Model robust designs for survival trials

Maria Konstantinou, Stefanie Biedermann and Alan Kimber

Abstract

The exponential-based proportional hazards model is often assumed in time-to-event experiments but may only approximately hold. Deviations in different neighbourhoods of this model are considered that include other widely used parametric proportional hazards models and the data are assumed to be subject to censoring. Minimax designs are then found explicitly, based on criteria corresponding to classical c- and D-optimality. Analytical characterisations of optimal designs are provided which, unlike optimal designs for related problems in the literature, have finite support and thus avoid the issues of implementing a density-based design in practice. Finally, the proposed designs are compared with the balanced design that is traditionally used in practice, and recommendations for practitioners are given.

Key words:

proportional hazards models, minimax optimal designs, D-optimality, c-optimality, Type-I censoring

1. Introduction

Optimal experimental designs are often constructed assuming that the model generating the data is known, up to the values of the parameters involved. In many practical situations, however, the proposed parametric model may only be approximately true and thus may cause the vector of parameter estimators to be biased. As illustrated by Box and Draper (1959) for the case of a linear regression model, the advantages of using an optimal design that minimises just the variance are lost even if the deviations from the assumed model are small.

Following Box and Draper (1959), robust designs for approximately linear regression have been constructed by Wiens (1992) based on classical optimality criteria but involving the mean squared error matrix. He finds minimax designs which are optimal in that they minimise the criteria functions for the worst possible deviation from the linear regression model. Prediction and extrapolation problems with possible heteroscedasticity are studied by Wiens (1998) and Fang and Wiens (1999) respectively among others. Sinha and Wiens (2002) consider the construction of sequential designs which are robust against model uncertainty for nonlinear models.

Further results on misspecified nonlinear regression include Woods et al. (2006), Wiens and Xu (2008) and Xu (2009a) for prediction and extrapolation problems.

However, none of these authors considers the case where the data are subject to censoring. This arises in many time-to-event experiments when a particular event of interest is not observed for some of the subjects utilised in the experiment. Censoring is often a result of the fact that the experiments are not run as long as necessary in order to obtain complete data, that is, event times for all the subjects, because of time and cost limitations. Therefore, it is of interest to find optimal designs which are robust to misspecifications of the assumed model and which allow for the possibility of the data being censored.

The available literature on model robust designs for time-to-event data is focused on accelerated life tests for which the subjects are put under extreme conditions in order for the event of interest to occur sooner than under normal circumstances. In this case, extrapolation to lower covariate values and prediction problems is often of interest; see, for example, Pascual and Montepiedra (2003), Xu (2009b) and McGree and Eccleston (2010).

An alternative class of models used for the modelling of time-to-event data is studied, namely that of proportional hazards models. Such models satisfy the proportional hazards assumption of constant hazard ratio over time and are frequently used in practice because of the simple interpretation of the regression coefficients in terms of hazard ratios. When a specific distribution is assumed for the event times, the resulting parametric models is referred to as distribution-based proportional hazards models. Cox's proportional hazards model, on the other hand, leaves the underlying distribution unspecified and therefore inference is based on the partial likelihood function (see Collett (2003) for further details).

Konstantinou et al. (2015) consider Cox's model and show that in the presence of Type-I censoring an exponential distribution can be assumed without greatly affecting the optimal choice of design for partial likelihood estimation. They also find that the full and partial likelihood approaches result in very similar designs for the same assumed model.

Following these findings, small deviations in a neighbourhood of the exponential-based proportional hazards model are considered. The model uncertainty is formulated via a contamination function and the data are assumed to be subject to Type-I censoring. Then following along the same lines as in Xu (2009b), both censoring and model uncertainty are incorporated to obtain the asymptotic properties of the maximum likelihood estimator. Based on the asymptotic mean squared error matrix, minimax optimal designs for full likelihood estimation are constructed which protect against the worst possible misspecification of the assumed exponential model. Note that Xu (2009b) considers various prediction and extrapolation problems for normally distributed data and investigates the construction of designs that are continuous with respect to the Lebesgue measure. However, the focus of the present paper is on designs with finite support. This allows for explicit solutions to be obtained and then compared with the corresponding results of Konstantinou et. al. (2014) for the case of the assumed model being true.

In Section 2 the assumed and true models considered are introduced and two different classes of contamination functions are defined to account for the various forms of the true distribution for the data. Then in Section 3 the asymptotic properties of the maximum likelihood estimator for the parameter vector are derived under model uncertainty and Type-I censoring. Analytical characterisations of minimax c- and D-optimal designs are given in Section 4. These designs are found using criteria corresponding to the classical c- and D-optimality criteria but are based on the mean squared error matrix rather than just the information matrix. In Section 5 the behaviour of the proposed designs is illustrated and they are compared with the balanced design traditionally used in practice. Finally, the main conclusions are discussed in Section 6.

2. Models and contamination functions

Time-to-event experiments are usually conducted in order to evaluate a particular intervention or treatment. Therefore, in what follows the focus is on models that involve one explanatory variable x. The mean squared error matrix for general designs is derived, and then design search is illustrated for the situation in which x takes values in the binary design space denoted by $\mathcal{X} = \{0, 1\}$, corresponding, for example, to a placebo and an active treatment in a clinical trial.

The aim of the experiment is assumed to be the estimation of one or both of the two model parameters. Let c be the predetermined duration of the experiment at which point the observations of subjects for which the event of interest has not occurred are said to be right-censored. Possibly censored data are summarised mainly using the hazard function which expresses the risk of the event of interest occurring at any time after the commencement of the experiment (Collett (2003)).

Consider the situation where the experimenter assumes the exponential-based proportional hazards model specified by the hazard function

$$h_1(t) = \exp\{\alpha + \beta x\}, \quad t > 0, \ x \in \mathcal{X} \subseteq I\!R, \tag{1}$$

where α and β are real parameters, when in fact this is only an approximation to the true underlying model. Denote the hazard function of the unknown true model by

$$h_2(t) = \exp\left\{\alpha + \beta x + \frac{g(t)}{\sqrt{n}}\right\}, \quad t > 0, \ x \in \mathcal{X} \subseteq I\!R, \ g(t) \in \mathcal{G}, \tag{2}$$

where n denotes the sample size. The function $g(t)$ represents uncertainty about the exact form of the underlying distribution for the data and, following the literature, it is called the contamination function or just the contaminant. It is assumed that $g(t)$ is unknown and ranges in a neighbourhood specified by the class $\mathcal{G}$.

The parametrisation in (2) allows one to remain within a proportional hazards framework and ensures that the model parameters are well defined. In particular, unlike the existing literature, see, for example, Wiens (1992), the contamination function is independent of the covariate value x. Therefore, the parameter β corresponds to the effect of the explanatory variable. For identifiability reasons it is further required that $g(t)$ does not involve an additive constant. If this were not the case, the constant term would be absorbed in the quantity $\exp\{\alpha\}$ that represents the baseline hazard for model (1), that is, the hazard function for a subject with $x = 0$.

The factor $n^{-1/2}$ is included so that the deviations are of the order $O(1/\sqrt{n})$, resulting in models that are in a neighbourhood of the exponential model (1). At the same time, the dependence of g on the time t ensures that the general form of the true model includes widely used parametric proportional hazards models, such as, for example, the Weibull and Gompertz distributions with known shape parameter γ. These distributions correspond to the cases where $g(t)$ is equal to $(\gamma - 1)\log t$ and γt respectively.

Two classes of contamination functions are defined which allow various forms of g, including those that correspond to the Weibull and Gompertz distributions. With the exception of Li and Notz (1982), the existing literature on model robustness considers the construction of designs that are absolutely continuous with respect to the Lebesgue measure. On the other hand, the formulation of the classes considered ensures the use of designs with finite support on the design space $\mathcal{X}$. This allows for exact solutions to be obtained which can then be compared with the corresponding solutions one would have in the case of the assumed model being true (see Section 4).

The first class of contaminants under study is specified by

$$\mathcal{G}_1 = \left\{ g : \max_{t \in [0,c]} |g(t)| \leq c_1 \right\}, \tag{3}$$

where c_1 is a specified positive constant. This class includes contamination functions $g(t)$ which are bounded on the time interval $[0, c]$ and is also used in Li and Notz (1982). They, however, considered extrapolation and interpolation problems for linear regression models with complete data.

Now consider the case of unbounded contamination functions such as $g(t) = (\gamma - 1)\log t$ for which $\lim_{t\to 0} g(t) = -\infty$. A class that can be used to include such contaminants is

$$\mathcal{G}_2 = \left\{ g : \left| \int_0^c e^{-te^{\alpha+\beta x}} g(t)\, dt \right| \leq c_2, \quad \forall x \in \mathcal{X} \text{ and } \left| \int_0^c g(s)\, ds \right| < \infty \right\}, \tag{4}$$

where, as before, c_2 is a specified positive constant. This class is defined so that the integral expression involved in the asymptotic expectation of the score function $b(\xi, g)$, evaluated in Section 3, is bounded in the design space $\mathcal{X}$.

3. Estimation under model uncertainty and Type-I censoring

For the estimation of the model parameters the full likelihood approach is adopted since the assumed parametric model is completely specified as the exponential-based proportional hazards model. Throughout this section a similar procedure to that used in Xu (2009b) is adopted in order to incorporate both censoring and model misspecification in the maximum likelihood estimation method and to obtain the asymptotic properties of the resulting estimator vector.

Let $T_1, \ldots, T_n$ be the independent random variables indicating the times to the occurrence of the event of interest for the n subjects utilised in the experiment with corresponding observed values $t_1, \ldots, t_n$. Under the right-censored data scenario considered, Type-I censoring corresponds to the case where all the subjects enter the experiment at the same time, indicated by zero, and so the censoring time c is common for all the subjects. This situation occurs commonly in reliability applications. Alternatively, as is more common in clinical studies, subjects may enter the study at different calendar times but each be followed up for c time units. Therefore, in the presence of Type-I censoring, what is actually observed are the values y_j of the random variables $Y_j = \min\{T_j, c\}$, $j = 1, \ldots, n$. This is formulated using an indicator variable δ_j that is equal to unity if observation j is an event time and zero if it is a right-censored observation. That is,

$$\delta_j = \begin{cases} 1, & \text{if } Y_j = T_j \\ 0, & \text{if } Y_j = c \end{cases} .$$

Note that $\delta_j \sim Bin(1, P_j)$, where $P_j = P(\delta_j = 1) = P(Y_j = T_j)$.

The likelihood function for possibly censored data involves the survivor and probability density functions which are defined in terms of the hazard function as

$$S_k(y) = \exp\left\{ -\int_0^y h_k(s)\, ds \right\} \quad \text{and} \quad f_k(y) = h_k(y) S_k(y), \qquad y \in [0, c]; k = 1, 2,$$

respectively (Collett (2003)). Therefore, assuming that model (1) is correct, the corresponding log-likelihood function of the jth observation y_j with covariate value x_j is given by

$$\begin{aligned} l := l(x_j, \alpha, \beta) &= \delta_j \log f_1(y_j) + (1 - \delta_j) \log S_1(c) \\ &= \delta_j \left(\alpha + \beta x_j - y_j e^{\alpha + \beta x_j} \right) - (1 - \delta_j) c e^{\alpha + \beta x_j}. \end{aligned}$$

To find the limiting properties of the maximum likelihood estimator for the vector of model parameters requires the evaluation of the asymptotic expectation and variance-covariance matrix of the score function where

$$\frac{\partial l}{\partial \alpha} = \delta_j \left(1 - y_j e^{\alpha + \beta x_j} \right) - (1 - \delta_j) c e^{\alpha + \beta x_j}, \qquad \frac{\partial l}{\partial \beta} = x_j \frac{\partial l(x_j, \alpha, \beta)}{\partial \alpha},$$

for the jth observation and also the calculation of the asymptotic information matrix involving the second order derivatives

$$\frac{\partial^2 l}{\partial \alpha^2} = -e^{\alpha+\beta x_j}\left[\delta_j y_j + c(1-\delta_j)\right], \qquad \frac{\partial^2 l}{\partial \alpha \partial \beta} = x_j \frac{\partial^2 l}{\partial \alpha^2}, \qquad \frac{\partial^2 l}{\partial \beta^2} = x_j^2 \frac{\partial^2 l}{\partial \alpha^2}.$$

At this stage the fact that the true model is actually specified by (2) must be taken into account. Assuming this true model and observing that the above expressions involve only two random quantities via δ_j and $\delta_j Y_j$, gives

$$\begin{aligned} E\left[\frac{\partial l}{\partial \alpha}\right] &= e^{\alpha+\beta x_j}\int_0^c e^{-y_j e^{\alpha+\beta x_j}} \frac{g(y_j)}{\sqrt{n}}\, dy_j + o\left(\frac{1}{\sqrt{n}}\right), \\ Var\left(\frac{\partial l}{\partial \alpha}\right) &= 1 - e^{-ce^{\alpha+\beta x_j}} + e^{\alpha+\beta x_j} e^{-ce^{\alpha+\beta x_j}} \int_0^c \frac{g(s)}{\sqrt{n}}\, ds \\ &\quad - \left(e^{\alpha+\beta x_j}\right)^2 \int_0^c 2\frac{y_j g(y_j)}{\sqrt{n}} e^{-y_j e^{\alpha+\beta x_j}}\, dy_j + o\left(\frac{1}{\sqrt{n}}\right), \\ E\left[-\frac{\partial^2 l}{\partial \alpha^2}\right] &= 1 - e^{-ce^{\alpha+\beta x_j}} + e^{\alpha+\beta x_j} e^{-ce^{\alpha+\beta x_j}} \int_0^c \frac{g(s)}{\sqrt{n}}\, ds \\ &\quad - e^{\alpha+\beta x_j} \int_0^c \frac{g(y_j)}{\sqrt{n}} e^{-y_j e^{\alpha+\beta x_j}}\, dy_j + o\left(\frac{1}{\sqrt{n}}\right), \end{aligned} \tag{5}$$

using Taylor expansions. The calculations for the derivation of the set of expressions (5) can be found in the appendix.

Now let $\boldsymbol{\theta} = (\alpha, \beta)^T$ be the vector of model parameters and $\boldsymbol{\theta}_0$ the vector of their true values. Also let

$$\xi = \begin{Bmatrix} x_1 & x_2 & \dots & x_m \\ \omega_1 & \omega_2 & \dots & \omega_m \end{Bmatrix}, \quad 0 < \omega_i \leq 1, \quad \sum_{i=1}^m \omega_i = 1, \tag{6}$$

where $x_1, \dots, x_m$ $(m \leq n)$ are the distinct experimental points where observations are taken and $\omega_1, \dots, \omega_m$ represent the relative proportions of observations taken at the corresponding point x_i. Using the expressions in (5) gives the asymptotic information matrix of $\boldsymbol{\theta}_0$

$$\begin{aligned} M(\xi) = M(\xi, \boldsymbol{\theta}_0) &= \lim_{n\to\infty} \frac{1}{n} E\left[-\sum_{j=1}^n \frac{\partial^2 l}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\boldsymbol{T}}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right] \\ &= \sum_{i=1}^m \omega_i \left(1 - e^{-ce^{\alpha+\beta x_i}}\right) \begin{pmatrix} 1 & x_i \\ x_i & x_i^2 \end{pmatrix}, \end{aligned}$$

the asymptotic expectation of the score function evaluated at $\boldsymbol{\theta}_0$

$$\begin{aligned} \tilde{b}(\xi, g) = \tilde{b}(\xi, g, \boldsymbol{\theta}_0) &= \frac{1}{\sqrt{n}} \lim_{n\to\infty} \frac{1}{n} E\left[\sqrt{n} \sum_{j=1}^n \frac{\partial l}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right] \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^m \omega_i e^{\alpha+\beta x_i} \int_0^c e^{-y_j e^{\alpha+\beta x_i}} g(y_j)\, dy_j \begin{pmatrix} 1 \\ x_i \end{pmatrix} := \frac{1}{\sqrt{n}} b(\xi, g), \end{aligned}$$

and finally the asymptotic variance-covariance matrix of the score function evaluated at $\boldsymbol{\theta}_0$ which is given by

$$\begin{aligned} C(\xi) = C(\xi, \boldsymbol{\theta_0}) &= \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} Cov\left(\frac{\partial l}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right) \\ &= \sum_{i=1}^{m} \omega_i (1 - e^{-ce^{\alpha+\beta x_i}}) \begin{pmatrix} 1 & x_i \\ x_i & x_i^2 \end{pmatrix}. \end{aligned}$$

Note that the asymptotic matrices $M(\xi)$ and $C(\xi)$ are identical. Now expanding the score function $s(\boldsymbol{\theta})$ around $\boldsymbol{\theta}_0$ gives

$$s(\boldsymbol{\theta}) = s(\boldsymbol{\theta}_0) + s'(\boldsymbol{\theta}_0)(\boldsymbol{\theta} - \boldsymbol{\theta}_0) + \dots,$$

and, since the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ is a root of the score function,

$$\begin{aligned} 0 &\approx s(\boldsymbol{\theta}_0) + s'(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &\approx M^{-1}(\xi, \boldsymbol{\theta}_0) s(\boldsymbol{\theta}_0). \end{aligned}$$

Now $\sqrt{n}\ s(\boldsymbol{\theta}_0) \sim AN(b(\xi, g), C(\xi))$ and therefore the asymptotic distribution of the maximum likelihood estimator is described by

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \sim AN\left(M^{-1}(\xi) b(\xi, g), M^{-1}(\xi)\right). \tag{7}$$

4. Minimax optimal designs

The optimal planning of time-to-event experiments is concerned with finding the experimental points and the number of subjects that should be assigned to each point so that the parameters are estimated as precisely as possible. To illustrate the proposed methodology, the binary design spaces $\mathcal{X} = \{0, 1\}$ is considered, corresponding, for example, to placebo and active treatment, respectively. Following Kiefer (1974), the problem is formulated through an approximate design of the form (6) with support points 0 and 1 and corresponding weights ω and $1-\omega$. In practice, if an approximate design is available and a total number of n observations can be taken, the quantities ωn and $(1-\omega)n$ are rounded to integers using an efficient rounding procedure in order for the design to be used (see Pukelsheim and Rieder (1992)).

As mentioned in the introduction and as can be seen in (7), fitting the exponential-based proportional hazards model given in (1) when in fact the true underlying model is specified by (2) adds a bias to the maximum likelihood estimator for the vector of parameters. Therefore, a suitable measure for the precision of the parameter estimates is the mean squared error matrix which, using (7), is given by

$$\begin{aligned} MSE(\xi, g) &= \left(M^{-1}(\xi) b(\xi, g)\right) \left(M^{-1}(\xi) b^T(\xi, g)\right) + M^{-1}(\xi) \\ &= M^{-1}(\xi) \left(b(\xi, g) b^T(\xi, g) + M(\xi)\right) M^{-1}(\xi). \end{aligned} \tag{8}$$

Furthermore, the minimax approach is adopted in order to find designs that ensure precise parameter estimation for the worst case scenario among all possible model departures in the class of contamination functions (either $\mathcal{G}_1$ or $\mathcal{G}_2$).

4.1. Optimality criteria

The optimality criteria used correspond to classical optimality criteria but they are based on the mean squared error matrix rather than just the information matrix. The resulting minimax optimal designs minimise the corresponding criteria functions with respect to the design, for the worst possible contamination function g.

The first criterion to be studied corresponds to the c-optimality criterion for estimating only the parameter β, treating α as a nuisance parameter. This is often the case in time-to-event experiments since β represents the explanatory variable effect and is therefore of primary interest. A design ξ^* is a minimax c-optimal design for estimating β if $(0 \;\; 1)^T$ is in the range of $MSE(\xi^*, g)$ and

$$\xi^* = \arg\min_{\xi} \max_{g \in \mathcal{G}_1 \, or \, \mathcal{G}_2} (0 \;\; 1)\, MSE(\xi, g) \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{9}$$

The case corresponding to D-optimality is also considered, that is, when one is interested in estimating both model parameters α and β. A design ξ^* is minimax D-optimal if

$$\xi^* = \arg\min_{\xi} \max_{g \in \mathcal{G}_1 \, or \, \mathcal{G}_2} det \left\{ MSE(\xi, g) \right\}. \tag{10}$$

Note that under both optimality criteria the resulting optimal designs depend on the parameter values and therefore, following Chernoff (1953), these are referred to as locally optimal designs. The corresponding locally optimal designs for the case of the exponential-based proportional hazards model being the true model are readily available in Konstantinou et al. (2014).

4.2. Minimax c-optimal designs for β

Design search is illustrated through the special case of a binary design space $\mathcal{X} = \{0, 1\}$. Hence a candidate design for estimating the parameter β can have either one or two support points. However, in the former case the mean squared error matrix cannot be defined since the information matrix $M(\xi)$ is singular. Therefore, the designs must be supported at both 0 and 1, and let ω and $1 - \omega$ be their corresponding weights. The following theorem gives the minimax c-optimal design for estimating β for both the cases of $g \in \mathcal{G}_1$ and $g \in \mathcal{G}_2$ (see the appendix for a proof).

Theorem 1. *Regardless of whether the contamination function belongs in $\mathcal{G}_1$ or $\mathcal{G}_2$, the minimax c-optimal design for estimating β on $\mathcal{X} = \{0, 1\}$ allocates a proportion of ω^* of observations at point $x = 0$, where*

$$\omega^* = \frac{\sqrt{1 - e^{-ce^{\alpha+\beta}}}}{\sqrt{1 - e^{-ce^{\alpha}}} + \sqrt{1 - e^{-ce^{\alpha+\beta}}}}. \tag{11}$$

The minimax c-optimal weight given in (11) is the same as the c-optimal weight for estimating β when the exponential-based proportional hazards model is true (see Konstantinou et al. (2014)). Therefore, the contamination function g does not affect the minimax c-optimal design for β and the exponential distribution can be assumed without loss of generality. This result is in line with the findings of Konstantinou et al. (2015) for partial likelihood estimation.

4.3. Minimax D-optimal designs To allow estimation of both parameters a design must have at least two support points. For $\mathcal{X} = \{0, 1\}$ this means that both points 0 and 1 must be support points of the minimax D-optimal design. However, now the choice of contamination class and therefore the worst possible contaminant affects the optimal choice of design. Theorems 2 and 3 provide analytical characterisations of the minimax D-optimal designs when $g \in \mathcal{G}_1$ and $g \in \mathcal{G}_2$ respectively and are proven in the appendix.

Theorem 2. *Let $g \in \mathcal{G}_1$. The minimax D-optimal design on $\mathcal{X} = \{0, 1\}$ allocates a proportion of ω^* observations at point $x = 0$, where*

$$\omega^* = \frac{\sqrt{c_1^2(1 - e^{-ce^{\alpha+\beta}}) + 1}\left[\sqrt{c_1^2(1 - e^{-ce^{\alpha}}) + 1} - \sqrt{c_1^2(1 - e^{-ce^{\alpha+\beta}}) + 1}\right]}{c_1^2(e^{-ce^{\alpha+\beta}} - e^{-ce^{\alpha}})}. \tag{12}$$

Theorem 3. *Let $g \in \mathcal{G}_2$. The minimax D-optimal design on $\mathcal{X} = \{0, 1\}$ allocates a proportion of ω^* observations at point $x = 0$, where*

$$\omega^* = \frac{\sqrt{\frac{c_2^2(e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})} + 1}\left[\sqrt{\frac{c_2^2(e^{\alpha})^2}{(1-e^{-ce^{\alpha}})} + 1} - \sqrt{\frac{c_2^2(e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})} + 1}\right]}{c_2^2\left[\frac{(e^{\alpha})^2}{(1-e^{-ce^{\alpha}})} - \frac{(e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})}\right]}. \tag{13}$$

Note that the D-optimal design when model (1) is true allocates equal proportions of observations at point 0 and 1. Furthermore, it is easy to check that both minimax D-optimal weights have limiting values as c_1 or c_2, increases. These are

$$\lim_{c_1 \to \infty} \omega^* = \frac{\sqrt{1 - e^{-ce^{\alpha+\beta}}}}{\sqrt{1 - e^{-ce^{\alpha}}} + \sqrt{1 - e^{-ce^{\alpha+\beta}}}}, \quad \text{when } g \in \mathcal{G}_1,$$

and

$$\lim_{c_2 \to \infty} \omega^* = \frac{e^{\beta}\sqrt{1 - e^{-ce^{\alpha}}}}{e^{\beta}\sqrt{1 - e^{-ce^{\alpha}}} + \sqrt{1 - e^{-ce^{\alpha+\beta}}}}, \quad \text{when } g \in \mathcal{G}_2.$$

Also note that that the minimax D-optimal weight for $g \in \mathcal{G}_1$ given in (12) tends to the c-optimal weight for β when the exponential-based proportional hazards model is true.

5. Numerical results

For time-to-event experiments comparing two treatments, or equivalently a placebo with an active treatment, practitioners traditionally use the balanced design allocating equal proportions of observations at the two treatments. The aim is to illustrate the theoretical results on minimax optimal designs found in the previous section and also to examine the efficiency of the balanced design in the presence of model uncertainty and possibly censored data.

5.1. Minimax c-optimal designs for β As shown in Section 4.2, the minimax c-optimal design for estimating β does not depend on the contamination function g but is locally optimal through the parameter values (see Theorem 1). To illustrate this parameter dependence β-values corresponding to small, moderate and large covariate effects are used along with various proportions of censored observations. Following Kalish and Harrington (1988), the proportion of censoring is characterised as the overall probability of censoring for model (1) had a balanced design been used. That is,

$$\text{proportion of censoring} = 1 - 0.5(1 - e^{-ce^{\alpha}}) - 0.5(1 - e^{-ce^{\alpha+\beta}}). \tag{14}$$

Setting $\alpha = 0$, for illustration purposes, this equation provides the value of the censoring time c for a given combination of β-value and censoring proportion.

Two different contamination scenarios are considered. For $g \in \mathcal{G}_1$, the Gompertz distribution is selected for which $g(t) = \gamma t$, where γ is the shape parameter. A value of $\gamma = 0$ would correspond to the exponential regression model. For the class $\mathcal{G}_2$ of possibly unbounded contamination functions the Weibull distribution with shape parameter γ is studied for which $g(t) = (\gamma - 1) \ln t$, so a value of $\gamma = 1$ would correspond to the exponential distribution.

The case of $\gamma = 1$ for the Gompertz model and $\gamma = 2$ for the Weibull model is presented. For both contamination types, various different values for γ gave similar results, and are thus omitted. Table 1 shows the minimax c-optimal design weights at point $x = 0$ for several combinations of β-values and proportions of censoring. For each combination, the value of c is determined using equation (14) by setting $\alpha = 0$. Note that the minimax c-optimal weights are the same regardless of whether the Gompertz or the Weibull distribution is used.

It can be observed that the minimax c-optimal design for β allocates more observations at point $x = 0$ when $\beta > 0$, that is, when the probability of occurrence of the event increases with x, and less when $\beta < 0$. Therefore, the design puts more weight at the experimental point where censoring is more likely.

The efficiencies of the balanced design for the Gompertz and the Weibull models under consideration are given in Table 2 for the various explanatory effect and proportion of censoring scenarios.

Table 1: Minimax c-optimal weights ω^* at point $x = 0$.

prop. of cens.	$e^{\beta}(\beta)$							
	0.03 (-3.51)	0.1 (-2.30)	0.25 (-1.39)	0.5 (-0.69)	2 (0.69)	4 (1.39)	10 (2.30)	33.3 (3.51)
0.1	0.47	0.47	0.47	0.48	0.52	0.53	0.53	0.53
0.3	0.39	0.39	0.42	0.46	0.54	0.58	0.61	0.61
0.5	0.22	0.31	0.38	0.44	0.56	0.62	0.69	0.78
0.7	0.17	0.27	0.36	0.43	0.57	0.64	0.73	0.83
0.9	0.15	0.25	0.34	0.43	0.58	0.66	0.75	0.85

It turns out that the balanced design is highly efficient for small proportions of censoring whereas its efficiency drops below 90% for absolute β-values of 2.3 or more and proportion of censoring of 50% or more. Furthermore, the efficiencies are almost identical for both contamination functions. This can be explained by the form of the objective function defined in (9) which is given by

$$\left[\frac{e^{\alpha+\beta}}{(1-e^{-ce^{\alpha+\beta}})}\int_0^c e^{-y_j e^{\alpha+\beta}} g(y_j)\,dy_j - \frac{e^{\alpha}}{(1-e^{-ce^{\alpha}})}\int_0^c e^{-y_j e^{\alpha}} g(y_j)\,dy_j\right]^2 + \frac{1}{\omega(1-e^{-ce^{\alpha}})} + \frac{1}{(1-\omega)(1-e^{-ce^{\alpha+\beta}})}$$

for a design of the form $\xi = \{0, 1; \omega, 1-\omega\}$. The above expression shows that the objective function is dominated by the terms involving ω but not g.

5.2. Minimax D-optimal designs Besides the parameter dependence, the minimax D-optimal designs also depend on the choice of contamination class and therefore on the values of the positive constants c_1 or c_2 (see Theorems 2 and 3). In order to illustrate the contaminant dependence a numerical example is used that is based on the study reported by Freireich et al. (1963), for which the maximum likelihood estimates are $\hat{\alpha} = -2.163$ and $\hat{\beta} = -1.526$ with approximately 30% of the observations right-censored. Using this proportion of censoring and the estimates $\hat{\alpha}$ and $\hat{\beta}$ for the α and β values, the value $c = 30$ is obtained from the characterisation of the proportion of censoring defined in section (14). Figures 1 and 2 illustrate the limiting behaviour of the minimax D-optimal weights ω^* on $x = 0$ given in (12) and (13) respectively as c_1 and c_2 increase.

For both cases of $\mathcal{G}_1$ and $\mathcal{G}_2$, the weight at point $x = 0$ is smaller than 0.5 (the D-optimal weight when model (1) is true) and its value decreases as c_1 and c_2 increase, with the limiting weight for $g \in \mathcal{G}_1$ being larger than that for $g \in \mathcal{G}_2$.

To investigate the performance of the D-optimal minimax design with respect to contamination functions $g \in \mathcal{G}_1$ and $\mathcal{G}_2$, the Gompertz and the Weibull distributions respectively are considered

Table 2: Efficiency, in percent, of the balanced design, for the Gompertz model with $\gamma = 1$ and the Weibull model with $\gamma = 2$ (in brackets).

prop. of cens.	$e^{\beta}(\beta)$ 0.03 (-3.51)	0.1 (-2.30)	0.25 (-1.39)	0.5 (-0.69)	2 (0.69)	4 (1.39)	10 (2.30)	33.3 (3.51)
0.1	100.0 (99.9)	100.0 (99.8)	99.8 (99.8)	99.8 (99.9)	99.8 (99.9)	99.7 (99.8)	99.7 (99.8)	99.7 (99.9)
0.3	99.4 (97.2)	96.4 (96.1)	97.4 (97.5)	99.2 (99.2)	99.2 (99.2)	97.4 (97.5)	95.4 (96.1)	95.2 (97.2)
0.5	76.5 (76.7)	87.2 (87.3)	94.7 (94.7)	98.6 (98.6)	98.6 (98.6)	94.7 (94.7)	87.1 (87.3)	76.3 (76.7)
0.7	70.1 (70.1)	82.6 (82.6)	92.5 (92.5)	98.0 (98.0)	98.0 (98.0)	92.5 (92.5)	82.6 (82.6)	70.1 (70.1)
0.9	67.6 (67.6)	79.8 (79.8)	90.8 (90.8)	97.4 (97.4)	97.4 (97.4)	90.8 (90.8)	79.8 (79.8)	67.7 (67.7)

again. In this situation the results turn out to be less clear cut than for c-optimality. In particular, the relative efficiencies of the balanced design are close to 1 in all scenarios, and sometimes even exceed 1. It is not surprising that an optimal minimax design can be less efficient than a non-optimal design in some scenarios, since the minimax designs protect against a whole class of contamination functions whereas each scenario is characterised by just one function from this class. This phenomenon is similar in nature to situations where parameter robust designs over a range of values are outperformed for specific values in this range.

As proposed by a Referee, the case of the continuous design space $\mathcal{X} = [0, 1]$ is also investigated, for the same combinations of β-values and censoring proportions as in Table 1. Using either the c- or D-optimality criterion, the support points of the resulting minimax optimal designs for either contamination class are found to always be the points 0 and 1. Furthermore, the corresponding minimax optimal weights are the same as in the case of the binary design space $\mathcal{X} = \{0, 1\}$ described above.

Overall, the conclusion is that if estimation of both parameters, α and β, is of interest, the balanced design is highly efficient, and can be used in practice. If, however, the main focus is on the treatment effect β, then the minimax c-optimal designs are recommended.

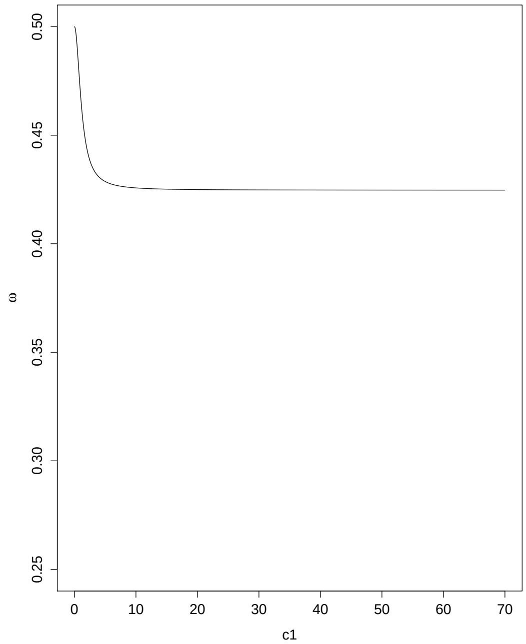

Figure 1: Minimax D-optimal weight at point 0 for $g \in \mathcal{G}_1$

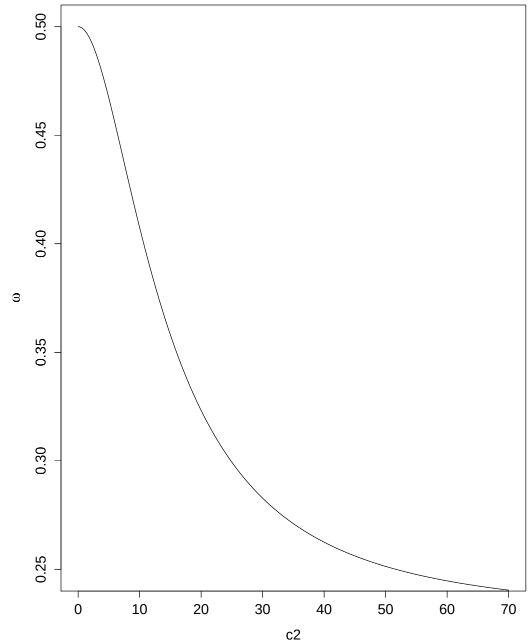

Figure 2: Minimax D-optimal weight at point 0 for $g \in \mathcal{G}_2$

6. Conclusions

In practice when parametric models are used for time-to-event experiments, often the exponential distribution is naturally assumed for the event times along with the proportional hazards assumption. However, this assumed model may only be an approximation of the true underlying parametric proportional hazards model.

Following this practical scenario, deviations in a neighbourhood of the exponential-based proportional hazards model are considered which are specified by a contamination function g. Two different classes of contamination functions are defined which can be used to include various forms of g but most importantly they include the commonly used parametric proportional hazards models based on the Weibull and Gompertz distributions.

Since time-to-event experiments are usually conducted in order to evaluate a particular intervention or treatment, the focus is on models involving one explanatory variable. Nevertheless, the model misspecification introduced in Section 2 and the results of Section 3 regarding the maximum likelihood estimator can be easily generalised to models with more than one explanatory variables.

Assuming that the time-to-event data are subject to Type-I censoring, the construction of designs which are robust to model misspecifications is investigated. Following Wiens (1992), optimality criteria corresponding to the classical c- and D-optimality criteria but based on the mean squared error matrix are used and minimax optimal designs are constructed which guard against the worst possible deviation from the assumed model. Therefore, previous results on minimax optimal designs are extended by considering both possibly censored data as well as the class of proportional hazards models. However, the choice of contamination classes enables the use of designs with finite support and therefore analytical characterisations of minimax c- and D-optimal designs are provided.

This framework is established for general designs, hence optimal designs on continuous design spaces $\mathcal{X}$ corresponding to, for example, doses of a drug can easily be found numerically. The design search is illustrated for the important special case of a binary design space, and some analytical results have been presented.

The results on minimax c-optimal designs for estimating the covariate coefficient β, show that the deviations from the exponential distribution do not affect the optimal choice of design if one remains in a proportional hazards framework. This is in accordance with the result for partial likelihood estimation, stating that under Type-I censoring the exponential distribution can be assumed for design search without loss of generality (see Konstantinou et al. (2015)).

If estimation of both parameters is required, that is, if D-optimality is the desired criterion, then Theorems 2 and 3 give the minimax optimal weights for deviations in the class $\mathcal{G}_1$ and $\mathcal{G}_2$ respectively. Both of these weights have limiting values if we allow the deviations to become large and in particular when $g \in \mathcal{G}_1$ the minimax D-optimal weight tends to the c-optimal weight corresponding to the case of the assumed model being true, as $c_1 \to \infty$. This again highlights the importance of the latter design in a model uncertainty situation.

The analytical characterisations of minimax optimal designs and the numerical results of Section 5 suggest that if the main interest is in estimating the treatment effect one has to move away from the traditional balanced design to guard against misspecifications of the assumed exponential model. A suitable candidate for practical use would appear to be the classical c-optimal design for estimating the covariate effect assuming the exponential-based proportional hazards model. It is minimax c-optimal for both contamination classes, is (in the limit) minimax D-optimal for $\mathcal{G}_1$, and is also highly efficient if Cox's proportional hazard model is fitted via partial likelihood estimation (see Konstantinou et al. (2015)). An analytical characterisation of the locally c-optimal design for estimating β is given in Konstantinou et al. (2014).

The designs derived are locally optimal hence, while being robust against model misspecifications, they depend on the values of the unknown model parameters. Finding designs which are robust to both sources of uncertainty is an interesting area of future research. A promising starting point for such an investigation could be the parameter robust designs derived in Konstantinou et al. (2014). They show that for a binary design space and a given range of possible

β-values, the optimal weight at point 0 of the standardised maximin c-optimal design for β (for parameter robustness) is the average of the two locally c-optimal weights at point 0 corresponding to the end-points of the given interval of β-values. As it is shown in Section 4, the (locally) minimax c- and D-optimal weights (for model robustness) are the same or tend, respectively, to the locally c-optimal weight. Therefore, it is of interest to investigate whether such an averaging of the locally optimal weights result also holds both for minimax c- and D-optimal designs for model robustness.

Acknowledgements

The work of M. Konstantinou has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823, Teilproject C2) of the German Research Foundation (DFG).

Appendix

A.1. Proof of set of expressions in (5) The true underlying model specified by (2) has corresponding probability density function given by

$$f_2(y_j) = \exp\left\{\alpha + \beta x_j + \frac{g(y_j)}{\sqrt{n}}\right\} \exp\left\{-e^{\alpha+\beta x_j} \int_0^{y_j} e^{g(s)/\sqrt{n}}\, ds\right\}, \qquad j = 1, \ldots, n.$$

Taking this into account gives

$$E(\delta_j) = P_j = P(Y_j = T_j) = \int_0^c f_2(y_j)\, dy_j = 1 - \exp\left\{-e^{\alpha+\beta x_j} \int_0^c e^{g(s)/\sqrt{n}}\, ds\right\}.$$

Since small deviations from the exponential-based proportional hazards model are considered, the Taylor expansion of $e^{g(s)/\sqrt{n}}$ around $g(s) = 0$ can be used. Then the above expression becomes

$$\begin{aligned} E(\delta_j) &= 1 - \exp\left\{-e^{\alpha+\beta x_j} \int_0^c \left[1 + \frac{g(s)}{\sqrt{n}} + o\left(\frac{g(s)}{\sqrt{n}}\right)\right] ds\right\} \\ &= 1 - \exp\left\{-ce^{\alpha+\beta x_j}\right\} \exp\left\{-e^{\alpha+\beta x_j} \left[\int_0^c \frac{g(s)}{\sqrt{n}}\, ds + o\left(\frac{1}{\sqrt{n}}\right)\right]\right\}. \end{aligned}$$

By further expanding around $\int_0^c \frac{g(s)}{\sqrt{n}}\, ds + o\left(\frac{1}{\sqrt{n}}\right) = 0$, the expectation of the random variable δ_j becomes

$$E(\delta_j) = 1 - e^{-ce^{\alpha+\beta x_j}} + e^{\alpha+\beta x_j} e^{-ce^{\alpha+\beta x_j}} \int_0^c \frac{g(s)}{\sqrt{n}}\, ds + o\left(\frac{1}{\sqrt{n}}\right),$$

where the first term, $1-e^{-ce^{\alpha+\beta x_i}}$, corresponds to the expectation if the assumed exponential model was in fact the true model. Using this expression the variance of δ_j can be found without making any further calculations and is given by

$$\begin{aligned}Var(\delta_j) = P_j(1-P_j) &= e^{-ce^{\alpha+\beta x_j}}(1-e^{-ce^{\alpha+\beta x_j}})\\ &+ e^{\alpha+\beta x_j}e^{-ce^{\alpha+\beta x_j}}(2e^{-ce^{\alpha+\beta x_j}}-1)\int_0^c \frac{g(s)}{\sqrt{n}}\,ds + o\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

Note that

$$\delta_j Y_j = \begin{cases} Y_j, & \text{if } Y_j = T_j \\ 0, & \text{if } Y_j = c \end{cases}.$$

Following along the same lines as for the random quantity δ_j, that is, using two consecutive Taylor expansions, the following expression can be obtained

$$\begin{aligned}E(\delta_j Y_j) = &\frac{1-e^{-ce^{\alpha+\beta x_j}}}{e^{\alpha+\beta x_j}} - ce^{-ce^{\alpha+\beta x_j}} + e^{-ce^{\alpha+\beta x_j}}(ce^{\alpha+\beta x_j}+1)\int_0^c \frac{g(s)}{\sqrt{n}}\,ds\\ &- \int_0^c \frac{g(y_i)}{\sqrt{n}}e^{-y_j e^{\alpha+\beta x_j}}\,dy_j + o\left(\frac{1}{\sqrt{n}}\right),\end{aligned}$$

$$\begin{aligned}Var(\delta_j Y_j) = &- c^2 e^{-ce^{\alpha+\beta x_j}}(1+e^{-ce^{\alpha+\beta x_j}}) + \frac{1-e^{-2ce^{\alpha+\beta x_j}}}{(e^{\alpha+\beta x_j})^2} - \frac{2ce^{-2ce^{\alpha+\beta x_j}}}{e^{\alpha+\beta x_j}} + e^{-ce^{\alpha+\beta x_j}}\\ &\left(c^2 e^{\alpha+\beta x_j} + 4ce^{-ce^{\alpha+\beta x_j}} + \frac{2e^{-ce^{\alpha+\beta x_j}}}{e^{\alpha+\beta x_j}} + 2c^2 e^{\alpha+\beta x_j}e^{-ce^{\alpha+\beta x_j}}\right)\int_0^c \frac{g(s)}{\sqrt{n}}\,ds\\ &- \int_0^c 2e^{-y_j e^{\alpha+\beta x_j}}\left(y_j + \frac{e^{-ce^{\alpha+\beta x_j}}}{e^{\alpha+\beta x_j}} + ce^{-ce^{\alpha+\beta x_j}}\right)\frac{g(y_j)}{\sqrt{n}}\,dy_j + o\left(\frac{1}{\sqrt{n}}\right),\end{aligned}$$

$$\begin{aligned}Cov(\delta_j, \delta_j Y_j) = &\, e^{-ce^{\alpha+\beta x_j}}(1-e^{-ce^{\alpha+\beta x_j}})/e^{\alpha+\beta x_j} - ce^{-2ce^{\alpha+\beta x_j}}\\ &+ e^{-ce^{\alpha+\beta x_j}}\left(2ce^{\alpha+\beta x_j}e^{-ce^{\alpha+\beta x_j}} + 2e^{-ce^{\alpha+\beta x_j}} - 1\right)\int_0^c \frac{g(s)}{\sqrt{n}}\,ds\\ &- e^{-ce^{\alpha+\beta x_j}}\int_0^c \frac{g(y_j)}{\sqrt{n}}e^{-y_j e^{\alpha+\beta x_j}}\,dy_j + o\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

The set of expressions given in (5) then follow since

$$\begin{aligned}E\left(\frac{\partial l}{\partial \alpha}\right) &= E(\delta_j) - e^{\alpha+\beta x_j}E(\delta_j Y_j) - ce^{\alpha+\beta x_j}E(1-\delta_j),\\ Var\left(\frac{\partial l}{\partial \alpha}\right) &= Var(\delta_j) + e^{2(\alpha+\beta x_j)}Var(\delta_j Y_j) + c^2 e^{2(\alpha+\beta x_j)}Var(\delta_j)\\ &\quad - 2e^{\alpha+\beta x_j}(1+ce^{\alpha+\beta x_j})Cov(\delta_j, \delta_j Y_j),\\ E\left(-\frac{\partial^2 l}{\partial \alpha^2}\right) &= -e^{\alpha+\beta x_j}\left[E(\delta_j Y_j) + cE(1-\delta_j)\right].\end{aligned}$$

A.2. Proof of Theorem 1 Let $\xi = \{0, 1; \omega, 1-\omega\}$. The objective function defined in (9) becomes

$$\left[\frac{e^{\alpha+\beta}}{(1-e^{-ce^{\alpha+\beta}})}\int_0^c e^{-y_j e^{\alpha+\beta}} g(y_j)\,dy_j - \frac{e^{\alpha}}{(1-e^{-ce^{\alpha}})}\int_0^c e^{-y_j e^{\alpha}} g(y_j)\,dy_j\right]^2 + \frac{1}{\omega(1-e^{-ce^{\alpha}})} + \frac{1}{(1-\omega)(1-e^{-ce^{\alpha+\beta}})}. \tag{15}$$

The minimax c-optimal design for β is found by minimising the above expression with respect to ω for the worst possible contaminant. Observe that the term involving the contamination function g is independent of the weight ω and therefore, it is enough to minimise

$$\frac{1}{\omega(1-e^{-ce^{\alpha}})} + \frac{1}{(1-\omega)(1-e^{-ce^{\alpha+\beta}})},$$

which gives the optimal weight

$$\omega^* = \frac{\sqrt{1-e^{-ce^{\alpha+\beta}}}}{\sqrt{1-e^{-ce^{\alpha}}} + \sqrt{1-e^{-ce^{\alpha+\beta}}}}.$$

A.3. Proof of Theorem 2 For $\xi = \{0, 1; \omega, 1-\omega\}$ the determinant of the mean squared error matrix is given by

$$\frac{1}{\omega(1-\omega)(1-e^{-ce^{\alpha}})(1-e^{-ce^{\alpha+\beta}})}\left\{1 + \omega\frac{\left[e^{\alpha}\int_0^c e^{-y_j e^{\alpha}} g(y_j)\,dy_j\right]^2}{(1-e^{-ce^{\alpha}})} + (1-\omega)\frac{\left[e^{\alpha+\beta}\int_0^c e^{-y_j e^{\alpha+\beta}} g(y_j)\,dy_j\right]^2}{(1-e^{-ce^{\alpha+\beta}})}\right\}.$$

Since $g \in \mathcal{G}_1$, then $\max_{y_j\in[0,c]} |g(y_j)| \leq c_1\ \forall j = 1, \ldots, n$ and so

$$\left|\int_0^c e^{-y_j e^{\alpha+\beta x}} g(y_j)\,dy_j\right| \leq \int_0^c e^{-y_j e^{\alpha+\beta x}} |g(y_j)|\,dy_j \leq \int_0^c e^{-y_j e^{\alpha+\beta x}} c_1\,dy_j = c_1(1-e^{-ce^{\alpha+\beta x}})/e^{\alpha+\beta x}, \quad \forall x \in \{0, 1\}$$

Therefore, for contamination functions g in the class $\mathcal{G}_1$ the maximum value of the determinant of the mean squared error matrix is given by

$$\frac{c_1^2}{\omega(1-e^{-ce^{\alpha}})} + \frac{c_1^2}{(1-\omega)(1-e^{-ce^{\alpha+\beta}})} + \frac{1}{\omega(1-\omega)(1-e^{-ce^{\alpha}})(1-e^{-ce^{\alpha+\beta}})}.$$

Taking the first order derivative of this expression with respect to ω and equating it to zero gives

$$c_1^2\omega^2(1-e^{-ce^{\alpha}}) - c_1^2(1-\omega)^2(1-e^{-ce^{\alpha+\beta}}) - (1-2\omega) = 0$$
$$\iff \omega_{1,2} = \frac{-[c_1^2(1-e^{-ce^{\alpha+\beta}})+1] \pm \sqrt{c_1^2(1-e^{-ce^{\alpha}})+1}\sqrt{c_1^2(1-e^{-ce^{\alpha+\beta}})+1}}{c_1^2(e^{-ce^{\alpha+\beta}} - e^{-ce^{\alpha}})}.$$

When β is positive, it is easy to see that both the numerator and the denominator of the above expression are non-positive. The negative root of the numerator is rejected since

$$\begin{aligned}&-c_1^2(1-e^{-ce^{\alpha+\beta}})-1-\sqrt{c_1^2(1-e^{-ce^{\alpha}})+1}\sqrt{c_1^2(1-e^{-ce^{\alpha+\beta}})+1}\\&<-c_1^2(1-e^{-ce^{\alpha+\beta}})+c_1^2(1-e^{-ce^{\alpha}})=c_1^2(e^{-ce^{\alpha+\beta}}-e^{-ce^{\alpha}}),\end{aligned}$$

and the weight must be always less than or equal to unity. In the case of negative β-values the denominator is positive and since $\omega>0$, again the positive root is accepted.

Therefore, whatever the sign of the parameter β, the minimax D-optimal weight at point 0 is given by (12).

A.4. Proof of Theorem 3 Since $g\in\mathcal{G}_2$, then $\left|\int_0^c e^{-y_j e^{\alpha+\beta x}}g(y_j)\,dy_j\right|\le c_2\ \forall x\in\{0,1\}$. Therefore, for a fixed design ξ supported at 0 and 1 with corresponding weights ω and $1-\omega$ the determinant of the mean squared error matrix is smaller than or equal to

$$\frac{1}{\omega(1-\omega)(1-e^{-ce^{\alpha}})(1-e^{-ce^{\alpha+\beta}})}\left\{1+\omega\frac{(c_2e^{\alpha})^2}{(1-e^{-ce^{\alpha}})}+(1-\omega)\frac{(c_2e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})}\right\}.$$

Taking the first order derivative of this expression with respect to ω and equating it to zero gives

$$\begin{aligned}&\frac{(c_2e^{\alpha})^2}{(1-e^{-ce^{\alpha}})}\omega^2-\frac{(c_2e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})}(1-\omega)^2-(1-2\omega)=0\\ \iff\ &\omega_{1,2}=\frac{-\left[\frac{(c_2e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})}+1\right]\pm\sqrt{\frac{(c_2e^{\alpha})^2}{(1-e^{-ce^{\alpha}})}+1}\sqrt{\frac{(c_2e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})}+1}}{c_2^2\left(\frac{(e^{\alpha})^2}{(1-e^{-ce^{\alpha}})}-\frac{(e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})}\right)}.\end{aligned}$$

When β is positive, it is easy to check that both the numerator and the denominator of the above expression are non-positive. The negative root of the numerator is rejected since

$$\begin{aligned}&-\frac{(c_2e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})}-1-\sqrt{\frac{(c_2e^{\alpha})^2}{(1-e^{-ce^{\alpha}})}+1}\sqrt{\frac{(c_2e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})}+1}\\&<-\frac{(c_2e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})}<c_2^2\left(\frac{(e^{\alpha})^2}{(1-e^{-ce^{\alpha}})}-\frac{(e^{\alpha+\beta})^2}{(1-e^{-ce^{\alpha+\beta}})}\right),\end{aligned}$$

and the weight must be always less than or equal to unity. In the case of negative β-values the denominator is positive and since $\omega>0$, again the positive root is accepted.

Therefore, whatever the sign of the parameter β, the minimax D-optimal weight at point 0 is given by (13).

References

Box, G. E. P., Draper, N. R., 1959. A basis for the selection of a response surface design. Journal of the American Statistical Association 54 (287), 622–654.

Chernoff, H., 1953. Locally optimal designs for estimating parameters. The Annals of Mathematical Statistics 24 (4), 586–602.

Collett, D., 2003. Modelling survival data in medical research. Chapman & Hall, London, New York.

Fang, Z., Wiens, D. P., 1999. Robust extrapolation designs and weights for biased regression models with heteroscedastic errors. The Canadian Journal of Statistics 27 (4), pp. 751–770.

Freireich, E. J., Gehan, E., Frei III, E., Schroeder, L. R., Wolman, I. J., Anbari, R., Burgert, O. E., Mills, S. D., Pinkel, D., Selawry, O. S., Moon, J. H., Gendel, B. R., Spurr, C. L., Storrs, R., Haurani, F., Hoogstraten, B., Lee, S., 1963. The effect of 6-mercaptopurine on the duration of steroid-induced remissions in acute leukimia: A model for evaluation of other potentially useful therapy. Blood 21 (6), 699–716.

Kalish, L. A., Harrington, D. P., 1988. Efficiency of balanced treatment allocation for survival analysis. Biometrics 44 (3), 815–821.

Kiefer, J., 1974. General equivalence theory for optimum designs (approximate theory). The Annals of Statistics 2 (5), 849–879.

Konstantinou, M., Biedermann, S., Kimber, A. C., 2014. Optimal designs for two-parameter nonlinear models with application to survival models. Statistica Sinica 24, 415–428.

Konstantinou, M., Biedermann, S., Kimber, A. C., 2015. Optimal designs for full and partial likelihood information - with application to survival models. Journal of Statistical Planning and Inference 165, 27–37.

Li, K., Notz, W., 1982. Robust designs for nearly linear regression. Journal of Statistical Planning and Inference 6 (2), 135 – 151.

McGree, J., Eccleston, J., 2010. Investigating design for survival models. Metrika 72 (3), 295–311.

Pascual, F. G., Montepiedra, G., 2003. Model-robust test plans with applications in accelerated life testing. Technometrics 45 (1), 47–57.

Pukelsheim, F., Rieder, S., 1992. Efficient rounding of approximate designs. Biometrika 79 (4), 763–770.

Sinha, S., Wiens, D. P., 2002. Robust sequentlial designs for nonlinear regression. The Canadian Journal of Statistics 30 (4), 601–618.

Wiens, D. P., 1992. Minimax designs for approximately linear regression. Journal of Statistical Planning and Inference 31 (3), 353 – 371.

Wiens, D. P., 1998. Minimax robust designs and weights for approximately specified regression models with heteroscedastic errors. Journal of the American Statistical Association 93 (444), pp. 1440–1450.

Wiens, D. P., Xu, X., 2008. Robust prediction and extrapolation designs for misspecified generalized linear regression models. Journal of Statistical Planning and Inference 138 (1), 30–46.

Woods, D. C., Lewis, S. M., Eccleston, J. A., Russell, K. G., 2006. Designs for generalized linear models with several variables and model uncertainty. Technometrics 48 (2), pp. 284–292.

Xu, X., 2009a. Robust designs for misspecified exponential regression models. Applied Stochastic Models in Business and Industry 25 (2), 179–193.

Xu, X., 2009b. Robust prediction and extrapolation designs for censored data. Journal of Statistical Planning and Inference 139 (2), 486–502.