

Working Paper M09/21

Methodology

Block Designs For Experiments

With Non-Normal Response

D.C. Woods, P. Van De Ven

Abstract

Many experiments measure a response that cannot be adequately described by a linear model with normally distributed errors and are often run in blocks of homogeneous experimental units. We develop the first methods of obtaining efficient block designs for experiments with an exponential family response described by a marginal model fitted via Generalized Estimating Equations. This methodology is appropriate when the blocking factor is a nuisance variable as, for example, occurs in industrial experiments. A D- optimality criterion is developed for finding designs robust to the values of the marginal model parameters and applied using three strategies: unrestricted algorithmic search, use of minimum- support designs, and blocking of an optimal design for the corresponding Generalized Linear Model. Designs obtained from each strategy are critically compared and shown to be much more efficient than designs that ignore the blocking structure. The designs are compared for a range of values of the intra- block working correlation and for exchangeable, autoregressive and nearest neighbor structures. An analysis strategy is developed for a binomial response that allows estimation from experiments with sparse data, and its effectiveness demonstrated. The design strategies are motivated and demonstrated through the planning of an experiment from the aeronautics industry.

Block Designs for Experiments with Non-Normal Response

D.C. WOODS

Southampton Statistical Sciences
Research Institute
University of Southampton
Southampton SO17 1BJ, UK
D.Woods@soton.ac.uk

P. VAN DE VEN

Department of Epidemiology and
Biostatistics
VU University Medical Centre
1007 MB Amsterdam, Netherlands
p.vandeven@vumc.nl

Many experiments measure a response that cannot be adequately described by a linear model with normally distributed errors and are often run in blocks of homogeneous experimental units. We develop the first methods of obtaining efficient block designs for experiments with an exponential family response described by a marginal model fitted via Generalized Estimating Equations. This methodology is appropriate when the blocking factor is a nuisance variable as, for example, occurs in industrial experiments. A D -optimality criterion is developed for finding designs robust to the values of the marginal model parameters and applied using three strategies: unrestricted algorithmic search, use of minimum-support designs, and blocking of an optimal design for the corresponding Generalized Linear Model. Designs obtained from each strategy are critically compared and shown to be much more efficient than designs that ignore the blocking structure. The designs are compared for a range of values of the intra-block working correlation and for exchangeable, autoregressive and nearest neighbor structures. An analysis strategy is developed for a binomial response that allows estimation from experiments with sparse data, and its effectiveness demonstrated. The design strategies are motivated and demonstrated through the planning of an experiment from the aeronautics industry.

KEY WORDS: Binary response; Block designs; D -optimality; Generalized Estimating Equations; Generalized Linear Model; Robust design; separation.

1. INTRODUCTION

Many experiments in science and engineering aim to understand a process by modeling a response variable using a generalized linear model (GLM) with several explanatory variables. Methods recently developed for constructing efficient and optimal designs for such experiments using homogeneous experimental units include Woods, Lewis, Eccleston, and Russell (2006), Dror and Steinberg (2006, 2008), and Russell, Woods, Lewis, and Eccleston (2009); see Khuri, Mukherjee, Sinha, and Ghosh (2006) for a review of earlier work. Woods et al. (2006) demonstrated the potential poor performance of standard linear model designs for GLMs. In this paper, we develop methods for constructing designs where the heterogeneous units are grouped into homogeneous sets or blocks to increase the accuracy of inferences made from the experimental data. We focus on the exponential family of distributions and models for the marginal response which, in the data analysis, are fitted by the method of generalized estimating equations (GEEs; Liang and Zeger, 1986). GEE models extend GLMs by allowing dependence between observations from units in the same block, characterized through an intra-block working correlation structure, and have been widely applied in the social and medical sciences. They are particularly appropriate when block effects are not of interest in themselves. See Robinson, Myers, and Montgomery (2004) for the analysis of an industrial split-plot experiment, and a comparison of the use of marginal GEE models with the alternative conditional paradigm of Generalized Linear Mixed Models (GLMMs; Breslow and Clayton, 1993).

In common with GLMs and other nonlinear models, designing an experiment for GEE models is complicated by the dependence of a design's performance on the values of the model

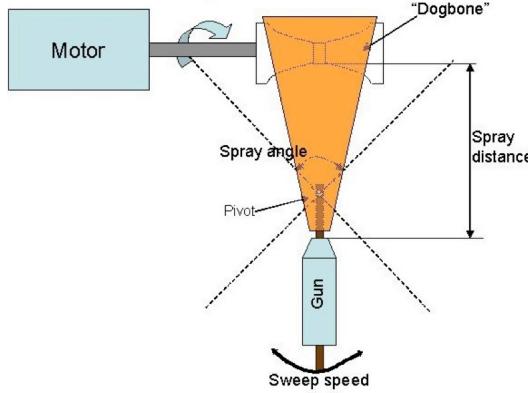


Figure 1: Schematic of the bearing spraying process

parameters. Local optimal designs for given values of the model parameters are effective when strong prior information is available from pilot experiments or historical data or, where practicable, as part of a sequential experiment strategy. Various approaches have been proposed to obtain robust designs that overcome the parameter dependence for GLMs, including maximin and Bayesian methods. The most common design selection criterion used is D -optimality which minimizes the determinant of the (asymptotic) variance-covariance matrix of the model parameters, and is appropriate when the aim of the experiment is accurate parameter estimation. We use a variant of this criterion for design selection and assess design performance through D -efficiency.

Our approach is motivated through an example from the aeronautics industry on investigating the occurrence of cracking in a coating applied to engine bearings. As part of the manufacturing process, a powder-form, nickel-based, bond coating is thermo-sprayed onto machined liners. In this spraying process, shown in Figure 1, two bearings are arranged in a “dogbone” and coated simultaneously. After manufacturing, cracking of the coating on each bearing is assessed by visual inspection with a binary outcome (pass/fail); failed bearings are rejected.

Variables thought likely to affect the cracking of the coating are the distance between the spray gun and the dogbone (x_1), the angle through which the spray gun sweeps (x_2), and the sweep speed (x_3). The smallest unit to which a combination of variable values (individual design point or treatment) can be applied is a pair of bearings and the response variable is the number of bearings in a pair that pass inspection. Typically, time and resource constraints allow only small experiments to be run. The runs are carried out in sessions (mornings or afternoons) that are regarded as blocks in the experiment, for example, four sessions of four runs. Through taking account of any systematic differences between runs in the different sessions, the efficiency of the design and accuracy of the findings from the subsequent analysis may be improved. As the sessions have no intrinsic engineering meaning, estimation of the between session variation is not of interest. Previous use of a surrogate continuous response, such as mechanical strength of the coating, has proved inadequate in explaining the cracking mechanism. There is therefore a need to develop methods of finding efficient block designs for non-normal data including binary and binomial observations.

This problem has received little attention in the literature, with results restricted to conditional models with (i) binary data, a logit link, and dichotomous explanatory variables (Moerbeek, Van Breukelen, and Berger, 2001; Moerbeek and Maas, 2005; Ouwens, Tan, and Berger, 2006); (ii) longitudinal binary data with a single variable (Tekle, Tan, and Berger, 2008); and (iii) Poisson data and a single variable (Niaparast, 2009). The current paper is an advance on previous work as it investigates general methods for finding optimal and efficient designs

for multi-variable experiments with continuous or categorical explanatory variables and for any form of within-block working correlation structure.

Section 2 describes the GEE model and defines the criterion for selecting a robust design used in this paper. Three strategies are proposed for finding efficient designs in Section 3. The strategies are evaluated and compared through their application to the bearing coating example in Section 4 for exchangeable, autoregressive and nearest neighbor working correlation structures. The robustness of the designs to (a) the values of the model parameters and (b) the choice of correlation structure is investigated through simulation studies. Finally, an adaptation of the GEE method is developed in Section 5 that provides valid inference for small experiments, such as the bearing example, thus overcoming the problem of “separation” (Albert and Anderson, 1984).

2. MODEL AND DESIGN CRITERION

Suppose the response $Y(\mathbf{x})$ may depend on m explanatory variables $\mathbf{x}^T = (x_1, \dots, x_m)$ and the experiment has N units, where the j th unit receives treatment $\mathbf{x}_j^T = (x_{1j}, \dots, x_{mj})$. The \mathbf{x}_j are chosen from a standardized design space $\mathcal{X} = [-1, 1]^m$ and are not necessarily distinct. Further, suppose that the units are arranged in b blocks of size k_l ($l = 1, \dots, b$), so that $k_1 + k_2 + \dots + k_b = N$. We order the entries in $\mathbf{Y} = (Y(\mathbf{x}_1), \dots, Y(\mathbf{x}_N))^T$ by block and by unit within block.

2.1. Model

We assume the marginal mean and variance of each observation is taken from an appropriate GLM. That is, $E[Y(\mathbf{x}_j)] = \mu(\mathbf{x}_j)$ and $\text{Var}[Y(\mathbf{x}_j)] = v[\mu(\mathbf{x}_j)]/\phi$, where ϕ is a constant scale parameter and $v(\cdot)$ is the variance function for the exponential family corresponding to the GLM ($j = 1, \dots, N$). The mean response is related to \mathbf{x} through $g[\mu(\mathbf{x})] = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta}$, where $g(\cdot)$ is the link function and $\mathbf{f}^T(\mathbf{x})\boldsymbol{\beta}$ is the linear predictor. The respective p -vectors $\mathbf{f}(\mathbf{x})$ and $\boldsymbol{\beta}$ hold known functions of \mathbf{x} , such as main effects and interactions, and unknown model parameters.

Liang and Zeger (1986) suggested the GEE approach for estimating $\boldsymbol{\beta}$ from dependent data. Heterogeneity of the blocks is modeled via a “working correlation” structure for the observations within each block, together with the assumption of independence of observations made on units in different blocks. The generalized estimating equations are an extension of the score functions for a GLM with independent data, and are given by

$$X^T \Delta V^{-1} (\mathbf{Y} - \boldsymbol{\mu}) = \mathbf{0}_p, \quad (1)$$

where X is the $N \times p$ model matrix with rows $\mathbf{f}^T(\mathbf{x}_j)$, $\Delta = \text{diag}\{1/g'[\mu(\mathbf{x}_j)]\}$ and $\boldsymbol{\mu} = (\mu(\mathbf{x}_1), \dots, \mu(\mathbf{x}_N))^T$. The matrix $V = A^{1/2}R(\alpha)A^{1/2}$, where $A = \text{diag}\{\text{Var}[Y(\mathbf{x}_j)]\}$, is an adaptation of the usual variance matrix for a GLM, with $R(\alpha)$ an $N \times N$ working correlation matrix that is assumed known up to the value of α and represents the assumed dependence structure in the data. For block designs, structures appropriate for a wide range of applications include exchangeable, autoregressive and nearest neighbor, each of which has intra-block correlations parameterized by a single parameter $\alpha \geq 0$. Arbitrary dependence structures with more unknown parameters could also be considered.

For a discrete response, the values of the actual pairwise correlations, in contrast to the working correlation, may depend on \mathbf{x} . We therefore follow Chaganty and Joe (2004) in interpreting $R(\alpha)$, which is independent of \mathbf{x} , as a matrix that adjusts the weight given to each observation in the estimation of $\boldsymbol{\beta}$, and choose the value of α according to their guidelines. For example, under exchangeable and autoregressive (order 1) structures, a value of α between .2 and .3 is appropriate for moderately dependent data.

For the bearing experiment of Section 1, the binomial data may result from a thresholding process. Hence a plausible model for the marginal response is probit regression with $g[\mu(\mathbf{x})] = \Phi^{-1}[\mu(\mathbf{x})]$, where $\Phi(\cdot)$ is the standard Normal distribution function. The marginal distribution for the observation from each unit is then binomial with $E[Y(\mathbf{x}_j)] = 2\Phi[\mathbf{f}^T(\mathbf{x}_j)\boldsymbol{\beta}]$ and $\text{Var}[Y(\mathbf{x}_j)] = \Phi[\mathbf{f}^T(\mathbf{x}_j)\boldsymbol{\beta}]\{1 - \Phi[\mathbf{f}^T(\mathbf{x}_j)\boldsymbol{\beta}]\}/\phi$. If the $N = 16$ units in the experiment are divided into $b = 4$ blocks of $k = 4$ units, then

$$R(\alpha) = I_b \otimes [(1 - \alpha)I_k + R_B(\alpha)] , \quad (2)$$

where \otimes denotes Kronecker product and I_b is the $b \times b$ identity matrix. A reasonable assumption for the working correlation is an exchangeable structure with $R_B(\alpha) = \alpha \mathbf{1}_k \mathbf{1}_k^T$, and $\mathbf{1}_k$ a k -vector with every entry 1.

The asymptotic variance-covariance matrix for the GEE estimator $\hat{\boldsymbol{\beta}}$ is given by

$$\text{Var}(\hat{\boldsymbol{\beta}}) = (X^T \Delta V^{-1} \Delta X)^{-1} X^T \Delta V^{-1} \Sigma V^{-1} \Delta X (X^T \Delta V^{-1} \Delta X)^{-1} , \quad (3)$$

where Σ is the true, unknown variance-covariance matrix for \mathbf{Y} ; see Liang and Zeger (1986). It follows that $\text{Var}(\hat{\boldsymbol{\beta}})$ depends on $\boldsymbol{\beta}$ through Δ and V , and on α through V .

2.2. Design Criterion

We adopt a pseudo-Bayesian approach to constructing designs for GEE models robust to the values of model parameters, as first suggested for single variable logistic regression by Chaloner and Larntz (1989) and used for multi-variable GLMs by Woods et al. (2006) and Dror and Steinberg (2006). For GEE models, a design d is found that maximizes the objective function

$$\Psi(d; \mathcal{B}, \alpha) = \int_{\mathcal{B}} \log \psi^D(d; \boldsymbol{\beta}, \alpha) dF(\boldsymbol{\beta}) , \quad (4)$$

where $\mathcal{B} \subset \mathbb{R}^p$ is the space of possible parameter values, $F(\boldsymbol{\beta})$ is a proper prior distribution function for $\boldsymbol{\beta}$, and $\psi^D(d; \boldsymbol{\beta}, \alpha)$ is the local D -optimality objective function for $\boldsymbol{\beta}$ and α . For the purpose of selecting a design, we make the assumption that the true correlation matrix equals the working correlation matrix $R(\alpha)$, so that

$$\Sigma = A^{1/2} R(\alpha) A^{1/2} . \quad (5)$$

Often this assumption will not hold. However, it can be demonstrated that departures from (5) have little impact on the choice of design (see Section 4.3) and hence we take

$$\psi^D(d; \boldsymbol{\beta}, \alpha) = \det [M(d, \boldsymbol{\beta}, \alpha)] , \quad (6)$$

where $M(d, \boldsymbol{\beta}, \alpha) = X^T \Delta V^{-1} \Delta X$ is the inverse of the asymptotic variance-covariance matrix for $\hat{\boldsymbol{\beta}}$ and X , Δ and V all depend on design d . In the following section we describe strategies for constructing optimal and near-optimal designs robust to the value of $\boldsymbol{\beta}$. These strategies are applied to the bearing example in Section 4, where the robustness of the designs to α is also addressed.

3. DESIGN CONSTRUCTION STRATEGIES

The main emphasis in this paper is on finding exact designs robust to the values of the model parameters, that is, designs having a pre-specified number, N , of runs. Continuous, or approximate, designs, defined as measures across the design region, can also be found and a general equivalence theorem established. The theorem is given in Appendix A together with an example of its application. An exact block design d may be specified as (i) design points

$\mathbf{x}_1, \dots, \mathbf{x}_N$, and (ii) an ordering of these points that defines their allocation to the units in each block.

Three strategies are proposed and investigated for finding exact robust D -optimal block designs for GEE models.

Strategy 1: Direct optimization of (4) via algorithmic search. We use a simulated annealing algorithm (Haines, 1987) to find designs for a given number of blocks of fixed, but not necessarily equal, sizes. The algorithm performs a continuous optimization through adjustment of each design point by a random perturbation. If the adjustment gives an increased value of (4), then it is accepted. Otherwise, the adjustment is accepted with non-zero probability, obtained from the Boltzmann-Gibbs distribution. This probability decreases with decreasing value of (4), and also with decrease in a “temperature” parameter. The temperature is reduced (cooled) as the search progresses, to end with a greedy algorithm. Our experience is that the algorithm finds efficient designs in reasonable time for an initial acceptance probability of between .5 and .9 and geometrical cooling.

To search for an exact design with a large number of large blocks can be a prohibitive computational burden for the regular use of these designs. We therefore propose two further methods of finding designs that require algorithmic search only for sub-problems of reduced complexity.

Strategy 2: Restriction to minimum-support designs. This strategy restricts the search to designs for which the number of distinct points is equal to the number, p , of unknown parameters in β . Minimal-support designs allow the use of known results on optimal designs for both GLMs and linear model block designs for the comparison of treatments. We extend the work of Cheng (1995) on designs for linear models to robust designs for nonlinear GEE models through similarly establishing a decomposition of the optimization problem into two smaller problems, see Appendix B. This allows a design d_m^* to be constructed that is robust D -optimal for the class of minimum-support designs by a three-step procedure:

1. find a robust D -optimal saturated design, d_1^* , for the corresponding GLM with points $\mathbf{x}_1^*, \dots, \mathbf{x}_p^*$
2. find a design, d_2^* , for comparing p treatments, labeled t_1, \dots, t_p , that is D -optimal for the set of designs with b blocks of sizes k_1, \dots, k_b under the working correlation structure
3. select an allocation of the points in d_1^* to the treatment labels of d_2^* .

As all allocations in the final step give designs that are equivalent under (4), the choice of allocation may be made using a secondary selection criterion, such as A - or D_s -optimality (Section 4).

Established algorithmic methods can be applied in steps 1 and 2 to obtain a minimal support design. Strategy 2 may also allow an appropriate design to be derived analytically through the use of known optimal designs at each step. For example, for first order Poisson regression with log link, the minimum-support designs of Russell et al. (2009) have been shown to meet the step 1 requirements by McGree, Eccleston, and Russell (2009). For step 2, balanced incomplete block designs can be used for exchangeable correlation structures, and the universally optimal designs of Azzalini and Giovagnoli (1987) for autoregressive structures.

Strategy 3: Allocation to blocks of runs from an optimal unblocked design. We now restrict the search procedure to finding a block design whose design points are those of a robust D -optimal design for the corresponding GLM:

1. find a robust D -optimal design for the GLM with design points $\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_N$

2. find an allocation of $\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_N$ to b blocks with sizes k_1, \dots, k_b that maximizes (6) for a given value of $\boldsymbol{\beta}$.

Both steps can be achieved using available algorithms: for step 1, the methods of Woods et al. (2006) or Dror and Steinberg (2006), and for step 2, a standard interchange algorithm. A natural choice of $\boldsymbol{\beta}$ in step 2 is a measure of location of the prior distribution $F(\boldsymbol{\beta})$, and expectation is used in this paper. This strategy has the advantage of not restricting to minimum-supported designs and the disadvantage that computation must generally be used for both steps.

4. EVALUATION OF THE DESIGN STRATEGIES

We investigate the performance of robust D -optimal exact designs found by each of the three strategies of Section 3 through examining designs for the bearing experiment in four blocks of size four and a probit regression model with linear predictor

$$g[\mu(\mathbf{x})] = \beta_0 + \sum_{i=1}^3 \beta_i x_i + \sum_{i=1}^3 \sum_{j>i}^3 \beta_{ij} x_i x_j. \quad (7)$$

We use a uniform prior distribution for $\boldsymbol{\beta}$ across the parameter space \mathcal{B} given by the product of the intervals in Table 1. The calculation of (4) involves a seven-dimensional integral that can be approximated by the average across a 21-point discrete Latin Hypercube sample; a recent alternative was given by Gotwalt, Jones, and Steinberg (2009). Throughout this section, designs are evaluated through their efficiencies under a given criterion, relative to an appropriate reference design. The D -efficiency of a design d relative to an optimal design d^* is

$$[\psi^D(d; \boldsymbol{\beta}, \alpha) / \psi^D(d^*; \boldsymbol{\beta}, \alpha)]^{1/p}. \quad (8)$$

4.1. Exchangeable working correlation structures

Figures 2 and 3 show the designs, d_o and d_a , found by maximizing (4) for an exchangeable working correlation structure with $\alpha = .2$ using Strategies 1 and 3 respectively. Each design has 16 distinct points including two points in the interior of the design region; the two interior points are different for the two designs.

The three components required for Strategy 2 are given in Table 2: a balanced incomplete block design for 7 treatments, the points of a minimum-support robust D -optimal GLM design and two allocations of design points to treatment labels. The first allocation produces an optimal design, d_{m_1} , under a robust version of A -optimality; the second, d_{m_2} , is optimal under a robust D_s -criterion for estimating the three interaction parameters in (7). The objective functions for these criteria are given by

$$\Psi^A(d; \boldsymbol{\beta}, \alpha) = \int_{\mathcal{B}} \text{tr} [M(d, \boldsymbol{\beta}, \alpha)^{-1}] \, dF(\boldsymbol{\beta}),$$

and

Table 1: Parameter space \mathcal{B} for linear predictor (7)

Parameter	β_0	β_1	β_2	β_3	$\beta_{12}, \beta_{13}, \beta_{23}$
Range	[-2,0]	[1,3]	[0,2]	[-2,0]	[-0.5,1.5]

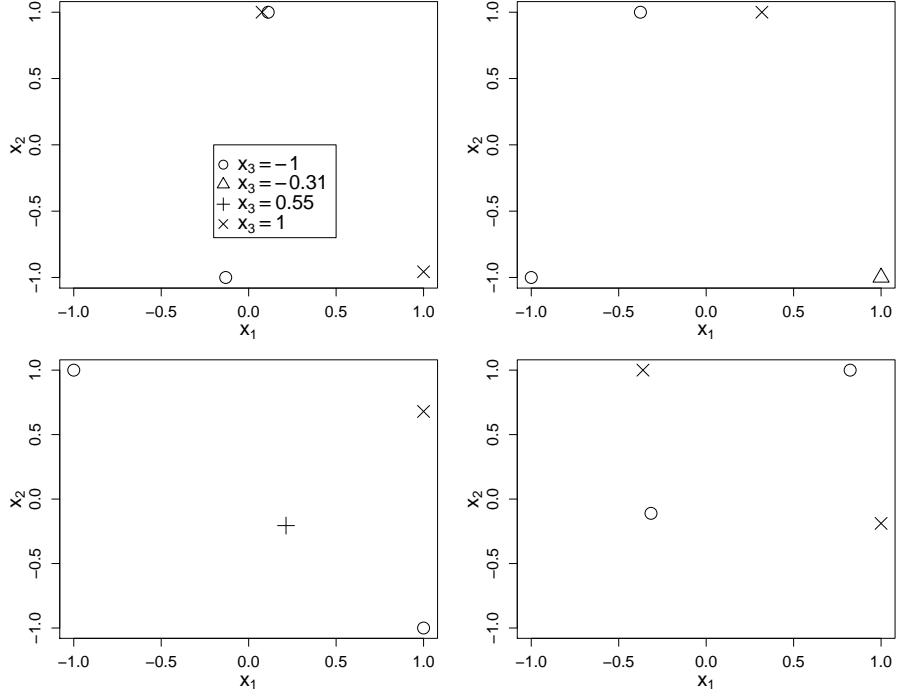


Figure 2: Design points in each block of design d_o for $\alpha = .2$ found by algorithmic search.

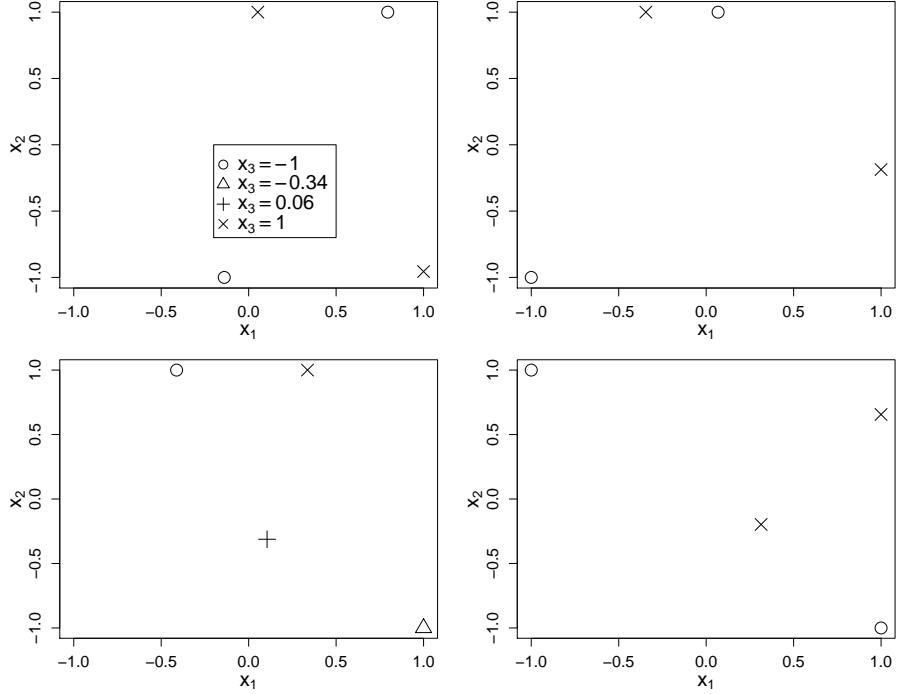


Figure 3: Design points in each block of design d_a for $\alpha = .2$ found by D -optimal allocation of points from a GLM design to blocks.

$$\begin{aligned} \Psi^{D_s}(d; \boldsymbol{\beta}, \alpha) = & \int_{\mathcal{B}} \log \det [X_s^T \Delta V^{-1} \Delta X_s \\ & - X_s^T \Delta V^{-1} \Delta X_n (X_n^T \Delta V^{-1} \Delta X_n)^{-1} X_n^T \Delta V^{-1} \Delta X_s] dF(\boldsymbol{\beta}), \end{aligned}$$

where $X = [X_s | X_n]$, X_s is the model matrix for the subset of parameters of interest and X_n is the model matrix for the nuisance parameters. The integrals were approximated by averaging the integrand across the parameter values from the sample used to evaluate (4). For

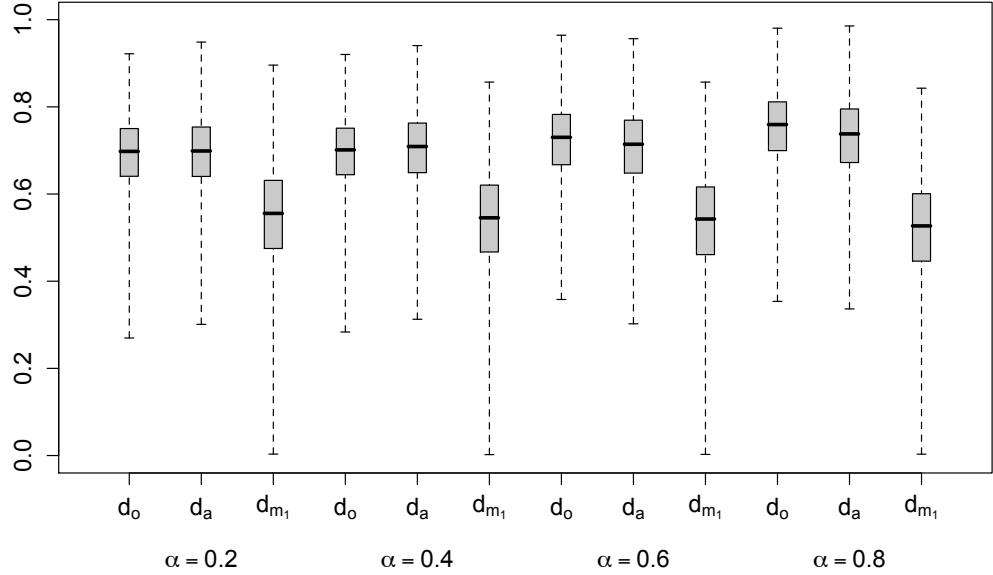


Figure 4: Boxplots of D -efficiencies of robust block designs, d_o , d_a , d_{m_1} , for 4 values of α .

in this small example, the designs were obtained by evaluating all possible 5040 distinct block designs; for larger examples, an interchange algorithm may be used. The maximum differences in efficiency between the best and worst of these designs are 8% and 12% for A - and D_s -optimality, respectively. The two optimal designs have no blocks in common.

To investigate the performance of these designs across the parameter space \mathcal{B} , the designs found for $\alpha = .2, .4, .6, .8$ were assessed via a simulation study in which a random sample of 10,000 parameter vectors was drawn from \mathcal{B} , as in all simulation studies in this paper. A local (near) D -optimal design for the GEE model was found for each vector using algorithmic search and the D -efficiencies (8) of d_o , d_a and d_{m_1} were calculated. (d_{m_2} has the same D -efficiency as d_{m_1}). Figure 4 shows boxplots of the D -efficiency distributions. Designs d_o and d_a perform similarly, and are fairly robust for all four values of α , with median efficiencies between .70 and .76, and lower quartiles between 0.64 and 0.70. For larger values of α , where the advantage of blocking is greatest, d_o offers a small but consistent improvement over d_a ; for the smaller α

Table 2: Components of the two minimum-support block designs d_{m_1} and d_{m_2} for $\alpha = .2$: (a) Balanced incomplete block design for 7 treatments in blocks of size 4; (b) Robust D -optimal minimum-support GLM design for 3 variables, with mappings of design points to treatment labels to obtain a design using A - or D_s -optimality as a secondary criterion

(a)				(b)				
Treatment				Mapping for		Variables		
block 1	t_1	t_2	t_3	A	D_s	x_1	x_2	x_3
block 2	t_1	t_2	t_5	t_6		.43	.64	-.90
block 3	t_1	t_4	t_5	t_7		.35	-.44	.53
block 4	t_2	t_3	t_6	t_7		1	-1	-.74
				t_4	t_5	-.49	-1	-1
				t_5	t_7	-.61	1	-1
				t_6	t_2	1	-.05	1
				t_7	t_4	.06	1	1

Table 3: D -efficiencies (minimum, median) for 5 values of α for designs found for $\alpha = .2$: D -optimal design from algorithmic search (d_o); allocation of a D -optimal GLM design to blocks (d_a); minimum-support designs (d_{m_1}, d_{m_2})

Design	α				
	0	.2	.4	.6	.8
d_o	(.27, .67)	(.31, .70)	(.30, .71)	(.34, .73)	(.36, .75)
d_a	(.30, .68)	(.30, .70)	(.28, .70)	(.30, .72)	(.30, .74)
d_{m_1}, d_{m_2}	(.00, .56)	(.00, .56)	(.00, .55)	(.00, .54)	(.00, .53)

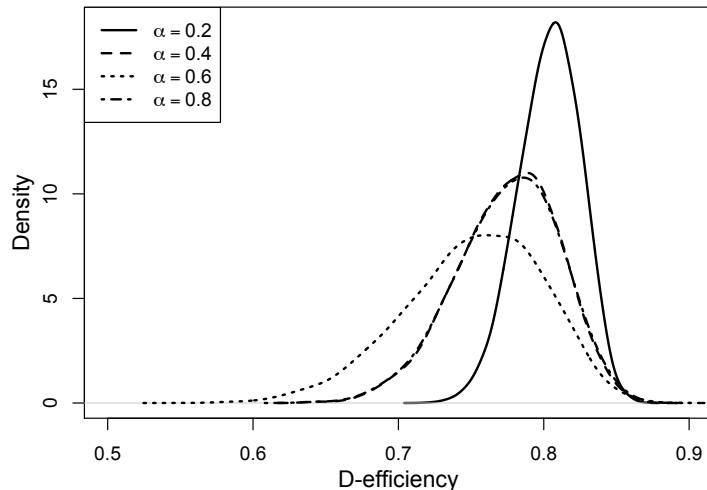


Figure 5: Smoothed density estimates of D -efficiencies of 10,000 randomly selected allocations of a 16-point robust D -optimal GLM design to 4 blocks of 4 points, for 4 values of α and β as for the aeronautical example; curves for $\alpha = 0.4$ and $\alpha = 0.6$ almost coincide.

values, this relationship is reversed.

The minimum-support design, d_{m_1} , has the worst performance of the three designs, with median efficiencies of .53–.56, and lower quartiles of .45–.48. For stronger prior information on β , other studies (not presented) show that the lack of distinct points is less of a problem. In the extreme when $F(\beta)$ is the point prior at the centroid of \mathcal{B} , local optimal designs from all three strategies perform similarly.

Robustness to the value of α of designs d_o , d_a and d_{m_1} found for $\alpha = .2$ was assessed by simulation from \mathcal{B} with $\alpha = 0, .2, .4, .6, .8$. Table 3 indicates that, for $\alpha \neq 0.2$, the D -efficiency of the designs from all three strategies is at least maintained relative to $\alpha = .2$; for d_o and d_a , efficiency actually improves for $\alpha = .6$ and $.8$ in agreement with results from the previous simulation (Figure 4). This is evidence that, for designs with stronger intra-block dependency (larger α), the values of β have less impact on design performance.

The strong performance of d_a across all values of α raises the question of whether *any* allocation of the points to blocks will suffice in Strategy 3. This issue was investigated by an assessment of 10,000 allocations chosen at random from the 2.63 million possible allocations for each of $\alpha = .2, \dots, .8$. Figure 5 shows the D -efficiency of these allocations for a point-mass prior on $\beta = (-1, 2, 1, -1, 0.5, 0.5, 0.5)$. An arbitrary allocation can be inefficient, particularly for large α . For example, when $\alpha \geq .4$, the median efficiency is more than 10% lower than the highest efficiency, and for $\alpha = .8$ (.6), the minimum efficiency is less than .5 (.6). This demonstrates the importance of careful choice of blocking scheme.

Table 4: D -efficiencies (minimum, median) for robust D -optimal designs for an autoregressive (nearest neighbor) working correlation structure under exchangeable and nearest neighbor (autoregressive) structures. Note that the NN correlation matrix is not positive definite for $\alpha = 0.8$

α	Autoregressive		Nearest neighbor	
	Exchangeable	Nearest neighbor	Exchangeable	Autoregressive
.2	(.25, .68)	(.27, .70)	(.25, .67)	(.29, .69)
.4	(.26, .68)	(.33, .68)	(.31, .64)	(.35, .69)
.6	(.30, .70)	(.26, .64)	(.31, .62)	(.37, .70)
.8	(.36, .72)	—	—	—

4.2. Autoregressive and nearest neighbor working correlation structures

GEE designs are investigated for experiments that produce dependent observations in which unit labels are ordered in one dimension, for example, by time in a longitudinal study, or by transact of land in an agricultural experiment. In contrast to the exchangeable correlation structure, units are no longer exchangeable within block, and hence the allocation of treatments to units within each block affects design efficiency. The models considered have working correlation matrices for an autoregressive order 1 (AR) and a nearest neighbor (NN) correlation structure. The strategies of Section 3 and methods of Section 4.1 may be used to find and compare designs. For brevity, we present results for Strategy 1 for the bearing experiment, linear predictor (7), and a uniform prior distribution across the parameter space of Table 1. The working correlation matrix is given by (2) with $R_B(\alpha)$ having diagonal entries α , with off-diagonal entries $\alpha^{|i-j|}$ for AR, and α for $|i-j| = 1$ (and 0 otherwise) for NN ($i, j = 1, \dots, 4; i \neq j$).

The robustness of an optimal design for each working correlation structure is evaluated (i) under the other working correlation structure, and (ii) under an exchangeable working correlation structure, using the approach of Section 4.1. From Table 4, the AR design performs well under the exchangeable structure, and also well under the NN structure provided α is small. Although the NN design performs well under the AR structure, it is inefficient under the exchangeable structure particularly for larger α values. These results are intuitive since, for small α , AR and NN working correlation structures are quite similar, while for larger α the exchangeable working correlation structure is closer to the AR than the NN structure.

4.3. Robustness of design selection to the assumed correlation matrix

We now assess the impact on the choice of design of departures from the assumption that the true correlation matrix is equal to the working correlation matrix, $R(\alpha)$, and, in particular, that it may depend on the treatments applied. To achieve this for an exchangeable working correlation structure we make an alternative assumption that the true correlation matrix is R^* in which observations in different blocks are uncorrelated and the intra-block correlation between responses $Y(\mathbf{x}_j)$ and $Y(\mathbf{x}_k)$ does not depend on the block in which they are observed. This correlation is given by $\rho_{jk} = \min(r_{jk}, \gamma) \geq 0$, where $0 < \gamma < 1$ and

$$r_{jk} = \min_{j \neq k} \left\{ \sqrt{\frac{\mu(\mathbf{x}_j)[1 - \mu(\mathbf{x}_k)]}{[1 - \mu(\mathbf{x}_j)]\mu(\mathbf{x}_k)}}, \sqrt{\frac{[1 - \mu(\mathbf{x}_j)]\mu(\mathbf{x}_k)}{\mu(\mathbf{x}_j)[1 - \mu(\mathbf{x}_k)]}} \right\}, \quad (9)$$

is the maximum achievable correlation between $Y(\mathbf{x}_j)$ and $Y(\mathbf{x}_k)$ (Joe, 1997, p. 210). We con-

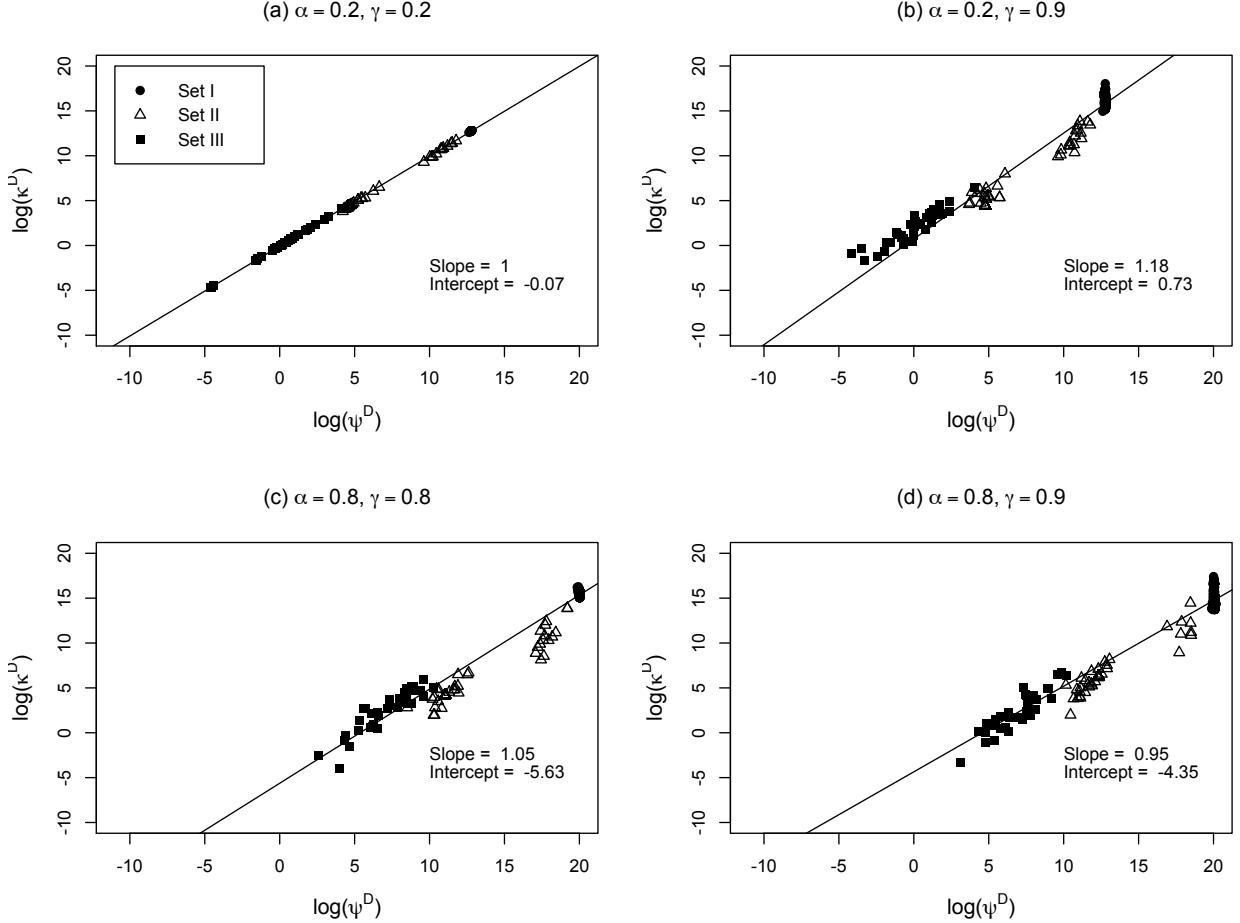


Figure 6: $\log \psi^D$ against $\log \kappa^D$ for selected α and γ values, β as for the aeronautics example, and three sets of designs: set I (\bullet); set II (\triangle); set III (\blacksquare).

sider different values of $\gamma \geq \alpha$ to evaluate the impact of having maximum pairwise correlations as large or larger than specified in the working correlation matrix. This form of R^* introduces a dependence of correlation on treatments \mathbf{x}_j and \mathbf{x}_k and typically will not have the same structure as $R(\alpha)$. A similar approach can be applied to AR and NN correlation structures.

We now compare designs for the bearing experiment by comparing values of the local objective functions $\psi^D(d; \beta, \alpha)$, defined in (6), and

$$\kappa^D(d; \beta, \alpha, \Sigma^*) = \det \left[X^T \Delta V^{-1} \Delta X (X^T \Delta V^{-1} \Sigma^* V^{-1} \Delta X)^{-1} X^T \Delta V^{-1} \Delta X \right], \quad (10)$$

obtained from (3) using $\Sigma^* = A^{1/2} R^* A^{1/2}$. For each value of α considered, three sets of 16-run designs are assessed: (I) 1000 local D -optimal designs for $\beta = (-1, 2, 1, -1, 0.5, 0.5, 0.5)$, found via simulated annealing, resulting in most designs having different design points but equivalent performance under (6); (II) 500 randomly generated designs with $x_{ij} = -1$ or 1 ; and (III) 500 randomly generated designs with $x_{ij} \in [-1, 1]$ ($i=1,2,3$). Figure 6 enables a comparison of objective functions (6) and (10) for each set of designs. For $\gamma = \alpha$, the working correlation matrix is as close as possible to the true correlation matrix. Figure 6(a) shows that each design has near identical performance under (6) and (10) for $\alpha = \gamma = 0.2$. For larger $\alpha = \gamma = 0.8$, R^* may still differ greatly from $R(\alpha)$ but the designs have similar rankings under each objective function (Figure 6(c)). For $\gamma = 0.9$, Figures 6(b) and 6(d), the ranking of designs is also mostly maintained. Similar results were obtained for $\alpha = .4$ and $.6$ (not shown). The D -optimal designs (set I) have a greater spread in the values of κ^D for larger γ . However, under both objective functions and for all α and γ values, each of these designs is the best or close to the

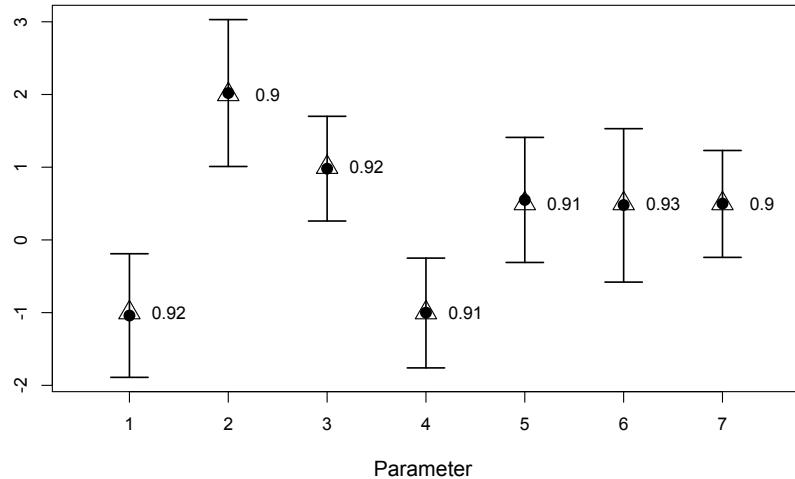


Figure 7: Simulation study results: true values (Δ) of $\beta_0, \dots, \beta_{23}$ (labeled 1, …, 7); average estimates (\bullet); average asymptotic 95% confidence intervals; achieved coverage (as a proportion).

best, and the optimal design under both criteria from the 2000 designs belongs to this set for each case. Hence, in this study, design selection is little affected by departures of the working correlation matrix from the true correlation matrix.

5. ANALYSIS OF SMALL EXPERIMENTS

For sparse binary data with small numbers of Bernoulli trials at each design point, the issue of separation may arise in the analysis (Albert and Anderson, 1984). This complication results in infinite estimates of model parameters (with infinite standard errors), fitted values near 0 or 1, and the solution to the estimating equations lying on the boundary of the parameter space.

The problem has been overcome for GLMs through alternative analyses which are mainly Bayesian in flavor, including the penalized methods of Clogg, Rubin, Schenker, Schultz, and Weidman (1991) and Firth (1993). We adopt a similar approach for GEE models through the addition of a penalty to (1) to give new estimating equations

$$X'\Delta V^{-1}(\mathbf{Y} - \boldsymbol{\mu}) + X'\Delta^*(V^*)^{-1}(\mathbf{m}_s^* - \boldsymbol{\mu}) = \mathbf{0}_p. \quad (11)$$

The N -vector \mathbf{m}_s^* is a vector of pseudo-observations m_{sj}^* ($j = 1, \dots, N$) constructed by assuming that m^* additional, hypothetical Bernoulli trials have been made at each design point. Here m^* is not necessarily an integer, and Δ^* and V^* are defined analogously to Δ and V in (1) for the “prior” data. This adjustment is equivalent to placing a shrinkage prior on $\boldsymbol{\beta}$. Empirical evidence from simulation studies (not presented) suggests the use of $m^* = p/n$ and $m_{sj}^* = 0.5m^*$. These values provide p additional Bernoulli trials in total and shrink the model parameters towards zero (Clogg et al., 1991). This choice of “pseudo-prior” ensures the existence of unique estimates for the model parameters under the prior estimating equations $X'\Delta^*(V^*)^{-1}(\mathbf{m}_s^* - \boldsymbol{\mu}) = \mathbf{0}_p$, and hence leads to a unique solution of (11).

A variety of simulation studies have been carried out to validate this adjustment. One study is presented here for illustration, based on the bearing example. Binomial data were generated according to a probit model with linear predictor (7) and $\boldsymbol{\beta} = (-1, 2, 1, -1, 0.5, 0.5, 0.5)$ for an experiment with 16 binomial observations, each for two trials, arranged in four blocks of four. The exchangeable working correlation structure was adopted with $\alpha = .2$. A local D -optimal design was used, with data generated from the discretization into two classes of partial

sums of independent Poisson random variables (Park, Park, and Shin, 1996) with a constant pairwise correlation of .12, the largest possible value for this design; see (9). The results from 10,000 simulations are summarized in Figure 7 which shows the average point estimates and asymptotic 95% confidence intervals for β . The method performs well with average estimates close to the values from the model used for data generation, and confidence intervals having close to nominal coverage. In addition, the average confidence interval does not contain zero for each parameter β with $|\beta| \geq 1$.

Two issues arise from the adjustment. The adjusted data are underdispersed, as the constant added to the mean does not change the variance. Hence the scale parameter $\phi < 1$ and must be estimated from the data. In addition, the adjustment may result in differing degrees of shrinkage of the coefficients, dependent on the number of prior observations employed. Finding an optimal choice of the number of prior observations is an area for future research.

6. DISCUSSION

There is an increasing recognition of the need to design experiments for non-normal data. This paper has developed the first general methods for block designs and distributions from the exponential family, and has investigated their efficiency by finding designs for an aeronautical application. Three strategies have been proposed for finding efficient designs robust to the model parameters, and an assessment made of the robustness to block-to-block heterogeneity of the designs produced. The methods can be extended to incorporate uncertainty in the functional form of the linear predictor and the link function, see Woods et al. (2006) for GLM designs. The methods also apply to linear response surface models with random block effects.

The comparison of the strategies shows clearly the need for additional support points in a robust design. We conclude that a minimum-support design should not be used when there is substantial uncertainty in the model parameters. Strategies 1 and 3 produce similarly robust designs, with Strategy 3 offering some computational advantages, particularly for local optimal designs where a closed form, optimal unblocked design may be available (Russell et al., 2009; Yang et al., 2009). We have also assessed the robustness of local optimal designs to the value of β . For the bearing example, exchangeable working correlation structure and $\alpha = .2$, the local optimal design for $\beta = (-1, 2, 1, -1, 0.5, 0.5, 0.5)$ has minimum efficiency over the parameter space of Section 4 that is approximately half that of the robust design d_o . The impact on the ranking of designs of the assumption that the true and working correlation matrices are identical was found to be minimal. Further simulation studies (not presented) show that the block designs found using asymptotic theory still outperformed non-blocked optimal designs under the empirical variance-covariance matrix for $\alpha \neq 0$. Also, the asymptotic and empirical parameter variances were found to be close, even for 16-run designs.

We have used marginal models and generalized estimating equations, and proposed a data adjustment that overcomes separation in binary data and ensures finite parameter estimates. We believe marginal modeling is an appropriate strategy for a range of designed experiments, particularly from industry, in the situation when blocks can be regarded as nuisance variables. For other experiments, such as clinical trials and longitudinal studies where the blocks are human subjects, conditional models may be more appropriate (Lee and Nelder, 2004). An area for future research is the robustness of designs to the chosen modeling method.

ACKNOWLEDGEMENTS

This work was supported by EPSRC grant EP/C008863/1. The authors are grateful to Goodrich Engine Control Systems for the example and for providing Figure 1.

APPENDIX A

The use of continuous, or approximate, designs allows the possibility of establishing necessary and sufficient conditions for a design to be optimal, as well as guaranteeing a convex optimization problem for D -optimality. When blocks have equal size k , a continuous block design can be defined as a probability measure on the space of blocks \mathcal{X}^k (see, for example, Cheng, 1995). That is,

$$\xi = \left\{ \begin{array}{cccc} \zeta_1 & \zeta_2 & \cdots & \zeta_b \\ w_1 & w_2 & \cdots & w_b \end{array} \right\}, \quad (12)$$

where $\zeta_l \in \mathcal{X}^k$ is the set of k points which form the l th block, $\sum_{l=1}^b w_l = 1$, and $0 < w_l \leq 1$. The sets ζ_l ($l = 1, \dots, b$) form the support of the block design.

Let $\zeta \in \mathcal{X}^k$ denote a set of points from a single block which may or may not be included in the design. Then the sum of the asymptotic standardized variances for the predicted responses at the points in ζ is $v(\zeta, \xi; \beta, \alpha) = \text{tr} [K(\zeta; \beta, \alpha) M(\xi; \beta, \alpha)^{-1}]$, with $K(\zeta; \beta, \alpha) = X_\zeta^T \Delta_\zeta V_\zeta^{-1} \Delta_\zeta X_\zeta$ and X_ζ , Δ_ζ and V_ζ defined as in (1). The robust D -optimality of a given continuous block design ξ^* for a GEE model can be verified from the following general equivalence theorem.

Theorem. For a GEE model with given linear predictor, link function and working correlation, the following conditions on a continuous design ξ^* with blocks of equal size k are equivalent:

1. ξ^* maximizes $\Psi(\xi; \mathcal{B}, \alpha)$.
2. ξ^* minimizes the maximum of $\Upsilon(\zeta, \xi; \mathcal{B}, \alpha)$ over \mathcal{X}^k , where

$$\Upsilon(\zeta, \xi; \mathcal{B}, \alpha) = \int_{\mathcal{B}} v(\zeta, \xi; \beta, \alpha) \, dF(\beta) - p.$$

3. $\Upsilon(\zeta, \xi^*; \beta, \alpha)$ attains its maximum value of 0 over \mathcal{X}^k at the support points ζ_l of ξ^* .

The proof is similar to that of Chaloner and Larntz (1989) for completely randomized designs; see also Atkinson (2008) for local optimal designs for nonlinear models with random effects.

Example 1. Consider a Poisson regression model with log link, linear predictor $\beta_0 + \beta_1 x + \beta_2 x^2$, and a design in blocks of size $k = 2$ with exchangeable working correlation. A robust D -optimal continuous design for $\alpha = .5$ and a uniform prior distribution on the parameter space $[-1, 1] \times [4, 5] \times [0.5, 1.5]$, found by quasi-Newton computer search, is

$$\xi^* = \left\{ \begin{array}{ccc} (.03, 1) & (1, .60) & (-.40, .78) \\ .355 & .310 & .335 \end{array} \right\},$$

where, for example, block 1 has the two design points $x = .03$ and $x = 1$. Objective function (4) was calculated using a product Gauss rule (Evans and Swartz, 2000, chap. 5). The theorem confirms the D -optimality of this design, with the maximum value of $\Upsilon(\zeta, \xi^*; \beta, \alpha)$ being 0 and occurring at the support points $\zeta_1, \zeta_2, \zeta_3$ of ξ^* , shown in Figure 8.

The theorem is most useful for designs with small block sizes. For large blocks, the need to evaluate $\Upsilon(\zeta, \xi; \mathcal{B}, \alpha)$ for an arbitrary block can lead to a considerable combinatorial problem.

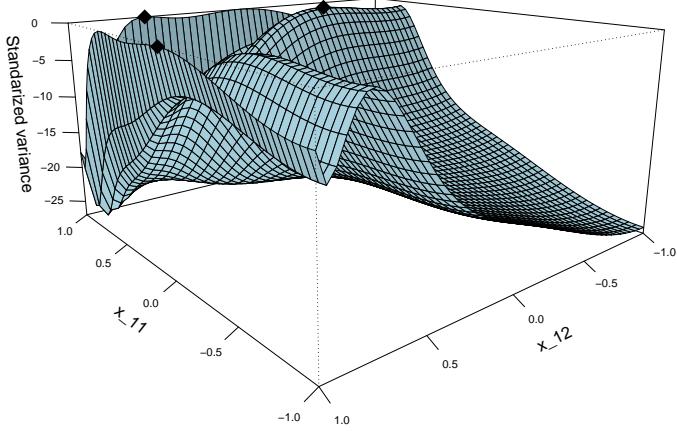


Figure 8: $\Upsilon(\zeta, \xi^*; \boldsymbol{\beta}, \alpha)$ surface for Example 1; x_{11} and x_{12} are the values of the variable x assigned to the same block. Diamonds identify the pairs of values for the three blocks ζ_1 , ζ_2 and ζ_3 of the robust D -optimal design.

APPENDIX B

The decomposition is established by expressing the objective function (4) for minimum-support designs as the sum of two independent functions, namely

$$\Psi(d_m; \mathcal{B}, \alpha) = \int_{\mathcal{B}} \log \det [M_1(d_1; \boldsymbol{\beta})] dF(\boldsymbol{\beta}) + \log \det [M_2(d_2; \alpha)], \quad (13)$$

where $M_1(d_1; \boldsymbol{\beta}) = X_1^T \Delta_1 A_1^{-1} \Delta_1 X_1$ is the information matrix for the GLM; X_1 , Δ_1 and A_1 are defined as in (1) for design d_1 ; $M_2(d_2; \alpha) = Z^T [R(\alpha)]^{-1} Z$ is proportional to the information matrix for the p -vector $\boldsymbol{\tau}$ of treatment effects under the linear model $E(\mathbf{Y}) = Z\boldsymbol{\tau}$ and $\text{Var}(\mathbf{Y}) = \sigma^2 R(\alpha)$. Here $Z = (Z_1^T : \dots : Z_b^T)^T$ where Z_l is the $k_l \times p$ unit-treatment incidence matrix for block l whose (i, j) th element is 1 if the i th unit is allocated to \mathbf{x}_j^* and 0 otherwise ($i = 1, \dots, k_l$; $j = 1, \dots, p$; $l = 1, \dots, b$).

The derivation of (13) from (4) follows from using $A^{-1/2} \Delta X = Z A_1^{-1/2} \Delta_1 X_1$ to show that

$$X^T \Delta A^{-1/2} [R(\alpha)]^{-1} A^{-1/2} \Delta X = X_1^T \Delta_1 A_1^{-1/2} Z^T [R(\alpha)]^{-1} Z A_1^{-1/2} \Delta_1 X_1,$$

and the fact that $A_1^{-1/2} \Delta_1 X_1$ and $Z^T [R(\alpha)]^{-1} Z$ are square matrices. The maximization of $M_2(d_2; \alpha)$ through choice of a D -optimal block design for comparing treatment effects was established by Cheng (1995, lemma 2.1).

References

Albert, A. and Anderson, J. A. (1984), “On the existence of maximum likelihood estimates in logistic regression models,” *Biometrika*, 71, 1–10.

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.

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

Breslow, N. E. and Clayton, D. G. (1993), "Approximate inference in Generalized Linear Mixed Models," *Journal of the American Statistical Association*, 88, 9–25.

Chaganty, N. R. and Joe, H. (2004), "Efficiency of generalized estimating equations for binary responses," *Journal of the Royal Statistical Society Series B*, 66, 851–860.

Chaloner, K. and Larntz, K. (1989), "Optimal Bayesian design applied to logistic regression experiments," *Journal of Statistical Planning and Inference*, 21, 191–208.

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

Clogg, C. C., Rubin, D. B., Schenker, N., Schultz, B., and Weidman, L. (1991), "Multiple imputation of industry and occupation codes in census public-use samples using Bayesian logistic regression," *Journal of the American Statistical Association*, 86, 68–78.

Dror, H. A. and Steinberg, D. M. (2006), "Robust experimental design for multivariate generalized linear models," *Technometrics*, 48, 520–529.

— (2008), "Sequential experimental designs for generalized linear models," *Journal of the American Statistical Association*, 103, 288–298.

Evans, M. and Swartz, T. (2000), *Approximating Integrals via Monte Carlo and Deterministic Methods*, Oxford: Oxford Science Publications.

Firth, D. (1993), "Bias Reduction of Maximum Likelihood Estimates," *Biometrika*, 80, 27–38.

Gotwalt, C. M., Jones, B. A., and Steinberg, D. M. (2009), "Fast computation of designs robust to parameter uncertainty for nonlinear settings," *Technometrics*, 51, 88–95.

Haines, L. M. (1987), "The application of the annealing algorithm to the construction of exact optimal designs for linear regression models," *Technometrics*, 29, 439–447.

Joe, H. (1997), *Multivariate Models and Dependence Concepts*, Boca Raton: Chapman and Hall.

Khuri, A. I., Mukherjee, B., Sinha, B. K., and Ghosh, M. (2006), "Design issues for generalized linear models: a review," *Statistical Science*, 21, 376–399.

Lee, Y. and Nelder, J. (2004), "Conditional and marginal models: another view," *Statistical Science*, 19, 219–238.

Liang, K. and Zeger, S. (1986), "Longitudinal data analysis using generalized linear models," *Biometrika*, 73, 13–22.

McGree, J. M., Eccleston, J. A., and Russell, K. G. (2009), "Robust designs for Poisson regression models," *Submitted for publication*.

Moerbeek, M. and Maas, C. J. M. (2005), "Optimal experimental designs for multilevel logistic models with two binary predictors," *Communications in Statistics - Theory and Methods*, 34, 1151–1167.

Moerbeek, M., Van Breukelen, G. J. P., and Berger, M. P. F. (2001), "Optimal experimental designs for multilevel logistic models," *Journal of the Royal Statistical Society Series D*, 50, 17–30.

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

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

Park, C. G., Park, T., and Shin, D. W. (1996), “A simple method for generating correlated binary variates,” *American Statistician*, 50, 306–310.

Robinson, T., Myers, R., and Montgomery, D. (2004), “Analysis considerations in industrial split-plot experiments with non-normal responses,” *Journal of Quality Technology*, 36, 180–191.

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

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

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

Yang, M., Zhang, B., and Huang, S. (2009), “Optimal designs for binary response experiments with multiple design variables,” *Submitted for publication*.