Journal of Statistical Planning and Inference 225 (2023) 219–242

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Covariate-adjusted response-adaptive designs for censored survival responses

Ayon Mukherjee [a], D. Stephen Coad [b], Sayantee Jana [c,*]

[a] Regulatory Affairs and Drug Development Solutions (RADDS), IQVIA, Unterschweinstiege 2-14, 60549, Frankfurt am Main, Germany
[b] School of Mathematical Sciences, Queen Mary University of London, Mile End Rd, London E1 4NS, UK
[c] Department of Mathematics, Indian Institute of Technology, Hyderabad, Kandi, Sangareddy, Telangana, India

ARTICLE INFO

Article history:
Received 17 April 2022
Received in revised form 20 November 2022
Accepted 2 January 2023
Available online 6 January 2023

Keywords:
Optimal allocation
Ethics
Power
Variability
Wald test
Weibull response

ABSTRACT

Covariate-adjusted response-adaptive designs use the available responses to skew treatment allocation in a clinical trial, in favour of the treatment, found at an interim stage of the trial, to be best for a patient's covariate profile. In this study, such designs are developed for censored Weibull responses. The designs are based on the covariate-adjusted doubly-adaptive biased coin design and the covariate-adjusted efficient randomized-adaptive design. The treatment allocation proportion for these designs, converges empirically to the expected target value. An extensive simulation study of the operating characteristics of these designs demonstrates that they can be considered as suitable alternatives to traditional balanced randomization designs, provided responses related to the primary endpoint are available during the recruitment phase to enable adaptations in the design. An existing clinical trial is redesigned using the proposed methodology to illustrate its implementability in real-life scenario.

1. Introduction

Contemporary clinical trial designs aim to increase the efficiency of studies, consequently increasing focus on adaptive designs . Recognizing this importance, the United States Food and Drug Administration (FDA) has issued a guidance for industry on implementation of adaptive designs for clinical trials of drugs and biologics (U.S. Department of Health and Human Services Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER) (2019)). The adaptation process allows the trial to adjust to information that was not available when the trial began. Clinical trials are often designed with adaptive features to achieve balance in sequential allocation of patients across competing treatment arms. It is also designed to force imbalance by allocating more patients to the superior treatment arm.

Larger number of patients receiving the better treatment leads to sequential experiments in which data are analysed dynamically and new allocations are made in light of the estimated parameters. However, advocates of traditional balanced randomized designs argue that having a balanced allocation of patients across the treatment arms, helps estimate treatment effects efficiently. Since clinical trials involve human patients, balanced allocation can be a serious ethical problem as it assigns almost half of the patients to the worse treatment arm. In order to balance the competing goals of ethics and statistical efficiency in a clinical trial, response-adaptive designs were developed and used. However,

* Corresponding author.
E-mail addresses: ayon.mukherjee@iqvia.com (A. Mukherjee), d.s.coad@qmul.ac.uk (D.S. Coad), sayantee.jana@math.iith.ac.in (S. Jana).

https://doi.org/10.1016/j.jspi.2023.01.001
0378-3758/

such designs do not consider the fact that human patients are invariably heterogeneous with regards to their baseline concomitant information. In the field of personalized medicine, consideration of such information becomes important when allocating a patient to a treatment arm.

Covariate-adjusted response-adaptive designs address this issue by randomizing more patients to the better treatment, based on their covariate profiles. Investigators are often aware of critical baseline covariates that might have a strong influence on patient responses, and they may wish to adjust the randomization procedure for these covariates. Rosenberger and Sverdlov (2008) gave an overview of different techniques for handling covariates in the design of clinical trials and distinguished between two main approaches, namely, the covariate-adaptive randomization and the covariate-adjusted response-adaptive randomization (CARA) procedures. CARA randomization is applicable to clinical trials where non-linear and heteroscedastic models determine the relationship between responses, treatments and covariates, and when multiple experimental objectives are pursued in the trial. Moreover, in some clinical trials the degree and direction of treatment effect may differ for patient subgroups within a treatment, and the research design should account for such covariate-specific treatment effects.

The goal of a CARA procedure in a phase III clinical trial, is to skew allocation in the direction of the better-performing treatment arm, for a given patient's covariate profile, while maintaining the power of the statistical test for treatment comparisons. These designs heavily rely on correctly specified parametric models. Biswas et al. (2016) developed these designs for phase III clinical trials when the treatment response is a survival time and there is random censoring. Asymptotic properties of allocation proportions for CARA designs, and the maximum likelihood estimators have been established by Zhang et al. (2007). Although the exponential survival model with independent right-censoring has been previously considered by Sverdlov et al. (2013) as a useful starting point for developing CARA randomization procedures, its constant hazard property limits its application in real-life clinical trials. To extend the scope of application of these designs, methods are developed in this paper for Weibull survival model, considering independent right-censoring mechanism. Sverdlov et al. (2013) also explored the robustness of the proposed CARA procedures to model mis-specification. This also applies for Weibull distributed responses and is discussed briefly in this article.[1]

The outline of this paper is as follows. Section 2 explains the basic background relating to the Weibull survival model, the maximum likelihood estimation approach of its parameters for right-censored data, and introduces the derived target allocation proportions developed in this paper based on the Weibull survival model. Section 3 discusses the CARA randomization procedures to achieve the derived targets, and also derives a theoretical template for the comparison of CARA designs in clinical trials. The designs proposed in Section 2 have been validated using extensive simulations in Section 4, and an example of a real-life clinical trial is provided to illustrate the implementability of the proposed CARA designs in clinical practice. Section 5 presents a discussion of the overall findings and suggests scopes of further work including regulatory perspectives. The article is concluded with an overall summary in Section 6. Appendices provide the derivation of the approximate variances of model parameter estimates, the derivation of the probability of an event for the Weibull model, and also the derivation of a proposed CARA optimal allocation proportion, from this study.

2. Materials and methods

2.1. The Weibull model

Associated with patient i is a vector of baseline covariates $\mathbf{z}_i = (1, z_{1i}, \ldots, z_{pi})^T$ for $i = 1, 2, \ldots, n$. For treatment $j = A, B$, let γ_j be the shape parameter for the distribution of the responses, let $\boldsymbol{\beta}_j = (\beta_{j0}, \beta_{j1}, \ldots, \beta_{jp})^T$ be the vector of unknown model parameters and let $\mu_j(\mathbf{z}_i) = \exp(\boldsymbol{\beta}_j^T \mathbf{z}_i)$ be the scale parameter, given the ith covariate vector $\mathbf{z}_i$. Then, given $\mathbf{z}_i$, a patient's survival time on treatment j, T_{ij}, follows a Weibull distribution with probability density function

$$f_j(t|\mathbf{z}_i; \boldsymbol{\beta}_j, \gamma_j) = \left\{\frac{\gamma_j}{\mu_j(\mathbf{z}_i)}\right\}\left\{\frac{t}{\mu_j(\mathbf{z}_i)}\right\}^{\gamma_j - 1} \exp\left[-\left\{\frac{t}{\mu_j(\mathbf{z}_i)}\right\}^{\gamma_j}\right], \text{ for } t > 0.$$

The survivor function is $S_j(t|\mathbf{z}_i; \boldsymbol{\beta}_j, \gamma_j) = \exp\left[-\{t/\mu_j(\mathbf{z}_i)\}\right]^{\gamma_j}$ and the hazard function can be expressed as $h_j(t|\mathbf{z}_i; \boldsymbol{\beta}_j, \gamma_j) = \{\gamma_j/\mu_j(\mathbf{z}_i)\}\{t/\mu_j(\mathbf{z}_i)\}^{\gamma_j - 1}$. It is also known that $S_j(t|\mathbf{z}_i; \boldsymbol{\beta}_j, \gamma_j) = \exp\{-\int_0^t h_j(u|\mathbf{z}_i; \boldsymbol{\beta}_j, \gamma_j)du\}$ where

$$\int_0^t h_j(u|\mathbf{z}_i; \boldsymbol{\beta}_j, \gamma_j)du = H_j(t|\mathbf{z}_i; \boldsymbol{\beta}_j, \gamma_j) = -\log\{S_j(t|\mathbf{z}_i; \boldsymbol{\beta}_j, \gamma_j)\}$$

is called the cumulative hazard, which therefore is $H_j(t|\mathbf{z}_i; \boldsymbol{\beta}_j, \gamma_j) = \{t/\mu_j(\mathbf{z}_i)\}^{\gamma_j}$. The mean survival time is $\lambda_j(\mathbf{z}_i) = \mu_j(\mathbf{z}_i)\Gamma(1 + 1/\gamma_j)$, where Γ denotes the gamma function.

[1] **Abbreviations**: CARA: covariate-adjusted response-adaptive randomization, ERADE: efficient randomized-adaptive designs, CAERADE: covariate-adjusted ERADE, CID: Complex Innovative Trial Design, DBCD: doubly-adaptive biased coin design, CADBCD: covariate-adjusted DBCD, EGFR: Epidermal growth factor receptor, FDA : The US Food and Drug Administration, ICH E9: International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, MAMS: Multi-Arm Multi Stage trial designs, MLE : Maximum Likelihood Estimates, PFS: progression-free survival, NDA: New Drug Application, PD: Pharmacodynamics, PK: Pharmacokinetics, POC: Proof of Concept, TKI: Tyrosine Kinase Inhibitor, RECIST v1.1.: Response Evaluation Criteria in Solid Tumours version 1.1, SLE : Systemic Lupus Erythematosus.

In what follows, it is useful to work with the logarithm of the survival time. If $Y_{ij} = \log(T_{ij})$, then Y_{ij} has an extreme value distribution with probability density function

$$f_j(y|\mathbf{z}_i; \boldsymbol{\beta}_j, \gamma_j) = \gamma_j \exp\left([y - \log\{\mu_j(\mathbf{z}_i)\}]\gamma_j\right) \exp\left(-\exp\left[y - \log\{\mu_j(\mathbf{z}_i)\}\right]\gamma_j\right), \text{ for } -\infty < y < \infty. \tag{1}$$

2.2. Maximum likelihood estimation

Censoring in a common phenomenon in survival trials, where, a critical assumption is that the survival times and the censoring times are independent. Since patients arrive sequentially in a clinical trial and are observed until the end of the treatment period or the follow-up period, the generalized Type I right censoring scheme is considered here. When subjects enter a study at different time-points, and are all observed to a fixed moment in time, it is termed as generalized Type I right censoring. Here time is measured from a different origin for each subject. Let C_{ij} be the censoring time for patient i on treatment j. Then the observed outcome is a bivariate random vector (X_{ij}, δ_{ij}), where $X_{ij} = \min(T_{ij}, C_{ij})$, and $\delta_{ij} = 0$ if X_{ij} is a right-censored time, and $\delta_{ij} = 1$ if an event occurred at time X_{ij}.

For a random sample of pairs (x_{ij}, δ_{ij}) of n_j patients on treatment j, the likelihood function is given by

$$L_j(\boldsymbol{\beta}_j, \gamma_j) = \prod_{i=1}^{n_k} \{f_j(x_{ij}|\mathbf{z}_i; \boldsymbol{\beta}_j, \gamma_j)\}^{\delta_{ij}} \{S_j(x_{ij}|\mathbf{z}_i; \boldsymbol{\beta}_j, \gamma_j)\}^{1-\delta_{ij}}. \tag{2}$$

Since in survival analysis, $f_j(x_{ij}|\mathbf{z}_i; \boldsymbol{\beta}_j, \gamma_j) = S_j(x_{ij}|\mathbf{z}_i; \boldsymbol{\beta}_j, \gamma_j) h_j(x_{ij}|\mathbf{z}_i; \boldsymbol{\beta}_j, \gamma_j)$. Therefore, (2) can be further simplified to

$$L_j(\boldsymbol{\beta}_j, \gamma_j) = \prod_{i=1}^{n_j} \{h_j(x_{ij}|\mathbf{z}_i; \boldsymbol{\beta}_j, \gamma_j)\}^{\delta_{ij}} S_j(x_{ij}|\mathbf{z}_i; \boldsymbol{\beta}_j, \gamma_j).$$

For a random sample of observations (x_{ij}, δ_{ij}) from the Weibull distribution with independent right censoring, let $y_{ij} = \log(x_{ij})$. It was seen earlier that the density of Y_{ij} is given by (1). Therefore, the likelihood function in this case is

$$L_j(\boldsymbol{\beta}_j, \gamma_j) = \prod_{i=1}^{n_j} \left\{\gamma_j \exp([y_{ij} - \log\{\mu_j(\mathbf{z}_i)\}]\gamma_j)\right\}^{\delta_{ij}} \exp\left(-\exp\left[y_{ij} - \log\{\mu_j(\mathbf{z}_i)\}\right]\gamma_j\right).$$

So the log-likelihood function here is given by

$$l_j(\boldsymbol{\beta}_j, \gamma_j) = \log(\gamma_j) \sum_{i=1}^{n_j} \delta_{ij} + \sum_{i=1}^{n_j} \left(\delta_{ij}[y_{ij} - \log\{\mu_j(\mathbf{z}_i)\}]\gamma_j - \exp\left[y_{ij} - \log\{\mu_j(\mathbf{z}_i)\}\right]\gamma_j\right). \tag{3}$$

The first partial derivatives of (3) with respect to $\log\{\mu_j(\mathbf{z}_i)\}$ and $1/\gamma_k$ are

$$\frac{\partial l_j}{\partial \log\{\mu_j(\mathbf{z}_i)\}} = -\sum_{i=1}^{n_j} \left(\delta_{ij} - \exp\left[y_{ij} - \log\{\mu_j(\mathbf{z}_i)\}\right]\gamma_j\right)\gamma_j \tag{4}$$

and

$$\frac{\partial l_j}{\partial(1/\gamma_j)} = -\sum_{i=1}^{n_j} \delta_{ij}\gamma_j - \sum_{i=1}^{n_j} \left(\delta_{ij} - \exp\left[y_{ij} - \log\{\mu_j(\mathbf{z}_i)\}\right]\gamma_j\right)[y_{ij} - \log\{\mu_j(\mathbf{z}_i)\}]\gamma_j^2. \tag{5}$$

Equating (4) and (5) to zero and numerically solving them for $1/\gamma_j$ and $\log\{\mu_j(\mathbf{z}_i)\}$ gives the maximum likelihood estimates (MLEs) for $1/\gamma_j$ and $\log\{\mu_j(\mathbf{z}_i)\}$. These, in turn, give the MLEs of the regression coefficients. The variances of the estimated parameters are obtained by calculating the second partial derivatives of the log-likelihood function, with respect to $1/\gamma_j$ and $\log\{\mu_j(\mathbf{z}_i)\}$. The details are provided in Appendix A.

2.3. The proposed optimal target allocation designs

One of the approaches for developing CARA randomization designs, is to derive the optimal target allocation proportion for a model without covariates and use its covariate-adjusted version for the sequential allocation of patients (Zhang et al., 2007). For a clinical trial with censored Weibull survival times, the optimal target allocation proportions can be obtained by using the maximum likelihood approach presented in Section 2.2. Let $\varsigma_j = \gamma_j[Y_j - \log\{\mu_j(\mathbf{z})\}]$ and let $\epsilon_j(\boldsymbol{\beta}_j, \gamma_j, \mathbf{z})$ be the probability of an event on treatment j. Then, as will be seen below, the optimal allocation proportion depends on the function

$$G_j = \frac{\epsilon_j(\boldsymbol{\beta}_j, \gamma_j, \mathbf{z}) + E(\varsigma_j^2 \exp \varsigma_j)}{\epsilon_j^2(\boldsymbol{\beta}_j, \gamma_j, \mathbf{z}) + \epsilon_j(\boldsymbol{\beta}_j, \gamma_j, \mathbf{z})E(\varsigma_j^2 \exp \varsigma_j) - \{E(\varsigma_j \exp \varsigma_j)\}^2}, \tag{6}$$

where $\epsilon_j(\boldsymbol{\beta}_j, \gamma_j, \mathbf{z}) = P(T_j \leq C_j|\boldsymbol{\beta}_j, \gamma_j, \mathbf{z})$ for $j = A, B$. The derivation of the formal analytical form for $\epsilon_j(\boldsymbol{\beta}_j, \gamma_j, \mathbf{z})$ is provided in Appendix B. The asymptotic variance, σ_j^2, of the logarithm of the estimated scale parameter is derived in Appendix A.

Consider minimizing the average hazard in the trial subject to the constraint that the asymptotic variance of the difference in the logarithms of the estimated scale parameters is a constant (Zhang and Rosenberger, 2007). This would ensure that, for any choice of the number of patients allocated to each treatment, the power for testing for a difference in the treatment effects would remain fixed. For a survival trial of overall duration D, the cumulative hazard at time D will be $(D/\mu_j)^{\gamma_j}$. Therefore, minimizing the total cumulative hazard $n_A(D/\mu_A)^{\gamma_A} + n_B(D/\mu_B)^{\gamma_B}$, subject to the constraint that the asymptotic variance $\{G_A/(n_A\gamma_A^2) + G_B/(n_B\gamma_B^2)\}$ is constant, the optimal target allocation proportion for treatment A is

$$\varrho_{A1}(\mu_A, \mu_B, \gamma_A, \gamma_B) = \frac{\gamma_B\sqrt{(D/\mu_B)^{\gamma_B}G_A}}{\gamma_B\sqrt{(D/\mu_B)^{\gamma_B}G_A} + \gamma_A\sqrt{(D/\mu_A)^{\gamma_A}G_B}}. \tag{7}$$

Consequently, the corresponding covariate-adjusted target allocation is expressed as

$$\pi_{A1}(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \gamma_A, \gamma_B, \mathbf{z}) = \frac{\gamma_B\sqrt{\{D/\mu_B(\mathbf{z})\}^{\gamma_B}G_A}}{\gamma_B\sqrt{\{D/\mu_B(\mathbf{z})\}^{\gamma_B}G_A} + \gamma_A\sqrt{\{D/\mu_A(\mathbf{z})\}^{\gamma_A}G_B}}. \tag{8}$$

An important feature of (8) is that, irrespective of the value of $\mathbf{z}$, the allocations are skewed towards the better treatment, yet the degree of skewing depends on the parameters of the Weibull distribution. The derivation of the proposed optimal target allocation proportion (8) is provided in Appendix C.

Other functions could also be similarly optimized to obtain different allocation proportions. Maximizing the total expected survival time subject to a constraint on the asymptotic variance of the estimated treatment difference fixed to a constant, similarly, leads to the covariate-adjusted optimal target allocation proportion given by

$$\pi_{A2}(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \gamma_A, \gamma_B, \mathbf{z}) = \frac{\gamma_B\sqrt{\mu_A(\mathbf{z})\Gamma(1+1/\gamma_A)G_A}}{\gamma_B\sqrt{\mu_A(\mathbf{z})\Gamma(1+1/\gamma_A)G_A} + \gamma_A\sqrt{\mu_B(\mathbf{z})\Gamma(1+1/\gamma_B)G_B}}. \tag{9}$$

This approach does not take into account the fact that the hazard of a Weibull distribution depends on survival times. It would be more appropriate when patients do not have a common follow-up time. Furthermore, minimizing overall trial size, subject to asymptotic variance of estimated treatment difference remaining fixed, produces the Neyman allocation function. This allocation proportion is directly proportional to the standard deviation of the log-transformed estimated scale parameter. Neyman allocation maximizes power of the Wald test for treatment comparisons, for a given sample of size n. The proposed optimal Neyman allocation proportion is

$$\pi_{A3}(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \gamma_A, \gamma_B, \mathbf{z}) = \frac{\gamma_B\sqrt{G_A}}{\gamma_A\sqrt{G_B} + \gamma_B\sqrt{G_A}}. \tag{10}$$

Biswas and Mandal (2004) proposed a procedure that results in an allocation which is a generalization of optimal allocation for normal responses. They generalized the binary optimal allocation for normal responses in terms of failures. Zhang and Rosenberger (2007) applied their approach for exponentially distributed survival times, to develop a response-adaptive randomization method without considering the covariate information of the patients. This approach can be further developed for the CARA randomization procedure when the survival times conform to a Weibull distribution. If the survival times are dichotomized based on some threshold constant c, that is, a survival time less than the threshold c is considered a failure, and otherwise a success, then the function $n_A[1-\exp\{-c/\mu_A(\mathbf{z})\}^{\gamma_A}] + n_B[1-\exp\{-c/\mu_B(\mathbf{z})\}^{\gamma_B}]$ can be minimized, subject to the constraint on the asymptotic variance $G_A/(n_A\gamma_A^2) + G_B/(n_B\gamma_B^2)$ fixed to a constant, to obtain the optimal target allocation proportion

$$\pi_{A4}(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \gamma_A, \gamma_B, \mathbf{z}) = \frac{\gamma_B\sqrt{G_A[1-\exp\{-c/\mu_B(\mathbf{z})\}^{\gamma_B}]}}{\gamma_B\sqrt{G_A[1-\exp\{-c/\mu_B(\mathbf{z})\}^{\gamma_B}]} + \gamma_A\sqrt{G_B[1-\exp\{-c/\mu_A(\mathbf{z})\}^{\gamma_A}]}}. \tag{11}$$

3. Theory and calculations

3.1. Randomization procedures

Various randomization procedures can be used to target a specific derived optimal target allocation proportion. Each of them consists of an allocation function, whose arguments asymptotically approach the derived allocation proportion.

Hu et al. (2009) proposed a family of randomization procedures that attain the Cramér–Rao lower bound on the allocation variances for any allocation proportion, however, the allocation function for this procedure is discrete and discontinuous. The asymptotic theory for response-adaptive designs that relies on a Taylor series expansion for the allocation functions is not applicable to non-differentiable cases. This family of efficient randomized-adaptive designs (ERADE) are discrete probability functions and can adapt to any desired allocation proportion and can be implemented in practice, for both discrete and continuous responses. Hu et al. (2009) overcame the difficulties of discontinuity by introducing a stopping time of a martingale process. However, the ERADE method ignores the fact that the patients in clinical trials are heterogeneous and therefore does not take into account the information related to the covariate profiles of the patients. To address this issue, a covariate-adjusted version of ERADE is introduced here, intuitively, following the

results of Zhang and Hu (2009) and Hu et al. (2009), to target the derived optimal target allocation proportions for a given covariate profile of the patients.

After the allocation of m patients to each of the two treatment arms and observing their responses, let $N_A(m)$ and $N_B(m) = m - N_A(m)$ denote the numbers of patients assigned to the two treatments. When the $(m+1)^{th}$ patient enters the clinical trial with covariate vector $\mathbf{z}_{m+1}$, let $\hat{\pi}_m = \hat{\pi}_A(\hat{\boldsymbol{\beta}}_{Am}, \hat{\boldsymbol{\beta}}_{Bm}, \hat{\gamma}_{Am}, \hat{\gamma}_{Bm}, \mathbf{z}_{m+1})$ represent the estimate of $\pi_A(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \gamma_A, \gamma_B, \mathbf{z})$ based on the responses from the m patients, adjusted for the covariate $\mathbf{z}_{m+1}$ of the incoming patient. Further, let $\hat{\rho}_m = \sum_{i=1}^{m} \hat{\pi}_A(\hat{\boldsymbol{\beta}}_{Am}, \hat{\boldsymbol{\beta}}_{Bm}, \hat{\gamma}_{Am}, \hat{\gamma}_{Bm}, \mathbf{z}_i)/m$ be an estimate of the average target allocation for treatment A based on the data for the first m patients. Then, for the covariate-adjusted ERADE (CAERADE), the $(m+1)^{th}$ patient is assigned to treatment A with probability

$$\phi_{m+1} = \begin{cases} \alpha\hat{\pi}_m & \text{if } \frac{N_A(m)}{m} > \hat{\rho}_m, \\ \hat{\pi}_m & \text{if } \frac{N_A(m)}{m} = \hat{\rho}_m, \\ 1 - \alpha(1-\hat{\pi}_m) & \text{if } \frac{N_A(m)}{m} < \hat{\rho}_m, \end{cases}$$

where $0 \leq \alpha < 1$ is a constant that reflects the degree of randomization. Hu et al. (2009) recommended a value of α between 0.4 and 0.7. This gives a family of CARA designs that are fully randomized and also asymptotically efficient as they attain the Cramér–Rao lower bound. As mentioned in Hu et al. (2009), one can show that the maximum likelihood estimates (MLEs) satisfy the Bahadur-type representation. Following Zhang et al. (2007) and Hu et al. (2009), it can be stated that if the response distribution conditional on the available covariates belongs to exponential family, CAERADE is efficient of the first order for any $\alpha \in [0, 1)$. When survival outcomes conform to Weibull distribution and the shape parameter is known, the distribution belongs to the exponential family. Therefore, in such cases, CAERADE also generates a first-order efficient allocation design. Following Zhang et al. (2007) and Hu et al. (2009), since the parameter estimates of the Weibull model is obtained using the maximum likelihood estimation method, they satisfy the Bahadur type representation. Following the assumption of Condition 3.1 of Zhang and Hu (2009) and the Condition A of Theorem 2.1 in Zhang et al. (2007), the CAERADE is expected to be asymptotically first-order efficient and its observed allocation proportion converges to the expected optimal target allocation proportion. The theoretical derivation of these asymptotic properties of CAERADE is beyond the scope of this study and requires further investigation. However, their performance targeting the expected proposed targets has been evaluated in Section 4, using extensive simulations.

When the allocation probability function is a continuous and differentiable function of $\hat{\rho}_m$, $\hat{\pi}_m$ and the current sample proportion, the asymptotic properties of adaptive designs are obtained using a Taylor series expansion. The expected sample proportions in such cases cannot be efficiently approximated by $\hat{\rho}_m$. Consequently, the variances of the allocation proportions do not attain the Cramér–Rao lower bound.

The covariate-adjusted doubly-adaptive biased coin design (CADBCD) procedure (Zhang and Hu, 2009) can also be used. For this design, the allocation probability is

$$\phi_{m+1} = \begin{cases} \frac{\hat{\pi}_m \left\{\frac{m\hat{\rho}_m}{N_A(m)}\right\}^{\alpha}}{\hat{\pi}_m \left\{\frac{m\hat{\rho}_m}{N_A(m)}\right\}^{\alpha} + (1-\hat{\pi}_m)\left\{\frac{m(1-\hat{\rho}_m)}{m-N_A(m)}\right\}^{\alpha}} & \text{if } 0 < \frac{N_A(m)}{m} < 1, \\ 1 - \frac{N_A(m)}{m} & \text{if } \frac{N_A(m)}{m} = 0 \text{ or } 1, \end{cases}$$

where $0 \leq \alpha' < \infty$ is a constant that controls the degree of randomization. A value of $\alpha' = 0$ corresponds to the procedure being most random and it is the most deterministic when $\alpha' \to \infty$. Following Theorem 3.1 of Zhang and Hu (2009), under mild conditions, $N_A(m)/m$ and $\hat{\pi}_m$ are strongly consistent and follow an asymptotic bivariate normal distribution, with the asymptotic means being the expected value of the target allocation proportion. Also, $\sqrt{m}(\hat{\boldsymbol{\beta}}_{jm} - \boldsymbol{\beta}_j)$ is asymptotically multivariate normal with zero mean vector. However, since the survival times conform to a Weibull distribution and there is no closed form for the maximum likelihood estimators of the scale and shape parameters, an explicit asymptotic variance for the CADBCD procedure cannot be obtained. When $\alpha' \to \infty$, the allocation function achieves the Cramér–Rao lower bound in terms of its asymptotic variance.

3.2. Asymptotic relationship between power and variance of CARA designs

It is essential to assess the performance of the target allocations and the CARA randomization procedures, to find out which randomization procedure, targeting a specific allocation proportion, outperforms the others. Hu and Rosenberger (2003) provided a theoretical template for the comparison of different response-adaptive randomization procedures and different target allocations in terms of power and expected failure rates, when the response of the patients to a treatment is binary. Zhang and Rosenberger (2006) further developed this idea for continuous responses following a normal distribution. However, there has rarely been any work on the asymptotic properties for comparing CARA designs. In this section, a theoretical template is provided for the comparison of different CARA designs and different target allocation proportions, when the survival responses conform to a Weibull distribution.

The optimal allocation proportions depend on the choice of measure of difference between the treatments. In this paper, the focus has been on the simple difference $\Delta = \log\{\mu_A(\mathbf{z})\} - \log\{\mu_B(\mathbf{z})\}$. The entire theory behind CARA designs relies on asymptotic approximation of the observed allocation proportion and the estimated target allocation proportion.

In real clinical trials, Wald tests are a popular choice for comparison across treatments. Wald test is used here to test $H_0 : \Delta = 0$, against $H_A : \Delta \neq 0$.

Using a Taylor series expansion of the non-centrality parameter of the usual chi-squared test, for testing the effect of the treatments, an explicit relationship can be derived between the target allocation proportion, the bias of the randomization procedure from the target and the variability induced by the randomization procedure for any allocation proportion. The Wald test statistic is given by

$$T_n = \frac{\log\{\hat{\mu}_A(\mathbf{z})\} - \log\{\hat{\mu}_B(\mathbf{z})\}}{\sqrt{\frac{G_A}{n_A\hat{\gamma}_A^2} + \frac{G_B}{n_B\hat{\gamma}_B^2}}}. \tag{12}$$

Under H_0, T_n has an asymptotic standard normal distribution.

For a design with fixed n_A and n_B, and independent Weibull survival outcomes, T_n^2 is asymptotically chi-squared with one degree of freedom. Under the alternative hypothesis, power can be expressed as an increasing function of the non-centrality parameter of the chi-squared distribution for a fixed target allocation proportion π on treatment A. Using the simple difference measure, the non-centrality parameter can be expressed as

$$\Lambda = \frac{[\log\{\mu_A(\mathbf{z})\} - \log\{\mu_B(\mathbf{z})\}]^2}{\frac{G_A}{n_A\gamma_A^2} + \frac{G_B}{n_B\gamma_B^2}},$$

which can be rewritten as

$$\frac{\Lambda(x)}{n} = \frac{[\log\{\mu_A(\mathbf{z})\} - \log\{\mu_B(\mathbf{z})\}]^2}{\frac{G_A}{(\pi+x)\gamma_A^2} + \frac{G_B}{(1-\pi-x)\gamma_B^2}},$$

where $x = n_A/\text{n} - \pi$. Expanding $\Lambda(x)/n$ in a Taylor series about π yields

$$\begin{aligned}\frac{\Lambda(n_A/n)}{n} &= \frac{[\log\{\mu_A(\mathbf{z})\} - \log\{\mu_B(\mathbf{z})\}]^2}{\frac{G_A}{\pi\gamma_A^2} + \frac{G_B}{(1-\pi)\gamma_B^2}} \\ &+[\log\{\mu_A(\mathbf{z})\} - \log\{\mu_B(\mathbf{z})\}]^2 \frac{(1-\pi)^2\frac{G_A}{\gamma_A^2} - \pi^2\frac{G_B}{\gamma_B^2}}{\{(1-\pi)\frac{G_A}{\gamma_A^2} + \pi\frac{G_B}{\gamma_B^2}\}^2}\left(\frac{n_A}{n} - \pi\right) \\ &-[\log\{\mu_A(\mathbf{z})\} - \log\{\mu_B(\mathbf{z})\}]^2 \frac{(\frac{G_A}{\gamma_A^2}\frac{G_B}{\gamma_B^2})}{\{(1-\pi)\frac{G_A}{\gamma_A^2} + \pi\frac{G_B}{\gamma_B^2}\}^3}\left(\frac{n_A}{n} - \pi\right)^2 + o\left\{\left(\frac{n_A}{n} - \pi\right)^2\right\}.\end{aligned}$$

The first term is determined by π and represents the non-centrality parameter for a fixed design. The Neyman allocation for patients with a given set of covariates $\mathbf{z}$ as in (10) maximizes this term. This term can be used to compare different target allocation proportions in terms of their powers. The second term represents the bias of the actual allocation from the target allocation. With the design shifting to different sides from the target allocation proportion π, the non-centrality parameter will increase or decrease according to the coefficient of $n_A/n - \pi$, and this coefficient equals zero if and only if $(1-\pi)^2 G_A/\gamma_A^2 - \pi^2 G_B/\gamma_B^2 = 0$, which yields the Neyman allocation given in (10).

In a real-life scenario, especially in the field of personalized medicine, scientists may be interested to know the proportion of patients on a particular treatment for a given set of covariates $\mathbf{z}$. CARA randomization procedures involve $N_{A|\mathbf{z}}(m)$, the number of patients with covariate $\mathbf{z}$ allocated to treatment A, after m allocations. Given a covariate $\mathbf{z}$, the proportion of patients allocated to treatment A is $N_{A|\mathbf{z}}(m)/N_{\mathbf{z}}(m)$, where $N_{\mathbf{z}}(m)$ is the total number of patients with covariate $\mathbf{z}$. Substituting n_A/n and n_B/n with $N_{A|\mathbf{z}}(m)/N_{\mathbf{z}}(m)$ and $N_{B|\mathbf{z}}(m)/N_{\mathbf{z}}(m)$ in (12), the test statistic T_n^2 still has an asymptotic chi-squared distribution. The justification for the asymptotic chi-squared distribution of the test statistic can be concluded from Hu and Zhang (2004). The critical condition to ensure the chi-squared limit, is that the covariate-adjusted allocation proportion on each treatment converges almost surely to a constant between 0 and 1 for the specific procedure. This substitution makes $\Lambda\{N_{A|\mathbf{z}}(m)/N_{\mathbf{z}}(m)\}$ a random non-centrality parameter, and therefore, its expectation can be considered. For example, since $N_{A|\mathbf{z}}(m)/N_{\mathbf{z}}(m)$ is asymptotically unbiased for π, $E\{N_{A|\mathbf{z}}(m)/N_{\mathbf{z}}(m) - \pi\} \to 0$. Therefore, the average power of the test directly relates to the variance $E[\{N_{A|\mathbf{z}}(m)/N_{\mathbf{z}}(m) - \pi\}^2]$ of the CARA procedure. It is this explicit relationship that would be mostly used for power comparisons of different CARA randomization procedures.

4. Results

4.1. Simulation study

4.1.1. Choice of design parameters

To compare the different randomization procedures, a survival trial with 400 patients is considered. Patients' arrival time is simulated from a Uniform (0, 365) distribution, so that the length of the recruitment period is $R = 365$ days. The

Table 1
Theoretical values of model parameters, for increasing and decreasing Hazard A vs B; theoretical values of population log hazard ratio (*HR*), expected target allocation proportions for CARA designs.

Hazard Direction	Models	Treatments	log(*HR*)	E(π_{A1})	E(π_{A2})	E(π_{A3})	E(π_{A4})
Decreasing Hazard A vs B	Neutral	A	0	0.5	0.5	0.5	0.5
		B		0.5	0.5	0.5	0.5
	Positive	A	−2.071	0.55	0.53	0.39	0.58
		B		0.45	0.47	0.61	0.42
	Negative	A	2.075	0.44	0.46	0.47	0.49
		B		0.56	0.54	0.53	0.51
Increasing Hazard A vs B	Neutral	A	0	0.5	0.5	0.5	0.5
		B		0.5	0.5	0.5	0.5
	Positive	A	−7.419	0.55	0.54	0.44	0.55
		B		0.45	0.46	0.56	0.44
	Negative	A	7.487	0.44	0.45	0.47	0.49
		B		0.56	0.55	0.53	0.51

response time of a patient is added to the recruitment time of the patient, and for patients whose outcomes have not been observed within a specified time, $D > R$, are right censored. The censoring scheme followed here is the generalized Type I right censoring. Therefore, for a fixed overall trial duration, the recruitment period has been considered to be sufficiently long to enable efficient parameter estimation. The overall trial duration is taken to be $D = 581.66$ days and the censoring time of a patient is simulated from a Uniform (0, 581.66) distribution. If the recruitment stage in the confirmatory trial is relatively long, and the amount of accumulating response data during the recruitment phase is substantial, then it is possible to facilitate CARA randomization.

A covariate structure of three independent covariates has been generated (Rosenberger et al., 2001). These are gender (Bernoulli, $p = 0.5$), age (Uniform(30, 75)) and cholesterol level (Normal (200, 400)). Treatment-covariate interactions are not considered at the simulation stage, as the covariate-adjusted treatment effects can be quite effective as compared to the individual treatment effects, even in the absence of such interaction terms. In real-life clinical trials, even though information might be available about known covariates or strata during the design stage, however, the treatment-covariate interaction might not be significant or such information might be unknown to the investigator. Even in such a situation, accounting for the known covariates would make the covariate-adjusted treatment effect estimate more precise which would also make the estimated target allocation proportions, a function of the estimated model parameters, to be more accurate and with reduced variability. An useful example to this is pointed out by (Gilson et al., 2017) highlighting that for STAMPEDE (Systemic Therapy in Advancing or Metastatic Prostate Cancer: Evaluation of Drug Efficacy) trial (Sydes et al., 2012), prostate cancers exhibit significant intra-tumoural genetic heterogeneity, which increases in advanced disease in response to multiple lines of therapy. This provides the rationale to evaluate precision medicine approaches earlier by incorporating biomarker selection, with the aim of achieving the greatest impact on patient outcome even without treatment-biomarker interaction information, as biomarker–treatment pairings may fail to translate to patient benefit due to an incomplete understanding of the biology of the biomarker or the interaction between the biomarker and the therapy during the randomization stage. Moreover, unnecessary additional interaction terms beat the purpose of parsimonious model which is ultimately essential in real-life clinical trials. To keep the interpretation of the treatment effects simple in the final clinical study report before submission for marketing application, the clinician or the investigator at times also ignore including interaction terms between the treatment and the covariates. Zhu (2015) have demonstrated that there is not much of a persistent difference between CARA designs with or without interactions.

The survival time of a patient with covariate vector $\mathbf{z} = (1, z_1, z_2, z_3)^T$ in treatment group j is simulated from the Weibull distribution with scale parameter $\mu_j(\mathbf{z})$ and shape parameter $\gamma_j = 2.07527$ when assessing the situation for a monotonically increasing hazard, and $\gamma_j = 0.57527$ when assessing for a monotonically decreasing hazard. Three choices of the treatment effects vector $\boldsymbol{\beta}$ are considered in the simulation study: (1.896, 0.810, 0.038, 0.001) for the neutral effect of each treatment; (5.504, 0.810, 0.038, 0.001) for the positive effect of treatment A; and (−1.171, 0.810, 0.038, 0.001) for the negative effect of treatment A. The values of simulation model parameters were chosen by trial and error, taking into account that there are 3 covariates with known distributions, Weibull distributed time to events, and targeting the approximate values of natural log transformed hazard ratio in the population as given in Table 1. The natural log transformed hazard ratio (HR) for a Weibull survival model represent the covariate-adjusted treatment effect based on the numerator of (12) multiplied by the negative shape parameter. Therefore this expression is given by,

$$\log(HR) = [\gamma_B \log\{\hat{\mu}_B(\mathbf{z})\} - \gamma_A \log\{\hat{\mu}_A(\mathbf{z})\}].$$

In clinical trials where clear understanding exists among the clinical trial team members about the biomarker–treatment interaction prior to the initiation of the trial, the values of hazard ratio will vary across individual patients depending on their covariate values. Such a trial has been discussed in Section 4.2. The theoretical expected target values for CARA designs targeting each of the derived optimal allocation proportions (8)–(11) based on the simulation model parameters are given in Table 1.

Table 2
List of the Competing Designs used in this Study.

Design	Competing Designs
I	Completely randomized design (CRD)
II	Efron's biased coin design with p = 2/3
III	Pocock and Simon design with p = 3/4
IV	CARA CADBCD with (8) as the target
V	CARA CADBCD with (9) as the target
VI	CARA CADBCD with (10) as the target
VII	CARA CADBCD with (11) as the target
VIII	CARA CAERADE with (8) as the target
IX	CARA CAERADE with (9) as the target
X	CARA CAERADE with (10) as the target
XI	CARA CAERADE with (11) as the target
XII	Response-adaptive DBCD with (7) as the target
XIII	Response-adaptive ERADE with (7) as the target

The neutral treatment effect refers to the hypothetical experimental scenario where treatments A and B are equally effective. The positive treatment effect refers to the hypothetical experimental scenario where treatment A is more effective than treatment B and vice-versa for the negative treatment effect. The procedures re-calculate the randomization probabilities after arrival of each patient and 5,000 simulations are performed.

In order to compare the different competing designs, two response-adaptive rules have also been considered for which the covariates are ignored at the design stage, but the final estimates of the treatment effects can be adjusted for all covariates. The competing randomization procedures are listed in Table 2 against their corresponding design numbers.

For the implementation of the CARA and the response adaptive designs presented in Table 2, initially $2m_0$ patients are equally allocated to the two treatments, using Efron's biased coin design. This initial stage is referred to as the "burn-in" stage of the design. At a given stage, the treatment imbalance is computed, and, with probability 2/3, an incoming patient is assigned to the underrepresented treatment, to reduce the overall imbalance. Here, m_0 is a positive number, and following recommendations from Sverdlov et al. (2013), $2m_0$ is chosen to be 110 for the CARA designs, and 80 for the response-adaptive (RA) designs, which are sufficiently large for accurate estimation of the model parameters. This is because for the response-adaptive designs, adaptive allocation begins early on in the trial, compared to CARA designs, because the RA designs estimate the main treatment effects only, whereas the CARA method involves complex nonlinear estimation of the full vector of treatment effects. If a large number of covariates are included in the model, the "burn-in" period may be very long, and the implementation of CARA randomization method may become meaningless, because recruitment may stop before treatment effects can be estimated. Unlike the designs based on the exponential regression model, the designs proposed in this article, require longer time to initialize the adaptation process, because of the extra shape parameter in the Weibull model. Following Sverdlov et al. (2013) the average proportion of the subjects whose data are considered to have been observed during the recruitment phase and are used for the purpose of design adaptation, ranges from 0.52 to 0.61 across different competing designs.

At a particular step, the Newton–Raphson method for fitting the Weibull regression model to the data, may not converge and MLEs may not be attainable. In that case, the treatment assignment for the patient is determined using Efron's biased coin design. After the model parameters are estimated from the initial stage of the design, a randomization probability is calculated after each new patient who arrives sequentially into the trial. This randomization probability can be based on any one of the derived allocation functions. A pseudo random number generator is then used to draw a random number between 0 and 1. If the derived randomization probability is greater than or equal to this generated random number, the patient is assigned to treatment A or else the patient is assigned to treatment B. At the end of the trial, the covariate-adjusted treatment effect estimator from the fitted Weibull regression model is used to perform the Wald test for significance of the covariate-adjusted treatment difference at the nominal two-sided level of significance, which is considered to be 0.05 in this study. The randomization procedures which are being compared with the derived CARA designs are as follows:

- A completely randomized design, for which every patient is randomized to treatment A or B with probability 0.5.
- Efron (1971)'s (1971) biased coin design, in which an incoming patient is allocated to the under-represented treatment arm with a probability of 2/3.
- Pocock and Simon (1975)'s (1975) covariate-adaptive randomization procedure: For its implementation, all covariates are required to be categorical. Therefore, the continuous covariate age has been dichotomized as age $<$ 53 years and age $\geq$ 53 years, whereas the covariate cholesterol level has been dichotomized to level $<$ 200 and level $\geq$ 200. For an incoming patient, the treatment imbalance is computed at each level of the patient's covariates and with probability 3/4 the patient is assigned to the treatment arm that reduces the overall covariate imbalance. If the imbalance is zero, the patient is equally likely to be assigned to treatment A or B.
- A response-adaptive rule with (7) as its target, for which the covariates are only ignored at the design stage, however, the final estimates of the treatment effects can be adjusted for all covariates. The response-adaptive rule is implemented by means of the doubly-adaptive biased coin design with $\alpha' = 2$ (Hu and Zhang, 2004).

Table 3
Performances of each of the competing designs for the neutral model.

Hazard	Design	$\frac{N_A}{n}$ (SE)	$N_{AM} - N_{BM}$	$N_{AF} - N_{BF}$	Event	Type I Error
Decreasing	I	0.50 (0.025)	100–100	100–100	325	0.05
	II	0.50 (0.003)	100–100	100–100	325	0.05
	III	0.50 (0.012)	100–100	100–100	325	0.06
	IV	0.50 (0.038)	100–100	100–100	325	0.03
	V	0.50 (0.033)	100–100	100–100	325	0.03
	VI	0.50 (0.027)	100–100	100–100	325	0.03
	VII	0.50 (0.025)	100–100	100–100	325	0.04
	VIII	0.50 (0.026)	100–100	100–100	325	0.03
	IX	0.50 (0.021)	100–100	100–100	325	0.04
	X	0.50 (0.016)	100–100	100–100	325	0.04
	XI	0.50 (0.006)	100–100	100–100	325	0.04
	XII	0.50 (0.037)	100–100	100–100	325	0.05
	XIII	0.50 (0.030)	100–100	100–100	325	0.05
Increasing	I	0.50 (0.025)	100–100	100–100	306	0.05
	II	0.50 (0.003)	100–100	100–100	306	0.05
	III	0.50 (0.012)	100–100	100–100	306	0.06
	IV	0.50 (0.038)	100–100	100–100	306	0.04
	V	0.50 (0.034)	100–100	100–100	306	0.04
	VI	0.50 (0.029)	100–100	100–100	306	0.05
	VII	0.50 (0.026)	100–100	100–100	306	0.05
	VIII	0.50 (0.033)	100–100	100–100	306	0.05
	IX	0.50 (0.028)	100–100	100–100	306	0.05
	X	0.50 (0.020)	100–100	100–100	306	0.04
	XI	0.50 (0.009)	100–100	100–100	306	0.05
	XII	0.50 (0.055)	100–100	100–100	306	0.05
	XIII	0.50 (0.049)	100–100	100–100	306	0.05

- A response-adaptive strategy with (7) as its target, which is implemented by means of the efficient randomized-adaptive design with $\alpha = 0.55$ (Hu et al., 2009).

For the DBCD and CADBCD in Table 1, following Zhang and Hu (2009), Hu and Zhang (2004), and Rosenberger and Hu (2004), the trade-off parameter for randomness is taken to be $\alpha' = 2$. Similarly, Hu et al. (2009) recommended that, for appropriate implementation of the ERADE, it is reasonable to choose α between 0.4 and 0.7. The value of α' determines the degree of randomness of the design. When α' is smaller, the ERADE is more deterministic and has a smaller variability. Here, α for CAERADE and ERADE is chosen to be 0.55.

For sensitivity analysis of the effect of α on the variability of a CAERADE, simulation with 5,000 runs was considered. Following Hu et al. (2009), a simulation study was conducted with different values of α, namely, $\alpha = 0.125, 0.25, 0.50, 0.67$, and 0.75. Samples with overall sizes of 50, 100 and 200 were considered. The simulated results for $\alpha = 0.125$ and $\alpha = 0.25$ were very similar to those for $\alpha = 0.50$, in terms of obtaining the target allocation proportion and its variability. However, the CAERADE with $\alpha = 0.75$ has slightly higher variability than the others.

A similar simulation study was also conducted, to assess the sensitivity of the randomization parameter α' on the CADBCD. With $\alpha' = 0$, the CADBCD becomes the design proposed by Zhang et al. (2007), which is the covariate-adjusted version of the adaptive randomized design proposed by Melfi et al. (2001). Such designs have the highest asymptotic variability, and consequently low power for treatment comparison. The CADBCD with $\alpha' > 0$ always has smaller asymptotic variance than the adaptive randomized design proposed by Zhang et al. (2007) . As $\alpha' \to \infty$, the CADBCD assigns the incoming patient to treatment A with probability 1 if $N_A(m)/m < \pi$, and to treatment B with probability 1 if $N_A(m)/m > \pi$. This is a deterministic procedure, except when $N_A(m)/m = \pi$. It turns out that the deterministic procedure has the smallest variability, that can be attained by any procedure targeting the optimal allocation proportion. Decreasing α' ensures greater randomization, but also increased variability. Simulation results for the sensitivity analysis of α' show that $\alpha' = 2$ is an optimum trade-off that yields almost the same results as $\alpha' = \infty$, but is slightly better than the adaptive randomized design of Zhang et al. (2007). Results have also been simulated for $\alpha' = 5$ and they were nearly identical to those for $\alpha' = 2$. Overall, for all scenarios considered here, the CADBCD with $\alpha' = 2$ works better than complete randomization, in terms of reducing the number of events in a clinical trial. This also is in line with the recommendations from Hu and Zhang (2004) and Rosenberger and Hu (2004) of using $\alpha' = 2$ in practice.

4.1.2. No difference in treatment effects

Let us consider the case when there is no difference between treatment effects. Table 3 presents operating characteristics of the designs in Table 2 under the neutral model.

The third column gives the observed allocation proportion of patients assigned to treatment A and its standard error. The fourth and fifth columns provide average numbers of patients, categorized by their gender, allocated to each of the two treatments. The sixth column shows the average total number of events in a trial. The final column presents the type

Table 4
Performances of the competing designs for the positive model.

Hazard	Design	$\frac{N_A}{n}$ (SE)	$N_{AM} - N_{BM}$	$N_{AF} - N_{BF}$	Event	Power
Decreasing	I	0.50 (0.025)	100–100	100–100	168	0.98
	II	0.50 (0.003)	100–100	100–100	168	0.99
	III	0.50 (0.012)	100–100	100–100	168	0.99
	IV	0.54 (0.035)	106–94	111–89	166	0.95
	V	0.53 (0.031)	104–97	106–93	166	0.94
	VI	0.39 (0.028)	77–123	80–120	168	0.97
	VII	0.57 (0.027)	113–87	113–87	165	0.96
	VIII	0.57 (0.029)	107–93	121–79	166	0.97
	IX	0.53 (0.028)	106–94	107–93	165	0.97
	X	0.38 (0.025)	74–126	80–120	168	0.98
	XI	0.58 (0.020)	116–84	116–84	166	0.98
	XII	0.55 (0.034)	110–90	110–90	165	0.96
	XIII	0.55 (0.029)	110–90	110–90	166	0.97
Increasing	I	0.50 (0.025)	100–100	100–100	194	0.99
	II	0.50 (0.003)	100–100	100–100	194	0.99
	III	0.50 (0.012)	100–100	100–100	194	0.99
	IV	0.55 (0.027)	111–91	108–90	192	0.96
	V	0.54 (0.027)	109–88	111–92	192	0.96
	VI	0.44 (0.030)	88–112	90–110	194	0.96
	VII	0.55 (0.028)	110–90	110–90	192	0.96
	VIII	0.55 (0.024)	110–90	110–90	192	0.99
	IX	0.53 (0.023)	105–95	107–93	192	0.98
	X	0.44 (0.024)	85–115	91–119	195	0.98
	XI	0.56 (0.020)	113–87	113–87	192	0.97
	XII	0.54 (0.038)	108–92	108–92	192	0.96
	XIII	0.53 (0.029)	105–95	106–94	192	0.97

I error rate of the Wald test which is performed using the covariate-adjusted treatment effect estimator from the fitted Weibull regression model after the completion of the randomization procedure for the total pre-specified sample size .

Table 3 shows that, under the neutral model, all the randomization designs result in an equal allocation of patients to treatments *A* and *B*. Therefore, in this scenario, the observed allocation proportions of all the competing designs converge to the expected derived target values given in Table 1, irrespective of the shape of the hazard function. Moreover, all the CAERADE and ERADE procedures are less variable than the corresponding CADBCD and DBCD ones. On average, all the competing designs result in an equal number of events. Therefore, when the difference between covariate-adjusted treatment effects is zero, CARA designs ethically perform as good as the traditional balanced or response-adaptive randomization procedures.

While maintaining the type I error rate of the Wald test, it can be seen that CRD, Efron's biased coin design and the response-adaptive designs achieve the nominal significance level as compared to the other competing procedures, when the value of the shape parameter is less than 1. When the value of the shape parameter is greater than 1, most of the competing designs have a type I error rate closer to the nominal value. The Pocock–Simon (minimization) design, however, in both cases gives a slightly inflated error rate. Most of the type I error rates for the CARA designs where the shape parameter is less than 1, are slightly conservative. This is improved for the case when the shape parameter is greater than 1. This behaviour was further explored by simulation for larger sample sizes and other values of the shape parameter. The empirical type I error rate for the CARA designs continued to be conservative when the shape parameter is less than 1. This might be because, the density function of the Weibull model is closer to normality when the hazard function is increasing compared to when the hazard function is decreasing. The standard errors of the Type I error rates are found to vary between 0.002 and 0.004. All of the standard errors of the average numbers of events in a trial are found to be 8 for an increasing hazard, and 9 for a decreasing hazard.

4.1.3. Differences in treatment effects

Table 4 presents the operating characteristics of the competing designs in Table 2, for the positive model.

Unlike the neutral model, most of the designs result in a skewed treatment allocation towards the better-performing treatment arm, which in this case is *A*. Evidently from Table 4, the CAERADE or the ERADE is less variable than the CADBCD or the DBCD, respectively. Almost all of the CARA designs result in a skewed allocation of patients to the better treatment arm, however, the degree of skewness varies between the different designs. It must be noted that irrespective of the shape of the hazard function, the proposed CARA designs converges to their corresponding theoretical expected target allocation proportion values given in Table 1. The slight deviation can be attributed to the rounding of the values. It can be observed that when the hazard rate is monotonically decreasing, the CAERADE design VIII converges to a value of 0.57, which is slightly higher than the theoretical expected target value of 0.55. This is because the CAERADE being a discrete allocation function, converges to the expected target values at a slower rate than the corresponding CADBCD, although its finite-sample variances are always small. However, all the values which the respective CARA designs converge to, are

Table 5
Performances of the competing designs in the case of the negative model.

Hazard	Design	$\frac{N_A}{n}$ (SE)	$N_{AM} - N_{BM}$	$N_{AF} - N_{BF}$	Event	Power
Decreasing	I	0.50 (0.025)	100–100	100–100	361	0.99
	II	0.50 (0.003)	100–100	100–100	361	0.99
	III	0.50 (0.012)	100–100	100–100	360	0.99
	IV	0.44 (0.034)	88–112	89–111	358	0.93
	V	0.46 (0.030)	90–109	93–108	359	0.94
	VI	0.47 (0.028)	94–106	95–105	360	0.95
	VII	0.49 (0.027)	96–104	97–103	359	0.95
	VIII	0.46 (0.025)	90–110	94–106	358	0.95
	IX	0.48 (0.018)	92–106	99–103	359	0.96
	X	0.48 (0.014)	94–106	99–101	360	0.96
	XI	0.49 (0.007)	98–102	98–102	358	0.97
	XII	0.45 (0.033)	89–111	89–111	359	0.93
	XIII	0.46 (0.024)	92–108	92–108	359	0.94
Increasing	I	0.50 (0.025)	100–100	100–100	351	0.99
	II	0.50 (0.003)	100–100	100–100	351	0.99
	III	0.50 (0.012)	100–100	100–100	351	0.99
	IV	0.44 (0.036)	87–113	88–112	348	0.96
	V	0.45 (0.031)	90–111	91–109	348	0.93
	VI	0.47 (0.029)	93–117	94–116	348	0.98
	VII	0.48 (0.026)	96–107	93–104	348	0.95
	VIII	0.46 (0.029)	89–111	94–106	348	0.97
	IX	0.47 (0.023)	90–110	96–114	348	0.96
	X	0.48 (0.016)	93–107	99–101	349	0.99
	XI	0.49 (0.009)	98–102	98–102	348	0.96
	XII	0.44 (0.034)	89–111	89–111	348	0.94
	XIII	0.46 (0.027)	92–108	91–109	348	0.94

within twice the standard error interval of the theoretical expected targets. It was observed that design VIII stabilizes to its expected target value of 0.55 with 10,000 simulation runs. This phenomenon is also seen for the CARA designs when the hazard is monotonically increasing. The powers of the balanced designs are the highest, compared to the other ones. However, they result in more events, and are therefore, ethically not as attractive as the CARA or the response-adaptive designs. It can be observed that, using the optimal adaptive designs results in slightly fewer events compared to the traditional balanced randomization designs. The standard error of these for all of the designs is 10.

The final column of Table 4 shows the power of the Wald test. It can be seen that using a CARA randomization procedure leads to more patients being allocated to the better treatment without compromising much on the power of the Wald test, compared to that of the traditional balanced designs. In both hazard scenarios, the response-adaptive designs have the most variable power, whereas, the balanced randomization designs have the least variable power. For decreasing hazard, the variability in power for design VIII is similar to that of the balanced randomization designs. In the case of increasing hazard, the variability of the power for design V is similar to that of the response-adaptive procedures. All of the standard errors for the power ranged from 0.001 to 0.004.

Unlike the case of exponential survival responses, when comparing CARA and response-adaptive designs with targets, the former perform slightly better for Weibull distributed responses with an increasing hazard rate. The CARA designs allocate more patients to the better treatment arm at each of the patient subgroup levels. It can also be seen that CARA designs targeting the Neyman allocation proportion result in more events compared to the other CARA designs. This is mainly because the Neyman allocation proportion does not account for any ethical criteria.

Sometimes, treatment *B* performs better than treatment *A*. The operating characteristics in Table 5 show the performance of the competing designs in such a situation.

It can be seen from Table 5, that all of the competing designs are fairly powerful for the negative model. Similar to the positive model, both the CARA and the response-adaptive designs result in slightly fewer events compared to the traditional balanced randomization procedures. Irrespective of the shape of the hazard function, the derived CARA designs converge to its expected target values. The CADBCD being a continuous allocation function converges to the expected target values faster than the CAERADE. The standard error of the average number of events for almost all of the designs for the increasing hazard scenario is 6, except for design I, whose standard error is 7. For the decreasing hazard scenario, almost all of the designs have a standard error of 7, for the average number of events, apart from designs III and XI whose standard error is 6. The negative model also shows that, irrespective of the shape of the hazard function, using a CARA randomization procedure results in a skewing of the treatment allocation probabilities in favour of the better treatment arm, while maintaining a high statistical power of the Wald test. It can be seen that the CAERADE or ERADE designs are substantially less variable than their CACBCD or DBCD counterparts.

Since all of the CARA designs result in skewing of the treatment allocation probabilities towards the better treatment, and achieve high statistical efficiency in estimating treatment effects in the presence of covariates, they can be considered to be suitable alternatives to the traditional balanced designs. Similar to the positive model, for both increasing and

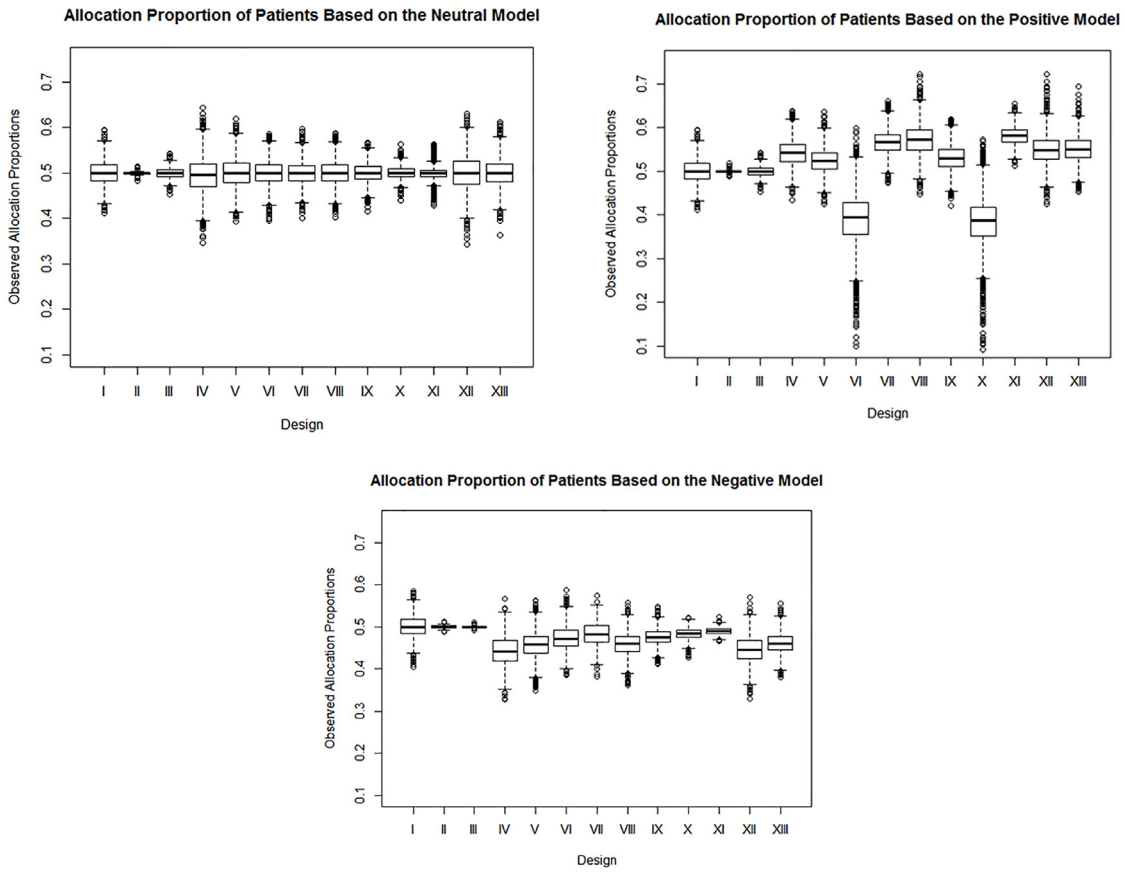

Fig. 1. Boxplots of the observed allocation proportions for the competing designs.

decreasing hazard scenarios, the response-adaptive designs have the most variability in power, whereas the balanced randomization designs have the least variable power. For decreasing hazard, the variability in power for design X, is similar to those of the balanced randomization designs. In the case of an increasing hazard, the variability in power for design IV, is similar to that of the response-adaptive procedures. The standard errors for the powers ranged from 0.001 to 0.004.

4.1.4. Distribution of the allocation proportions

In addition to evaluating the performances of the designs over 5,000 simulation runs, it is sometimes useful to assess the performance of the individual trials. The overall performance of the individual trials for the competing designs in Table 1, when the shape parameter of the Weibull model is less than 1, is illustrated in Fig. 1.

The boxplots (Fig. 1) depict the observed allocation proportions of the competing randomization procedures for $n = 400$ patients, sequentially arriving in the trials. The distributions of the observed allocation proportions appear to be very close to a symmetric distribution, but with different means and with different variances. For the response-adaptive procedures, the adaptive allocation starts early on in the trial as compared to the CARA designs, because the former estimate the main treatment effects only and the latter involve estimation of the full vector of treatment effects. For all three models, Efron's biased coin design, and the Pocock and Simon covariate-adaptive randomization procedure seem to be the least variable, among the competing designs. However, they along with the completely randomized design, allocate patients equally between the two treatment arms, irrespective of their performance based on patient responses. Therefore, these traditional balanced designs suffer from the disadvantage of allocating more patients to the worse treatment arm, during the course of the trial. In contrast, the CARA as well as the response-adaptive designs skew the patient allocation, on average, towards the better-performing treatment arm. The boxplots also confirm the finding, that the CAERADE and ERADE designs are more efficient than the corresponding CADBCD and DBCD procedures.

The simulation results in this section clearly suggest that, for Weibull distributed survival responses, the CARA designs would significantly skew the allocation probabilities away from balance, however, the degree and the direction of the skewness may vary depending on the expected target allocation proportions that the CARA designs converge to. The variabilities of the designs may also be different. It has been observed earlier that the CAERADE designs are the most

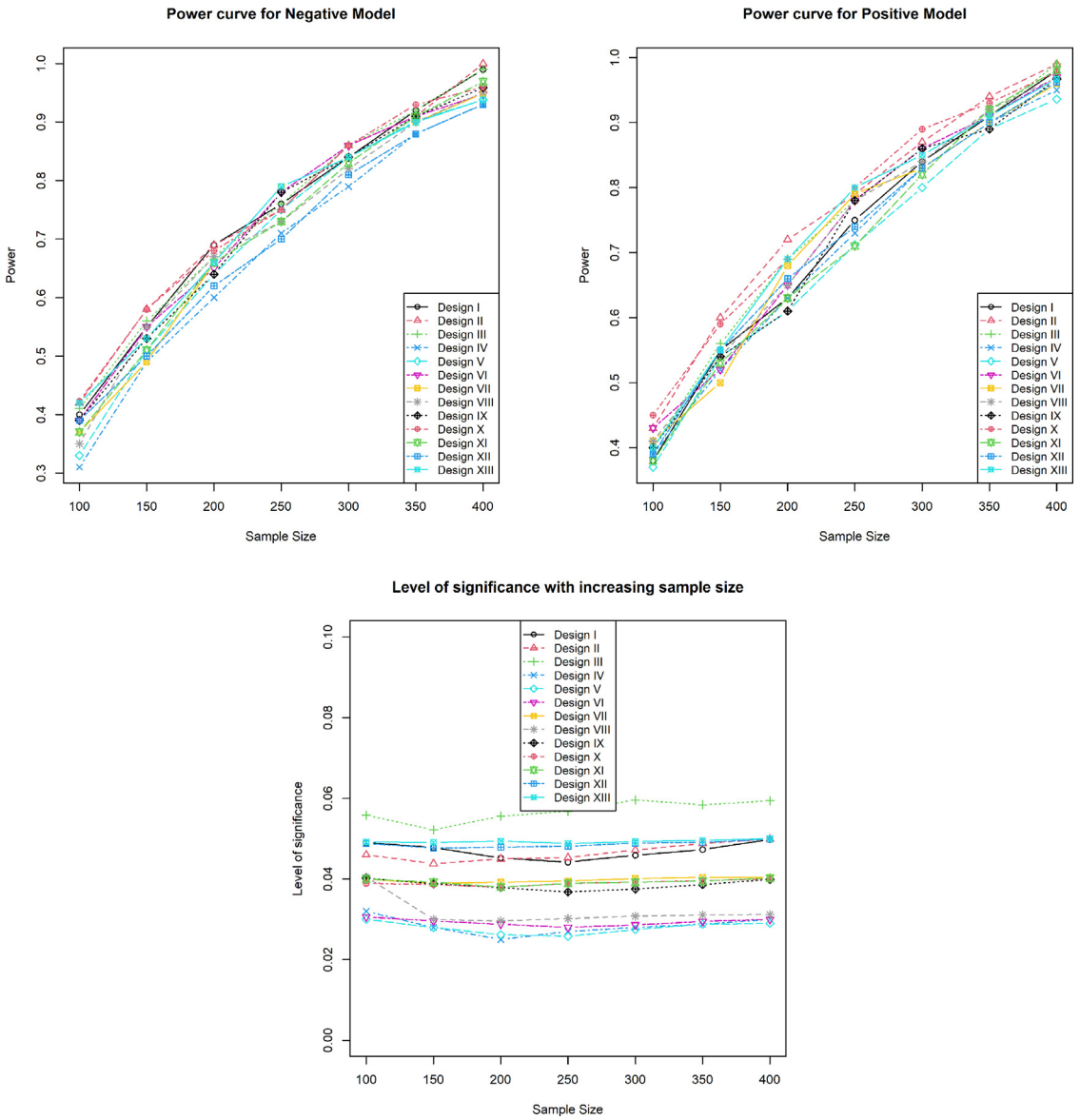

Fig. 2. Empirical level of significance and Power of the competing designs .

efficient among the CARA designs. Therefore, when balancing the competing goals of statistical efficiency and of treating more patients with the better-performing treatment, the CARA designs outperform the balanced randomization designs, provided such response data related to the primary endpoint are available during the recruitment stage, to enable adaptations in the design.

For the Wald test for treatment difference, the nominal level of significance is set to 0.05, two-sided. Evidently from Fig. 2, when the shape parameter of the Weibull model is less than 1, the minimization method gives slightly inflated Type I error rates for some sample sizes. However, the overall Type I error rates, are well controlled in the range 0.025 to 0.06, and that the power values are very similar across the competing designs. For statistical power of the Wald test, for covariate-adjusted treatment differences, to be greater than 0.9, most of the competing CARA designs are equally efficient as the traditional balanced designs and provides valid statistical inference for treatment comparison, when the shape parameter of the Weibull model is less than 1. This finding is also true, if the power for treatment difference is relaxed slightly by the clinical trial team to be greater than 0.8, which is often the case in confirmatory clinical trials, due to patient recruitment problems. It must be noted that the simulations in this study considered a phase 3 clinical trial with $n = 400$ patients, since, such a sample size is large enough to enable asymptotic properties of CARA procedures to hold, with power less than 1 (Fig. 2). When the hazard function is monotonically decreasing and the distribution of

Table 6
Recruitment Patterns.

Recruitment Pattern	Model Scenario	Positive		Negative			Neutral	
	Design	I	IV	I	IV		I	IV
Uniform	Power	0.979	0.951	0.989	0.931	Type I Error	0.051	0.03
	Events (Var)	168 (65.94)	166 (64.28)	361 (194.30)	358 (182.13)	Events (Var)	325 (116.28)	325 (112.38)
	$\hat{p}_R$	N/A	0.554	N/A	0.594		N/A	0.605
Beta(1,5)	Power	0.976	0.948	0.982	0.927	Type I Error	0.054	0.046
	Events (Var)	158 (66.34)	155 (66.20)	339 (192.22)	334 (194.64)	Events (Var)	325 (116.12)	325 (118.56)
	$\hat{p}_R$	N/A	0.299	N/A	0.306		N/A	0.317
Beta(5,1)	Power	0.979	0.949	0.987	0.930	Type I Error	0.045	0.026
	Events (Var)	180 (65.53)	174 (67.00)	369 (202.10)	363 (192.66)	Events (Var)	325 (114.23)	325 (115.7)
	$\hat{p}_R$	N/A	0.802	N/A	0.856		N/A	0.797
Beta(5,5)	Power	0.978	0.950	0.988	0.928	Type I Error	0.049	0.031
	Events (Var)	160 (68.08)	155 (65.91)	345 (197.09)	339 (189.6)	Events (Var)	325 (115.77)	325 (116.99)
	$\hat{p}_R$	N/A	0.532	N/A	0.522		N/A	0.541
Beta(1/5,1/5)	Power	0.981	0.952	0.991	0.931	Type I Error	0.048	0.039
	Events (Var)	174 (68.71)	169 (65.38)	364 (197.61)	360 (195.28)	Events (Var)	325 (114.07)	325 (115.48)
	$\hat{p}_R$	N/A	0.620	N/A	0.674		N/A	0.617

the time-to-event outcomes follow a Weibull model, the change in the power as the sample size increases from 100 to 400 patients is similar across all the competing designs. For Weibull model with shape parameter greater than 1 similar results were observed about the change in the performance, among the different designs as the sample size was varied.

It is known from Sverdlov et al. (2013) that, building the CARA designs assuming an exponential regression model for the survival responses, leads to the Type I error rates for the Wald test to be slightly inflated. However, the simulation results here, indicate that, for the designs in Section 2.3, if the responses of the patients follow a Weibull regression model, gives slightly conservative Type I error rates of the Wald test, when the hazard is decreasing. However, when the hazard rate is increasing, the empirical level of significance is closer to the nominal value of 0.05, for most of the optimal allocation proportions. This phenomenon may be due to the fact that, when the hazard rate is increasing, the density function for the Weibull model is closer to normality, rather than the exponential density function. Therefore, using the correct model for Weibull distributed survival responses, helps to control the Type I error rate of the Wald test, which is a primary concern in any clinical trial. The theoretical justification of this phenomenon deserve further investigation, however, they are beyond the scope of the current work.

4.1.5. Effect of varying recruitment patterns

The interplay between the recruitment period and the length of the trial, plays an important role in the performance of an adaptive method. Here, the trial recruitment period has been considered to be $R = 365$ days, and the trial duration to be $D = 581.66$ days. For the given fixed trial duration, the recruitment period here has been considered to be sufficiently long, to allow for the time to gather response data, sufficient to enable initiation of the adaptation process in the trial. Shorter recruitment period may lead to inefficient treatment effect estimation, due to the number of parameters in a Weibull AFT model. The trial duration is considered here from a practical point of view for an usual Phase 3 trial before regulatory submission. Longer trial duration is undesirable, since that would require more resources and delay the market launch of the drug in a confirmatory clinical trial, before regulatory submission of a New Drug Application (NDA). The proposed CARA randomization procedures rely on the assumption of uniform recruitment pattern. In practice, the accrual rate may be nonuniform and it is important to assess the impact of this phenomenon on the inferential properties of the designs. Table 6 compares operating characteristics of CRD (I) and CARA design IV, when the shape parameter of the Weibull model is less than 1, with delayed responses, under three experimental Model scenarios presented in Table 1.

Five recruitment patterns have been explored in this study: (i) Uniform; (ii) Beta(1, 5) (right skewed; accrual rate decreases over time); (iii) Beta(5, 1) (left skewed; accrual rate increases over time); (iv) Beta(1/5, 1/5) (symmetric; recruitment is accelerated at the onset and at the outset of the trial); (v) Beta(5, 5) (symmetric; recruitment peaks in the middle of the trial). Similar recruitment patterns were investigated by Sverdlov et al. (2013), for CARA designs based on exponential regression model. Here, they are explored further, for the proposed CARA designs based on Weibull model. The average proportion of patients $\hat{p}_R$ whose outcomes are observed during the recruitment phase and are used for the adaptation purposes of design IV, is noted to be different across five accrual patterns. However, the validity of the statistical inference, for treatment comparison for the proposed designs, across different recruitment patterns is the primary focus in this section.

Both the competing designs seem to have well controlled Type I error rates under the null hypothesis, and sufficiently high power under the alternative hypothesis, under each of the three experimental Model scenarios across five accrual patterns in Table 6. However, under the Positive and the Negative model scenarios, design IV results in modest average ethical gains, in terms of fewer number of events as compared with design I. Though the CARA design IV gives slightly conservative results under the null hypothesis, it leads to valid statistical inference for a confirmatory clinical trial and

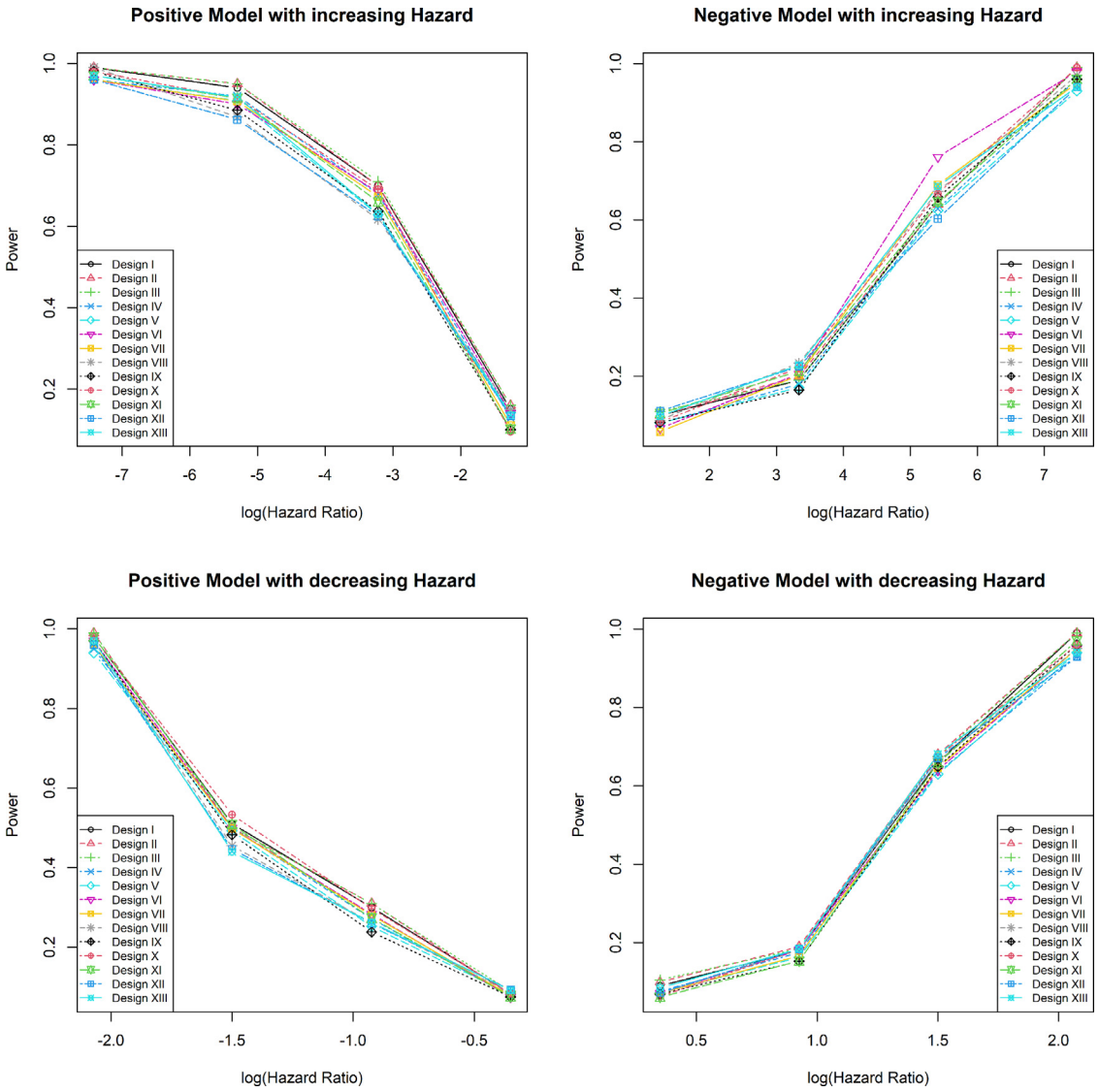

Fig. 3. Empirical level of Power With Change in Treatment Effects .

can therefore, be considered as suitable alternative to balanced randomization methods, when sufficient response data is observed during the recruitment period. Similar results were obtained for Weibull model with monotonically increasing hazard rate.

4.1.6. Effect of change in the treatment effects

The inferential properties of the proposed designs become more prominent, when the absolute benefit obtained from the experimental agent is varied. For a Phase III clinical trial with 400 patients, Fig. 3 depicts the change in the power of the competing designs given in Table 2, to identify a statistically superior treatment regimen, as the covariate-adjusted natural log transformed hazard ratios between the competing treatment arms are varied. As in Section 4.1.1, the $\log(HR)$ represents the difference between the covariate-adjusted treatment effects, based on the rate of change in the hazard function. As the survival responses conform to a Weibull model, simulation results are presented for both increasing and decreasing hazard rates.

In real-life survival trials, the hazard ratio or the $\log(HR)$ is often used to summarize the treatment effect benefit for the primary endpoint and is reported in the final clinical study report. Fig. 3 summarizes the behaviour of the competing designs of Table 2, for Weibull distributed time-to-event outcomes conditional on the covariate vector $\mathbf{z} = (1, z_1, z_2, z_3)^T$ as given in Section 4.1.1 considering the positive and negative models of Table 1. It considers the shape parameter of the

Weibull survival model, for treatment group j, as $\gamma_j = 2.07527$ when assessing the situation for a monotonic increasing hazard, and $\gamma_j = 0.57527$ when assessing for a monotonic decreasing hazard.

It is observed that all the competing designs in all the scenarios have very low power for detecting a superior treatment regimen, when the difference between the treatments is less pronounced. However, the power increases as this effect size, represented here by the log hazard ratio, increases between the two treatment arms. As we move further from the null hypothesis ($HR = 1 \rightarrow \log(HR) = 0$), regardless of the Negative or Positive Model, and regardless of the direction of hazard, power increases, indicating that the designs can detect larger departures from the null with higher accuracy. However, note that, the rate of increase of power differs across the different scenarios (Fig. 3) as the $\log(HR)$ is dependent on the shape of the hazard function.

This phenomenon follows from the derivation provided in Section 3.2, where the non-centrality parameter in a Wald test for a CARA design is used as an increasing function of the power of the Wald test. Therefore, the power of the designs increases with an increase in the non-centrality parameter. When the hazard rate is monotonically increasing over time and treatment A is performing better than treatment B, it can be clearly observed that sufficiently high power is achieved much earlier, for a smaller treatment difference as compared to all the other scenarios.

In this scenario, the power curves for all the competing designs increases steeply and a power close to 0.65 to 0.70 is achieved, with a covariate-adjusted log hazard rate of -3.219 between the two treatment arms. However, the rate of increase of the power is slightly slower across the other scenarios. The rate of increase of the power based on the effect size between the treatment arms is quite slow initially for the Negative model, until the difference between the covariate-adjusted log scale parameters between the competing arms is -1.607685 after which the power changes at a quicker rate. In this scenario, the competing designs achieve a power over 0.90, when the difference between the covariate-adjusted log scale parameters between the competing arms is -3.607685, irrespective of the shape of the hazard function. When the hazard rate is monotonically decreasing over time, the rate of increase of the power is the slowest. In this scenario, when the log hazard rate between the competing arms is -1.457, the power of the competing designs to detect a superior treatment regimen is about 0.50. The difference in the rate of change of power, across different scenarios, makes it critical for framing suitable assumption about the treatment effect, during the planning stage of the study; and for the sample size calculation in the trial protocol of a clinical study.

According to various regulatory guidelines such as the International Conference On Harmonisation Off Technical Requirements For Registration Of Pharmaceuticals For Human Use (ICH), the sample size calculation and the primary analysis method of a clinical trial must conform with the study design. Therefore, to design a confirmatory clinical trial using a CARA procedure, proper planning and inputs from the clinical team is required about the performance of the regimen under consideration. An effective inter-disciplinary collaboration within the clinical trial team is crucial to comprehend the behaviour of a drug under consideration. This would help in the use of sufficient clinical information about the treatment regimen, from earlier phases to ensure that practical considerations are a part of the statistical design and planning; and to decide the power requirement of the primary analysis, during the sample size calculations for the clinical study. A Phase Ib dose expansion study is usually performed at the initial stage of a clinical trial to identify efficacy, pharmacokinetic (PK) or pharmacodynamic (PD) signals in a treatment regimen. This is followed by a proof of concept (POC) study in a Phase IIa trial. Hence, at the onset of the confirmatory Phase III clinical trial, an investigator can use the information from the dose expansion and the POC study, to understand the behaviour of the new treatment, and utilize such information to plan for the relevant statistical trial design and use practical assumptions of the effect size during the power calculation, while preparing the clinical trial protocol for the Phase III study.

4.2. Re-designing a real clinical trial

In order to evaluate the performance of the derived methodologies, in a real-life clinical trial, a confirmatory survival trial from Mok et al. (2009) in pulmonary adenocarcinoma has been re-designed. The study has been first explored in the CARA randomization context by Zhu and Hu (2010) and in Sverdlov et al. (2013), in terms of redesigning the trial using CARA designs based on exponential models. Since the exponential model is a special category of the Weibull model, and the simulation of progression free survival (PFS) for sample size calculation, in real clinical trials, is often assumed to be from an exponential or a Weibull model, the study here is re-designed based on the derived methodologies for Weibull distributed survival responses. It must be noted that, as per guidelines of the International Conference On Harmonisation Off Technical Requirements For Registration Of Pharmaceuticals For Human Use (ICH) for real clinical trials, the primary analysis and the sample size calculation for any study must be coherent with the statistical design, being used for the trial. In this phase III open-label study, during a 20-month period, 1,217 adult patients, from East Asia between the ages of 18 and 50, were randomly assigned in a 1:1 ratio between gefitinib (treatment A) and paclitaxel (treatment B).

These patients were diagnosed with pulmonary adenocarcinoma, who are eligible for treatment with Tyrosine kinase inhibitors (TKI). The study excluded all patients who were former smokers, and also the ones suffering from long QT syndrome. The patients were followed up for a period of 12 months after the treatment phase. The primary objective of this trial was to test if treating patients with gefitinib would increase the time to progression from pulmonary adenocarcinoma or time to death, as compared to those patients who are being treated with paclitaxel. The primary endpoint was PFS which is defined as the time from randomization to disease progression or death due to any cause, to be the event of interest. Patients who do not experience the disease progression or death until the end of the study period, or if they

Table 7
Performances of the competing designs for the pulmonary adenocarcinoma trial.

Designs	N_A/n (SE)	$N_{A-} - N_{B-}$	$N_{A+} - N_{B+}$	Events	Power
I	0.50 (0.014)	243–243	365–365	1208	0.99
II	0.50 (0.001)	244–244	365–365	1208	0.99
IV	0.46 (0.019)	229–499	333–156	1196	0.98
VIII	0.45 (0.018)	241–489	305–182	1197	0.99
XII	0.46 (0.017)	335–358	229–295	1205	0.97
XIII	0.47 (0.016)	320–356	251–290	1205	0.98

are lost to follow-up during the study are considered to be censored. Time from randomization to the first documented disease progression per Response Evaluation Criteria in Solid Tumours version 1.1 (RECIST v1.1) or death due to any cause, whichever occurs first, was considered as the primary endpoint in the trial by Eisenhauer et al. (2009). The primary estimand here, following International Conference On Harmonisation Of Technical Requirements For Registration Of Pharmaceuticals For Human Use (ICH), is to compare the PFS as per RECIST v1.1 in adult patients with pulmonary adenocarcinoma, treated with Gefitinib versus adult patients treated with Paclitaxel.

Epidermal growth factor receptor (EGFR) gene mutation was considered to be one of the significant factors affecting patients' response. The study demonstrated that there was a significant interaction between the treatments and EGFR. Treatment A was superior to treatment B in the EGFR+ subgroup (hazard rate for progression 0.48), and inferior in the EGFR- subgroup (hazard rate for progression 2.85).

Patient survival times for treatment j were simulated from an exponential distribution with mean $\exp(\beta_{j0} + \beta_{j1}z)$ for $j = A, B$ and $z = 0,1$. Following Sverdlov et al. (2013), the parameters were chosen as follows: $R = 20$ months, $D = 26.5$ months; $z = 0$ (EGFR+) with probability 0.6, and $z = 1$ (EGFR-) with probability 0.4; $\beta_{A0} = 1.62$, $\beta_{A1} = 0.98$, $\beta_{B0} = 2.35$ and $\beta_{B1} = 0.80$. For appropriate implementation of the derived CARA designs, 315 patients were initially equally randomized to the two treatment arms using Efron (1971)'s (1971) biased coin design, before the adaptive randomization process was started. It must be noted that, unlike the scenarios considered in Section 4.1, the degree and direction of the covariate-adjusted treatment effect differ for patients in the EGFR subgroups within the competing treatment arms. Therefore, in this scenario, at the end of the trial, the covariate-adjusted interaction effect estimator from the fitted Weibull regression model is used to perform the asymptotic Wald test for significance of the covariate-adjusted interaction effect at the nominal two-sided 0.05 level of significance. The performances of the designs are presented in Table 7.

Table 7 summarizes the performances of some of the competing designs in Table 2. Clearly, the completely randomized design and Efron's biased coin design allocate patients equally between the two treatment arms, resulting in more events, irrespective of the performance of the treatment arms. They also do not take into account the difference between the covariate-adjusted treatment effects within the levels of EGFR. Hence, they allocate equal numbers of patients within both the EGFR subgroups. The average allocation proportions for both the CARA and the response-adaptive designs are similar. This is because of the different direction of the treatment effect in the EGFR+ and EGFR- subgroups. The overall allocation proportion is therefore close to 0.5, for both the CARA and the response-adaptive designs. However, the CARA designs IV and VIII, unlike the response-adaptive designs XII and XIII, account for the difference in the direction of the treatment effect in the EGFR subgroups.

One of the primary motivation for developing CARA designs is that, in some clinical trials (see, e.g., Karapetis et al., 2008; Behrendt and Gehan, 2009; Mok et al., 2009), the degree and direction of the treatment effect differ across patient subgroups, within a treatment arm, and the research design should ideally account for such covariate-specific treatment effects. CARA randomization is intended to achieve ethical and statistical objectives of the trial, without undermining validity of the trial results, and maintaining randomized nature of the experiment. Therefore, within each of the EGFR subgroups, the CARA procedure allocates more patients to the better treatment arm and has, on average, fewer events than each of the response-adaptive randomization procedures and the two balanced designs. The response-adaptive designs also result in a skewed allocation towards treatment B, but the degree of skewing is similar across the EGFR subgroups. It can be seen that design VIII is less variable than design IV. Hence, if the clinical trial team aims is to have an ethical design with minimum variability, the CAERADE design seems to be the preferred candidate. This is often the case while handling indications for rare disease in a pivotal study, or in studies developing a compound for personalized medicine which takes into account the efficacy of a drug, for the specific patient under consideration. The empirical Type I error rates for the designs vary between 0.03 and 0.05. All the competing randomization procedures considered, have similarly powerful Wald test for significance of the covariate-adjusted interaction effect from the fitted Weibull regression model, at the end of the trial. The standard errors of the average numbers of events were no more than 10. All the randomization procedures also yield consistent estimates for the model parameters.

5. Discussion

This article proposes CARA as a novel trial design, which is applicable beyond the constant hazard scenario, in survival trial. The CARA designs here, are based on the covariate-adjusted doubly-adaptive biased coin design (Zhang and Hu, 2009) and the covariate-adjusted version of the efficient randomized-adaptive design (Hu et al., 2009). The

operating characteristics of the proposed adaptive designs as well as the balanced randomization designs, have been compared using extensive simulations for a two-arm survival trial with three predictive covariates and right-censored data. Apparently,almost all of the proposed CARA designs generate skewed allocations towards the better treatment, according to covariate-specific treatment effects, and consequently result in fewer events in the trial without considerable compromise on the statistical efficiency compared to the balanced randomization designs. The only exceptions to this, are the CARA designs targeting the Neyman allocation proportion. This is because, the objective function for the Neyman allocation does not address any ethical criteria. Its objective is to minimize the overall trial size. The skewness in the treatment allocation proportions, in favour of the better treatment, establishes the ethical gain of using the CARA designs compared to the traditional balanced randomization procedures. Using the correct model for adaptation when the responses of the patients follow a Weibull model, helps to control the Type I error rate, which was slightly inflated when using the designs based on an exponential model.

An asymptotic relationship between the power of the Wald test and the variance of the CARA designs, has been derived using a Taylor series expansion of the non-centrality parameter of the Wald test. It has been illustrated that the variance of the CARA designs is inversely proportional to the power of the Wald test (Section 3.2). The CAERADE being the asymptotically most efficient CARA design, increases the power for treatment comparison as compared to the corresponding CADBCD. In situations where efficiency is critically important while addressing the ethical criterion of allocating more patients to the covariate specific superior treatment arms for a given patient, theoretically the CAERADE should be the best choice among all of the CARA randomization procedures. Examples of such a situation, are rare diseases and personalized medicine, where the number of patients is often too small. The proposed CARA designs converges to the derived optimal target allocation proportions, on an average. However, sometimes the CAERADE does not converge to the expected target allocation proportion as fast as the CADBCD, although its finite-sample variances are always small. This is primarily because, the allocation probabilities for the CAERADE are not stable. The allocation function being discrete, they always fluctuate. A continuous allocation function like the CADBCD can make the allocation probabilities stable, and speed up the convergence of the sample allocation proportions to the expected target values. Despite the marginal loss of statistical power of the Wald test, a CARA design would often outperform a traditional balanced randomization design in terms of ethical considerations, while achieving reasonably high efficiency in estimating covariate-adjusted treatment differences.

In an era, when the regulatory authorities advocate the innovation of novel trial designs with a patient centric approach, the CARA designs provide a suitable alternative to the traditional balanced designs considering how patients respond to specific treatment for their personalized covariate profile. The performance of CAERADE proposed here has been tested using extensive simulation work, targeting the derived allocation proportions. The theoretical derivation of the asymptotic properties of the proposed CAERADE require further investigation, and would be a suitable advancement to the work of Hu et al. (2009).

The simulation results in Sections 4.1 and 4.2 suggest that the type I error rate of the Wald test is reasonably controlled at the nominal two-sided 0.05 level of significance. This was not the case with the proposed CARA designs of Sverdlov et al. (2013) based on the exponential regression model. For real-life confirmatory clinical trials, it is common to use a one-sided 0.025 or a two-sided 0.05 level of significance as per regulatory guidance in U.S. Department of Health and Human Services Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER) (2022). Since simulated operating characteristics are not always sufficient in pivotal regulatory contexts, some theoretical justification for the control of type I error rate would be valuable. Theoretical proof of the controlled Type I error rates for different shape parameters of a Weibull model require further investigation, and are beyond the scope of the current work.

The methodology proposed in this article relies on the assumption of Weibull distributed survival responses. Sverdlov et al. (2013) suggests, empirically, that for cases where the distribution of the time-to-event outcomes is different from an exponential model, their designs lead to valid statistical inference, provided that the final data are analysed with the correctly specified accelerated failure time (AFT) model for the event times. Similar findings resonate for the CARA designs proposed here, where, even if the true distribution of the event times does not conform to a Weibull model, the designs proposed here can be safely used in practice, provided the final data analysis is performed using the correct AFT model, with the original covariate structure.

Though in real clinical trials an exponential or a Weibull model is often assumed for event times during simulation or sample size calculation, such recommendation of performing the final data analysis using the correct AFT model is rarely applicable in practice. This is because, a real pivotal trial hardly encounter situations where the analysis of the primary or the key secondary estimand is performed using an AFT model. If such an analysis is carried out in a study, it would invoke critical interrogation by the regulatory authorities. A semi-parametric approach or a non-parametric approach, such as a Cox proportional hazard model or a stratified log rank test respectively, is often used to analyse the key estimands in an oncology trial. Therefore, future direction of research requires development of designs free of distributional assumption of the time-to-event responses. Nevertheless, to preserve the robustness of the statistical inference for treatment comparison, it can be developed based on a lighter assumption of the hazard between the treatment arms, being proportional to each other.

For adaptive clinical trials, patient recruitment plays a critical role in determining the performance of the designs. In multi-centre trials, patient recruitment might be skewed because of their lack of availability. Various accrual rates

have been explored to ascertain the robustness of the inference from the proposed designs. Discernibly, even when the recruitment pattern is not uniform the proposed CARA designs provide valid statistical inference for treatment comparison (Table 6).

Though most randomized controlled trials, with a time-to-event outcome, are designed and analysed using the log-rank test and the Cox (1972) model, under the assumption of proportional hazards, patterns of delayed treatment effects have been observed recently across immuno-oncology trials. There could be multiple underlying causes for the delayed treatment effects, for example, the unique mechanism of action of the treatment, heterogeneous underlying population subgroups, and study design. Immuno-oncology trials can also encounter scenarios such as crossing hazards, and diminishing treatment effects over time. In such cases, there is heterogeneity in time about a successful treatment arm where there in a drift in the probability of success for a treatment arm, during the trial period. A possible future direction of research is to develop a suitable CARA design obviating any parametric distributional assumption of event times, but accounting for such drift during the process of the pivotal oncology trial.

The CARA designs proposed in this paper, have been developed for clinical trial comparing two treatment arms, assuming a Weibull model for event times. The suggested methods can be further extended for potential application of CARA randomization procedures, in time-to-event trials, with more than two treatment arms. Such adaptive randomization procedures are a frequent point of discussion among industry, academia and regulatory authorities in the Complex Innovative Trial Design (CID) pilot program. The CID pilot program is an initiative from the United States Food and Drug Administration (FDA) to support the goal of facilitating and advancing the use of complex adaptive, Bayesian, and other novel clinical trial designs. However, in practice, multi-arms treatment is more prevalent among real-life clinical trials. The Multi-Arm Multi-Stage (MAMS) trial designs have seen frequent implementation of response-adaptive randomization methods, with the option to drop futile arm or add an effective arm, during the process of the clinical trial considering the patient specific response. Amgen Inc. has recently participated in the FDA Complex Innovative Trial Designs (CID) Pilot Program implementing response-adaptive randomization method in a randomized, double-blind, Phase 2 study in patients with systemic lupus erythematosus (SLE), a rare disease with a high unmet need, in U.S. Department of Health and Human Services Food and Drug Administration (FDA) (2022). However, since the implementation of the response-adaptive randomization method did not consider the superior treatment arm for a given patient's covariate profile, it can impact the estimand being targeted in the final analysis. Biswas and Bhattacharya (2018) developed CARA designs considering binary treatment outcomes in a phase III clinical trial set up, involving multiple treatment arms. Optimal target allocations are available for ($j \geq 2$) treatment arm survival trials, with independent right-censored exponential survival outcomes (Sverdlov et al., 2011), and one can use covariate-adjusted versions of suggested optimal allocation. However, to our knowledge, such work for Weibull or other parametric time-to-event outcomes, have not been explored. Further research exploring such designs for multi-arms trials, assuming a more flexible parametric model (e.g., Weibull model) or devoid of any distributional assumption, for event times need to be developed, which will be beneficial for the CID pilot program of the FDA.

The CARA randomization designs merit investigation in clinical trials where treatment failures are costly, ethical issues are at stake and the primary outcomes follow heteroscedastic or nonlinear models. Sverdlov et al. (2013) considered only those scenarios in their simulation where the degree and direction of treatment effect differ for patient subgroups within a treatment arm. Although this is one of the most popular scenario for applications of CARA designs, however, the methodology is also applied in scenarios when the treatment-covariate interaction is absent. A regression model to predict the treatment effect would provide a more precise estimate of the covariate-adjusted treatment effect as compared to a response-adaptive design not considering significant covariates in the regression model. The design proposed in this article, are applicable irrespective of the presence or absence of treatment-covariate interaction. Therefore, to illustrate the performance of the proposed designs, under both scenarios, the treatment-covariate interactions have not been considered in the simulation work of Section 4.1, however, they have been considered while re-designing the real clinical trial in Section 4.2. Results from both the sections depict the suitability and patient-centric approach of CARA designs in confirmatory clinical trials as an alternative to the traditional balanced designs.

Sverdlov et al. (2013) in their simulation study did not consider restricted randomization as a competing procedure to their proposed CARA methods. Restricted randomization rules such as the permuted block design or the Efron (1971)'s (1971) design are quite popular in real clinical trials. Design proposed by Efron (1971) is also a special case of the ERADE design suggested by Hu et al. (2009) and its asymptotic variance has been shown to converge to a constant. Comparative assessments of the proposed designs, in this article, with restricted randomization method such as Efron (1971)'s (1971) design, have demonstrated its comparability with such randomization rule targeting to achieve balance.

The simulations to investigate the performance of the proposed designs assume an uniformly distributed recruitment pattern. In this study, the validity of the statistical inference about the treatment effect difference, between the competing arms, has been investigated under varying recruitment patterns. The proposed designs provide robust and valid statistical inference, even when recruitment patterns are quite different from an uniform distribution (Table 6). The proposed methodology is only applicable for confirmatory survival trials with long recruitment periods, where most patients contribute towards outcome data, during the recruitment phase. Due to delayed responses, parameter estimates for CARA designs may be obtained late, and the intended skewness in allocation proportions, may not be achieved. The number of covariates in the model impacts sequential estimation at the design stage. Unmeasured covariates at the design stage can be considered during the analysis. One should consider implementing CARA randomization only with a limited number of the most predictive baseline covariates that are well-known to the investigators, at the onset of the trial and it should be well documented in the clinical trial protocol and the statistical analysis plan.

6. Conclusion

For high statistical power and accuracy in estimating treatment effects, traditional theoretical literature and clinical trial protocols incline towards on balanced randomized designs, for comparing two or more treatments. However, besides achieving statistical power, clinical trials also emphasize on ethics, since they involve human subjects, and the ultimate objective is to benefit as many patients as possible, during the process of the trial. Adaptive randomization rules considering responses of the patients, skew allocation of patients to the better-performing treatment arm during a clinical trial, therefore, balancing ethics of the trial along with statistical efficiency. CARA designs are a practical improvement over traditional balanced randomized designs, as they provide a balance between ethics and statistical efficiency, by incorporating information about covariate profiles of patients, while at the same time maintaining high statistical power. This is a patient-centric approach while maintaining randomness in treatment allocation. Additionally, it aims to improve the rate of patient recruitment in the trial, by addressing the ethical criteria of an adaptive patient allocation perspective of the regulatory authorities.

According to International Conference On Harmonisation Off Technical Requirements For Registration Of Pharmaceuticals For Human Use (ICH) guidelines for statistical principles in clinical trials, for every clinical trial with a marketing application, all important details of its design and conduct, and principal features of its proposed statistical analysis need to be clearly specified in a protocol written and approved before the trial begins. For implementation of the proposed CARA designs in practice, clear indication needs to be laid out in the trial protocol and the statistical analysis plan about the sample size for the initial interim stage, the randomization method to be followed during this initial stage, randomization method to be used after the initial interim stage, optimal target allocation proportion, response model considered for event times, cohort size for sequential analysis, timing of interim analyses, baseline covariates under consideration.

Earlier CARA designs for time-to-event outcome have been developed for exponential regression models, which is although an useful starting point for developing CARA randomization procedures, however is not much practically applicable in real clinical trials. This article goes beyond exponential models, and proposes a more robust CARA design, by considering censored Weibull time-to-event response, making it more relevant for practical applications for increasing and decreasing hazards. The designs proposed here are based on CADBCD and the CAERADE designs. This design is applicable for models with and without interactions. The proposed designs specially if enhanced for multi-arm trials can also fit into the regulatory interest of CID pilot program. Considering the points mentioned in this article, the proposed allocation rules provide a suitable alternative to the traditional balanced rules which are quite popular in clinical trials.

Declarations

This work is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically, without the written consent of the copyright-holder.

CRediT authorship contribution statement

Ayon Mukherjee: Conceptualization, Methodology, Software, Writing – original draft writing. **D. Stephen Coad:** Reviewing, Validation of results, Collaboration. **Sayantee Jana:** Writing, Reviewing, Simulations and editing, Collaboration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors wish to specially thank Professor A.C. Atkinson, Dr. J. Constant, Professor W.F. Rosenberger and Dr. O. Sverdlov for their consultative inputs and constructive review comments. The authors specially thank the department of Regulatory Affairs and Drug Development Solutions (RADDS) within IQVIA for providing the resources for completing this research work. The authors also thank the Chief Editor and the anonymous referees for the careful review and constructive comments that have significantly improved an earlier version of the article.

Funding

This research was supported by the Queen Mary Principal's Research Fund, UK. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The analyses, conclusions, opinions and statements expressed herein are those of the author.

Appendix A. Deriving the variances of the parameter estimates

Consider two random samples (x_{ij}, δ_{ij}) for $j = A, B$ from Weibull distributions. Let $\hat{\mu}_j(\mathbf{z})$ and $\hat{\gamma}_j$ be the MLEs of $\mu_j(\mathbf{z})$ and γ_j. Then the Fisher information matrix is given by

$$I[\log\{\mu_j(\mathbf{z})\}, 1/\gamma_j] = \begin{bmatrix} n_j\epsilon_j(\boldsymbol{\beta}_j, \gamma_j, \mathbf{z})\gamma_j^2 & n_jE(\varsigma_j \exp\{\varsigma_j\})\gamma_j^2 \\ n_jE(\varsigma_j \exp\{\varsigma_j\})\gamma_j^2 & n_j\epsilon_j(\boldsymbol{\beta}_j, \gamma_j, \mathbf{z})\gamma_j^2 + n_jE(\varsigma_j^2 \exp\{\varsigma_j\})\gamma_j^2 \end{bmatrix},$$

where $\varsigma_j = \gamma_j[y_j - \log\{\mu_j(\mathbf{z})\}]$, $\epsilon_j(\boldsymbol{\beta}_j, \gamma_j, \mathbf{z})$ is the probability of an event and n_j is the number of patients on treatment j. The determinant of the above matrix is given by $n_j^2\gamma_j^4\{\epsilon_j^2(\boldsymbol{\beta}_j, \gamma_j, \mathbf{z}) + \epsilon_j(\boldsymbol{\beta}_j, \gamma_j, \mathbf{z})E(\varsigma_j^2 \exp\{\varsigma_j\}) - E(\varsigma_j \exp\{\varsigma_j\})^2\}$. Hence, the approximate covariance matrix of $\log\{\hat{\mu}_j(\mathbf{z})\}$ and $1/\hat{\gamma}_j$ is $M = I^{-1}[\log\{\mu_j(\mathbf{z})\}, 1/\gamma_j]$, which can be written as

$$M = \frac{1}{n_j\gamma_j^2}\begin{bmatrix} G_j & -\frac{E(\varsigma_j e^{\varsigma_j})}{\epsilon_j^2(\boldsymbol{\beta}_j,\gamma_j,\mathbf{z})+\epsilon_j(\boldsymbol{\beta}_j,\gamma_j,\mathbf{z})E(\varsigma_j^2 e^{\varsigma_j})-E(\varsigma_j e^{\varsigma_j})^2} \\ -\frac{E(\varsigma_j e^{\varsigma_j})}{\epsilon_j^2(\boldsymbol{\beta}_j,\gamma_j,\mathbf{z})+\epsilon_j(\boldsymbol{\beta}_j,\gamma_j,\mathbf{z})E(\varsigma_j^2 e^{\varsigma_j})-E(\varsigma_j e^{\varsigma_j})^2} & \frac{\epsilon_j(\boldsymbol{\beta}_j,\gamma_j,\mathbf{z})}{\epsilon_j^2(\boldsymbol{\beta}_j,\gamma_j,\mathbf{z})+\epsilon_j(\mathbf{z};\boldsymbol{\beta}_j,\gamma_j)E(\varsigma_j^2 e^{\varsigma_j})-E(\varsigma_j e^{\varsigma_j})^2} \end{bmatrix},$$

where (6) gives the mathematical form of G_j. Therefore,

$$\sigma_j^2 = \text{var}[\log\{\hat{\mu}_j(\mathbf{z})\}] = \frac{G_j}{n_j\gamma_j^2}$$

and

$$\text{var}(1/\hat{\gamma}_j) = \frac{\sigma_j^2\epsilon_j(\boldsymbol{\beta}_j, \gamma_j, \mathbf{z})}{\epsilon_j(\boldsymbol{\beta}_j, \gamma_j, \mathbf{z}) + E(\varsigma_j^2 \exp\varsigma_j)}.$$

Appendix B. Deriving the analytical form for $\epsilon_j(\boldsymbol{\beta}_j, \boldsymbol{\gamma}_j, \mathbf{z})$

It is known that $\epsilon_j(\boldsymbol{\beta}_j, \gamma_j, \mathbf{z}) = \text{P}(T_{ij} \le C_{ij}|\boldsymbol{\beta}_j, \gamma_j, \mathbf{z})$. The survival outcomes are assumed to conform to a Weibull distribution with scale parameter $\mu_j(\mathbf{z})$ and shape parameter γ_j, and the censoring times C_{ij} are assumed to follow a Uniform $(0, D)$ distribution, where D is the trial duration. Let $\text{A} = \{(t_{ij}, c_{ij}) : t_{ij} \le c_{ij}\}$. Then $\epsilon_j(\boldsymbol{\beta}_j, \gamma_j, \mathbf{z}) = \text{P}\{(T_{ij}, C_{ij}) \in \text{A}\}$ or

$$\epsilon_j(\boldsymbol{\beta}_j, \gamma_j, \mathbf{z}) = \int_0^D \int_0^{c_{ij}} f(t_{ij}, c_{ij})dt_{ij}dc_{ij}.$$

Substituting the joint density function of the event times and the censoring times, we get

$$\epsilon_j(\boldsymbol{\beta}_j, \gamma_j, \mathbf{z}) = \int_0^D \frac{1}{D}\int_0^{c_{ij}} \frac{\gamma_j}{\mu_j(\mathbf{z})}\{t_{ij}/\mu_j(\mathbf{z})\}^{\gamma_j-1}\exp\{-t_{ij}/\mu_j(\mathbf{z})\}^{\gamma_j}dt_{ij}dc_{ij}.$$

Performing a change of coordinates in the set A,

$$\epsilon_j(\boldsymbol{\beta}_j, \gamma_j, \mathbf{z}) = \int_0^D \frac{1}{D}\int_{t_{ij}}^D \frac{\gamma_j}{\mu_j(z)}\{t_{ij}/\mu_j(\mathbf{z})\}^{\gamma_j-1}\exp\{-t_{ij}/\mu_j(\mathbf{z})\}^{\gamma_j}dc_{ij}dt_{ij}.$$

Integrating with respect to c_{ij}, we get

$$\begin{aligned}\epsilon_j(\boldsymbol{\beta}_j, \gamma_j, \mathbf{z}) &= \int_0^D f(t_{ij})dt_{ij} - \frac{1}{D}\int_0^D t_{ij}f(t_{ij})dt_{ij} \\ &= 1 - S_j(D|\mathbf{z}; \boldsymbol{\beta}_j, \gamma_j) - \frac{1}{D}E[T_{ij}1_{\{T_{ij}\le D\}}],\end{aligned}$$

where $1_{\{T_{ij}\le D\}}$ denotes the indicator function of the event $\{T_{ij} \le D\}$.

Appendix C. Deriving the optimal allocation proportion for minimizing the total cumulative hazard function

The total cumulative hazard at time $D > 0$ can be minimized subject to the asymptotic variance for the covariate adjusted treatment difference remaining fixed to a constant. This yields the optimal allocation proportion of (8) as follows:

$$\text{min: } n_A\left\{\frac{D^{\gamma_A}}{\mu_A^{\gamma_A}(\mathbf{z})}\right\} + n_B\left\{\frac{D^{\gamma_B}}{\mu_B^{\gamma_B}(\mathbf{z})}\right\}$$

$$\text{subject to : } \left\{\frac{G_A}{n_A\gamma_A^2}\right\} + \left\{\frac{G_B}{n_B\gamma_B^2}\right\} = \mathrm{k} > 0$$

Re-arranging the constraint we get

$$\mathrm{k} - \left\{\frac{G_B}{n_B\gamma_B^2}\right\} = \left\{\frac{G_A}{n_A\gamma_A^2}\right\},$$

Solving for n_A,

$$n_A = \frac{G_A n_B \gamma_B^2}{\mathrm{k} n_B \gamma_B^2 \gamma_A^2 - G_B \gamma_A^2}.$$

The ethical objective here is to minimize,

$$n_A\left\{\frac{D^{\gamma_A}}{\mu_A^{\gamma_A}(\mathbf{z})}\right\} + n_B\left\{\frac{D^{\gamma_B}}{\mu_B^{\gamma_B}(\mathbf{z})}\right\}.$$

However,

$$\frac{G_A n_B \gamma_B^2}{\mathrm{k} n_B \gamma_B^2 \gamma_A^2 - G_B \gamma_A^2}\left\{\frac{D^{\gamma_A}}{\mu_A^{\gamma_A}(\mathbf{z})}\right\} + n_B\left\{\frac{D^{\gamma_B}}{\mu_B^{\gamma_B}(\mathbf{z})}\right\}$$

can be minimized in order to achieve the ethical objective.

To achieve the minimum value of the objective function

$$\frac{\partial}{\partial n_B}\left[\frac{G_A n_B \gamma_B^2}{\mathrm{k} n_B \gamma_B^2 \gamma_A^2 - G_B \gamma_A^2}\left\{\frac{D^{\gamma_A}}{\mu_A^{\gamma_A}(\mathbf{z})}\right\} + n_B\left\{\frac{D^{\gamma_B}}{\mu_B^{\gamma_B}(\mathbf{z})}\right\}\right] = 0.$$

Differentiating with respect to n_B we get

$$\left\{\frac{D^{\gamma_B}}{\mu_B^{\gamma_B}(\mathbf{z})}\right\} + \left\{\frac{D^{\gamma_A}}{\mu_A^{\gamma_A}(\mathbf{z})}\right\}\frac{-G_A G_B \gamma_A^2 \gamma_B^2}{(\mathrm{k} n_B \gamma_B^2 \gamma_A^2 - G_B \gamma_A^2)^2} = 0$$

$$\Rightarrow \left\{\frac{D^{\gamma_A}}{\mu_A^{\gamma_A}(\mathbf{z})}\right\} G_A G_B \gamma_A^2 \gamma_B^2 = (\mathrm{k} n_B \gamma_B^2 \gamma_A^2 - G_B \gamma_A^2)^2 \left\{\frac{D^{\gamma_B}}{\mu_B^{\gamma_B}(\mathbf{z})}\right\}.$$

Taking the positive square-root on both sides we get

$$\sqrt{\left\{\frac{D^{\gamma_A}}{\mu_A^{\gamma_A}(\mathbf{z})}\right\} G_A G_B \gamma_A^2 \gamma_B^2} = (\mathrm{k} n_B \gamma_B^2 \gamma_A^2 - G_B \gamma_A^2)\sqrt{\left\{\frac{D^{\gamma_B}}{\mu_B^{\gamma_B}(\mathbf{z})}\right\}}.$$

Substituting for k we get,

$$\sqrt{\left\{\frac{D^{\gamma_A}}{\mu_A^{\gamma_A}(\mathbf{z})}\right\} G_A G_B \gamma_A^2 \gamma_B^2} = \left\{\frac{G_A n_B \gamma_B^2}{n_A} - \frac{G_B n_B \gamma_A^2}{n_B} - G_B \gamma_A^2\right\}\sqrt{\left\{\frac{D^{\gamma_B}}{\mu_B^{\gamma_B}(\mathbf{z})}\right\}}.$$

$$= \frac{n_B}{n_A} G_A \gamma_B^2 \sqrt{\left\{\frac{D^{\gamma_B}}{\mu_B^{\gamma_B}(\mathbf{z})}\right\}}.$$

Solving for $\frac{n_B}{n_A}$ we get

$$\frac{n_B}{n_A} = \frac{\sqrt{\left\{\frac{D^{\gamma_A}}{\mu_A^{\gamma_A}(\mathbf{z})}\right\} G_A G_B \gamma_A^2 \gamma_B^2}}{\sqrt{\left\{\frac{D^{\gamma_B}}{\mu_B^{\gamma_B}(\mathbf{z})}\right\}} G_A \gamma_B^2}.$$

Now,

$$\pi_{A1}(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \gamma_A, \gamma_B, \mathbf{z}) = \frac{n_A}{n_A + n_B}.$$

$$= \frac{n_A/n_B}{1 + n_A/n_B}.$$

Substituting for $\frac{n_A}{n_B}$ we get

$$\pi_{A1}(\boldsymbol{\beta}_A, \boldsymbol{\beta}_B, \gamma_A, \gamma_B, \mathbf{z}) = \frac{\sqrt{\left\{\frac{D^{\gamma_B}}{\mu_B^{\gamma_B}(\mathbf{z})}\right\}\frac{G_A}{\gamma_A^2}}}{\sqrt{\left\{\frac{D^{\gamma_B}}{\mu_B^{\gamma_B}(\mathbf{z})}\right\}\frac{G_A}{\gamma_A^2}} + \sqrt{\left\{\frac{D^{\gamma_A}}{\mu_A^{\gamma_A}(\mathbf{z})}\right\}\frac{G_B}{\gamma_B^2}}}.$$
$$= \frac{\gamma_B\sqrt{\{D/\mu_B(\mathbf{z})\}^{\gamma_B}G_A}}{\gamma_B\sqrt{\{D/\mu_B(\mathbf{z})\}^{\gamma_B}G_A} + \gamma_A\sqrt{\{D/\mu_A(\mathbf{z})\}^{\gamma_A}G_B}}.$$

References

Behrendt, C.E., Gehan, E.A., 2009. Treatment-subgroup interaction: An example from a published, phase II clinical trial. Contemp. Clin. Trials 30, 279–281. http://dx.doi.org/10.1016/j.cct.2009.02.002.

Biswas, A., Bhattacharya, R., 2018. A class of covariate-adjusted response-adaptive allocation designs for multitreatment binary response trials. J. Biopharm. Statist. 28, 809–823. http://dx.doi.org/10.1080/10543406.2018.1485683.

Biswas, A., Bhattacharya, R., Park, E., 2016. On a class of covariate-adjusted response adaptive designs for survival outcomes. Stat. Methods Med. Res. 25, 2444–2456. http://dx.doi.org/10.1177/0962280214524177.

Biswas, A., Mandal, S., 2004. Optimal adaptive designs in phase III clinical trials for continuous responses with covariates. In: Bucchianico, A. Di, Lauter, H., Wynn, H.P. (Eds.), mODa 7—Advances in Model-Oriented Design and Analysis. Physica-Verlag, Heidelberg, pp. 51–58. http://dx.doi.org/10.1016/j.jkss.2010.04.004.

Cox, D.R., 1972. Regression models and life-tables (with discussion). J. R. Stat. Soc. Ser. B Stat. Methodol. 34, 187–220. http://dx.doi.org/10.1111/j.2517-6161.1972.tb00899.x.

Efron, B., 1971. Forcing a sequential experiment to be balanced. Biometrika 58, 403–417. http://dx.doi.org/10.1093/biomet/58.3.403.

Eisenhauer, E.A., Therasse, P., Bogaerts, J., Schwartz, L.H., Sargent, D., Ford, R., Dancey, J., Arbuck, S., Gwyther, S., Mooney, M., Rubinstein, L., 2009. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). Eur. J. Cancer 45, 228–247. http://dx.doi.org/10.1016/j.ejca.2008.10.026.

Gilson, C., Chowdhury, S., Parmar, M., Sydes, M., STAMPEDE Investigators, 2017. Incorporating biomarker stratification into STAMPEDE: an adaptive multi-arm, multi-stage trial platform. Clin. Oncol. 29, 778–786. http://dx.doi.org/10.1016/j.clon.2017.10.004.

Hu, F., Rosenberger, W.F., 2003. Optimality, variability, power: evaluating response-adaptive randomization procedures for treatment comparisons. J. Amer. Statist. Assoc. 98, 671–678. http://dx.doi.org/10.1198/016214503000000576.

Hu, F., Zhang, L.-X., 2004. Asymptotic properties of doubly adaptive biased coin designs for multitreatment clinical trials. Ann. Statist. 32, 268–301. http://dx.doi.org/10.1214/aos/1079120137.

Hu, F., Zhang, L.-X., He, X., 2009. Efficient randomized-adaptive designs. Ann. Statist. 37, 2543–2560. http://dx.doi.org/10.1214/08-AOS655.

International Conference On Harmonisation Of Technical Requirements For Registration Of Pharmaceuticals For Human Use(ICH), 2019. Guideline for Good Clinical Practice E9 (R1): Addendum on Estimands and Sensitivity Analysis in Clinical Trials To the Guideline on Statistical Principles for Clinical Trials. https://www.gmp-navigator.com/files/guidemgr/E9-R1_Step4_Guideline_2019_1203.pdf.

International Conference On Harmonisation Off Technical Requirements For Registration Of Pharmaceuticals For Human Use(ICH), 1998. Guideline for Good Clinical Practice E9. https://database.ich.org/sites/default/files/E9_Guideline.pdf.

Karapetis, C.S., Khambata-Ford, S., Jonker, D.J., O'Callaghan, C.J., Tu, D., Tebbutt, N.C., Simes, R.J., Chalchal, H., Shapiro, J.D., Robitaille, S., Price, T.J., Shepherd, L., Au, H.-J., Langer, C., Moore, M.J., Zalchberg, J.R., 2008. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. N. Engl. J. Med. 359, 1757–1765. http://dx.doi.org/10.1056/NEJMoa0804385.

Melfi, V.F., Page, C., Geraldes, M., 2001. An adaptive randomized design with application to estimation. Canad. J. Statist. 20, 328–342. http://dx.doi.org/10.2307/3316054.

Mok, T.S., Wu, Y.-L., Thongprasert, S., Yang, C.-H., Chu, D.-T., Saijo, N., Sunpaweravong, P., Han, B., Margono, B., Ichinose, Y., Nishiwaki, Y., Ohe, Y., Yang, J.-J., Chewaskulyong, B., Jiang, H., Duffield, E.L., Watkins, C.L., Armour, A.A., Fukuoka, M., 2009. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. N. Engl. J. Med. 361, 947–957. http://dx.doi.org/10.1056/NEJMoa0810699.

Pocock, S.J., Simon, R., 1975. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics 31, 103–115. http://dx.doi.org/10.2307/2529712.

Rosenberger, W.F., Hu, F., 2004. Maximizing power and minimizing treatment failures in clinical trials. Clin. Trials 1, 141–147. http://dx.doi.org/10.1191/1740774504cn016oa.

Rosenberger, W.F., Sverdlov, O., 2008. Handling covariates in the design of clinical trials. Statist. Sci. 23, 404–419. http://dx.doi.org/10.1214/08-STS269.

Rosenberger, W.F., Vidyashankar, A.N., Agarwal, D.K., 2001. Covariate-adjusted response-adaptive designs for binary response. J. Biopharm. Statist. 11, 227–236. http://dx.doi.org/10.1081/BIP-120008846.

Sverdlov, O., Rosenberger, W.F., Ryeznik, Y., 2013. Utility of covariate-adjusted response-adaptive randomization in survival trials. Stat. Biopharm. Res. 5, 38–53. http://dx.doi.org/10.1080/19466315.2012.754376.

Sverdlov, O., Tymofyeyev, Y., Wong, W.K., 2011. Optimal response-adaptive randomized designs for multi-armed survival trials. Stat. Med. 30, 2890–2910. http://dx.doi.org/10.1002/sim.4331.

Sydes, M.R., Parmar, M.K., Mason, M.D., Clarke, N.W., Amos, C., Anderson, J., Bono, J.de., Dearnaley, D.P., Dwyer, J., Green, C., Jovic, G., 2012. Flexible trial design in practice-stopping arms for lack-of-benefit and adding research arms mid-trial in STAMPEDE: a multi-arm multi-stage randomized controlled trial. Trials 13, 1–14. http://dx.doi.org/10.1186/1745-6215-13-168.

U.S. Department of Health and Human Services Food and Drug Administration (FDA), 2022. CID Case Study: A Study in Patients with Systemic Lupus Erythematosus.. https://www.fda.gov/media/155404/download.

U.S. Department of Health and Human Services Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), 2019. Adaptive Designs for Clinical Trials of Drugs and Biologics Guidance for Industry. https://www.fda.gov/media/78495/download.

U.S. Department of Health and Human Services Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), 2022. A Multiple Endpoints in Clinical Trials Guidance for Industry. https://www.fda.gov/media/162416/download.

Zhang, L.-X., Hu, F., 2009. A new family of covariate-adjusted response-adaptive designs and their properties. Appl. Math. J. Chinese Univ. Ser. A 24, 1–13. http://dx.doi.org/10.1007/s11766-009-0001-6.
Zhang, L.-X., Hu, F., Cheung, S.H., Chan, W.S., 2007. Asymptotic properties of covariate-adjusted response-adaptive designs. Ann. Statist. 35, 1166–1182. http://dx.doi.org/10.1214/009053606000001424.
Zhang, L., Rosenberger, W.F., 2006. Response-adaptive randomization in clinical trials with continuous outcomes. Biometrics 62, 562–569. http://dx.doi.org/10.1111/j.1541-0420.2005.00496.x.
Zhang, L., Rosenberger, W.F., 2007. Response-adaptive randomization for survival trials: the parametric approach. Appl. Stat. 56, 153–165. http://dx.doi.org/10.1111/j.1467-9876.2007.00571.x.
Zhu, H., 2015. Covariate-adjusted response adaptive designs incorporating covariates with and without treatment interactions. Canad. J. Statist. 43, 534–553. http://dx.doi.org/10.1002/cjs.11260.
Zhu, H., Hu, F., 2010. Covariate-adjusted response adaptive randomization design with survival responses. In: 2010 Joint Statistical Meetings, Vancouver, Canada.