Journal of Biopharmaceutical Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lbps20

Efficient and Ethical Response-Adaptive Randomization Designs for Multi-Arm Clinical Trials With Weibull Time-to-Event Outcomes

Oleksandr Sverdlov[a], Yevgen Ryeznik[b] & Weng-Kee Wong[c]

[a] Translational Sciences, Novartis Pharmaceuticals Corporation, East Hanover, New Jersey, USA

[b] CS Ltd., Kharkov, Ukraine

[c] Department of Biostatistics, Fielding School of Public Health, University of California Los Angeles, Los Angeles, California, USA

Accepted author version posted online: 03 Apr 2014.Published online: 14 May 2014.

To cite this article: Oleksandr Sverdlov, Yevgen Ryeznik & Weng-Kee Wong (2014) Efficient and Ethical Response-Adaptive Randomization Designs for Multi-Arm Clinical Trials With Weibull Time-to-Event Outcomes, Journal of Biopharmaceutical Statistics, 24:4, 732-754, DOI: 10.1080/10543406.2014.903261

To link to this article: http://dx.doi.org/10.1080/10543406.2014.903261

Journal of Biopharmaceutical Statistics, 24: 732–754, 2014

ISSN: 1054-3406 print/1520-5711 online
DOI: 10.1080/10543406.2014.903261

Taylor & Francis
Taylor & Francis Group

EFFICIENT AND ETHICAL RESPONSE-ADAPTIVE RANDOMIZATION DESIGNS FOR MULTI-ARM CLINICAL TRIALS WITH WEIBULL TIME-TO-EVENT OUTCOMES

Oleksandr Sverdlov[1], Yevgen Ryeznik[2], and Weng-Kee Wong[3]

[1]*Translational Sciences, Novartis Pharmaceuticals Corporation, East Hanover, New Jersey, USA*
[2]*CS Ltd., Kharkov, Ukraine*
[3]*Department of Biostatistics, Fielding School of Public Health, University of California Los Angeles, Los Angeles, California, USA*

We consider a design problem for a clinical trial with multiple treatment arms and time-to-event primary outcomes that are modeled using the Weibull family of distributions. The* D*-optimal design for the most precise estimation of model parameters is derived, along with compound optimal allocation designs that provide targeted efficiencies for various estimation problems and ethical considerations. The proposed optimal allocation designs are studied theoretically and are implemented using response-adaptive randomization for a clinical trial with censored Weibull outcomes. We compare the merits of our multiple-objective response-adaptive designs with traditional randomization designs and show that our designs are more flexible, realistic, generally more ethical, and frequently provide higher efficiencies for estimating different sets of parameters.

Key Words: Censoring; *D*-optimal design; Ethical concern; Response-adaptive randomization; Weibull distribution.

1. INTRODUCTION

Many randomized clinical trials have multiple treatment groups and pursue multiple experimental objectives. Response-adaptive randomization (RAR) is used in clinical trials to achieve selected experimental objectives while preserving the validity of the trial results (Hu and Rosenberger, 2006). Modern research on RAR has focused on finding optimal allocation strategies and implementing them using response-adaptive randomization procedures with minimal variability (Rosenberger et al., 2001; Zhang and Rosenberger, 2007; Tymofyeyev et al., 2007; Liang and Carriere Chough, 2009; Li and Wang, 2013). Most of these papers deal with binary or continuous outcomes. However, many clinical trials use censored time-to-event outcomes as primary measures of treatment safety and/or efficacy. Time-to-event outcomes arise in medical studies in oncology (progression-free survival; overall survival), virology (duration of viral shredding; time until resolution of all

Received August 3, 2012; Accepted May 4, 2013
Address correspondence to Dr. Oleksandr Sverdlov, Translational Sciences, Novartis Pharmaceuticals Corporation, One Health Plaza, East Hanover, NJ 07936, USA; E-mail: alex.sverdlov@novartis.com
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lbps.

signs and symptoms), dentistry (time to onset and duration of anesthesia), and some other therapeutic areas.

The research on optimal RAR designs in time-to-event trials has been limited. The first paper that provided optimal RAR procedures for survival trials is by Zhang and Rosenberger (2007), where they focused on two-armed survival trials with exponential and Weibull outcomes with a goal to minimize the total hazard in the study while maintaining power. Several other papers also deal with optimal design problems for time-to-event experiments with two treatment arms (Jóźwiak and Moerbeek, 2010; Konstantinou et al., 2014).

Zhang and Rosenberger (2007) acknowledged the challenges and affirmed the usefulness of optimal allocation designs for survival trials with more than two arms. Sverdlov et al. (2011) proposed optimal RAR designs for multi-arm survival trials with censored exponential outcomes. In the current article we develop optimal RAR designs for ($K \geq 2$)-treatment multiple-objective clinical trials with censored time-to-event outcomes that can be modeled using the Weibull family of distributions. We focus on the Weibull distributions because this family is widely used in survival analysis and its utility is well documented (Carroll, 2003). The significance of our optimal RAR designs is that they handle simultaneously the objective of efficient estimation of different sets of parameters of interest and the goal of assigning more patients to clinically best treatments, while maintaining randomization. In addition, in our development we directly address the issue of using inherently delayed responses in the design adaptations.

The next section contains statistical background material. In section 3, we obtain the D-optimal allocation for a multi-arm clinical trial with censored Weibull outcomes and propose two approaches for obtaining multi-objective optimal allocations. In section 4, we discuss implementation of the proposed optimal allocation designs in practice by means of RAR. In section 5, we present results of simulation studies to evaluate operating characteristics of the proposed optimal RAR designs and assess how delayed responses impact statistical properties of the designs. In section 6, we present an application of our methodology for a three-arm survival trial. In section 7, we give conclusions and outline future research in this direction. All technical considerations are relegated to the appendix.

2. STATISTICAL BACKGROUND

Throughout we assume we have a predetermined sample size n, either from budget considerations or based on prior experience of the number of patients that could be realistically recruited for the trial in the given time frame. Our main design question is then how to allocate these n patients into K groups in some optimal way. For our setup, we only determine the optimal proportion of patients ρ_k to assign to group k for $k = 1, \ldots, K$ and do so in practice by allocating roughly $n_k = n\rho_k$ patients to group k subject to $\rho_k \in [0, 1]$ with $\sum_{k=1}^{K} \rho_k = 1$. More specifically, given a statistical model, our design problem is to find an allocation vector $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_K)$ that minimizes some convex criterion of the inverse of the Fisher information matrix $\boldsymbol{M}(\boldsymbol{\rho}, \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a vector of unknown model parameters.

Suppose there are $K \geq 2$ treatment groups and the event time T_k from group k follows a Weibull distribution such that

$$\log T_k = \mu_k + bW, \quad k = 1, \ldots, K. \tag{1}$$

In equation (1), μ_k is the effect of treatment k, $b > 0$ is an unknown scale parameter assumed to be common for the K groups and W is the standard extreme value random

variable with probability density function $f(w) = e^w \exp(-e^w)$. In what is to follow, we let $\boldsymbol{\theta} = (\mu_1, \ldots, \mu_K, b)$. Note that when $b = 1$, we have the usual exponential model; otherwise, we have a general Weibull model that allows for different hazard patterns.

In trials with time-to-event outcomes, observations are likely to be censored. We assume that censoring time C is independent of T_k. For the ith patient in treatment group k, one observes a pair (t_{ik}, δ_{ik}), where $t_{ik} = \min(T_{ik}, C_{ik})$ is the observed time and $\delta_{ik} = 1_{\{t_{ik}=T_{ik}\}}$ is an indicator of the event of interest for $i = 1, \ldots, n_k$ and $k = 1, \ldots, K$.

Following convention, we focus on the D-optimal allocation problem where we want to minimize $|\boldsymbol{M}^{-1}(\boldsymbol{\rho}, \boldsymbol{\theta})|$ or, equivalently, $\log |\boldsymbol{M}^{-1}(\boldsymbol{\rho}, \boldsymbol{\theta})|$, by choosing $\boldsymbol{\rho}$ in order to obtain the most accurate inference for $\boldsymbol{\theta}$.

Let $z_{ik} = (\log t_{ik} - \mu_k)/b$ and assume that event times from all patients are independent. The log-likelihood function is

$$\log L(\boldsymbol{\theta}) = \log \prod_{k=1}^{K} \prod_{i=1}^{n_k} \left\{ b^{-1} e^{z_{ik}} \exp\left(-e^{z_{ik}}\right) \right\}^{\delta_{ik}} \left\{ \exp\left(-e^{z_{ik}}\right) \right\}^{1-\delta_{ik}}.$$

The value of the maximum likelihood estimator of $\boldsymbol{\theta}$ is found by solving the system of score equations $\frac{\partial \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}$, and its asymptotic variance–covariance matrix is inversely proportional to the Fisher information matrix. For an allocation vector $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_K)$, the $(K + 1) \times (K + 1)$ Fisher information matrix for $\boldsymbol{\theta}$ is found as

$$\boldsymbol{M}(\boldsymbol{\rho}, \boldsymbol{\theta}) = \frac{n}{b^2} \begin{pmatrix} \operatorname{diag}\{\rho_1 \epsilon_1, \ldots, \rho_K \epsilon_K\} & \boldsymbol{x} \\ \boldsymbol{x}^{\mathrm{T}} & \sum_{k=1}^{K} \rho_k (\epsilon_k + c_k) \end{pmatrix}, \tag{2}$$

where $\boldsymbol{x} = (\rho_1 a_1, \ldots, \rho_K a_K)^{\mathrm{T}}$, $\epsilon_k = \Pr(\delta_{ik} = 1)$, $a_k = E\left(z_{ik} e^{z_{ik}}\right)$, $c_k = E\left(z_{ik}^2 e^{z_{ik}}\right)$, and each of the ϵ_k, a_k, and c_k is a function of $\boldsymbol{\theta}$ and the censoring mechanism in the trial. Accordingly, we shall first consider locally optimal designs that require nominal values of the parameter $\boldsymbol{\theta}$ be available. Such optimal allocation designs serve as "calibrating" designs for comparative purposes in the ideal case, and they represent a useful starting point for finding optimal RAR designs at a later stage.

3. OPTIMAL ALLOCATION DESIGNS

3.1. The D-Optimal Allocation

From equation (2) we obtain

$$|\boldsymbol{M}(\boldsymbol{\rho}, \boldsymbol{\theta})| = (n/b^2)^{K+1} \prod_{k=1}^{K} \rho_k \epsilon_k \sum_{k=1}^{K} \rho_k d_k,$$

where $d_k = \epsilon_k + c_k - a_k^2/\epsilon_k > 0$ for $k = 1, \ldots, K$. The quantity d_k is inversely proportional to the asymptotic variance of the maximum likelihood estimator $\hat{b}$ if treatment k is considered alone. A larger value of d_k means more information for b in treatment group k. The D-optimal allocation is found by solving the following minimization problem:

$$\begin{cases} \min\limits_{\rho_1,\ldots,\rho_K} & (\prod_{k=1}^K \rho_k\epsilon_k)^{-1}(\sum_{k=1}^K \rho_k d_k)^{-1} \\ \text{subject to} & \sum_{k=1}^K \rho_k = 1, \end{cases} \tag{3}$$

where $0 \le \rho_k \le 1$, $k = 1, \ldots, K$.

Using Lagrange multipliers, one determines that the D-optimal allocation vector $\boldsymbol{\rho}^* = (\rho_1^*, \ldots, \rho_K^*)$ satisfies the following system of K equations:

$$\frac{1}{\rho_k} + \frac{d_k}{\sum_{k=1}^K \rho_k d_k} = K + 1, \quad k = 1, \ldots, K. \tag{4}$$

From equation (4), it is clear that the D-optimal allocation vector $\boldsymbol{\rho}^*$ depends on the model parameters $\boldsymbol{\theta}$ via d_k, $k = 1, \ldots, K$, which are unknown at the trial onset. In section 4 we discuss sequential estimation of $\boldsymbol{\theta}$ and construct RAR procedures to target the D-optimal allocation.

The D-optimal allocation has the following interesting properties:

(i) If there is no censoring in any treatment group, that is, $\epsilon_1 = \ldots = \epsilon_K = 1$, the D-optimal allocation vector is $\boldsymbol{\rho}^* = (1/K, \ldots, 1/K)$, which is the balanced allocation.
(ii) For $K = 2$, the D-optimal allocation is as follows: If $d_1 = d_2$, $\rho_1^* = \rho_2^* = 1/2$. If $d_1 \ne d_2$,

$$\rho_1^* = \frac{d_1 - 2d_2 + (d_1^2 - d_1 d_2 + d_2^2)^{1/2}}{3(d_1 - d_2)}, \quad \rho_2^* = 1 - \rho_1^*.$$

(iii) The D-optimal proportions satisfy $1/(K+1) \le \rho_k^* \le 2/(K+1)$, $k = 1, \ldots, K$.
(iv) Assume that treatments $1, \ldots, K$ are such that $d_1 \ge d_2 \ge \ldots \ge d_K$. Then the D-optimal allocation proportions satisfy $\rho_1^* \ge \rho_2^* \ge \ldots \ge \rho_K^*$.

This last property has a clear interpretation. Since we assume that parameter b is common to the K treatment groups, one has greater allocation proportions for the treatment groups that contribute a larger amount of information for b.

To fix ideas, suppose that shorter responses are clinically favorable (e.g., healing times). For a pair of treatments (i, j), where $i \ne j$, the condition $\mu_i < \mu_j$ implies that treatment i is more efficacious than treatment j. For many censoring schemes, including the censoring scheme with constant follow-up time for each patient (subsection 5.1), d_k is monotonically decreasing in μ_k. In this case, the condition $\mu_1 \le \mu_2 \le \ldots \le \mu_K$ implies $d_1 \ge d_2 \ge \ldots \ge d_K$, and from property (iv) it follows that $\rho_1^* \ge \rho_2^* \ge \ldots \ge \rho_K^*$. Hence, the D-optimal allocation proportions are ordered consistently with the magnitude of treatment effects such that larger proportions of patients are assigned to the treatments with shorter healing times. This implies that the balanced allocation induces losses on both ethical and inferential grounds, and hence, in this case, it is *not D-admissible* (Baldi Antognini and Giovagnoli, 2010).

In trials where the aim is to prolong life (e.g., survival trials), the objectives of D-efficiency and ethics are in conflict for our model. In this case a balanced allocation $\boldsymbol{\rho}_B = (1/K, \ldots, 1/K)$ could be one compromise strategy. In subsection 3.2 we present a

more rigorous approach to determine allocation rules to achieve trade-off between inferential efficiency and ethical considerations, and in section 6 we explore an application of this approach for survival trials where longer event times imply higher treatment efficacy.

3.2. Weighted Optimal Allocation

We have shown in subsection 3.1 that the D-optimal allocation is an attractive strategy in trials where shorter event times are clinically favorable (e.g., pain relief in migraine studies; viral shredding in virology studies; time to onset of anesthesia in dentistry). However, the D-optimal allocation is skewed not far away from the balanced allocation; see property (iii). An investigator may want to apply a greater degree of skewing toward better treatments, while maintaining statistical efficiency. Better treatments here mean the time to recovery is shorter, and an allocation strategy that places more patients to better treatments is desirable from an ethical standpoint (a similar idea can be applied in survival trials where longer event times are clinically favorable; see section 6).

Suppose one of the trial objectives is to identify a "clinically best" treatment group with shortest recovery times. Let $\boldsymbol{\rho}_{\mathrm{E}} = (\rho_{\mathrm{E}1}, \ldots, \rho_{\mathrm{E}K})$ denote an allocation vector for which condition $\mu_i \leq \mu_j$ must imply $\rho_{\mathrm{E}i} \geq \rho_{\mathrm{E}j}$, with equality if an only if $\mu_i = \mu_j$. We define the components of $\boldsymbol{\rho}_{\mathrm{E}}$ as follows:

$$\rho_{\mathrm{E}k} = \frac{\{\exp(-\mu_k/b)\}^{\nu}}{\sum_{j=1}^{K}\{\exp(-\mu_j/b)\}^{\nu}}, \quad k = 1, \ldots, K, \tag{5}$$

where $\nu \geq 0$ is a user-specified parameter controlling the degree of skewness to the better treatment. When $\nu \to \infty$, $\rho_{\mathrm{E}k} \to 1$ for $k = \arg\min_{j=1,\ldots,K} \mu_j$, and one achieves the most ethical allocation for which all patients are assigned to the treatment arm having the smallest value of μ_k (or, if there are several such treatments, these treatments receive equal proportions). A large but finite value of ν ensures that $\boldsymbol{\rho}_{\mathrm{E}}$ is a smooth function. Other treatment effect mappings can be considered for $\boldsymbol{\rho}_{\mathrm{E}}$ (Rosenberger, 1993). In practice, it is helpful to have a graphical review of the allocation proportions of the selected $\boldsymbol{\rho}_{\mathrm{E}}$ across different scenarios. This should be done collaboratively by clinician and statistician.

We now present a simple yet general approach for constructing allocation designs that balance inferential and ethical considerations. We assume that the inferential objective is implemented by the allocation $\boldsymbol{\rho}_{\mathrm{I}} = (\rho_{\mathrm{I}1}, \ldots, \rho_{\mathrm{I}K})$, and the ethical objective is implemented by the allocation $\boldsymbol{\rho}_{\mathrm{E}} = (\rho_{\mathrm{E}1}, \ldots, \rho_{\mathrm{E}K})$. Note that both $\boldsymbol{\rho}_{\mathrm{I}}$ and $\boldsymbol{\rho}_{\mathrm{E}}$ are functions of $\boldsymbol{\theta}$.

Let $\lambda(\boldsymbol{\rho}, \tilde{\boldsymbol{\rho}})$ be some measure of the distance between allocation vectors $\boldsymbol{\rho}$ and $\tilde{\boldsymbol{\rho}}$, and let $0 \leq \alpha \leq 1$ be a user-selected constant. We define a weighted optimality allocation vector $\boldsymbol{\rho}_{\alpha}$ as the one that minimizes the weighted distance between $\boldsymbol{\rho}_{\mathrm{I}}$ and $\boldsymbol{\rho}_{\mathrm{E}}$, that is,

$$\boldsymbol{\rho}_{\alpha} = \arg\min_{\boldsymbol{\rho}} \{\alpha\lambda(\boldsymbol{\rho}, \boldsymbol{\rho}_{\mathrm{I}}) + (1-\alpha)\lambda(\boldsymbol{\rho}, \boldsymbol{\rho}_{\mathrm{E}})\}.$$

Different choices of the distance function will yield different solutions. For example, if we use the Kullback–Leibler directed divergence measure $\lambda_{KL}(\boldsymbol{\rho}, \tilde{\boldsymbol{\rho}}) = \sum_{k=1}^{K} \rho_k \log(\rho_k/\tilde{\rho}_k)$, it is easily shown that $\boldsymbol{\rho}_{\alpha}$ has components

$$\rho_{\alpha k} = \frac{(\rho_{\mathrm{I}k})^{\alpha}(\rho_{\mathrm{E}k})^{1-\alpha}}{\sum_{j=1}^{K}(\rho_{\mathrm{I}j})^{\alpha}(\rho_{\mathrm{E}j})^{1-\alpha}}, \quad k = 1, \ldots, K. \tag{6}$$

If we select a quadratic distance metric $\lambda_2(\boldsymbol{\rho}, \tilde{\boldsymbol{\rho}}) = \sum_{k=1}^{K}(\rho_k - \tilde{\rho}_k)^2$, then $\boldsymbol{\rho}_{\alpha}$ has components

$$\rho_{\alpha k} = \alpha\rho_{\mathrm{I}k} + (1-\alpha)\rho_{\mathrm{E}k}, \quad k = 1, \ldots, K. \tag{7}$$

If $\alpha = 1$, the optimal solution is $\boldsymbol{\rho}_{\mathrm{I}}$. If $\alpha = 0$, the optimal solution is $\boldsymbol{\rho}_{\mathrm{E}}$. For $0 < \alpha < 1$ we have an allocation that provides a compromise between efficiency and ethics. The choice of the value of α will have to be determined by the experimenter and scientists in the field.

3.3. Compound Optimal Allocation

Here we present another approach for achieving trade-off between inferential and ethical objectives. This approach borrows ideas from a general methodology of compound optimal designs (Cook and Wong, 1994). As in subsection 3.2, we assume that shorter event times are clinically favorable. Let us show that for model (1) with censoring, an allocation that leads to the most accurate inference for the parameter b is also attractive from the ethical perspective. By property (iv), a larger value of d_k leads to the greater allocation proportion for treatment k. The asymptotic variance of $\hat{b}$ is $\mathrm{avar}(\hat{b}) = b^2(n\Delta)^{-1}$, where $\Delta = \sum_{k=1}^{K}\rho_k d_k$. The allocation minimizing avar $(\hat{b})$, or, equivalently, minimizing $\Phi_1(\boldsymbol{\rho}) = -\log\Delta$, places all subjects on the treatment arm with the maximum value of d_k (or assigns equal proportions if there are several such treatments). The merits of such an allocation are twofold: It is optimal for estimating b, and it assigns all subjects to the treatment(s) with shortest recovery times. A disadvantage of this allocation is that it provides no information about other treatment arms.

To achieve trade-off between the criteria $\Phi_1(\boldsymbol{\rho}) = -\log\Delta$ and $\Phi_2(\boldsymbol{\rho}) = -\log|\boldsymbol{M}(\boldsymbol{\rho}, \boldsymbol{\theta})|$, we consider the following convex minimization problem:

$$\begin{cases} \min\limits_{\rho_1, \ldots, \rho_K} & \alpha\Phi_1(\boldsymbol{\rho}) + (1-\alpha)\Phi_2(\boldsymbol{\rho}) \\ \text{subject to} & \sum_{k=1}^{K}\rho_k = 1, \end{cases} \tag{8}$$

and

$$\rho_k \in [0, 1], \quad k = 1, \ldots, K$$

where $0 \le \alpha \le 1$ is a user-selected constant that determines the importance of the two objectives. The problem (8) is a well-defined convex optimization problem with a unique solution that is found numerically by solving the following nonlinear system of K equations:

$$\frac{1-\alpha}{\rho_k} + \frac{d_k}{\sum_{k=1}^{K}\rho_k d_k} = (1-\alpha)K + 1, \quad k = 1, \ldots, K.$$

If $\alpha = 0$, we have the D-optimal allocation. If $\alpha = 1$, the problem is to minimize $\Phi_1(\boldsymbol{\rho}) = -\log\Delta$, and we obtain the allocation that places all patients to the treatment(s) with

shortest recovery times. For $0 < \alpha < 1$ we have an allocation that provides a compromise between the two criteria.

4. IMPLEMENTING OPTIMAL ALLOCATION USING RESPONSE-ADAPTIVE RANDOMIZATION

The optimal allocation designs we have discussed thus far cannot be directly implemented because they depend on model parameters $\boldsymbol{\theta}$ that are unknown at the trial onset. For implementing optimal allocation in practice we use a doubly adaptive biased coin design (DBCD) (Eisele, 1994; Hu and Zhang, 2004), a RAR procedure with established statistical properties. The key assumption for applying the DBCD procedure to target $\boldsymbol{\rho} = (\rho_1(\boldsymbol{\theta}), \ldots, \rho_K(\boldsymbol{\theta}))$ is that $\boldsymbol{\rho}$ is a continuously differentiable vector function of $\boldsymbol{\theta}$ that is twice continuously differentiable in a small neighborhood of the true value of the parameter. The *D*-optimal allocation function meets this regularity assumption, and trade-off allocation schemes in subsections 3.2 and 3.3 also have smooth allocation functions. Therefore, the DBCD procedure is applicable for our proposed allocation schemes.

With the DBCD procedure, first Km_0 patients (where m_0 is some small positive integer) are randomized among the K groups with equal probability. Consider a point in the trial when j ($j \geq Km_0$) patients have been randomized and their outcome data have been obtained. Based on the data, one computes $\hat{\boldsymbol{\theta}}_j$ and estimates the target allocation proportions as $\hat{\rho}_{kj} = \rho_k(\hat{\boldsymbol{\theta}}_j)$ for $k = 1, \ldots, K$. Let $\boldsymbol{N}_j/j = (N_{1j}/j, \ldots, N_{Kj}/j)$ denote the vector of treatment proportions after j assignments. Then the probability of assigning treatment k for the $(j+1)$th patient is

$$\psi_{j+1,k} = \frac{\hat{\rho}_{kj}\left(\frac{\hat{\rho}_{kj}}{N_{kj}/j}\right)^2}{\sum_{i=1}^{K}\hat{\rho}_{ij}\left(\frac{\hat{\rho}_{ij}}{N_{ij}/j}\right)^2}, \quad k = 1, \ldots, K. \tag{9}$$

For randomization procedure (9), as $j \to \infty$, $\hat{\boldsymbol{\theta}}_j$ is strongly consistent and asymptotically normal, and the vector of treatment allocation proportions $\boldsymbol{N}_j/j$ is strongly consistent for $\boldsymbol{\rho}$ and follows an asymptotically normal distribution. Note that in time-to-event trials, responses are inherently delayed, and at the point of entry of the $(j+1)$th patient, outcome data from some of the patients $1, \ldots, j$ may not be available yet, and therefore, $\hat{\boldsymbol{\theta}}_j$ will be potentially computed based on data from fewer than j patients. Hu et al. (2008) showed that large-sample properties of DBCD are unaffected by delayed response under the condition that the outcomes occur "not too far out" in the accrual pattern. A "rule of thumb" is that 60% or more of the patients' outcomes should be observed during the accrual period. The next two subsections describe practical implementation of equation (9) with delayed responses.

4.1. "Cohort" Response-Adaptive Randomization

One approach to facilitate procedure (9) with delayed responses is a "cohort" RAR (Chappell and Karrison, 2006). For this group sequential procedure, every new cohort of patients is enrolled only after all patients in the previous cohort have responded (Fig. 1).

For a trial with K treatment groups we establish cohorts of size Km, where m is some positive integer. The Km subjects in the first cohort are randomized to the treatment groups

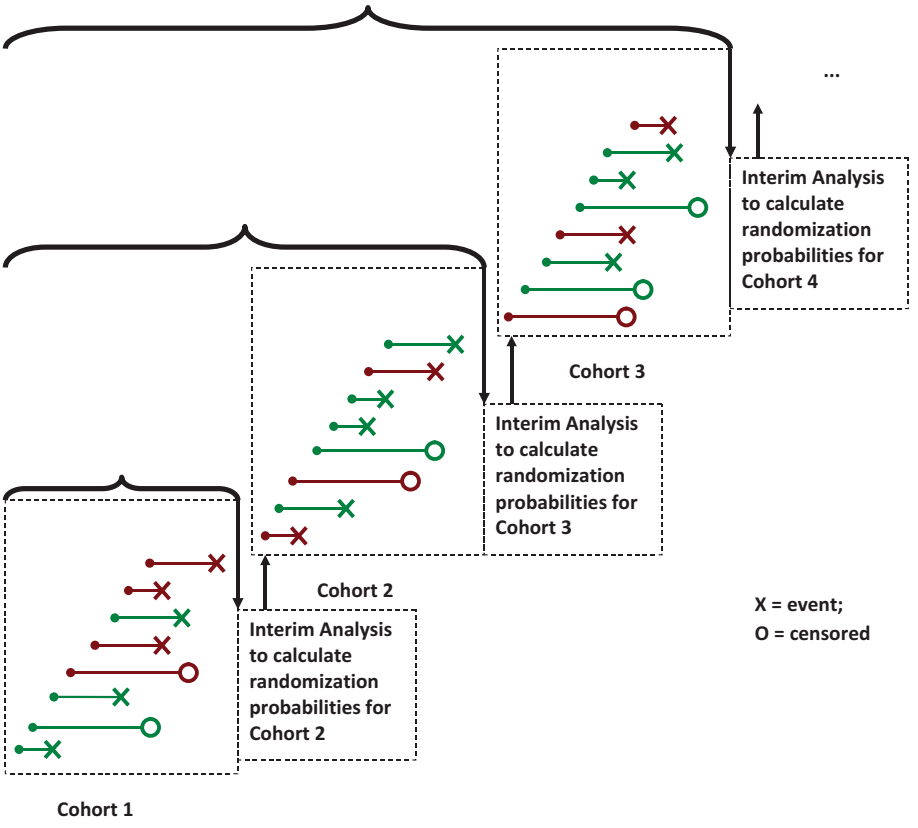

Figure 1 "Cohort" response-adaptive randomization design.

with equal probability to ascertain initial data, estimate model parameters and target allocation proportions, and use equation (9) to determine randomization probabilities for the second cohort. These randomization probabilities are applied individually to every subject in the second cohort. After all subjects in the second cohort have been randomized and their outcome data have been recorded, the full history of treatment assignments and responses from the $2Km$ subjects in the study is retrieved to determine randomization probabilities for the third cohort. The procedure is then repeated for the remaining cohorts in the study.

4.2. Response-Adaptive Randomization With Staggered Entry and Delayed Responses

A disadvantage of the "cohort" RAR scheme is that one has to wait until all subjects in a given cohort respond to make an adaptation for the next cohort. This may delay the enrollment. An alternative approach is to use an adaptation scheme that accounts for patient staggered entry and delayed responses (Latta, 1981; Zhang and Rosenberger, 2007). Such a scheme can be described using a priority queue model as follows. The trial has a fixed recruitment period of length $R > 0$. Patients enter the trial sequentially and their enrollment times are random. The observed time of a patient (minimum between the event time and the censoring time) is added to the patient's enrollment time. The initial Km_0 patients are

equally randomized among the K treatments. At prespecified points in the trial, interim data analyses are conducted to reestimate treatment randomization probabilities using data from those trial participants whose outcomes were observed prior to the time of interim data analysis.

Figure 2 shows a schematic of a RAR design with two interim analyses. One can also consider a fully sequential RAR procedure: When patient j ($j \geq Km_0$) enters the trial, data from $(j-1)$ (or fewer) patients whose outcomes were observed before the jth patient arrival are used to determine treatment randomization probabilities for the jth patient. Note that a substantial amount of patient outcome data must be observed during the recruitment phase to enable meaningful adaptation in the design; mathematically, the probability that a patient's outcome (event time or censoring time) will be observed before additional n patients arrive should be of order n^{-c} for some constant $c > 0$ (Hu et al., 2008). This condition is satisfied for clinical trials with uniform or exponential patient entry and Weibull event times. Simulations show that when 60% or more of the study patients contribute outcomes throughout the recruitment period, large-sample properties of the DBCD procedure still hold (Zhang and Rosenberger, 2007).

5. NUMERICAL RESULTS

5.1. Theoretical Comparison of Optimal Allocation Rules

It is instructive to compare efficiencies of different allocation strategies under an "idealized" situation when the model parameters $\boldsymbol{\theta} = (\mu_1, \ldots, \mu_K, b)$ are known and need not be estimated adaptively. We consider a clinical trial with $K = 3$ treatment arms, and consider four choices for the vector $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)$ and three choices for b: $b = \frac{1}{2}$ (increasing hazard), $b = 1$ (constant hazard; exponential distribution), and $b = 1.5$ (decreasing hazard). For the censoring scheme, we assume that each subject has a constant predetermined follow-up time $\tau > 0$ such that an observation occurs if $T_k \leq \tau$ and the observation is censored otherwise. Applications of such a scheme with constant censoring time can be found in Cheung et al. (2006) and Zhang and Rosenberger (2007). In our example, we set $\tau = 1/(-\log(0.1))$, the 10th percentile of the standard extreme value distribution. We compare five allocation rules for our three-arm trial:

I. Balanced allocation, $\boldsymbol{\rho} = (1/3, 1/3, 1/3)$.
II. D-optimal allocation, $\boldsymbol{\rho}^*$ (subsection 3.1).
III. Compound optimal allocation with $\alpha = 1/2$ (subsection 3.3).
IV. Weighted optimal allocation with $\alpha = 1/2$ using Euclidean metric (7).
V. Weighted optimal allocation with $\alpha = 1/2$ using Kullback–Leibler metric (6).

For allocation rules IV and V, we set $\nu = 2$ in equation (5). For each allocation we compute the D-efficiency and the efficiency for estimating the parameter b. The D-efficiency of an allocation $\boldsymbol{\rho}$ relative to $\boldsymbol{\rho}^*$ is defined as follows:

$$E_1(\boldsymbol{\rho}) = \left\{ \frac{|\boldsymbol{M}^{-1}(\boldsymbol{\rho}^*, \boldsymbol{\theta})|}{|\boldsymbol{M}^{-1}(\boldsymbol{\rho}, \boldsymbol{\theta})|} \right\}^{1/4}.$$

Let $\boldsymbol{\rho}^\dagger = (\rho_1^\dagger, \rho_2^\dagger, \rho_3^\dagger)$ be the compound optimal allocation with $\alpha = 1$ (from subsection 3.3) for the most efficient estimation of b. The efficiency of $\boldsymbol{\rho}$ relative to $\boldsymbol{\rho}^\dagger$ is defined as

$$E_2(\boldsymbol{\rho}) = \frac{\sum_{k=1}^{3} \rho_k d_k}{\sum_{k=1}^{3} \rho_k^{\dagger} d_k}.$$

Clearly, $E_1(\boldsymbol{\rho})$ and $E_2(\boldsymbol{\rho})$ are between 0 and 1 and high values of each efficiency criterion are desirable.

Table 1 shows that under all scenarios, allocation rules II, III, IV, and V have higher values of $E_2(\boldsymbol{\rho})$ than allocation rule I. In terms of $E_1(\boldsymbol{\rho})$ (D-efficiency), the ranking of the allocation rules in the order of their performance is: 1, D-optimal (allocation II); 2, Compound optimal (allocation III); 3, Balanced (allocation I); 4, Weighted optimality (WO) with Euclidean metric (allocation IV); and 5, WO with Kullback–Leibler metric (allocation V).

5.2. Simulation of Optimal Response-Adaptive Designs With "Cohort" RAR Scheme

Here, we present results of simulation studies to evaluate utility of RAR procedures targeting optimal allocation with a "cohort" RAR scheme described in subsection 4.1. We consider 12 experimental scenarios for $\boldsymbol{\theta} = (\mu_1, \mu_2, \mu_3, b)$ as in subsection 5.1. For each scenario, a trial with $n = 150$ patients was simulated 1000 times. We implemented five randomization designs: complete randomization for which every subject is randomized to treatment groups with equal probability (design I), and the DBCD procedure targeting four different optimal allocation schemes described in subsection 5.1 (designs II, III, IV, and V respectively). For implementing designs II–V, we used 10 cohorts of size 15.

Table 1 Efficiencies of the five allocation rules when $\boldsymbol{\theta} = (\mu_1, \mu_2, \mu_3, b)$ is known

		$b = 0.5$		$b = 1$		$b = 1.5$	
(μ_1, μ_2, μ_3)	Allocation	$E_1(\boldsymbol{\rho})$	$E_2(\boldsymbol{\rho})$	$E_1(\boldsymbol{\rho})$	$E_2(\boldsymbol{\rho})$	$E_1(\boldsymbol{\rho})$	$E_2(\boldsymbol{\rho})$
$(0, -1, -1)$	I	0.990	0.728	0.996	0.814	0.998	0.863
	II	1.000	0.783	1.000	0.840	1.000	0.876
	III	0.995	0.821	0.997	0.860	0.998	0.888
	IV	0.923	0.907	0.955	0.912	0.974	0.920
	V	0.782	0.961	0.930	0.927	0.967	0.925
$(0, -1, 0)$	I	0.974	0.455	0.993	0.629	0.997	0.725
	II	1.000	0.555	1.000	0.666	1.000	0.743
	III	0.985	0.630	0.994	0.702	0.997	0.762
	IV	0.833	0.801	0.922	0.792	0.962	0.810
	V	0.684	0.879	0.911	0.801	0.960	0.811
$(0, -0.5, -1)$	I	0.985	0.547	0.996	0.708	0.998	0.788
	II	1.000	0.613	1.000	0.732	1.000	0.800
	III	0.990	0.667	0.996	0.755	0.998	0.811
	IV	0.879	0.795	0.950	0.813	0.975	0.843
	V	0.785	0.837	0.939	0.820	0.973	0.844
$(0, -0.5, 1)$	I	0.977	0.485	0.991	0.636	0.996	0.725
	II	1.000	0.576	1.000	0.680	1.000	0.750
	III	0.987	0.645	0.994	0.717	0.997	0.772
	IV	0.867	0.786	0.925	0.800	0.955	0.824
	V	0.636	0.856	0.877	0.823	0.941	0.833

Note. $E_1(\boldsymbol{\rho})$, D-efficiency; $E_2(\boldsymbol{\rho})$, efficiency for estimating b.

Table 2 Simulated characteristics of the five randomization procedures with $K = 3$ treatments, $n = 150$ subjects, and a "cohort" response-adaptive randomization scheme using 10 cohorts of size 15, based on 1000 simulations

		$b = 0.5$			$b = 1$			$b = 1.5$		
(μ_1, μ_2, μ_3)	Design	$M(E_1)$	$M(E_2)$	TT (SD)	$M(E_1)$	$M(E_2)$	TT (SD)	$M(E_1)$	$M(E_2)$	TT (SD)
$(0, -1, -1)$	I	0.990	0.690	50.0 (1.5)	0.996	0.767	43.1 (1.9)	0.998	0.810	38.9 (2.1)
	II	1.000	0.747	48.8 (1.4)	1.000	0.796	42.4 (1.9)	1.000	0.826	38.4 (2.1)
	III	0.994	0.785	48.1 (1.3)	0.997	0.819	41.9 (1.8)	0.998	0.837	38.1 (2.0)
	IV	0.919	0.870	46.5 (1.4)	0.950	0.873	40.6 (1.8)	0.968	0.879	37.2 (2.1)
	V	0.749	0.924	45.6 (1.6)	0.914	0.893	40.2 (2.0)	0.958	0.890	36.9 (2.2)
$(0, -1, 0)$	I	0.974	0.455	55.6 (1.3)	0.993	0.630	47.9 (1.8)	0.997	0.725	42.8 (2.1)
	II	1.000	0.554	53.6 (1.3)	1.000	0.662	47.1 (1.8)	1.000	0.744	42.2 (2.1)
	III	0.985	0.630	52.1 (1.3)	0.993	0.701	46.2 (1.9)	0.997	0.764	41.7 (2.1)
	IV	0.831	0.800	48.9 (1.6)	0.915	0.795	43.8 (2.0)	0.956	0.812	40.4 (2.2)
	V	0.632	0.890	48.0 (1.8)	0.877	0.816	43.3 (2.1)	0.946	0.820	40.2 (2.4)
$(0, -0.5, -1)$	I	0.985	0.549	53.6 (1.4)	0.995	0.705	45.8 (1.8)	0.998	0.784	40.9 (2.2)
	II	1.000	0.612	52.4 (1.2)	1.000	0.732	45.2 (1.8)	1.000	0.796	40.6 (2.2)
	III	0.990	0.667	51.3 (1.3)	0.996	0.755	44.7 (1.8)	0.998	0.814	40.3 (2.1)
	IV	0.877	0.797	48.8 (1.5)	0.942	0.815	43.3 (1.9)	0.967	0.843	39.3 (2.1)
	V	0.743	0.851	48.1 (1.7)	0.918	0.830	42.7 (2.1)	0.960	0.849	39.2 (2.2)
$(0, -0.5, 1)$	I	0.977	0.485	60.5 (0.9)	0.991	0.632	53.1 (1.6)	0.995	0.723	47.6 (2.0)
	II	1.000	0.576	60.0 (1.0)	1.000	0.680	52.4 (1.6)	1.000	0.750	47.1 (2.0)
	III	0.987	0.645	59.7 (1.2)	0.993	0.716	51.8 (1.7)	0.996	0.769	46.5 (2.0)
	IV	0.867	0.786	59.1 (1.6)	0.918	0.798	50.4 (1.8)	0.947	0.824	45.4 (2.1)
	V	0.656	0.856	59.0 (1.7)	0.842	0.839	50.1 (1.9)	0.920	0.845	45.1 (2.2)

Note. $M(E_1)$, median D-efficiency; $M(E_2)$, median efficiency for estimating b; TT, total observed time; SD, standard deviation.

Table 2 presents key design characteristics: median values of the distributions of efficiency criteria $E_1(\boldsymbol{\rho})$ and $E_2(\boldsymbol{\rho})$ and average total time (TT) observed in the trial, with standard deviation. Since we assume that the response is time to recovery, designs with smaller TT are more ethically appealing. The detailed results including distributions and summary statistics of allocation proportions, and maximum likelihood estimates are not shown here for the sake of brevity but are available upon request.

Table 2 shows the efficiencies of all five designs are consistent with the the theoretical results in Table 1 when $\boldsymbol{\theta} = (\mu_1, \mu_2, \mu_3, b)$ was assumed known. While the balanced randomization (design I) has high median values of $E_1(\boldsymbol{\rho})$, it has lowest median values of $E_2(\boldsymbol{\rho})$, and therefore, it is least efficient for estimating the underlying hazard pattern among the five designs. Designs II, III, IV, and V achieve reasonable trade-off between the two efficiency criteria, and at the same time, they have smaller average values of TT compared to design I. Among the five designs considered, the compound optimal design (III) seems to provide the best compromise between randomization, efficiency, and ethics.

5.3. Simulation of Optimal RAR Designs With Staggered Entry and Delayed Response

In this subsection we report simulation results for optimal RAR designs with an adaptation scheme with staggered entries and delayed responses, as described in subsection 4.2. We consider 12 experimental scenarios for $\boldsymbol{\theta} = (\mu_1, \mu_2, \mu_3, b)$ as in subsection 5.1 and focus on the compound optimal design III with $n = 150$ patients. Patient arrival pattern was

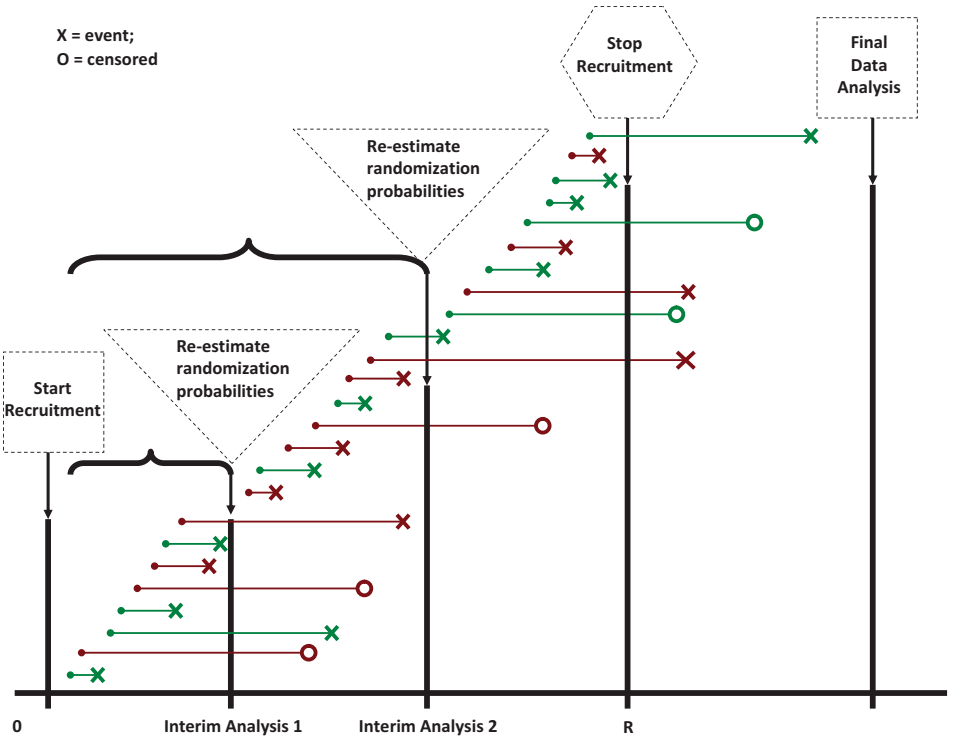

Figure 2 Response-adaptive randomization design with staggered entry and delayed responses.

generated using a Poisson process with the ordered recruitment times uniformly distributed over the recruitment period $(0, R)$. We set $R = 3$; with such a choice of R, the average proportions of study patients with observed outcomes before time R are between 80% to 89% for our considered experimental scenarios. For the adaptation scheme we require that treatment randomization probabilities are updated twice in the trial: after enrollment of 50 patients, and after enrollment of 100 patients. Note that some patients may not have outcome data by the time when the design adaptations are made (Fig. 2).

Table 3 reports theoretical and simulated (average with standard deviation) treatment allocation proportions. The simulated average allocation proportions agree well with the theoretical values. The simulated median values of $E_1(\boldsymbol{\rho})$ and $E_2(\boldsymbol{\rho})$ are very close to the theoretical values. These results show that our designs work as intended in trials where 80% to 89% of patients contribute outcome data during the recruitment phase.

5.4. Effect of Delayed Responses on the Convergence to Target Allocation

We further examined an impact of delayed responses on the convergence of RAR procedures to their target allocation. For the scenario with $\mu_1 = 0$ $\mu_2 = -0.5$, $\mu_3 = -1$, and $b = 0.5$, we simulated RAR design III with $n = 150$ patients as in subsection 5.3 for different lengths of the recruitment period R. We investigated $R = 1$, 1.5, 3, 7, and 9. For such choices of R, the estimated average proportion of study patients $\hat{p}_R$ with observed outcome before R was 47%, 57%, 87%, 91%, and 99%, respectively.

Table 3 Theoretical (Theor.) and simulated characteristics of the DBCD procedure targeting compound optimal allocation with $K = 3$ treatments, $n = 150$ subjects, and a response-adaptive randomization scheme incorporating patient staggered entry and delayed responses, based on 1000 simulations

		$b = 0.5$		$b = 1$		$b = 1.5$	
(μ_1, μ_2, μ_3)		Theor.	Simulated*	Theor.	Simulated*	Theor.	Simulated*
$(0, -1, -1)$	ρ_1	0.220	0.222 (0.027)	0.252	0.252 (0.033)	0.272	0.272 (0.037)
	ρ_2	0.390	0.389 (0.042)	0.374	0.372 (0.045)	0.364	0.364 (0.047)
	ρ_3	0.390	0.389 (0.042)	0.374	0.375 (0.045)	0.364	0.364 (0.045)
	$E_1(\boldsymbol{\rho})$	0.995	0.992	0.997	0.993	0.998	0.994
	$E_2(\boldsymbol{\rho})$	0.821	0.820	0.860	0.859	0.888	0.890
$(0, -1, 0)$	ρ_1	0.226	0.236 (0.031)	0.268	0.267 (0.037)	0.289	0.287 (0.039)
	ρ_2	0.547	0.528 (0.041)	0.465	0.465 (0.044)	0.421	0.420 (0.048)
	ρ_3	0.226	0.237 (0.032)	0.268	0.269 (0.038)	0.289	0.293 (0.042)
	$E_1(\boldsymbol{\rho})$	0.985	0.988	0.994	0.990	0.997	0.993
	$E_2(\boldsymbol{\rho})$	0.630	0.613	0.702	0.703	0.762	0.761
$(0, -0.5, -1)$	ρ_1	0.225	0.228 (0.029)	0.261	0.260 (0.035)	0.282	0.281 (0.038)
	ρ_2	0.276	0.279 (0.039)	0.313	0.316 (0.044)	0.324	0.323 (0.043)
	ρ_3	0.500	0.493 (0.040)	0.426	0.424 (0.047)	0.395	0.396 (0.045)
	$E_1(\boldsymbol{\rho})$	0.990	0.988	0.996	0.992	0.998	0.994
	$E_2(\boldsymbol{\rho})$	0.667	0.663	0.755	0.755	0.811	0.812
$(0, -0.5, 1)$	ρ_1	0.266	0.302 (0.049)	0.314	0.319 (0.056)	0.330	0.331 (0.055)
	ρ_2	0.527	0.415 (0.089)	0.453	0.444 (0.058)	0.415	0.415 (0.057)
	ρ_3	0.207	0.282 (0.058)	0.233	0.237 (0.033)	0.255	0.254 (0.037)
	$E_1(\boldsymbol{\rho})$	0.987	0.983	0.994	0.989	0.997	0.989
	$E_2(\boldsymbol{\rho})$	0.645	0.551	0.717	0.714	0.772	0.775

*Note.**Mean (SD) is reported for allocation proportions; medians are reported for $E_1(\boldsymbol{\rho})$ and $E_2(\boldsymbol{\rho})$.

Figure 3 is a plot of the average values (±S.D.) of the treatment allocation proportions. The target allocation is $\boldsymbol{\rho} = (0.225, 0.275, 0.500)$. We observe that as $\hat{p}_R$ increases from 47% to 99%, convergence of the design to its target improved. We also observed (results not shown here) that as $\hat{p}_R$ increases from 47% to 99%, the design results in more accurate estimates of $\boldsymbol{\theta}$ and simulated values of design efficiencies are getting closer to their theoretical values. These findings reinforce the importance of the assumption that a substantial amount of outcome data must accumulate during the recruitment phase.

6. APPLICATION OF THE PROPOSED METHODOLOGY FOR SURVIVAL TRIALS

In this section we demonstrate utility of our proposed methodology in survival trials. We use a Phase III survival trial in head and neck cancer reported by Fountzilas et al. (2004) as an illustration. The trial used three treatment groups: standard fractionated radiotherapy (RT) alone (group 1), RT concomitantly with cisplatin (group 2), and RT concomitantly with carboplatin (group 3). The objective was to compare each of the experimental treatments 2 and 3 versus control 1 with respect to overall survival.

To match the experimental setup of the trial in Fountzilas et al. (2004), we assume the trial duration is $D = 96$ months and the length of the recruitment period is $R = 55$ months. For the censoring scheme we assume a combination of uniform and administrative censoring (Latta, 1981; Rosenberger and Seshaiyer, 1997). Patient enrollment follows a Poisson process over $(0, R)$, and the observed time for the ith patient in group k is $t_{ik} =$

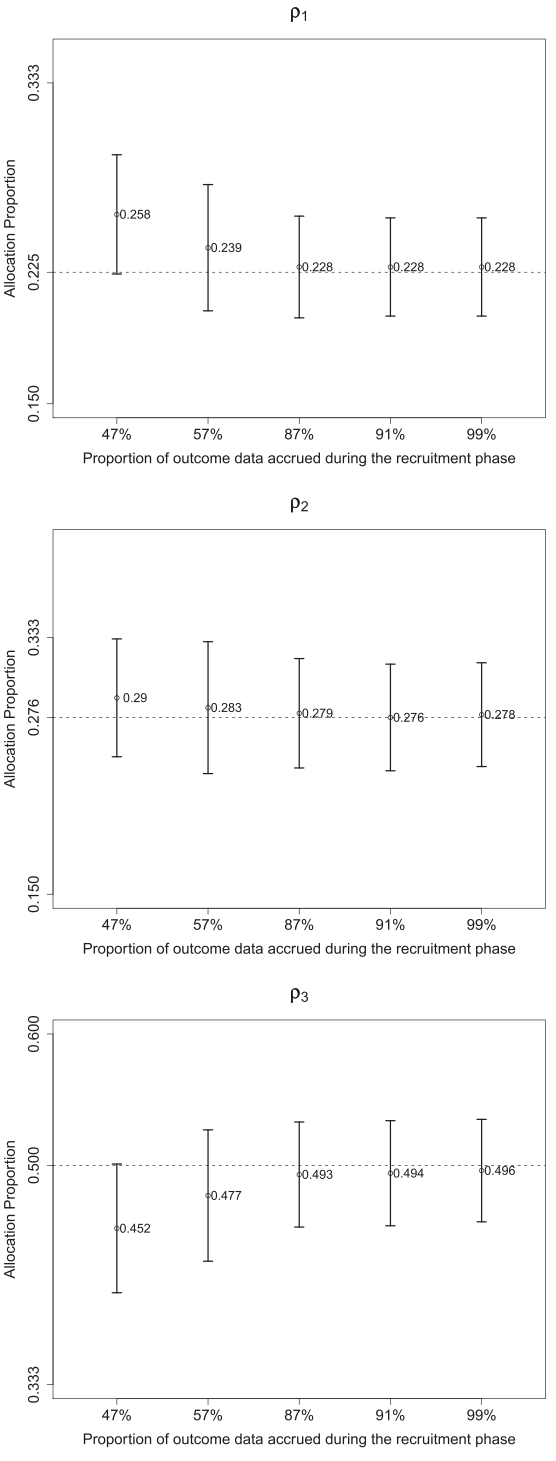

Figure 3 Effect of delayed response on convergence of the compound optimal design (III) to the target allocation $\boldsymbol{\rho} = (0.225, 0.275, 0.500)$. Displayed are means $\pm$SD.

Table 4 The parameter values (μ_1, μ_2, μ_3, b) and treatment hazard ratios to match the median overall survival times of 12.2, 48.6, and 24.5 months of the head and neck cancer trial in Fountzilas et al. (2004)

					Hazard ratio	
Scenario	μ_1	μ_2	μ_3	b	2 vs. 1	3 vs. 1
A	2.81	4.20	3.51	0.85	0.20	0.44
B	2.87	4.25	3.57	1.00	0.25	0.50
C	2.96	4.34	3.66	1.25	0.33	0.57

$\min(T_{ik}, C_{ik}, D - R)$, where T_{ik} is the survival time following a Weibull model (1) and C_{ik} is an independent censoring time uniformly distributed over $(0, D)$. Patients who are alive and have not dropped out by the end of the study are administratively censored.

From Fountzilas et al. (2004), the reported intent-to-treat median overall survival times for treatment groups 1, 2, and 3 are 12.2, 48.6, and 24.5 months, respectively. We consider three choices of $\boldsymbol{\theta} = (\mu_1, \mu_2, \mu_3, b)$ (referred to as scenarios A, B, and C) to match the reported treatment effects with Weibull distributions with cumulative hazard functions $H_k(t) = (te^{-\mu_k})^{1/b}$, $k = 1, 2, 3$. The parameter values and treatment differences expressed as hazard ratios of treatments 2 versus 1 and 3 versus 1 are summarized in Table 4.

Let $\boldsymbol{\rho}^* = (\rho_1^*, \rho_2^*, \rho_3^*)$ denote the D-optimal allocation and $\boldsymbol{\rho}_{\mathrm{E}} = (\rho_{\mathrm{E}1}, \rho_{\mathrm{E}2}, \rho_{\mathrm{E}3})$ denote the ethical allocation with components

$$\rho_{\mathrm{E}k} = \frac{\{\exp(\mu_k/b)\}^2}{\sum_{j=1}^{3}\{\exp(\mu_j/b)\}^2}, \quad k = 1, 2, 3. \tag{10}$$

The allocation proportions in equation (10) are skewed in favor of treatments with longer survival times ($\mu_i \geq \mu_j$ implies $\rho_{\mathrm{E}i} \geq \rho_{\mathrm{E}j}$, with equality if and only if $\mu_i = \mu_j$). To balance the requirements of statistical efficiency and ethics, we consider a weighted optimal allocation

$$\rho_{\alpha k} = \alpha\rho_k^* + (1 - \alpha)\rho_{\mathrm{E}k}, \quad k = 1, 2, 3, \tag{11}$$

where $0 \leq \alpha \leq 1$ is a prespecified trade-off parameter. To calibrate the design, we study operating characteristics of allocation (11) for various choices of α, for different values of $\boldsymbol{\theta}$, and we use balanced allocation $\boldsymbol{\rho}_{\mathrm{B}} = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ as the reference in our comparisons. The operating characteristics include:

- *D-efficiency* of $\boldsymbol{\rho}_\alpha$ relative to $\boldsymbol{\rho}_{\mathrm{B}}$:

$$E_1 = \left\{ \frac{|\boldsymbol{M}^{-1}(\boldsymbol{\rho}_{\mathrm{B}}, \boldsymbol{\theta})|}{|\boldsymbol{M}^{-1}(\boldsymbol{\rho}_\alpha, \boldsymbol{\theta})|} \right\}^{1/4}.$$

  Note that $0 < E_1 < 1$ implies $\boldsymbol{\rho}_\alpha$ is less efficient than $\boldsymbol{\rho}_{\mathrm{B}}$, and $E_1 \geq 1$ implies $\boldsymbol{\rho}_\alpha$ is at least as efficient as $\boldsymbol{\rho}_{\mathrm{B}}$ for estimating $\boldsymbol{\theta}$.
- *D_A-efficiency* of $\boldsymbol{\rho}_\alpha$ relative to $\boldsymbol{\rho}_{\mathrm{B}}$:

$$E_2 = \left\{ \frac{|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{M}^{-1}(\boldsymbol{\rho}_{\mathrm{B}}, \boldsymbol{\theta})\boldsymbol{A}|}{|\boldsymbol{A}^{\mathrm{T}}\boldsymbol{M}^{-1}(\boldsymbol{\rho}_{\alpha}, \boldsymbol{\theta})\boldsymbol{A}|} \right\}^{1/2},$$

where $\boldsymbol{A}^{\mathrm{T}}$ is a 2×4 matrix of contrasts such that $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\theta} = (\mu_2 - \mu_1, \mu_3 - \mu_1)^{\mathrm{T}}$. The $\mathrm{D_A}$-efficiency is a relative measure of precision for estimating the contrast vector. If $0 < E_2 < 1$ then $\boldsymbol{\rho}_{\alpha}$ is less efficient, and if $E_1 \geq 1$ then $\boldsymbol{\rho}_{\alpha}$ is at least as efficient as $\boldsymbol{\rho}_{\mathrm{B}}$.

- *Power* for testing the homogeneity hypothesis $H_0 : \boldsymbol{\mu}_c = 0$ versus $H_1 : \boldsymbol{\mu}_c \neq 0$, where $\boldsymbol{\mu}_c = (\mu_2 - \mu_1, \mu_3 - \mu_1)^{\mathrm{T}}$. We use the Wald test statistic $W_n = \hat{\boldsymbol{\mu}}_c{}^{\mathrm{T}} \hat{\boldsymbol{\Sigma}}_n^{-1} \hat{\boldsymbol{\mu}}_c$, where $\hat{\boldsymbol{\mu}}_c$ is the maximum likelihood estimator of $\boldsymbol{\mu}_c$ and $\hat{\boldsymbol{\Sigma}}_n$ is a consistent estimator of $\boldsymbol{\Sigma}_n = \mathrm{var}\,(\hat{\boldsymbol{\mu}}_c)$. Given n and $\boldsymbol{\theta}$, the power is computed as

$$\mathrm{Power} = \Pr(Y > \chi^2_{2,0.95}),$$

where Y follows a noncentral chi-squared distribution with 2 degrees of freedom and noncentrality parameter $\boldsymbol{\mu}_c^{\mathrm{T}} \boldsymbol{\Sigma}_n \boldsymbol{\mu}_c$ and $\chi^2_{2,0.95}$ is the 95th percentile of the central chi-squared distribution with 2 degrees of freedom.

- *Expected total hazard in the trial.* The cumulative hazard over the interval $(0, D)$ for a single patient in group k is $H_k(D) = (D\mathrm{e}^{-\mu_k})^{1/b}$, $k = 1, 2, 3$. We use the expected total cumulative hazard in the trial as an ethical characteristic of the design:

$$H(D) = n \sum_{k=1}^{3} \rho_k H_k(D).$$

For each of the scenarios A, B, and C from Table 4 we compute the operating characteristics of allocation (11) for the values of α from 0 to 1. The graphical summary of the results is presented in Fig. 4. For $\alpha = 0$ the design has lowest expected total hazard but low values of statistical efficiency measures. As α increases, the D-efficiency, D_A-efficiency, and power are increasing, but the expected hazard is increasing as well. The D_A-efficiency and power curves are not monotone in α. Overall, from Fig. 4 we conclude that $\alpha = 0.5$ is a good trade-off value. Across the considered scenarios, the design with $\alpha = 0.5$ yields the same power, $\geq 89\%$ D-efficiency, $\geq 92\%$ D_A-efficiency, and it is expected to achieve 14% to 23% reductions in total hazard compared to the balanced allocation.

Next we apply the DBCD procedure (9) to target allocation (11) with $\alpha = 0.5$ and compare this RAR design with balanced randomization. The requisite sample size for scenarios A, B, and C to achieve approximately 90% power is 60, 72, and 123 patients, respectively. For the RAR design, randomization probabilities are recalculated twice: after enrollment of $n/3$ and $2n/3$ patients. Each interim analysis is based only on data observed up to that point (see Fig. 2). For the balanced randomization, every patient is randomized to the treatments with probabilities $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ (a completely randomized design, CRD). For each design and scenario combination, a trial is simulated 10,000 times.

Table 5 is the summary of the simulation results. The RAR procedure results in skewed allocations favoring the treatment arms with longer survival times but it has more variable allocation proportions than the CRD. Due to delayed responses, the target allocations for the RAR procedure are not attained perfectly. With regard to the statistical characteristics, the RAR procedure has $\geq 94\%$ median D-efficiency, $\geq 92\%$ median D_A-efficiency, and the same average power as the CRD. At the same time, the RAR procedure

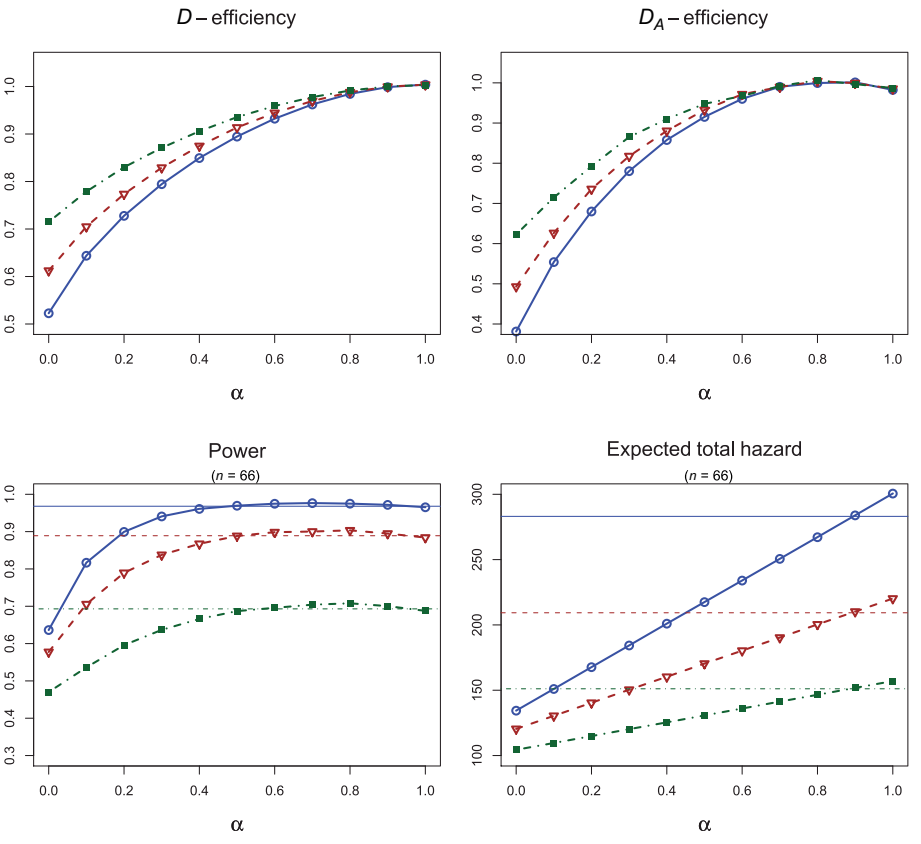

Figure 4 Theoretical operating characteristics of a weighted optimal (WO) allocation design described in section 6 for a three-arm survival trial under three experimental scenarios (A, B, C) from Table 4. The D-efficiency and D_A-efficiency are calculated for the WO allocation relative to the balanced allocation (values less than 1 indicate WO is less efficient than balanced). Scenario A, WO allocation. Scenario A, balanced allocation. Scenario B, WO allocation. Scenario B, balanced allocation. Scenario C, WO allocation. Scenario C, balanced allocation.

has 7% to 9% lower average total hazard, 2 to 4 fewer deaths on average, and longer average total observed survival time compared to the CRD. We conclude that the proposed RAR procedure is both an efficient and an ethical research design for the considered survival trial.

7. SUMMARY

In this article we proposed a practical approach to construct optimal RAR designs for multiple objective clinical trials with $K \geq 2$ treatment arms and censored Weibull outcomes. We derived the D-optimal allocation and gained insights into its properties. As noted by Gwise et al. (2008, p.128), "How to combine the idea of D-optimal or D_A-optimal with minimizing the number of patients into the inferior treatment group remains a further research topic." The methodology of the current article addresses this question for time-to-event trials, and it can be also applied for trials with other types of outcomes, such as binary or continuous (normal). We presented two distinct approaches for constructing

Table 5 Simulated characteristics of completely randomized design (CRD) and a response-adaptive randomization (RAR) design targeting allocation (11) with $\alpha = 0.5$ for a three-arm survival trial under three experimental scenarios (A, B, C) from Table 4, based on 10,000 simulations

	Scenario A ($n = 66$)		Scenario B ($n = 72$)		Scenario C ($n = 123$)	
	CRD	RAR*	CRD	RAR†	CRD	RAR§
ρ_1	0.335 (0.041)	0.281 (0.083)	0.334 (0.032)	0.264 (0.078)	0.334 (0.025)	0.265 (0.072)
ρ_2	0.331 (0.041)	0.390 (0.117)	0.332 (0.032)	0.409 (0.099)	0.333 (0.026)	0.404 (0.080)
ρ_3	0.334 (0.041)	0.330 (0.103)	0.334 (0.032)	0.328 (0.093)	0.333 (0.025)	0.331 (0.080)
$M(D)$	0.995	0.947	0.997	0.944	0.998	0.953
$M(D_A)$	0.999	0.915	0.999	0.930	0.999	0.945
Power	0.916	0.917	0.903	0.901	0.883	0.881
TD (SD)	39 (4)	37 (4)	50 (4)	48 (5)	75 (5)	71 (6)
TH (SD)	305 (107)	281 (104)	283 (70)	258 (68)	295 (49)	274 (48)
TT (SD)	1331 (178)	1375 (153)	1797 (167)	1874 (185)	2757 (220)	2841 (241)

Note. $M(D)$, median D-efficiency; $M(D_A)$, median D_A-efficiency; TD, total number of deaths; TH, total hazard; TT, total time; SD, standard deviation.

* Target allocation is $(0.203, 0.551, 0.246)$.

† Target allocation is $(0.211, 0.526, 0.262)$.

§ Target allocation is $(0.224, 0.495, 0.281)$.

allocation designs that provide a trade-off between study objectives. For the first approach, the trade-off is achieved by taking a mixture of several optimal allocations that reflect the objectives of inferential efficiency and ethical considerations. For the second approach, a compound optimal allocation is obtained by solving a convex optimization problem that includes selected optimality criteria. While optimal allocation designs in this article are based on D-optimality (which may not be optimal for inference using treatment contrasts), our numerical studies show that the proposed designs also achieve high values of D_A-efficiency and power, comparable to those of the balanced allocation design. Finding the D_A-optimal allocation and the allocation maximizing power for Weibull models are important open problems.

The designs proposed in this article depend on the user-defined parameter $\alpha \in [0, 1]$ that weighs the importance of the study objectives. How to judiciously select α is an open problem. In this article we used a graphical approach to determine the value of α to achieve desired trade-off among selected competing criteria. This is essentially the approach proposed by Cook and Wong (1994). Another possibility is to use an adaptive weight $\alpha = \alpha(\boldsymbol{\theta})$, as Baldi Antognini and Giovagnoli (2010) did. One may require that the weight is more skewed toward the inferential objective (D-optimality) at initial stages of the trial, and it becomes skewed in favor of the ethical objective as the trial progresses. This approach requires more calibration, and this is a topic for future research.

In our study we used simulation under various experimental scenarios to demonstrate that the proposed optimal allocation designs can be successfully implemented using RAR. For the considered setups, optimal RAR designs outperform balanced randomization designs when multiple considerations, including selected inferential efficiency criteria and ethical concerns, are incorporated at the study onset. We also presented an application of our methodology using a Phase III survival trial in head and neck cancer as an example. The MATLAB code for finding optimal allocations

and performing simulation studies in this article is fully documented and is available upon request. We are currently developing a user-friendly software interface for simulating and implementing optimal RAR procedures in multi-arm time-to-event trials.

APPENDIX: PROOFS

A.1. Derivation of the Fisher Information Matrix in Equation (2)

The log-likelihood for $\boldsymbol{\theta} = (\mu_1, \ldots, \mu_K, b)$ given data from the patients in the K treatment groups is

$$\log L(\boldsymbol{\theta}) = \sum_{k=1}^{K} \sum_{i=1}^{n_k} \left(-\delta_{ik} \log b + \delta_{ik} z_{ik} - \mathrm{e}^{z_{ik}}\right). \tag{A1}$$

The maximum likelihood estimator $\hat{\boldsymbol{\theta}} = (\hat{\mu}_1, \ldots, \hat{\mu}_K, \hat{b})$ is found by solving the system of score equations

$$0 = \frac{\partial \log L(\boldsymbol{\theta})}{\partial \mu_k} = \sum_{i=1}^{n_k} (\delta_{ik} - \mathrm{e}^{z_{ik}}) \left(-\frac{1}{b}\right), \quad k = 1, \ldots, K,$$

$$0 = \frac{\partial \log L(\boldsymbol{\theta})}{\partial b} = -\frac{1}{b} \sum_{k=1}^{K} \sum_{i=1}^{n_k} \delta_{ik} + \sum_{k=1}^{K} \sum_{i=1}^{n_k} (\delta_{ik} - \mathrm{e}^{z_{ik}}) \left(-\frac{z_{ik}}{b}\right). \tag{A2}$$

Since $E\left(\frac{\partial \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) = \mathbf{0}$, from equation (A2) it follows that

$$E\left(\sum_{i=1}^{n_k} \delta_{ik}\right) = E\left(\sum_{i=1}^{n_k} \mathrm{e}^{z_{ik}}\right), \quad k = 1, \ldots, K, \tag{A3}$$

and

$$E\left(\sum_{k=1}^{K} \sum_{i=1}^{n_k} \delta_{ik}\right) = -E\left(\sum_{k=1}^{K} \sum_{i=1}^{n_k} (\delta_{ik} - \mathrm{e}^{z_{ik}})\, z_{ik}\right). \tag{A4}$$

For any subject $i = 1, \ldots, n_k$ in group k (where $k = 1, \ldots, K$) define the following parameters:

$$\epsilon_k = \Pr(\delta_{ik} = 1), \quad a_k = E\left(z_{ik} \mathrm{e}^{z_{ik}}\right), \quad c_k = E\left(z_{ik}^2 \mathrm{e}^{z_{ik}}\right). \tag{A5}$$

Note that each of the ϵ_k, a_k, and c_k in equation (A5) is a function of $\boldsymbol{\theta}$ and the censoring mechanism used in the trial. The Fisher information matrix is $E\left(-\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathrm{T}}}\right)$ with elements

$$E\left(-\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \mu_k^2}\right) = \frac{1}{b^2} n_k \epsilon_k, \quad k = 1, \ldots, K,$$

$$E\left(-\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \mu_k \partial b}\right) = \frac{1}{b^2} n_k a_k, \quad k = 1, \ldots, K,$$

and

$$E\left(-\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial b^2}\right) = \frac{1}{b^2} \sum_{k=1}^{K} n_k(\epsilon_k + c_k).$$

For an allocation $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_K)^{\mathrm{T}}$, the Fisher information matrix for $\boldsymbol{\theta} = (\mu_1, \ldots, \mu_K, b)$ is

$$M(\boldsymbol{\rho}, \boldsymbol{\theta}) = \frac{n}{b^2} \begin{pmatrix} \mathrm{diag}\{\rho_1\epsilon_1, \ldots, \rho_{\mathrm{K}}\epsilon_{\mathrm{K}}\} & \boldsymbol{x} \\ \boldsymbol{x}^{\mathrm{T}} & \sum_{k=1}^{K} \rho_k(\epsilon_k + c_k) \end{pmatrix}, \tag{A6}$$

where $\boldsymbol{x} = (\rho_1 a_1, \ldots, \rho_K a_K)^{\mathrm{T}}$, and ϵ_k, a_k, and c_k are as in equation (A5). When treatment k is considered alone, the Fisher information matrix for (μ_k, b) is

$$\boldsymbol{M}_k(\mu_k, b) = \frac{n\rho_k}{b^2} \begin{pmatrix} \epsilon_k & a_k \\ a_k & \epsilon_k + c_k \end{pmatrix},$$

and the asymptotic covariance matrix of $(\hat{\mu}_k, \hat{b})$ is

$$\boldsymbol{M}_k^{-1}(\mu_k, b) = \frac{b^2}{n\rho_k} \begin{pmatrix} (\epsilon_k + c_k)/(\epsilon_k d_k) & -a_k/(\epsilon_k d_k) \\ -a_k/(\epsilon_k d_k) & 1/d_k \end{pmatrix},$$

where $d_k = \epsilon_k + c_k - a_k^2/\epsilon_k$, $k = 1, \ldots, K$. Clearly, $d_k > 0$ for $k = 1, \ldots, K$, because d_k is inversely proportional to the asymptotic variance of the maximum likelihood estimator $\hat{b}$ of b, computed using data from treatment k alone.

To derive $|\boldsymbol{M}(\boldsymbol{\rho}, \boldsymbol{\theta})|$, we use the following result from matrix algebra. Let $\boldsymbol{A}$ be a nonsingular $K \times K$ matrix, $\boldsymbol{y}$ be a $K \times 1$ column vector, $\boldsymbol{x}^{\mathrm{T}}$ be a $1 \times K$ row vector, and c be a constant. Then

$$\left| \begin{pmatrix} \boldsymbol{A} & \boldsymbol{y} \\ \boldsymbol{x}^{\mathrm{T}} & c \end{pmatrix} \right| = |\boldsymbol{A}|(c - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{-1}\boldsymbol{y}).$$

It follows from (A6) that $|\boldsymbol{M}(\boldsymbol{\rho}, \boldsymbol{\theta})| = \left(n/b^2\right)^{K+1} \prod_{k=1}^{K} \rho_k \epsilon_k \sum_{k=1}^{K} \rho_k d_k$. □

A.2. Justification of Equation (4)

To minimize $|\boldsymbol{M}^{-1}(\boldsymbol{\rho}, \boldsymbol{\theta})|$, we can equivalently minimize $\log |\boldsymbol{M}^{-1}(\boldsymbol{\rho}, \boldsymbol{\theta})|$. Letting λ be the Lagrange multiplier, we have

$$\frac{\partial}{\partial \rho_k}\left\{\log |\boldsymbol{M}^{-1}(\boldsymbol{\rho}, \boldsymbol{\theta})| + \lambda\left(\sum_{k=1}^{K} \rho_k - 1\right)\right\} = -\frac{1}{\rho_k} - \frac{d_k}{\sum_{k=1}^{K} \rho_k d_k} + \lambda, \quad k = 1, \ldots, K.$$

Setting the derivatives to zero gives

$$\frac{1}{\rho_k} + \frac{d_k}{\sum_{k=1}^{K} \rho_k d_k} = \lambda, \quad k = 1, \ldots, K.$$

Multiplying both sides of the above identity by ρ_k and summing over $k = 1, \ldots, K$, we obtain $\lambda = K + 1$ and equation (4) follows.

Let us show that the obtained solution $\boldsymbol{\rho}^* = (\rho_1^*, \ldots, \rho_K^*)$ is indeed the point of minimum. Let $\mathbf{diag}(1/\boldsymbol{\rho}^{*2})$ be the $K \times K$ diagonal matrix with the jth diagonal element equal to $1/\rho_j^{*2}$, $j = 1, \ldots, K$, let $\boldsymbol{d} = (d_1, \ldots, d_K)^{\mathrm{T}}$, and let $\Delta^* = \sum_{j=1}^{K} \rho_j^* d_j$. To show that the matrix $\boldsymbol{J} = \frac{\partial^2 \log |\boldsymbol{M}^{-1}(\boldsymbol{\rho}, \boldsymbol{\theta})|}{\partial \boldsymbol{\rho} \partial \boldsymbol{\rho}^{\mathrm{T}}}$ evaluated at $\boldsymbol{\rho}^*$ is positive definite, we observe that

$$\boldsymbol{J}(\boldsymbol{\rho}^*, \boldsymbol{d}) = \begin{pmatrix} \frac{1}{\rho_1^{*2}} + \frac{d_1^2}{\Delta^{*2}} & \frac{d_1 d_2}{\Delta^{*2}} & \cdots & \frac{d_1 d_K}{\Delta^{*2}} \\ \frac{d_1 d_2}{\Delta^{*2}} & \frac{1}{\rho_2^{*2}} + \frac{d_2^2}{\Delta^{*2}} & \cdots & \frac{d_2 d_K}{\Delta^{*2}} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{d_1 d_K}{\Delta^{*2}} & \frac{d_2 d_K}{\Delta^{*2}} & \cdots & \frac{1}{\rho_K^{*2}} + \frac{d_K^2}{\Delta^{*2}} \end{pmatrix} = \mathbf{diag}(1/\boldsymbol{\rho}^{*2}) + \frac{1}{\Delta^{*2}} \boldsymbol{d}\boldsymbol{d}^{\mathrm{T}}. \quad \text{(A7)}$$

Therefore, we have $|\boldsymbol{J}(\boldsymbol{\rho}^*, \boldsymbol{d})| = (\prod_{j=1}^{K} \rho_j^{*2})^{-1} \left(1 + \frac{1}{\Delta^{*2}} \sum_{j=1}^{K} (\rho_j^* d_j)^2\right) > 0$, and replacing K by an arbitrary integer $s = 1, \ldots, (K-1)$ we see that all main minors of $\boldsymbol{J}(\boldsymbol{\rho}^*, \boldsymbol{d})$ are positive definite. Hence, $\boldsymbol{J}(\boldsymbol{\rho}^*, \boldsymbol{d})$ is positive definite and $\boldsymbol{\rho}^*$ is the point of local minimum. Since the criterion $\log |\boldsymbol{M}^{-1}(\boldsymbol{\rho}, \boldsymbol{\theta})|$ is strictly convex, any local optimal solution is also a global optimal solution, and this completes the proof. □

A.3. Justification of Property (i) in Section 3.1

Without censoring, $z_{ik} = (\log t_{ik} - \mu_k)/b$ follow a standard extreme value distribution Z with density $f(z) = \mathrm{e}^z \exp(-\mathrm{e}^z)$. Its moment-generating-function is $M_Z(t) = E\left(\mathrm{e}^{tZ}\right) = \Gamma(t+1)$, $t > -1$, where $\Gamma(t) = \int_0^\infty x^{t-1} \mathrm{e}^{-x} dx$ is the gamma function. This implies that for $k = 1, \ldots, K$ we have

$$a_k = E\left(Z\mathrm{e}^Z\right) = \frac{d}{dt} M_Z(t)|_{t=1} = 1 - \gamma,$$

$$c_k = E\left(Z^2\mathrm{e}^Z\right) = \frac{d^2}{dt^2} M_Z(t)|_{t=1} = \frac{\pi^2}{6} - 1 + (1-\gamma)^2,$$

and since $\epsilon_k = 1$ for $k = 1, \ldots, K$, it follows that $d_k = \pi^2/6$ for $k = 1, \ldots, K$. Therefore, $|\boldsymbol{M}(\boldsymbol{\rho}, \boldsymbol{\theta})| = \pi^2/6 \prod_{k=1}^{K} \rho_k$ and the optimization problem (3) simplifies to minimizing $(\prod_{k=1}^{K} \rho_k)^{-1}$ subject to $\sum_{k=1}^{K} \rho_k = 1$. By the arithmetic–geometric mean inequality, the optimal solution is $\boldsymbol{\rho}^* = (1/K, \ldots, 1/K)$. □

A.4. Justification of Property (ii) in Section 3.1

When $K = 2$, we let $\rho_1^* = \rho = 1 - \rho_2^*$ and from equation (4) we have

$$3(d_1 - d_2)\rho^2 - 2(d_1 - 2d_2)\rho - d_2 = 0.$$

If $d_1 = d_2$, the preceding equation is linear with the single root $\rho = \frac{1}{2}$. If $d_1 \neq d_2$, we have two roots given by

$$\rho_\pm = \frac{d_1 - 2d_2 \pm (d_1^2 - d_1 d_2 + d_2^2)^{1/2}}{3(d_1 - d_2)}.$$

If $d_1 > d_2$, we have $\rho_+ > \rho_-$ and the parabola $Q(\rho) = 3(d_1 - d_2)\rho^2 - 2(d_1 - 2d_2)\rho - d_2$ at $\rho = \rho_+$ changes its sign from "–" to "+", which implies that ρ_+ is the point of minimum of the D-optimal criterion. Similarly, if $d_1 < d_2$, we have $\rho_+ < \rho_-$ and the parabola $Q(\rho)$ at $\rho = \rho_+$ changes its sign from "–" to "+," which again implies that ρ_+ is the point of minimum. Further algebra shows that $\rho_+ \in (0, 1)$ (in fact, $\frac{1}{3} \leq \rho_+ \leq \frac{2}{3}$) and, therefore, ρ_+ is the optimal solution in the case when $d_1 \neq d_2$. □

A.5. Justification of Property (iii) in Section 3.1

From equation (4) it follows that

$$(K + 1)\rho_k^* - 1 = \frac{\rho_k^* d_k}{\sum_{j=1}^K \rho_j^* d_j}, \quad k = 1, \ldots, K. \tag{A8}$$

Since $d_k > 0$ for $k = 1, \ldots, K$, we have $0 \leq \frac{\rho_k^* d_k}{\sum_{j=1}^K \rho_j^* d_j} \leq 1$, which, together with equation (A8), implies that $\frac{1}{K+1} \leq \rho_k^* \leq \frac{2}{K+1}$ for $k = 1, \ldots, K$. □

A.6. Justification of Property (iv) in Section 3.1

From the ordering $d_1 \geq d_2 \geq \ldots \geq d_K$ and the facts that $d_j > 0$ and $\frac{1}{K+1} \leq \rho_j^* \leq \frac{2}{K+1}$ for $j = 1, \ldots, K$, it follows that

$$\frac{\rho_2^*}{\rho_1^*} = \frac{d_1}{d_2 + (K + 1)(d_1 - d_2)\rho_2^*} \leq 1,$$

and

$$\frac{\rho_{j+1}^*}{\rho_j^*} = \frac{d_j + (K + 1)(d_1 - d_j)\rho_1^*}{d_{j+1} + (K + 1)(d_1 - d_{j+1})\rho_1^*} \leq 1.$$

The limiting values of $1/(K + 1)$ and $2/(K + 1)$ are achieved as $d_j/d_k \to \infty$ for all $j \neq k$, and $d_j/d_k \to 0$ for all $j \neq k$, respectively. □

REFERENCES

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

Carroll, K. J. (2003). On the use and utility of the Weibull model in the analysis of survival data. *Controlled Clinical Trials* 24:682–701.

Chappell, R., Karrison, T. (2006). Letter to the editor (in response to "Continuous Bayesian adaptive randomization based on event times with covariates"). *Statistics in Medicine* 26:3050–3052.

Cheung, Y. K., Lurdes, Y. T., Wathen, J. K., Thall, P. F. (2006). Continuous Bayesian adaptive randomization based on event times with covariates. *Statistics in Medicine* 25:55–70.

Cook, R. D., Wong, W. K. (1994). One the equivalence of constrained and compound optimal designs. *Journal of the American Statistical Association* 89:687–692.

Eisele, J. (1994). The doubly adaptive biased coin design for sequential clinical trials. *Journal of Statistical Planning and Inference* 38:249–262.

Fountzilas, G., Ciuleanu, E., Dafni, U., Plataniotis G., Kalogera-Fountzila, A., Samantas, E., Athanassiou, E., Tzitzikas, J., Ciuleanu, T., Nikolaou, A., Pantelakos, P., Zaraboukas, T., Zamboglou, N., Daniilidis, J., Ghilezan N. (2004). Concomitant radiochemotherapy vs radiotherapy alone in patients with head and neck cancer. *Medical Oncology* 21(2):95–107.

Gwise, T., Hu, J., and Hu, F. (2008). Optimal biased coins for two-arm clinical trials. *Statistics and Its Interface* 1:125–135.

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

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

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

Jóźwiak, K., Moerbeek, M. (2011). Power analysis for trials with discrete-time survival endpoints. *Journal of Educational and Behavioral Statistics*. doi:10.3102/1076998611424876.

Konstantinou, M., Biedermann, S., Kimber, A. (2014). Optimal designs for two-parameter nonlinear models with application to survival models. *Statistica Sinica* 24(1):415–428.

Latta, R. B. (1981). A Monte Carlo simulation of some two-sample rank tests with censored data. *Journal of the American Statistical Association* 76:713–719.

Li, X, Wang, X. (2013). Response adaptive designs with misclassified responses. *Communications in Statistics-Theory and Methods* 42:2071–2083.

Liang, Y. Y., Carriere Chough, K. (2009). Multiple-objective response-adaptive repeated measurement designs for clinical trials. *Journal of Statistical Planning and Inference* 139:1134–1145.

Rosenberger, W. F. (1993). Asymptotic inference with response-adaptive treatment allocation designs. *Annals of Statistics* 21:2098–2107.

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

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

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

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

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