Journal of Statistical Software

August 2015, Volume 66, Issue 1. *http://www.jstatsoft.org/*

RARtool: A MATLAB Software Package for Designing Response-Adaptive Randomized Clinical Trials with Time-to-Event Outcomes

Yevgen Ryeznik
Uppsala University

Oleksandr Sverdlov
EMD Serono, Inc.

Weng Kee Wong
University of California, Los Angeles

Abstract

Response-adaptive randomization designs are becoming increasingly popular in clinical trial practice. In this paper, we present **RARtool**, a user interface software developed in MATLAB for designing response-adaptive randomized comparative clinical trials with censored time-to-event outcomes. The **RARtool** software can compute different types of optimal treatment allocation designs, and it can simulate response-adaptive randomization procedures targeting selected optimal allocations. Through simulations, an investigator can assess design characteristics under a variety of experimental scenarios and select the best procedure for practical implementation. We illustrate the utility of our **RARtool** software by redesigning a survival trial from the literature.

Keywords: adaptive design, censored data, delayed responses, MATLAB, multi-objective clinical trial, optimal allocation, response-adaptive randomization, survival trial.

1. Introduction

Response-adaptive randomization (RAR) in clinical trials refers to sequential modification of treatment randomization probabilities based on the history of treatment assignments and responses from patients in the trial, with the goal of assigning more patients to the better treatment while maintaining important statistical properties of the trial design (Hu and Rosenberger 2006). Modern research on RAR has focused on the development of optimal RAR procedures for multi-objective clinical trials with two or more treatment arms using a template of Hu and Rosenberger (2003). The idea is to determine an optimal allocation for the given experimental objectives (Sverdlov and Rosenberger 2013) and find a fully randomized RAR procedure with minimal variability that converges to the chosen optimal allocation (Hu,

Rosenberger, and Zhang 2006). Although optimal RAR designs frequently have more desirable statistical properties than traditional fixed randomization designs, implementation of the former designs in clinical trials has been limited. One of the reasons is lack of user-friendly statistical software which would allow an experimenter to visualize design characteristics under a variety of hypothetical experimental scenarios and select an optimal procedure for use in practice.

In this paper, we present **RARtool**, a user-friendly software package, developed in MATLAB (The MathWorks, Inc. 2011), to facilitate the design of randomized comparative clinical trials with time-to-event outcomes, by implementing statistical methodology from several recent papers (Zhang and Rosenberger 2007; Sverdlov, Tymofyeyev, and Wong 2011; Sverdlov, Ryeznik, and Wong 2012, 2014). The highlights of **RARtool** are as follows. (1) It can compute optimal allocation designs and values of different statistical efficiency criteria for user-selected sets of experimental parameters. Such optimal allocations provide "benchmarks" for comparison of various allocation designs. (2) It can perform Monte Carlo simulations of RAR procedures targeting selected optimal allocations. Through simulations, an investigator can assess the performance of RAR procedures under a variety of standard to worst-case scenarios and select the best procedure for practical implementation. Therefore, the **RARtool** package is intended to fill the gap between methodology and implementation of optimal RAR designs in time-to-event trials.

The outline of the paper is as follows. Section 2 gives statistical background material. In Section 3, we describe the structure of the **RARtool** package, and in Section 4, we illustrate its utility by redesigning a survival trial from the literature. In Section 5, we give a summary and discuss possible extensions.

2. Statistical background

Hu and Rosenberger (2003) proposed a mathematical template for developing optimal RAR procedures. Their template consists of three major steps:

1. Deriving an optimal allocation to satisfy selected experimental objectives. The objectives may include most accurate estimation of treatment contrasts, maximizing power of a statistical test, or minimizing total hazard in the study subject to appropriate constraints.

2. Constructing a RAR procedure with minimal variability and high speed of convergence to the chosen optimal allocation.

3. Analyzing clinical trial data following the chosen RAR procedure.

Our software development process follows this template for clinical trials with censored time-to-event outcomes.

2.1. Optimal allocation

Consider a clinical trial with $K \geq 2$ treatment arms and time-to-event primary outcomes. We assume a parallel group design with n subjects (n is fixed and pre-determined by budgetary and logistical considerations) for which n_k subjects are to be assigned to treatment k, where

$k = 1, \ldots, K$ and $\sum_{k=1}^{K} n_k = n$. Many survival trials in the literature use designs with $K = 2$ or $K = 3$ treatment arms.

Throughout the paper we assume that event times follow a parametric distribution with probability density function $f(t|\boldsymbol{\theta})$ and survivor function $S(t|\boldsymbol{\theta}) = \int_t^\infty f(s|\boldsymbol{\theta})ds$, where $\boldsymbol{\theta}$ is a vector of unknown model parameters. In time-to-event trials observations are likely to be right-censored by some fixed or random censoring time (Lawless 2003). For the ith patient in group k, let T_{ik} denote the event time, C_{ik} denote the censoring time, $t_{ik} = \min(T_{ik}, C_{ik})$ denote the actual observed time, and $\delta_{ik} = \mathbf{1}_{\{t_{ik}=T_{ik}\}}$ denote the event indicator ($\delta_{ik} = 1$, if $T_{ik} \leq C_{ik}$ and $\delta_{ik} = 0$ otherwise). The individual observations (t_{ik}, δ_{ik}) are assumed to be independent for $i = 1, \ldots, n_k$ and $k = 1, \ldots, K$. The likelihood function is given by

$$L(\text{Data}|\boldsymbol{\theta}) = \prod_{k=1}^{K} \prod_{i=1}^{n_k} \{f(t_{ik}|\boldsymbol{\theta})\}^{\delta_{ik}} \{S(t_{ik}|\boldsymbol{\theta})\}^{1-\delta_{ik}}.$$

The maximum likelihood estimator can be determined by solving the system of score equations $\frac{\partial}{\partial \boldsymbol{\theta}} \log L(\text{Data}|\boldsymbol{\theta}) = \mathbf{0}$, and it is important to note that $\mathsf{E}\{\frac{\partial}{\partial \boldsymbol{\theta}} \log L(\text{Data}|\boldsymbol{\theta})\} = \mathbf{0}$. The key object in optimal design theory is the Fisher information matrix given by $\boldsymbol{M}(\boldsymbol{\theta}) = -\mathsf{E}\left\{\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top} \log L(\text{Data}|\boldsymbol{\theta})\right\}$, whose inverse provides the lower bound on the variance of an unbiased estimator of $\boldsymbol{\theta}$. By minimizing the inverse of the Fisher information matrix by choice of design one can achieve the most accurate inference for the parameters of interest.

In this paper we consider two different parametric model for event times and different optimization problems.

Exponential model

The exponential distribution is frequently used in survival analysis because it represents a natural starting point for development of optimal allocation designs for time-to-event trials (Lawless 2003).

For the ith patient in group k we assume that the patient's event time T_{ik} follows an exponential distribution with mean θ_k, $k = 1, \ldots, K$. Then $f(t_{ik}|\boldsymbol{\theta}) = \theta_k^{-1} \exp(-t_{ik}/\theta_k)$, $S(t_{ik}|\boldsymbol{\theta}) = \exp(-t_{ik}/\theta_k)$, and $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)^\top$. The likelihood function is

$$L(\text{Data}|\boldsymbol{\theta}) = \prod_{k=1}^{K} \theta_k^{-\Delta_k} \exp(-T_k/\theta_k), \tag{1}$$

where $\Delta_k = \sum_{i=1}^{n_k} \delta_{ik}$ is the number of events in group k and $T_k = \sum_{i=1}^{n_k} t_{ik}$ is the total observed time for group k. The maximum likelihood estimator of θ_k is $\hat{\theta}_k = T_k/\Delta_k$. Define $\epsilon_k = \mathsf{P}(T_{ik} \leq C_{ik})$, the probability of observing the event before censoring in group k. The expression for ϵ_k depends on the censoring mechanism used in the trial; in general ϵ_k is a function of $\boldsymbol{\theta}$. Let $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_K)^\top$ denote the design that allocates the proportion ρ_k of the total subjects to treatment group k, with the conditions $0 \leq \rho_k \leq 1$ and $\sum_{k=1}^{K} \rho_k = 1$. For a trial with n patients, this means roughly $n_k = n\rho_k$ are assigned to treatment k subject to $\sum_{k=1}^{K} n_k = n$. Then the Fisher information matrix for $\boldsymbol{\theta}$ using design $\boldsymbol{\rho}$ is the diagonal matrix

$$\boldsymbol{M}(\boldsymbol{\rho}, \boldsymbol{\theta}) = n \cdot \text{diag}\left\{\frac{\rho_1 \epsilon_1}{\theta_1^2}, \ldots, \frac{\rho_K \epsilon_K}{\theta_K^2}\right\}. \tag{2}$$

An experimental design problem is to find an optimal allocation vector $\boldsymbol{\rho}^* = (\rho_1^*, \ldots, \rho_K^*)^\top$ as a solution to some formal optimization problem involving the inverse of (2). We will consider four different optimal allocation rules that address different study objectives.

Suppose the primary objective of the study concerns efficient estimation of the contrasts of $(K-1)$ experimental treatments $2, \ldots, K$ versus the control treatment 1. Let $\boldsymbol{A}^\top\boldsymbol{\theta} = (\theta_2 - \theta_1, \ldots, \theta_K - \theta_1)^\top$, where $\boldsymbol{A}^\top$ is the appropriate $(K-1) \times K$ matrix of contrasts. Let $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_K)^\top$. The variance-covariance matrix of $\boldsymbol{A}^\top\hat{\boldsymbol{\theta}}$ is $\mathsf{VAR}(\boldsymbol{A}^\top\hat{\boldsymbol{\theta}}) = \boldsymbol{A}^\top\boldsymbol{M}^{-1}(\boldsymbol{\rho}, \boldsymbol{\theta})\boldsymbol{A}$, and we want to minimize it, in some sense, to achieve most accurate inference for $\boldsymbol{A}^\top\boldsymbol{\theta}$.

(Exp-D$_A$) The D$_A$-optimal allocation minimizes the criterion $\log(\det\{\boldsymbol{A}^\top\boldsymbol{M}^{-1}(\boldsymbol{\rho}, \boldsymbol{\theta})\boldsymbol{A}\})$ (Atkinson 1982). Such an allocation results in the smallest volume of the confidence ellipsoid for $\boldsymbol{A}^\top\boldsymbol{\theta}$. The analytical description of the D$_A$-optimal allocation is given in Sverdlov *et al.* (2011, p. 2893).

(Exp-A$_A$) The A$_A$-optimal allocation minimizes the criterion $\mathrm{trace}\{\boldsymbol{A}^\top\boldsymbol{M}^{-1}(\boldsymbol{\rho}, \boldsymbol{\theta})\boldsymbol{A}\}$, and the A$_A$-optimal treatment allocation proportions are found as

$$\rho_1^* = \frac{\theta_1\sqrt{(K-1)/\epsilon_1}}{\theta_1\sqrt{(K-1)/\epsilon_1} + \sum_{i=2}^K \theta_i/\sqrt{\epsilon_i}},$$
$$\rho_k^* = \frac{\theta_k/\sqrt{\epsilon_k}}{\theta_1\sqrt{(K-1)/\epsilon_1} + \sum_{i=2}^K \theta_i/\sqrt{\epsilon_i}}, \quad k = 2, \ldots, K.$$

When $K = 2$, the D$_A$-optimal and A$_A$-optimal allocation designs coincide and are referred to as *Neyman* allocation:

$$\rho_1^* = \frac{\theta_1/\sqrt{\epsilon_1}}{\theta_1/\sqrt{\epsilon_1} + \theta_2/\sqrt{\epsilon_2}} \quad \text{and} \quad \rho_2^* = 1 - \rho_1^*. \tag{3}$$

Suppose the primary objective of the study concerns testing the hypothesis of homogeneity among treatment effects H_0: $\boldsymbol{\theta}_c = \boldsymbol{A}^\top\boldsymbol{\theta} = \boldsymbol{0}$ versus H_A: $\boldsymbol{\theta}_c \neq \boldsymbol{0}$. Let $\hat{\boldsymbol{\theta}}_c = \boldsymbol{A}^\top\hat{\boldsymbol{\theta}}$ and consider the Wald test statistic $W_n = \hat{\boldsymbol{\theta}}_c\hat{\boldsymbol{\Sigma}}_n^{-1}\hat{\boldsymbol{\theta}}_c$, where $\hat{\boldsymbol{\Sigma}}_n$ is a consistent estimator of $\boldsymbol{\Sigma}_n = \mathsf{VAR}(\hat{\boldsymbol{\theta}}_c) = \boldsymbol{A}^\top\boldsymbol{M}^{-1}(\boldsymbol{\rho}, \boldsymbol{\theta})\boldsymbol{A}$. Let $B \in [0, 1/K]$ be the minimum desired proportion of patients for each treatment group and $\boldsymbol{w} = (w_1, \ldots, w_K)^\top$ be a vector of user-defined positive weights which may be functions of $\boldsymbol{\theta}$. Consider the following optimization problem:

$$\begin{cases} \text{minimize} & \sum_{i=1}^K w_i n_i \\ \text{subject to} & n_i / \sum_{j=1}^K n_j \geq B, \quad i = 1, \ldots, K, \\ \text{and} & \boldsymbol{\theta}_c\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\theta}_c \geq \eta, \end{cases} \tag{4}$$

where $\eta > 0$ is some constant (the optimal solution will not depend on η). This is a well-defined nonlinear programming optimization problem with a unique solution (Tymofyeyev, Rosenberger, and Hu 2007). We are interested in two choices of the vector of weights $\boldsymbol{w}$:

(Exp-NP1) Nonlinear Programming 1 (NP-1) allocation solving (4) with $\boldsymbol{w} = (1, \ldots, 1)^\top$. Such an allocation maximizes power of the Wald test (for a given sample size n) under the restriction that the proportion for each treatment group is at least B. The closed-form solution is available in Sverdlov *et al.* (2011, Theorem 2, p. 2895).

(Exp-NP2) Nonlinear Programming 2 (NP-2) allocation solving (4) with $\boldsymbol{w} = (\theta_1^{-1}, \ldots, \theta_K^{-1})^\top$. Such an allocation minimizes total expected hazard in the study subject to the minimum constraints on the treatment proportions and power. The analytical solution is unknown, but a numerical solution can be found using standard optimization software.

Notice that when $K = 2$, allocation *(Exp-NP1)* reduces to Neyman allocation (3), and allocation *(Exp-NP2)* is as follows (Zhang and Rosenberger 2007):

$$\rho_1^{**} = \frac{\sqrt{\theta_1^3/\epsilon_1}}{\sqrt{\theta_1^3/\epsilon_1} + \sqrt{\theta_2^3/\epsilon_2}} \quad \text{and} \quad \rho_2^{**} = 1 - \rho_1^{**}. \tag{5}$$

Weibull model

The Weibull distribution is a flexible and useful model in parametric survival analysis which allows for different hazard shapes (Carroll 2003; Cheung, Lurdes, Wathen, and Thall 2006). We shall consider several optimal allocation designs for Weibull time-to-event outcomes.

As in Sverdlov *et al.* (2014), let $T_{ik} > 0$ denote the event time for the ith patient in group k. For a log-transformed T_{ik} we assume the following linear model:

$$\log T_{ik} = \mu_k + bW_{ik}, \tag{6}$$

where W_{ik} are independent, identically distributed errors with probability density $f(w) = e^w \exp(-e^w)$. In (6), μ_k represents the effect of treatment k, b is the scale parameter assumed to be common to the K groups and $\boldsymbol{\theta} = (\mu_1, \ldots, \mu_K, b)^\top$ is the vector of unknown parameters. Note that when $b = 1$, we have an exponential model; otherwise we have a Weibull model which allows for different hazard patterns. Let $z_{ik} = (\log t_{ik} - \mu_k)/b$ be standardized log-transformed observed times. The likelihood function for $\boldsymbol{\theta}$ is

$$L(\text{Data}|\boldsymbol{\theta}) = \prod_{k=1}^{K} \prod_{i=1}^{n_k} \{b^{-1} e^{z_{ik}} \exp(-e^{z_{ik}})\}^{\delta_{ik}} \{\exp(-e^{z_{ik}})\}^{1-\delta_{ik}}. \tag{7}$$

The maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ is found numerically from the $(K+1)$-system of score equations $\frac{\partial}{\partial \boldsymbol{\theta}} \log L(\text{Data}|\boldsymbol{\theta}) = \mathbf{0}$. The Fisher information matrix for $\boldsymbol{\theta}$ is

$$\boldsymbol{M}(\boldsymbol{\rho}, \boldsymbol{\theta}) = \frac{n}{b^2} \begin{pmatrix} \text{diag}\{\rho_1\epsilon_1, \ldots, \rho_K\epsilon_K\} & \boldsymbol{x} \\ \boldsymbol{x}^\top & \sum_{k=1}^{K} \rho_k(\epsilon_k + c_k) \end{pmatrix}, \tag{8}$$

where $\boldsymbol{x} = (\rho_1 a_1, \ldots, \rho_K a_K)^\top$, $\epsilon_k = \mathsf{P}(\delta_{ik} = 1)$, $a_k = \mathsf{E}(z_{ik} e^{z_{ik}})$, $c_k = \mathsf{E}\left(z_{ik}^2 e^{z_{ik}}\right)$, and each of the ϵ_k, a_k and c_k is a function of $\boldsymbol{\theta}$ and the censoring mechanism used in the trial. Our goal is to find a vector of treatment allocation proportions which minimizes some convex criterion of the inverse of (8).

(Weib-CO) *Compound optimal allocation* provides trade-off between D-optimality and efficiency in estimating the underlying hazard pattern (via the parameter b) (Sverdlov *et al.* 2014). Consider the following dual-objective optimization problem:

$$\begin{cases} \min\limits_{\rho_1, \ldots, \rho_K} & \alpha\Phi_1(\boldsymbol{\rho}) + (1-\alpha)\Phi_2(\boldsymbol{\rho}) \\ \text{subject to} & \sum_{k=1}^{K} \rho_k = 1, \end{cases} \tag{9}$$

where $\Phi_1(\boldsymbol{\rho}) = -\log\left|\boldsymbol{M}^{-1}(\boldsymbol{\rho},\boldsymbol{\theta})\right|$ is the D-optimality criterion,

$$\Phi_2(\boldsymbol{\rho}) = -\log(\sum_{k=1}^{K}\rho_k d_k)$$

($d_k = \epsilon_k + c_k - a_k^2/\epsilon_k,\ k = 1,\ldots,K$) is a criterion whose optimization leads to the most accurate inference for the parameter b, and $0 \le \alpha \le 1$ is a user-defined constant determining the trade-off between $\Phi_1(\boldsymbol{\rho})$ and $\Phi_2(\boldsymbol{\rho})$. The values of $\alpha = 0$ and $\alpha = 1$ correspond to minimization of $\Phi_2(\boldsymbol{\rho})$ and $\Phi_1(\boldsymbol{\rho})$, respectively.

If $\alpha = 0$, then the optimal allocation places all patients to the treatment group with maximum value of d_k (or assigns equal proportions if there are several such treatments). For $0 < \alpha \le 1$, the optimization problem (9) has the unique solution which can be found from the system of K nonlinear equations

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

(Weib-WDO) *Weighted distance optimal allocation* provides trade-off between D-optimality and some ethical criterion (Sverdlov *et al.* 2014). Let $\boldsymbol{\rho}_D = (\rho_{D1},\ldots,\rho_{DK})^\top$ denote the D-optimal allocation vector and $\boldsymbol{\rho}_E = (\rho_{E1},\ldots,\rho_{EK})^\top$ denote an allocation vector that is desirable from an ethical point of view. For instance, if shorter event times signify clinical efficacy (e.g., recovery), then an investigator may want to choose the components of $\boldsymbol{\rho}_E$ as follows:

$$\rho_{Ek} = \frac{\{\exp(-\mu_k/b)\}^\nu}{\sum_{j=1}^{K}\{\exp(-\mu_j/b)\}^\nu}, \quad k = 1,\ldots,K, \tag{10}$$

where $\nu \ge 0$ is a user-specified parameter controlling the degree of skewness to the better treatment. With allocation (10), the condition $\mu_i \le \mu_j$ implies $\rho_{Ei} \ge \rho_{Ej}$, with equality if and only if $\mu_i = \mu_j$. A similar idea can be applied for trials where longer event times are clinically desirable (e.g., survival trials).

To provide trade-off between $\boldsymbol{\rho}_D$ and $\boldsymbol{\rho}_E$, one can consider minimizing a weighted distance $\alpha\lambda(\boldsymbol{\rho},\boldsymbol{\rho}_D) + (1-\alpha)\lambda(\boldsymbol{\rho},\boldsymbol{\rho}_E)$, where $\lambda(\cdot,\cdot)$ is some distance measure between two vectors of allocation proportions and $0 \le \alpha \le 1$ is a user-selected constant which determines the trade-off between the inferential and ethical criteria (if $\alpha = 1$, the optimal solution is $\boldsymbol{\rho}_D$; if $\alpha = 0$, the optimal solution is $\boldsymbol{\rho}_E$; and if $0 < \alpha < 1$, then we have an allocation that provides a trade-off between efficiency and ethics).

The choice of $\lambda_{KL}(\boldsymbol{\rho},\tilde{\boldsymbol{\rho}}) = \sum_{k=1}^{K}\rho_k\log(\rho_k/\tilde{\rho}_k)$ (Kullback-Leibler directed divergence) yields the optimal solution as a weighted geometric mean of the components of $\boldsymbol{\rho}_D$ and $\boldsymbol{\rho}_E$:

$$\boldsymbol{\rho}_{\alpha k} = \frac{(\rho_{Dk})^\alpha(\rho_{Ek})^{1-\alpha}}{\sum_{j=1}^{K}(\rho_{Dj})^\alpha(\rho_{Ej})^{1-\alpha}}, \quad k = 1,\ldots,K. \tag{11}$$

If we select a quadratic distance metric $\lambda_2(\boldsymbol{\rho},\tilde{\boldsymbol{\rho}}) = \sum_{k=1}^{K}(\rho_k - \tilde{\rho}_k)^2$, the optimal solution is an arithmetic mean of the components of $\boldsymbol{\rho}_D$ and $\boldsymbol{\rho}_E$:

$$\boldsymbol{\rho}_{\alpha k} = \alpha\rho_{Dk} + (1-\alpha)\rho_{Ek}, \quad k = 1,\ldots,K. \tag{12}$$

When we have $K = 2$ treatment groups, we can explicitly obtain some useful allocations for model (6); (see Zhang and Rosenberger 2007; Sverdlov *et al.* 2012):

(Weib-D$_A$) D$_A$-optimal allocation for most accurate estimation of the treatment contrast $\mu_2 - \mu_1$, which minimizes

$$\mathsf{VAR}(\hat{\mu}_2 - \hat{\mu}_1) = \frac{1}{\rho_1\epsilon_1} + \frac{1}{(1-\rho_1)\epsilon_2} + \frac{(a_1/\epsilon_1 - a_2/\epsilon_2)^2}{\rho_1 d_1 + (1-\rho_1)d_2},$$

with respect to $\rho_1 \in (0, 1)$. The point of the minimum can be found numerically once we have nominal values of ϵ_k, a_k and d_k, $k = 1, 2$.

(Weib-HR) Hazard ratio-optimal allocation for most accurate estimation of the log-hazard ratio $(\mu_2 - \mu_1)/b$, which minimizes

$$\mathsf{VAR}\left(\frac{\hat{\mu}_2 - \hat{\mu}_1}{\hat{b}}\right) = \frac{1}{\rho_1\epsilon_1} + \frac{1}{(1-\rho_1)\epsilon_2} + \frac{(a_1/\epsilon_1 - a_2/\epsilon_2 + (\mu_1 - \mu_2)/b)^2}{\rho_1 d_1 + (1-\rho_1)d_2},$$

with respect to $\rho_1 \in (0, 1)$. The point of the minimum can be found numerically once we have nominal values of b, ϵ_k, a_k, d_k and μ_k, $k = 1, 2$.

(Weib-ZR1) The allocation minimizing the average hazard of a Weibull distribution subject to the restriction on the non-centrality parameter of the Wald test; see Formula 14 on page 162 in Zhang and Rosenberger (2007).

(Weib-ZR2) The allocation minimizing the average hazard of a Weibull distribution assuming the common constant follow-up time for the patients subject to the restriction on the non-centrality parameter of the Wald test; see Formula 15 on page 163 in Zhang and Rosenberger (2007).

2.2. Response-adaptive randomization

The optimal allocation designs discussed in Section 2.1 depend on model parameters $\boldsymbol{\theta}$ which are unknown at the trial onset. For implementing optimal allocation in practice, one can sequentially estimate $\boldsymbol{\theta}$ using accumulating outcome data from patients in the trial and use RAR for treatment allocation. The RAR procedure should have minimal variability and high speed of convergence to the chosen optimal allocation (Hu and Rosenberger 2003).

In time-to-event trials, there is a natural delay in observing responses. The delay time for a patient is the patient's time-to-event outcome (minimum between the event time and the censoring time). A RAR procedure should account for such delays. In general, RAR is appropriate only when responses are "not significantly delayed" in the accrual pattern; e.g., when 60% or more of study subjects contribute outcome data throughout the recruitment period (Hu, Zhang, Cheung, and Chan 2008). In particular, for uniform or exponential patient accrual patterns and censored exponential or Weibull time-to-event outcomes, the asymptotic properties of certain RAR procedures are still valid (Zhang and Rosenberger 2007).

Figure 1 is a schematic illustration of a RAR procedure. Suppose the trial objectives have been quantified and the optimal target allocation is chosen as $\boldsymbol{\rho} = (\rho_1(\boldsymbol{\theta}), \ldots, \rho_K(\boldsymbol{\theta}))^\top$. At the beginning of the trial there are no data; therefore initial treatment allocations must be

non-adaptive. The first Km_0 patients (where m_0 is some small positive integer) are randomized among treatments $1, \ldots, K$ with equal probability. The constant m_0 is user-defined; it should be sufficiently large so that maximum likelihood estimators are attainable. In practice, simulations can be used to elicit the value of m_0. Consider the point in the trial when $i-1$ ($\geq Km_0$) patients have been randomized among K treatment groups and suppose patient i enrolls into the trial. Let $\hat{\boldsymbol{\theta}}_{i-1}$ denote the maximum likelihood estimator of $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\rho}}_{i-1} = (\rho_1(\hat{\boldsymbol{\theta}}_{i-1}), \ldots, \rho_K(\hat{\boldsymbol{\theta}}_{i-1}))^\top$ denote the estimated target allocation based on the treatment assignments and responses from $i-1$ patients in the trial. Then the ith patient is randomized among the treatment groups with probabilities $\boldsymbol{\Psi}_i = (\Psi_1(\hat{\boldsymbol{\rho}}_{i-1}), \ldots, \Psi_K(\hat{\boldsymbol{\rho}}_{i-1}))^\top$, where $\Psi_k(\cdot)$ are some appropriately chosen functions satisfying $0 \leq \Psi_k(\cdot) \leq 1$ and $\sum_{k=1}^K \Psi_k = 1$.

The choice of $\boldsymbol{\Psi} = (\Psi_1(\cdot), \ldots, \Psi_K(\cdot))^\top$ is essential to ensure convergence to the target allocation. There are two types of optimal RAR procedures with established statistical properties that can be used for this purpose: the doubly-adaptive biased coin design (DBCD; Hu and Zhang 2004) and optimal RAR designs based on urn models (Zhang, Hu, Cheung, and Chan 2011). In our software development we implement the former approach. Let $\boldsymbol{\rho} = (\rho_1(\boldsymbol{\theta}), \ldots, \rho_K(\boldsymbol{\theta}))^\top$ be the target allocation, $\hat{\boldsymbol{\theta}}_{i-1}$ be the MLE of $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\rho}}_{i-1} = (\hat{\rho}_{1,i-1}, \ldots, \hat{\rho}_{K,i-1})^\top$ be an estimate of $\boldsymbol{\rho}$ based on available data from $i-1$ patients in the trial, where $\hat{\rho}_{k,i-1} = \rho_k(\hat{\boldsymbol{\theta}}_{i-1})$, $k = 1, \ldots, K$. Also, let $\boldsymbol{N}_{i-1}/(i-1) = (N_{1,i-1}/(i-1), \ldots, N_{K,i-1}/(i-1))^\top$ denote the treatment allocation proportions after $i-1$ patient assignments. Then the DBCD treatment randomization probabilities for the ith patient are as follows:

$$\Psi_{k,i} = \frac{\hat{\rho}_{k,i-1}\left(\frac{\hat{\rho}_{k,i-1}}{N_{k,i-1}/(i-1)}\right)^\gamma}{\sum_{j=1}^K \hat{\rho}_{j,i-1}\left(\frac{\hat{\rho}_{j,i-1}}{N_{j,i-1}/(i-1)}\right)^\gamma}, \quad k = 1, \ldots, K, \tag{13}$$

where γ is a user-specified parameter controlling the degree of randomness of the procedure. Hu and Zhang (2004) showed that under widely satisfied conditions ($\boldsymbol{\rho}(\boldsymbol{\theta})$ must be a continuously differentiable vector function and twice continuously-differentiable in a small neighborhood of the true value of $\boldsymbol{\theta}$), the procedure (13) works as intended: the allocation proportion vector $\boldsymbol{N}_i/i$ converges almost surely (as $i \to \infty$) to $\boldsymbol{\rho}$ and has an asymptotically normal distribution. Hu *et al.* (2008) further showed that these large-sample properties 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 study patients should have their outcomes throughout the accrual period.

We make some further important remarks here. First, at the point of entry of the ith patient, some of the previous $i-1$ patients may not have responded yet, and therefore $\hat{\boldsymbol{\theta}}_{i-1}$ will be potentially computed based on data from fewer than $i-1$ patients. Second, since estimation of $\boldsymbol{\theta}$ frequently involves numerical optimization of the log-likelihood function, the algorithm may not converge and $\hat{\boldsymbol{\theta}}_{i-1}$ may not be attainable. Should this happen, the treatment assignment of the ith patient will be made with equal probability. Third, an investigator must decide how frequently treatment randomization probabilities will be updated throughout the course of the trial. One possibility is a fully sequential procedure, for which the RAR algorithm in Figure 1 is applied for every patient entering the trial. This approach utilizes all available data in the design; however, it requires that data are unblinded throughout the study, which may compromise the integrity of the trial results. Another possibility is a two-stage design: at the first stage, Km_0 patients are randomized equally among K treatments, and based on

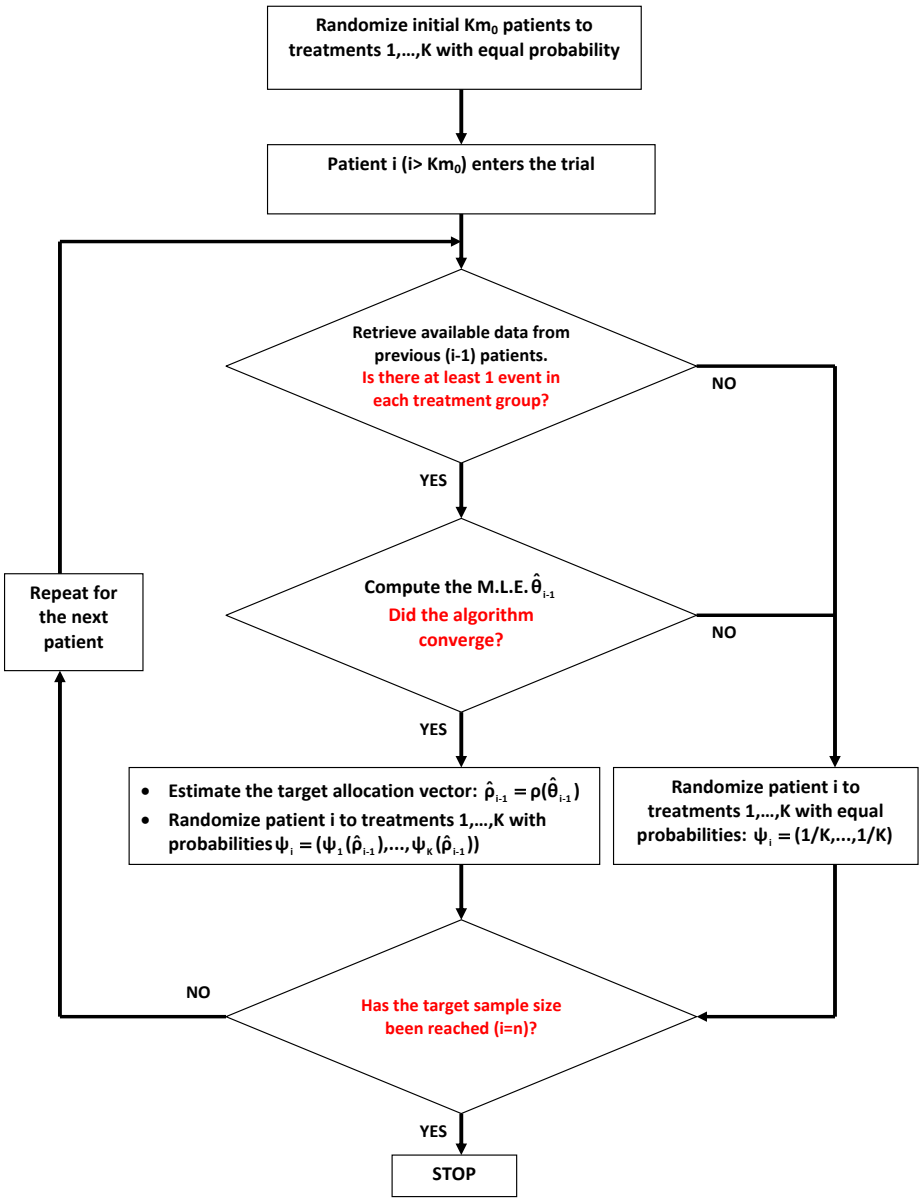

Figure 1: An algorithm for response-adaptive randomization with time-to-event outcomes. In the algorithm, m_0 is a user-defined, small positive integer elicited via simulation.

their data randomization probabilities are adjusted for the patients at the second stage. We pursue a multi-stage design by requiring that the RAR algorithm in Figure 1 is applied after every m, $2m$, $3m$, ... patients, where m is some positive integer. This approach includes both a fully sequential procedure (if $m = 1$) and a two-stage design (if $m = n/2$) as special cases.

2.3. Data analysis following response-adaptive randomization

While RAR generates complex data structures as treatment assignments are no longer independent, the asymptotic properties of estimators and tests are valid under widely satisfied

conditions (Hu *et al.* 2006). The DBCD procedure (13) has the following important asymptotic properties:

- The maximum likelihood estimator $\hat{\boldsymbol{\theta}}_n$ is strongly consistent for $\boldsymbol{\theta}$ and follows an asymptotically normal distribution.
- The vector of treatment allocation proportions $\boldsymbol{N}_n/n$ is strongly consistent for $\boldsymbol{\rho}$ and has an asymptotically normal distribution with variance-covariance matrix

$$\boldsymbol{\Sigma} = \frac{1}{1+2\gamma}\boldsymbol{\Sigma}_1 + \frac{2(1+\gamma)}{1+\gamma}\boldsymbol{\Sigma}_{\text{LB}},$$

  where $\boldsymbol{\Sigma}_1 = \text{diag}\{\rho_1(\boldsymbol{\theta}), \ldots, \rho_K(\boldsymbol{\theta})\} - \boldsymbol{\rho}\boldsymbol{\rho}^\top$ and $\boldsymbol{\Sigma}_{\text{LB}}$ is the lower bound on the variance of a RAR procedure targeting $\boldsymbol{\rho}$. The expression for $\boldsymbol{\Sigma}_{\text{LB}}$ depends on the gradient of $\boldsymbol{\rho}(\boldsymbol{\theta})$ and it can be found using the methodology of Hu and Zhang (2004). Note that as $\gamma \to \infty$, we have $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\text{LB}}$, that is, the DBCD procedure is asymptotically best for $\boldsymbol{\rho}$ (Hu *et al.* 2006), but such a procedure is almost deterministic. In our development we use $\gamma = 2$, as suggested by Rosenberger and Hu (2004).

Since the DBCD procedure has established asymptotic properties, one can perform statistical inference for $\boldsymbol{\theta}$ using standard asymptotic techniques. For instance, treatment contrasts can be tested using the Wald test and asymptotic confidence intervals for different subsets of $\boldsymbol{\theta}$ can be constructed using normal approximations. It is important, however, to investigate accuracy of large-sample approximations for small and moderate sample sizes via simulation.

3. The RARtool package

The **RARtool** package has been developed in MATLAB. Our main motivation for using MATLAB as a basis for software development is that MATLAB has an excellent software optimization toolbox which is well-suited for fast and accurate implementation of numerical methods that are invoked repeatedly in RAR algorithms. MATLAB also has an easy-to-use tool to compile a GUI interface and create stand-alone executable files and display results in an appealing way.

There are two ways to run the **RARtool** package. The first way is to launch the application directly from MATLAB. The second way is to create a stand-alone executable file from MATLAB and launch this executable file. In both cases, it is assumed that the MATLAB Compiler Runtime (**MCR**) version 7.16 or higher is installed; if it is not, the user should run **MCR** Installer, e.g., located in `<matlabroot>\toolbox\compiler\deploy\win32\MCRInstaller.exe` on Windows. For more information about **MCR**, see `http://www.mathworks.com/products/compiler/mcr/`.

Before running **RARtool**, a user should copy the **RARtool** project files into some folder that must be added to the MATLAB path (`File->Set Path` in the main MATLAB window). The project files are as follows:

1. `allocExponential.m` – calculates optimal allocation designs for different optimality criteria in trials with exponential event times.
2. `allocWeibull.m` – calculates optimal allocation designs for different optimality criteria in trials with Weibull event times.

3. `effExponential.m` – calculates efficiency of an allocation design in trials with exponential event times.

4. `effWeibull.m` – calculates efficiency of an allocation design in trials with Weibull event times.

5. `mle.m` – calculates maximum likelihood estimates of $\boldsymbol{\theta}$ for exponential and Weibull event times.

6. `powerExponential.m` – calculates power of an allocation design in trials with exponential event times.

7. `powerWeibull.m` – calculates power of an allocation design in trials with Weibull event times.

8. `rarsimtool.m` – provides the graphical user interface (GUI) for trial simulations.

9. `rartool.m` – provides the GUI to get theoretical and simulated characteristics of the trial.

10. `setrsp.m` – generates random time-to-event outcomes.

11. `trsetup.m` – sets up simulated trial options (used by `rarsimtool.m`).

12. `trsimulate.m` – simulates response-adaptive trials (used by `rarsimtool.m`).

To run **RARtool** directly from MATLAB, one should type the command

```
rartool
```

in the MATLAB Command Window and press `ENTER`. The GUI window will open.

To create a stand-alone executable file, one should type the command

```
mcc -m rartool
```

in the MATLAB Command Window and press `ENTER`. The executable file `rartool.exe` will be created in the main MATLAB folder for Windows. After one launches `rartool.exe`, the GUI window will open. The following subsections give a detailed description of the GUI components of **RARtool**.

3.1. Configuration of clinical trial parameters

Figure 2 displays the first level menu options of **RARtool**. A user should specify:

(i) Time-to-event distribution (`exponential` or `Weibull`).

(ii) Number of treatment arms (`K = 2` or `K = 3`).

Once the choices in (i) and (ii) are made, the following options become available:

(iii) Specify the vector of underlying model parameters $\boldsymbol{\theta}$.

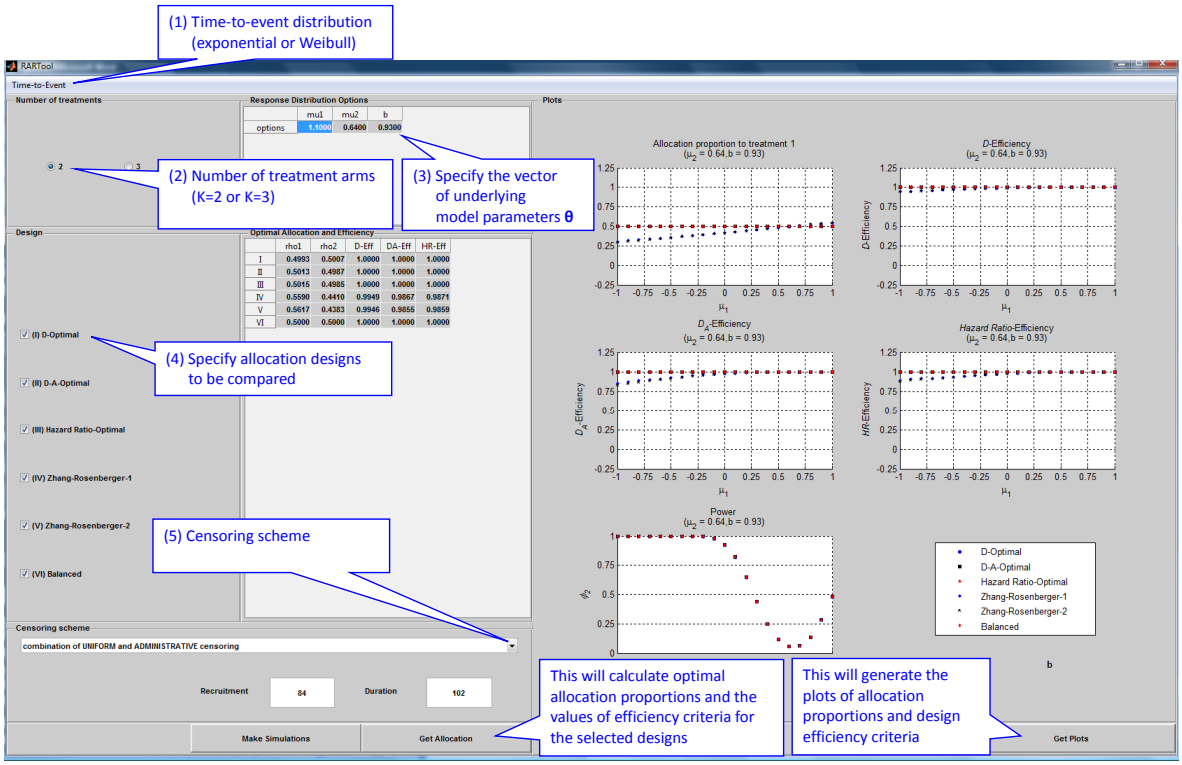

Figure 2: **RARtool** interface (first level GUI) shows the configuration of trial parameters, and tabular and graphical displays of operating characteristics of selected optimal allocation designs.

(iv) Allocation design (see Section 2.1). Note that a user can check/uncheck boxes to include/exclude a particular allocation design from the comparison.

- For an exponential model with $K = 2$ or $K = 3$ treatments, five allocation designs are available:
  - `D-A-optimal` *(Exp-D_A)*.
  - `A-A-optimal` *(Exp-A_A)*.
  - `NP-1-optimal` *(Exp-NP1)*.
  - `NP-2-optimal` *(Exp-E2)*.
  - `Balanced`.
- For a Weibull model with $K = 2$ treatments, six allocation designs are available:
  - `D-optimal` (allocation *(Weib-CO)* with $\alpha = 1$).
  - `D-A-optimal` *(Weib-D_A)*.
  - `Hazard ratio-optimal` *(Weib-HR)*.
  - `Zhang-Rosenberger-1` *(Weib-ZR1)*.
  - `Zhang-Rosenberger-2` *(Weib-ZR2)*.
  - `Balanced`.

- For a Weibull model with $K = 3$ treatments, six allocation designs are available:
  - `D-optimal` (allocation *(Weib-CO)* with $\alpha = 1$).
  - `Compound optimal` (allocation *(Weib-CO)* with a user-defined $\alpha \in [0, 1]$).
  - `Weighted optimal-KL` (allocation *(Weib-WD)* with Kullback-Leibler divergence (Equation 11 with a user-defined $\alpha \in [0, 1]$).
  - `Weighted optimal-Euclid` (allocation *(Weib-WD)* with Euclidean distance (Equation 12 with a pre-defined $\alpha \in [0, 1]$).
  - `Ethical` (Equation 11 with $\alpha = 0$).
  - `Balanced`.

(v) Censoring scheme:

- `with constant follow-up time` (each patient in the study is followed-up for a fixed time period $\tau > 0$. The observed time is $t_{ik} = \min(T_{ik}, \tau)$).
- `combination of uniform and administrative censoring` (A trial has a fixed recruitment period $R > 0$ and a fixed duration $D > R$. Patient arrival times follow a Poisson process over $(0, R)$. The observed time is $t_{ik} = \min(T_{ik}, C_{ik}, D - R)$, where C_{ik} is the ith patient's censoring time, assumed to be uniform over $(0, D)$).

The reason why we focused on the two aforementioned censoring schemes in the current version of **RARtool** is that these schemes are quite common in practice, and they serve as useful starting points to construct optimal designs for time-to-event trials. In particular, for a censoring scheme combining uniform and administrative censoring, the censoring probabilities can be derived in closed form in the exponential response case (Zhang and Rosenberger 2007); this substantially simplifies calculation of optimal allocation designs and simplifies the development of corresponding optimal RAR designs. Our software can potentially be extended to incorporate more complex censoring schemes, such as when both accrual and censoring time distributions are not uniformly distributed. Sverdlov, Rosenberger, and Ryeznik (2013) considered such a situation, where one of the goals was to evaluate robustness of the designs to misspecification of the recruitment pattern. Implementing these additional, more complex accrual patterns and censoring schemes is one important possible extension of the **RARtool** package which we intend to pursue in subsequent versions of our software.

3.2. Operating characteristics of optimal allocation designs

Once the clinical trial parameters have been configured, the operating characteristics of various optimal allocation designs can be computed and visualized graphically. A user has the following options (see Figure 2):

`Get Allocation` – The values of allocation proportions and corresponding statistical efficiency criteria for the selected designs are calculated and displayed in a tabular format. Treatment 1 is the control group, and Treatments 2 and 3 are experimental groups. For a given allocation, statistical efficiency criteria (which take values from 0 to 1, and higher values are desirable) include:

- `D-eff` (D-efficiency for estimating the vector of model parameters $\boldsymbol{\theta}$).

- `DA-eff` (D_A-efficiency for estimating the vector of treatment contrasts (experimental vs. control).
- `HR-eff` (Efficiency for estimating the vector of hazard ratios (experimental vs. control), for the Weibull distribution with $K = 2$ groups).
- `b-eff` (Efficiency for estimating the Weibull hazard pattern via the parameter b).
- `Power` (Statistical power of the Wald test of homogeneity of treatment effects).

`Get Plots` – The plots of optimal allocation proportions and design efficiency criteria are generated for selected optimal allocation designs and the chosen parameter values. To assess changes in allocation and design efficiency criteria, the range of values is taken for the parameter of Treatment 1, whereas the values of other parameters (as defined by the user) are kept fixed.

3.3. Simulation of a response-adaptive randomization trial

In **RARtool** we have implemented the DBCD randomization procedure (13) with $\gamma = 2$. To evaluate performance of a RAR design over multiple simulation runs, one should first configure the clinical trial parameters as described in Section 3.1. After that, one should click the `Make Simulations` button in the first level menu of **RARtool**.

To illustrate the utility of the simulation tool, we reproduce the results from Zhang and Rosenberger (2007) who used the DBCD procedure with Weibull time-to-event outcomes to redesign a phase III survival trial in metastatic breast cancer originally reported by Jones *et al.* (2005). In this trial, $n = 449$ patients were randomized equally between two treatment regimens, docetaxel (Treatment 1) and paclitaxel (Treatment 2). In the intent-to-treat population, median time to progression was significantly longer for patients treated with docetaxel (5.7 months) than with paclitaxel (3.6 months). As in Zhang and Rosenberger (2007), we assume the sample size $n = 449$, $\mu_1 = 1.1$, $\mu_2 = 0.64$, $b = 0.93$, the length of the recruitment period $R = 84$ months, and the overall study duration $D = 102$ months. In the first level menu of **RARtool** (Figure 2) we configure the trial parameters as follows:

(i) `Time-to-Event = Weibull`;

(ii) `Number of treatments = 2`;

(iii) `Response distribution` options are: `mu1 = 1.1`, `mu2 = 0.64`, `b = 0.93`;

(iv) `Design = Zhang-Rosenberger-1`;

(v) `Censoring scheme` is `combination of uniform and administrative censoring` with `Recruitment = 84` and `Duration = 102`.

The second level GUI menu will open (Figure 3). The additional parameters must be specified at this point:

- `Number of patients involved in the trial` (selected is `449`).
- `Number of patients randomized with CRD` (selected is `20`). This determines the initial number of patients randomized with equal probability between two treatments.

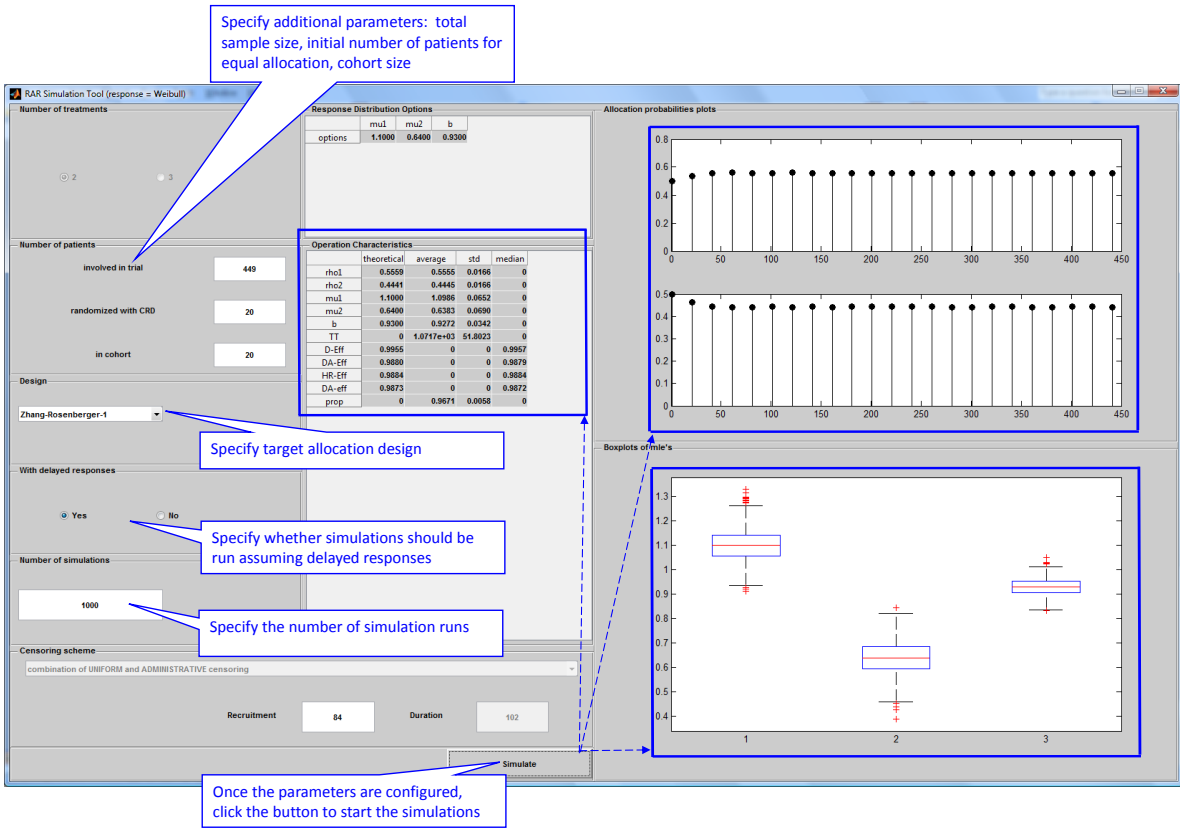

Figure 3: **RARtool** interface (second level GUI) shows simulation of a selected RAR procedure with results displayed in both tabular and graphical formats on the right side.

- `Number of patients in a cohort` (selected is `20`). This determines how frequently randomization probabilities are updated based on accrued data in the trial. In our example, randomization probabilities are updated after every 20 patients.
- `Design` (selected is `Zhang-Rosenberger-1`).
- `With delayed responses?` (`Yes/No`) (Selected is `Yes`). If `No`, then simulations will be performed under an "idealized" scenario assuming that responses become available immediately after the treatment assignment. The `No` option may be useful for assessing theoretical properties of RAR procedures.
- `Number of simulations` (selected is `1000`).

After all these selections have been made, we click on the `Simulate` button, after which the progress status bar will appear. The generated results will be output in tabular and graphical formats (as shown in Figure 3).

In the tabular format (left-hand part of the display), one can see theoretical and simulated (average with standard deviation) values of target allocation proportions and model parameters, simulated (average with standard deviation) total time observed in the study (`TT`), simulated median values of various design efficiency criteria, and the average proportion of

"responders", i.e., subjects in the study whose response data become available throughout the recruitment phase to enable adaptations in the design (`prop`). The simulation results show that the allocation proportion (S.D.) for Treatment 1 is 0.556 (0.017). This matches well the the results reported in Zhang and Rosenberger (2007).

In the right-hand part of the display are the plots of treatment randomization probabilities (averaged across simulation runs) at the points in the trial when the response-adaptive algorithm is applied to modify randomization probabilities. One can see that allocation becomes skewed in favor of Treatment 1 as the trial progresses. Also displayed is the plot of the simulated distribution of the maximum likelihood estimators of model parameters. In theory, model parameter estimators are consistent and have asymptotically normal distributions; however, due to delayed responses the simulated results may deviate from the theoretical ones. Therefore, it may be useful to perform simulations with the `No` option for delayed responses to obtain the design characteristics in the "idealized" no-delay case. In our example, the distributions look approximately normal around the "true" assumed values of the parameters.

4. An application

In this section we illustrate how to use the **RARtool** software to redesign a phase III survival trial for greater efficiency and ethical measures. We consider a phase III survival trial in patients with unresectable squamous cell head and neck cancer reported by Adelstein *et al.* (2003). Between March 1992 and December 1999, 295 eligible patients were randomized equally among three treatment arms: Radiotherapy Alone (Treatment 1), Radiotherapy plus Cisplatin (Treatment 2), and Radiotherapy (Split Course) plus Cisplatin/5FU (Treatment 3). The original study goal was 462 patients; however, because of a slow accrual rate, the study was closed in December 1999, after 295 patients had been enrolled. The data were analyzed as of January 2001. The primary endpoint was overall survival defined as the time from randomization to death from any cause. The median survival times were 12.6, 19.1, and 13.8 months, for Treatments 1, 2, and 3, respectively. The study recruitment duration was $R = 94$ months and the overall study duration was $D = 106$ months. To gain insights of various possible allocation designs for this study, we examine both exponential and Weibull time-to-event distributions. In both cases, we assume a combination of uniform and administrative censoring scheme.

4.1. Exponential event times

For an exponential distribution, Mean = Median/log(2). Under this assumption, based on data in Adelstein *et al.* (2003) we elicit that the mean survival times in the three treatment groups are 18.2, 27.6, and 19.9 months. In the first level menu of **RARtool** we configure the trial parameters as follows: (1) `Time-to-Event = Exponential`; (2) `Number of treatments = 3`; (3) `Response distribution` options are: `theta1 = 18.2`, `theta2 = 27.6`, `theta3 = 19.9`; (4) `Design` options: checked are `D-A-optimal`, `A-A-optimal`, `NP-1 with B = 0.1`, and `balanced`; (5) `Censoring scheme` is `combination of uniform and administrative censoring` with `Recruitment = 94` and `Duration = 106`.

The `Get Allocation` calculates the target allocation proportions as follows: $(0.29, 0.39, 0.32)$ for the `D-A-optimal` design; $(0.34, 0.39, 0.27)$ for the `A-A-optimal` design; $(0.32, 0.58, 0.10)$ for the `NP-1 with B = 0.1` design; and $(0.33, 0.33, 0.33)$ for the `balanced` design. Note the

different degree of skewing in the allocation proportions. The most skewed is the `NP-1 with B = 0.1` design which should assign 58% of study patients to the most efficacious Treatment 2, 32% of study patients to the "control" Treatment 1, and 10% of study patients to Treatment 3. Let us consider D_A-efficiency as a measure of goodness of the designs. From **RARtool**, the values of D_A-efficiency for the `D-A-optimal`, `A-A-optimal`, `NP-1 with B = 0.1`, and `balanced` designs are 1, 0.97, 0.58, and 0.98, respectively.

Next, we click on the `Make Simulations` button in **RARtool**. In addition to the already specified parameters, we specify that `Number of patients involved in the trial = 295`, `Number of patients randomized with CRD = 30`, `Number of patients in cohort = 30`, and `Number of simulations = 5000`. We perform seven sets of simulations consecutively for the four designs (`D-A-optimal`, `A-A-optimal`, `NP-1 with B = 0.1`, and `balanced`). For the first three designs, we explore both the cases of "without delay" and "with delay" (options `No` and `Yes`, respectively, in the `With delayed responses` field). For the balanced randomization design, which is non-adaptive, only the `No` option is used.

Table 1 summarizes the key operating characteristics of the four randomization designs. For each of the three response-adaptive designs, there is overall a good agreement between the operating characteristics (average allocation proportions and median values of D_A-efficiency) in the simulations "without delay" and the corresponding theoretical values. Some discrepancy is present for the "most skewed" `NP-1 with B = 0.1` design. In the simulations "with delay", while the adaptive designs are skewed towards their intended targets, the convergence is not fully achieved due to delayed responses. It is worth noting that the `D-A-optimal` design is advantageous over the `balanced` design both in terms of D_A-efficiency and the total survival time, both in the "without delay" and the "with delay" cases.

4.2. Weibull event times

For the Weibull model (6), the median of the event time in group k is $e^{\mu_k}(\log 2)^{1/b}$. To match the median survival times reported in Adelstein *et al.* (2003) and to be consistent with the setting in Section 4.1, we assume $b = 1$ (exponential distribution) and elicit $\mu_1 = 2.90$, $\mu_2 = 3.32$, $\mu_1 = 2.99$. We want to compare the merits of three allocation designs developed under the Weibull model. These designs are:

- `D-Optimal` – D-optimal allocation (Equation 12 with $\alpha = 1$);
- `WO-Euclid with alpha = 0.5` – Weighted distance optimal allocation that provides equal trade-off between D-efficiency and ethics (Equation 12 with $\alpha = 0.5$);
- `Ethical` – Ethical allocation (Equation 12 with $\alpha = 0$).

As in the exponential case (Section 4.1), in the first level menu we configure the trial parameters: (1) `Time-to-Event = Weibull`; (2) `Number of treatments = 3`; (3) `Response distribution` options are: `mu1 = 2.90`, `mu2 = 3.32`, `mu3 = 2.99`, `b = 1`; (4) `Design` options: checked are `D-Optimal`, `WO-Euclid with alpha = 0.5`, and `Ethical` (also one should check the `Long trial` option which indicates that this is a survival trial where long times are desirable); (5) `Censoring scheme` is `combination of uniform and administrative censoring` with `Recruitment = 94` and `Duration = 106`.

The `Get Allocation` calculates the target allocation proportions as follows: $(0.34, 0.32, 0.34)$

Design			Treatment Allocation Proportions			D_A-efficiency		Total Time	
			1	2	3				
`D-A-optimal`	Theoretical		0.29	0.39	0.32	Theoretical	1.00	Simulated	
	Delay = "No"	Mean	0.29	0.39	0.32	Simulated†	0.99	Mean	4045
		(S.D.)	(0.03)	(0.03)	(0.03)			(S.D.)	(199)
	Delay = "Yes"	Mean	0.31	0.37	0.32	Simulated†	0.99	Mean	4020
		(S.D.)	(0.03)	(0.03)	(0.03)			(S.D.)	(200)
			1	2	3				
`A-A-optimal`	Theoretical		0.34	0.39	0.27	Theoretical	0.97	Simulated	
	Delay = "No"	Mean	0.34	0.39	0.26	Simulated†	0.96	Mean	4026
		(S.D.)	(0.05)	(0.05)	(0.04)			(S.D.)	(199)
	Delay = "Yes"	Mean	0.37	0.35	0.28	Simulated†	0.95	Mean	3993
		(S.D.)	(0.04)	(0.04)	(0.04)			(S.D.)	(203)
			1	2	3				
`NP-1 with B = 0.1`	Theoretical		0.32	0.58	0.10	Theoretical	0.58	Simulated	
	Delay = "No"	Mean	0.26	0.51	0.23	Simulated†	0.78	Mean	4122
		(S.D.)	(0.08)	(0.11)	(0.10)			(S.D.)	(217)
	Delay = "Yes"	Mean	0.29	0.42	0.29	Simulated†	0.95	Mean	4053
		(S.D.)	(0.06)	(0.08)	(0.07)			(S.D.)	(201)
			1	2	3				
`balanced`	Theoretical		1/3	1/3	1/3	Theoretical	0.98	Simulated	
	Delay = "No"	Mean	0.33	0.33	0.33	Simulated†	0.98	Mean	3997
		(S.D.)	(0.01)	(0.01)	(0.01)			(S.D.)	(187)

† Median value of the simulated distribution.

Table 1: Redesign of a phase III survival trial reported by Adelstein *et al.* (2003) using **RARtool**, assuming exponential outcomes, based on 5,000 simulation runs.

for `D-Optimal`; $(0.28, 0.42, 0.30)$ for `WO-Euclid with alpha = 0.5`; and $(0.22, 0.51, 0.27)$ for `Ethical`.

Next, we click on the `Make Simulations` button and specify additional parameters as in the exponential case (Section 4.1): `Number of patients involved in the trial = 295`, `Number of patients randomized with CRD = 30`, `Number of patients in cohort = 30`, and `Number of simulations = 5000`. We run six sets of simulation studies consecutively for the three designs (`D-Optimal`, `WO-Euclid with alpha = 0.5`, and `Ethical`), each with the options `No` and `Yes` in the `With delayed responses` field.

Table 2 summarizes the key results. The `D-Optimal` design is almost identical to the balanced completely randomized design. It is highly efficient, but does not allocate patients to better performing treatment arms. Designs `WO-Euclid with alpha = 0.5` and `Ethical` result in skewed allocations in favor of treatments with longer event times. The `Ethical` design is most skewed and has highest average total survival time; yet it is also most variable and least efficient among the three designs. The `WO-Euclid with alpha = 0.5` design achieves a reasonable trade-off between inferential efficiency (D_A-efficiency $= 0.98$) and ethics (total survival time). Note that because of delayed responses the intended skewing is not achieved for the latter two designs. This is consistent with the findings in the exponential case (Section 4.1).

4.3. Some further remarks

The examples considered in this paper are phase III trials. One may argue that the regulatory agencies give very limited flexibility to adaptive phase III confirmatory trials, in particular

Design			Treatment Allocation Proportions			D_A-efficiency‡		Total Time	
			1	2	3				
`D-Optimal`	Theoretical		0.34	0.32	0.34	Theoretical	0.99	Simulated	
	Delay = "No"	Mean	0.34	0.32	0.34	Simulated†	1.00	Mean	4661
		(S.D.)	(0.01)	(0.01)	(0.01)			(S.D.)	(239)
	Delay = "Yes"	Mean	0.34	0.33	0.33	Simulated†	1.00	Mean	4630
		(S.D.)	(0.01)	(0.01)	(0.01)			(S.D.)	(240)
			1	2	3				
`WO-Euclid with alpha = 0.5`	Theoretical		0.28	0.42	0.30	Theoretical	0.98	Simulated	
	Delay = "No"	Mean	0.28	0.42	0.30	Simulated†	0.98	Mean	4776
		(S.D.)	(0.05)	(0.05)	(0.04)			(S.D.)	(246)
	Delay = "Yes"	Mean	0.30	0.38	0.32	Simulated†	0.98	Mean	4699
		(S.D.)	(0.03)	(0.04)	(0.03)			(S.D.)	(245)
			1	2	3				
`Ethical`	Theoretical		0.22	0.51	0.27	Theoretical	0.92	Simulated	
	Delay = "No"	Mean	0.22	0.52	0.27	Simulated†	0.89	Mean	4856
		(S.D.)	(0.07)	(0.09)	(0.07)			(S.D.)	(274)
	Delay = "Yes"	Mean	0.27	0.42	0.30	Simulated†	0.95	Mean	4811
		(S.D.)	(0.06)	(0.07)	(0.06)		0.90	(S.D.)	(259)

† Median value of the simulated distribution.
‡ Computed relative to the balanced allocation.

Table 2: Redesign of a phase III survival trial reported by Adelstein *et al.* (2003) using **RARtool**, assuming Weibull outcomes, based on 5,000 simulation runs.

trials with response-adaptive randomization (US Food and Drug Administration 2010). The main reasons we considered phase III trials in this paper are three-fold. First, we wanted to illustrate that our software can successfully validate simulation results previously reported in the literature – the example in Section 3.3 shows that **RARtool** accurately replicates the results from Zhang and Rosenberger (2007). Second, the examples from Sections 4.1 and 4.2 show that RAR can be an attractive strategy for phase III survival trials which exhibit significant differences among the treatment arms – in this case the allocation is skewed to treatment arms with longer survival times. Third, and most important, we believe that the examples presented here add to the body of knowledge on how RAR works in phase III survival trials; we hope that this work along with others will eventually pave the way for implementation of RAR trials in practice.

We would also like to mention that our software can simulate design operating characteristics for any user-specified sample size. As such the software can potentially handle designs of randomized phase II time-to-event trials with small or moderate sample sizes. While RAR methodology relies on large sample approximations, simulations show that moderate sample sizes ($n = 50$ to 100) are sufficient for the large sample approximations to be reasonably valid if the doubly-adaptive biased coin design of Hu and Zhang (2004), implemented in the current paper, is used; see Rosenberger, Sverdlov, and Hu (2012, p. 728) for further details.

5. Conclusions and future work

In this paper, we have presented a user-friendly software **RARtool** which should aid an investigator in planning response-adaptive randomized comparative clinical trials with censored time-to-event outcomes. Our software implements various optimal allocation schemes (Zhang

and Rosenberger 2007; Sverdlov *et al.* 2011, 2012, 2014), under the assumptions of exponential and Weibull distribution for the event times. The implemented randomization designs are scientifically sound and can provide trade-off between statistical efficiency and ethical considerations. While the current version of **RARtool** can be useful in designing randomized time-to-event clinical trials, it can be improved in a number of ways. We outline a few possible extensions:

- Explore additional design optimality criteria and additional censoring schemes.
- Implement optimal allocation designs and corresponding RAR procedures in a ($K > 3$)-treatment case.
- Implement urn-based optimal RAR procedures (Zhang *et al.* 2011) and compare them with the the DBCD procedure (Hu and Zhang 2004) implemented in this paper.
- Incorporate group-sequential monitoring schemes on top of RAR to allow early stopping of a trial for efficacy or futility.
- Develop a utility that allows interactive sequential data entry to recalculate treatment randomization probabilities and facilitate implementation of RAR designs in practice.

We intend to tackle the above mentioned problems in our future work.

Acknowledgments

The authors would like to thank two anonymous referees whose comments have led to an improved version of this paper. The research of Weng Kee Wong reported in this paper was partially supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R01GM107639. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Oleksandr Sverdlov completed work on this paper during his employment at Novartis Pharmaceuticals Corporation. He would like to acknowledge Novartis Statistical Methodology and Translational Sciences groups for very helpful discussions on adaptive clinical trial designs.

References

Adelstein DJ, Adams GL, Wagner H, Kish JA, Ensley JF, Schuller DE, Forastiere AA (2003). "An Intergroup Phase III Comparison of Standard Radiation Therapy and two Schedules of Concurrent Chemoradiotherapy in Patients with Unresectable Squamous Cell Head and Neck Cancer." *Journal of Clinical Oncology*, **21**(1), 92–98.

Atkinson A (1982). "Optimum Biased Coin Designs for Sequential Clinical Trials with Prognostic Factors." *Biometrika*, **69**(1), 69–67.

Carroll KJ (2003). "On the Use and Utility of the Weibull Model in the Analysis of Survival Data." *Controlled Clinical Trials*, **24**(6), 682–701.

Cheung YK, Lurdes YT, Wathen JK, Thall PF (2006). "Continuous Bayesian Adaptive Randomization Based on Event Times with Covariates." *Statistics in Medicine*, **25**(1), 55–70.

Hu F, Rosenberger WF (2003). "Optimality, Variability, Power: Evaluating Response-Adaptive Randomization Procedures for Treatment Comparisons." *Journal of the American Statistical Association*, **98**(463), 671–678.

Hu F, Rosenberger WF (2006). *The Theory of Response-Adaptive Randomization*. John Wiley & Sons.

Hu F, Rosenberger WF, Zhang LX (2006). "Asymptotically Best Response-Adaptive Randomization Procedures." *Journal of Statistical Planning and Inference*, **136**(6), 1911–1922.

Hu F, Zhang LX (2004). "Asymptotic Properties of Doubly-Adaptive Biased Coin Designs for Multitreatment Clinical Trials." *The Annals of Statistics*, **32**(1), 268–301.

Hu F, Zhang LX, Cheung SH, Chan WS (2008). "Doubly Adaptive Biased Coin Designs with Delayed Responses." *Canadian Journal of Statistics*, **36**(4), 541–559.

Jones SE, Erban J, Overmoyer B, Budd GT, Hutchings L, Lower E, Laufman L, Sundaram S, Urba WJ, Pritchard KI, Mennel R, Richards D, Olsen S, Meyers ML, Ravdin PM (2005). "Randomized Phase III Study of Docetaxel Compared with Paclitaxel in Metastatic Breast Cancer." *Journal of Clinical Oncology*, **23**(24), 5542–5551.

Lawless JF (2003). *Statistical Models and Methods for Lifetime Data*. 2nd edition. John Wiley & Sons.

Rosenberger WF, Hu F (2004). "Maximizing Power and Minimizing Treatment Failures in Clinical Trials." *Clinical Trials*, **1**(2), 141–147.

Rosenberger WF, Sverdlov O, Hu F (2012). "Adaptive Randomization for Clinical Trials." *Journal of Biopharmaceutical Statistics*, **22**(4), 719–736.

Sverdlov O, Rosenberger WF (2013). "On Recent Advances in Optimal Allocation Designs in Clinical Trials." *Journal of Statistical Theory and Practice*, **7**(4), 753–773.

Sverdlov O, Rosenberger WF, Ryeznik Y (2013). "Utility of Covariate-Adjusted Response-Adaptive Randomization in Survival Trials." *Statistics in Biopharmaceutical Research*, **5**(1), 38–53.

Sverdlov O, Ryeznik Y, Wong WK (2012). "Doubly-Adaptive Biased Coin Designs for Balancing Competing Objectives in Time-to-Event Trials." *Statistics and Its Interface*, **5**(4), 401–414.

Sverdlov O, Ryeznik Y, Wong WK (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.

Sverdlov O, Tymofyeyev Y, Wong WK (2011). "Optimal Response-Adaptive Randomized Designs for Multi-Armed Survival Trials." *Statistics in Medicine*, **30**(24), 2890–2910.

The MathWorks, Inc (2011). *MATLAB – The Language of Technical Computing, Version R2011b*. The MathWorks, Inc., Natick, Massachusetts. URL http://www.mathworks.com/products/matlab/.

Tymofyeyev Y, Rosenberger WF, Hu F (2007). "Implementing Optimal Allocation in Sequential Binary Response Experiments." *Journal of the American Statistical Association*, **102**(477), 224–234.

US Food and Drug Administration (2010). *Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics (Draft Document)*. URL http://www.fda.gov/downloads/Drugs/Guidances/ucm201790.pdf.

Zhang L, Rosenberger WF (2007). "Response-Adaptive Randomization for Survival Trials: the Parametric Approach." *Journal of the Royal Statistical Society C*, **56**(2), 153–165.

Zhang LX, Hu F, Cheung SH, Chan WS (2011). "Immigrated Urn Models – Theoretical Properties and Applications." *The Annals of Statistics*, **39**(1), 643–671.

Affiliation:

Yevgen Ryeznik
Department of Mathematics
Uppsala University
Lägerhyddsvägen 1, Hus 1, 5 och 7
Box 480, 751 06, Uppsala, Sweden
E-mail: yevgen.ryeznik@math.uu.se

Oleksandr Sverdlov
R&D Global Biostatistics
EMD Serono, Inc.
45A Middlesex Turnpike
Billerica, MA 01821, United States of America
E-mail: alex.sverdlov@emdserono.com

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

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume 66, Issue 1 *Submitted:* 2012-12-03
August 2015 *Accepted:* 2014-10-19