd) of ρ as n grows shows the improvement in the convergence of the RAR procedure. The same conclusions still hold also for $(\theta_A, \theta_B) = (15, 10)$, since the higher difference in the treatment effects makes the ethical skew induced by the adoption of the suggested methodology even more pronounced (STs highlight even more the increased ethics in the trial). For what concerns the estimation efficiency - not shown here for brevity - the compound optimal target exhibits good performance, with a maximum loss up to 2%, confirming the previous results of Section 5.

© 0 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 8 Simulated operating characteristics of the considered targets with $\theta_B = 10$ and $\theta_A = 12,\ 15$.

	$n = 300$				$n = 400$				$n = 500$			
$(\theta_A, \theta_B) = (12, 10)$	$\rho\ (sd)$	Wald	LR	ST	$\rho\ (sd)$	Wald	LR	ST	$\rho\ (sd)$	Wald	LR	ST
$\tilde{\rho}^*_{0.3} = 0.60$	0.59 (.06)	0.45	0.32	3035	0.59 (.06)	0.54	0.41	4048	0.59 (.05)	0.62	0.49	5062
$\tilde{\rho}^*_{\omega_1} = 0.56$	0.56 (.06)	0.45	0.33	3025	0.56 (.05)	0.54	0.41	4032	0.56 (.04)	0.62	0.49	5040
$\tilde{\rho}^*_{\omega_{1.5}} = 0.57$	0.57 (.06)	0.45	0.32	3029	0.57 (.06)	0.54	0.41	4038	0.57 (.05)	0.62	0.49	5047
$\tilde{\rho}^*_{\omega_2} = 0.58$	0.58 (.07)	0.45	0.32	3034	0.58 (.06)	0.54	0.41	4044	0.58 (.06)	0.62	0.49	5055
$\tilde{\rho}_{BM} = 0.56$	0.56 (.05)	0.45	0.32	3024	0.56 (.05)	0.54	0.41	4031	0.56 (.04)	0.62	0.49	5039
$\tilde{\rho}_N = 0.55$	0.55 (.04)	0.45	0.33	3017	0.55 (.04)	0.54	0.41	4022	0.55 (.03)	0.62	0.49	5028
$\tilde{\rho}_{ZR} = 0.57$	0.57 (.06)	0.45	0.32	3028	0.57 (.05)	0.55	0.41	4037	0.57 (.05)	0.62	0.49	5047
$\rho^*_{CR} = 0.50$	0.50 (.04)	0.43	0.32	2993	0.50 (.03)	0.53	0.40	3992	0.50 (.03)	0.61	0.48	4989
$(\theta_A, \theta_B) = (15, 10)$	$\rho\ (sd)$	Wald	LR	ST	$\rho\ (sd)$	Wald	LR	ST	$\rho\ (sd)$	Wald	LR	ST
$\tilde{\rho}^*_{0.3} = 0.65$	0.65 (.05)	0.95	0.90	3517	0.65 (.05)	0.98	0.96	4690	0.65 (.04)	1.00	0.99	5864
$\tilde{\rho}^*_{\omega_1} = 0.64$	0.64 (.06)	0.95	0.91	3502	0.64 (.06)	0.99	0.97	4669	0.64 (.05)	1.00	0.99	5836
$\tilde{\rho}^*_{\omega_{1.5}} = 0.66$	0.66 (.06)	0.95	0.90	3526	0.66 (.06)	0.99	0.96	4701	0.66 (.06)	1.00	0.99	5877
$\tilde{\rho}^*_{\omega_2} = 0.68$	0.68 (.07)	0.95	0.90	3550	0.68 (.07)	0.98	0.96	4733	0.68 (.06)	1.00	0.99	5917
$\tilde{\rho}_{BM} = 0.63$	0.63 (.05)	0.95	0.91	3494	0.63 (.05)	0.99	0.97	4659	0.63 (.05)	1.00	0.99	5824
$\tilde{\rho}_N = 0.61$	0.61 (.04)	0.96	0.91	3464	0.61 (.04)	0.99	0.97	4617	0.61 (.04)	1.00	0.99	5771
$\tilde{\rho}_{ZR} = 0.65$	0.65 (.06)	0.95	0.91	3520	0.65 (.06)	0.99	0.96	4692	0.65 (.05)	1.00	0.99	5865
$\rho^*_{CR} = 0.50$	0.50 (.03)	0.95	0.91	3338	0.50 (.03)	0.98	0.97	4452	0.50 (.03)	1.00	0.99	5564

6.3 Robustness to model misspecification

Albeit the proposed targets have been derived for exponential responses, in practical applications this assumption may not hold, so the purpose of this section is to investigate the robustness of our proposals to model misspecification. We considered three alternative distributions for the generation of the survival times: Weibull, log-logistic and log-normal. For each distribution, we consider two different values for the shape parameter h modelling the behavior of the hazard function. In particular, we set $h = 0.8$ and 1.5 for the Weibull model (corresponding to monotone decreasing/increasing hazard, respectively), $h = 0.4$ and 1 for the log-logistic (encompassing both non-monotone and decreasing hazard), while $h = 0.8$ and 1.2 for the log-normal distribution (having non-monotone hazards). Table 9 summarizes the performance - in terms of simulated type-I error/power and total observed survival time - of $\tilde{\rho}^*_{\omega}$ ($\omega = 0.3$), $\tilde{\rho}^*_{\omega_a}$ ($a = 1,\ 1.5$ and 2) implemented via DBCD (with randomization parameter 2) compared to CR design with $n = 400$. As is well-known (Sverdlov et al., 2011), since the model misspecification strongly affects both the power and type-I error of Wald test, compromising its reliability, we focus only on LR test.

For $\tilde{\rho}^*_{\omega_1}$ and $\tilde{\rho}^*_{\omega_{1.5}}$, type-I error is generally preserved, while a slight inflation may be observed for $\tilde{\rho}^*_{0.3}$ and $\tilde{\rho}^*_{\omega_2}$ (nevertheless it is always lower or at most equal to 6%). As is well-known, this behaviour is quite typical of RAR rules, especially if combined with delayed responses and targets with strong ethical skew (Sverdlov et al., 2011; Rosenberger and Lachin, 2015; Baldi Antognini et al., 2016, 2018).

In general, for a given statistical model, the power induced by CR design coincides substantially with that of DBCD, regardless of the chosen target, while the ethical skew of the optimal compound target (see Table 8) translates into an increase in the observed total ST wrt CR. As to be expected, the performances of the suggested methodology are strongly affected by model misspecifications. If the statistical model is characterized by a monotone decreasing (increasing, respectively) hazard, then this translates into an increased (decreased) total ST combined with a loss (gain) of power. For instance, taking into account the scenario $(\theta_A, \theta_B) = (12, 10)$, $\tilde{\rho}^*_{\omega_{1.5}}$ assigns 57% of patients to the best treatment and, under exponential outcomes, the power is 0.41 with a total observed ST of 4038 months. With Weibull responses instead, for $h = 1.5$ the consistent gain of power ($+0.33$ wrt the exponential model) is balanced with a loss of

© 0 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 9 Simulated power of LR test and total observed ST.

	Power					ST				
$(\theta_A, \theta_B) = (10, 10)$	$\tilde{\rho}^*_{0.3}$	$\tilde{\rho}^*_{\omega_1}$	$\tilde{\rho}^*_{\omega_{1.5}}$	$\tilde{\rho}^*_{\omega_2}$	ρ^*_{CR}	$\tilde{\rho}^*_{0.3}$	$\tilde{\rho}^*_{\omega_1}$	$\tilde{\rho}^*_{\omega_{1.5}}$	$\tilde{\rho}^*_{\omega_2}$	ρ^*_{CR}
Exponential	0.06	0.05	0.05	0.06	0.05	3666	3666	3666	3666	3666
Weibull, $h = 0.8$	0.06	0.05	0.06	0.06	0.05	3964	3964	3964	3964	3964
Weibull, $h = 1.5$	0.06	0.05	0.05	0.05	0.05	3412	3412	3412	3412	3412
Log-logistic, $h = 0.4$	0.06	0.05	0.05	0.05	0.05	4684	4684	4684	4684	4683
Log-logistic, $h = 1$	0.06	0.06	0.06	0.06	0.05	6717	6717	6717	6717	6718
Log-normal, $h = 0.8$	0.05	0.05	0.05	0.06	0.05	4900	4900	4900	4900	4899
Log-normal, $h = 1.2$	0.06	0.06	0.05	0.06	0.05	5858	5858	5858	5858	5861
$(\theta_A, \theta_B) = (12, 10)$	$\tilde{\rho}^*_{0.3}$	$\tilde{\rho}^*_{\omega_1}$	$\tilde{\rho}^*_{\omega_{1.5}}$	$\tilde{\rho}^*_{\omega_2}$	ρ^*_{CR}	$\tilde{\rho}^*_{0.3}$	$\tilde{\rho}^*_{\omega_1}$	$\tilde{\rho}^*_{\omega_{1.5}}$	$\tilde{\rho}^*_{\omega_2}$	ρ^*_{CR}
Exponential	0.41	0.41	0.41	0.41	0.40	4048	4032	4038	4044	3992
Weibull, $h = 0.8$	0.28	0.28	0.28	0.28	0.28	4338	4325	4330	4336	4289
Weibull, $h = 1.5$	0.73	0.74	0.74	0.74	0.74	3795	3775	3781	3788	3730
Log-logistic, $h = 0.4$	0.61	0.62	0.62	0.61	0.62	5170	5147	5155	5163	5087
Log-logistic, $h = 1$	0.17	0.16	0.17	0.17	0.16	7100	7090	7094	7099	7056
Log-normal, $h = 0.8$	0.51	0.52	0.52	0.51	0.52	5398	5375	5383	5392	5315
Log-normal, $h = 1.2$	0.27	0.28	0.27	0.27	0.27	6321	6305	6311	6319	6257
$(\theta_A, \theta_B) = (15, 10)$	$\tilde{\rho}^*_{0.3}$	$\tilde{\rho}^*_{\omega_1}$	$\tilde{\rho}^*_{\omega_{1.5}}$	$\tilde{\rho}^*_{\omega_2}$	ρ^*_{CR}	$\tilde{\rho}^*_{0.3}$	$\tilde{\rho}^*_{\omega_1}$	$\tilde{\rho}^*_{\omega_{1.5}}$	$\tilde{\rho}^*_{\omega_2}$	ρ^*_{CR}
Exponential	0.96	0.97	0.96	0.96	0.97	4690	4669	4701	4733	4452
Weibull, $h = 0.8$	0.85	0.85	0.85	0.85	0.85	4944	4919	4945	4973	4732
Weibull, $h = 1.5$	1.00	1.00	1.00	1.00	1.00	4456	4442	4478	4516	4191
Log-logistic, $h = 0.4$	1.00	1.00	1.00	1.00	1.00	6008	5992	6038	6084	5656
Log-logistic, $h = 1$	0.55	0.56	0.56	0.55	0.57	7690	7664	7687	7710	7491
Log-normal, $h = 0.8$	0.99	0.99	0.99	0.99	0.99	6241	6223	6268	6313	5893
Log-normal, $h = 1.2$	0.83	0.84	0.83	0.83	0.84	7054	7027	7061	7096	6780

257 months in ST, while for $h = 0.8$ a loss of power (-0.13) along with an ethical gain of $+292$ months of ST is observed. The log-logistic distribution with $h = 1$ exhibits the maximum power loss (-0.24) matched with the maximum ethical improvement ($+3056$ months) in ST. Interestingly, in the case of non-monotone hazard, an improvement in terms of both ethics and power can be observed. For instance, under the log-logistic distribution with $h = 0.4$ the power induced by $\tilde{\rho}^*_{\omega_{1.5}}$ grows ($+0.21$) and, at the same time, there is an improvement of 1117 months in the total observed ST (the same behaviour still holds under the log-normal model with $h = 0.8$ and for the case $(\theta_A, \theta_B) = (15, 10)$).

7 Case Studies

In this section, we apply our proposed methodology to redesign two phase III randomized clinical trials. The first one is the ITACa trial: this study was a first-line phase III randomized clinical multicenter trial on metastatic colorectal cancer, promoted and supervised by Istituto Scientifico Romagnolo per lo Studio e la Cura dei Tumori (IRST) - Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS). It aimed at evaluating the effectiveness of adding bevacizumab (arm A) to standard first-line chemotherapy (arm B). Between November 2007 and March 2012 ($R = 52$), 370 patients were randomized (176 to treatment A and 194 to B) and after a 24 month follow-up ($S = 76$), the estimated median progression free survival times (times from randomization to objective disease progression, or death from any cause, whichever occurs first) were 9.6 and 8.4 months for arm A and B, respectively (i.e. $\hat{\theta}_A = 13.8$ and $\hat{\theta}_B = 12.1$) (Passardi et al., 2015).

© 0 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

The adoption of the proposed compound optimal target may have still induced a consistent ethical gain, also improving the inferential precision. Indeed, by taking into account the right censoring scheme of Section 4, $\tilde{\rho}^*_\omega$ with fixed weight $\omega = 0.3$ would have allocated 219 patients to arm A ($\tilde{\rho}^*_{0.3} = 0.59$), namely 43 more patients to the treatment showing a higher benefit. Greater values of fixed weight would have accentuated the ethical skew even more, reaching the value of 289 patients assigned to A for $\tilde{\rho}^*_{0.69} = 0.78$. The adoption of log-normal weighting function (6) would have mitigated the ethical skew: for a between 1 to 4, from 55% to 59% patients would have been assigned to A, instead of 48% actually used, that also induces a gain in terms of inferential precision since the compound optimal target tends to be close to the Neyman's one.

The second case study is a trial on metastatic breast cancer reported by Jones et al. (2005) where 225 patients had been randomized to docetaxel (arm A) and 224 to paclitaxel (arm B). The recruitment period (R) lasted 84 months for a total study duration (S) of 102 months: the follow-up time was 18 months. In the intention-to-treat population, the estimated overall median survival times (times from randomization to death by any cause) were 16.1 months for arm A (i.e., $\hat{\theta}_A = 23.2$) and 12.7 months for B (i.e., $\hat{\theta}_B = 18.3$).

Assuming exponential distributions for the survival times, we redesigned the trial incorporating the censoring scheme. Along with the compound optimal target $\tilde{\rho}^*_\omega$ with fixed weight, the log-normal weighting function with $a = 1, 1.5$ and 2 are provided. We investigate the operating characteristics ensuing by the adoption of $\tilde{\rho}_N$, $\tilde{\rho}_{ZR}$ and $\tilde{\rho}_{BM}$ (with threshold $c = 20$). As a sequential analogue of the $1:1$ randomization considered in Jones et al. (2005), we also take into account the Completely Randomized (CR) design aimed at targeting the balanced allocation $\rho^*_{CR} = 1/2$.

For each choice of the target, 30000 trials were replicated by adopting the DBCD (with randomization parameter 2). As a new patient enters the trial, the target is estimated on the basis of the information accrued up to that time and the allocation probabilities are skewed toward the better performing treatment. Results by Hu et al. (2008) and Zhang and Rosenberger (2007) ensure that the DBCD is not affected by the intrinsic delay induced by exponentially distributed time-to-event responses subject to the independent right censoring. In the simulations, in order to derive non trivial estimates of the treatment effects, patients are assigned with block randomization up until at least one event is observed in both arms, then the DBCD procedure starts. In addition to Wald test, we also take into account the Log-Rank (LR) test, namely a common nonparametric procedure aimed to compare the survival functions among treatment groups.

Table 10 collects the values of the chosen targets and the simulated operating characteristics of the redesigned trial: observed allocation proportion to A (ρ) and its standard deviation ($sd(\rho)$), power of Wald and LR tests, total observed Survival Time (ST) and estimation efficiency C_I.

Table 10 Redesigning Jones et al. (2005) trial: targets ($\tilde{\rho}$), observed allocation proportions to arm A (ρ) and their standard deviation (sd) within brackets, power of Wald and LR tests, ST and C_I.

	ρ (sd)	Wald	LR	ST	C_I
$\tilde{\rho}_N = 0.57$	0.57 (.04)	0.68	0.45	5005	1.00
$\tilde{\rho}_{ZR} = 0.60$	0.60 (.06)	0.68	0.45	5018	1.00
$\tilde{\rho}_{BM} = 0.59$	0.59 (.06)	0.67	0.45	5015	1.00
$\tilde{\rho}^*_{0.3} = 0.62$	0.61 (.06)	0.67	0.44	5023	0.99
$\tilde{\rho}^*_{0.4} = 0.65$	0.63 (.07)	0.67	0.44	5032	0.98
$\tilde{\rho}^*_{0.5} = 0.68$	0.65 (.08)	0.67	0.43	5043	0.95
$\tilde{\rho}^*_{\omega_1} = 0.59$	0.59 (.06)	0.67	0.45	5015	1.00
$\tilde{\rho}^*_{\omega_{1.5}} = 0.60$	0.60 (.07)	0.67	0.45	5018	1.00
$\tilde{\rho}^*_{\omega_2} = 0.61$	0.61 (.07)	0.67	0.45	5024	0.99
$\rho^*_{CR} = 0.50$	0.50 (.02)	0.67	0.45	4976	0.98

Regardless of the choice of the weights, the compound optimal target tends to improve ethics wrt to the other competitors, clearly even more wrt the balanced allocation adopted by Jones et al. (2005). Due to the

© 0 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

moderate difference between the treatment effects, this aspect becomes more apparent for higher values of the fixed weight ω instead of the log-normal weights. In general, a more pronounced ethical component in the compound optimal target seems to slow the convergence of the simulated allocation proportion to the target, as also confirmed by a slight increase of the corresponding standard deviation. This is likely due to i) the sensitivity of the ethical skew of the targets to even slight differences in the survival estimates and ii) the effect of delayed responses (as already stressed by (Zhang and Rosenberger, 2007; Hu et al., 2008)). Indeed, simulation results not shown here for the sake of brevity highlight that for a higher sample size the convergence of the allocation proportion to the target consistently improves and its standard deviation decreases (see also Section 6).

Even if $\tilde{\rho}_N$ optimizes inference, all the considered targets lead to simulated powers and estimation efficiencies which are substantially the same as Neyman's. Moreover, when the weighting function ω_a is adopted, the convergence of the simulated allocation proportion to the target is guaranteed along with a moderate gain in terms of overall survival time wrt CR, regardless of the choice of a. As expected, under the assumption of exponential outcomes, the power of Wald test is higher wrt that of LR test, which instead is more appropriate in the presence of model misspecification (see Section 6.3).

8 Discussion

The present paper deals with the problem of finding optimal allocations for comparative clinical trials with survival endpoints. In order to obtain an appropriate trade-off among ethical demands and inferential precision, a compound optimization approach is proposed. The ensuing compound optimal target guarantees very good performances in terms of both ethical gain and statistical efficiency: our results show that in the case of a fixed weight, a value of $\omega < 0.4$ can be used in order to obtain a good trade-off between those objectives. A greater value of ω, although further emphasizing the ethical gain, may induce inflated type-I errors, especially for low sample sizes.

To further improve the flexibility of the proposed approach, a weighting function depending on the unknown treatment effects can be adopted: this will allow for the possibility to adaptively adjust the weights during the trial. This flexibility, along with the peculiarity of ω_a to improve ethics even for small treatment differences, makes the implementation of $\tilde{\rho}^*_{\omega_a}$ via RAR procedures preferable to the one based on possible best guesses carried out in the planning phase. Based on our extensive simulation study, a value of $a \in [1; 2)$ is suggested to obtain a good compromise between ethics and inference, which also preserves the type-I errors.

The ethical improvement induced by the proposed target makes it a valid candidate especially for oncological trials, where the ethical concern represents a crucial aspect. Indeed, the results here obtained underline that the adoption of an ethical allocation implemented with DBCD can provide a consistent ethical improvement wrt the balanced allocation, without compromising inference. Clearly, in trials where the ethical concern plays a less important role, CR design could still represent a good choice as it guarantees the control of type-I error with smaller sample sizes.

The analytical properties of the newly introduced targets are compared to those of the already existing allocations proposed in the literature, showing a strong ethical gain with at worst only a slight loss in inferential performances. The procedure is also illustrated by redesigning two real oncological trials: in both cases the adoption of our proposal has induced a considerable ethical benefit without adversely affecting neither power nor estimation precision.

Although the present methodology is derived for exponentially distributed outcomes, where Wald statistics is used to test the equality of treatment effects, our results highlight that the adoption of RAR procedures with the compound optimal target could be suitably matched with LR test, especially in the case of possible model misspecification. Indeed, an extensive simulation study has been carried out in order to investigate the operating characteristics of the new proposal, taking also into account its robustness and

© 0 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

type-I error control. In this regard, our proposal turned out to be quite similar to CR in terms of power, while the improvement in ST is remarkable.

To extend the applicability of our methodology, one of the fundamental directions is the inclusion of the covariates. For example, the approach proposed by Sverdlov et al. (2013) could be a possible starting point: by linking the expected survival times of the treatments with a linear function of important covariates, the optimal compound target would incorporate their effects as well. However, this natural approach has not yet a formal mathematical justification and therefore optimal properties of targets accounting for covariates/prognostic factors should be explored. Another interesting future research consists in combining RAR procedures with group sequential methods, in order to improve both individual and collective ethics by allowing early stopping of the trial for futility/efficacy and, at the same time, skewing more patients to the best performing treatment.

Acknowledgements Research supported by the Italian Ministry of Education, University and Research under PRIN 2015 "Environmental processes and human activities: capturing their interactions via statistical methods (EphaStat)". We are grateful to the referees and the AE for their helpful suggestions and comments, which led to a substantially improved version of the paper.

Conflict of Interest

The authors have declared no conflict of interest.

Appendix *(Proof of Theorem 3.2)*

Firstly note that the compound criterion $\mathcal{C}_\omega$ is strictly concave in ρ, since $\mathcal{C}_E$ is linear and $\mathcal{C}_I$ is concave; indeed, from (3),

$$\mathcal{C}_I''(\rho) = \frac{-2\gamma^2(\gamma+1)^2}{\{\rho(1-\gamma^2)+\gamma^2\}^3} < 0,$$

since $\gamma \in \mathbb{R}^+$ and $\rho(1-\gamma^2)+\gamma^2 = \rho + (1-\rho)\gamma^2 > 0$. Moreover, by differentiating (5), we obtain the following equation

$$\left(\frac{\omega}{1-\omega}\right)\operatorname{sgn}(\gamma-1) = \frac{(\gamma+1)^2(x^2-\gamma^2)}{[x+\gamma^2]^2}, \tag{10}$$

where $x = \rho/(1-\rho) \in \mathbb{R}^+$. Letting $f(x,\gamma) = (\gamma+1)^2(x^2-\gamma^2)[x+\gamma^2]^{-2}$, then f is monotonically increasing in x with $\lim_{x\to 0} f(x,\gamma) = -[(\gamma+1)/\gamma]^2$, $\lim_{x\to\infty} f(x,\gamma) = (\gamma+1)^2$ and $f(x,\gamma) = 0$ iff $x = \gamma$ (i.e., $\rho = \rho_N$). If $\gamma > 1$, when $\omega/(1-\omega) \geq (\gamma+1)^2 = (1-\rho_N)^{-2}$, i.e. $\omega \geq [1+(1-\rho_N)^2]^{-1}$, then $\rho^*_\omega = 1$. Analogously for $\gamma < 1$, when $\omega/(1-\omega) \geq (\gamma+1)^2/\gamma^2 = \rho_N^{-2}$, i.e., $\omega \geq (1+\rho_N^2)^{-1}$, then $\rho^*_\omega = 0$. Therefore, (7) follows after simple algebra by observing that, when $\gamma > 1$, $\max\{\rho_N; 1-\rho_N\} = \rho_N > 1/2$, while for $\gamma < 1$, $\max\{\rho_N; 1-\rho_N\} = 1-\rho_N > 1/2$, so that $\omega/(1-\omega) < 4$ implies that $\omega < [1+(1-\max\{\rho_N; 1-\rho_N\})^2]^{-1}$. Thus, $\rho^*_\omega \geq \rho_N$ for $\gamma \geq 1$ and $\rho^*_\omega < \rho_N$ for $\gamma < 1$, since f is monotonically increasing in x with $f(\gamma,\gamma) = f(\rho_N/(1-\rho_N),\gamma) = 0$.

From now on we take into account the case $\gamma \geq 1$ (the other scenario could be derived analogously). Property P2 follows immediately by the fact that $4/5 \leq [1+(1-\rho_N)^2]^{-1}$, since $\rho_N \geq 1/2$. As regards P3, it is sufficient to notice that the LHS of (10) is monotonically increasing in ω.

Finally, for the proof of P4, let

$$\widetilde{f}(x,\gamma) = \begin{cases} f(x,\gamma), & x \geq \gamma; \\ 0, & 1 \leq x < \gamma, \end{cases}$$

© 0 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

The derivative of $\widetilde{f}$ wrt to γ is, for $x \geq \gamma$

$$\frac{2(\gamma+1)}{\gamma^2+x}\left[\frac{(x-\gamma^2-2\gamma)(x^2-\gamma^2)}{\gamma^2+x} - \frac{\gamma(\gamma+1)}{\gamma^2+x}\right]$$

and 0 otherwise. Firstly, note that for every $x \in [1; 2+\sqrt{3}]$ the function $\widetilde{f}$ is decreasing in γ, since

$$(x-\gamma)^2(x+\gamma) \leq \gamma(\gamma+1)x(x+1) \qquad \forall \gamma \geq 1. \tag{11}$$

Moreover, since f is monotonically increasing in x, for every $\gamma \geq 1$

$$\widetilde{f}(x,\gamma) \leq \widetilde{f}(2+\sqrt{3},\gamma) \leq \widetilde{f}(2+\sqrt{3},1) = 4/\sqrt{3}, \qquad \forall x \in [1; 2+\sqrt{3}]. \tag{12}$$

Thus, from (10), by choosing the weight ω such that $\omega/(1-\omega) < 4/\sqrt{3}$, i.e. $\omega < 0.698$, guarantees that the optimal compound target is monotonically increasing in γ. Notice that this proof encompasses also the case $\omega = \omega(\gamma)$ since the LHS of (10) does not depend on x, namely it is still a constant wrt to x.

References

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

Baldi Antognini, A. and A. Giovagnoli (2015). *Adaptive Designs for Sequential Treatment Allocation*. Boca Raton, FL: Chapman & Hall/CRC Biostatistics.

Baldi Antognini, A., A. Vagheggini, and M. Zagoraiou (2016). Is the classical wald test always suitable under response-adaptive randomization? *Statistical Methods in Medical Research 27*, 2294–2311.

Baldi Antognini, A., A. Vagheggini, M. Zagoraiou, and M. Novelli (2018). A new design strategy for hypothesis testing under response adaptive randomization. *Electronic Journal of Statistics 12*, 2454–2481.

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

Biswas, A. and R. Bhattacharya (2009). Optimal response-adaptive designs for normal responses. *Biometrical Journal 51*, 193–202.

Biswas, A., R. Bhattacharya, and L. Zhang (2007). Optimal response-adaptive designs for continuous responses in phase III trials. *Biometrical Journal 49*, 928–940.

Biswas, A. and S. Mandal (2004). Optimal adaptive designs in phase iii clinical trials for continuous responses with covariates. In *moDa 7–Advances in model-oriented design and analysis*, pp. 51–59. Springer.

Hallstrom, A., M. Brooks, and M. Peckova (1996). Logrank, play the winner, power and ethics. *Statistics in Medicine 15*(20), 2135–2142.

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

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

Hu, F., L.-X. Zhang, S. H. Cheung, and W. S. Chan (2008). Doubly adaptive biased coin designs with delayed responses. *The Canadian Journal of Statistics 36*, 541–559.

Jones, M. (2008). On reciprocal symmetry. *Journal of Statistical Planning and Inference 138*(10), 3039–3043.

Jones, S., J. Erban, B. Overmoyer, G. Budd, L. Hutchins, E. Lower, L. Laufman, S. Sundaram, W. Urba, K. Pritchard, et al. (2005). Randomized phase iii study of docetaxel compared with paclitaxel in metastatic breast cancer. *Journal of Clinical Oncology 23*(24), 5542–5551.

© 0 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Lawless, J. F. (2011). *Statistical Models and Methods for Lifetime Data* (2nd ed.). New York: John Wiley & Sons.

Passardi, A., O. Nanni, D. Tassinari, D. Turci, L. Cavanna, A. Fontana, S. Ruscelli, C. Mucciarini, V. Lorusso, A. Ragazzini, et al. (2015). Effectiveness of bevacizumab added to standard chemotherapy in metastatic colorectal cancer: final results for first-line treatment from the itaca randomized clinical trial. *Annals of Oncology 26*(6), 1201–1207.

Rosenberger, W. and P. Seshaiyer (1997). Adaptive survival trials. *Journal of Biopharmaceutical Statistics 7*, 617–624.

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

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

Royall, R. M. (1991). Ethics and statistics in randomized clinical trials. *Statistical Science 6*, 52–62.

Seshadri, V. (1965). On random variables which have the same distribution as their reciprocals. *Canadian Mathematical Bulletin 8*(6), 819–824.

Sverdlov, O., W. F. Rosenberger, and Y. Ryeznik (2013). Utility of covariate-adjusted response-adaptive randomization in survival trials. *Statistics in Biopharmaceutical Research 5*(1), 38–53.

Sverdlov, O., Y. Tymofyeyev, and W. K. Wong (2011). Optimal response-adaptive designs for multi-armed trials. *Statistics in Medicine 30*, 2890–2910.

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

Yao, Q. and L. Wei (1996). Play the winner for phase II/III clinical trials. *Statistics in Medicine 15*, 2413–2423.

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

© 0 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim