

Doubly adaptive biased coin designs for balancing competing objectives in time-to-event trials

OLEKSANDR SVERDLOV, YEVGEN RYEZNIK AND WENG KEE WONG*

Many clinical trials have multiple objectives and have a time-to-event outcome that may be modeled using a Weibull distribution. For two-arm trials, we obtain the optimal allocations for a few design criteria and for multi-arm trials, we provide a general approach for finding the optimal allocations. These multi-objective optimal designs meet user-defined tradeoffs among the objectives. We focus on two-objective design problems for estimating model parameters and discriminating whether the treatments have constant hazard (exponential distribution) or non-constant hazard (general Weibull distribution). To target the desired allocations designs, we implement the doubly adaptive biased coin design (DBCD) of Hu and Zhang (2004) and evaluate its effectiveness. We compare performance of the various response-adaptive allocation strategies in an exemplary four-arm trial using a simulation study and show that our proposed response-adaptive randomization designs generally outperform a balanced design when ethics, randomization and estimation efficiency are incorporated at the onset.

KEYWORDS AND PHRASES: Dual-objective clinical trial, Compound-optimal design, Doubly adaptive biased coin design, Ethical concern, Hazard ratio, Randomization design.

1. INTRODUCTION

Response-adaptive randomization for clinical trials has received a lot of attention recently. The theoretical developments are covered in a monograph by Hu and Rosenberger (2006) and a recent summary of the statistical research on response-adaptive randomization and its applications can be found in Pong and Chow (2011, Chapter 15) and in the review paper by Rosenberger, Sverdlov and Hu (2012). Response-adaptive randomization designs are becoming increasingly popular in clinical trials because such designs are increasingly viewed as being more ethical, efficient and practical than one-stage locally single-objective optimal designs for estimating treatment effects.

Hu and Rosenberger (2003) proposed a mathematical template for the development and evaluation of various response-adaptive randomization procedures. For clinical trials with two treatment arms, Jennison and Turnbull (2000) proposed a general approach for deriving optimal allocations

that minimize a weighted sum of treatment sample sizes subject to a constraint on the asymptotic variance of the estimated treatment difference. Their approach led to a series of papers proposing optimal response-adaptive randomization designs. The main references include papers by Rosenberger et al. (2001), Biswas and Mandal (2004), Zhang and Rosenberger (2006, 2007), Gwise, Hu and Hu (2008), Biswas and Bhattacharya (2009, 2010), Bandyopadhyay and Bhattacharya (2010). For implementing optimal allocations in two-treatment trials one can use the ERADE procedure of Hu, Zhang and He (2009), which is an asymptotically best response-adaptive randomization procedure for two-arm trials. In our work, we employ the doubly-adaptive biased coin design (DBCD) of Hu and Zhang (2004), which is currently a popular sequential method for finding optimal allocation rules in a multi-arm trial with several optimal properties.

For multi-objective clinical trials with more than two treatment arms, obtaining an optimal allocation can be challenging. Besides technical difficulties, a main difficulty is that designs optimal under a given design criterion can have low efficiencies under another criterion. A few design approaches have been advocated to handle this problem and most sought to balance the trade-offs among the competing objectives. Tymofeyev et al. (2007) proposed an approach for finding optimal allocations for multi-arm clinical trials based on minimization of some clinically relevant measure (e.g., expected number of treatment failures) subject to constraints on the size of the non-centrality parameter of the chi-square test and the treatment allocation proportions. Their approach was applied to develop optimal response-adaptive randomization procedures (Zhu and Hu, 2009; Jeon and Hu, 2010). For clinical trials where the primary concern is estimation efficiency rather than hypothesis testing, the optimality criterion can be based on the Fisher information matrix. For example, Wong and Zhu (2008) considered normal responses with heteroscedasticity and obtained locally D_A -optimal designs that maximize efficiency for estimating treatment contrasts. Their approach allowed different sets of contrasts to be estimated with different levels of user-specified efficiencies, with more important ones requiring higher efficiencies. In a similar vein, Gwise, Zhu and Hu (2011) proposed the D -optimal and the D_A -optimal biased coin designs for multi-treatment trials with heteroscedastic outcomes. Biswas, Mandal and Bhattacharya (2011) obtained optimal allocations for trials with $K \geq 2$ treatments and heteroscedastic outcomes assuming at least one and at most

*Corresponding author.

$(K - 1)$ restrictions on the variances of the treatment contrasts.

Most of the aforementioned papers provide response-adaptive randomization designs for binary or continuous outcomes. Many clinical trials have censored time-to-event outcomes and research on response-adaptive randomization designs for such trials is limited. Some recent references are Cheung et al. (2006), Zhang and Rosenberger (2007), Bandyopadhyay, Biswas, and Bhattacharya (2009, 2010). Most recently, Sverdlov, Ryznik and Wong (2012) derived locally D -optimal designs for multi-arm trials with a censored Weibull time-to-event outcome, and proposed a way to construct compound optimal designs that target inference efficiency and skew more patients to more efficacious doses at the same time. They also explored weighted optimality designs that provide tradeoffs between several optimal allocations by minimizing the Euclidean distance and the Kullback-Leibler distance between the allocations. They did not further discuss how to find the “optimal” weights in the combined criterion, except that there was a general sense that larger weights should be given to the more important criteria.

This paper focuses on optimal designs for a multiple-arm trial with a censored Weibull outcome and differs from the work in Sverdlov, Ryznik and Wong (2012) in a number of ways. Here, we first obtain optimal allocations for a single-objective trial with Weibull models in which efficient estimation of treatment contracts is the primary concern. For a multiple-objective trial, we provide a systematic and meaningful way to select the proper weights in the convex combination of the criteria to find the optimal designs that achieve user-specified levels of the design efficiencies. Additionally, we provide a general formal design strategy to incorporate multiple objectives in the trial. These objectives may include parameter estimation, hypothesis testing, protection against model uncertainty and ethical concerns and they may also have different levels of importance to the researcher. We also discuss implementation issues of the various adaptive optimal allocation schemes using the DBCD procedure and present a few new simulation studies for a multiple-arm dual-objective trial with censored Weibull outcomes.

Section 2 provides the statistical background, model specification for a multi-arm trial and discusses optimality criteria that include D_A -optimal allocation, trace-optimal allocation, and optimal allocation for estimating treatment hazard ratios for Weibull models. In Section 3, we focus on different optimal allocations for estimating treatment contrasts in a two-arm single-objective trial and study their operating characteristics. In Section 4, we formulate the design problem for a two-objective trial and show how to systematically determine dual-objective optimal designs that meet user-specified efficiency requirements. We also discuss implementation issues using a doubly adaptive biased coin design procedure of Hu and Zhang (2004) and compare the performance of the various response-adaptive allocation strategies and a balanced randomization design in an exemplary four-arm trial. We conclude with a discussion in Section 5.

2. STATISTICAL BACKGROUND AND OPTIMALITY CRITERIA

We consider a multi-arm clinical trial with K treatment groups and a time-to-event outcome that comes from a member of the family of Weibull distributions. Let T_{ik} be the response time from subject i in the k th treatment group. The statistical model of interest is

$$(1) \quad \log T_{ik} = \mu_k + bW_{ik},$$

where $b > 0$, μ_k represents the effect of treatment $k = 1, \dots, K$ and i indexes the patient within the group. The error terms W_{ik} are assumed to be independent and follow the standard extreme value distribution with probability density function $f(w) = e^w \exp(-e^w)$. The random variable T_{ik} follows a Weibull distribution with survivor function $S(t) = \exp\{-(e^{-\mu_k t})^{1/b}\}$, $t > 0$ and we denote all unknown model parameters by $\boldsymbol{\theta} = (\mu_1, \dots, \mu_K, b)'$.

Let $C_i > 0$ be the censoring time for the i th patient, let $t_{ik} = \min(T_{ik}, C_i)$ be the observed time and let $\delta_{ik} = \mathbf{1}_{\{t_{ik}=T_{ik}\}}$ be the indicator of the event of interest. It is assumed that C_i is independent of T_{ik} and the total sample size n is pre-determined in advance either by cost or the number of patients that can be realistically included in the study. The design question is how to optimally allocate $n_k \geq 0$ patients to the k th group so that $n = \sum_{k=1}^K n_k$. Letting $z_{ik} = (\log t_{ik} - \mu_k)/b$ be the standardized $\log t_{ik}$, the log-likelihood for $\boldsymbol{\theta}$ given data from all subjects in the K treatment groups is

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

The maximum likelihood estimator $\hat{\boldsymbol{\theta}} = (\hat{\mu}_1, \dots, \hat{\mu}_K, \hat{b})'$ of $\boldsymbol{\theta}$ is obtained by solving the following system of score equations $\frac{\partial \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}$ numerically:

$$\begin{aligned} 0 &= \frac{\partial \log L(\boldsymbol{\theta})}{\partial \mu_k} = \sum_{i=1}^{n_k} (\delta_{ik} - e^{z_{ik}}) \left(-\frac{1}{b} \right), \quad k = 1, \dots, K, \\ 0 &= \frac{\partial \log L(\boldsymbol{\theta})}{\partial b} = -\frac{1}{b} \sum_{k=1}^K \sum_{i=1}^{n_k} \delta_{ik} + \sum_{k=1}^K \sum_{i=1}^{n_k} (\delta_{ik} - e^{z_{ik}}) \left(-\frac{z_{ik}}{b} \right). \end{aligned}$$

To determine the Fisher information matrix for $\boldsymbol{\theta}$, it is instructive to work with designs defined by the proportions of subjects allocated to the K groups. Let $\boldsymbol{\rho} = (\rho_1, \dots, \rho_K)'$ be the design that allocates ρ_k proportion of the total subjects to treatment group k , $0 \leq \rho_k \leq 1$ and $\sum_{k=1}^K \rho_k = 1$. For a trial with n patients, this means roughly $n_i = n\rho_i$ patients are assigned to treatment i subject to $n = n_1 + \dots + n_K$. A direct calculation shows the Fisher information matrix for $\boldsymbol{\theta} = (\mu_1, \dots, \mu_K, b)'$ using design $\boldsymbol{\rho}$

is

$$\begin{aligned} \mathbf{M}(\boldsymbol{\rho}, \boldsymbol{\theta}) &= E \left(-\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right) \\ &= \frac{n}{b^2} \begin{pmatrix} \text{diag}\{\rho_1 \epsilon_1, \dots, \rho_K \epsilon_K\} & \mathbf{x} \\ \mathbf{x}' & \sum_{k=1}^K \rho_k (\epsilon_k + c_k) \end{pmatrix}. \end{aligned}$$

Here $\mathbf{x} = (\rho_1 a_1, \dots, \rho_K a_K)'$, $\epsilon_k = \Pr(\delta_{ik} = 1)$, $a_k = E(z_{ik} e^{z_{ik}})$, $c_k = E(z_{ik}^2 e^{z_{ik}})$ and ϵ_k , a_k , c_k all depend on μ_k , b and the censoring mechanism used in the trial.

Many design optimality criteria are formulated in terms of the information matrix. A popular criterion for estimating all model parameters in the model is D -optimality that requires the determinant of the information matrix to be maximized to minimize the volume of the confidence ellipsoid for $\boldsymbol{\theta}$. This is the same as finding a design to minimize the determinant or some monotonic increasing function of the determinant of the inverse of the Fisher Information matrix given by

$$\mathbf{M}^{-1}(\boldsymbol{\rho}, \boldsymbol{\theta}) = \frac{b^2}{n} \begin{pmatrix} \text{diag}\{\frac{1}{\rho_1 \epsilon_1}, \dots, \frac{1}{\rho_K \epsilon_K}\} + \frac{1}{\Delta} \mathbf{y} \mathbf{y}' & -\frac{1}{\Delta} \mathbf{y} \\ -\frac{1}{\Delta} \mathbf{y}' & \frac{1}{\Delta} \end{pmatrix},$$

where $\mathbf{y} = (a_1/\epsilon_1, \dots, a_K/\epsilon_K)'$, $\Delta = \sum_{k=1}^K \rho_k d_k$, and

$$d_k = \epsilon_k + c_k - \frac{a_k^2}{\epsilon_k} > 0, \quad k = 1, \dots, K.$$

The value of d_k is inversely proportional to the variance of \hat{b} when \hat{b} is computed using data from treatment group k alone.

In practice, we frequently wish to compare $(K-1)$ experimental treatments with a placebo or an active placebo. For the $(K-1)$ pairwise comparisons, we want to estimate the vector $\mathbf{A}'\boldsymbol{\theta} = (\mu_2 - \mu_1, \dots, \mu_K - \mu_1)'$, where \mathbf{A}' is an appropriately chosen $(K-1) \times (K+1)$ matrix of contrasts. The Fisher information matrix for $\mathbf{A}'\boldsymbol{\theta}$ is

(2)

$$\begin{aligned} \mathbf{A}'\mathbf{M}^{-1}(\boldsymbol{\rho}, \boldsymbol{\theta})\mathbf{A} &= \frac{b^2}{n} \left(\text{diag} \left\{ \frac{1}{\rho_2 \epsilon_2}, \dots, \frac{1}{\rho_K \epsilon_K} \right\} + \frac{1}{\rho_1 \epsilon_1} \mathbf{1} \mathbf{1}' + \frac{1}{\Delta} \mathbf{u} \mathbf{u}' \right), \end{aligned}$$

where $\mathbf{u} = (a_1/\epsilon_1 - a_2/\epsilon_2, \dots, a_1/\epsilon_1 - a_K/\epsilon_K)'$ and $\mathbf{1} = (1, \dots, 1)'$. This matrix can now be minimized in some sense by choice of design so that we have maximal information of $\mathbf{A}'\boldsymbol{\theta}$. Some possibilities are:

2.1 The D_A -optimal allocation

The D_A -optimal allocation minimizes the criterion $|\mathbf{A}'\mathbf{M}^{-1}(\boldsymbol{\rho}, \boldsymbol{\theta})\mathbf{A}|$. In general, it is quite challenging to find the analytical formula for the D_A -optimal allocation. To build intuition, let us consider the case when there is no censoring in any of the treatment groups. Then $\epsilon_k = 1$ for

$k = 1, \dots, K$ and $a_1 = \dots = a_K$, and so $\mathbf{u} = \mathbf{0}$ and the third summand in (2) is zero. It follows that when there is no censoring, we have $|\mathbf{A}'\mathbf{M}^{-1}(\boldsymbol{\rho}, \boldsymbol{\theta})\mathbf{A}| = (b^2/n)^{K+1} (\prod_{k=1}^K \rho_k)^{-1}$, and it is easily found using the arithmetic-geometric mean inequality that the D_A -optimal allocation is the balanced allocation vector $\boldsymbol{\rho}^* = (1/K, \dots, 1/K)'$.

In the case with censoring, let us assume that we have $K = 2$ treatment groups. If we let $\rho_1 = p = 1 - \rho_2$, we have

$$g(p) = |\mathbf{A}'\mathbf{M}^{-1}(\boldsymbol{\rho}, \boldsymbol{\theta})\mathbf{A}| = \frac{1}{p\epsilon_1} + \frac{1}{(1-p)\epsilon_2} + \frac{\left(\frac{a_1}{\epsilon_1} - \frac{a_2}{\epsilon_2}\right)^2}{pd_1 + (1-p)d_2}.$$

To determine the D_A -optimal allocation, we find the minimum of $g(p)$ over the interval $(0, 1)$. A direct calculation shows

$$g'(p) = -\frac{1}{p^2 \epsilon_1} + \frac{1}{(1-p)^2 \epsilon_2} - \frac{(d_1 - d_2) \left(\frac{a_1}{\epsilon_1} - \frac{a_2}{\epsilon_2}\right)^2}{(pd_1 + (1-p)d_2)^2},$$

$\lim_{p \rightarrow 0^+} g'(p) = -\infty$, $\lim_{p \rightarrow 1^-} g'(p) = +\infty$ and $g''(p) > 0$ for all p in $(0, 1)$. This implies that there is at least one zero of $g'(p)$ in the interval $(0, 1)$ and because $g'(p)$ is strictly monotone increasing in p , the equation $g'(p) = 0$ has a unique root p^* , which is the point of minimum of $g(p)$. The value of p^* can be found numerically once we have nominal values of ϵ_k , a_k , and d_k , for $k = 1, 2$.

2.2 The trace-optimal allocation

The trace-optimal allocation minimizes trace of $\mathbf{A}'\mathbf{M}^{-1}(\boldsymbol{\rho}, \boldsymbol{\theta})\mathbf{A}$. From (2), we have

$$\begin{aligned} \text{tr}\{\mathbf{A}'\mathbf{M}^{-1}(\boldsymbol{\rho}, \boldsymbol{\theta})\mathbf{A}\} &= \frac{b^2}{n} \left\{ \frac{K-1}{\rho_1 \epsilon_1} + \sum_{k=2}^K \frac{1}{\rho_k \epsilon_k} + \frac{1}{\Delta} \sum_{k=2}^K \left(\frac{a_1}{\epsilon_1} - \frac{a_k}{\epsilon_k} \right)^2 \right\}. \end{aligned}$$

If there is no censoring in any of the treatment groups, we find the proportions ρ_1, \dots, ρ_K that minimize

$$\text{tr}\{\mathbf{A}'\mathbf{M}^{-1}(\boldsymbol{\rho}, \boldsymbol{\theta})\mathbf{A}\} = \frac{b^2}{n} \left\{ \frac{K-1}{\rho_1} + \sum_{k=2}^K \frac{1}{\rho_k} \right\}$$

subject to $\sum_{k=1}^K \rho_k = 1$. Using the harmonic-arithmetic mean inequality, the trace-optimal allocation proportions are found as

$$\begin{aligned} \rho_1^* &= \frac{\sqrt{K-1}}{\sqrt{K-1} + K-1}, \\ \rho_k^* &= \frac{1}{\sqrt{K-1} + K-1}, \quad k = 2, \dots, K. \end{aligned}$$

This is a well-known result given in Fleiss (1986, p. 96). In the case with censoring and $K = 2$ treatment groups, the trace-optimal allocation is the same as the D_A -optimal allocation described in subsection 2.1.

2.3 Optimal designs for hazard ratio estimation

In a survival trial with K treatment groups, the main interest is often in estimating hazard ratios of $(K - 1)$ experimental treatments versus “control”. For model (1), the hazard function in the k th treatment group is equal to $h_k(y) = b^{-1} \exp((y - \mu_k)/b)$, and the hazard ratio comparing treatment k versus treatment 1 is given by $\exp((\mu_1 - \mu_k)/b)$, $k = 2, \dots, K$. We are interested in estimating with maximum precision the vector of log-hazard ratios

$$\boldsymbol{\nu}(\boldsymbol{\theta}) = \left(\frac{\mu_1 - \mu_2}{b}, \dots, \frac{\mu_1 - \mu_K}{b} \right)',$$

which is a nonlinear function of $\boldsymbol{\theta} = (\mu_1, \dots, \mu_K, b)'$. By the invariance property of the maximum likelihood estimates, we have $\boldsymbol{\nu}(\hat{\boldsymbol{\theta}}) = ((\hat{\mu}_1 - \hat{\mu}_2)/\hat{b}, \dots, (\hat{\mu}_1 - \hat{\mu}_K)/\hat{b})'$ and the delta method approximates the asymptotic variance-covariance matrix of $\boldsymbol{\nu}(\hat{\boldsymbol{\theta}})$ by

$$\mathbf{Avar} \left(\boldsymbol{\nu}(\hat{\boldsymbol{\theta}}) \right) = \left(\frac{\partial \boldsymbol{\nu}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)' \mathbf{M}^{-1}(\boldsymbol{\rho}, \boldsymbol{\theta}) \left(\frac{\partial \boldsymbol{\nu}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right),$$

where the $(K - 1) \times (K + 1)$ matrix $\left(\frac{\partial \boldsymbol{\nu}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)'$ is given by

$$\left(\frac{\partial \boldsymbol{\nu}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)' = \begin{pmatrix} \frac{1}{b} & -\frac{1}{b} & 0 & \dots & 0 & -\frac{\mu_1 - \mu_2}{b^2} \\ \frac{1}{b} & 0 & -\frac{1}{b} & \dots & 0 & -\frac{\mu_1 - \mu_3}{b^2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{1}{b} & 0 & 0 & \dots & -\frac{1}{b} & -\frac{\mu_1 - \mu_K}{b^2} \end{pmatrix}.$$

When $K = 2$, we have $\rho_1 = p = 1 - \rho_2$, $\boldsymbol{\nu} = (\mu_1 - \mu_2)/b$, $\frac{\partial \boldsymbol{\nu}}{\partial \boldsymbol{\theta}} = (1/b, -1/b, -(\mu_1 - \mu_2)/b^2)'$, and

$$(3) \quad \mathbf{Avar} \left(\frac{\hat{\mu}_1 - \hat{\mu}_2}{\hat{b}} \right) = \frac{1}{p\epsilon_1} + \frac{1}{(1-p)\epsilon_2} + \frac{\left(\frac{a_1}{\epsilon_1} - \frac{a_2}{\epsilon_2} + \frac{\mu_1 - \mu_2}{b} \right)^2}{pd_1 + (1-p)d_2}.$$

To minimize the asymptotic variance of the estimated hazard ratio, we now find the value of $p \in (0, 1)$ that minimizes the expression in (3). This problem is similar to the problem of finding the D_A -optimal allocation in subsection 2.1 but the optimal allocation proportions ρ_1 and ρ_2 for both cases may not be the same.

In the next section, we focus on two-arm trials which are ubiquitous in medical studies and make specific comparisons.

3. COMPARISON OF OPTIMAL ALLOCATION STRATEGIES FOR TWO-ARM TRIALS

For two-arm trials, it is possible to provide formulae for different optimal allocation schemes. We consider a two-arm

trial in which patients are followed up during a fixed predetermined period of time $\tau > 0$. We set $\tau = 1/(-\log(0.1))$, which means that an observation occurs if the logarithm of the event time is less than or equal to the 10th percentile of the standard extreme value distribution, and the observation is censored otherwise.

Before any design is implemented, it is always instructive to understand its operating characteristics and its relative performance compared with alternative designs under a broad range of assumptions and different measures of goodness. This is because design can affect the quality of inference and its performance can vary substantially from one criterion to another. For example, consider the Weibull distribution with parameters $b = 0.75$ (increasing hazard), $\mu_1 = 0$, and let μ_2 range from -1 to 1 in a two-arm trial and we want to compare the following six allocation strategies using four optimality measures:

I The D -optimal allocation (Sverdlov, Ryznik and Wong, 2012):

$$\rho_1 = \frac{d_1 - 2d_2 + \sqrt{d_1^2 - d_1d_2 + d_2^2}}{3(d_1 - d_2)}, \quad \rho_2 = 1 - \rho_1.$$

II The D_A -optimal allocation (subsection 2.1)

III The optimal allocation for hazard ratio estimation (subsection 2.3)

IV The allocation minimizing the average hazard of a Weibull distribution subject to the restriction on the non-centrality parameter (Zhang and Rosenberger (2007), formula (14), p. 162). For our model this allocation is given by

$$\rho_1^* = \frac{\sqrt{e^{-\mu_2} \left(1 + \frac{c_1}{\epsilon_1} \right) / d_1}}{\sqrt{e^{-\mu_2} \left(1 + \frac{c_1}{\epsilon_1} \right) / d_1} + \sqrt{e^{-\mu_1} \left(1 + \frac{c_2}{\epsilon_2} \right) / d_2}},$$

$$\rho_2^* = 1 - \rho_1^*.$$

V The allocation minimizing the average hazard of a Weibull distribution assuming the common constant follow-up time for the patients (Zhang and Rosenberger (2007), formula (15), p. 163):

$$\rho_1^{**} = \frac{\sqrt{e^{-\mu_2/b} \left(1 + \frac{c_1}{\epsilon_1} \right) / d_1}}{\sqrt{e^{-\mu_2/b} \left(1 + \frac{c_1}{\epsilon_1} \right) / d_1} + \sqrt{e^{-\mu_1/b} \left(1 + \frac{c_2}{\epsilon_2} \right) / d_2}},$$

$$\rho_2^{**} = 1 - \rho_1^{**}.$$

VI The balanced allocation $\boldsymbol{\rho} = (1/2, 1/2)$

The four optimality measures are (i) D -efficiency, (ii) D_A -efficiency, (iii) Efficiency of hazard ratio estimation, and (iv) the size of the non-centrality parameter of the statistical test based on the maximum likelihood estimator $(\hat{\mu}_1, \hat{\mu}_2, \hat{b})$ for testing the hypotheses $H_0 : \mu_1 = \mu_2$ vs. $H_A : \mu_1 \neq \mu_2$.

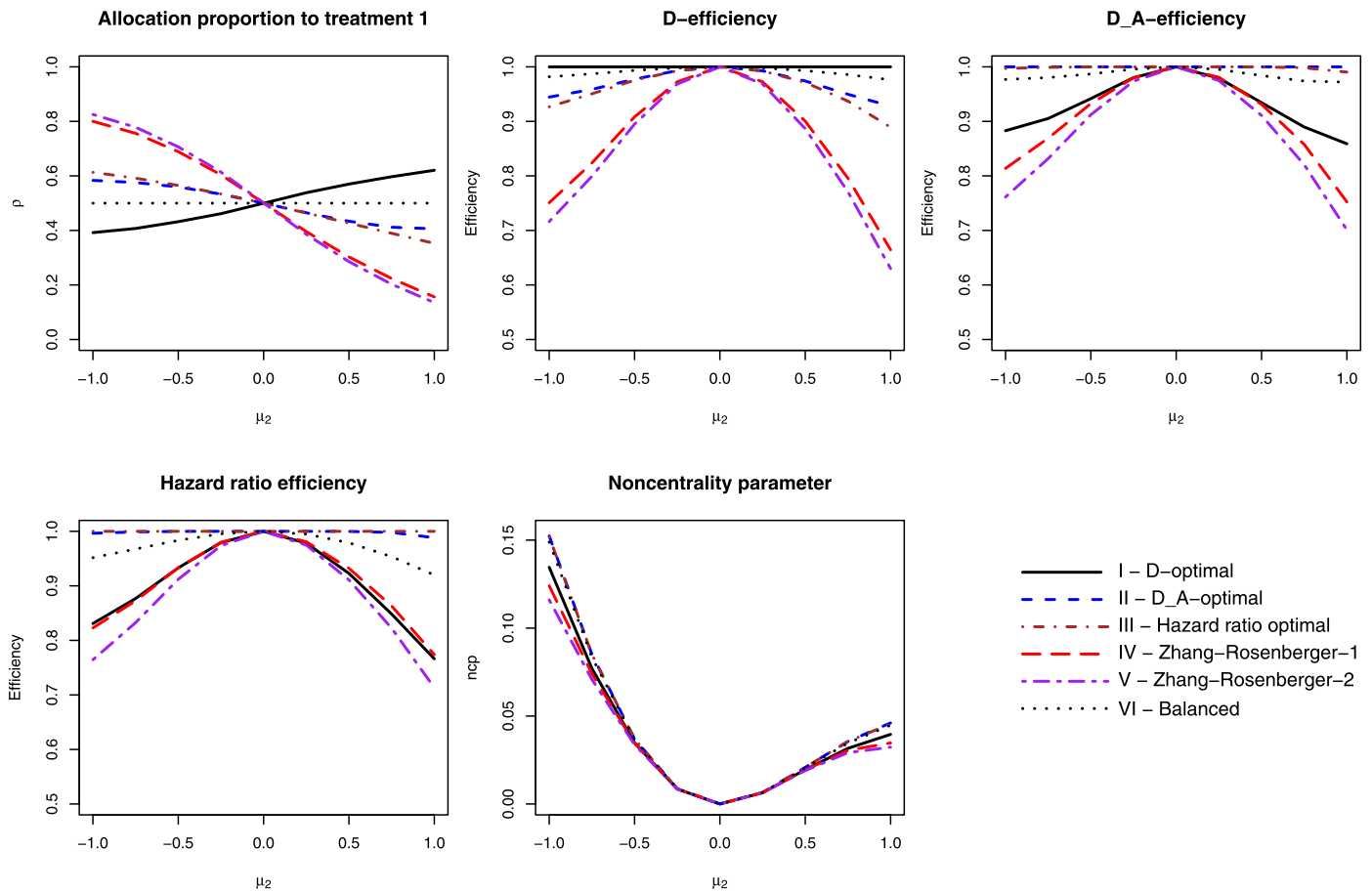


Figure 1. Operating characteristics of the various allocation strategies for a hypothetical two-arm trial with constant follow-up time for each patient.

Figure 1 is a graphical summary of the allocations. The upper left plot of the treatment allocation proportions shows the D -optimal allocation (allocation I) is skewed in favor of the treatment with a lower value of the parameter μ_k , $k = 1, 2$. In other words, $\rho_1 < 1/2$ for $\mu_2 < \mu_1 = 0$ and $\rho_1 \geq 1/2$ for $\mu_2 \geq 0$. In contrast, each of the allocations II, III, IV, and V are skewed towards the treatment group with higher value of the μ_k . The optimal allocations II and III seem always closer to 0.5 than the optimal allocations IV and V.

From the three efficiency plots (D -efficiency, D_A -efficiency, and hazard ratio efficiency) we observe that each of the allocations I, II, and III is best according to the one criterion it optimizes. We also note that the balanced allocation is quite efficient, and the two allocations of Zhang and Rosenberger (2007) (allocations IV and V) exhibit substantial loss in the efficiencies as the treatment difference increases (μ_2 deviates from $\mu_1 = 0$). The plot of the noncentrality parameter (ncp) also shows that the six allocation strategies have similar values of the ncp when the two treatments have similar performance, i.e. μ_2 is close to μ_1 , and as the treatment difference increases, allocations IV and V have smaller values of the ncp, and hence are potentially less

powerful, than the other allocations. Figure 1 also demonstrates that as the treatment difference increases, each of these allocations can have very different efficiencies; for instance, the D_A -efficiencies of the allocation rules IV and V are more than 20% lower than the D_A -efficiencies of the balanced design, and more than 10% lower than D_A -efficiencies of the D -optimal design as μ_2 approaches ± 1 .

These results reinforce that design considerations are important and one should implement a design only after a careful comparison of the alternatives. In our example, unsurprisingly, no single allocation is best in terms of all the four criteria. The operating characteristics of competing allocation strategies should be evaluated under a range of hypothetical experimental scenarios to appreciate the design characteristics and tradeoff among all candidate designs.

4. OPTIMAL ALLOCATIONS FOR DUAL-OBJECTIVE TRIALS

Nowadays, it is unrealistic that trials are designed and carried out with only one single goal in mind. Today the rising cost of experimentation has increasingly led researchers

to look for more efficient designs that enable them to reliably answer multiple scientific questions at the same time without raising cost. In particular, they want efficient designs that are able to incorporate study objectives and concerns more accurately. Frequently, not all parameters are equally interesting and objectives may have different levels of interest. For example, a researcher wants to estimate model parameters but is concerned about model inadequacy, or some parameters in a model have more meaningful biological interpretations than others and so should be estimated with greater accuracy than the rest. A multiple-objective optimal design incorporates these aims and practical concerns at the design stage and delivers user-specified efficiencies for one or more objectives.

Sverdlov, Ryzhnik and Wong (2012) considered a dual-objective optimization problem using a Weibull model for a time-to-event trial. The first objective was to maximize information for $\theta = (\mu_1, \dots, \mu_K, b)'$, which can be achieved by minimizing $\Phi_1(\rho) = \log |\mathbf{M}^{-1}(\rho, \theta)|$. The second objective was to have the most accurate inference for the parameter b by minimizing $\Phi_2(\rho) = -\log \Delta$ over all possible allocation schemes for the K -treatment groups. For a user-selected constant $0 \leq \alpha \leq 1$, they considered minimizing a convex combination

$$(4) \quad \alpha \Phi_1(\rho) + (1 - \alpha) \Phi_2(\rho)$$

subject to $\sum_{k=1}^K \rho_k = 1$. If $\alpha = 1$, we have the D -optimal allocation, which guarantees maximum information for θ . If $\alpha = 0$, the optimization problem reduces to minimizing $\Phi_2(\rho) = -\log \Delta$, and the optimal allocation 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 two-fold. First, it is optimal for discriminating between the exponential model ($b = 1$) and a general Weibull model ($b \neq 1$). Second, it is attractive from an individual ethics perspective in clinical trials where shorter response times are desirable (e.g., time to recovery or time until pain symptoms disappear). The matter is that for many censoring schemes (including a censoring scheme with constant follow-up period for each patient, considered in this paper), the d_k is monotonically decreasing in μ_k when b is fixed. Therefore, the treatment group with the minimum value of μ_k has the maximum value of d_k , and the optimization problem with $\alpha = 0$ results in allocating all subjects to this treatment group, thereby achieving an allocation which is “best” from an individual ethics perspective.

If $0 < \alpha < 1$, then we have an allocation providing a tradeoff between two optimality criteria. For $0 < \alpha \leq 1$, the minimizer of (4) is a vector $\rho^* = (\rho_1^*, \dots, \rho_K^*)'$ whose components satisfy the following nonlinear system of K equations:

$$\frac{\alpha}{\rho_k^*} + \frac{d_k}{\sum_{k=1}^K \rho_k^* d_k} = \alpha K + 1, \quad k = 1, \dots, K,$$

and for $k = 1, \dots, K$ we have

$$(5) \quad \frac{\alpha}{\alpha K + 1} \leq \rho_k^* \leq \frac{1 + \alpha}{\alpha K + 1}.$$

Note that the optimal allocation vector ρ^* depends on unknown model parameters $\theta = (\mu_1, \dots, \mu_K, b)'$ through d_1, \dots, d_K , and therefore it is *locally* optimal, as are all optimal designs in Section 3.

4.1 Choosing the value of the tradeoff parameter

An important question is how to determine the right value of α that provides the desired tradeoff between the objectives. One can use inequalities (5) to calculate the lower and upper bounds on the allocation proportions for a given value of α . For a more rigorous approach, it is useful to consider the relative efficiencies under different criteria.

Let ρ_1^* denote the D -optimal allocation rule obtained by minimizing $\log |\mathbf{M}^{-1}(\rho, \theta)|$, and ρ_2^* denote the allocation rule obtained by minimizing $-\log \Delta$. For an arbitrary allocation $\rho = (\rho_1, \dots, \rho_K)'$, define the efficiency of ρ relative to ρ_1^* as

$$E_1(\rho) = \left\{ \frac{|\mathbf{M}^{-1}(\rho_1^*, \theta)|}{|\mathbf{M}^{-1}(\rho, \theta)|} \right\}^{1/(K+1)},$$

and the efficiency of ρ relative to $\rho_2^* = (\rho_{21}^*, \dots, \rho_{2K}^*)'$ as

$$E_2(\rho) = \frac{\sum_{k=1}^K \rho_k d_k}{\sum_{k=1}^K \rho_{2k}^* d_k}.$$

If the efficiency is equal to 0.5, the design ρ has to be replicated twice to do as well as the optimal design. Clearly, both $E_1(\rho)$ and $E_2(\rho)$ are between 0 and 1 and high values are desirable.

To find a design that delivers user-prescribed levels of efficiency for the two criteria, we use the approach of Cook and Wong (1994). They showed that when we have convex criteria, two sets of optimization problems are equivalent. The first problem considers a convex combination of the two convex criteria, which is still a convex criterion. Minimizing the new criterion is straightforward but it is not clear what is the interpretation of the weights used in the convex combination. For each weight used in the convex combination, we determine the compound optimal design that minimizes the new criterion. The other problem is a constrained optimization problem. If one objective is more important, we want to ensure that the design has at least a certain level of user-specified efficiency and subject to this requirement, does as well as possible for the second objective. It follows that for each user-specified level of efficiency for the more important criterion, we have a constrained optimal design. Cook and Wong (1994) showed that these two classes of optimal design are equivalent in the sense that solving one solves

Table 1. Compound optimal allocation vector $\rho^* = (\rho_1^*, \rho_2^*, \rho_3^*, \rho_4^*)'$ for different values of α in a hypothetical four-arm trial, assuming that the parameter values are known

$b = 0.5$		α										
		0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$\mu_1 = 0$	ρ_1^*	0	0.085	0.130	0.157	0.174	0.186	0.195	0.202	0.207	0.211	0.215
$\mu_2 = -0.25$	ρ_2^*	0	0.097	0.145	0.172	0.189	0.200	0.208	0.214	0.218	0.222	0.225
$\mu_3 = -0.5$	ρ_3^*	0	0.121	0.175	0.202	0.217	0.226	0.231	0.235	0.238	0.240	0.241
$\mu_4 = -1$	ρ_4^*	1	0.696	0.550	0.469	0.420	0.388	0.366	0.349	0.337	0.327	0.319

a corresponding problem in the other. The key is to find which weight α in the convex combination corresponds to the user-specified efficiency sought for the more important criterion. To answer this question, they proposed using an efficiency plot that simultaneously graphs the two efficiencies of the compound optimal design as a function of α . In other words, we plot $E_1(\rho)$ and $E_2(\rho)$ versus α in the same figure as illustrated in Wong (1995) where he considered polynomial models and model misspecification and parameter estimation were the two objectives of interest. Assuming estimating the model parameter θ is the more important objective and the user has specified a minimal value for $E_1(\rho)$, say e_1 , the value of α to use to generate the compound optimal design is the one that corresponds to the point of intersection of the graph of $E_1(\rho)$ and the horizontal line at $E_1(\rho) = e_1$. This resulting compound optimal design is the one that solves the more intuitive constraint optimization problem formulated at the beginning of the problem with user-specified efficiency.

To illustrate this idea, consider a trial with four treatment arms with constant follow-up times of the patients. We first construct locally optimal designs, for which nominal values of the parameters are available. Later on we sequentially estimate these parameters and perform response-adaptive randomization of patients to the treatment groups. Here we choose $\mu_1 = 0$, $\mu_2 = -0.25$, $\mu_3 = -0.5$, $\mu_4 = -1$, and $b = 0.5$. As in Section 3, the follow-up time is chosen to be $\tau = 1/(-\log(0.1))$. We have an observation if the event time is less than or equal to τ ; otherwise the event time is censored if the event time is larger than τ .

Table 1 displays the compound optimal design $\rho^* = (\rho_1^*, \rho_2^*, \rho_3^*, \rho_4^*)'$ for different values of α between 0 and 1. Figure 2 is the efficiency plot obtained by graphing $E_1(\rho)$ and $E_2(\rho)$ versus values of α in the interval $[0, 1]$. The plot provides information on the competitive nature of the two criteria; a steep curve represents that much efficiency has to be given up for a small gain in the efficiency of the other criterion. Given the constrained optimization problem, the plot finds the constrained optimal design indirectly by finding the compound optimal design using the right α determined from the efficiency plot. For example, if both criteria are equally important in our example, the figure shows the correct α to use to generate the compound optimal design is $\alpha = 0.1$. If one criterion is more important than the other, then the value of α must be chosen differently.

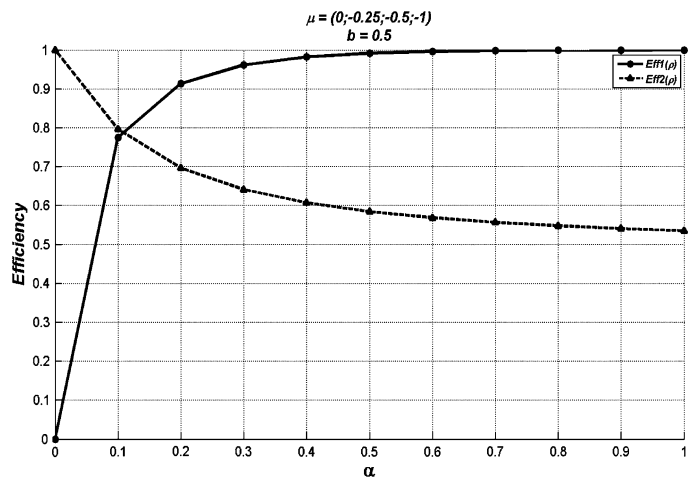


Figure 2. Efficiency plot for the four-arm trial for estimating θ and b , assuming that θ is known.

For instance, if estimating the vector of all model parameters is the main objective, we want a design that guarantees high efficiency for estimating θ , say $e_1 = E_1(\rho) = 90\%$. The plot shows if we draw a horizontal line at $e_1 = 0.90$, this line meets the graph of $E_1(\rho)$ at around $\alpha = 0.2$. This means that the constrained optimal design is found by generating the compound optimal design with $\alpha = 0.2$. The plot also shows that $E_2(\rho) \approx 70\%$ implying that the two criteria are not very competitive. Further applications of efficiency plots to find dual-objective optimal designs for biomedical studies can be found in Huang and Wong (1998a, 2004).

The above designs are all locally optimal and because θ is unknown, they cannot be implemented. One may generate many compound optimal designs for a range of hypothetical θ values and hope that they are not sensitive to the choice of the nominal values. In our example, we changed the nominal values of $(\mu_1, \mu_2, \mu_3, \mu_4)'$ and constructed similar plots (not shown here). The patterns were similar to the one shown in Figure 2 while keeping $\tau = 1/(-\log(0.1))$ and $b = 0.5$. However, when different nominal values result in different optimal designs, alternative methods are required. One such method is the response-adaptive approach that sequentially estimates the unknown parameters using all currently available responses from patients to generate the compound optimal designs.

4.2 Implementing optimal allocations by response-adaptive randomization

In practice the true parameter values are unknown. Response-adaptive randomization designs can be used to learn from accumulating data in the trial and enhance the optimal allocations. In this subsection we present results of a simulation study comparing several response-adaptive randomization designs with a balanced randomization design. In our simulations the true parameter vector θ is assumed to be unknown and it will be sequentially estimated. We consider a four-arm trial with $n = 200$ patients and constant follow-up time $\tau = 1/(-\log(0.1))$ for each patient. Throughout, we assume that $b = 0.5$, and treatment effects vary across the four groups as signified by their mean values μ_1, μ_2, μ_3 and μ_4 . In our simulation, we consider four hypothetical dose response profiles:

- A** Flat dose-response: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$
- B** Monotone dose-response: $\mu_1 = 0, \mu_2 = -0.25, \mu_3 = -0.5, \mu_4 = -1$
- C** “U-shape” dose-response: $\mu_1 = 0, \mu_2 = -0.25, \mu_3 = -0.5, \mu_4 = -0.25$
- D** “Threshold” dose-response: $\mu_1 = 0, \mu_2 = \mu_3 = \mu_4 = -0.5$

The following four allocation strategies will be compared using efficiencies and also a measure based on ethical considerations which will be discussed momentarily.

- I** A compound optimal allocation with $\alpha = 0.1$
- II** A compound optimal allocation with $\alpha = 0.2$
- III** The D -optimal allocation with $\alpha = 1$
- IV** The balanced allocation $\rho = (0.25, 0.25, 0.25, 0.25)'$

Allocation IV is implemented using a completely randomized design (CRD) for which every subject is randomized to treatment groups with equal probabilities. Due to its unpredictability, the CRD mitigates the chance of selection bias in the design. Note that the CRD balances treatment assignments asymptotically, but the treatment numbers may not be perfectly balanced for a finite sample size due to randomization in the design. However, we use the CRD as a reference procedure in order to facilitate a comparison of the variability of allocation proportions with response-adaptive randomization procedures.

Allocations I, II, and III are implemented using the doubly-adaptive biased coin design procedure (DBCD, Hu and Zhang, 2004). For a trial with K treatment groups ($K = 4$ in our case), let (ρ_1, \dots, ρ_K) denote a target allocation vector corresponding to one of the allocations I, II or III. For the DBCD procedure, the $(j + 1)$ th patient is randomized to treatment k with probability

$$(6) \quad \psi_{j+1,k} = \frac{\hat{\rho}_k(j) \left(\frac{\hat{\rho}_k(j)}{N_k(j)/j} \right)^\gamma}{\sum_{i=1}^K \hat{\rho}_i(j) \left(\frac{\hat{\rho}_i(j)}{N_i(j)/j} \right)^\gamma}, \quad k = 1, \dots, K,$$

where $\hat{\rho}(j) = (\hat{\rho}_1(j), \dots, \hat{\rho}_K(j))'$ is the estimated target allocation based on the observed data from the first j patients,

$N_i(j)/j, i = 1, \dots, K$ are current treatment proportions, and $\gamma \geq 0$ is a parameter controlling the degree of randomness of an allocation procedure ($\gamma = 0$ is most random, and $\gamma = \infty$ is almost a deterministic procedure). Rosenberger and Hu (2004) found by simulation that $\gamma = 2$ provides a reasonable balance between randomness and optimality and accordingly, our allocations I, II and III are implemented using the value $\gamma = 2$.

We also evaluate these designs by their ethical value. One common ethical measure increasingly used in the literature is the average value of the total number of “successes” (TNS) in the trial from the design (Rosenberger et al., 2001). In our example we assume that short responses are clinically favorable, and an observation is a “success” if it is less than the threshold $\tau = 1/(-\log(0.1))$. Therefore, TNS in our case is simply the total number of subjects in the trial with observed (uncensored) event times, i.e.

$$\text{TNS} = \sum_{k=1}^4 \sum_{i=1}^{n_k} \mathbf{1}_{\{T_{ik} \leq \tau\}}.$$

A major stumbling block in implementing adaptive designs is the time delay, as one needs to wait for outcomes of currently enrolled patients before allocating the next cohort of patients to the treatment groups. Hu et al. (2008) showed that under widely applicable conditions, large sample results for the DBCD procedure are unaffected by delayed response. Simulations of adaptive survival trials with moderate sample sizes (e.g. Zhang and Rosenberger, 2007) show that when 60 percent or more of the patient’s responses accrue during the recruitment stage of the trial, inferential properties response-adaptive designs is negligibly affected compared to the case when there is no delay. Clearly, response-adaptive designs are inappropriate in trials with fast recruitment and very long follow-up periods.

In our example we assume that recruitment is not fast so that adaptations in the design are feasible. There are several ways to facilitate response-adaptive randomization with delayed outcomes. One approach is to simulate a priority queue data structure, assuming some distribution for patient arrival times and use a continuous monitoring scheme for updating history of responses from patients in the trial and implementing response-adaptive allocation (Rosenberger and Seshaiyer, 1997; Zhang and Rosenberger, 2007). Another approach is a “cohort” response-adaptive randomization (Chappell and Karrison, 2006), for which treatment randomization probabilities are updated after cohorts of subjects respond. We use the latter approach in simulations for this section. For a trial with $n = 200$ patients, first 20 patients are randomized equally among the four treatment arms. The next wave of patients are enrolled in cohorts of 20 such that every cohort of size 20 is enrolled only after all patients in the previous cohort have responded. Thus the model parameters and treatment randomization probabilities are recalculated after every 20 patients, based on the vector of accumulated responses.

Table 2. Theoretical and simulated allocation proportion vector $(\rho_1, \rho_2, \rho_3, \rho_4)$, and maximum likelihood estimates of the model parameters $(\mu_1, \mu_2, \mu_3, \mu_4, b)$ of designs I, II, III, and IV for scenarios A, B, C, and D, based on 1,000 simulations

		Scenario A (flat dose-response)								
Design		ρ_1	ρ_2	ρ_3	ρ_4	μ_1	μ_2	μ_3	μ_4	b
I	Theoretical	0.250	0.250	0.250	0.250	0	0	0	0	0.5
	Average	0.249	0.247	0.245	0.246	0.098	0.109	0.117	0.101	0.495
	S.D.	0.130	0.135	0.134	0.132	0.289	0.294	0.297	0.282	0.081
II	Theoretical	0.250	0.250	0.250	0.250	0	0	0	0	0.5
	Average	0.245	0.248	0.250	0.246	0.062	0.058	0.062	0.070	0.493
	S.D.	0.083	0.088	0.088	0.085	0.260	0.266	0.267	0.273	0.081
III	Theoretical	0.250	0.250	0.250	0.250	0	0	0	0	0.5
	Average	0.247	0.247	0.247	0.248	0.045	0.037	0.037	0.035	0.499
	S.D.	0.034	0.034	0.034	0.033	0.250	0.242	0.253	0.240	0.084
IV	Theoretical	0.250	0.250	0.250	0.250	0	0	0	0	0.5
	Average	0.250	0.249	0.251	0.250	0.016	0.021	0.019	0.030	0.497
	S.D.	0.030	0.030	0.031	0.031	0.241	0.245	0.235	0.252	0.086
		Scenario B (monotone dose-response)								
Design		ρ_1	ρ_2	ρ_3	ρ_4	μ_1	μ_2	μ_3	μ_4	b
I	Theoretical	0.085	0.097	0.121	0.696	0	-0.25	-0.5	-1	0.5
	Average	0.095	0.105	0.127	0.663	0.076	-0.194	-0.443	-1.000	0.497
	S.D.	0.018	0.022	0.031	0.081	0.301	0.249	0.188	0.050	0.041
II	Theoretical	0.130	0.145	0.175	0.550	0	-0.25	-0.5	-1	0.5
	Average	0.135	0.148	0.178	0.530	0.044	-0.215	-0.483	-0.999	0.496
	S.D.	0.020	0.023	0.034	0.067	0.282	0.201	0.150	0.054	0.042
III	Theoretical	0.215	0.225	0.241	0.319	0	-0.25	-0.5	-1	0.5
	Average	0.213	0.222	0.238	0.315	0.038	-0.243	-0.491	-1.001	0.496
	S.D.	0.026	0.028	0.030	0.038	0.237	0.159	0.121	0.069	0.048
IV	Theoretical	0.250	0.250	0.250	0.250	0	-0.25	-0.5	-1	0.5
	Average	0.249	0.251	0.251	0.249	0.018	-0.243	-0.497	-1.002	0.499
	S.D.	0.031	0.029	0.031	0.031	0.196	0.156	0.122	0.082	0.051
		Scenario C ("U-shape" dose-response)								
Design		ρ_1	ρ_2	ρ_3	ρ_4	μ_1	μ_2	μ_3	μ_4	b
I	Theoretical	0.110	0.161	0.567	0.161	0	-0.25	-0.5	-0.25	0.5
	Average	0.124	0.185	0.497	0.185	0.103	-0.158	-0.473	-0.155	0.499
	S.D.	0.042	0.090	0.142	0.096	0.303	0.239	0.129	0.238	0.059
II	Theoretical	0.157	0.210	0.423	0.210	0	-0.25	-0.5	-0.25	0.5
	Average	0.162	0.217	0.398	0.217	0.078	-0.205	-0.484	-0.205	0.499
	S.D.	0.034	0.060	0.087	0.062	0.273	0.192	0.115	0.193	0.062
III	Theoretical	0.226	0.246	0.282	0.246	0	-0.25	-0.5	-0.25	0.5
	Average	0.226	0.243	0.278	0.244	0.032	-0.225	-0.494	-0.242	0.496
	S.D.	0.026	0.030	0.034	0.030	0.240	0.172	0.122	0.163	0.065
IV	Theoretical	0.250	0.250	0.250	0.250	0	-0.25	-0.5	-0.25	0.5
	Average	0.250	0.250	0.250	0.250	0.020	-0.244	-0.489	-0.244	0.494
	S.D.	0.031	0.030	0.030	0.030	0.220	0.174	0.128	0.169	0.064
		Scenario D ("threshold" dose-response)								
Design		ρ_1	ρ_2	ρ_3	ρ_4	μ_1	μ_2	μ_3	μ_4	b
I	Theoretical	0.103	0.299	0.299	0.299	0	-0.5	-0.5	-0.5	0.5
	Average	0.111	0.288	0.295	0.296	0.093	-0.449	-0.455	-0.456	0.501
	S.D.	0.030	0.125	0.128	0.131	0.309	0.156	0.161	0.164	0.055
II	Theoretical	0.147	0.284	0.284	0.284	0	-0.5	-0.5	-0.5	0.5
	Average	0.150	0.280	0.281	0.278	0.049	-0.477	-0.479	-0.470	0.494
	S.D.	0.027	0.075	0.077	0.075	0.275	0.125	0.128	0.134	0.056
III	Theoretical	0.220	0.260	0.260	0.260	0	-0.5	-0.5	-0.5	0.5
	Average	0.219	0.257	0.257	0.258	0.043	-0.488	-0.493	-0.491	0.498
	S.D.	0.025	0.031	0.031	0.030	0.235	0.125	0.125	0.125	0.056
IV	Theoretical	0.250	0.250	0.250	0.250	0	-0.5	-0.5	-0.5	0.5
	Average	0.250	0.250	0.251	0.249	0.027	-0.492	-0.496	-0.498	0.497
	S.D.	0.030	0.031	0.031	0.031	0.219	0.121	0.121	0.125	0.057

Table 3. Theoretical and simulated design efficiencies $E_1(\rho)$, $E_2(\rho)$ and the total number of successes (TNS) of the four randomization procedures (I, II, III, IV) for the four dose response scenarios (A, B, C, D) based on 1,000 simulation runs. The table reports the median values of $E_1(\rho)$ and $E_2(\rho)$, and the mean value of TNS, with its standard deviation in parentheses

Design		Dose-response scenario							
		A		B		C		D	
		Theor.	Sim.	Theor.	Sim.	Theor.	Sim.	Theor.	Sim.
I	$E_1(\rho)$	1.000	0.911	0.775	0.772	0.871	0.857	0.949	0.911
	$E_2(\rho)$	1.000	0.847	0.796	0.796	0.817	0.816	0.938	0.956
	TNS	34	35 (5)	123	123 (6)	67	66 (7)	76	76 (7)
II	$E_1(\rho)$	1.000	0.978	0.913	0.913	0.964	0.960	0.983	0.974
	$E_2(\rho)$	1.000	0.809	0.696	0.697	0.753	0.754	0.912	0.818
	TNS	34	35 (5)	109	109 (6)	62	62 (7)	73	74 (7)
III	$E_1(\rho)$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	$E_2(\rho)$	1.000	0.771	0.535	0.534	0.686	0.686	0.868	0.756
	TNS	34	34 (5)	87	87 (5)	57	58 (6)	70	70 (7)
IV	$E_1(\rho)$	1.000	0.998	0.990	0.990	0.997	0.997	0.998	0.997
	$E_2(\rho)$	1.000	0.759	0.483	0.485	0.669	0.673	0.850	0.735
	TNS	34	34 (5)	80	80 (6)	55	55 (6)	69	69 (6)

Let us first examine the performance of the procedures under a flat dose-response, i.e. when all four treatments are equally effective (scenario A). From Table 2, we see that under scenario A our adaptive designs (I, II, III) and the completely randomized design (CRD, design IV) all converged to the equal allocation. The D -optimal design (III) had similar variability of allocation proportions compared to CRD, whereas two compound optimal designs (I and II) had slightly higher standard deviations of allocation proportions compared to CRD. From Table 3, the D -optimal design yielded higher median values of efficiencies $E_1(\rho)$ and $E_2(\rho)$ compared to CRD; the two compound optimal designs had higher median values of $E_2(\rho)$, but lower values of $E_1(\rho)$ compared to CRD. The key message here is that when we have flat dose-responses from the four groups, our adaptive designs generally have comparable characteristics with CRD; however, due to response-adaptive randomization in the design, our designs under scenario A are somewhat inferior to the “ideal” equal allocation which divides patients equally among 4 arms. In the latter case, both efficiencies $E_1(\rho)$ and $E_2(\rho)$ are 100% under scenario A.

Table 2 also gives a summary of the designs’ characteristics under scenarios B, C, and D, when treatment effects vary across the treatment arms. The conclusions from these scenarios are generally similar and so we discuss the common findings using scenario B as an illustrative case when we have a monotone dose-response. One can see that under scenario B, all four designs converged successfully to their corresponding target allocations. The compound optimal designs with $\alpha = 0.1$ and $\alpha = 0.2$ (designs I and II) were skewed away from the balanced allocation more than the D -optimal design (III). For the adaptive designs, treatment groups with lower proportions of subjects had smaller standard deviations of allocation proportions than treatment

groups with higher proportions of subjects. Since our adaptive designs place more subjects at the doses with shorter event times, the estimation at these doses is more accurate with adaptive designs than with the balanced design. In particular, all four designs estimated the values of $\mu_4 = -1$ and $b = 0.5$ unbiasedly, but the adaptive designs had smaller standard deviations compared to the balanced design. Figures 3 and 4 show simulated distributions of the maximum likelihood estimators ($\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3, \hat{\mu}_4$) and \hat{b} , respectively. In theory, maximum likelihood estimators are strongly consistent and asymptotically normal (Hu and Zhang, 2004). From the plots it is clear that all the designs estimated the true parameters consistently, although the distributions of the estimates in some cases are slightly right-skewed.

Table 3 shows theoretical and simulated (median) values of efficiencies $E_1(\rho)$ and $E_2(\rho)$, and Figure 5 shows distributions of these efficiencies based on 1,000 simulations. From Table 3, design III had higher median values of $E_1(\rho)$ and $E_2(\rho)$ compared to CRD; designs I and II provided trade-offs between the two efficiency criteria and had consistently higher median values of $E_2(\rho)$ than CRD, which implies that these designs were more efficient estimating the underlying hazard pattern (via the parameter b) than the balanced design. It is important to note that for all four designs, simulated median values of $E_1(\rho)$ and $E_2(\rho)$ were close to the corresponding theoretical values from the local optimal designs. In our additional simulations (not reported here) we increased the sample size from $n = 200$ to $n = 400$ patients, and we observed that with the increased sample size, median values of $E_1(\rho)$ and $E_2(\rho)$ were getting even closer to the theoretical values. Further, with $n = 400$ patients, the standard deviations of both allocation proportions and the maximum likelihood estimated were smaller than in the case with $n = 200$ patients.

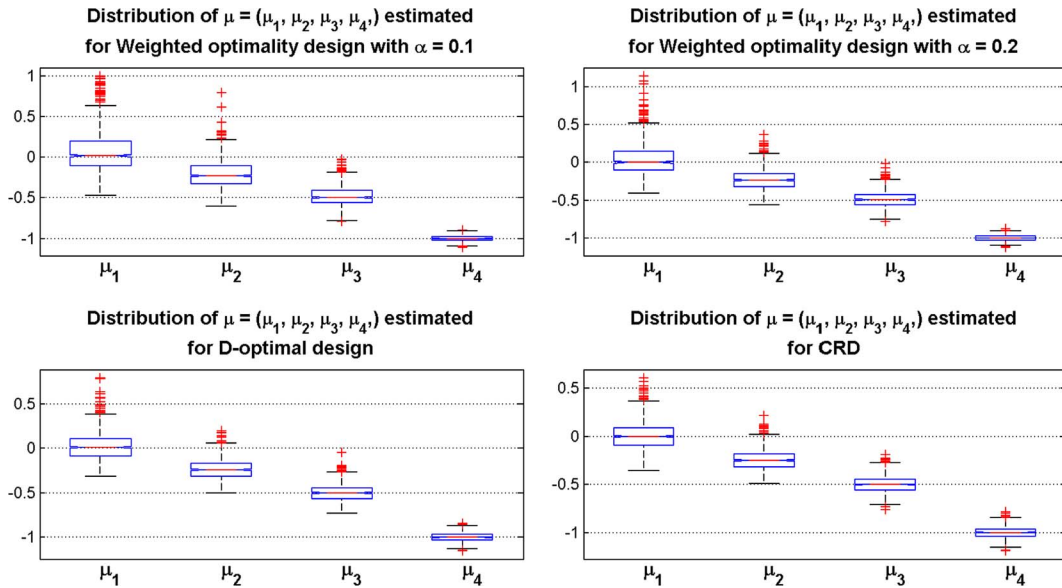


Figure 3. Distributions of the maximum likelihood estimators $\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3, \hat{\mu}_4$ for the four randomization procedures based on 200 patients and 1,000 simulations with $\mu_1 = 0, \mu_2 = -0.25, \mu_3 = -0.5, \mu_4 = -1$ and $b = 0.5$.

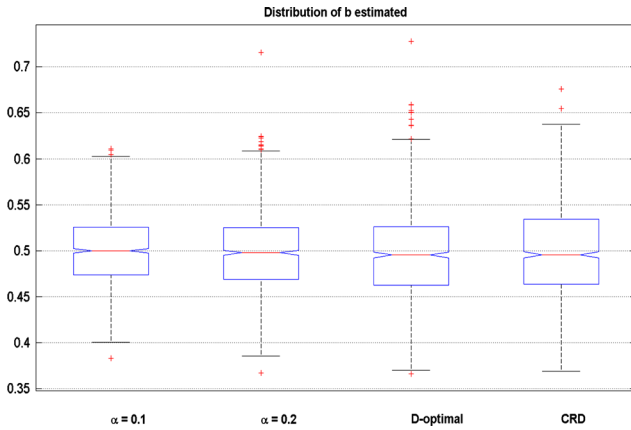


Figure 4. Distributions of the maximum likelihood estimator \hat{b} for the four randomization procedures based on 200 patients and 1,000 simulations with $\mu_1 = 0, \mu_2 = -0.25, \mu_3 = -0.5, \mu_4 = -1$ and $b = 0.5$.

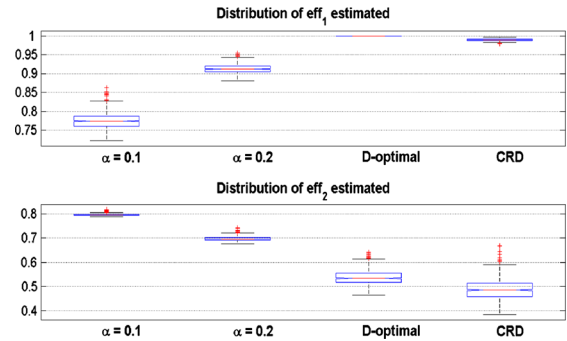


Figure 5. Distributions of the two efficiencies $E_1(\rho)$ and $E_2(\rho)$ for the four randomization procedures based on 200 patients and 1,000 simulations with $\mu_1 = 0, \mu_2 = -0.25, \mu_3 = -0.5, \mu_4 = -1$ and $b = 0.5$.

Finally, we note that under each of the scenarios B, C, and D, the balanced design (IV) was least ethical among the four designs (Table 3). For instance, under scenario B, on average, the D -optimal design (III) resulted in 7 more successes; the design with $\alpha = 0.2$ (design II) had 29 more successes; and the design with $\alpha = 0.1$ (design I) had 43 more successes than the balanced design. All four designs had similar standard deviations of TNS.

The overall conclusion from the simulation study is that the proposed adaptive designs are fully randomized and do a good job targeting optimal allocations. At the same time, our adaptive designs achieve the prescribed levels of efficiency and tend to allocate more patients to the more effi-

cient doses. Our examples here also show that adaptive designs, when properly implemented, can outperform balanced randomization designs in terms of efficiency and ethical criteria.

5. DISCUSSION

In this paper, several design issues have been addressed for multi-arm time-to-event clinical trials with censored Weibull outcomes. We explored several optimality criteria for estimating treatment contrasts, including D_A -optimality, trace-optimality, and optimality for estimation of treatment hazard ratios. If there is no censoring in the model, the optimal allocations are easily found. Unfortunately, matters are not so mathematically neat when censoring is present. In the $K = 2$ treatment case, we found optimal allocations numer-

ically by minimizing objective functions which depend on model parameters. Numerical studies show that optimal allocations are very close to 0.5, the balanced allocation. These results are consistent with some previous findings in the literature on efficiency of balanced allocation for an exponential model and Cox's proportional hazards model (Kalish and Harrington, 1988).

For trials with $K \geq 2$ treatment arms we discussed how to select weights in a compound optimality criterion to generate allocations that achieve prescribed levels of two selected efficiency criteria. The resulting designs are locally optimal. We used response-adaptive randomization to sequentially target selected optimal allocations in a hypothetical clinical trial with four treatment arms. Simulations show that the doubly-adaptive biased coin design procedures (DBCD; Hu and Zhang, 2004) converge well to the selected optimal targets and thereby achieve the desired levels of statistical efficiency. The described methodology may be useful in planning phase II dose ranging studies with primary time-to-event outcomes where smaller event times signify higher treatment efficacy (e.g. recovery).

The design approach discussed here can be extended to problems for more than two objectives and the objectives have unequal interest to the researcher. After formatting all objectives as convex functional of the information matrix and prioritizing the objectives in terms of their importance, the researcher specifies the efficiencies required for each criterion. The most important one should have the highest efficiency requirement, followed by the next most important criterion and so on. Of course the sought optimal design may not exist if all the efficiencies are unreasonably high and the objectives are competitive. The researcher does not have to specify the efficiency required for the least important criterion because it is only determined after all the more important objectives have met their efficiencies requirements. Computationally, no technical problems arise because the compound optimal design for the multiple-objective problem is still found straightforwardly by minimizing a convex combination of convex criteria. The only issue is the interpretation of the weights used in the convex combination, i.e. which convex combination corresponds to the set of user-specified efficiency requirements? We can resort to a high dimensional efficiency plot and used the same technique as for a dual-objective optimal design problem. However, appreciating high dimensional graphics can be both tricky and difficult and so there are still implementing and interpreting issues for finding multiple objective optimal design problems. Some work in this direction are Huang and Wong (1998b), and Zhu and Wong (2001).

We close with a note that we dealt with the problem of finding optimal allocation proportions for a multi-arm trial. An obvious analog of the design problem just discussed is to have the K treatment groups be different dose levels of the same drug. In this case, it is sensible to select the best doses on a continuous scale, rather than finding the best proportions. This is a promising research topic for the future work.

ACKNOWLEDGEMENTS

Weng Kee Wong worked on this manuscript when he was a visiting fellow and a member of the scientific advisory board for a six-month workshop on the design and analysis of experiments at The Sir Isaac Newton Institute at Cambridge, England. He would like to thank the Institute for the support during his repeated visits in the second half of 2011. The authors would like to thank the associate editor and an anonymous referee for the constructive comments, which led to an improved version of the paper.

Received 11 October 2011

REFERENCES

- BANDYOPADHYAY, U., BISWAS, A., and BHATTACHARYA, R. (2009). A Bayesian adaptive design for two-stage clinical trials with survival data. *Lifetime Data Analysis* **15** 468–492. [MR2563199](#)
- BANDYOPADHYAY, U., BISWAS, A., and BHATTACHARYA, R. (2010). A covariate-adjusted adaptive design for two-stage clinical trials with survival data. *Statistica Neerlandica* **64** 202–226. [MR2756486](#)
- BANDYOPADHYAY, U. and BHATTACHARYA, R. (2010). Response adaptive procedures with dual optimality. *Statistica Neerlandica* **63** 353–367. [MR2751897](#)
- BISWAS, A. and BHATTACHARYA, R. (2009). Optimal response-adaptive designs for normal responses. *Biometrical Journal* **51** 193–202. [MR2667522](#)
- BISWAS, A. and BHATTACHARYA, R. (2010). An optimal response-adaptive design with dual constraints. *Statistics and Probability Letters* **80** 177–185. [MR2575443](#)
- BISWAS, A. and MANDAL, S. (2004). Optimal adaptive designs in phase III clinical trials for continuous responses with covariates. In *mODA 7-Advances in Model-Oriented Design and Analysis*. A. Di Bucchianico, H. Läuter, and H. P. Wynn (eds), 51–58. Physica-Verlag, Heidelberg. [MR2089325](#)
- BISWAS, A., MANDAL, S., and BHATTACHARYA, R. (2011). Multi-treatment optimal response-adaptive designs for phase III clinical trials. *Journal of the Korean Statistical Society* **40** 33–44. [MR2760006](#)
- CHAPPELL, R. and KARRISON, T. (2006). Letter to the editor. *Statistics in Medicine* **26** 3050–3052. [MR2370985](#)
- CHEUNG, Y. K., LURDES, Y. T., WATHEN, J. K., and THALL, P. F. (2006). Continuous Bayesian adaptive randomization based on event times with covariates. *Statistics in Medicine* **25** 55–70. [MR2222074](#)
- COOK, R. D. and WONG, W. K. (1994). On the equivalence of constrained and compound optimal designs. *Journal of the American Statistical Association* **89** 687–692. [MR1294092](#)
- FLEISS, J. L. (1986). *The Design and Analysis of Clinical Experiments*. Wiley-Interscience.
- GWISE, T., HU, J., and HU, F. (2008). Optimal biased coins for two-arm clinical trials. *Statistics and its Interface* **1** 125–135. [MR2425350](#)
- GWISE, T., ZHOU, J., and HU, F. (2011). An optimal response adaptive biased coin design with k heteroscedastic treatments. *Journal of Statistical Planning and Inference* **141** 235–242. [MR2719490](#)
- HU, F. and ROSENBERGER, W. F. (2003). Optimality, variability, power: Evaluating response-adaptive randomization procedures for treatment comparisons. *Journal of the American Statistical Association* **98** 671–678. [MR2011680](#)
- HU, F. and ROSENBERGER, W. F. (2006). *The Theory of Response-Adaptive Randomization in Clinical Trials*. Wiley, New York. [MR2245329](#)
- HU, F. and ZHANG, L.-X. (2004). Asymptotic properties of doubly-adaptive biased coin designs for multitreatment clinical trials. *The Annals of Statistics* **32** 268–301. [MR2051008](#)

- HU, F., ZHANG, L.-X., CHEUNG, S. H., and CHAN, W. S. (2008). Doubly-adaptive biased coin designs with delayed responses. *Canadian Journal of Statistics* **36** 541–559. [MR2532251](#)
- HU, F., ZHANG, L.-X., and HE, X. (2009). Efficient randomized-adaptive designs. *The Annals of Statistics* **37** 2543–2560. [MR2543702](#)
- HUANG, Y. C. and WONG, W. K. (1998a). Multiple-objective optimal designs. *Journal of Biopharmaceutical Statistics* **8** 635–643.
- HUANG, Y. C. and WONG, W. K. (1998b). Sequential construction of multiple-objective designs. *Biometrics* **54** 1388–1397.
- HUANG, Y. C. and WONG, W. K. (2004). Robustness properties of multiple-objective optimal designs. *Drug Information Journal* **39** 223–232.
- JENNISON, C. and TURNBULL, B. W. (2000). *Group Sequential Methods with Applications to Clinical Trials*. Chapman and Hall/CRC, Boca Raton. [MR1710781](#)
- JEON, Y. and HU, F. (2010). Optimal adaptive designs for binary response trials with three treatments. *Statistics in Biopharmaceutical Research* **2** 310–318.
- KALISH, L. and HARRINGTON, D. P. (1988). Efficiency of balanced treatment allocation for survival analysis. *Biometrics* **44** 815–821.
- PONG, A. and CHOW, S.-C., eds. (2011). *Handbook of Adaptive Designs in Pharmaceutical and Clinical Development*. CRC Press, Boca Raton. [MR2789104](#)
- ROSENBERGER, W. F. and HU, F. (2004). Maximizing power and minimizing treatment failures in clinical trials. *Clinical Trials* **1** 141–147.
- 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., and RICKS, M. (2001). Optimal adaptive designs for binary response trials. *Biometrics* **57** 909–913. [MR1863454](#)
- ROSENBERGER, W. F., SVERDLOV, O., and HU, F. (2012). Adaptive randomization for clinical trials. *Journal of Biopharmaceutical Statistics* **22**(4) 719–736.
- SVERDLOV, O., RYEZNIK, Y., and WONG, W. K. (2012). Optimal response-adaptive designs for multiple-objective clinical trials with Weibull time-to-event outcomes, submitted.
- TYMOFYEVYEV, Y., ROSENBERGER, W. F., and HU, F. (2007). Implementing optimal allocation in sequential binary response experiments. *Journal of the American Statistical Association* **102** 224–234. [MR2345540](#)
- WONG, W. K. and ZHU, W. (2008). Optimum treatment allocation rules under a variance heterogeneity model. *Statistics in Medicine* **27** 4581–4595. [MR2528532](#)
- ZHANG, L. and ROSENBERGER, W. F. (2006). Response-adaptive randomization for clinical trials with continuous outcomes. *Biometrics* **62** 562–569. [MR2236838](#)
- ZHANG, L. and ROSENBERGER, W. F. (2007). Response-adaptive randomization for survival trials: the parametric approach. *Applied Statistics* **56** 153–165. [MR2359239](#)
- ZHU, H. and HU, F. (2009). Implementing optimal allocation for sequential continuous responses with multiple treatments. *Journal of Statistical Planning and Inference* **139** 2420–2430. [MR2508003](#)
- ZHU, W. and WONG, W. K. (2001). Bayesian optimal designs for estimating a set of symmetric quantiles. *Statistics in Medicine* **20** 123–137.

Oleksandr Sverdlov
 Novartis Pharmaceuticals Corporation
 One Health Plaza
 East Hanover, NJ 07936-1080
 USA
 E-mail address: alex.sverdlov@novartis.com

Yevgen Ryznyk
 Department of Mathematics
 Kharkov National University of Economics
 9A Lenin Ave., Kharkov 610001
 Ukraine
 E-mail address: yevgen.ryznyk@gmail.com

Weng Kee Wong
 Department of Biostatistics
 School of Public Health
 University of California, Los Angeles
 Los Angeles, CA 90095
 USA
 E-mail address: wkwong@ucla.edu