Optimal blocked minimum-support designs for non-linear models

P.M. van de Ven[⋆] & D.C. Woods[†1]

[⋆]Department of Epidemiology and Biostatistics, VU University Medical Center,
Amsterdam, The Netherlands
[†]Southampton Statistical Sciences Research Institute, University of Southampton,
Southampton SO17 1BJ, UK

Finding optimal designs for experiments for non-linear models and dependent data is a challenging task. We show how the problem simplifies when the search is restricted to designs that are minimally supported; that is, the number of distinct runs (treatments) is equal to the number of unknown parameters, p, in the model. Under this restriction, the problem of finding a locally or pseudo-Bayesian D-optimal design decomposes into two simpler problems that are more widely studied. The first is that of finding a minimum-support D-optimal design d_1 with p runs for the corresponding model for the mean but assuming independent observations. The second problem is finding a D-optimal block design for assigning the treatments in d_1 to the experimental units. We find and assess optimal minimum-support designs for three examples, each assuming a mean model from a different member of the exponential family: binomial, Poisson and normal. In each case, the efficiencies of the designs are compared to the optimal design where the restriction on the number of distinct support points is relaxed. The optimal minimum-support designs are found to often perform satisfactorily under both local and Bayesian D-optimality for concentrated prior distributions. The results are also relatively insensitive to the assumed degree of dependence in the data.

Keywords: Binary data; block designs; count data; D-optimality; Generalized Estimating Equations; Generalized Linear Models; Michaelis Menten model.

1. Introduction

Many experiments aim to model a non-linear relationship between a response and several explanatory variables. If a binary or count response variable is observed, an appropriate Generalized Linear Model (GLM) can be assumed to describe this relationship. In other cases, the outcome may be continuous and normally distributed, but the relationship between the mean and the explanatory variables may be non-linear. The data observed from different runs in the experiment are often assumed to be independent and the parameters in the model are estimated using maximum likelihood estimation techniques. It is, however, not uncommon for experiments to be performed in blocks; that is, different runs are performed, for example, on different days, by different scientists or operators, or using different batches of material. Such situations may induce dependence between the

observations within a block, whilst observations in different blocks remain independent. An estimation method that takes into account this dependence structure increases the accuracy of inferences made from the experimental data. The blocking structure should also be taken into account when designing the experiment.

Most literature on design for non-linear models and dependent data has been concerned with nonlinear mixed effect models, see the seminal paper of Mentré et al. (1997), typically in a clinical or biological setting (see also Han and Chaloner, 2004; Atkinson, 2008, and references therein). Such methodology assumes a conditional, or subject-specific, modeling approach, with potentially differing model parameters for each block. In this paper, we find designs for marginal, or population-averaged, models where the dependencies in the data do not arise from subject-specific parameters; see also Hughes-Oliver (1998) and Atkinson and Ucinski (2004). For the linear model, where there has been substantial work on designs with random block effects (see, for example, Goos and Vandebroek, 2001), these two modeling paradigms coincide.

Recently, the first general results and methods for finding optimal blocked designs for discrete data have been published. Niaparast (2009) presented local D-optimal designs for Poisson regression models with a random intercept in the linear predictor and a single explanatory variable. Woods and van de Ven (2011) presented approaches for finding D-optimal designs for non-normal data and generalized estimating equations (GEE models) and any number of explanatory variables.

Designs for nonlinear models generally suffer from their performance depending on the values of the unknown model parameters. Hence, either an initial guess of these parameters is required (Chernoff, 1953), perhaps as part of a sequential strategy (Dror and Steinberg, 2008), or a Bayesian, or pseudo-Bayesian, design is required; see Han and Chaloner (2004), Woods et al. (2006) and Section 2 of this paper.

The focus in this paper is on designs that are minimally supported, i.e. the number of distinct design points, or treatments, is equal to the number of unknown parameters in the model; see also Cheng (1995). We present methodology for finding minimum-support block designs to estimate the model parameters in a nonlinear model with dependent observations for both continuous and discrete responses. The restriction to minimally supported designs allows a decomposition of the pseudo Bayesian D-optimality objective function, leading to two, simpler, optimization problems (Section 3). For some experiments, there may also be practical advantages in reducing the number of treatments employed, such as reduced cost; for example, when the experiment relies on the construction of templates or formulations in manufacturing or chemistry. The decomposition of the objective function allows either the analytic derivation of optimal designs or the computational complexity of finding optimal designs to be reduced and we show, through a series of examples (Section 4), that minimum-support designs can suffer only a minor loss in performance compared to unrestricted designs.

2. Models and design criteria

Consider a continuous or discrete response $Y(\mathbf{x})$ that may depend on the values taken by m explanatory variables $\mathbf{x}^{\mathrm{T}} = (x_1, x_2, \ldots, x_m)$. In an experiment, responses are observed for different settings of the explanatory variables according to a design d consisting of N experimental runs. In run j, the jth experimental unit receives treatment $\mathbf{x}_j^{\mathrm{T}} = (x_{1j}, \ldots, x_{mj})$, chosen from a bounded design space $\mathcal{X} \subset \mathbb{R}^m$ $(j = 1, \ldots, N)$. We assume that the units are arranged in b blocks of size k_l $(l = 1, \ldots, b)$ so that $k_1 + k_2 + \ldots + k_b = N$. The entries in the design $d = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ and the observations $\mathbf{Y} = (Y(\mathbf{x}_1), Y(\mathbf{x}_2), \ldots, Y(\mathbf{x}_N))^{\mathrm{T}}$ are ordered by block and by unit within block. All pairs of observations made in different blocks are assumed independent but observations within the same block may be dependent.

We are interested in finding efficient designs for estimating the unknown parameters $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^{\mathrm{T}}$ in a marginal model for the mean response $\mathrm{E}[Y(\mathbf{x})] = \mu(\mathbf{x}, \boldsymbol{\beta})$. We consider a general class of models for which the inverse of the model-based asymptotic variance-covariance matrix for an estimator $\widehat{\boldsymbol{\beta}}$ is of the form

$$\mathbf{M} = \mathbf{F}^T \mathbf{V}^{-1/2} \mathbf{R}^{-1} \mathbf{V}^{-1/2} \mathbf{F}, \tag{1}$$

where only the matrices $\mathbf{F}$ and $\mathbf{V}$ depend on the design d and the parameter values $\boldsymbol{\beta}$. The matrix $\mathbf{R}$ is a block-diagonal correlation matrix. Common examples include marginal models for correlated discrete data estimated using the GEE approach, under some assumptions, and nonlinear models with additive correlated normally distributed errors.

Under a GEE model for a discrete response, the mean and variance of the observations are assumed to come from an appropriate GLM such that $\mathrm{Var}[Y(\mathbf{x})] = \nu[\mu(\mathbf{x}, \boldsymbol{\beta})]/\phi$, where ϕ is a constant scale parameter and $\nu(\cdot)$ is the variance function of the GLM (Liang and Zeger, 1986). The mean is related to $\mathbf{x}$ through $g[\mu(\mathbf{x}, \boldsymbol{\beta})] = \mathbf{f}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta}$, where $g(\cdot)$ is the link function of the GLM and the product $\mathbf{f}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta}$ is the linear predictor, with the p-vector $\mathbf{f}(\mathbf{x})$ holding known functions of $\mathbf{x}$. The dependence in the data is modeled by means of a "working correlation" matrix $\mathbf{R}$, which is assumed known. This matrix will typically have a standard structure and is not necessarily equal to the actual correlation structure in $\mathbf{Y}$. The model-based estimator for the asymptotic variance-covariance matrix for the GEE estimator $\widehat{\boldsymbol{\beta}}$ is given by

$$\mathrm{Var}\left(\widehat{\boldsymbol{\beta}}\right) = \left(\mathbf{X}^T \boldsymbol{\Delta} \mathbf{V}^{-1/2} \mathbf{R}^{-1} \mathbf{V}^{-1/2} \boldsymbol{\Delta} \mathbf{X}\right)^{-1}, \tag{2}$$

where $\mathbf{X}$ is the $N \times p$ model matrix with rows $\mathbf{f}^T(\mathbf{x}_j)$, $\boldsymbol{\Delta} = \mathrm{diag}\{1/g'[\mu(\mathbf{x}_j, \boldsymbol{\beta})]\}$ and $\mathbf{V} = \mathrm{diag}\{\mathrm{Var}[Y(\mathbf{x}_j)]\}$; see Lee et al. (2006, p.75). The inverse of the variance-covariance matrix given in (2) is of the form specified in (1) with $\mathbf{F} = \boldsymbol{\Delta}\mathbf{X}$. This model-based estimator is derived under the assumption that the working correlation is exactly equal to the true correlation structure, which may not hold. However, the choice of the design has been shown to be relatively insensitive to the exact correlation structure (see Woods and van de Ven, 2011).

In addition, we consider nonlinear models for continuous responses of the form

$$Y(\mathbf{x}_j) = \mu(\mathbf{x}_j, \boldsymbol{\beta}) + \varepsilon_j, \quad (j = 1, \ldots, N), \tag{3}$$

with the errors $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N)$ distributed according to a multivariate normal distribution $\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{R})$, where $\mathbf{R}$ is the correlation matrix modeling the block heterogeneity. Under the assumption that the true correlation matrix $\mathbf{R}$ is known, the information matrix for the generalized least squares estimator $\widehat{\boldsymbol{\beta}}$ is of the form specified in (1) with $\mathbf{F}$ the $N \times p$ matrix containing the parameter sensitivities, i.e. $\mathbf{F}_{j,i} = \partial \mu(\mathbf{x}_j, \boldsymbol{\beta}) / \partial \beta_j$, and $\mathbf{V} = \sigma^2 \mathbf{I}$.

We find designs that maximize functionals of the matrix $\mathbf{M}$ in (1), which depends on the unknown parameters $\boldsymbol{\beta}$ and any parameters, $\boldsymbol{\alpha}$, required for the specification of $\mathbf{R}$. For a particular choice of parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, a locally D-optimal design maximizes the objective function

$$\psi(d; \boldsymbol{\beta}, \boldsymbol{\alpha}) = \log \det [\mathbf{M}(d; \boldsymbol{\beta}, \boldsymbol{\alpha})], \tag{4}$$

where $\mathbf{M}(d; \boldsymbol{\beta}, \boldsymbol{\alpha})$ is the inverse covariance matrix for d evaluated at $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$.

To overcome the dependence of the design on the value of $\boldsymbol{\beta}$, we apply a pseudo-Bayesian criterion for constructing designs that overcome the dependence on the values of the model parameters $\boldsymbol{\beta}$. The Bayesian designs maximize the objective function

$$\Psi(d; \mathcal{B}, \boldsymbol{\alpha}) = \int_{\mathcal{B}} \psi(d; \boldsymbol{\beta}, \boldsymbol{\alpha}) \, \mathrm{d}F(\boldsymbol{\beta}), \tag{5}$$

where $\mathcal{B} \subset \mathbb{R}^p$ is the space of possible parameter values and $F(\boldsymbol{\beta})$ is a proper prior distribution function for $\boldsymbol{\beta}$. Woods and van de Ven (2011) found design performance for binary data models to be robust to the values of $\boldsymbol{\alpha}$; in Section 4 we investigate this robustness for other models, and hence we do not include the correlation parameters in the pseudo-Bayesian criterion.

In addition, we also consider the performance of designs under pseudo-Bayesian versions of the D_s- and A-criteria. The objective functions for these criteria are given by, respectively,

$$\Psi^A(d; \mathcal{B}, \boldsymbol{\alpha}) = -\int_{\mathcal{B}} \mathrm{tr}\left[\mathbf{M}(d; \boldsymbol{\beta}, \boldsymbol{\alpha})^{-1}\right] \mathrm{d}F(\boldsymbol{\beta}), \tag{6}$$

and

$$\begin{aligned}
\Psi^{D_s}(d; \mathcal{B}, \boldsymbol{\alpha}) = \int_{\mathcal{B}} \log \det \Big[&\mathbf{F}_1^T \mathbf{W} \mathbf{F}_1 \\
&- \mathbf{F}_1^T \mathbf{W} \mathbf{F}_2 \left(\mathbf{F}_2^T \mathbf{W} \mathbf{F}_2\right)^{-1} \mathbf{F}_2^T \mathbf{W} \mathbf{F}_1 \Big] \mathrm{d}F(\boldsymbol{\beta}),
\end{aligned} \tag{7}$$

where $\mathbf{W} = \mathbf{V}^{-1/2} \mathbf{R}^{-1} \mathbf{V}^{-1/2}$ and $\mathbf{F} = [\mathbf{F}_1 | \mathbf{F}_2]$. For a GEE model, $\mathbf{F}_i = \Delta \mathbf{X}_i$ $(i = 1, 2)$ where $\mathbf{X}_1$ is the model matrix for the subset of parameters of interest and $\mathbf{X}_2$ is the model matrix for the nuisance parameters. For a normal-theory model, $\mathbf{F}_1$ is the matrix of parameter sensitivities for the parameters of interest, with $\mathbf{F}_2$ similarly defined for the nuisance parameters.

3. D-optimal minimum-support designs

We now restrict attention to minimum-support designs with exactly p treatments, each of which may be replicated in the experiment. For such designs, Theorem 3.1, given below, establishes a decomposition of each of objective functions (4) and (5) into two additive components, each of which can be optimized separately making use of existing results. The theorem is an extension to Bayesian designs for nonlinear models of the work of Cheng (1995) on designs for linear models.

Theorem 3.1. *Let d be a minimally supported design with distinct points denoted by $\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_p$. If the model is such that the asymptotic variance-covariance matrix has the form (1), the objective functions in (4) and (5) can be decomposed into two independent functions as follows:*

$$\psi(d; \boldsymbol{\beta}, \boldsymbol{\alpha}) = \log \det\left[\mathbf{M}_1\left(d_1; \boldsymbol{\beta}\right)\right] + \log \det\left[\mathbf{M}_2\left(d_2; \boldsymbol{\alpha}\right)\right], \tag{8}$$

and

$$\Psi(d; \mathcal{B}, \boldsymbol{\alpha}) = \int_{\mathcal{B}} \log \det\left[\mathbf{M}_1\left(d_1; \boldsymbol{\beta}\right)\right] \, \mathrm{d}F(\boldsymbol{\beta}) + \log \det\left[\mathbf{M}_2\left(d_2; \boldsymbol{\alpha}\right)\right], \tag{9}$$

where $\mathbf{M}_1\left(d_1; \boldsymbol{\beta}\right) = \mathbf{F}_p^T \mathbf{V}_p^{-1} \mathbf{F}_p$ is the information matrix for an exact design $d_1 = \{\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_p\}$ under the corresponding GLM or non-linear model with independent observations; $\mathbf{F}_p$ and $\mathbf{V}_p$ denote the $\mathbf{F}$ and $\mathbf{V}$ matrices defined in Section 2 for the minimum-support design d_1; $\mathbf{M}_2\left(d_2; \boldsymbol{\alpha}\right) = \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z}$ is the information matrix for the p-vector $\boldsymbol{\tau}$ of treatment effects under the linear model with $E(\boldsymbol{Y}) = \mathbf{Z}\boldsymbol{\tau}$ and $Var(\boldsymbol{Y}) = \sigma^2 \mathbf{R}$. Here, $\mathbf{Z}$ denotes the $n \times p$ unit-treatment incidence matrix for the block design d_2 whose (i,j)th element is 1 if the ith unit is allocated treatment $\tilde{\mathbf{x}}_j$ and 0 otherwise $(i = 1, \ldots, N; j = 1, \ldots, p)$.

Proof. As $\mathbf{M}\left(d; \boldsymbol{\beta}, \boldsymbol{\alpha}\right)$ from (1) can be written as $\mathbf{M}\left(d; \boldsymbol{\beta}, \boldsymbol{\alpha}\right) = \mathbf{F}_p^T \mathbf{V}_p^{-1/2} \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z} \mathbf{V}_p^{-1/2} \mathbf{F}_p$, the determinant of $\mathbf{M}\left(d; \boldsymbol{\beta}, \boldsymbol{\alpha}\right)$ for a minimum-support design d can be decomposed as

$$
\begin{aligned}
\det[\mathbf{M}\left(d; \boldsymbol{\beta}, \boldsymbol{\alpha}\right)] &= \det\left(\mathbf{F}_p^T \mathbf{V}_p^{-1/2} \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z} \mathbf{V}_p^{-1/2} \mathbf{F}_p\right) \\
&= \det\left(\mathbf{V}_p^{-1/2} \mathbf{F}_p \mathbf{F}_p^T \mathbf{V}_p^{-1/2} \mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z}\right) \\
&= \det\left(\mathbf{V}_p^{-1/2} \mathbf{F}_p \mathbf{F}_p^T \mathbf{V}_p^{-1/2}\right) \cdot \det\left(\mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z}\right) \\
&= \det\left(\mathbf{F}_p^T \mathbf{V}_p^{-1} \mathbf{F}_p\right) \cdot \det\left(\mathbf{Z}^T \mathbf{R}^{-1} \mathbf{Z}\right) \\
&= \det\left[\mathbf{M}_1\left(d_1; \boldsymbol{\beta}\right)\right] \cdot \det\left[\mathbf{M}_2\left(d_2; \boldsymbol{\alpha}\right)\right].
\end{aligned}
$$

Substitution of this equation in (4) and (5) gives (8) and (9), respectively.

Theorem 3.1 allows designs $d^\star$ to be constructed that are locally or Bayesian D-optimal for the class of minimum-support designs by a three-step procedure:

1. Find a locally D-optimal or Bayesian D-optimal minimum-supported design, d_1, for the corresponding GLM or non-linear model with points $\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_p$.

2. Find a block design, d_2, for comparing p treatments, labeled $t_1, \ldots, t_p$, that is D-optimal for the set of designs with b blocks of sizes $k_1, \ldots, k_b$ under the correlation structure in $\mathbf{R}$.

3. Select an allocation of the points in d_1 to the treatment labels of d_2.

Established methods and theoretical results can be applied in steps 1 and 2 to a minimum-support D-optimal design. For step 1 algorithmic approaches (Woods et al., 2006) or theoretical results (Russell et al., 2009; Yang et al., 2011; McGree and Eccleston, 2012) can be used. For step 2, depending on the number of treatments and blocks, balanced (incomplete) block designs can be used for exchangeable correlation structures (see, for example, Shah and Sinha, 1989, p.86), and the universally optimal designs of Azzalini and Giovagnoli (1987) for autoregressive structures. In other cases, a simple interchange algorithm can be used to find to D-optimal block design with p treatments in the b blocks. As all allocations in the final step give designs that are equivalent under (4) and (5), the choice of allocation may be made using a secondary selection criterion, such as A- or D_s-optimality.

4. Examples

4.1. Binary response and generalized estimating equations

Woods and van de Ven (2011) described an example of a designed experiment from the aeronautics industry aimed at investigating the occurrence of cracking in a coating applied to engine bearings. The experiment investigated three variables, x_1, x_2 and x_3, and a binary response was measured (pass/fail on a visual inspection); no surrogate continuous response was available. A probit regression model was postulated to describe the response, with linear predictor

$$g\left[\mu\left(\mathbf{x}, \boldsymbol{\beta}\right)\right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3. \tag{10}$$

We assume each $x_i \in [-1, 1]$, and so $g^{-1}(\beta_0)$ is the probability of success at the center of the design region. There was a need to take account of blocks as the runs of the experiment were performed in sessions, with potential changes in operator or resetting of equipment between sessions. We consider the situation where four runs are performed within each session (block) and assume seven sessions are available for the whole experiment. We consider a marginal (GEE) model with an exchangeable working correlation structure, with common working correlation α within blocks and independence of observations in different blocks. We assume independent uniform prior distributions for each parameter in $\boldsymbol{\beta}$ across intervals defined as mean $\pm \kappa$ (Table 1) and, assuming $\alpha = 0.2$, find Bayesian D-optimal designs with minimum support for $\kappa = 0.5, 1, 2$.

Bayesian D-optimal minimum-support designs for the corresponding non-blocked problem (probit regression) were found numerically; Table 2 gives the design for $\kappa = 0.5$, along

6

Table 1: Example 4.1: Parameter space $\mathcal{B}$ for linear predictor (10); $\kappa = 0.5, 1, 2$

Parameter	β_0	β_1	β_2	β_3	$\beta_{12}, \beta_{13}, \beta_{23}$
Interval	$-1 \pm \kappa$	$2 \pm \kappa$	$1 \pm \kappa$	$-1 \pm \kappa$	$0.5 \pm \kappa$

with a balanced incomplete block design with seven treatments in seven blocks of size four (which is independent of the choice of κ). Any allocation of the seven support points to treatments will produce a pseudo-Bayesian D-optimal design, d_m. The first allocation in Table 2 gives an A-optimal design, maximizing (6); the second allocation is a D_s-optimal design, maximizing (7), for estimating the intercept and the three "main effect" parameters in (10).

Table 2: Example 4.1: Components of the minimum-support block design d_m for $\alpha = 0.2$ and $\kappa = 0.5$: (a) Balanced incomplete block design for 7 treatments in 7 blocks of size 4 (independent of κ); (b) Bayesian D-optimal minimum-support GLM design for three variables, with mappings of design points to treatment labels to obtain a design using A-optimality or D_s-optimality for $(\beta_0, \beta_1, \beta_2, \beta_3)$ as a secondary criterion

(a)

	Treatment						
block 1	t_1	t_2	t_3	t_4			
block 2	t_1	t_2			t_5	t_6	
block 3	t_1		t_3			t_6	t_7
block 4	t_1			t_4	t_5		t_7
block 5		t_2	t_3		t_5		t_7
block 6		t_2		t_4		t_6	t_7
block 7			t_3	t_4	t_5	t_6	

(b)

Mapping for		Variables		
A	D_s	x_1	x_2	x_3
t_1	t_3	-1	-1	-1
t_2	t_4	0.53	1	1
t_3	t_1	1	-1	1
t_4	t_5	0.79	1	-1
t_5	t_7	-0.79	1	-1
t_6	t_6	-0.53	1	1
t_7	t_2	1	-1	-1

For each value of κ, we assess the D-efficiency of the minimum-support design, d_m, through a simulation study with 5000 parameter vectors randomly drawn from the prior distribution. For each parameter vector $\boldsymbol{\beta}_i$, the locally D-optimal (not necessarily minimally supported) design, $d^\star$, was found numerically, and the efficiency

$$\text{eff}(d_m) = \exp\left\{ \left[\psi(d_m; \boldsymbol{\beta}_i, \boldsymbol{\alpha}) - \psi(d^\star; \boldsymbol{\beta}_i, \boldsymbol{\alpha}) \right] / p \right\}$$

was calculated ($i = 1, \ldots, 5000$). This sample of efficiencies can then be summarised to assess and compare designs.

For $\kappa = 0.5$, Figure 1(a) presents boxplots of $\text{eff}(d_m)$ and also of $\text{eff}(d_o)$, where d_o is the unrestricted pseudo-Bayesian D-optimal block design found using the methods of Woods
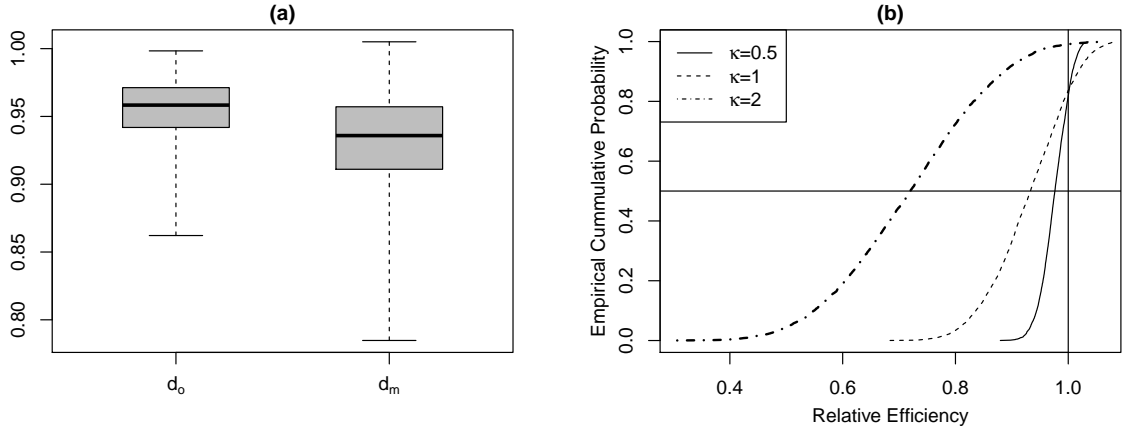
Figure 1: Example 4.1: assessment via simulation (a) Boxplots of D-efficiency for designs d_o and d_m with $\kappa = 0.5$; (b) Relative efficiency (11) of d_m compared to d_o for $\kappa = 0.5, 1, 2$.

and van de Ven (2011). The median efficiencies of both designs are high, 0.94 for d_m and 0.96 for d_o. The minimum efficiencies for d_m and d_o are greater than 0.78 and 0.86 respectively.

Designs d_m and d_o can also be directly compared for each value of κ through the relative efficiency

$$\exp\left\{[\psi(d_m; \boldsymbol{\beta}_i, \boldsymbol{\alpha}) - \psi(d_o; \boldsymbol{\beta}_i, \boldsymbol{\alpha})]/p\right\}, \tag{11}$$

for $i = 1, \ldots, 5000$, see Figure 1(b) for empirical cumulative probability plots for $\kappa = 0.5, 1, 2$ estimated from the simulated parameter vectors and corresponding designs. From both plots, it can be seen that the loss in efficiency from restricting to minimum-support designs is only moderate for concentrated prior information ($\kappa = 0.5$); the median relative efficiency of d_m compared to d_o is around 0.98, and d_m is only very rarely less than 88% as efficient as d_o. For more diffuse prior information, $\kappa = 1, 2$, the minimum-support designs perform rather less well. For $\kappa = 1$, d_m has median and minimum relative efficiencies of 0.93 and 0.68 respectively when compared to d_o; for $\kappa = 2$, these efficiencies are 0.72 and 0.31. Many, although not all, instances of high relative efficiency for d_m occur when the parameter vectors are close to the mean of the prior distribution.

Optimal designs under more diffuse prior distributions, for example, $k = 1, 2$, tend to have a larger number of support points, putting the minimum-support design at an immediate disadvantage. Hence, minimum-support designs are not recommended for situations when there is little prior knowledge about $\boldsymbol{\beta}$, i.e. when a diffuse prior distribution is assumed.

Table 3: Example 4.2: support points and allocation to 8 blocks of six 6.

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	Blocks							
1	1	1	1	1	1	1	1	1	1	1	1	2			5	6	7	
2	-1	1	1	1	1	1	1	1	1	1	1	2	3				7	
3	1	-1	1	1	1	1	1	1	1	1	1	2				6		8
4	1	1	-1	1	1	1	1	1	1	1	1		3	4		6		8
5	1	1	1	0	1	1	1	1	1	1	1			4	5		7	8
6	1	1	1	1	0	1	1	1	1	1	1		3	4	5			
7	1	1	1	1	1	0	1	1	1	1		2		4			7	8
8	1	1	1	1	1	1	0	1	1	1		2	3		5			8
9	1	1	1	1	1	1	1	0.33	1	1		2		4	5	6		
10	1	1	1	1	1	1	1	1	0.33	1			3	4		6	7	
11	1	1	1	1	1	1	1	1	1	0.33			3		5	6	7	8

4.2. Poisson response and generalized estimating equations

As a second example we consider a Poisson response possibly depending on 10 variables, $x_1,\ldots,x_{10}$, through the linear predictor

$$\log\left[\mu\left(\mathbf{x},\boldsymbol{\beta}\right)\right] = \beta_0 + \sum_{j=1}^{10} \beta_j x_j\,.$$

For illustration, we assume $\boldsymbol{\beta} = (1,1,1,1,2,2,2,2,3,3,3)$, $\mathcal{X} = [-1,1]^{10}$ and that a locally D-optimal design maximizing (4) is sought. In this case, the unblocked D-optimal design for the corresponding log-linear regression model is minimally supported, and can be obtained analytically using the results of Russell et al. (2009). The support points are given in Table 3, along with a D-optimal allocation into eight blocks of size six assuming an exchangeable working correlation structure which was found via an interchange algorithm. The same allocation to blocks was optimal for working correlation $\alpha = 0.1,\ldots,0.6$.

We can again assess the loss in efficiency from using a minimum-support design by comparing its performance to the unrestricted D-optimal design found from algorithmic search for each of $\alpha = 0.1,\ldots,0.6$. The relative efficiencies for the minimum-support design compared to the unrestricted designs are displayed in Figure 2. Note that the minimum-support design is D-optimal for $\alpha = 0$ and that for $\alpha > 0$, the unrestricted designs have, on average, 23 support points. There is a slow decrease in the efficiency of the minimum-support design as α increases, although it never drops below 0.75. For small values of α, there is little loss in efficiency from using the minimum-support design.

In addition to the support points in Table 3 forming a locally D-optimal design, McGree and Eccleston (2012) showed that such a design is also the minimum-support pseudo-Bayesian D-optimal design for any discrete prior distribution on $\boldsymbol{\beta}$ with mean vector $(1,1,1,1,2,2,2,2,3,3,3)$. Hence, from Theorem 3.1, the support points and allocation
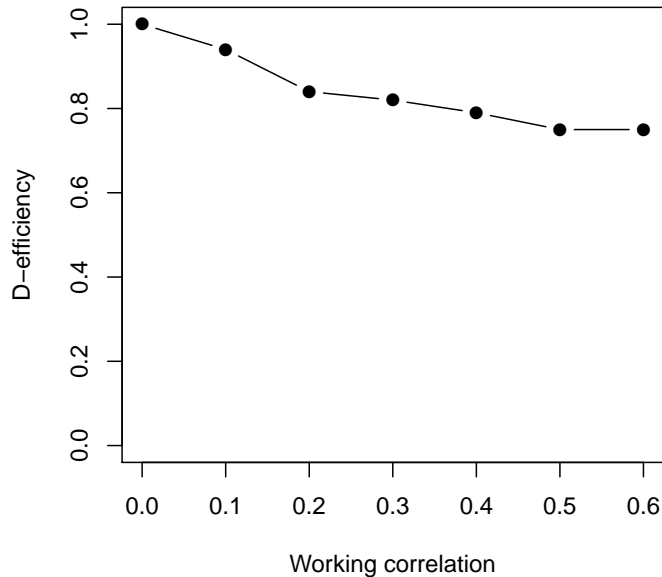
Figure 2: Example 4.2: relative efficiencies of the minimum-support D-optimal design compared to the unrestricted D-optimal design.

to blocks in Table 3 also gives a minimum-support pseudo-Bayesian D-optimal block design.

4.3. Non-linear models with correlated errors

Bates and Watts (1988, Section 3.12) described the analysis of data on the utilization of nitrite in bush beans as a function of light-intensity. The experiment involved subjecting leaves from three 16-day-old bean plants to eight different levels of light intensity; each intensity level was applied to one leaf from each plant. The Michaelis-Menten relationship was used to define a suitable class of models. The two-parameter Michaelis-Menten model is given by

$$\mu\left(x, \boldsymbol{\beta}\right) = \frac{\beta_1 x}{\beta_2 + x}, \tag{12}$$

where $x \in [0, x_{\max}]$ is the light-intensity (in $\mu\mathrm{E/m}^2\mathrm{s}$). An alternative three-parameter model is

$$\mu\left(x, \boldsymbol{\beta}\right) = \frac{\beta_1 x}{\beta_2 + x + \beta_3 x^2}. \tag{13}$$

The 24-run experiment was repeated on two days, giving 48 runs in total. Bates and

Table 4: Example 4.3: Parameter estimates and parameter values for models (12) and (13)

| | | Estimates | | Parameter values for |
	Parameter	Day 1	Day 2	locally optimal design
Two parameter model (12)	β_1	24734	22414	23500
	β_2	35.27	33.10	34
Three parameter model (13)	β_1	89846	50890	70360
	β_2	186.7	103.5	145
	β_3	0.01626	0.0078	0.012

Watts (1988) took account of differences between days in their analysis by including an additional dummy variable in the model, equaling one when the observation was made on the second day, that resulted in 2 and 3 additional parameters in models (12) and (13), respectively. This parameterization allowed the estimation of separate Michaelis-Menten models for each day; parameter estimates from these models are given in Table 4.

To illustrate our design methods, we take a different approach when finding minimum-support locally D-optimal designs for this problem and assume that the mean on both days, which are treated here as blocks, follows the same Michaelis-Menten model (with equal parameter values). Correlation between observations made on the same day is taken into account by assuming an additive and normally distributed error term with exchangeable intra-block correlation structure; see model (3).

We consider designs consisting of two blocks of 24 runs that are locally D-optimal for the parameter values given in Table 4. It is well-known that an exact locally D-optimal design for model (12) has two treatments, $x = \beta_2 / (1 + 2\beta_2 / x_{\max})$ and $x = x_{\max}$ (see, for example, López-Fidalgo and Wong, 2002). In this experiment, $x_{\max} = 175$ and hence $d_1 = \{24.49, 175\}$. The minimum-support locally D-optimal blocked design, d_{m1}, is then a balanced complete block design with these two treatments replicated 12 times on each day.

Analytical solutions for exact locally D-optimal designs have not been derived for the three-parameter model (13). Dette and Kiss (2009) gave the form of the approximate locally D-optimal design for this model; however, although the design has three support points, it is not equally-weighted. Therefore, the exact locally D-optimal design with three support points was calculated numerically using a Nelder-Mead algorithm to be $d_1 = \{23.78, 80.74, 175\}$. The minimum-support locally D-optimal blocked design, d_{m2}, for the three parameter model has eight replications of this design on each day.

The minimum-support locally D-optimal designs were compared to locally D-optimal designs that were not restricted to be minimally supported, found using a Nelder-Mead algorithm and 1000 random starting designs. The D-efficiencies of the minimum-support locally D-optimal designs are plotted in Figure 3 for models (12) and (13), and a range of intra-block correlations. For low correlation (≤ 0.2), the minimum-support designs d_{m1} and d_{m2} are highly efficient (greater than 90%). For higher correlations, this efficiency drops rapidly, when the unrestricted designs have many more than two or three support
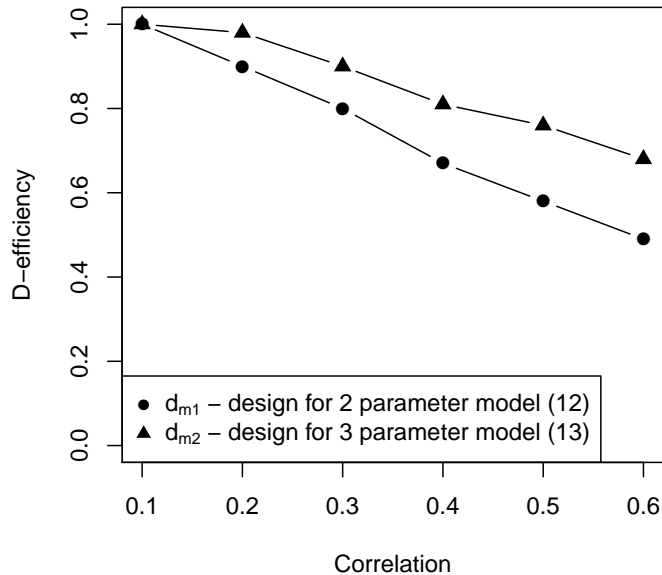
11

Figure 3: Example 5.3: *D*-efficiency of minimum-support locally *D*-optimal designs, d_{m1} and d_{m2}, for models (12) and (13) respectively

points.

5.   Concluding remarks

The theoretical results in this paper allow minimum-support *D*-optimal blocked designs to be found that can exploit the increasing number of results on analytically available completely randomised designs for non-linear models, both locally optimal and Bayesian designs for continuous and discrete responses. Even in situations where analytical results on the form of the completely randomised design are not available, decomposing the optimal design problem into finding a treatment design and its allocation to blocks separately can provide a substantial increase in computational efficiency compared to an unrestricted design search.

The examples in this paper demonstrate the efficiency that can be achieved using minimum-support designs, particularly for Bayesian designs with informative prior distributions on the model parameters and when there are low intra-block dependencies. For other cases, we would recommend the more general design methods proposed by Woods and van de Ven (2011).

References

Atkinson, A.C., 2008. Examples of the use of an equivalence theorem in constructing optimum experimental designs for random-effects nonlinear regression models. Journal of Statistical Planning and Inference 138, 2595–2606.

Atkinson, A.C., Ucinski, D., 2004. Experimental design for time-dependent models with correlated observations. Studies in Nonlinear Dynamics and Econometrics 8, Article 13.

Azzalini, A., Giovagnoli, A., 1987. Some optimal designs for repeated measurements with autoregressive errors. Biometrika 74, 725–734.

Bates, D.M., Watts, D.G., 1988. Nonlinear Regression Analysis and its Applications. Wiley, New York.

Cheng, C.S., 1995. Optimal regression designs under random block-effects models. Statistica Sinica 5, 485–497.

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

Dette, H., Kiss, C., 2009. Optimal designs for inverse regression models. Statistica Sinica 19, 1567–1586.

Dror, H.A., Steinberg, D.M., 2008. Sequential experimental designs for generalized linear models. Journal of the American Statistical Association 103, 288–298.

Goos, P., Vandebroek, M., 2001. D-optimal response surface designs in the presence of random block effects. Computational Statistics and Data Analysis 37, 433–453.

Han, C., Chaloner, K., 2004. Bayesian experimental design for nonlinear mixed-effects models with application to HIV dynamics. Biometrics 60, 25–33.

Hughes-Oliver, J.M., 1998. Optimal designs for nonlinear models with correlated errors, in: New Developments and Applications in Experimental Design. Institute of Mathematical Sciences. volume 34 of *IMS Lecture Notes - Monograph Series*, pp. 163–174.

Lee, Y., Nelder, J.A., Pawitan, Y., 2006. Generalized Linear Models with Random Effects. Chapman and Hall, Boca Raton.

Liang, K.Y., Zeger, S.L., 1986. Longitudinal data analysis using generalized linear models. Biometrika 73, 13–22.

López-Fidalgo, J., Wong, W.K., 2002. Design issues for the Michaelis-Menton model. Journal of Theoretical Biology 215, 1–11.

McGree, J.M., Eccleston, J.A., 2012. Robust designs for Poisson regression models. Technometrics 54, 64–72.

Mentré, F., Mallet, A., Baccar, D., 1997. Optimal design in random-effects regression models. Biometrika 84, 429–442.

Niaparast, M., 2009. On optimal design for a Poisson regression model with random intercept. Statistics and Probability Letters 79, 741–747.

Russell, K.G., Woods, D.C., Lewis, S.M., Eccleston, J.A., 2009. D-optimal designs for Poisson regression models. Statistica Sinica 19, 721–730.

Shah, K.R., Sinha, B.K., 1989. Theory of Optimal Designs. Springer, New York.

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, 284–292.

Woods, D.C., van de Ven, P., 2011. Blocked designs for experiments with correlated non-normal response. Technometrics 53, 173–182.

Yang, M., Zhang, B., Huang, S., 2011. Optimal designs for generalized linear models with multiple design variables. Statistica Sinica 21, 1415–1430.