
Research Article

Adaptive Optimal Designs for Dose-Finding Studies with Time-to-Event Outcomes

Yevgen Ryzhnik,^{1,2,4} Oleksandr Sverdlov,³ and Andrew C. Hooker²

Received 23 June 2017; accepted 28 October 2017; published online 28 December 2017

Abstract. We consider optimal design problems for dose-finding studies with censored Weibull time-to-event outcomes. Locally D-optimal designs are investigated for a quadratic dose–response model for log-transformed data subject to right censoring. Two-stage adaptive D-optimal designs using maximum likelihood estimation (MLE) model updating are explored through simulation for a range of different dose–response scenarios and different amounts of censoring in the model. The adaptive optimal designs are found to be nearly as efficient as the locally D-optimal designs. A popular equal allocation design can be highly inefficient when the amount of censored data is high and when the Weibull model hazard is increasing. The issues of sample size planning/early stopping for an adaptive trial are investigated as well. The adaptive D-optimal design with early stopping can potentially reduce study size while achieving similar estimation precision as the fixed allocation design.

KEY WORDS: adaptive design; censoring; D-optimal design; dose finding; Weibull distribution.

INTRODUCTION

Dose–response studies play an important role in clinical drug development. Such studies are typically randomized, multi-armed, placebo-controlled parallel group designs involving several dose levels of an investigational drug. The study goals may be to estimate the drug’s dose–response profile with respect to some primary outcome measure and to identify a dose or doses to be tested in subsequent confirmatory phase III trials. Optimization of a trial design can allow an experimenter to achieve study objectives most efficiently with a given sample size.

Many clinical trials use time-to-event outcomes as primary study endpoints. The outcome could be, for example, progression-free survival in oncology, duration of viral shredding in virology, time from treatment administration until pain symptoms disappear in studies of migraine, time to onset/duration of anesthesia in dentistry, or time to first relapse in multiple sclerosis. Optimal experimental designs

for multi-arm time-to-event outcome trials are warranted, but finding and implementing such designs in practice may be challenging due to uncertainty about the model for event times, delayed and potentially censored outcomes, and dependence of optimal designs for survival models on model parameters that are unknown at the trial outset (the so-called *locally optimal* designs).

Recently, there has been an increasing interest in research and application of optimal designs for experiments with time-to-event outcomes. For problems where the design space is discrete (e.g., treatment is a classification factor), the design optimization involves finding optimal allocation proportions to the given treatment groups to maximize efficiency of treatment comparisons for selected study objectives (1). Optimal allocation designs for survival trials with two or more treatment arms were studied in (2–6) to name a few. These optimal allocation designs can be implemented in practice by means of response–adaptive randomization with established statistical properties (7).

Another class of problems involve dose–response studies where the design space is an interval and therefore the dose level is measured on a continuous scale. In this case, for a given dose–response regression model, an optimal design problem is to determine a set of optimal doses and the probability mass distribution at these doses to maximize some convex criterion of the model Fisher information matrix. Optimal designs for two-parameter exponential regression models with different censoring mechanisms were investigated in (8–10). Optimal designs for most efficient estimation of specific quantiles of censored Weibull or log-normal observations were developed in (11). Optimal design

Electronic supplementary material The online version of this article (<https://doi.org/10.1208/s12248-017-0166-5>) contains supplementary material, which is available to authorized users.

¹ Department of Mathematics, Uppsala University, Room Å14133 Lägerhyddsvägen 1, Hus 1, 6 och 7, 751 06, Uppsala, Sweden.

² Department of Pharmaceutical Biosciences, Uppsala University, Uppsala, Sweden.

³ Early Development Biostatistics – Translational Medicine, Novartis Pharmaceuticals Corporation, East Hanover, New Jersey, USA.

⁴ To whom correspondence should be addressed. (e-mail: yevgen.ryzhnik@math.uu.se)

problems for accelerated failure time (AFT) models with log-logistic distribution for event times and type I and random censoring were considered in (12). Optimal designs for partial likelihood (Cox’s proportional hazard) regression models were investigated in (13) and (14). One common and important observation from these papers is that censoring impacts the structure of optimal designs. Also, most of the optimal designs developed in these papers are locally optimal and cannot be implemented unless reliable guesstimates of the model parameters are available to an experimenter. Solutions to this problem include: (i) *maximin* designs that maximize the worst efficiency with respect to the locally optimal designs over the range of potential model parameter values, (ii) *Bayesian* optimal designs that maximize average efficiency with respect to the locally optimal designs for a given prior distribution on the parameters, or (iii) *adaptive* optimal designs which sequentially or periodically update the estimates of the model parameters and direct future dose assignments to the targeted optimal design.

In this paper, we overcome the limitation of locally optimal designs by developing adaptive optimal designs for dose-finding trials with Weibull time-to-event outcomes. We focus on the Weibull family of distributions because this family is widely used in the analysis of time-to-event data and its utility is well documented (15). The “Materials and Methods” section provides necessary statistical background. We study locally D-optimal designs for the most precise estimation of the dose–response relationship, both in situations with and without censoring. In the same section, we propose multi-stage adaptive optimal designs using maximum likelihood (MLE) updating. In the “Results” section, we investigate the locally D-optimal design structure and its efficiency compared to the popular equal allocation design. A striking result is that equal allocation designs can be highly inefficient in the presence of heavy censoring. We also report results of a simulation study to investigate operating characteristics of two non-adaptive designs (uniform and locally D-optimal) and a two-stage adaptive D-optimal design using MLE updating under 24 different dose–response scenarios and different amount of censoring in the model. We find that the proposed adaptive D-optimal designs can significantly improve efficiency of a dose-finding trial; in fact, adaptive optimal designs can be nearly as efficient as the locally optimal designs. Furthermore, we present simulation results in a more complex setting, assuming the trial has pre-specified criteria which can potentially enable stopping of the trial once model parameters have been estimated with due precision, thereby potentially reducing the total sample size in the study. The “Discussion” section provides some concluding remarks and outlines future work.

The R code used to generate all results in this paper is fully documented and is available for download from the journal website.

MATERIALS AND METHODS

Statistical Background

Let $T > 0$ denote the event time, $\mathbf{x} = (1, x_1, \dots, x_p)'$ be a vector of covariates (1 corresponds to the baseline) and $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)'$ be the vector of regression coefficients.

Consider the following *accelerated failure time* (AFT) model:

$$\log T = \boldsymbol{\beta}'\mathbf{x} + bW, \tag{1}$$

where $b > 0$ is an unknown parameter and W is a random error term with a continuous probability density function (p.d.f.) $f(w)$ on the real line.

In time-to-event trials, observations are likely to be censored. Standard methods of survival analysis are based on the assumption of non-informative right censoring. The event time T is right-censored by a fixed or random time $C > 0$, if T and C are independent and one observes (t, δ) , where $t = \min(T, C)$ and $\delta_i = I\{T_i \leq C_i\}$, where $I\{\cdot\}$ is an indicator function. The random variables t and δ are, in general, not independent; in fact, the joint distribution of (t, δ) can be quite complex as it depends on the censoring mechanism in the trial.

For an experiment of size n , the data structure is $\{(t_i, \delta_i, \mathbf{x}_i), i = 1, \dots, n\}$, where $t_i = \min(T_i, C_i)$, $\delta_i = I\{T_i \leq C_i\}$, and $\mathbf{x}_i = (1, x_{1i}, \dots, x_{pi})'$. The likelihood function for $\boldsymbol{\theta} = (\boldsymbol{\beta}, b)$ is:

$$\mathcal{L}(\boldsymbol{\theta}) = \prod_{i=1}^n \{b^{-1} f(w_i)\}^{\delta_i} \{S(w_i)\}^{1-\delta_i}, \tag{2}$$

where $w_i = (\log t_i - \boldsymbol{\beta}'\mathbf{x}_i)/b$ is the standardized log-transformed time, $S(t) = \Pr(T > t)$ is the survivor function, and $f(t) = \frac{d}{dt}\{1 - S(t)\}$, p.d.f. The log-likelihood is:

$$\ell(\boldsymbol{\theta}) = -r \log b + \sum_{i=1}^n \{\delta_i \log f(w_i) + (1 - \delta_i) \log S(w_i)\},$$

where $r = \sum_{i=1}^n \delta_i$ is the total number of events. The maximum likelihood estimator $(\hat{\boldsymbol{\beta}}, \hat{b})$ of $(\boldsymbol{\beta}, b)$ is obtained by solving the system of $(p + 1)$ score equations $\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}$, which can be written as:

$$\begin{aligned} \mathbf{0} &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = -\frac{1}{b} \sum_{i=1}^n \left\{ \delta_i \frac{\partial \log f(w_i)}{\partial w_i} + (1 - \delta_i) \frac{\partial \log S(w_i)}{\partial w_i} \right\} \mathbf{x}_i, \\ 0 &= \frac{\partial \ell(\boldsymbol{\theta})}{\partial b} = -\frac{r}{b} - \frac{1}{b} \sum_{i=1}^n \left\{ \delta_i \frac{\partial \log f(w_i)}{\partial w_i} + (1 - \delta_i) \frac{\partial \log S(w_i)}{\partial w_i} \right\} w_i. \end{aligned}$$

The Fisher information matrix is given by $\mathbf{M}(\boldsymbol{\theta}) = -E\left\{\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right\}$, and its inverse provides an asymptotic lower bound on the variance of an unbiased estimator of $\boldsymbol{\theta}$. For the model in Eq. (1), $\mathbf{M}(\boldsymbol{\theta})$ is a $(p + 1)$ -square matrix (16):

$$\mathbf{M}(\boldsymbol{\theta}) = \frac{1}{b^2} \begin{pmatrix} -\sum_{i=1}^n E(\lambda_i) \mathbf{x}_i \mathbf{x}_i' & -\sum_{i=1}^n E(\lambda_i w_i) \mathbf{x}_i \\ -\sum_{i=1}^n E(\lambda_i w_i) \mathbf{x}_i' & E(r) - \sum_{i=1}^n E(\lambda_i w_i^2) \end{pmatrix}, \tag{3}$$

where

$$\lambda_i = \delta_i \frac{\partial^2 \log f(w_i)}{\partial w_i^2} + (1 - \delta_i) \frac{\partial^2 \log S(w_i)}{\partial w_i^2}. \tag{4}$$

It should be noted that our considered structure of the likelihood function in Eq. (2) using standardized log-

transformed times as observations is in line with (16), and this enables development of optimal designs for a broad class of parametric models. Specifically, the Fisher information matrix in Eq. (3) has the same general structure for the Weibull, log-logistic, log-normal, and some other distributions, and the key step is derivation of the quantity in Eq. (4) which is determined by the distribution of the model error term for the standardized log-transformed event times.

Alternatively, one can consider the likelihood function using time-to-event data on the original (untransformed) scale as was done, for example in (17). In this case, the elemental Fisher information matrix is readily available for the common time-to-event models, including censored cases (17,18). Applying the generalized regression approach developed in (18) (cf. Chapters 1.6 and 5.4 from (19)), one can derive similar results and facilitate construction of optimal designs in a streamlined manner. This approach merits further investigation but it is beyond the scope of our current work.

Weibull Model

Through the rest of the paper, we shall focus on a second-order polynomial Weibull model for event times, in which case Eq. (1) takes the form:

$$\log T = \beta_0 + \beta_1 x + \beta_2 x^2 + bW, \tag{5}$$

with the error term W following a standard extreme value distribution with p.d.f. $f(w) = \exp(w - e^w)$ and survival function $S(w) = \exp(-e^w)$.

For such a model, the hazard function of T conditional on x is $h(t|x) = b^{-1} \exp(-(\beta_0 + \beta_1 x + \beta_2 x^2)/b) t^{1/b-1}$. Therefore, for a given x , the hazard is monotone increasing if $0 < b < 1$, it is constant if $b = 1$, and it is monotone decreasing if $b > 1$. Also, $\text{Median}(T|x) = \exp(\beta_0 + \beta_1 x + \beta_2 x^2) \{\log(2)\}^b$. Our motivation for choosing the model in Eq. (5) is that such a model is very flexible and it covers various (non)linear dose-response shapes. This model was considered, for instance, in (11) in the context of optimal estimation of specific quantiles of the Weibull distribution, in application to reliability studies.

A direct calculation shows that for the model in Eq. (5), Eq. (4) is simplified to $\lambda_i = -e^{w_i}$.

Without loss of generality, we assume that the dose level x is chosen from the interval $\mathcal{X} = [0, 1]$ (which can be achieved after an appropriate dose-transformation). Let $\mathbf{x} = (1, x, x^2)'$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$, $w_x = (\log t - \mathbf{x}'\boldsymbol{\beta})/b$ (standardized log-transformed time for a subject assigned to dose x), and $\delta_x = I\{T \leq C|x\}$ (event indicator for a subject assigned to dose x). Denote $A_x = E(e^{w_x})$, $B_x = E(w_x e^{w_x})$, and $D_x = E(w_x^2 e^{w_x})$. We have $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, b)$, and the Fisher information matrix for $\boldsymbol{\theta}$ for a single observation at dose $x \in \mathcal{X}$ is a 4×4 matrix:

$$\mathbf{M}_x(\boldsymbol{\theta}) = \frac{1}{b} \begin{pmatrix} A_x & xA_x & x^2A_x & B_x \\ xA_x & x^2A_x & x^3A_x & xB_x \\ x^2A_x & x^3A_x & x^4A_x & x^2B_x \\ B_x & xB_x & x^2B_x & E(\delta_x) + D_x \end{pmatrix} \tag{6}$$

Type I Censoring

The expressions for A_x, B_x, D_x , and $E(\delta_x)$ in Eq. (6) are, in general, functions of $\boldsymbol{\theta}$ and the censoring mechanism in the trial. Although the methodology in this paper is applicable for any generic right-censoring scheme, for consistency of presentation throughout the paper, we consider a *type I censoring* scheme (2,10,12) which can be described as follows. Each subject in the study has a fixed follow-up time $\tau > 0$. At dose level x , the subject's event time is T_x , but one observes $t_x = \min(T_x, \tau)$ and $\delta_x = I\{t_x = T_x\}$ (i.e., $\delta_x = 1$ if $t_x = T_x$ and $\delta_x = 0$ if $t_x = \tau$). Let $L_x = (\log \tau - \boldsymbol{\beta}'\mathbf{x})/b$ and $W = (\log T_x - \boldsymbol{\beta}'\mathbf{x})/b$. Then, the random variable $w_x = \min(W, L_x)$ has the distribution function $F_x(w) = \int_{-\infty}^w \exp(z - e^z) dz$ for $w < L_x$ and $F_x(w) = 1$ for $w \geq L_x$. Therefore, we derive:

$$\begin{aligned} A_x &= E(\delta_x) = 1 - \exp(-e^{L_x}), \\ B_x &= \int_{-\infty}^{L_x} z \exp(2z - e^z) dz + L_x \exp(L_x - e^{L_x}), \\ D_x &= \int_{-\infty}^{L_x} z^2 \exp(2z - e^z) dz + L_x^2 \exp(L_x - e^{L_x}). \end{aligned}$$

As $\tau \rightarrow \infty$, there is no censoring in the model, in which case $A_x = E(\delta_x) = 1$, $B_x = 1 - \gamma$, and $D_x = \pi^2/6 - 1 + (1 - \gamma)^2$, where $\gamma = 0.577215\dots$ (Euler's constant). Therefore, without censoring, the Fisher information matrix in Eq. (6) does not depend on $\boldsymbol{\theta}$.

OPTIMAL DESIGNS

Locally D-optimal Design

Let n be a fixed predetermined sample size, x_1, \dots, x_K be distinct dose levels in $\mathcal{X} = [0, 1]$, n_k be the number of observations to be sampled at dose x_k , and $\rho_k = n_k/n$ be the corresponding allocation proportion, $0 \leq \rho_k \leq 1$ and $\sum_{k=1}^K \rho_k = 1$. A K -point design is determined by a discrete probability measure:

$$\xi = \begin{pmatrix} x_1 & x_2 & \dots & x_K \\ \rho_1 & \rho_2 & \dots & \rho_K \end{pmatrix}.$$

The design Fisher information matrix for the model in Eq. (5) is a weighted sum of information matrices in Eq. (6) evaluated at doses x_1, \dots, x_K :

$$\mathbf{M}(\xi, \boldsymbol{\theta}) = \sum_{k=1}^K \rho_k \mathbf{M}_{x_k}(\boldsymbol{\theta}).$$

The locally D-optimal design ξ^* minimizes the criterion $-\log |\mathbf{M}(\xi, \boldsymbol{\theta})|$. Such a design yields the smallest volume of the confidence ellipsoid for $\boldsymbol{\theta}$, which leads to most accurate estimation of the dose-response relationship. In practice, the general equivalence theorem (GET) (20) can be used to

determine the D-optimal design. For the model in Eq. (5), the GET asserts the equivalence of the following conditions:

- a) The design ξ^* minimizes $-\log |\mathbf{M}(\xi, \boldsymbol{\theta})|$.
- b) The design ξ^* minimizes $\max_{x \in \mathcal{X}} \text{trace}\{\mathbf{M}^{-1}(\xi, \boldsymbol{\theta})\mathbf{M}_x(\boldsymbol{\theta})\} - 4$.
- c) For all $x \in \mathcal{X}$, the derivative function $d(x, \xi, \boldsymbol{\theta}) = \text{trace}\{\mathbf{M}^{-1}(\xi, \boldsymbol{\theta})\mathbf{M}_x(\boldsymbol{\theta})\} - 4 \leq 0$, with the equality holding at each support point of ξ^* .

A particularly interesting scenario is when no censoring is present in the model. In this case, one can easily verify that a uniform design:

$$\xi_U = \begin{pmatrix} 0 & 0.5 & 1 \\ 1/3 & 1/3 & 1/3 \end{pmatrix} \tag{7}$$

is D-optimal. A direct calculation shows that the derivative function for the model in Eq. (5) without censoring is $d(x, \xi_U, \boldsymbol{\theta}) = \text{trace}\{\mathbf{M}^{-1}(\xi_U, \boldsymbol{\theta})\mathbf{M}_x(\boldsymbol{\theta})\} = 72x(x - 0.5)^2(x - 1)$ which satisfies condition c) of the GET, which confirms that ξ_U is indeed D-optimal.

With censoring, the D-optimal design is more complex as it depends on $\boldsymbol{\theta}$ and the censoring mechanism in the trial. In this work, it is found using a first-order (exchange) algorithm (21) implemented using the R software; the code is fully documented and available in the supplementary online materials.

Adaptive D-optimal Designs

A major limitation of the locally D-optimal design is its dependence on the true values of model parameters and the amount of censoring in the experiment—these aspects are frequently unknown at the study planning stage. An adaptive design is a natural approach to handle such uncertainty. In practice, many dose–response trials are performed in a staged manner. At each stage, a cohort of eligible patients is enrolled and allocated to the study treatments. Patient outcome data can be monitored sequentially or periodically throughout the trial to update knowledge on the underlying dose–response relationship and facilitate informed decisions (e.g., to change allocation of subsequent cohorts to the “most informative” dose levels).

The idea of constructing an adaptive strategy for updating the design under model uncertainty can be traced to the work of Box and Hunter (22) where the authors proposed choosing design points sequentially to maximize an incremental increase of information at each step. See also Chapter 5.3 of (19) and references therein for further background on adaptive optimal designs.

The implementation of a *multi-stage adaptive design* involves the following steps:

- **Stage 1:** $n^{(1)}$ patients are allocated to doses according to some initial design $\xi^{(1)}$.
- **Interim updating:** For $k = 2, \dots, v$, fit model in Eq. (5) using accrued data from stages 1, ..., $(k - 1)$ to obtain an updated estimate $\hat{\boldsymbol{\theta}}^{(k-1)}$ of $\boldsymbol{\theta}$. Compute the optimal design for the k th stage, $\xi^{(k)}$, based on $\hat{\boldsymbol{\theta}}^{(k-1)}$

and the information accumulated up to this point, $\xi^{(1, \dots, k-1)}$.

- **Stage $k = 2, \dots, v$:** $n^{(k)}$ patients are allocated to doses according to $\xi^{(k)}$.

Several notes should be made here. First, an initial design $\xi^{(1)}$ should be chosen judiciously. One reasonable choice is $\xi^{(1)} = \xi_U$, the uniform allocation design (cf. Eq. (7)). Second, an experimenter should decide upfront on the number of stages v , the cohort sizes $n^{(k)}$, $k = 1, \dots, v$, and the maximum total sample size for the experiment $n = n^{(1)} + n^{(2)} + \dots + n^{(v)}$. Third, one should decide upfront on the interim updating algorithm. In this paper, we shall consider updating based on *Maximum Likelihood Estimation* (MLE) (23), which can be described as follows:

1. At the interim analysis $k = 2, \dots, v$, fit the model in Eq. (5) to the cumulative outcome data from cohorts 1, ..., $(k - 1)$ to obtain $\hat{\boldsymbol{\theta}}_{MLE}^{(k-1)}$, the MLE of $\boldsymbol{\theta}$.
2. Amend the cohort design as

$$\xi^{(k)} = \arg \min_{\xi} \left\{ -\log \left| \mathbf{M}_{\text{obs}} \left(\xi^{(1, \dots, k-1)}, \hat{\boldsymbol{\theta}}_{MLE}^{(k-1)} \right) + \mathbf{M} \left(\xi, \hat{\boldsymbol{\theta}}_{MLE}^{(k-1)} \right) \right| \right\}, \tag{8}$$

where $\mathbf{M}_{\text{obs}} \left(\xi^{(1, \dots, k-1)}, \hat{\boldsymbol{\theta}}_{MLE}^{(k-1)} \right)$ is the observed Fisher information matrix of the previous cohorts, with their realized design $\xi^{(1, \dots, k-1)}$, evaluated at $\hat{\boldsymbol{\theta}}_{MLE}^{(k-1)}$.

In practice, an adaptive design with one interim analysis (i.e., $v = 2$, a two-stage design) is a reasonable choice both from statistical and operational perspectives (e.g., (24,25)). Each interim analysis requires database lock, data cleaning, analysis, review, and interpretation of results—if multiple interim looks are planned then study timelines can be delayed which may be undesirable from a business perspective.

RESULTS

Locally D-optimal Design

To investigate the locally D-optimal design structure under different experimental setups, we consider 24 scenarios for the dose–response, which are displayed in Fig. 1. Our rationale for choosing these dose–response scenarios was to cover different combinations of shapes of dose–response (determined by the parameter vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$) and the Weibull hazard pattern (determined by the parameter b). The values of the parameter vector $\boldsymbol{\beta}$ were selected such that all median time-to-event profiles fit in the same range from 0 to 180 (on the y-axis).

Specifically, we consider six choices for the parameter b in Eq. (1) which defines the hazard pattern of a Weibull distribution: four cases of $0 < b < 1$ (monotone increasing hazard): $b = 0.4$, $b = \gamma = 0.57721\dots$ (Euler’s gamma), $b = 0.65$, and $b = 0.8$; the case of an exponential distribution (constant hazard): $b = 1$; and one case of $b > 1$ (monotone decreasing hazard): $b = 1.5$. In addition, we consider four choices for $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$ that determines the shape of a dose–response, namely monotone increasing (Shape I ($\beta_0 = 1.9, \beta_1 = 0.6, \beta_2 =$

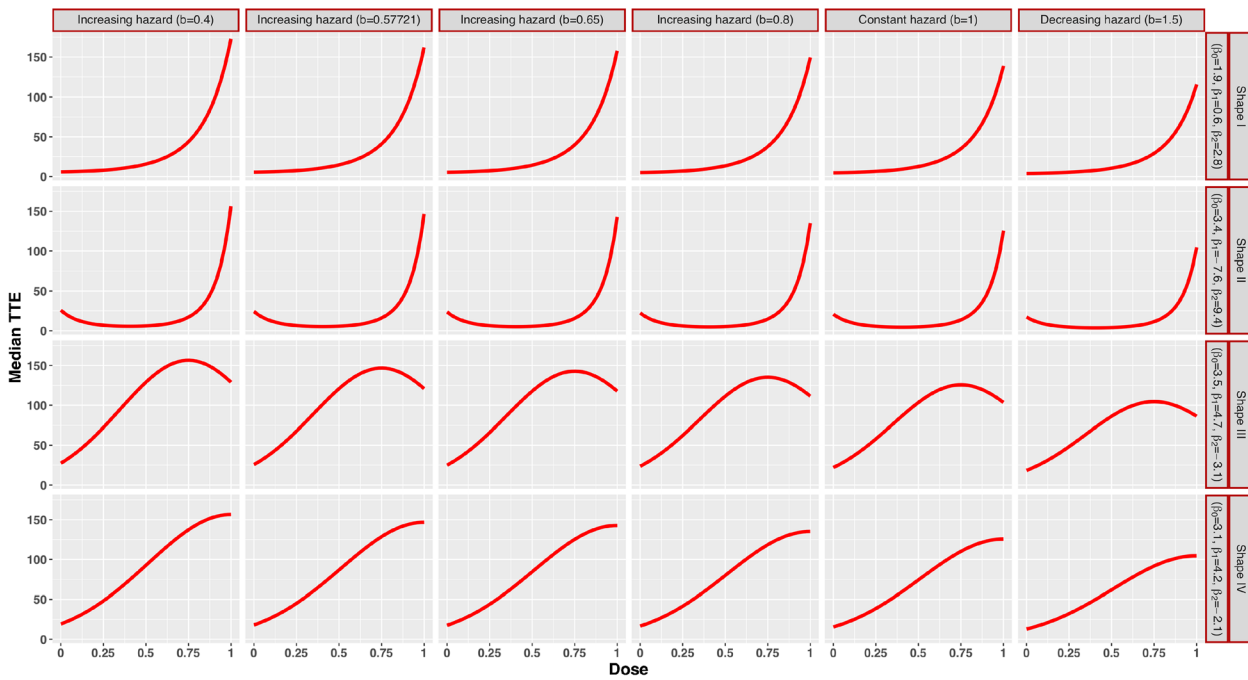


Fig. 1. Dose–response relationships considered. $\text{Median}(T|x) = \exp(\beta_0 + \beta_1x + \beta_2x^2)\{\log(2)\}^b$

2.8); U-shape with a minimum in $[0,1]$ (Shape II ($\beta_0 = 3.4, \beta_1 = -7.6, \beta_2 = 9.4$); unimodal with a maximum in $[0, 1]$ (Shape III ($\beta_0 = 3.5, \beta_1 = 4.7, \beta_2 = -3.1$); and S-shape (Shape IV ($\beta_0 = 3.1, \beta_1 = 4.2, \beta_2 = -2.1$).

Figure 2 is a graphical summary of the D-optimal design structure for the considered dose–response scenarios. From Fig. 2, we see that the D-optimal design always has three points; however, the location of these points in the allowed interval $[0,1]$ as well as the probability mass at these points vary for different values of τ . When τ is small (amount of censored data is high), the D-optimal design can be quite different from the three-point uniform design in Eq. (7); as τ increases, the D-optimal design becomes closer to the uniform design, which is consistent with theory. For every given shape of the dose–response (I, II, III, and IV), the D-optimal design is most shifted from the uniform design when the hazard is increasing ($0 < b < 1$) and least shifted when the hazard is decreasing ($b = 1.5$).

To compare the estimation precision of different designs, we compute the D-efficiency of the uniform design ξ_U in Eq. (7) relative to the D-optimal design ξ^* as follows:

$$D_{\text{eff}}(\boldsymbol{\theta}) = \left\{ \frac{|\mathbf{M}^{-1}(\xi^*, \boldsymbol{\theta})|}{|\mathbf{M}^{-1}(\xi_U, \boldsymbol{\theta})|} \right\}^{\frac{1}{4}}$$

Clearly, $0 < D_{\text{eff}}(\boldsymbol{\theta}) \leq 1$ for any value of $\boldsymbol{\theta}$. $D_{\text{eff}}(\boldsymbol{\theta}) = 1$ indicates that ξ_U is as efficient as ξ^* . A value of $D_{\text{eff}}(\boldsymbol{\theta}) = 0.90$ implies that ξ_U is 90% as efficient as ξ^* ; in other words, an experiment using ξ_U would require 10% more subjects than an experiment using ξ^* to achieve the same level of estimation precision of $\boldsymbol{\theta}$.

Figure 3 shows D-efficiency values of the uniform design vs. D-optimal design for the 24 considered dose–

response scenarios (blue curve), along with the mean values of event probability for the D-optimal design (red curve) and the uniform design (green curve). One can see that for small values of τ (which correspond to low event probability for both designs), the uniform design can be quite inefficient compared to the D-optimal design. The reason for the improved efficiency of the D-optimal design may be, in part, because the D-optimal designs typically increase the probability of events for a given censoring time, which is optimal from a parameter estimation point of view but may not be optimal clinically, if an event is not a desired outcome (26). We also see in Fig. 3 that as τ increases, the two designs are nearly the same (D-efficiency is close to unity), as expected from the result without censoring (Eq. (7)). Further, D-efficiency is also sensitive to the value of b ; as b increases, the two designs are nearly the same regardless of the value of τ (results not shown here).

Simulation Study to Compare Fixed and Adaptive Designs

Here, we present results of several simulation studies to compare operating characteristics of various experimental designs. In the subsection “Fixed Total Sample Size”, we investigate two non-adaptive designs (uniform and locally D-optimal) and a two-stage adaptive D-optimal design using MLE updating, if the total sample size for the experiment is fixed and predetermined (e.g. according to budgetary and/or logistical considerations). In the subsection “Designs with Early Stopping Criteria”, we present results in a more complex setting, assuming the trial can potentially be stopped early (before the total planned sample size has been reached), provided that pre-defined requirements of estimation accuracy have been achieved.

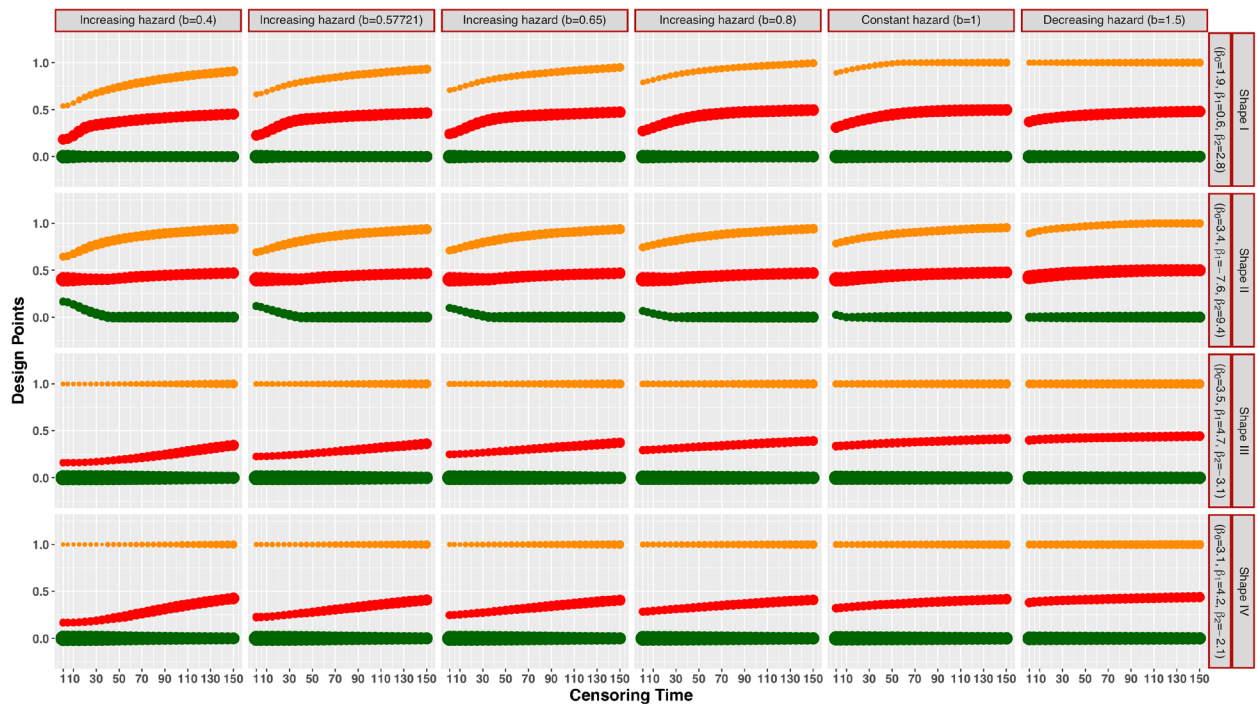


Fig. 2. D-optimal designs for the considered dose–response relationships, with different values of the censoring time. For each scenario, the plot displays the support points of the D-optimal design (green, yellow, and red) for a range of values of the censoring time τ : from $\tau = 1$ (a lot of censoring) to $\tau = 150$ (little censoring). The radius of the circle around the design point is proportional to the probability mass at this point

Fixed Total Sample Size

Operating characteristics of three designs (single-stage uniform, single-stage locally D-optimal, and two-stage adaptive D-optimal) were evaluated under various experimental scenarios, using 1000 simulation runs for each design/scenario

combination. The scenarios included 24 choices of dose–response relationships (Fig. 1), different choices of the total sample size ($n = 150; 300; 450$), different amounts of censoring in the study and, for the adaptive design, different initial cohort sizes. For censoring, we considered type I censoring schemes with parameter τ selected in such a way that the total

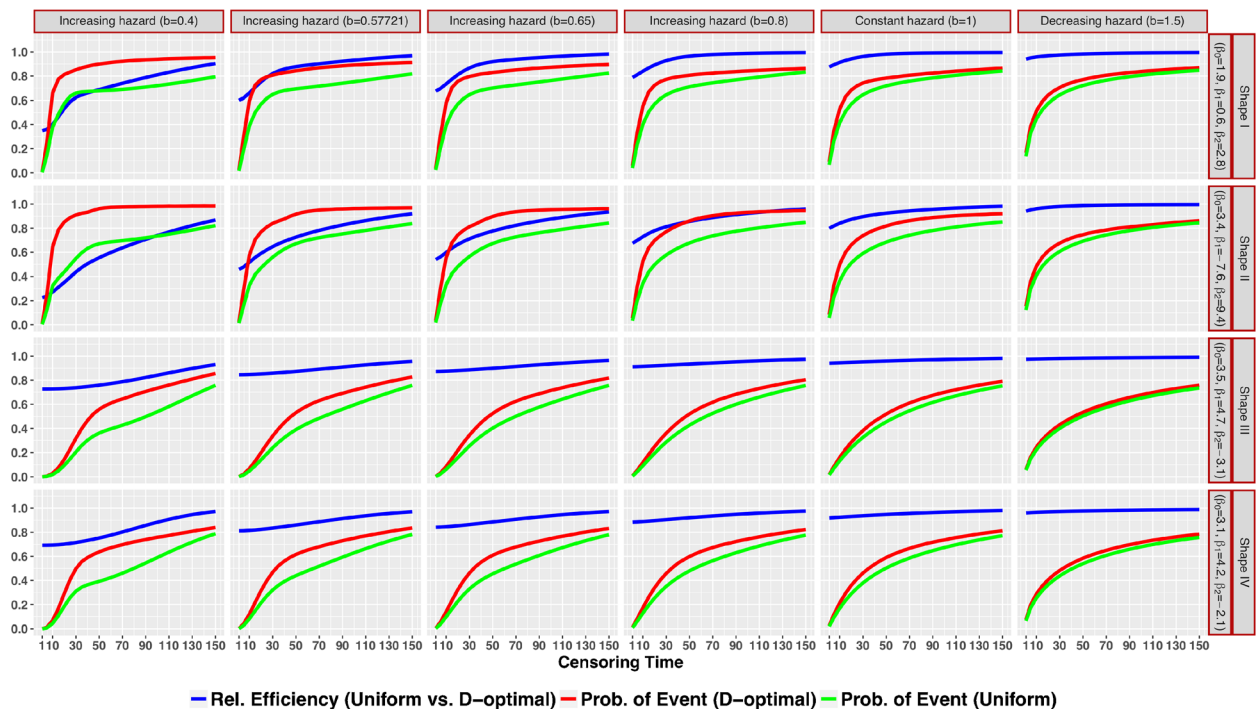


Fig. 3. D-efficiency of a three-point uniform design ξ_U relative to the D-optimal design ξ^*

average probability of event in the experiment for the uniform design in Eq. (7) is one of 25%, 50%, or 75%. Note that the single-stage locally D-optimal design should be viewed here as a theoretical benchmark; in practice, it cannot be implemented because the true model parameter values are unknown at the beginning of the study.

Let us consider in detail a scenario with $n=300$ and 50% total average probability of event for the uniform design. Figure 4 shows both theoretical and estimated (via 1000 simulations) median time-to-event profiles as well as 5th and 95th quantile time-to-event curves, for two-stage adaptive D-optimal design, locally D-optimal design, and uniform design, given four different shapes and the hazard parameter $b = \gamma$. It can be seen that quality of estimation of the dose-response is better for the D-optimal design than for the uniform design. Additional simulations show that this difference is more pronounced in scenarios when the hazard is increasing and b is small ($b \leq \gamma$) (cf. Fig. S1, S2, S3 in the Supplemental Appendix). This is consistent with the theory which suggests that D-optimal designs can be highly skewed from uniform in certain settings. The two-stage adaptive D-optimal design was implemented with an initial cohort of size $n^{(1)}=90$. The first cohort was allocated to doses according to the uniform design, whereas the second cohort was allocated to doses according to the estimated D-optimal design based on the outcome data from the first cohort. From Fig. 4, one can see that the adaptive D-optimal design performs very stably and delivers similar quality of estimation compared to the locally D-optimal design across the considered scenarios. Therefore, the adaptive design learns from experimental data and can be advantageous over a single-stage uniform design, particularly when the hazard pattern is increasing.

We also examined design estimation accuracy for other values of event probabilities. Outputs similar to those in Fig. 4 and Fig. S1, S2, and S3 in the Supplemental Appendix have been generated for average event probabilities of 25% and 75% (results not shown here). In the case of 75% event probability, the designs have, overall, better estimation accuracy than in the case of 50% probability. However, when only 25% of observations, on average, are events, estimation of dose-response is challenging. Our simulations showed that all three designs (uniform, locally D-optimal, and two-stage adaptive D-optimal) in the 25% event probability case had lower estimation accuracy (higher bias and higher variance of estimated mean/median dose-response profiles) than in the 50% event probability case. Therefore, in situations when a high amount of censoring is expected, increasing the study size is necessary.

To assess an impact of an initial cohort size on the design performance of the two-stage D-optimal design, additional simulations assuming the total sample size $n=300$, but with different values of initial cohort size ($n^{(1)}=30, 60, 90, 120, 150, 180, 210, 240,$ and 270) were considered. For larger $n^{(1)}$ values, the design resulted in improved quality of estimation of the dose-response curve: estimated mean/median event times were more closely matched to the true mean/median TTE curve (results not shown here). Figure 5 quantifies gains in efficiency due to a larger size of the initial cohort. The red curve in Fig. 5 displays the simulated average D-efficiency of a two-stage adaptive D-optimal design relative to the locally D-optimal design with true parameters (practically impossible) versus initial cohort size $n^{(1)}$, given four different shapes and a hazard parameter $b = \gamma$ (the corresponding plots for the remaining 20 models can be found in the Supplemental Appendix, Fig. S4). The average D-efficiency of the uniform design is also displayed (blue line in Fig. 5). The gain in

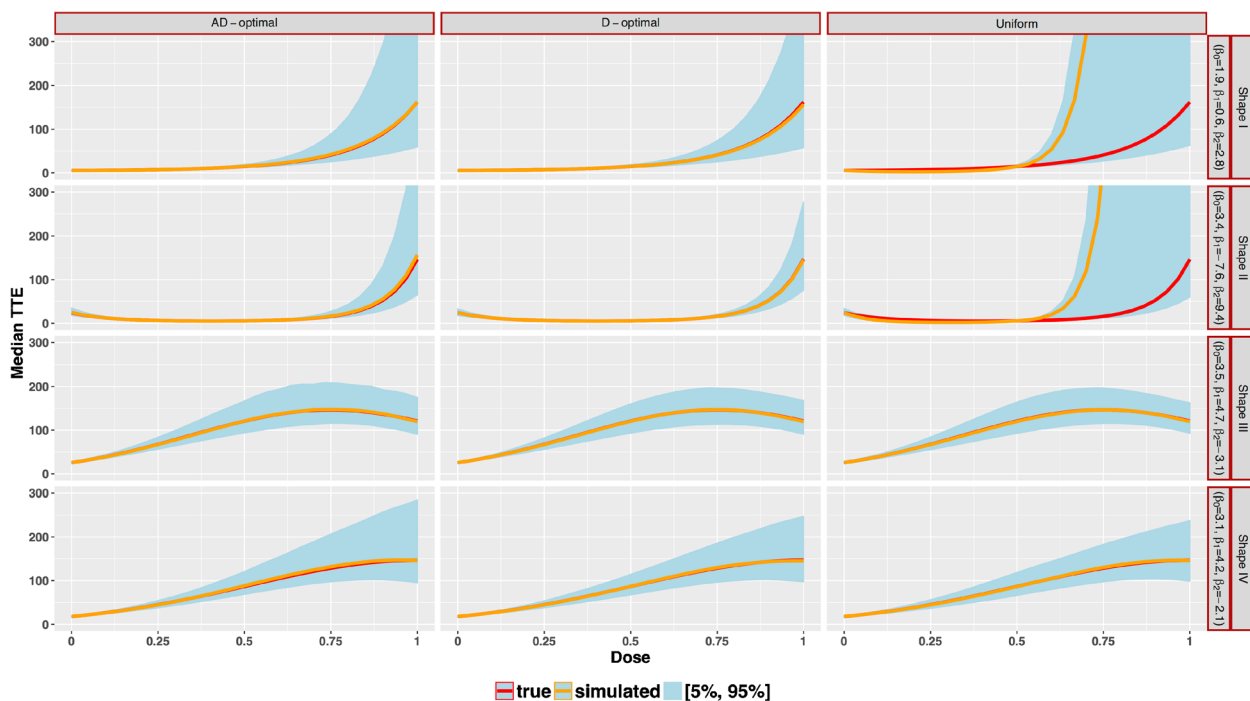


Fig. 4. Estimated dose-response relationship for the two-stage adaptive D-optimal, D-optimal, and uniform designs ($b = \gamma$, $n = 300$ and 50% total average probability of event)

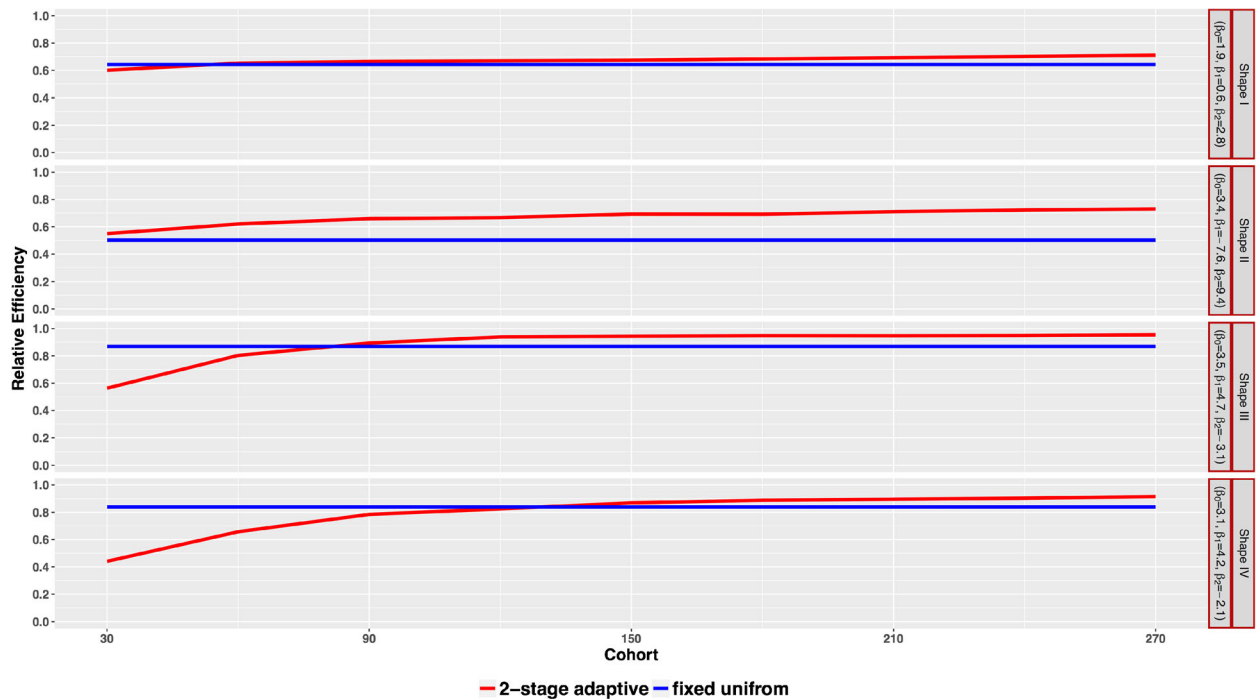


Fig. 5. Average D-efficiency of the two-stage D-optimal design vs. initial cohort size $n^{(1)}(b = \gamma, n = 300$ and 50% total average probability of event)

efficiency due to using a larger initial cohort for the adaptive D-optimal design can be pronounced. For instance, the bottom plot in Fig. 5 (for Shape IV) shows that average D-efficiency can increase from $\sim 45\%$ when $n^{(1)}=30$ to $\sim 90\%$ when $n^{(1)}=270$. We also see that the uniform design can have better relative performance if the initial cohort is too small in the two-stage adaptive design. However, it has been observed that the results also vary across experimental scenarios, which reinforces the importance of careful simulation studies at the planning stage. For some of the dose–response scenarios, e.g., the lower left panel on Fig. S4 in the Supplemental Appendix, we ran an additional simulation with a larger total sample size, $n = 450$, and initial cohort sizes $n^{(1)} = 300, 330, 360, 390, 420$. We observed that starting from $n^{(1)} = 360$, the efficiency of the two-stage adaptive design exceeded that of the uniform design, and the efficiency pattern was increasing with an increase of the initial cohort size.

Designs with Early Stopping Criteria

So far, we have assumed that the study sample size is fixed and predetermined. However, an investigator may want to have a more flexible adaptive design which allows for the possibility to stop the trial early, before the target sample size is reached. Here, we propose an adaptive design with early stopping based on the accuracy of estimation precision. Similar ideas have been explored in the context of extensions of a continual reassessment method (27), by several authors (28–30).

Let $\hat{\theta}_{MLE} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{b})$ denote the maximum likelihood estimate of $\theta = (\beta_0, \beta_1, \beta_2, b)$, and let $(SD_{\hat{\beta}_0}, SD_{\hat{\beta}_1}, SD_{\hat{\beta}_2}, SD_{\hat{b}})$ denote a vector of standard deviations of components of $\hat{\theta}_{MLE}$. We want to stop the experiment once the model

parameters have been estimated with a pre-specified level of accuracy. If we assume that there is no correlation between the parameters and all the parameters have a constant coefficient of variation $\eta \times 100\%$ ($0 < \eta < 1$), i.e.:

$$\frac{SD_{\hat{\beta}_0}}{|\hat{\beta}_0|} = \frac{SD_{\hat{\beta}_1}}{|\hat{\beta}_1|} = \frac{SD_{\hat{\beta}_2}}{|\hat{\beta}_2|} = \frac{SD_{\hat{b}}}{|\hat{b}|} = \eta,$$

then the value of a volume of the confidence ellipsoid under this assumption is:

$$\left(\eta^4 |\hat{\beta}_0| |\hat{\beta}_1| |\hat{\beta}_2| |\hat{b}| \right)^2.$$

Let $M_{obs}(\xi, \hat{\theta}_{MLE})$ be an observed Fisher information matrix, given $\hat{\theta}_{MLE}$ and the current design ξ . The following rule is proposed as a stopping criterion:

$$|M_{obs}^{-1}(\xi, \hat{\theta}_{MLE})| \leq \left(\eta^4 |\hat{\beta}_0| |\hat{\beta}_1| |\hat{\beta}_2| |\hat{b}| \right)^2, \tag{9}$$

for some small predetermined $\eta > 0$. For the purpose of simulation, $\eta = 0.2$ was chosen.

To use the stopping rule Eq. (9) in an adaptive design, the first cohort is randomized to dose groups according to the uniform (non-adaptive) design in Eq. (7). Based on observed data, the stopping criterion in Eq. (9) is checked, and if it is met then the study is stopped; otherwise, the next cohort is

randomized to dose groups according to the updated D-optimal design in Eq. (8). This procedure is repeated until either the stopping criterion is met or the maximum sample size is reached.

We performed simulations to compare three designs using the stopping criterion in Eq. (9). The experimental scenarios included four choices of dose–response relationships (cf. Fig. 1) (four shapes and parameter $b = \gamma$) and three choices for the total average probability of event (25%, 50%, or 75%). The three designs were: (i) *Uniform design*: for each cohort of patients, each subject is randomized to one of the doses 0, 0.5, or 1 with equal probability; (ii) *Locally D-optimal design*: assuming that the true model parameters are known, each cohort of patients is randomized among the optimal doses according to the true D-optimal design; and (iii) *Adaptive D-optimal design with a stopping rule*: described above. For each design, interim analyses were made after every 90 subjects (i.e., the number of subjects in each cohort is 90) and the trial was stopped when the stopping criteria was met or when the maximum sample size in the study was reached (set to $n_{\max} = 2,500$). The randomization probabilities were applied individually to every subject in the given cohort. All results were obtained based on 1000 simulation runs for each design/scenario combination.

Since all three designs aimed at achieving similar quality of estimation but with potentially different sample sizes, a particularly important operating characteristic is final sample size at study termination. Figure 6 shows boxplots of the final sample size distribution for the three designs under the four dose–response scenarios when the average event probability is ~ 50%. One can see that the adaptive D-optimal design has very similar performance to the (practically impossible) locally D-optimal design. These two designs require

substantially lower sample size than the uniform design in scenarios when hazard is increasing and parameter b is “small” ($b \leq \gamma$), and they will have similar sample sizes to the uniform design in scenarios when the hazard is constant or decreasing or $b > \gamma$, since in this case the efficiency of adaptive design is approaching the efficiency of the uniform design when b increases (Fig. S4 in supplemental appendix).

We ran additional simulations with other choices of the parameter η (results not shown here). In particular, we found that for $\eta = 0.25$, the final sample size upon termination was overall lower than in the case of $\eta = 0.20$, and for some scenarios it was equal to the size of the initial cohort. This makes good sense as the targeted quality of estimation with $\eta = 0.25$ is a lower bar than in the case of $\eta = 0.20$, and therefore, a smaller sample size can fulfill the experimental objectives. For $\eta = 0.15$, we observed that the final sample size was higher and required more than one adaptation.

Tables I and II provide a closer look at the relative merits of adaptive D-optimal design relative to the uniform and (practically impossible) locally D-optimal design. Here we investigate the difference (%) in median sample size for a fixed average event probability (50%) and different values of the cohort size (Table I), as well as for a fixed cohort size (90 subjects) and different values of average event probability (25%, 50%, 75%) (Table II). The main findings are, overall, consistent with the previous ones—for shapes I and II, adaptive design requires either the same or at most 50% larger sample size compared to the fixed D-optimal design; however, when comparing adaptive D-optimal design vs. fixed uniform design, the difference is more pronounced—from 100% to 675% more subjects are needed for the uniform design to satisfy the stopping criterion in Eq. (9). For shapes III and IV, all three designs require the same sample size in order to satisfy Eq. (9).

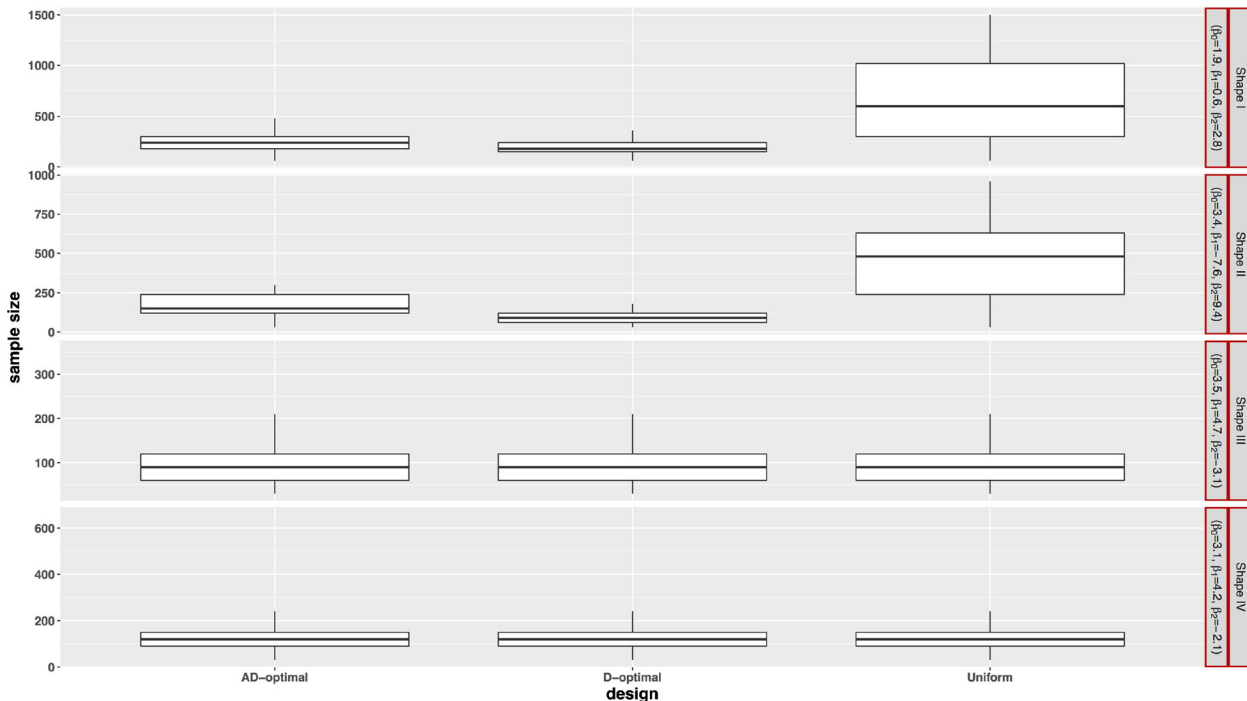


Fig. 6. Final sample size at study termination for three designs with the early stopping rule ($b = \gamma$, probability of event ~ 50%). *Shape I* ($\beta_0 = 1.9, \beta_1 = 0.6, \beta_2 = 2.8$), *Shape II* ($\beta_0 = 3.4, \beta_1 = -7.6, \beta_2 = 9.4$). *Shape III* ($\beta_0 = 3.5, \beta_1 = 4.7, \beta_2 = -3.1$), and *Shape IV* ($\beta_0 = 3.1, \beta_1 = 4.2, \beta_2 = -2.1$)

Table I. Difference in Median Sample Size Upon Termination for a Scenario with Early Stopping Rule ($b = \gamma$, average probability of event = 50%), with Various Initial Cohort Sizes

$(b = \gamma)$ average probability of event = 50%				
Initial cohort size	Shape I		Shape II	
	Adaptive vs. D-optimal	Adaptive vs. uniform	Adaptive vs. D-optimal	Adaptive vs. uniform
30	50%	-675%	50%	-675%
60	0%	-200%	50%	-350%
90	0%	-200%	50%	-200%
120	0%	-150%	50%	-150%
150	33%	-100%	50%	-100%

Shape I ($\beta_0 = 1.9, \beta_1 = 0.6, \beta_2 = 2.8$), Shape II ($\beta_0 = 3.4, \beta_1 = -7.6, \beta_2 = 9.4$)

DISCUSSION

The methodology proposed in this paper is applicable to dose–response studies with time-to-event outcomes, where events are assumed to follow a Weibull distribution and are subject to right-censoring. We focused on the Weibull family of distributions because this family is widely used in survival analysis and its utility is well documented (15). A quadratic regression for the log-transformed event times provides flexibility and covers various (non)linear dose–response shapes for the median time-to-event dose–response relationship. In our study, the experimental design settings corresponded to three-arm trials. Such designs are very common in randomized phase II clinical studies where the three arms correspond, for instance, to placebo, low dose of the drug, and high dose of the drug (e.g., the maximum tolerated dose).

In general, the D-optimal design can be quite different from the popular uniform (equal allocation) design due to the dependence on both model parameters and the amount of censoring in the model. However, the D-optimal design cannot be directly implemented unless reliable guesstimates of the model parameters are available. To overcome the limitation of local optimality, we proposed a two-stage adaptive D-optimal design which performs dose assignments adaptively, according to updated knowledge on the dose–response curve at an interim analysis. Simulations under various experimental scenarios show that the proposed two-stage adaptive design provides a very good approximation and it is nearly as efficient as the true D-optimal design. A

particular advantage of the adaptive D-optimal design compared to the uniform design has been observed in scenarios when the Weibull model hazard is increasing. Since the hazard pattern is frequently unknown at the trial outset, a two-stage adaptive D-optimal design provides a scientifically sound approach to dose finding in time-to-event settings. Higher statistical efficiency can potentially translate into reduction in study sample size. We showed that by adding a stopping criterion prescribing that the experiment should stop once model parameters have been estimated with due precision, one can add even more flexibility to adaptive D-optimal designs. In this paper, we explored one simple and practical stopping criterion (cf. Eq. (9)). Other criteria can be considered as well. In particular, we investigated a stopping criterion which prescribes stopping the study once the maximum value of the coefficient of variation for estimating each component of the model parameter vector is less than or equal to a predetermined constant. The results and conclusions were generally similar to the case of the early stopping based on the D-criterion. The detailed results are available from the first author upon request.

In addition to potential benefits of adaptive D-optimal designs, locally D-optimal designs themselves provide useful tools in real dose-finding experiments. For instance, they provide theoretical measures of statistical estimation precision against which other experimental designs (e.g., with different number of treatment arms and/or different allocation ratios) can be compared.

As noted above, the optimal designs considered in this paper are based on the Weibull family of distributions. If the

Table II. Difference in Median Sample Size Upon Termination for a Scenario with Early Stopping Rule ($b = \gamma$, initial cohort size = 90), with Various Values of Pr(Event)

$(b = \gamma)$ initial cohort size = 90				
Pr(Event)	Shape I		Shape II	
	Adaptive vs. D-optimal	Adaptive vs. uniform	Adaptive vs. D-optimal	Adaptive vs. uniform
25%	0%	-283%	50%	-650%
50%	0%	-200%	50%	-200%
75%	50%	0%	50%	0%

Shape I ($\beta_0 = 1.9, \beta_1 = 0.6, \beta_2 = 2.8$), Shape II ($\beta_0 = 3.4, \beta_1 = -7.6, \beta_2 = 9.4$)

Weibull model is misspecified, then loss in efficiency is possible. To handle model uncertainty at the design stage, one could consider several parametric candidate models (e.g., Weibull, log-logistic, log-normal), and consider a two-stage design for which the first-stage data are used to estimate each model from the candidate set, and then select the “best” one for the second stage. This idea is similar in the spirit to the MCP-Mod methodology (31), but its further development is beyond the scope of the current paper.

Another important consideration is the censoring scheme. In this paper, we focused on type I censoring. On the other hand, in many clinical trials, patient enrollment times are random (e.g., follow a Poisson process), and therefore, event times are censored by random follow-up times. This calls for using more complex censoring schemes in the study. Optimal designs for time-to-event trials with censoring driven by random enrollment were obtained recently in (17) and (32). These two papers provide general theoretical results applicable for a broad class of time-to-event models. An interesting and important future research topic is construction of adaptive optimal designs for studies with censoring driven by random enrollment.

In addition to model misspecification in the design phase, one will certainly have parameter misspecification. In this work, we considered MLE updating, where the calculated designs are based on point estimates of the MLE. Alternatively, one could consider uncertainty in parameter estimates, using, for example, robust design criteria (32) or Bayesian updating (21). In the case of Bayesian updating, Eq. (8) in the adaptive design would have to be modified as follows: Assume θ is a random vector with a prior p.d.f. $\pi(\theta)$. At the interim analysis $k=2, \dots, \nu$, obtain the posterior p.d.f. $\pi(\theta|data) \propto \mathcal{L}(\theta|data)\pi(\theta)$. Amend the design as

$$\xi^{(k)} = \arg \max_{\xi} \int_{\theta} \log \mathbf{M}(\xi, \theta) \pi(\theta|data) d\theta.$$

Two-stage designs with Bayesian updating were recently investigated in the context of adaptive MCP-Mod procedures (23), and they were found to outperform adaptive designs with MLE updating. Implementation of Bayesian updating and its comparison with MLE updating in time-to-event dose-finding trials is an important future work. Some preliminary results were obtained in (34).

In any experiment involving multiple treatment arms, it is important that treatment allocation involves randomization—this allows mitigation of various experimental biases (35). For adaptive D-optimal designs, both dose levels and target allocation proportions are calibrated through the course of the experiment. Treatment allocation ratio for a given cohort of subjects is frequently different from equal allocation. To implement an unequal allocation in practice, one can use randomization procedures with established statistical properties such as *brick tunnel randomization* (36) or *wide brick tunnel randomization* (37). These procedures preserve the allocation ratio at each step and lead to valid statistical inference at the end of the trial.

Finally, we would like to highlight that successful implementation of any methodology relies on validated statistical software. The R code used to generate results in

this paper is fully documented and can be used to generate additional results under user-defined experimental scenarios.

CONCLUSION

The current paper developed adaptive D-optimal designs for dose-finding experiments with censored time-to-event outcomes. The proposed designs overcome a limitation of local D-optimal designs by performing response-adaptive allocation to most informative dose levels according to pre-defined statistical criteria. These designs are flexible and maintain a high level of statistical estimation efficiency, which can potentially translate into reduction in study sample size. All results presented in this paper are fully reproducible with the R code which can be downloaded from the journal website.

ACKNOWLEDGEMENTS

The research leading to these results has received support from the Innovative Medicines Initiative Joint Undertaking under grant agreement No. 115156 (the DDMoRe project), resources of which were composed of financial contributions from the European Union’s Seventh Framework Programme (FP7/2007–2013) and EFPIA companies’ in kind contribution. The DDMoRe project was also supported by financial contributions from academic and SME partners. The authors would like to acknowledge three anonymous reviewers whose comments led to an improved version of the paper.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

REFERENCES

1. Sverdlov O, Rosenberger WF. On recent advances in optimal allocation designs in clinical trials. *J Stat Theory Prac.* 2013;7(4):753–73. <https://doi.org/10.1080/15598608.2013.783726>.
2. Kalish LA, Harrington DP. Efficiency of balanced treatment allocation for survival analysis. *Biometrics.* 1988;44(3):815–21. <https://doi.org/10.2307/2531593>.
3. Zhang L, Rosenberger WF. Response-adaptive randomization for survival trials: the parametric approach. *J R Stat Soc: Ser C: Appl Stat.* 2007;56(2):153–65. <https://doi.org/10.1111/j.1467-9876.2007.00571.x>.
4. Sverdlov O, Tymofeyev Y, Wong WK. Optimal response-adaptive randomized designs for multi-armed survival trials. *Stat Med.* 2011;30(24):2890–910. <https://doi.org/10.1002/sim.4331>.
5. Sverdlov O, Ryzhnik Y, Wong WK. Doubly-adaptive biased coin designs for balancing competing objectives in time-to-event trials. *Statistics and Its Interface.* 2012;5:401–13.
6. Sverdlov O, Ryzhnik Y, Wong WK. Efficient and ethical response-adaptive randomization designs for multi-arm clinical trials with

- Weibull time-to-event outcomes. *J Biopharm Stat.* 2014;24(4):732–54. <https://doi.org/10.1080/10543406.2014.903261>.
7. Ryznik Y, Sverdlov O, Wong WK. RARtool: a MATLAB software package for designing response-adaptive randomized clinical trials with time-to-event outcomes. *J Stat Softw.* 2015;66(1) 10.18637/jss.v066.i01.
 8. Lopez-Fidalgo J, Rivas-Lopez MJ. Optimal designs for Cox regression. *Statistica Neerlandica.* 2009;63(2):135–48. <https://doi.org/10.1111/j.1467-9574.2009.00415.x>.
 9. Müller CH. D-optimal designs for lifetime experiments with exponential distribution and censoring. In Ucinski D, editor. *mODa 9—advances in model-oriented design and analysis.*: Springer International Publishing Switzerland; 2010. p. 179–86.
 10. Konstantinou M, Biedermann S, Kimber AC. Optimal designs for two-parameter nonlinear models with application to survival models. *Stat Sin.* 2014;24:415–28.
 11. Chaloner K, Larntz K. Bayesian design for accelerated life testing. *Stat Plann Inference.* 1992;33(2):245–59. [https://doi.org/10.1016/0378-3758\(92\)90071-Y](https://doi.org/10.1016/0378-3758(92)90071-Y).
 12. Rivas-Lopez MJ, Lopez-Fidalgo J, del Campo R. Optimal experimental designs for accelerated failure time with type I and random censoring. *Biom J.* 2014;56(5):819–37. <https://doi.org/10.1002/bimj.201300209>.
 13. Konstantinou M, Biedermann S, Kimber AC. Optimal designs for full and partial likelihood information—with application to survival models. *J Stat Plann Inference.* 2015;165:27–37. <https://doi.org/10.1016/j.jspi.2015.03.007>.
 14. Lopez-Fidalgo J, Rivaz-Lopez MJ. Optimal experimental designs for partial likelihood information. *Comp Stat Data Anal.* 2014;71:859–67. <https://doi.org/10.1016/j.csda.2012.10.009>.
 15. Carroll KJ. On the use and utility of the Weibull model in the analysis of survival data. *Control Clin Trials.* 2003;24(6):682–701. [https://doi.org/10.1016/S0197-2456\(03\)00072-2](https://doi.org/10.1016/S0197-2456(03)00072-2).
 16. Lawless JF. *Statistical models and methods for lifetime data.* 2nd ed. New York: Wiley; 2003.
 17. Xue X, Fedorov VV. Optimal design of experiments with the observation censoring driven by random enrollment of subjects. *J Stat Theory Prac.* 2017;11(1):163–78. <https://doi.org/10.1080/15598608.2016.1263808>.
 18. Atkinson AC, Fedorov VV, Herzberg AM, Zhang R. Elemental information matrices and optimal experimental design for generalized regression models. *J Stat Plann Inference.* 2012;144:81–91.
 19. Fedorov VV, Leonov SL. *Optimal design for nonlinear response model.* Boca Raton: Chapman & Hall/CRC Press; 2014.
 20. Kiefer J, Wolfowitz J. The equivalence of two extremum problems. *The. Can J Math.* 1960;12(0):363–6. <https://doi.org/10.4153/CJM-1960-030-4>.
 21. Fedorov VV, Hackl P. *Model-oriented design of experiments.* Berlin: Springer; 1997. <https://doi.org/10.1007/978-1-4612-0703-0>.
 22. Box GEP, Hunter WG. Sequential design of experiments for nonlinear models. In: Korth JJ, editor. *Proceedings of IBM Scientific Computing Symposium.* NY: IBM, White Plains; 1965.
 23. McCallum E, Bornkamp B. Accounting for parameter uncertainty in two-stage designs for phase II dose-response studies. In: Sverdlov O, editor. *Modern adaptive randomized clinical trials: statistical and practical aspects.* Boca Raton: Chapman & Hall/CRC Press; 2015. p. 428–49.
 24. Dragalin V, Fedorov VV, Yu W. Two-stage design for dose-finding that accounts for both efficacy and safety. *Statistics in Medicine.* 2008;27:5156–76.
 25. Dette H, Bornkamp B, Bretz F. On the efficiency of two-stage response-adaptive designs. *Stat Med.* 2013;32(10):1646–60. <https://doi.org/10.1002/sim.5555>.
 26. Lledó-García R, Hennig S, Nyberg J, Hooker AC, Karlsson MO. Ethically attractive dose-finding designs for drugs with a narrow therapeutic index. *J Clin Pharmacol.* 2012;52(1):29–38. <https://doi.org/10.1177/0091270010390041>.
 27. O’Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase I clinical trials in cancer. *Biometrics.* 1990;46(1):33–48. <https://doi.org/10.2307/2531628>.
 28. O’Quigley J, Reiner E. A stopping rule for the continual reassessment method. *Biometrika.* 1998;85(3):741–8. <https://doi.org/10.1093/biomet/85.3.741>.
 29. Heyd JM, Carlin B. Adaptive design improvements in the continual reassessment method for phase I studies. *Stat Med.* 1999;18(11):1307–21. [https://doi.org/10.1002/\(SICI\)1097-0258\(19990615\)18:11<1307::AID-SIM128>3.0.CO;2-X](https://doi.org/10.1002/(SICI)1097-0258(19990615)18:11<1307::AID-SIM128>3.0.CO;2-X).
 30. O’Quigley J. Continual reassessment designs with early termination. *Biostatistics.* 2002;3(1):87–99. <https://doi.org/10.1093/biostatistics/3.1.87>.
 31. Bretz F, Pinheiro J, Branson M. Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics.* 2005;61(3):738–48. <https://doi.org/10.1111/j.1541-0420.2005.00344.x>.
 32. Fedorov VV, Xue X. Survival models with censoring driven by random enrollment. In Kunert J, Müller CH, Atkinson AC, editors. *mODa 11—advances in model-oriented design and analysis.*: Springer; 2016. p. 95–102.
 33. Dodds MG, Hooker AC, Vicini P. Robust population pharmacokinetic experiment design. *J Pharmacokinetic Pharmacodyn.* 2005 February;32(1):33–64. <https://doi.org/10.1007/s10928-005-2102-z>.
 34. Ryznik Y, Sverdlov O, Hooker A. PAGE. [Online]; 2016. Available from: www.page-meeting.org/?abstract=5932.
 35. Sverdlov O, Rosenberger WF. Randomization in clinical trials: can we eliminate bias? *Clinical Investigation.* 2013;3(1):37–47. <https://doi.org/10.4155/cli.12.130>.
 36. Kuznetsova OM, Tymofyeyev Y. Brick tunnel randomization for unequal allocation to two or more treatment groups. *Stat Med.* 2011;30(8):812–24. <https://doi.org/10.1002/sim.4167>.
 37. Kuznetsova O, Tymofyeyev Y. Wide brick tunnel randomization—an unequal allocation procedure that limits the imbalance in treatment totals. *Stat Med.* 2014;33(9):1514–30. <https://doi.org/10.1002/sim.6051>.