

## OPTIMAL DESIGNS FOR TWO-PARAMETER NONLINEAR MODELS WITH APPLICATION TO SURVIVAL MODELS

Maria Konstantinou, Stefanie Biedermann and Alan Kimber

*University of Southampton*

*Abstract:* Censoring occurs in many industrial or biomedical ‘time to event’ experiments. Finding efficient designs for such experiments can be problematic since the statistical models involved are usually nonlinear, making the optimal choice of design parameter dependent. We provide analytical characterisations of locally  $D$ - and  $c$ -optimal designs for a class of models, thus reducing the numerical effort for design search substantially. We also investigate standardised maximin  $D$ - and  $c$ -optimal designs. We illustrate our results using the natural proportional hazards parameterisation of the exponential regression model. Different censoring mechanisms are incorporated and the robustness of designs against parameter misspecification is assessed.

*Key words and phrases:*  $c$ -optimality,  $D$ -optimality, proportional hazards, survival analysis.

### 1. Introduction

There is a large literature on optimal designs for nonlinear models but there is little on designs for models with potentially censored data. Ford, Torsney, and Wu (1992) consider optimal designs for nonlinear models where the response is distributed as a member of the exponential family and Sebastiani and Settini (1997) prove the optimality of these designs for a logistic regression model. Sitter and Torsney (1992) study  $D$ -optimal designs for generalised linear models with multiple design variables using the geometry of the design space in Ford, Torsney, and Wu (1992), and Sitter and Torsney (1995) consider  $D$ - and  $c$ -optimal designs for binary response models with two design variables. Neither paper considers the case where the data are subject to censoring.

Becker, McDonald, and Khoo (1989) find  $D$ -optimal designs for proportional hazards models with one or two parameters and specified baseline hazard. They use geometric arguments and empirical values for the hazard to investigate how censoring affects the  $D$ -optimal designs for different shapes of the design region. López-Fidalgo, Rivas-López, and Del Campo (2009) propose an algorithm to find  $D$ -optimal designs conditional on arrival time, where the design space is binary.

They consider a two-parameter exponential regression model that requires constraints on the parameters. For recent results on accelerated life testing see, for example, Wu, Lin, and Chen (2006) and McGree and Eccleston (2010).

Our research was motivated by the following problem. Let  $T_1, \dots, T_n$  be independent survival times of  $n$  subjects in an experiment with  $t_1, \dots, t_n$  the corresponding observed values. Let  $\alpha$  and  $\beta$  be the unknown model parameters. In survival models involving one explanatory variable,  $\alpha$  relates to the baseline hazard whereas  $\beta$  describes how the hazard varies with the explanatory variable. Let  $x_j \in \mathcal{X}$  be the experimental condition at which the  $j$ th observation is taken. In what follows, the design space  $\mathcal{X}$  is either binary,  $\mathcal{X} = \{0, 1\}$ , corresponding, for example, to two different treatments, or an interval,  $\mathcal{X} = [u, v]$ , corresponding, for example, to the doses of a drug.

The period of the experiment is the interval  $[0, c]$ . We consider two types of censoring. Under Type I censoring all subjects enter the study at the same time and are observed until time  $c$  or until failure, whichever is earlier; survival times greater than  $c$  are therefore right-censored. Under random censoring, the  $j$ th individual enters the study at a random time in  $[0, c]$ , independent of the survival time; the censoring time for the individual is random. The example we use to illustrate our general results is the exponential regression model in its proportional hazards parameterisation, naturally used in survival analysis (see, for example, Collett (2003)) that is specified by the probability density function  $f(t_j, x_j)$  with corresponding survivor function  $S(t_j, x_j)$ ,

$$f(t_j, x_j) = e^{\alpha + \beta x_j} e^{-t_j e^{\alpha + \beta x_j}}, \quad S(t_j, x_j) = e^{-t_j e^{\alpha + \beta x_j}}, \quad (t_j > 0). \quad (1.1)$$

This parameterisation avoids the need to specify constraints on the parameters.

Optimal design is concerned with finding the experimental conditions at which measurements should be taken in order to draw the most precise conclusions. We consider approximate designs of the form

$$\xi = \left\{ \begin{array}{c} x_1 \dots x_m \\ \omega_1 \dots \omega_m \end{array} \right\}, \quad 0 < \omega_i \leq 1, \quad \sum_{i=1}^m \omega_i = 1,$$

where the support points  $x_i$ ,  $i = 1, \dots, m$ ,  $m \leq n$ , are the distinct experimental conditions in the design, and the weights  $\omega_i$  represent the proportion of observations taken at the corresponding support point.

A recent trend in optimal design literature is to solve problems in more generality. Hedayat, Zhong, and Nie (2004) characterise  $D$ -optimal designs for a class of two-parameter models. But these results are not applicable to many models such as at (1.1). Yang and Stufken (2009) consider Loewner optimality and a more general class of models, showing that under some conditions, for

each given design there is always a design from a simple class that is better in the Loewner sense. These results were generalised to models with more than two parameters by Yang (2010). Depending on the model, however, the conditions can be difficult to verify, even with symbolic computational software.

We provide characterisations of  $D$ - and  $c$ -optimal designs under assumptions that are somewhat less restrictive and easier to verify than those in Yang and Stufken (2009), and which are satisfied by a large class of models, including (1.1), for the censoring schemes considered. In Section 2 we develop this approach for  $D$ -optimality. Section 3 contains the corresponding results for  $c$ -optimality when only the slope parameter  $\beta$  is of interest. The results are applied to model (1.1) with Type I and random censoring in Section 4. Section 5 provides analytical characterisations of the standardised maximin  $D$ - and  $c$ -optimal designs when a parameter space can be specified, even when the locally optimal designs are not available in closed form. In Section 6, we assess the robustness of locally optimal and parameter robust designs for (1.1) and compare their efficiency with traditional designs currently in use. A brief discussion is given in Section 7. The more technical proofs are in the Appendix.

## 2. $D$ -optimal Designs

A  $D$ -optimal design maximises the determinant of the Fisher information  $M(\xi, \alpha, \beta)$  with respect to the design, thereby minimising the volume of the confidence ellipsoid for the parameter estimators. A design  $\xi^*$  is  $D$ -optimal if

$$\xi^* = \arg \max_{\xi} |M(\xi, \alpha, \beta)|.$$

We consider two-parameter models with Fisher information of the form

$$M(\xi, \alpha, \beta) = \sum_{i=1}^m \omega_i I(x_i, \alpha, \beta) = \sum_{i=1}^m \omega_i Q(\theta_i) \begin{pmatrix} 1 & x_i \\ x_i & x_i^2 \end{pmatrix}, \quad (2.1)$$

where  $I(x_i, \alpha, \beta)$  is the Fisher information at the point  $x_i$  and  $\theta_i = \alpha + \beta x_i$ , satisfying the conditions (a)–(d) below. Following Ford, Torsney, and Wu (1992), an equivalent problem to maximising  $|M(\xi, \alpha, \beta)|$  is to maximise the determinant of this matrix with  $x_i$  replaced by  $\theta_i = \alpha + \beta x_i$ ,  $i = 1, \dots, m$ , where  $\beta \neq 0$ , also denoted by  $M(\xi, \alpha, \beta)$  in what follows. The parameter dependence of the design problem thus enters only via the transformed design space  $\Theta = \alpha + \beta \mathcal{X}$  where  $\beta \neq 0$ . For  $\beta = 0$ ,  $Q(\theta) = Q(\alpha)$  and we have the trivial case of a linear model. The assumptions are therefore given for  $\theta \in \mathbb{R}$ , so they are valid for all possible ranges for  $\Theta$ .

- (a) The function  $Q(\theta)$  at (2.1) is positive for all  $\theta \in \mathbb{R}$  and twice continuously differentiable.

- (b) The function  $Q(\theta)$  is strictly increasing on  $\mathbb{R}$ .
- (c) The second derivative  $g''(\theta)$  of the function  $g(\theta) = 2/Q(\theta)$  is an injective function.
- (d) For any  $s \in \mathbb{R}$ , the function  $r(\theta) = Q(\theta)(s - \theta)^2$  satisfies  $r'(\theta) = 0$  for exactly two values of  $\theta \in (-\infty, s]$ .

For the case of  $c$ -optimality we require an extra condition

- (d1) : The function  $\log Q(\theta)$  is concave for  $\theta \in \mathbb{R}$ .

This implies (d) given that (a) and (b) hold. Our assumptions hold, for example, for the Poisson, Gamma and Inverse Gamma regression models and for parametric proportional hazards models with a hazard function of the form

$$h(t) = e^\alpha g(t) e^{\beta x}, \quad (2.2)$$

where  $e^\alpha g(t)$  is the baseline hazard. Further, our assumptions hold (but those of Yang and Stufken (2009) do not) for certain accelerated failure time models with two failure modes where the type of failure time distribution differs between models, such as Gamma with shape parameter 2 and exponential depending on the sign of  $\theta$ .

To allow estimation of both parameters, a design must have at least two support points. For the binary design space  $\mathcal{X} = \{0, 1\}$  this means that both 0 and 1 are support points of the  $D$ -optimal design. From Lemma 5.1.3 in Silvey (1980), it then follows that the  $D$ -optimal design has equal weights.

For the rest of this section we consider interval design spaces  $\mathcal{X} = [u, v]$ . The locally  $D$ -optimal design for given  $\alpha$  and  $\beta$ , on an arbitrary interval  $[u, v]$ , can be obtained from the locally  $D$ -optimal design on the interval  $[0, 1]$  for parameter values  $\tilde{\alpha} = \alpha + \beta u$  and  $\tilde{\beta} = \beta(v - u)$  by transforming its support points  $x_i^*$  via  $z_i^* = u + (v - u)x_i^*$ . Therefore without loss of generality we take  $\mathcal{X} = [0, 1]$ . A tool for characterising  $D$ -optimal designs and for checking the  $D$ -optimality of a candidate design is the Equivalence Theorem (see, for example, Silvey (1980)).

**Theorem 1.** *A design  $\xi^*$  is  $D$ -optimal for a model with information matrix (2.1) if the inequality*

$$d(\xi^*, \alpha, \beta) = \text{tr}\{M^{-1}(\xi^*, \alpha, \beta)I(x, \alpha, \beta)\} \leq 2,$$

*holds for all  $x \in [0, 1]$ , with equality in the support points of  $\xi^*$ .*

From Caratheodory's Theorem (see, for example, Silvey (1980)), there exists a  $D$ -optimal design with at most three support points. Lemma 1 shows that this number can be further reduced.

**Lemma 1.** *If  $\beta \neq 0$  and assumptions (a)–(c) are satisfied, then the  $D$ -optimal design for a model with Fisher information (2.1) is unique and has two equally weighted support points.*

The proof of Lemma 1 is in the Appendix.

**Theorem 2.** *Let assumptions (a)–(d) be satisfied.*

(a) *If  $\beta > 0$ , the design*

$$\xi^* = \left\{ \begin{array}{cc} x_1^* & 1 \\ 0.5 & 0.5 \end{array} \right\}$$

*is  $D$ -optimal on  $\mathcal{X}$ , where  $x_1^* = 0$  if  $\beta < 2Q(\alpha)/Q'(\alpha)$ ; if not  $x_1^*$  is the unique solution of the equation  $\beta(x_1 - 1) + 2Q(\alpha + \beta x_1)/Q'(\alpha + \beta x_1) = 0$ .*

(b) *If  $\beta < 0$ , the design*

$$\xi^* = \left\{ \begin{array}{cc} 0 & x_2^* \\ 0.5 & 0.5 \end{array} \right\}$$

*is  $D$ -optimal on  $\mathcal{X}$ , where  $x_2^* = 1$  if  $\beta > -2Q(\alpha + \beta)/Q'(\alpha + \beta)$ ; if not  $x_2^*$  is the unique solution of the equation  $\beta x_2 + 2Q(\alpha + \beta x_2)/Q'(\alpha + \beta x_2) = 0$ .*

Theorem 2 (proved in the Appendix) provides a complete classification of  $D$ -optimal designs. Depending on some easily verifiable conditions on the parameters, the design problem has been either reduced to an optimisation problem in one variable or solved entirely.

### 3. $c$ -optimal Designs

Interest often centers on estimating  $\beta$  while treating  $\alpha$  as a nuisance parameter. For example, at (1.1)  $\beta$  is a log hazard ratio. Then an appropriate optimality criterion is  $c$ -optimality for  $\beta$  which minimises the asymptotic variance of the maximum likelihood estimator  $\hat{\beta}$ . A design  $\xi^*$  is  $c$ -optimal for  $\beta$  if  $(0 \ 1)^T \in \text{range}(M(\xi^*, \alpha, \beta))$  and

$$\xi^* = \underset{\xi}{\text{argmin}} (0 \ 1) M^{-}(\xi, \alpha, \beta) \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad (3.1)$$

where  $M^{-}$  is a generalised inverse of the matrix  $M$ .

**Lemma 2.** *For any real  $\alpha$ ,  $\beta \neq 0$  and any model with Fisher information (2.1) there exists a  $c$ -optimal design for  $\beta$  with exactly two support points.*

From Pukelsheim and Torsney (1991), we obtain an expression for the optimal weights. A  $c$ -optimal design  $\xi^*$  for  $\beta$  with support points  $x_1^*$  and  $x_2^*$  is

$$\xi^* = \left\{ \begin{array}{cc} x_1^* & x_2^* \\ \frac{\sqrt{Q(\alpha + \beta x_2^*)}}{\sqrt{Q(\alpha + \beta x_1^*)} + \sqrt{Q(\alpha + \beta x_2^*)}} & \frac{\sqrt{Q(\alpha + \beta x_1^*)}}{\sqrt{Q(\alpha + \beta x_1^*)} + \sqrt{Q(\alpha + \beta x_2^*)}} \end{array} \right\}. \quad (3.2)$$

The design problem for  $\mathcal{X} = \{0, 1\}$  has thus been solved completely. It remains to find the optimal support points when  $\mathcal{X} = [u, v] \subset \mathbb{R}$ .

**Theorem 3.** *Let assumptions (a), (b) and (d1) be satisfied.*

- (a) *If  $\beta > 0$  the design  $\xi^*$ , with support points  $x_1^*$  and  $v$  and the optimal weights given in (3.2), is  $c$ -optimal for  $\beta$ , where  $x_1^* = u$  if*

$$\beta(u - v) + \frac{2Q(\alpha + \beta u)}{Q'(\alpha + \beta u)} \left( 1 + \frac{\sqrt{Q(\alpha + \beta u)}}{\sqrt{Q(\alpha + \beta v)}} \right) > 0. \quad (3.3)$$

*Otherwise  $x_1^*$  is the unique solution of*

$$\beta(x_1 - v) + \frac{2Q(\alpha + \beta x_1)}{Q'(\alpha + \beta x_1)} \left( 1 + \frac{\sqrt{Q(\alpha + \beta x_1)}}{\sqrt{Q(\alpha + \beta v)}} \right) = 0. \quad (3.4)$$

- (b) *If  $\beta < 0$  the design  $\xi^*$ , with support points  $u$  and  $x_2^*$  and the optimal weights given in (3.2), is  $c$ -optimal for  $\beta$ , where  $x_2^* = v$  if*

$$\beta(u - v) - \frac{2Q(\alpha + \beta v)}{Q'(\alpha + \beta v)} \left( 1 + \frac{\sqrt{Q(\alpha + \beta v)}}{\sqrt{Q(\alpha + \beta u)}} \right) < 0.$$

*Otherwise  $x_2^*$  is the unique solution of*

$$\beta(u - x_2) - \frac{2Q(\alpha + \beta x_2)}{Q'(\alpha + \beta x_2)} \left( 1 + \frac{\sqrt{Q(\alpha + \beta x_2)}}{\sqrt{Q(\alpha + \beta u)}} \right) = 0.$$

#### 4. Application to an Exponential Regression Model

We apply the previous results to model (1.1) for an interval design space. We briefly discuss the special case of no censoring, corresponding to  $c = \infty$ , a study running for as long as necessary to record all survival times. From (1.1), the log-likelihood at  $x_j$  is  $l(\alpha, \beta, x_j) = \alpha + \beta x_j - t_j e^{\alpha + \beta x_j}$  and thus the Fisher information at the point  $x_j$  is

$$I(x_j, \alpha, \beta) = \begin{pmatrix} 1 & x_j \\ x_j & x_j^2 \end{pmatrix},$$

since  $E(T_j) = 1/e^{\alpha + \beta x_j}$ . In this case the Fisher information is the same as for the linear model for iid errors. The  $D$ -optimal design for this model is equally supported at the end points of the design space  $\mathcal{X}$  (see, for example, Atkinson, Donev, and Tobias (2007)) and coincides with the  $c$ -optimal design for  $\beta$ .

### 4.1. Type I censoring

In Type I censoring the censoring time  $c$  is fixed and common for all individuals. This occurs, for example, when all individuals have been recruited at the same time to a study of duration  $c$ . If the event of interest has not occurred by the end of the study the observation is right-censored. Let  $Y_j = \min\{T_j, c\}$  be the  $j$ th possibly censored observation and let  $T_j$  follow model (1.1). Then

$$E(Y_j) = \int_0^c ye^{\alpha+\beta x_j} e^{-ye^{\alpha+\beta x_j}} dy + cP(Y_j = c) = \frac{(1 - e^{-ce^{\alpha+\beta x_j}})}{e^{\alpha+\beta x_j}}, \quad (4.1)$$

and the log-likelihood at  $x_j$  is  $l(\alpha, \beta, x_j) = \delta_j(\alpha + \beta x_j) - y_j e^{\alpha+\beta x_j}$ , where  $\delta_j$  is an event indicator which is zero if  $y_j$  is a censored observation and unity otherwise. Hence the Fisher information at  $x_j$  is

$$I(x_j, \alpha, \beta) = (1 - e^{-ce^{\alpha+\beta x_j}}) \begin{pmatrix} 1 & x_j \\ x_j & x_j^2 \end{pmatrix},$$

which yields (2.1) with  $Q(\theta) = (1 - e^{-ce^\theta})$ . It can be shown that assumptions (a)–(d) and (d1) hold here. Hence Theorems 2 and 3 hold for Type I censoring.

### 4.2. Random censoring

Random censoring occurs, for example, if the  $j$ th individual enters the study at random time  $Z_j \in [0, c]$ , where  $Z_j$  is independent of the survival time  $T_j$ , so the censoring time  $C_j = c - Z_j$  for this individual is random. We assume that  $Z_1, \dots, Z_n$  follow a uniform distribution on  $[0, c]$ , thus  $C_1, \dots, C_n$  also have a uniform distribution on  $[0, c]$  with probability density function  $f_c(c_j) = 1/c$ . We observe  $Y_j = \min\{T_j, C_j\}$  where  $E(Y_j|C_j = c_j)$  is given by the right hand side of (4.1) with  $c$  replaced by  $c_j$ . Thus

$$\begin{aligned} E(Y_j) &= E(E(Y_j|C_j = c_j)) = \int_0^c \frac{(1 - e^{-c_j e^{\alpha+\beta x_j}})}{c_j e^{\alpha+\beta x_j}} dc_j \\ &= \frac{(ce^{\alpha+\beta x_j} + e^{-ce^{\alpha+\beta x_j}} - 1)}{ce^{2(\alpha+\beta x_j)}}, \end{aligned}$$

and the log-likelihood at  $x_j$  is  $l(\alpha, \beta, x_j) = \delta_j(-\log c + \alpha + \beta x_j) - y_j e^{\alpha+\beta x_j}$ . Hence the Fisher information at point  $x_j$  is

$$I(x_j, \alpha, \beta) = \frac{(ce^{\alpha+\beta x_j} + e^{-ce^{\alpha+\beta x_j}} - 1)}{ce^{\alpha+\beta x_j} \begin{pmatrix} 1 & x_j \\ x_j & x_j^2 \end{pmatrix}}.$$

Again this is of the form (2.1) for  $Q(\theta) = 1 + (e^{-ce^\theta} - 1)/ce^\theta$  and assumptions (a)–(d) and (d1) hold.

For  $\beta > 0$  ( $< 0$ ) these  $Q$ -functions are increasing (decreasing) with  $x$ . Therefore from (3.2) the  $c$ -optimal weight corresponding to the smaller support point is greater (smaller) than the other weight if  $\beta > 0$  ( $< 0$ ). This means, for example, that more patients are allocated to the more effective dose.

## 5. Standardised Optimal Designs

The optimal designs found depend on model parameters that are unknown in practice. Nevertheless, in many practical situations some information about the parameter values can be provided by the experimenter. For example,  $\alpha$  may determine the baseline hazard for a standard treatment. Hence precise knowledge of its value might be available, whereas for  $\beta$  the experimenter can specify a range of values for a clinically significant improvement with new treatment. We further assume that the experimenter has no preference for specific  $\beta$ -values and that the total duration of the study,  $c$ , is known.

Following Dette (1997) we seek designs that maximise the worst efficiencies with respect to the locally optimal designs over a range of parameter values. This allows us to construct robust designs that protect against the worst case scenario. Dette and Sahn (1998) compare a standardised and a nonstandardised maximum variance optimality criterion and show that in some cases the optimal designs based on the latter criterion can be inefficient. A design  $\xi^*$  maximising

$$\Phi(\xi) = \min \left\{ \frac{|M(\xi, \alpha, \beta)|}{|M(\xi_\beta^*, \alpha, \beta)|} \mid \beta \in [\beta_0, \beta_1] \right\} \quad (5.1)$$

is called a standardised maximin  $D$ -optimal design, and a design  $\xi^*$  maximising

$$\Phi(\xi) = \min \left\{ \frac{\begin{pmatrix} 0 & 1 \end{pmatrix} M^{-1}(\xi_\beta^*, \alpha, \beta) \begin{pmatrix} 0 \\ 1 \end{pmatrix}}{\begin{pmatrix} 0 & 1 \end{pmatrix} M^{-1}(\xi, \alpha, \beta) \begin{pmatrix} 0 \\ 1 \end{pmatrix}} \mid \beta \in [\beta_0, \beta_1] \right\} \quad (5.2)$$

is called a standardised maximin  $c$ -optimal design for  $\beta$ , where  $\xi_\beta^*$  is the locally optimal design. Criteria (5.1) and (5.2) seek a design that maximises the worst  $D$ -efficiency and  $c$ -efficiency respectively, given by

$$eff_D(\xi) = \left( \frac{|M(\xi, \alpha, \beta)|}{|M(\xi_\beta^*, \alpha, \beta)|} \right)^{1/2}, \quad (5.3)$$

$$eff_c(\xi) = \frac{\begin{pmatrix} 0 & 1 \end{pmatrix} M^{-1}(\xi_\beta^*, \alpha, \beta) \begin{pmatrix} 0 \\ 1 \end{pmatrix}}{\begin{pmatrix} 0 & 1 \end{pmatrix} M^{-1}(\xi, \alpha, \beta) \begin{pmatrix} 0 \\ 1 \end{pmatrix}}. \quad (5.4)$$

For a binary design space the locally  $D$ -optimal design is equally supported at 0 and 1 for any parameter values, so no further investigation need be done. For



an interval design space  $\mathcal{X} = [0, 1]$ , the following theorem provides an analytical characterisation of the standardised maximin  $D$ -optimal two point design for a given range of negative  $\beta$ -values; its proof is given in the Appendix.

**Theorem 4.** *Let  $\beta \in [\beta_0, \beta_1]$  where  $\beta_1 < 0$ ,  $\alpha$  be fixed, and assumptions (a)-(d) and (d1) be satisfied. The standardised maximin  $D$ -optimal two-point design is equally supported at points 0 and  $x_2^*$  where  $x_2^* = 1$  if  $\beta_0 > -2Q(\alpha + \beta_0)/Q'(\alpha + \beta_0)$ . Otherwise  $x_2^*$  is the solution of*

$$Q(\alpha + \beta_0 x)Q(\alpha + \beta_1 x(\beta_1))x(\beta_1)^2 = Q(\alpha + \beta_1 x)Q(\alpha + \beta_0 x(\beta_0))x(\beta_0)^2, \quad (5.5)$$

where  $x(\beta_0)$ ,  $x(\beta_1)$  are the solutions of the equation  $\beta x + 2Q(\alpha + \beta x)/Q'(\alpha + \beta x) = 0$ ,  $0 < x \leq 1$ , for  $\beta_0$  and  $\beta_1$ , respectively.

Note that Theorem 4 applies when  $\beta < 0$ . The proof used in this case is not applicable when  $\beta > 0$  and this is a topic for further investigation.

As shown in Section 3, the locally  $c$ -optimal designs for  $\beta$  depend on the model parameters. Theorem 5, which is proven in the Appendix, gives an analytical characterisation of the standardised maximin  $c$ -optimal design for  $\beta$ , for a binary design space.

**Theorem 5.** *Let  $\beta \in [\beta_0, \beta_1]$ ,  $\alpha$  be fixed, and assumptions (a), (b) and (d1) be satisfied. If  $\mathcal{X} = \{0, 1\}$ , the standardised maximin  $c$ -optimal two-point design is*

$$\xi^* = \left\{ \begin{array}{cc} 0 & 1 \\ \omega^* & 1 - \omega^* \end{array} \right\},$$

where  $\omega^* = (\omega(\beta_0) + \omega(\beta_1))/2$ , and  $\omega(\beta_0)$  ( $\omega(\beta_1)$ ) is the optimal weight on zero for the locally  $c$ -optimal design for  $\beta$  given in (3.2) for  $\beta_0$  ( $\beta_1$ ).

## 6. Robustness Analysis

We assess the robustness of our designs by calculating their efficiency if the parameters have been misspecified. As a starting point we used the maximum likelihood estimates for  $\alpha$  and  $\beta$  from the Freireich data (see Collett (2003)), -2.163 and -1.526 respectively, and  $c = 30$ . To compare the performance of an arbitrary design  $\xi$  to a locally  $D$ -optimal design  $\xi^*$ , we used the  $D$ -efficiency (5.3), whereas for the comparison of  $\xi$  to a locally  $c$ -optimal design  $\xi^*$  we used the  $c$ -efficiency (5.4). Type I censoring is assumed.

### 6.1. Locally $D$ -optimal designs

We considered locally  $D$ -optimal designs for the vector of parameter values  $\gamma = (\alpha, \beta)$ . The value of the maximum likelihood estimator for  $\alpha$  was

Table 1.  $D$ -efficiencies for some selected locally  $D$ -optimal designs.

Parameter vector	Design			
	$\xi_{\gamma_0}$	$\xi_{\gamma_1}$	$\xi_{\gamma_2}$	$\xi_{\gamma_3}$
$\gamma_0 = (-2.163, -0.1)$	1	1	1	0.900
$\gamma_1 = (-2.163, -0.405)$	1	1	1	0.905
$\gamma_2 = (-2.163, -1.526)$	1	1	1	0.946
$\gamma_3 = (-2.163, -2.623)$	0.992	0.992	0.992	1

Table 2.  $c$ -efficiencies for some selected locally  $c$ -optimal designs.

Parameter vector	Weight on 0	Design			
		$\xi_{\gamma_0}$	$\xi_{\gamma_1}$	$\xi_{\gamma_2}$	$\xi_{\gamma_3}$
$\gamma_0$	0.498	1	0.9998	0.9782	0.8772
$\gamma_1$	0.491	0.9998	1	0.9824	0.8864
$\gamma_2$	0.425	0.9787	0.9828	1	0.9552
$\gamma_3$	0.323	0.8908	0.8991	0.9597	1

used, whereas the  $\beta$ -values were chosen to have small, medium and large treatment effect. Table 1 gives the parameter vectors used and the corresponding  $D$ -efficiencies of the locally  $D$ -optimal designs when the parameter values were misspecified.

For the first three sets of parameter values, the locally  $D$ -optimal design is the standard design supported at 0 and 1 with equal weights, whereas  $\xi_{\gamma_3}$  is equally supported at 0 and 0.9. The standard design has high  $D$ -efficiency for all values of the parameter vectors. Here  $\xi_{\gamma_3}$  seems to be a good alternative to the standard design if, for example, the experimenter does not want to expose patients to the highest drug doses.

## 6.2. Locally $c$ -optimal designs

For the parameter vectors used in Section 6.1, their locally  $c$ -optimal designs have support points 0 and 1. The weights corresponding to point 0 were found using (3.2) and are shown in Table 2 along with the  $c$ -efficiencies of each of these designs when the parameter values are misspecified.

The locally  $c$ -optimal designs have high  $c$ -efficiencies for all four sets of parameter values. The design  $\xi_{\gamma_2}$ , locally  $c$ -optimal for a parameter value near the center of the parameter space, has a lowest efficiency of 0.9597 and hence is more robust than the other three designs.

## 6.3. Standardised maximin optimal designs

We can find the standardised maximin  $D$ - and  $c$ -optimal designs for the range of  $\beta$ -values used above, denoted by  $\xi_{\gamma_4}$  in both cases. Although we consider the

case of an interval design space, the locally  $c$ -optimal designs found in Section 6.2 are supported at points 0 and 1 and so Theorem 5 can be used.

The standardised maximin  $D$ -optimal design is supported at 0 and 0.993, with equal weights, and is locally  $D$ -optimal for  $\gamma_4 = (-2.163, -2.380)$ , whereas the standardised maximin  $c$ -optimal design allocates 41.1% of the observations at the experimental point 0 and the rest at point 1, and is locally  $c$ -optimal for  $\gamma_4 = (-2.163, -1.690)$ . The minimum (median) efficiencies are 0.993 (0.993) for the  $D$ -optimal design and 0.969 (0.974) for the  $c$ -optimal design. For both designs the minimum efficiencies are obtained at  $\gamma_0$  and  $\gamma_3$ .

#### 6.4. Cluster designs

This is a modification (see Biedermann and Woods (2011)) of the method introduced by Dror and Steinberg (2006). For each of 1,000 parameter vectors, found by drawing 1,000  $\beta$ -values from a uniform distribution on the interval from  $-2.623$  to  $-0.1$ , the locally  $D$ - and  $c$ -optimal designs were obtained. Then a clustering algorithm was applied where the cluster centroids were chosen as support points and each weight was chosen proportional to the corresponding cluster size, reflecting the relative importance of each cluster.

The number of clusters for the  $D$ -optimal designs was chosen to vary from 2 to 6 and, for each value, the  $D$ -efficiency of a cluster design was calculated via (5.3) relative to each of the 1,000 locally  $D$ -optimal designs. The two-point cluster design was equally supported at 0 and 1 whereas the rest of the cluster designs with more than two support points allocated half the observations at point 0, very little weight at points other than 0 and 1 and the rest at point 1. The minimum and median efficiencies were found to be the same for all the cluster designs (0.993 and 0.997 respectively) and this may be a result of the low weight that all cluster designs gave to experimental points other than 0 and 1.

The support points of the 1,000 locally  $c$ -optimal designs were always 0 and 1, hence the cluster design had the same. The clustering here was applied to *design* points, rather than *support* points as the support points of the locally  $c$ -optimal points have differing weights. The resulting cluster design allocated 43% of the observations to 0 and the rest to 1, and performed well as the minimum (median) efficiencies found via (5.4) were 0.956 (0.990).

#### 6.5. Comparison of designs

We compare the performance of eleven designs: the locally  $D$ -optimal designs  $\xi_{\gamma_0}, \dots, \xi_{\gamma_3}$ , the standardised maximin  $D$ -optimal design  $\xi_{\gamma_4}$ , the cluster designs  $\xi_1, \dots, \xi_5$  and the equally spaced design  $\xi_0$  with support points 0, 0.5, 1 and equal weights. The  $D$ -efficiency (5.3) of each was calculated with respect to each of the 1,000 locally optimal designs and the results are summarised in Figure 1. Designs  $\xi_0$  and  $\xi_{\gamma_3}$  were omitted since they were clearly outperformed.

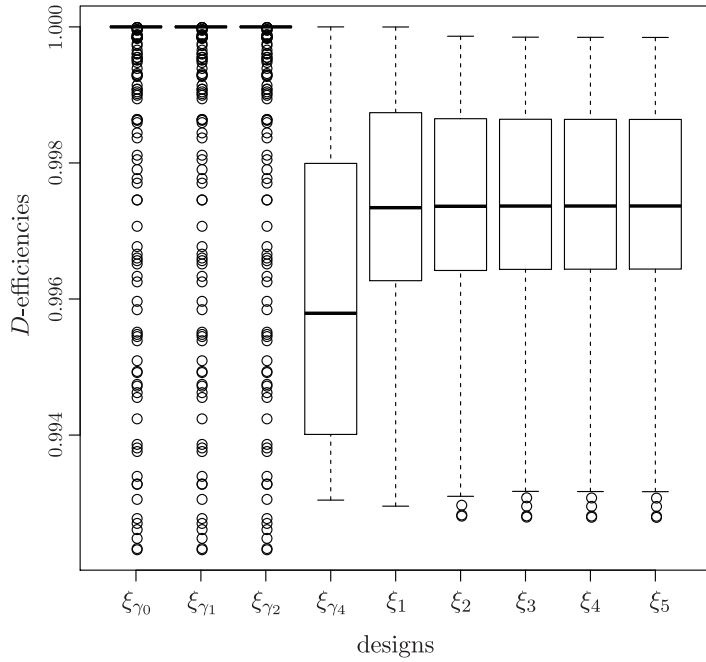


Figure 1. Boxplots of  $D$ -efficiencies calculated for 9 different designs for 1,000 parameter vectors.

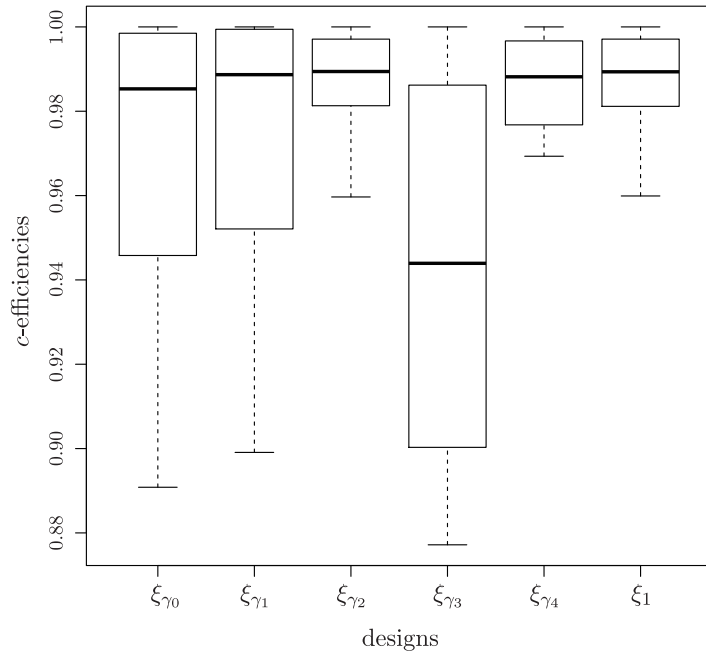


Figure 2. Boxplots of  $c$ -efficiencies calculated for 6 different designs for 1,000 parameter vectors.

Figure 1 shows that the standardised maximin  $D$ -optimal design  $\xi_{\gamma_4}$  has the highest minimum efficiency but lower median efficiency: there is a trade off between protecting against the worst case scenario and having a worse median efficiency. The cluster designs  $\xi_2, \dots, \xi_5$  with more than two support points are useful since they allow for linearity of the regression to be checked and do not perform worse than the two-point cluster design  $\xi_1$ . All cluster designs are good alternatives to locally optimal designs and perform similarly to the standardised maximin  $D$ -optimal design.

The locally  $c$ -optimal designs  $\xi_{\gamma_0}, \dots, \xi_{\gamma_3}$ , the standardised maximin  $c$ -optimal design  $\xi_{\gamma_4}$  and the two-point cluster design  $\xi_1$  are compared in Figure 2. Among the locally  $c$ -optimal designs only  $\xi_{\gamma_2}$  performs well across the parameter space. As for  $D$ -optimality, there is a trade off between best minimum efficiency and a lower median efficiency for the standardised maximin  $c$ -optimal design  $\xi_{\gamma_4}$ . Overall both  $\xi_{\gamma_4}$  and  $\xi_1$  are good alternatives to the locally optimal designs.

## 7. Discussion

Survival models used in applications are usually nonlinear, hence the optimal designs depend on the unknown model parameters. To overcome this difficulty robust designs must be constructed to perform well across a wide range of parameter values. A difficulty in finding optimal designs for these applications is that the data are often subject to censoring.

For models with Fisher information of the form (2.1) that satisfy assumptions (a)–(d) and (d1) we have provided a complete classification of locally  $D$ - and  $c$ -optimal designs. Our assumptions are somewhat less restrictive and easier to check than those provided by Yang and Stufken (2009) and are satisfied by many models. Our results were then applied to the proportional hazards parameterisation of the exponential regression model (1.1), for the cases of Type I and random censoring. Under some conditions on the parameters the optimal design is not the “standard design” supported at 0 and 1 with equal weights.

We have found optimal designs based on standardised maximin criteria, when there is some knowledge about the parameter values, that maximise the worst efficiency among all two-point designs. Cluster designs were built from locally optimal designs for a specific set of parameter values and their computation was facilitated by results for the locally optimal designs. In Section 6 we have shown that alternatives to the locally optimal designs are cluster designs that in some cases have more than two support points, thereby enabling the linearity of the regression function to be checked.

## References

- Atkinson, A. C., Donev, A. N. and Tobias, R. D. (2007). *Optimum Experimental Designs, with SAS*. Oxford University Press, Oxford.

- Becker, N., McDonald, B. and Khoo, C. (1989). Optimal designs for fitting a proportional hazards regression model to data subject to censoring. *Austral. J. Statist.* **31**, 449-468.
- Biedermann, S. and Woods, D. C. (2011). Optimal designs for generalized non-linear models with application to second-harmonic generation experiments. *Appl. Statist.* **60**, 281-299.
- Collett, D. (2003). *Modelling Survival Data in Medical Research*. Chapman and Hall/CRC.
- Dette, H. (1997). Designing experiments with respect to 'standardized' optimality criteria. *J. Roy. Statist. Soc. Ser. B* **59**, 97-110.
- Dette, H. and Salm, M. (1998). Minimax optimal designs in nonlinear regression models. *Statist. Sinica* **8**, 1249-1264.
- Dror, H. A. and Steinberg, D. M. (2006). Robust experimental design for multivariate generalized linear models. *Technometrics* **48**, 520-529.
- Ford, I., Torsney, B. and Wu, C. F. J. (1992). The use of a canonical form in the construction of locally optimal designs for non-linear problems. *J. Roy. Statist. Soc. Ser. B* **54**, 569-583.
- Hedayat, A. S., Zhong, J. and Nie, L. (2004). Optimal and efficient designs for 2-parameter nonlinear models. *J. Statist. Plann. Inference* **124**, 205-217.
- López-Fidalgo, J., Rivas-López, M. J. and Del Campo, R. (2009). Optimal designs for Cox regression. *Statist. Neerlandica* **63**, 135-148.
- McGree, J. M. and Eccleston, J. A. (2010). Investigating design for survival models. *Metrika* **72**, 295-311.
- Pukelsheim, F. and Torsney, B. (1991). Optimal weights for experimental designs on linearly independent support points. *Ann. Statist.* **19**, 1614-1625.
- Sebastiani, P. and Settini, R. (1997). A note on D-optimal designs for a logistic regression model. *J. Statist. Plann. Inference* **59**, 359-368.
- Silvey, S. D. (1980). *Optimal Designs: An Introduction to the Theory for Parameter Estimation*. Chapman and Hall, London.
- Sitter, R. R. and Torsney, B. (1992). D-optimal designs for generalized linear models. Technical Report No. 211, Laboratory for Research in Statistics and Probability, Carleton University, Ottawa, Canada.
- Sitter, R. R. and Torsney, B. (1995). Optimal designs for binary response experiments with two design variables, *Statist. Sinica* **5**, 405-419.
- Wu, S. J., Lin, Y. P. and Chen, Y. J. (2006). Planning step-stress life test with progressively type I group-censored exponential data. *Statist. Neerlandica* **60**, 46-56.
- Yang, M. (2010). On the de la Garza phenomenon. *Ann. Statist.* **38**, 2499-2524.
- Yang, M. and Stufken, J. (2009). Support points of locally optimal designs for nonlinear models with two parameters. *Ann. Statist.* **37**, 518-541.
- Southampton Statistical Sciences Research Institute, University of Southampton, University Road, Southampton SO17 1BJ, United Kingdom.  
E-mail: mk21g09@southampton.ac.uk
- Southampton Statistical Sciences Research Institute, University of Southampton, University Road, Southampton SO17 1BJ, United Kingdom.  
E-mail: S.Biedermann@southampton.ac.uk
- Southampton Statistical Sciences Research Institute, University of Southampton, University Road, Southampton SO17 1BJ, United Kingdom.  
E-mail: A.C.Kimber@southampton.ac.uk

(Received November 2011; accepted January 2013)