Statistica Neerlandica (2016) Vol. 70, nr. 2, pp. 146–171
doi:10.1111/stan.12085

D-optimal designs for a continuous predictor in longitudinal trials with discrete-time survival endpoints

Maryam Safarkhani* and Mirjam Moerbeek

Utrecht University, 3508 TC Utrecht, The Netherlands

In designing an experiment with one single, continuous predictor, the questions are composed of what is the optimal number of the predictor's values, what are these values, and how many subjects should be assigned to each of these values. In this study, locally D-optimal designs for such experiments with discrete-time event occurrence data are studied by using a sequential construction algorithm. Using the Weibull survival function for modeling the underlying time to event function, it is shown that the optimal designs for a linear effect of the predictor have two points that coincide with the design region's boundaries, but the design weights highly depend on the predictor effect size and its direction, the survival pattern, and the number of time points. For a quadratic effect of the predictor, three or four design points are needed.

Keywords and Phrases: design region, optimal design point, optimal design weight, predictor effect size, sequential construction algorithm, underlying survival function.

1 Introduction

The design and analysis of longitudinal studies is an increasingly important area in applied research in many fields of science. In social and behavioral sciences, for example, this kind of studies is widely conducted to follow up subjects over successive time points in order to study changes and differences that occur in their attitudes, performances, and behaviors. Examples are smoking intervention studies with repeated measurements of smoking behavior across time, or educational studies to evaluate critical transitions across educational levels, which measure student performances on academic and adaptive skills over time. Conducting longitudinal studies is always expensive, time-consuming, and intellectually challenging; it also requires a massive effort of proficient experts in recruiting and following subjects up over a long period of time. All these efforts can be in vain if the study does not have sufficient power and efficiency for estimating and testing the hypotheses of interest. Therefore, researchers must carefully design their study in the planning stage, and they must be certain their design is efficient for the objective or objectives of their study.

*m.safarkhani@uu.nl

In general, the efficiency of a design is indicated by the accuracy of estimators of the model parameters in terms of their variances: the smaller the variances, the more efficient the estimators. A simple way to increase efficiency is to increase the sample size of the study. However, researchers should always keep in mind that their designs must meet many practical constraints, such as ethical, feasibility, and more important cost constraints. Therefore, they should always find a right balance between costs of the study on the one hand and statistical power and efficiency on the other.

Optimal design theory is a standard procedure for finding highly efficient and cost-effective designs, and it mainly depends on theoretical considerations, along with practical constraints, of the problem at hand. Optimal design theory measures the efficiency of a design based on its variance–covariance matrix of the parameter estimators and defines various optimality criteria; each optimizes a certain function of the variance–covariance matrix.

Optimal designs for longitudinal studies have been investigated by various authors. Tekle *et al.* (2008a), Tekle *et al.* (2008b), Galbraith and Marschner (2002), Lima Passos *et al.* (2011), and Moerbeek (2008) are just a few among others who studied optimal designs for trials with dichotomous (e.g., daily smoking) and continuous (e.g., tumor reduction) responses.

A particular type of outcome in longitudinal studies is a survival endpoint, where the research interest centers on the occurrence and timing of events. The timing of events can be measured continuously using thin precise units (e.g., minutes or days) in time. Data that are recorded on a continuous scale are called continuous-time survival data, and they are often encountered in the biomedical sciences. In social and behavioral sciences, in contrast, far less likely is the possibility that the timing of events is recorded this precisely. This may be illustrated by an example of smoking intervention studies with the aim of preventing or delaying the onset of daily smoking during adolescence. Researchers will not be able to contact participants on a daily basis to record their smoking status. Instead, they might measure the onset of this event discretely using a set of discrete intervals. To do so, they might record the onset of daily smoking once each month and define having initiated daily smoking as having smoked at least one cigarette a day since the last time of measurement.

Time that is measured in discrete intervals is called discrete time, and survival data that are recorded in discrete time are referred to as discrete-time survival data. Discrete-time survival data can be encountered in retrospective studies, where, because of memory failure, respondents do not recall the exact time an event occurred, and also in studies where events can only occur at discrete points in time. For instance, graduation from university occurs at a few points in time during the academic year. Optimal designs for longitudinal studies with discrete-time survival endpoints have been recently studied by Jóźwiak and Moerbeek (2012, 2013). These studies solely focus on a comparison between two qualitative treatments (a control condition and an experimental condition). We aim to extend their results to studies, where the effect of a single and quantitative predictor variable is of interest. An example is a social work project that aims to ascertain the effect of a reduction in social welfare on the

occurrence and timing of finding a paid job. Another example can be found in marketing: a study that aims to evaluate the effect of a reduction in the price of a specific good or service on the occurrence and timing of buying that good or service. In both studies, the time to event occurrence (finding a paid job or buying the good or service) may be measured discretely in, for instance, months. It is worth noting that the predictor is continuous; it is, however, the underlying time to event occurrence that is measured discretely by using discrete-time intervals while in fact events can occur at any point in time.

A design for a continuous predictor variable is determined by the choice of design (or support) points and weights assigned to these points. The design points are the values of the continuous predictor, and the weights specify the proportion of subjects assigned to each value. In our example, design points can be thought of as the proportions of the standard amount of social welfare, and weights are the proportions of unemployed persons who are allocated to these design points. Different designs have different design points and/or different design weights. To find the most efficient design, we use optimal design theory to optimally select the design points and their weights. Optimal designs with a continuous predictor variable have been extensively examined in several studies for the logistic regression model. Some relevant references are Atkinson *et al.* (2007), Berger and Wong (2009), and King and Wong (2000). Moreover, Hsieh and Lavori (2000) examine this specific optimal design problem for trials with continuous-time survival outcomes.

For discrete-time survival analysis, however, there have been no studies to obtain optimal designs for a continuous predictor variable. The aim of the present paper is to determine such optimal designs. We focus our attention on D-optimality criterion, which minimizes the determinant of the variance–covariance matrix of a design; that is, it minimizes the volume of the confidence ellipsoid of the parameter estimators. Hence, the D-optimality criterion has a natural interpretation. It has also other properties. D-optimal designs do not depend on the chosen design space, and so they are not affected by a linear transformation of the design space (Ouwens *et al.*, 2006). They are usually quite robust with respect to other criteria (Lucas, 1974; Donev & Atkinson, 1988; Chasalow, 1992). Therefore, D-optimality criterion is the most important and accessible optimality criterion. As the designs depend on an initial best guess of the unknown parameters that we want to estimate, we construct locally optimal designs. An evaluation of the robustness of such designs is also given. We also study the effects of attrition and the number of time periods on D-optimal designs.

The remainder of this paper is organized as follows. The next section describes the generalized linear model that is widely used to analyze discrete-time survival data and gives the variance–covariance matrix of parameter estimates. Then, we provide theoretical explanations of constructing D-optimal designs and present numerical results in section 4. We then illustrate the use of our methodology by an example in section 5. Finally, we conclude with a discussion in section 6.

2 The discrete-time hazard model

This section describes a widely used model to analyze discrete-time hazard model. An extensive description of this model is found in Singer and Willett (1993, 2003). Let Y_{ik} denote the binary outcome variable for the ith ($i = 1, 2, \cdots, N$) subject in the kth ($k = 1, 2, \cdots, p$) period, where $Y_{ik} = 0$ if subject i in time period k does not experience the event and $Y_{ik} = 1$ if the subject experiences the event in time period k. The event indicator Y_{ik} is observed until and including the kth period if subject i experiences the event in period k or is lost to follow-up during period k, or until the pth period if subject i experiences the event in period p or does not experience the event during the course of the study and the study concludes.

The risk of event occurrence for subject i in the kth discrete interval is denoted by $h\left(t_{ik}\right)$ and depends on the duration of the interval and the underlying continuous-time survival function $S(t) = \Pr(T > t)$, where T is a continuous random variable that measures the survival time in discrete intervals $[t_{k-1}, t_k)$ for $k = 1, 2, \cdots, p$, with $S(t_0) = 1$. The $h\left(t_{ik}\right)$ gives the conditional probability that subject i experiences the target event in period k given that the event did not occur before period k and is defined as $h\left(t_{ik}\right) = \Pr(\Gamma_i = k | \Gamma_i \geq k)$, where Γ is a discrete random variable and indicates the time period when the target event occurs; that is, $\Gamma_i = k$ if $t_{k-1} \leq T_i < t_k$. The discrete-time hazard probability can also be expressed in terms of discrete-time survival probabilities as $h\left(t_k\right) = \frac{S(t_{k-1}) - S(t_k)}{S(t_{k-1})}$, where $S\left(t_k\right)$ is the survival probability at the end of time period k, and gives the probability of not experiencing the event through time period k.

The generalized linear model with a logit link function is used to model the discrete-time hazard probability for subject i in period k:

$$\text{logit}\, h\left(t_{ik}\right) = \log \frac{h\left(t_{ik}\right)}{1 - h\left(t_{ik}\right)} = \sum_{k=1}^{p} \alpha_k D_{ik} + \beta X_i. \tag{1}$$

The dummy variable D_{ik} is set to 1 in time interval k and 0 elsewhere. The independent variable X_i is a continuous variable that takes on the value x_i for subject i in the region of x values $x_{\min} \leq x \leq x_{\max}$, where $x_{\min}$ and $x_{\max}$ are the lower and upper boundaries for the variable X. These boundaries are dictated by ethical and practical considerations, and they may represent values of X at which there is no effect on the event of interest or values that cause a harmful damage. For example, a too large reduction in social welfare to an unemployed person may be considered unethical because he or she should at least be able to pay for basic needs such as food, clothes, housing, and medical care. The predictor X is only subscribed i, which means it is time-invariant and its values vary across subjects but do not change over time periods for a given subject. The intercept parameter α_k represents the value of the logit hazard probability in period k for the baseline group, which is the subset of subjects with value zero on the variable X. It should be noted that the baseline hazard probability in the kth period is computed from evaluating the inverse of the logit link function at α_k. The regression coefficient β

assesses the effect of a one-unit increase in the predictor X on event occurrence on the logit scale and is assumed to be constant across time intervals. In other words, Model 1 is a proportional odds model.

Suppose we decide to take a fixed number of N subjects in our study and each subject is allowed to have a different value of the independent variable X. We assume that there are m distinct values of the independent variable such as $x_1, x_2, \cdots, x_m$, where $x_{\min} \leq x_j \leq x_{\max}$ for all j. We consider the weight π_j as the proportion of subjects with value x_j, subject to $\sum_{j=1}^{m} \pi_j = 1$.

In matrix form, we can write Model 1 as follows:

$$\text{logit}(\mathbf{h}(\mathbf{t})) = \mathbf{X}\theta,$$

where $\mathbf{h}(\mathbf{t})$ is a vector of discrete-time hazard probabilities of event occurrence for all p time periods and all N subjects in the study until they experience the target event or drop out from the study or the study concludes (i.e., $k = p$). The design matrix $\mathbf{X}$ is a matrix with $(p+1)$ columns and $\sum_{k=1}^{p} N_k$ rows, where $N_k = N \sum_{j=1}^{m} \pi_j S\left(t_{j,k-1}\right)(1-r)^{k-1}$ denotes the number of subjects entering the kth period and neither experienced the event nor dropped out of the study past time period $k-1$. It should be noted that $S\left(t_{j,k-1}\right)$ indicates the probability of survival of a subject with the value of x_j on X through the $(k-1)$th time period, and $r \in [0, 1]$ is the attrition rate and denotes the proportion of subjects who leave the study between any two adjacent measurement points because of unforeseen reasons other than event occurrence such as moving out of town or changing jobs or schools. Therefore, the censoring mechanism is non-informative, which implies that all subjects who remain in the study are representative of everyone who would have remained in the study had censoring not occurred (Singer & Willett, 2003, section 9.3.2). Finally, $\theta = (\alpha_1, \alpha_2, \cdots, \alpha_p, \beta)'$ is the column vector of $p+1$ unknown regression parameters.

The vector θ can be estimated by iteratively reweighted least squares that is extensively described by McCullagh and Nelder (1989). The least squares estimator $\hat{\theta}$ has the following asymptotic variance–covariance matrix:

$$\text{Cov}\left(\hat{\theta}\right) = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1} = \frac{1}{N}\left(\sum_{j=1}^{m}\sum_{k=1}^{p} \mathbf{X}'_{jk} W\left(t_{jk}\right)\mathbf{X}_{jk}\pi_j S\left(t_{j,k-1}\right)(1-r)^{k-1}\right)^{-1}. \quad (2)$$

The vector $\mathbf{X}_{jk}$ corresponds to subjects with the value x_j on X in the kth time period and has $(p+1)$ elements with value 1 on the kth element and value x_j on the $(p+1)$th element and zeros elsewhere so that the first p elements represent the values on the dummies $D_1, D_2, \cdots, D_p$ and the $(p+1)$th element represents the value on X. The scalar $W\left(t_{jk}\right)$ is the least squares weight for subjects with the value of x_j on X in time period k, and for a logit link function, it is obtained as $W\left(t_{jk}\right) = h\left(t_{jk}\right)\left(1 - h\left(t_{jk}\right)\right)$, where $h\left(t_{jk}\right)$ is the probability of experiencing the event in period k. It should be mentioned that the variances of the parameter estimates are proportional to the diagonal elements of $\text{Cov}\left(\hat{\theta}\right)$.

The variance–covariance matrix in Equation 2 shows that the variances of the estimated parameters and the power of finding an existing predictor effect are specified by design factors such as the number of X values (m), their values (x_j), the proportion of subjects allocated to each value (π_j), the duration of the study (p), the predictor effect size (β), the attrition rate (r), and the underlying discrete-time survival function $S\left(t_k\right)$. In this study, we assume that p is fixed beforehand and prior estimates for β, r, and $S\left(t_k\right)$ are obtained from findings in the literature or estimated based on a pilot study or expert opinion. Then any different combination of m, x_j, and π_j results in a different design over the region $x_{\min} \leq x \leq x_{\max}$. An important design question is how to find the design that has highest efficiency. Doing so results in constructing an optimal design, which is the subject of the next section.

3 Optimal designs

We investigate locally D-optimal designs that minimize the determinant of the variance–covariance matrix or, equivalently, maximize the determinant of the information matrix of the parameter estimates of Model 1. A design for our study is characterized by the choice of the design points x_j and the design weights π_j at these points. The design points are selected within a predetermined design region that is feasible and ethical for the study at hand. We denote the design region as $\Delta = [x_{\min} \leq x \leq x_{\max}]$, and our aim is to choose m distinct points in Δ so that all the regression parameters in Equation 1 are estimated as precisely as possible.

A continuous design with m distinct points is presented by the measure ξ over Δ as follows:

$$\xi = \left\{ \begin{array}{c} x_1 \; x_2 \; \cdots \; x_m \\ \pi_1 \; \pi_2 \; \cdots \; \pi_m \end{array} \right\}.$$

The first line gives the design points in Δ with the associated design weights π_j in the second line, where $0 \leq \pi_j \leq 1$ for all j and $\sum_{j=1}^{m} \pi_j = 1$. Different designs ξ for Model 1 are determined by different design points and/or different design weights. An ideal experimental design is the one that results in small values of the variances and covariances of the parameter estimates. The variance–covariance matrix $\mathrm{Cov}\left(\hat{\theta}\right)$ in Equation 2 is inversely related to the Fisher information matrix $\mathbf{M} = \mathbf{X'WX}$ and the larger elements in $\mathbf{M}$, the larger the Fisher information of a design. Therefore, our aim is to find the D-optimum design ξ^* that maximizes the criterion $\Psi\{\mathbf{M}(\xi)\} = |\mathbf{M}(\xi)|$ among all ξ over Δ.

In this paper, we use a sequential construction algorithm, which sequentially adds the point in Δ at which the determinant of the information matrix after N points $|\mathbf{M}(\xi_N)|$ is a maximum (Atkinson *et al.*, 2007, section 11.2). To do this, we first choose $N_0 = p+1$ starting points from Δ and refer to these as the initial design ξ_{N_0}. We then compute the information matrix of the initial design $\mathbf{M}(\xi_{N_0})$ and evaluate the increase in this information matrix by adding each of the candidate points in Δ. Finally, we add the candidate design point that results in the largest possible increase in $|\mathbf{M}(\xi_{N_0})|$ to the initial design ξ_{N_0}. The result is a new $N_1 = N_0+1$-point design ξ_{N_1}, and it is further

improved by adding the point in Δ that results in the highest increase in $|\mathbf{M}(\xi_{N_1})|$. We replicate this procedure 1000 times to find an approximation of the D-optimal continuous design. The more replications are included in the algorithm, the better the approximation of the D-optimum design. It should be mentioned that as the design region Δ is continuous and the variable X can take any value on this continuous scale, the search of the sequential algorithm is carried over a grid on the design region Δ. In our calculations, we subdivided Δ with a step size of 0.001.

3.1 Weibull survival function

The variances and covariances of the parameter estimates depend on the underlying survival function $S\left(t_k\right)$. We remind the reader that the current study considers trials with a continuous underlying time to event variable T that is measured discretely in discrete intervals $[t_{k-1}, t_k)$ for $k = 1, 2, \cdots, p$. We further assume that all time intervals have equal length that is fixed beforehand in order to enable a comparison between hazard probabilities across time periods. Different continuous-time survival functions can be used to model survival in the baseline group. This study focuses on the continuous-time Weibull survival function, which constructs a flexible hazard function that decreases, increases, or remains constant over time. The Weibull survival function is given by $S(t) = e^{-\lambda t^{\tau}}$, and its hazard rate is given by $h(t) = \lambda\tau t^{\tau-1}$. The parameters τ and λ are shape and scale parameters, respectively, and t determines time that has elapsed in the study. Time is rescaled between 0 and 1 with 0 as the beginning of the trial (t_0) and 1 as the end of the trial (t_p). The scale parameter λ can be replaced by $-\log(1-\omega)$ with $\omega \in [0, 1]$ as the overall proportion of subjects in the baseline group who have experienced the event by the end of the study, that is, at $t_p = 1$.

The shape parameter τ determines the shape of the hazard function. For $\tau < 1$, the hazard rate decreases over time and corresponds to trials where the risk is concentrated toward the beginning of the study. For $\tau > 1$, the hazard rate increases over time and indicates trials with the highest risk concentrated toward the end. A value $\tau = 1$ implies a constant hazard rate over time, and the corresponding survival function is called the exponential survival function. The baseline discrete-time survival probability at the end of the kth discrete interval $S\left(t_k\right)$ is computed by evaluating the Weibull survival function at time point $t_k = \frac{k}{p}$, and the related hazard probability $h\left(t_k\right)$ is calculated by using the formula $h\left(t_k\right) = \frac{S(t_{k-1}) - S(t_k)}{S(t_{k-1})}$. We assume that the survival and hazard probabilities of the baseline group follow from the Weibull distribution and the predictor effect β quantifies the difference in the value of discrete-time hazard probability (on the logit scale) in every time period per unit difference in the predictor X.

In this paper, we assume the duration of trials is 12 time periods (i.e., $p = 12$) that might be considered as the length for 1-year trials given the time periods represent months or trials with weekly measurements and a duration of 12 weeks. We also find optimal designs for trials with a smaller number of time periods (i.e., $k < 12$) to study to what extent optimal designs depend on the number of time periods; recall that for each number of time periods, we use the same scale on the continuous predictor X. It should

be noticed that as the period length is fixed, a trial with 12 time intervals has maximum study duration and a trial with k ($k < 12$) time intervals has a shorter duration. We note that the first k elements of $S\left(t_k\right)$ and $h\left(t_k\right)$ for trials with 12 time periods are the survival and hazard probabilities of trials with a shorter duration of k time periods.

4 Numerical results

We choose values $\tau = 0.5, 1, 2$, and $\omega = 0.2, 0.5, 0.8$ for the parameters of the survival function. We take different negative ($\beta = -2.0, -1.5, -0.5$) and positive ($\beta = +2.0, +1.5, +0.5$) values of the predictor effect β into account. When $\beta > 0$, the probability of experiencing an event increases as X increases, while for $\beta < 0$, the risk of event occurrence decreases with increasing X. In addition, we study different values for the attrition rate r ($r = 0, 0.05, 0.1, 0.25$) and take the design region $\Delta = [0.75, 1]$ into account.

4.1 Locally D-optimal designs

Equation 2 shows that $\mathrm{Cov}\left(\hat{\theta}\right)$ is a complicated function of the design matrices $\mathbf{X}$ and weights π_j. It is therefore very cumbersome to find the optimal designs analytically, and so we compute them numerically and present our results using a series of graphs. All calculations were performed in the program R, and the codes are included in the Appendix. For $k = 1$, we observe that the D-optimal design points are equal to the boundaries of the design regions. This can be explained by the fact that for a bounded design region $\Delta = [x_{\min}, x_{\max}]$, the D-optimal design points will become equal to the boundaries $x_{\min}$ and $x_{\max}$ when $x_1^* < x_{\min} < x_{\max} < x_2^*$, where x_1^* and x_2^* are the optimal design points for an unbounded design region $\Delta = (-\infty, +\infty)$ (Berger & Wong, 2009, section 5.3.3). From our calculations, this also applies to more than one period ($k > 1$). Another observation is that D-optimal weights at the boundaries are always equal to 0.5 when $k = 1$, which confirms the findings in the literature for the logistic regression model for one time period (Sebastiani & Settimi, 1997; Mathew & Sinha, 2001). We now discuss results of the D-optimal design weights and start with the combination $\omega = 0.2$ and $\tau = 2$ of the Weibull survival parameters.

When $\omega = 0.2$, 20% of subjects with $X = 0$ have experienced the target event by the end of a study with 12 time points and the value $\tau = 2$ implies that the probability of experiencing the event is lowest at the beginning of the study and increases over time. Figure 1 displays the D-optimal design weights as a function of the number of time periods k (on the vertical axis) for the design region $\Delta = [0.75, 1]$. The design weights corresponding to negative values of β ($-2.0, -1.5,$ and -0.5) are displayed in the left plot, and the weights corresponding to positive values of β ($+2.0, +1.5, +0.5$) in the right. In each plot, the open symbols represent the D-optimal weights $\pi_{\min}$ at the lower boundary ($x_{\min} = 0.75$), and the closed symbols represent the D-optimal weights $\pi_{\max}$ at the upper boundary ($x_{\max} = 1$). Different types of symbols correspond to different values of β: $\beta = -0.5$ or $+0.5$ (circle), $\beta = -1.5$ or $+1.5$ (square), and $\beta = -2.0$ or $+2.0$ (triangle). We want to emphasize that as the sum of $\pi_{\min}$ and $\pi_{\max}$ is always equal to

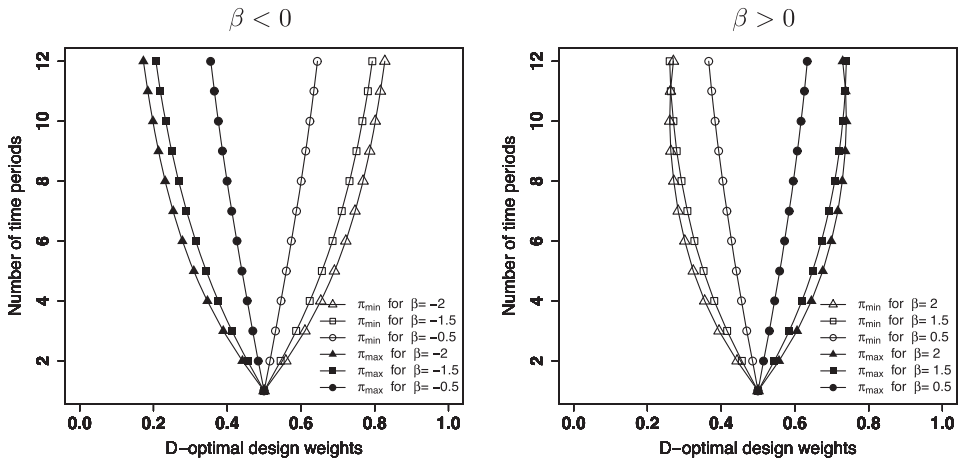

Fig. 1. Effect of β on D-optimal design weights when $\omega = 0.2$, $\tau = 2$, and $\Delta = [0.75, 1]$. In each plot, the vertical axis shows the number of time periods, and the horizontal axis shows the D-optimal design weights.

1, a decrease in $\pi_{\min}$ results in an increase in $\pi_{\max}$ and vice versa. Therefore, we only discuss results for $\pi_{\min}$, and the reader will observe the reverse pattern occurs for $\pi_{\max}$.

We first consider the plot for $\beta < 0$. As can be seen, $\pi_{\min}$ is equal to 0.5 for $k = 1$, and it increases when k increases. The increase in $\pi_{\min}$ depends on the value of β, and it is stronger for a more negative β. When $\beta > 0$, we observe that increasing the number of time periods from 1 to 12 results in a decrease in $\pi_{\min}$, and this decrease is larger for a larger β. However, the weights with $\beta = +2.0$ only slightly differ from those with $\beta = +1.5$, especially for higher values of k. In sum, we can conclude that the weights of the locally D-optimal design depend on how strongly the predictor X influences the outcome variable and on the direction of this effect. For instance, when $\beta < 0$, the chance of finding a paid job decreases by an increase in the proportion of the regular amount of social welfare, and it is then more efficient to offer the maximum reduction of 25% in social welfare to more than half of the unemployed people and welfare without any reduction to less than half of the people ($\pi_{\min} > \pi_{\max}$). The reverse occurs ($\pi_{\min} < \pi_{\max}$) if increasing the reduction in social welfare decreases the chance of finding a paid job ($\beta > 0$).

Now, we investigate to what extent the parameters ω and τ affect the optimal designs for a given β. Figure 2 displays the D-optimal design weights for $\omega = 0.2$. Each plot represents a different value of β; the plots in the upper half correspond to negative β values and those in the lower half correspond to positive β values. In this figure, the three different symbols indicate different τ values: $\tau = 2$ (circle), $\tau = 1$ (square), and $\tau = 0.5$ (triangle). The curves of $\pi_{\min}$ almost overlap when $\beta < 0$, which means that τ has a negligible effect on the optimal design weights when X has a negative effect on event occurrence. In contrast, if $\beta > 0$, $\pi_{\min}$ decreases as τ increases from 0.5 to 2 for a given k. Therefore, in trials where many subjects experience the event shortly after entering the study ($\tau < 1$), the weights at the boundaries are closer to 0.5 compared with cases

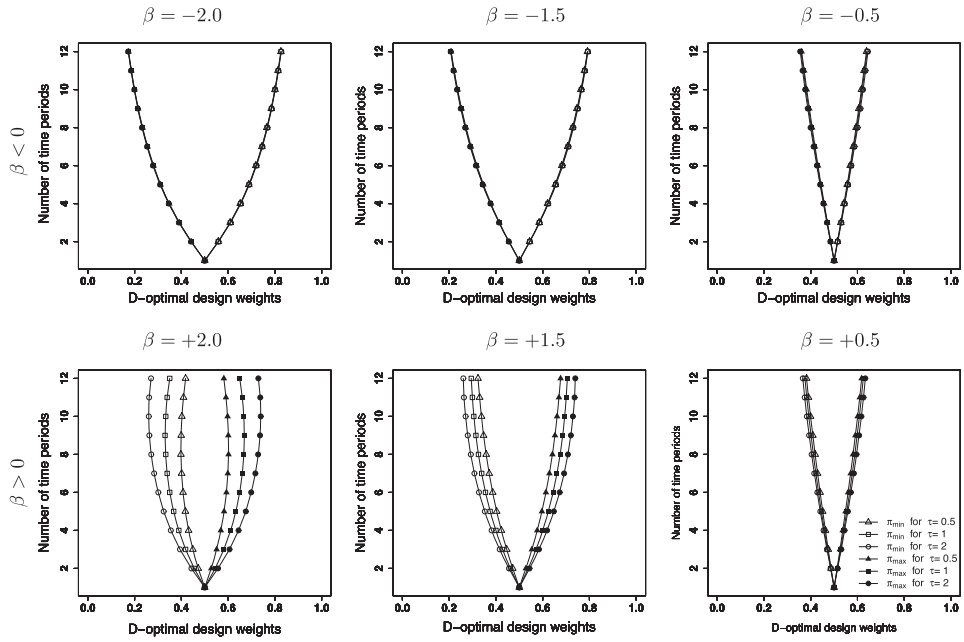

Fig. 2. Effect of τ on D-optimal design weights when $\omega = 0.2$ and $\Delta = [0.75, 1]$. In each plot, the vertical axis shows the number of time periods, and the horizontal axis shows the D-optimal design weights.

with $\tau > 1$. For instance, a lower proportion of subjects should be offered a reduction of 25% in social welfare when $\tau = 2$ in comparison with $\tau = 0.5$ if $\beta > 0$. Moreover, $\pi_{\min}$ fairly decreases with increasing the number of time periods k for any value of τ when $\beta > 0$, but as k becomes closer to 12, an increase in k causes a slight decrease in $\pi_{\min}$ when $\beta = +2.0$.

Now, we investigate the effect of ω on the optimal design weights for a given τ. Figure 3 shows plots similar to Figure 2 but for $\tau = 2$. Here, three different symbols represent three different values of ω: $\omega = 0.8$ (circle), $\omega = 0.5$ (square), and $\omega = 0.2$ (triangle). We observe that the curves of the design weights for different ω are fairly similar when $\beta = -2.0$ and $\beta = -1.5$. When $\beta = -0.5$, $\pi_{\min}$ slightly decreases with increasing ω for any given k. The design weights substantially depend on the value of ω when $\beta > 0$. It is observed that when $\beta = +0.5$, a larger $\pi_{\min}$ is found with a larger ω for any given k. In general, $\pi_{\min}$ decreases as k increases, but it becomes closer to 0.5 when $\omega = 0.8$ and k approaches 12.

The effect of ω on the D-optimal design weights becomes even more stunning when β increases further to +1.5 and +2.0. The pattern that we observe for $\omega = 0.2$ does not hold for larger values of ω. For $\omega = 0.2$, $\pi_{\min}$ mainly decreases if k increases from 1 to 12. For those cases with a larger ω, $\pi_{\min}$ first decreases if k increases, but at some value of k, it considerably increases with an increase in k. We also observe that the minimum value of $\pi_{\min}$ is found at a smaller value of k when $\omega = 0.8$ compared with that when $\omega = 0.5$. We remind the reader that when ω increases, the proportion of event

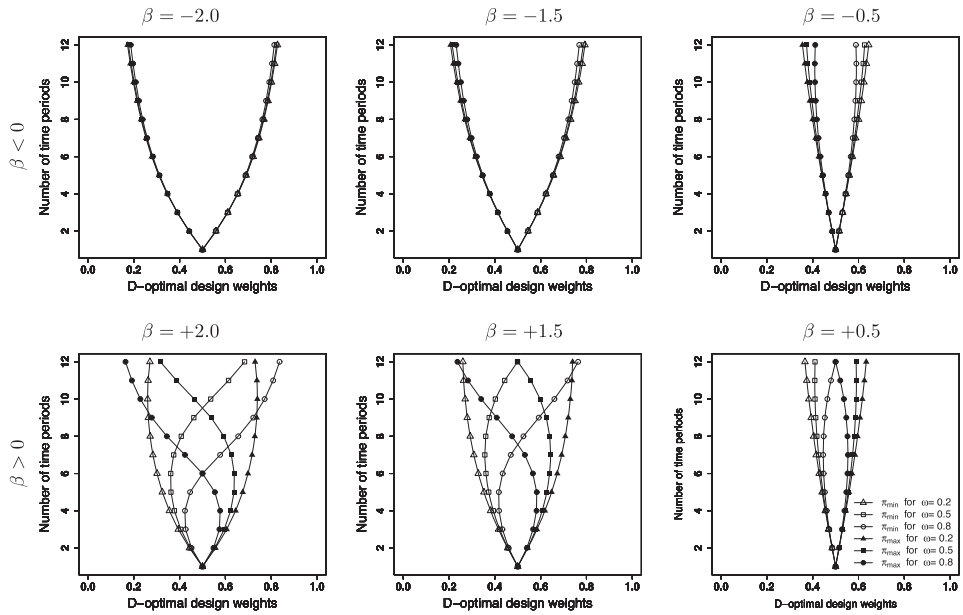

Fig. 3. Effect of ω on D-optimal design weights when $\tau = 2$ and $\Delta = [0.75, 1]$. In each plot, the vertical axis shows the number of time periods, and the horizontal axis shows the D-optimal design weights.

occurrence by the end of the study for $X=0$ increases. In studies with a larger ω and a small number of time periods ($k<5$), it is always most efficient to offer the total amount of social welfare to more than half of the subjects ($\pi_{\min} < \pi_{\max}$) if subjects receiving a lower amount of social welfare are less likely to find a paid job ($\beta \geq 1.5$). The reverse pattern becomes more efficient as the number of time periods increases. It should be emphasized that the consequence of the aberrant pattern for $\beta > 0$ is that the number of time periods should be fixed beforehand and should not be increased during the course of the study. If one does increase the number of time periods, then it may turn out the optimal design for the increased number of time period requires most weight at the lower boundary of the design space while the chosen design put most weight on the upper boundary. This could have consequences for design efficiency.

Figures 1–3 correspond to studies without attrition ($r=0$). We also investigated the effect of the attrition rate r on the D-optimal design weights. We observed that a slightly larger $\pi_{\min}$ is obtained with a larger attrition rate, which is more clear when ω is 0.5 or 0.8 (results not shown).

4.1.1 Robustness of the locally optimal designs

The asymptotic variance–covariance matrix Cov $\left(\hat{\theta}\right)$ in Equation 2, and consequently the D-optimality criterion, depends on the unknown parameters of interest, including β, ω, τ, and r. Thus, practitioners have to specify values of the parameters *a priori* before they can optimally design a study to estimate these parameters. One way to cir-

cumvent this difficulty is to assume that a good initial estimate is available based on a prestudy or subjective guesses and compute locally optimum designs as discussed earlier. It is, however, worth to study how robust these designs are against misspecification of the initial parameter estimate. We investigate the robustness of locally optimal designs presented in the previous figures against incorrect estimates of the true parameters β and ω. With respect to the parameter β, the robustness of a design based on an incorrect estimate for a given k is given in terms of RE as follows:

$$\mathrm{RE}\left(\xi^*_{\beta\neq\beta^*}|\xi^*_{\beta=\beta^*}\right) = \left[\frac{\mathrm{Det}\left[\mathbf{M}^{-1}\left(\xi^*_{\beta=\beta^*}\right)\right]}{\mathrm{Det}\left[\mathbf{M}^{-1}\left(\xi^*_{\beta\neq\beta^*}\right)\right]}\right]^{\frac{1}{k+1}},$$

where the numerator is the determinant of the variance–covariance matrix $\mathbf{M}^{-1}$ under the locally optimal design for the correct estimate of β (β^*) and the denominator is the determinant of $\mathbf{M}^{-1}$ under the locally D-optimal design for an incorrect estimate of β. They both are evaluated at the true $\beta = \beta^*$. A similar equation can be written with respect to the parameter ω.

The RE has a value between 0 and 1; it is equal to 1 for the design $\xi^*_{\beta=\beta^*}$. When $\mathrm{RE}(\xi^*_{\beta\neq\beta^*}|\xi^*_{\beta=\beta^*}) < 1$, $\xi^*_{\beta\neq\beta^*}$ is less efficient than $\xi^*_{\beta=\beta^*}$. The $\mathrm{RE}^{-1}(\xi^*_{\beta\neq\beta^*}|\xi^*_{\beta=\beta^*})$ indicates the relative amount of extra information that must be taken under design $\xi^*_{\beta\neq\beta^*}$ to be equally efficient as the other design $\xi^*_{\beta=\beta^*}$ (Atkinson *et al.*, 2007; Berger & Wong, 2009). For example, if $\mathrm{RE}(\xi^*_{\beta\neq\beta^*}|\xi^*_{\beta=\beta^*}) = 0.8$, then $(0.8^{-1} - 1) \times 100\% = 25\%$ more subjects are required under design $\xi^*_{\beta\neq\beta^*}$ to have the same efficiency as under design $\xi^*_{\beta=\beta^*}$. Therefore, relative efficiencies of 0.8 or 0.9 and closer to 1 are preferred.

The two plots of Figure 4 represent the RE as a function of initial estimates of β (on the left) and initial estimates of ω (on the right) for the design region $\Delta = [0.75, 1]$ and $r = 0$. The four curved lines represent the RE for different numbers of time periods k. We first discuss the plot on the left. Here, horizontal dotted lines represent 0.95 and 0.85 relative efficiency, and the vertical dotted line marks initial parameter estimate equal to the population parameter, which is $\beta = \beta^* = 1.0$. We presume researchers have some prior information about the shape of the survival function. For example, a Weibull distribution with an increasing hazard function $\tau = 2$ and $\omega = 0.2$ may be expected. The values of the initial β estimates are within the range $[-2, +2]$. The value of $\beta = +2$ results in a difference of about 17% in the survival probabilities $S_{\min}\left(t_k\right)$ and $S_{\max}\left(t_k\right)$ after 12 time periods, where $S_{\min}\left(t_k\right)$ gives the survival probability in period k for subjects with the value $x_{\min}$ on X, and $S_{\max}\left(t_k\right)$ denotes the survival probability for subjects with the value $x_{\max}$ on X. Correspondingly, the value of $\beta = -2$ leads to $S_{\max}(t_{12}) - S_{\min}(t_{12}) = 1.9\%$. As can be seen, even a severe misspecification of the true β results in a RE above 0.8, and so, locally optimal designs are rather robust to misspecification of true β. We observe that underestimating the true β affects the RE more severely than overestimation does especially so when the study duration is long. We also studied the same plot for other combinations of ω and τ and observed that the RE is always above 0.8, even with a large departure of the true β.

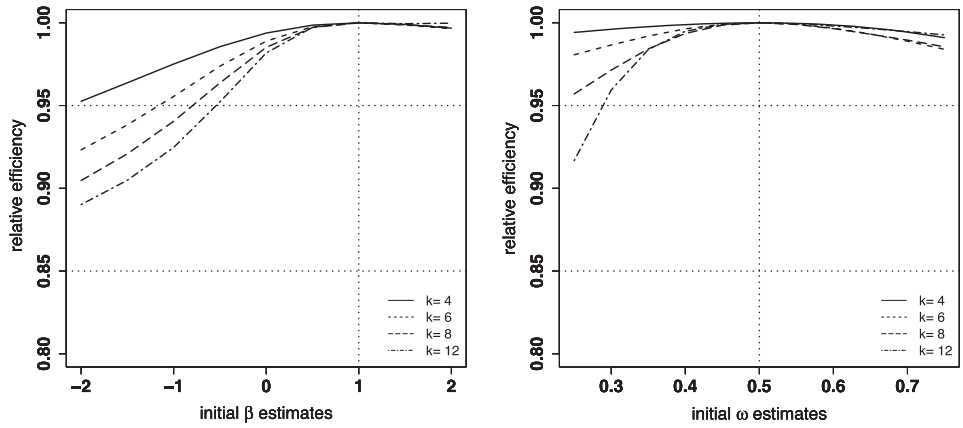

Fig. 4. Relative efficiency as a function of the initial β estimates (on the left) and the initial ω estimates (on the right) for different numbers of time periods k with the design space $\Delta = [0.75, 1]$ when $r = 0$. The dotted horizontal lines represent 0.95 and 0.85 relative efficiency and vertical dotted line marks initial parameter estimate equal to the population parameter, which is $\beta = \beta^* = +1.0$ (on the left) and $\omega = \omega^* = 0.5$ (on the right).

The plot on the right presents the RE as a function of initial estimates of the parameter ω when $\beta = 2.0$ and $\tau = 1$. Here, the true ω is $\omega^* = 0.5$, and the initial estimates are within the range $[0.25, 0.75]$. Even a large deviation from the true ω results in a RE above 0.9. The RE is very close to 1 for a minor departure from the true ω, and it drops as the incorrect estimate of the true ω deviates more from the population ω. We observe that underestimation of the true ω affects the RE more than overestimation. The decrease in RE is rather steep for a larger study duration k. We also studied the same plot for different β and τ values and observed that the decrease in RE is rather less sizable for a smaller β, but the RE does not change that much if another τ value is used. This means that if the population value of the predictor effect size β is smaller than +2, the loss in efficiency as a result of misspecification of ω becomes smaller; the population value of τ hardly influences the loss in efficiency, though.

4.1.2 Quadratic effect of the predictor

It is important to note that Model 1 only takes a linear effect of X into account. We also study optimal designs for a quadratic model by adding the term X^2 to Model 1 as follows:

$$\text{logit}\, h\left(t_{ik}\right) = \log \frac{h\left(t_{ik}\right)}{1 - h\left(t_{ik}\right)} = \sum_{k=1}^{p} \alpha_k D_{ik} + \beta_1 X_i + \beta_2 X_i^2,$$

where β_1 represents the linear effect of X on the probability of event occurrence and β_2 represents the quadratic effect of X. Figure 5 presents the optimal design points for different combinations of β_1 and β_2 as a function of the number of time periods k. Note that the horizontal axis in Figures 1–3 corresponds to the optimal design weights π. In

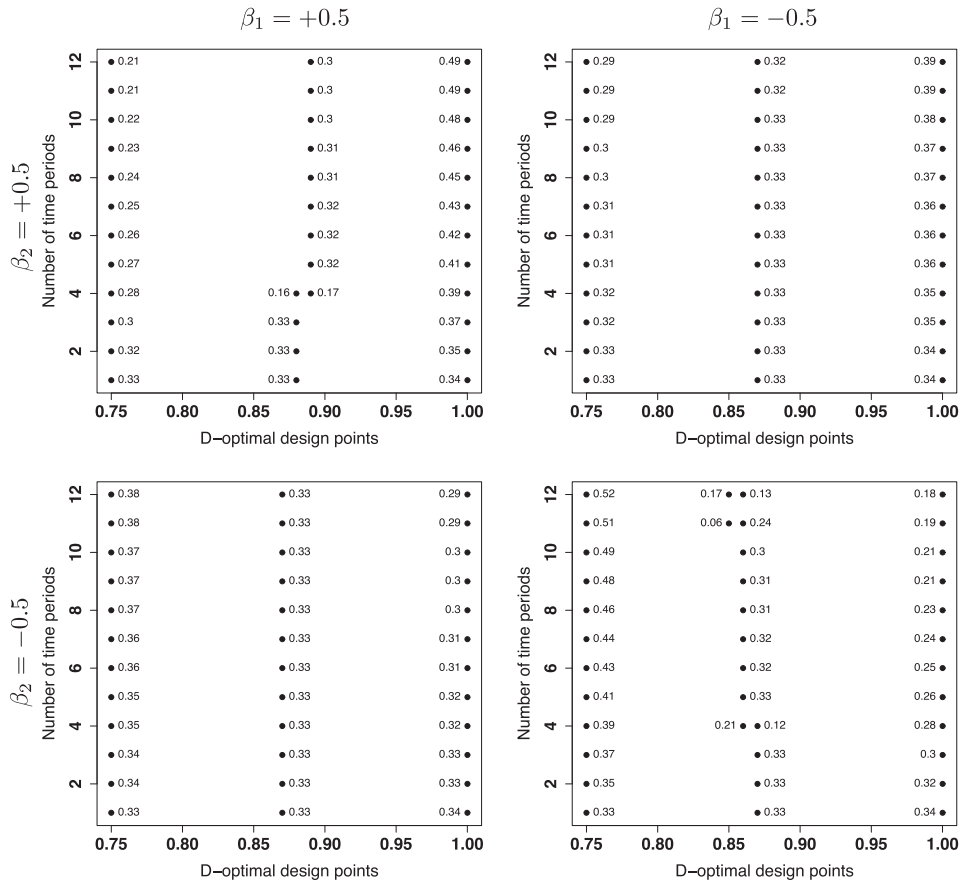

Fig. 5. Effect of $\boldsymbol{\beta} = (\beta_1, \beta_2)'$ on D-optimal design points and weights for the quadratic model when $\omega = 0.2$, $\tau = 2$ and $\Delta = [0.75, 1]$. In each plot, the vertical axis shows the number of time periods, and the horizontal axis shows the D-optimal design points. The numbers next to the dots show D-optimal design weights corresponding to the points.

Figure 5, however, this axis corresponds to the optimal design points x. The values next to the points in Figure 5 indicate the mass distributions (π_j) at these points. As it can be observed, more than two design points are needed to optimally estimate parameters in the quadratic model. When $k = 1$, optimal designs have three design points, two end points of the interval Δ and an additional point mainly located in the middle of Δ, with almost equal weights at these points. The weight at the lower boundary $\pi_{\min}$ decreases with an increase in the number of time periods (k) when $\beta_2 > 0$; this is stronger if $\beta_1 > 0$ compared with $\beta_1 < 0$. However, the weight $\pi_{\min}$ increases as k increases when $\beta_2 < 0$; this is stronger if $\beta_1 < 0$ compared with $\beta_1 > 0$. The opposite holds for $\pi_{\max}$. It is important to note that in designs for a linear effect of X, an increase or a decrease in $\pi_{\min}$ with an increase in k depends on the value of β_1 (Figure 1). For the quadratic model, however, it seems this depends on the β_2 value. In Figure 5, the quadratic effect

is high ($\beta_2 = 0.5$). We studied if this is also the case with a smaller β_2 (e.g., $\beta_2 = 0.2$) and observed that the relation between $\pi_{\min}$ and k still depends on β_1 in a quadratic model with a small quadratic effect. What is more, the mass at the middle point hardly changes as k changes. For some cases, it is also observed that the middle point moves to a slightly larger or smaller value as the number of time periods increases, and for few values of k, even four design points are needed.

4.2 Optimal number of time periods

In the previous section, we presented optimal designs for any given number of time periods $k \in \{1, 2, \cdots, 12\}$. Here, we compare the optimal designs for a different number of time periods to obtain the optimal number of time periods. A larger number of time periods result in more information and consequently larger efficiency of a design but require a higher budget. To make a fair comparison between optimal designs with a different number of time periods, we take the total costs of each design into account because increasing k leads to taking more measurements, which increases the study costs. Budget restrictions are widely used in optimum design constructions (e.g., Tekle *et al.*, 2008a; Berger & Wong, 2009; Tack & Vandebroek, 2004). Here, we compare the optimal designs with an equal sample size N at baseline. The relative efficiency (RE) of $\xi^*_{K_1}$ compared with $\xi^*_{K_2}$ is given by

$$\mathrm{RE}\left(\xi^*_{K_1}|\xi^*_{K_2}\right) = \frac{\mathrm{Det}\left[\mathbf{M}^{-1}\left(\xi^*_{K_2}\right)\right]^{\frac{1}{K_2+1}}}{\mathrm{Det}\left[\mathbf{M}^{-1}\left(\xi^*_{K_1}\right)\right]^{\frac{1}{K_1+1}}} \times \frac{C_{\xi^*_{K_2}} - C_0}{C_{\xi^*_{K_1}} - C_0}.$$

Here, $\xi^*_{K_1}$ represents the D-optimal design for a study with K_1 time periods, and $\xi^*_{K_2}$ corresponds to the optimal design of a study with K_2 time periods. $\mathbf{M}^{-1}\left(\xi^*_{K_1}\right)$ and $\mathbf{M}^{-1}\left(\xi^*_{K_2}\right)$ are the variance–covariance matrices of the unknown parameters under designs $\xi^*_{K_1}$ and $\xi^*_{K_2}$, respectively. If $\mathrm{RE}\left(\xi^*_{K_1}|\xi^*_{K_2}\right) = 1$, the design $\xi^*_{K_1}$ is as efficient as the design $\xi^*_{K_2}$; if $\mathrm{RE}\left(\xi^*_{K_1}|\xi^*_{K_2}\right) < 1$, $\xi^*_{K_2}$ is more efficient than $\xi^*_{K_1}$; if $\mathrm{RE}\left(\xi^*_{K_1}|\xi^*_{K_2}\right) > 1$, $\xi^*_{K_2}$ is less efficient than $\xi^*_{K_1}$. The total costs under designs $\xi^*_{K_1}$ and $\xi^*_{K_2}$ are denoted by $C_{\xi^*_{K_1}} - C_0$ and $C_{\xi^*_{K_2}} - C_0$, respectively, where C_{ξ^*} is the total amount of money required for conducting a study with design ξ^* and C_0 is the initial cost for setting up the design. We assume that the initial costs are defined as those costs that are not directly connected to repeatedly sampling and measuring subjects. Therefore, we assume the same initial costs for designs with different number of time periods.

In this paper, we define three different cost functions corresponding to different types of follow-up that allow the total costs of the designs to vary across the

number of time periods. Similar cost functions can be found in Jóźwiak and Moerbeek (2012, 2013). Cost function I assumes that all subjects are followed up until the end of the study even if they experience the event in earlier periods. This cost function is realistic for studies where not only the primary event is of interest but other secondary outcomes are important as well. Cost function I for a continuous D-optimum design $\xi_K^* = \left\{ \begin{array}{cc} x_{\min} & x_{\max} \\ \pi_{\min_K} & \pi_{\max_K} \end{array} \right\}$ with K time periods is given as

$$C_{\xi_K^*} = C_0 + \pi_{\min_K} C_{\min} + \pi_{\max_K} C_{\max} + C_2(K+1) = C_0 + C_2 \left[\pi_{\min_K} f_2 + \pi_{\max_K} f_1 f_2 + (K+1) \right].$$

Here, $C_{\min}$ and $C_{\max}$ are the costs of a subject assigned to the values $x_{\min}$ and $x_{\max}$ of the design region $\Delta = [x_{\min}, x_{\max}]$, respectively. In our example, $C_{\min}$ can be thought of as the cost of paying 75% of the regular amount of social welfare to an unemployed person and $C_{\max}$ is the cost of paying the regular amount of social welfare. Obviously, $C_{\max}$ is always larger than $C_{\min}$. In addition, C_2 is the cost of obtaining one measurement for a given subject. It should be noted that $K+1$ indicates the number of measurements for a given subject, which is equal to the number of time periods K plus a baseline measurement.

The cost ratio $f_1 = \frac{C_{\max}}{C_{\min}}$ represents the cost of a subject assigned to $x_{\max}$ relative to the cost of a subject assigned to $x_{\min}$. Correspondingly, the cost ratio $f_2 = \frac{C_{\min}}{C_2}$ indicates the relative cost of a subject assigned to $x_{\min}$ to the cost of obtaining one measurement for a given subject. For the design region $\Delta = [0.75, 1]$, f_1 is equal to $f_1 = \frac{1}{0.75}$. In our study, we assume that the cost of sampling a new subject at $x_{\min}$ is at least twice as high as the cost of obtaining one measurement per subject, and we take values $f_2 = 2, 10,$ and 100 into account.

The first cost function is not appropriate for studies where subjects immediately leave the study after event occurrence. In such studies, the number of measurements for a given subject can be smaller than $K+1$ and should be replaced by the summation $\pi_{\min_K} \sum_{k=0}^{K} S_{\min}(t_k) + \pi_{\max_K} \sum_{k=0}^{K} S_{\max}(t_k)$. Then, cost function II is given by

$$C_{\xi_K^*} = C_0 + C_2 \left[\pi_{\min_K} f_2 + \pi_{\max_K} f_1 f_2 + \left(\pi_{\min_K} \sum_{k=0}^{K} S_{\min}(t_k) + \pi_{\max_K} \sum_{k=0}^{K} S_{\max}(t_k) \right) \right].$$

Finally, cost function III assumes that subjects drop out of the study not only because of event occurrence but also because of other unforeseen reasons. Therefore, this cost function is used for examining costs in studies with attrition and is obtained by replacing $K+1$ in cost function I with $\pi_{\min_K} \sum_{k=0}^{K} S_{\min}(t_k)(1-r)^k + \pi_{\max_K} \sum_{k=0}^{K} S_{\max}(t_k)(1-r)^k$.

For each combination of β, τ, ω, r, f_1, f_2, and k ($k = 1, 2, \cdots, 12$), we compute the determinant of the variance–covariance matrix of the corresponding D-optimal design and its total costs. The D-optimum design with the smallest $\mathrm{Det}\left[\mathbf{M}^{-1}\left(\xi_k^*\right)\right]^{\frac{1}{k+1}} \times \left(C_{\xi_k^*} - C_0\right)$ is chosen as the reference design, and the efficiencies of the other designs relative to the reference design are evaluated by using the RE.

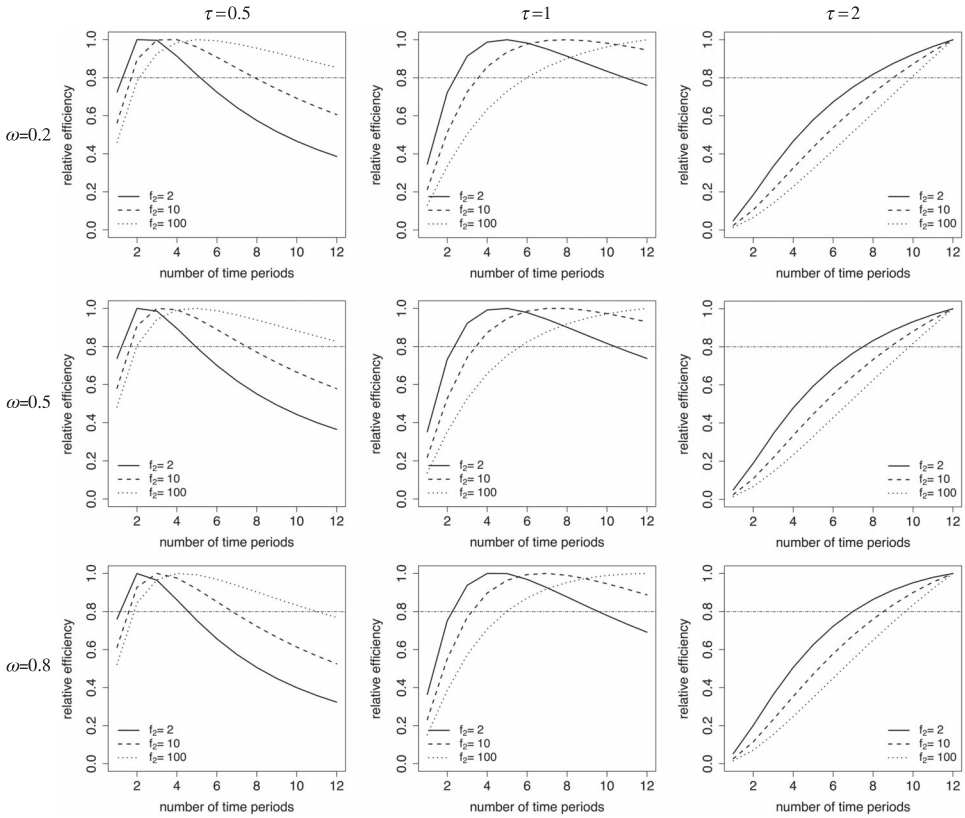

Fig. 6. Relative efficiency of D-optimal designs as a function of the number of time periods k when $\beta = -2$ and $\Delta = [0.75, 1]$.

The matrices of plots in Figures 6–7 show the relative efficiency as a function of the number of time periods k on the horizontal axes, the survival pattern τ in the columns ($\tau = 0.5, 1, 2$), and different ω values in the rows ($\omega = 0.2, 0.5, 0.8$) for the design region $\mathbf{\Delta} = [0.75, 1]$ and cost function I. Within each plot, the three lines represent different values of the cost of sampling a subject at $x_{\min}$ ($C_{\min}$) relative to the cost of once measuring a subject (C_2): $f_2 = 2$ (solid line), $f_2 = 10$ (dashed line), and $f_2 = 100$ (dotted line). Figure 6 corresponds to a negative value of β ($\beta = -2.0$), and Figure 7 displays the relative efficiency for a positive value of β ($\beta = +2.0$). In general, we observe that the optimal number of time periods either increases or remains constant as the cost ratio f_2 increases for a given combination of β, τ, and ω. Now, we discuss the results of each figure separately.

Figure 6 indicates that for a given ω, the optimal number of time points for a given cost ratio f_2 increases with an increase in τ. The reason is that for $\tau = 0.5$, the event is more likely to occur by the beginning of the study, and so it is more cost-effective to recruit more subjects instead of taking more measurements. But this is not a very efficient design when the event is more likely to occur at the end of the study ($\tau = 2$).

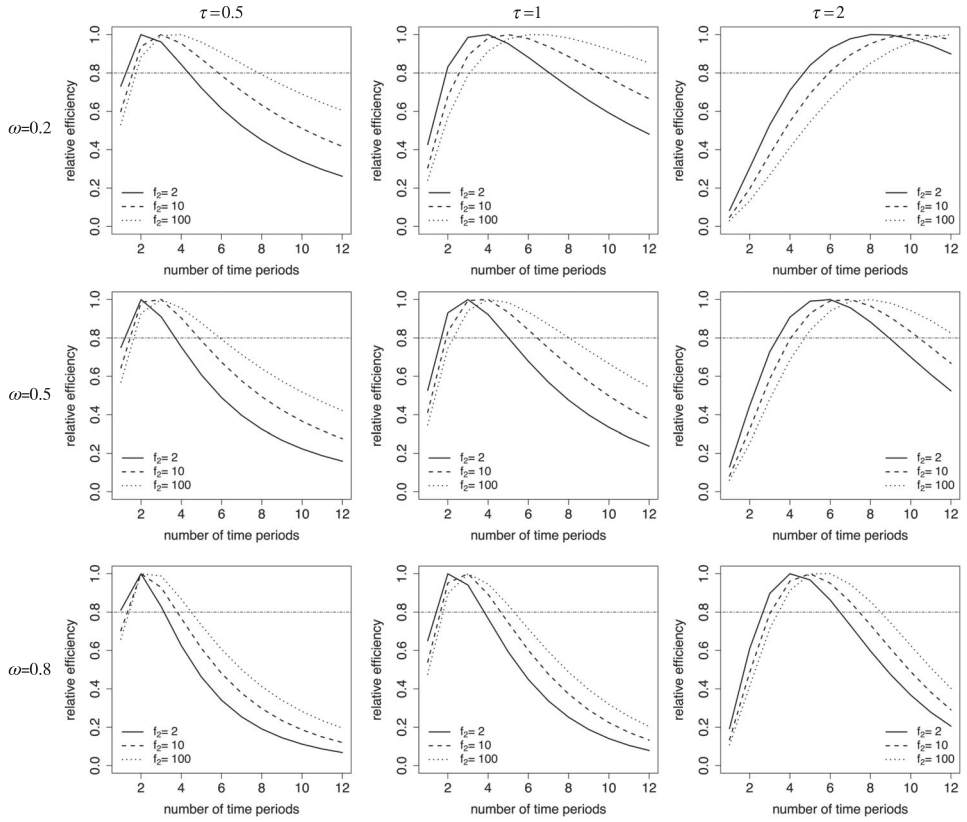

Fig. 7. Relative efficiency of D-optimal designs as a function of the number of time periods k when $\beta = +2$ and $\Delta = [0.75, 1]$.

For a given τ, changing ω only slightly changes the relative efficiency corresponding to a given cost ratio f_2. So, we can conclude that there is a main effect of τ but not a main effect of ω on the optimal number of time periods when $\beta < 0$.

As can be seen in Figure 7, increasing τ from 0.5 to 2 increases the optimal value of k for a given f_2; however, this effect highly depends on the value of ω. To illustrate this point, consider $\omega = 0.2$ and $f_2 = 100$. When $\tau = 0.5$, the optimal value of k is small. When $\tau = 1$, the optimal value of k is in the middle of the range $[1, 12]$, and as τ increases to 2, the best design is the one with a high number of time periods. The increase becomes less severe when ω becomes larger. We also observe that to what extent ω influences the optimal number of time periods highly depends on the value of τ. When $\tau = 0.5$ or $\tau = 1$, increasing ω from 0.2 to 0.8 results in a somewhat smaller optimal value of k for a given f_2, but an increase in ω largely decreases the optimal value of k when $\tau = 2$. In cases with $\tau = 2$, it is more efficient to choose a design with many measurements (i.e., many time periods) when $\omega = 0.2$, but such designs result in a high loss of efficiency when ω becomes larger. With a larger ω, a design with a smaller number of time periods should be taken, which results in recruiting more

subjects at $x_{\min}$ in favor of taking fewer measurements, even if it is more expensive to include subjects than to take measurements (i.e., $f_2 > 2$). Therefore, we can state that there is an interaction effect between the two factors τ and ω on the optimal value of k for $\beta = 2$.

Table 1 displays the optimal value of k for each cost function for different β, ω, τ, and f_2 values. In general, we observe that the optimal number of time periods does not change at all or slightly increases when cost function II is used instead of cost function I; the changes are somewhat larger when $\beta = +2.0$ than when $\beta = -2$. Therefore, subjects' leaving after event occurrence does not have a large impact on the optimal number of time periods. Table 1 also shows that the optimal value of k with cost function III is either equal to or smaller than those with the other two cost functions. The optimal k is in the range of 2 to 4 when $\tau = 0.5$ or $\tau = 1$, and it is between 4 and 9 when $\tau = 2$. We also observe that the cost ratio f_2 does not have as large an effect on the optimal number of time periods when cost function III is considered. As a result, in studies with dropout because of unforeseen reasons, a smaller number of time points should be taken into account compared with studies without attrition, even though the cost of sampling subjects at $x_{\min}$ relative to the cost of measuring them increases. Finally, it should be mentioned that for any combination of β, ω, f_2, and cost function type, the optimal number of time periods depends on τ, especially so when β is negative. So the optimal number of time periods is not robust against misspecification of τ.

5 An example: time to recovery from childhood malnutrition

Throughout this section, we consider the study of Le Roux *et al.* (2010) that investigated the effect of a paraprofessional home visiting program on improving childhood nourishment in South Africa. In this study, the outcome of interest was whether and when an underweight child reached an acceptable weight and time to rehabilitation to the target weight was measured at baseline and at 3-, 6-, 9-, and 12-month follow-ups ($p = 4$). This study showed that home visits by mentor mothers rehabilitated malnourished children to weights that are appropriate for their ages by providing nutrition education, improving feeding practices, and so on.

Suppose a specialist in nutrition wants to study the effect of the number of visits by mentor mothers on the occurrence and timing of achieving the normative weight. She assumes that each family is visited at least once and at most 12 times. So the design region is identified as $\Delta = [1, 12]$, with integer values only. To find initial estimates for the baseline logit hazard probabilities (i.e., $X = 0$), the researcher takes the probabilities of rehabilitation reported in Figure 2 of Le Roux *et al.* (2010) for underweight children in the control condition. The researcher believes that underweight children from families who are visited more often are more likely to reach a healthy weight. So, she presumes the value of $\hat{\beta} = +0.07$ for the effect of one-unit increase in the number of visits on the logit probability of rehabilitation of malnourished children. Figure 8 displays the discrete-time survival and logit hazard probabilities for 0, 1, 6, and 12 visits. As can be seen in the left panel, the value of $\beta = +0.07$ leads to a maximum increase of 24% in

Table 1. Optimal number of time periods for each cost function with different combinations of β, ω, τ, and f_2

		Predictor effect																	
		$\beta = -2$									$\beta = +2$								
Cost	Cost	$\omega = 0.2$			$\omega = 0.5$			$\omega = 0.8$			$\omega = 0.2$			$\omega = 0.5$			$\omega = 0.8$		
ratio	function	$\tau = 0.5$	$\tau = 1$	$\tau = 2$	$\tau = 0.5$	$\tau = 1$	$\tau = 2$	$\tau = 0.5$	$\tau = 1$	$\tau = 2$	$\tau = 0.5$	$\tau = 1$	$\tau = 2$	$\tau = 0.5$	$\tau = 1$	$\tau = 2$	$\tau = 0.5$	$\tau = 1$	$\tau = 2$
$f_2 = 2$	I	2	5	12	2	5	12	2	4	12	2	4	8	2	3	6	2	2	4
	II	2	5	12	2	5	12	2	5	12	3	4	11	3	4	7	2	3	5
	III	2	4	8	2	4	8	2	4	8	2	3	6	2	3	5	2	2	4
$f_2 = 10$	I	4	8	12	3	7	12	3	7	12	3	5	10	3	4	7	2	3	5
	II	4	8	12	3	8	12	3	7	12	3	6	12	3	4	8	2	3	5
	III	3	4	9	3	4	9	2	4	8	2	4	7	2	3	5	2	2	4
$f_2 = 100$	I	5	12	12	5	12	12	4	12	12	4	6	12	3	4	8	2	3	6
	II	5	12	12	5	12	12	4	12	12	4	7	12	3	4	8	2	3	6
	III	3	4	9	3	4	9	3	4	8	2	4	7	2	3	5	2	2	4

Note: Cost function III is considered with $r = 0.25$.

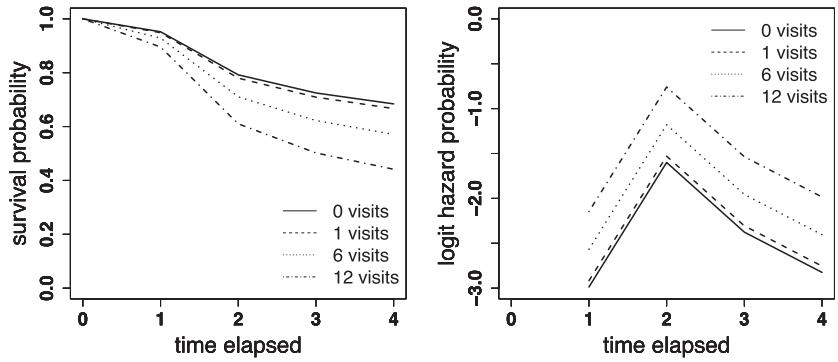

Fig. 8. Survivor function (left side) and logit hazard probability function (right side) for achieving normative weight corresponding to baseline children and those with different numbers of visits in the design region $\Delta = [1, 12]$.

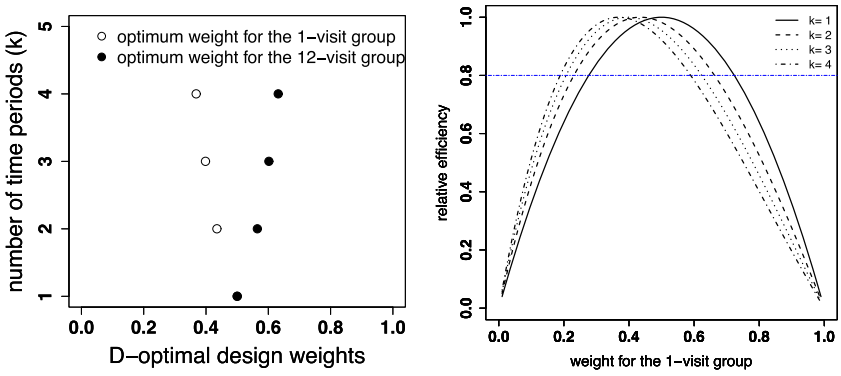

Fig. 9. D-optimal designs (left side) and relative efficiencies of other designs compared with the optimal designs (right side) for different number of time periods in the recovery from childhood malnutrition study.

the probability of an underweight child reaching a normative weight as compared with those who are not visited by mentor mothers. It also appears that the survival function does not follow the Weibull distribution, and so this example shows the use of the methodology presented in this paper for a different survival function.

D-optimal design weights of a study with 1-year follow-up ($k = 4$) or shorter ($k < 4$) are presented in the left panel in Figure 9, where the open circle indicates the optimal proportion of underweight children at the lower boundary and closed circle depicts the optimal proportion at the upper boundary. It is seen that the design for the model with one time period equally divides underweight children into two groups: one with children who are visited only once and the other with children who are visited 12 times. As the number of time periods increases from 1, the optimal designs include more underweight children in the 12-visit group and fewer children in the 1-visit group.

Suppose that the researcher is also interested in investigating the performance of other alternative designs compared with the optimal designs using RE as a measure

of comparison. The right panel in Figure 9 shows the REs of other designs with the same design points but different design weights relative to the optimal designs for a different number of time periods k. Note that the optimal designs presented in the left plot have a relative efficiency of 1 in the right plot. As can be illustrated from this plot, if the researcher wants to conduct a design with RE ≥ 0.8 when $k = 1$, she can take any value in the range of $[0.28, 0.72]$ as the proportion of underweight children for the one-visit group into account. However, this range decreases to smaller values as k increases. We also observe that the range of possible proportions at the one-visit group becomes slightly narrower with an increase in k.

6 Conclusion and discussion

Our results for the Weibull survival function indicate that the optimal designs for a linear model of X are two-point designs where the design points correspond to the design region's boundaries. The weight at the design points highly depends on both the value and the sign of β. More than this, the degree to which the optimal design weights depend on the survival parameters τ and ω is substantially determined by β: there is a very slight effect of τ and ω on the weights when $\beta < 0$, while these factors (especially ω) markedly influence the weights when $\beta > 0$, which becomes more noticeable when β becomes larger. We also observed that the attrition rate r does not have a large effect on the optimal design weights for any combination of the other factors. The other major finding was concerned with the effect of k on the optimal designs. We found that the optimal number of time points largely depends on τ but does not depend on ω when $\beta < 0$, whereas there is an interaction effect between τ and ω on the optimal value of k when $\beta > 0$, that is, the extent to which varying τ or ω changes the optimal k is sizably influenced by the other survival parameter. To efficiently estimate parameters in a quadratic model, however, more than two design points are needed, the two end points of the interval $\Delta = [x_{\min}, x_{\max}]$ and other points in the middle of the interval. The weights at the boundaries still depend on k, but the weights at the middle points hardly depend on k. It is worth noting that the procedure for obtaining D-optimal designs can be applied without the assumption of the Weibull survival function when data from the literature are available.

Jóźwiak and Moerbeek (2012, 2013) studied optimal designs for discrete-time survival data with one dichotomous predictor (e.g., treatment versus control), which are identified by the optimal combination of the number of subjects and time periods. We extended their findings to optimal designs with one continuous predictor, which are defined by optimal choices of design points and their weights for a given number of time periods. Although optimal designs in both studies have two design points, their corresponding weights are not comparable as they are based on a different optimality criterion. We also studied optimal designs for a quadratic effect of the continuous predictor, while Jóźwiak and Moerbeek (2012, 2013) could only focus on a linear effect. We want to emphasize that our results are in agreement with those of Sebastiani and Settimi (1997), Mathew and Sinha (2001), and others who have reviewed D-optimal designs in the logistic model with one continuous predictor and just one time period.

Our conclusion as to the effect of the continuous predictor is that the optimal design is sensitive to misspecification of the predictor effect (its value along with its sign). Moreover, the sensitivity of the optimal designs against misspecification of τ, ω, and k highly depends on the direction in which the predictor influences the outcome. Specifying the Weibull survival parameters requires more careful thought when the predictor has a positive effect on the outcome. Our results showed that a design that is optimal for one combination of τ and ω may not be so for another. Further, the fact that accounting for study costs is likely to influence the optimal value of k adds to our understanding that we must account for the cost differential when comparing optimal designs because an optimal design with many time periods is not always the most cost-efficient design for our study.

To compute locally optimal designs, researchers need to make initial estimates or educated guesses of the unknown model parameters based on findings in the literature, a pilot study, or a previous study. Although our study shows that locally optimal designs are relatively robust to misspecification of some parameter values, this is not necessarily the case in other applications. To overcome local optimality problem, a sequential procedure might be useful. To this end, we can divide the total number of iterations used to approximate optimal designs into some sets. The first set of iterations could be used to obtain locally optimal designs for the first initial parameter guess. This guess could be corrected by using the data to re-estimate the parameter and use the second set of iterations to correct the second guess and so on. Because these designs are not necessarily robust to poor initial parameter estimates, we suggest researchers to consider more advanced methods, such as minimax, maximin, or Bayesian design methods, if good initial estimates are not available. These alternative methods result in optimal designs with more and spread-out design points, which are reasonably robust within some prespecified regions of the values for unknown parameters. A study of minimax and maximin D-optimal designs for the logistic model (e.g., Sitter, 1992; King & Wong, 2000) and binary longitudinal trial responses (e.g., Tekle *et al.*, 2008b) is available in the literature. A review of optimal linear and non-linear Bayesian designs can be found in Chaloner and Verdinelli (1995). Additional research will need to be performed to study a procedure for obtaining more robust designs with discrete-time survival data if precise knowledge of the parameter estimates is unavailable.

Acknowledgement

This research was supported by a VIDI grant from the Netherlands Organization for Scientific Research (NWO) number 452-08-004.

Appendix A: R-syntax to run the sequential construction algorithm

This Appendix provides a code for running the sequential construction algorithm of optimum designs for trials with discrete-time survival endpoints. We assume an estimate of the Weibull survival parameters (τ, ω), as well as the predictor effect size (β), is available. The search of the sequential algorithm is carried over a grid on the design region $\Delta = [a_0, a_1]$.

```
# the main program
Sequential.Construction <-
function(initial, beta, alpha, period, b, a0, a1, step){
X <- seq(a0, a1, by=step)
W.X <- matrix(NA, nrow=length(X), ncol=period)
S.X <- matrix(NA, nrow=length(X), ncol=period)
for(i in 1:length(X)){
Xi <- X[i]
W.X[i, ] <- Maxperiod(p, beta, alpha, Xi)$Weight[1:period, ,drop=F]
S.X[i, ] <- Maxperiod(p, beta, alpha, Xi)$S.prob[1:period, ,drop=F]
                    }
x0 <- initial
for(q in 1:b){
W.x0 <- matrix(NA, nrow=nrow(x0), ncol=period)
S.x0 <- matrix(NA, nrow=nrow(x0), ncol=period)
for(j in 1:nrow(x0)){
xj <- x0[j, ]
W.x0[j, ] <- Maxperiod(p, beta, alpha, xj)$Weight[1:period, ,drop=F]
S.x0[j, ] <- Maxperiod(p, beta, alpha, xj)$S.prob[1:period, ,drop=F]
                    }
InformM <- matrix(0, period+1, period+1)
for(r in 1:nrow(x0)){
D.x0_r <- cbind(diag(1, nrow=period, ncol=period),x0[r, ])
W.x0_r <- as.vector(W.x0[r, ])
S.x0_r <- as.vector(S.x0[r, ])
contrib.x0_r <- matrix(0, period+1, period+1)
for (k in 1:period) {
contrib.x0_rk <- crossprod(D.x0_r[k,,drop=F],D.x0_r[k,,drop=F])
                 *W.x0_r[k]*S.x0_r[k]
contrib.x0_r <- contrib.x0_r + contrib.x0_rk
                     }
InformM <- InformM + contrib.x0_r
                   }
Det.InformM <- matrix(NA, nrow=length(X))
for(r in 1:length(X)){
D.X_r <- cbind(diag(1, nrow=period, ncol=period),X[r])
W.X_r <- as.vector(W.X[r, ])
S.X_r <- as.vector(S.X[r, ])
contrib.X_r <- matrix(0, period+1, period+1)
for (k in 1:period) {
contrib.X_rk <- crossprod(D.X_r[k,,drop=F],D.X_r[k,,drop=F])
                *W.X_r[k]*S.X_r[k]
contrib.X_r <- contrib.X_r + contrib.X_rk
      }
```

```
Inform_r <- InformM + contrib.X_r
Det.InformM[r, ] <- det(Inform_r)
                        }
Max.Det.InformM <- which(Det.InformM==max(Det.InformM),arr.ind=TRUE)[1,1]
New.support.point <- X[Max.Det.InformM]
x0 <- rbind(x0, New.support.point, deparse.level = 0)
                                                }
mytable <- rle(sort(x0))
mytable <- data.frame(number=mytable$values, n=mytable$lengths)
list(designPoints=mytable[ ,1],designWeights=prop.table(table(x0)))
                 }
# compute baseline logit hazard probabilities for trials with p intervals
Weibull <- function(p, w, tau){
t <- seq(0, 1, l=(p+1))
lambda <- -log(1-w)
survival <- matrix(NA, nrow=length(t), ncol=1)
hazard.prob <- matrix(NA, nrow=length(t)-1, ncol=1)
Talpha <- matrix(NA, nrow=length(t)-1, ncol=1)
survival[, 1] <- exp(-(lambda*(t^tau)))
for(j in 2:(p+1)){
hazard.prob[j-1, 1] <- (survival[j-1, 1] - survival[j, 1])/survival[j-1, 1]
Talpha[j-1, 1] <- log(hazard.prob[j-1, 1]/(1-hazard.prob[j-1, 1]))
                 }
list(alpha = Talpha[, 1])
                                    }
# compute survival probabilities and least squares weights for any value
of the design region
Maxperiod <- function(p, beta, alpha, x) {
theta <- matrix(c(alpha, beta), nrow=p+1)
x.temp <- cbind(diag(1, nrow=p, ncol=p),x[1])
logit.x <- x.temp %*% theta
hazard.x <- 1/(1 + exp(-logit.x))
weight.x <- hazard.x * (1 - hazard.x)
survival.x <- matrix(NA, nrow=p, ncol=1)
survival.x[1, 1] <- 1 - hazard.x[1, 1]
if(p > 1){
for(s in 2:p){survival.x[s, 1] <- survival.x[s-1, 1] * (1 - hazard.x[s, 1])}
}
survival.x.Final <- rbind(1, survival.x, deparse.level = 0)
list(Weight = weight.x , S.prob = survival.x.Final)
                                           }
```

References

Atkinson, A. C., A. N. Donev and R. D. Tobias (2007), *Optimum experimental design, with SAS*, Clarendon, Oxford.

Berger, M. P. F. and W. K. Wong (2009), *An introduction to optimal designs for social and biomedical research*, Wiley, Chichester.

Chaloner, K. and I. Verdinelli (1995), Bayesian experimental design: a review, *Statistical Science* **103**, 273–304.

Chasalow, S. D. (1992), Exact optimal response surface designs with random block effects, PhD.

Donev, A. N. and A. C. Atkinson (1988), An adjustment algorithm for the construction of exact D-optimum experimental designs, *Technometrics* **30**, 429–433.

Galbraith, S. and I. C. Marschner (2002), Guidelines for the design of clinical trials with longitudinal outcomes, *Controlled Clinical Trials* **23**, 257–273.

Hsieh, F. Y. and P. W. Lavori (2000), Sample-size calculations for the Cox proportional hazards regression model with nonbinary covariates, *Controlled Clinical Trials* **21**, 552–560.

Jóźwiak, K. and M. Moerbeek (2012), Cost-effective designs for trials with discrete-time survival endpoints, *Computational Statistics and Data Analysis* **56**, 2086–2096.

Jóźwiak, K. and M. Moerbeek (2013), Optimal treatment allocation and study duration for trials with discrete-time survival endpoints, *Journal of Statistical Planning and Inference* **143**, 971–982.

King, J. and W. K. Wong (2000), Minimax D-optimal designs for the logistic model, *Biometrics* **56**, 1263–1267.

Le Roux, I. M., K. Le Roux, W. S. Comulada, E. M. Greco, K. A. Desmond, N. Mbewu and M. J. Rotheram-Borus (2010), Home visits by neighborhood mentor mothers provide timely recovery from childhood malnutrition in South Africa: results from a randomized controlled trial, *Nutrition Journal* **9**, 1–56.

Lima Passos, V., F. E. S. Tan and M. P. F. Berger (2011), Cost-efficiency considerations in the choice of a microarray platform for time course experimental designs, *Computational Statistics and Data Analysis* **55**, 944–954.

Lucas, J. M. (1974), Optimum composite designs, *Technometrics* **16**, 561–567.

Mathew, T. and B. K. Sinha (2001), Optimal designs for binary data under logistic regression, *Journal of Statistical Planning and Inference* **93**, 295–307.

McCullagh, R. M. and J. A. Nelder (1989), *Generalized linear models*, Chapman and Hall, London.

Moerbeek, M. (2008), Powerful and cost-efficient designs for longitudinal intervention studies with two treatment groups, *Journal of Educational and Behavioral Statistics* **33**, 41–61.

Ouwens, M. J. N. M., F. E. S. Tan and M. P. F. Berger (2006), A maximin criterion for the logistic random intercept model with covariates, *Journal of Statistical Planning and Inference* **136**, 962–981.

Sebastiani, P. and R. Settimi (1997), A note on D-optimal designs for a logistic regression model, *Journal of Statistical Planning and Inference* **59**, 359–368.

Singer, J. D. and J. B. Willett (1993), Its about time: using discrete-time survival analysis to study duration and the timing of events, *Journal of Educational and Behavior Statistics* **18**(2), 155–195.

Singer, J. D. and J. B. Willett (2003), *Applied longitudinal data analysis. Modeling change and event occurrence*, Oxford University Press, Oxford.

Sitter, R. R. (1992), Robust designs for binary data, *Biometrics* **48**, 1145–1155.

Tack, L. and M. Vandebroek (2004), Budget constrained run orders in optimum design, *Journal of Statistical Planning and Inference* **124**, 231–249.

Tekle, F. B., F. E. S. Tan and M. P. F. Berger (2008a), D-optimal cohort designs for linear mixed-effects model, *Statistics in Medicine* **27**(14), 2586–2600.

Tekle, F. B., F. E. S. Tan and M. P. F. Berger (2008b), Maximin D-optimal designs for binary longitudinal responses, *Computational Statistics and Data Analysis* **52**(12), 5253–5262.

Received: 30 July 2013. Revised: 29 October 2014.