

# Designs for generalized linear models with random block effects via information matrix approximations

BY T. W. WAITE

*School of Mathematics, University of Manchester, Manchester M13 9PL, U.K.*  
timothy.waite@manchester.ac.uk

AND D. C. WOODS

*Southampton Statistical Sciences Research Institute, University of Southampton,  
Southampton SO17 1BJ, U.K.*  
d.woods@southampton.ac.uk

## SUMMARY

The selection of optimal designs for generalized linear mixed models is complicated by the fact that the Fisher information matrix, on which most optimality criteria depend, is computationally expensive to evaluate. We provide two novel approximations that reduce the computational cost of evaluating the information matrix by complete enumeration of response outcomes, or Monte Carlo approximations thereof: an asymptotic approximation that is accurate when there is strong dependence between observations in the same block; and an approximation via kriging interpolators. For logistic random intercept models, we show how interpolation can be especially effective for finding pseudo-Bayesian designs that incorporate uncertainty in the values of the model parameters. The new results are used to evaluate the efficiency, for estimating conditional models, of optimal designs from closed-form approximations to the information matrix derived from marginal models. Correcting for the marginal attenuation of parameters in binary-response models yields much improved designs, typically with very high efficiencies. However, in some experiments exhibiting strong dependence, designs for marginal models may still be inefficient for conditional modelling. Our asymptotic results provide some theoretical insights into why such inefficiencies occur.

*Some key words:* Bayesian design; Binary response; Blocked experiment; Count response; Generalized linear mixed model; Kriging; Outcome enumeration; Quasilikelihood.

## 1. INTRODUCTION

There is increasing recognition of the need to design experiments in situations where a linear model with only fixed effects cannot adequately capture the essential features of the data. In particular, there is a growing body of work on optimal design for generalized linear models (e.g., Chaloner & Larntz, 1989; Khuri et al., 2006; Woods et al., 2006; Yang et al., 2011) which can be used when the response variable follows a non-normal exponential family distribution. An even larger literature addresses optimal design when there is heterogeneity between blocks in an experiment, using a linear mixed model for normally distributed responses with random block effects (e.g., Cheng, 1995; Goos & Vandebroek, 2001). In many practical contexts, such as the industrial experiment described by Woods & Van de Ven (2011), both non-normality

and heterogeneity are present. For such experiments, particularly where the response is discrete, generalized linear mixed models may be appropriate. In this paper, we develop and compare optimal design methodologies for such models.

We find  $D$ -optimal designs, that is, designs that minimize the volume of the asymptotic confidence ellipsoid for the model parameters by maximizing the determinant of the Fisher information matrix. Dependence of an optimal design on the unknown values of the parameters is addressed using a pseudo-Bayesian approach. The key difficulty in the construction of  $D$ -optimal designs for generalized linear mixed models is that the information matrix is computationally expensive to evaluate.

The adoption of a mixed model implies a marginal distribution for the response with intra-block correlation: two responses from units within the same block are correlated and responses from units in different blocks are uncorrelated. Existing approaches to optimal design for correlated discrete responses (Moerbeek & Maas, 2005; Niaparast, 2009; Woods & Van de Ven, 2011) are tailored for estimation methods, such as quasilielihood and generalized estimating equations, that use only the first- and second-order moments of the marginal distribution. Our focus is on design for likelihood estimation of parameters in the conditional model, for which we present novel approximations to the information matrix. For binary data, we also adapt and extend marginal approximations to provide more efficient designs for the conditional model. We compare designs from the various approximations to those found from computationally expensive gold standard approximations throughout our examples, using either naïve outcome enumeration or Monte Carlo methods.

## 2. PRELIMINARIES

### 2.1. Generalized linear mixed models for blocked experiments

We denote the response for the  $j$ th unit in the  $i$ th block by  $y_{ij}$ , and the corresponding treatment vector of values taken by the  $q$  controllable variables by  $x_{ij} \in \mathcal{X} \subseteq \mathbb{R}^q$  ( $i = 1, \dots, n$ ;  $j = 1, \dots, m_i$ ). Also let  $\zeta_i = (x_{i1}, \dots, x_{im_i}) \in \mathcal{X}^{m_i}$  denote the  $m_i$  treatment vectors in the  $i$ th block. Then, for a generalized linear mixed model, there is a vector,  $u_i$ , of  $r$  random effects associated with the  $i$ th block in the experiment. Conditional on  $u_i$ , the responses in block  $i$  are independent and follow an exponential family distribution,  $y_{ij} | u_i \sim \pi(x_{ij}; u_i, \beta)$ , with mean  $\mu_{ij} = \mu(x_{ij}; u_i, \beta)$  and variance  $\phi v(x_{ij}; u_i, \beta)$ . For the models we consider, the dispersion parameter  $\phi$  equals unity. The mean function  $\mu(x; u, \beta)$  is defined by

$$g\{\mu(x; u, \beta)\} = v(x; u, \beta), \quad v(x; u, \beta) = f^T(x)\beta + z^T(x)u, \quad (1)$$

where  $f: \mathcal{X} \rightarrow \mathbb{R}^p$  and  $z: \mathcal{X} \rightarrow \mathbb{R}^r$  are known vectors of regressor functions, with  $z$  typically a subvector of  $f$ , and  $\beta$  is a  $p$ -vector of fixed regression parameters. The link function  $g$  relates the linear predictor  $v$  to the mean response. Let  $h$  denote the inverse link function,  $g^{-1}$ . To fully determine the model, assumptions about the distribution of  $u_i$  are necessary; we specify independent  $u_i \sim N(0, G)$ , with  $G$  an arbitrary  $r \times r$  covariance matrix. The presence of random effects in the linear predictor introduces correlation between observations from experimental units in the same block. In this paper, we focus principally on random intercept models appropriate for blocked experiments, where  $r = 1$ ,  $G = \sigma^2$ ,  $\sigma^2 > 0$ , and  $z(x) = 1$ .

We assume for simplicity that  $m_i = m$ , and say treatment blocks  $\zeta_1, \zeta_2 \in \mathcal{X}^m$  are equivalent if one can be obtained from the other by treatment permutation. Without loss of generality, we may assume that the blocks are ordered so that the design is supported on  $\zeta_1, \dots, \zeta_b$  ( $1 \leq b \leq n$ ), and that no pair from among the first  $b$  blocks is equivalent. Let  $w_k$  ( $k = 1, \dots, b$ ) be the proportion of blocks equivalent to the  $k$ th support block  $\zeta_k$ . Then we have the following concise notation for

a normalized design:

$$\xi = \left\{ \begin{array}{c} \zeta_1, \dots, \zeta_b \\ w_1, \dots, w_b \end{array} \right\}, \quad 0 < w_k \leq 1, \quad \sum_{k=1}^b w_k = 1. \quad (2)$$

As defined above,  $nw_k$  is a positive integer. We focus on approximate block designs, which relax this constraint (Cheng, 1995). We impose no restrictions on the form of  $\zeta_k$ , and designs may differ in the  $w_k$ , the  $\zeta_k$  and the value of  $b$ . We consider only designs with finite support,  $b \leq n < \infty$ .

## 2.2. Information matrix

Let  $\theta$  denote the complete vector of parameters for model (1). Thus  $\theta$  includes the fixed effects parameters  $\beta$  as well as any parameters specifying the distribution of  $u_i$ . Let  $M_\beta(\xi, \theta)$  denote the information matrix for  $\beta$  under design  $\xi$ , holding all other components of  $\theta$  fixed. The maximum likelihood estimator  $\hat{\beta}$  is distributed approximately as  $N\{\beta, M_\beta^{-1}(\xi, \theta)/n\}$ , assuming known variance components. In common with many papers on design for both linear mixed models (Cheng, 1995; Goos & Vandebroek, 2001) and specific examples of their generalized counterparts (Moerbeek & Maas, 2005; Tekle et al., 2008; Niaparast & Schwabe, 2013), we do not consider the additional variability in  $\hat{\beta}$  introduced when the variance components are also estimated.

For the approximate block design  $\xi$  in (2), the information matrix  $M_\beta$  depends on  $\theta$  and, as observations in different blocks are independent, it can be decomposed into a weighted sum of the information matrices for each support block,

$$M_\beta(\xi, \theta) = \sum_{k=1}^b w_k M_\beta(\zeta_k, \theta). \quad (3)$$

The information matrix for an arbitrary block  $\zeta = (x_1, \dots, x_m) \in \mathcal{X}^m$  is

$$M_\beta(\zeta, \theta) = F^T E_Y \left[ P(Y | \theta, \zeta)^{-2} \left\{ \frac{\partial P(Y | \theta, \zeta)}{\partial \eta} \right\} \left\{ \frac{\partial P(Y | \theta, \zeta)}{\partial \eta} \right\}^T \right] F, \quad (4)$$

where  $Y = (y_1, \dots, y_m)^T$  denotes the response vector or outcome corresponding to  $\zeta$ ,  $P(Y | \theta, \zeta)$  is the marginal likelihood of the model parameters,  $\eta = [f^T(x_1)\beta, \dots, f^T(x_m)\beta]^T$ , and  $F = [f(x_1), \dots, f(x_m)]^T$  is the model matrix. The likelihood and its derivative are of the form

$$P(Y | \theta, \zeta) = \int_{\mathbb{R}^r} P(Y | u, \theta, \zeta) f_u(u) du, \quad \frac{\partial P(Y | \theta, \zeta)}{\partial \eta} = \int_{\mathbb{R}^r} \frac{\partial P(Y | u, \theta, \zeta)}{\partial \eta} f_u(u) du, \quad (5)$$

where  $P(Y | u, \theta, \zeta)$  is the conditional probability density of  $Y$  given  $u$  and  $f_u$  is the density function of a  $N(0, G)$  random variable. Typically a closed form for the partial derivative of the conditional density is available. For random intercept models, the integrals in (5) can be evaluated numerically using Gauss–Hermite quadrature.

For models with binary response, the expectation in (4) can be evaluated by enumeration of outcomes  $Y \in \{0, 1\}^m$ . Expanding the expectation, we obtain

$$M_\beta(\zeta, \theta) = F^T \sum_{Y \in \{0, 1\}^m} P(Y | \theta, \zeta)^{-1} \left\{ \frac{\partial P(Y | \theta, \zeta)}{\partial \eta} \right\} \left\{ \frac{\partial P(Y | \theta, \zeta)}{\partial \eta} \right\}^T F, \quad (6)$$

where the sum is over all possible response patterns in block  $\zeta$ . An obvious approximation to information matrix (4) is via (6) with numerical approximation of (5) using quadrature. We call this approach naïve outcome enumeration. Clearly, for even moderately sized blocks, such an approximation will be computationally expensive.

For other response distributions, such as the Poisson, the expectation in (4) can be approximated by Monte Carlo sampling of response vectors  $Y$ . In practice, to obtain reasonable precision in the approximation of the information matrix using this method, it is necessary to consider many more than the  $2^m$  possible distinct outcomes obtained from a binary model.

### 2.3. Optimality criteria

We study both locally  $D$ -optimal designs, i.e.,  $\xi_D^* = \arg \max_{\xi} |M_{\beta}(\xi, \theta)|$  for an assumed value of  $\theta$ , and pseudo-Bayesian designs. From (3) and an application of Caratheodory's theorem (Silvey, 1980, p. 16), it follows that there is always a locally  $D$ -optimal design supported on at most  $p(p+1)/2 + 1$  distinct blocks. The pseudo-Bayesian approach may be used to construct a design that is more robust to misspecification of the model parameters, and requires specification of a prior distribution,  $\Lambda$ , for  $\theta$ . Given  $\Lambda$ ,  $\xi$  is Bayesian  $D$ -optimal if it maximizes  $\psi(\xi) = E_{\theta} \{\log |M_{\beta}(\xi, \theta)|\}$  (Chaloner & Larntz, 1989). We do not assume that the resulting analysis will be Bayesian, or that it will use prior distribution  $\Lambda$ . When the prior distribution has unbounded support, care must be taken to ensure that  $\psi(\xi) > -\infty$ .

## 3. APPROXIMATIONS VIA MARGINAL MODELS

### 3.1. Marginal quasilielihood

Breslow & Clayton (1993) discussed marginal quasilielihood as a computationally inexpensive, approximate method for estimating the parameters of a generalized linear mixed model. The method is indirect in that it applies standard quasilielihood equations for dependent data (McCullagh & Nelder, 1989, § 9.3) to a linearization of the model about the mean value of the random effects. An information matrix approximation corresponding to this method is

$$M_{\beta}^{\text{marg}}(\xi, \theta) = \sum_{k=1}^b w_k F_k^T V_k^{-1} F_k,$$

where  $F_k$  is the model matrix for  $\zeta_k$ ,  $V_k = \mathcal{V}(\zeta_k, \theta)$  is determined from  $\mathcal{V}(\zeta, \theta) = W(\zeta, \theta)^{-1} + Z(\zeta)GZ(\zeta)^T$ ,  $W(\zeta, \theta)$  is the diagonal matrix with entries  $v(x_1; 0, \beta), \dots, v(x_m; 0, \beta)$ , and  $Z(\zeta) = [z(x_1), \dots, z(x_m)]^T$ . For designs using similar methods, see Moerbeek & Maas (2005).

There are several higher-order marginal quasilielihood approximations in the literature, including that of Goldstein & Rasbash (1996). An approximation to the information matrix using a second-order method was derived in a 2012 University of Southampton PhD thesis by T. W. Waite. Use of this approximation does not result in better designs, so we omit the results here. The marginal quasilielihood approximation is similar to the first-order approximations used in the design of pharmacokinetic studies (Retout & Mentré, 2003).

### 3.2. Generalized estimating equations

Generalized estimating equations (Liang & Zeger, 1986) may be used to estimate parameters when the marginal distribution of the response follows a generalized linear model, making use of a working correlation matrix that need not equal the true correlation matrix. Typically, an exchangeable, autoregressive, or nearest neighbour structure is used for the working correlation.

However these assumptions are incompatible with most known probability models for dependent discrete responses, in which the correlation is a nontrivial function of the treatments and parameters. Indeed no probability model may achieve these simple correlation structures with the required univariate marginal distributions if, for example, the working correlation violates the bounds on correlation for binary data (Joe, 1997, Ch. 7). Nonetheless, the estimators are consistent under misspecification of the correlation structure and may still be highly efficient (Chaganty & Joe, 2004). Here we use generalized estimating equations only to obtain an approximation to the mixed model information matrix.

Woods & Van de Ven (2011) found designs for marginal generalized linear models that are  $D$ -optimal for the generalized estimating equation method under the assumption that the true correlation structure corresponds to a specified working correlation structure. They also found that the resulting designs were robust to a general class of departures from this correlation assumption. Denote the parameters of the assumed marginal model by  $\beta^*$ , the correlation parameter by  $\rho$ , and assume that the marginal model has the same link and variance functions as the conditional model. Then for exchangeable correlation, the normalized asymptotic inverse estimator covariance matrix is

$$M_{\beta}^{\text{gen}}(\xi, \beta^*, \rho) = \sum_{k=1}^b w_k F_k^T D_k \{ (V_k^*)^{1/2} R(\rho) (V_k^*)^{1/2} \}^{-1} D_k F_k,$$

where  $D_k$  is the diagonal matrix with entries  $1/g'(\mu_{k1}^*), \dots, 1/g'(\mu_{km}^*)$ ,  $\mu_{kj}^* = h\{f^T(x_{kj})\beta^*\}$ ,  $V_k^*$  is the diagonal matrix with entries  $v(x_{k1}; 0, \beta^*), \dots, v(x_{km}; 0, \beta^*)$ , and  $R(\rho) = (1 - \rho)I_m + \rho 1_m 1_m^T$ , with  $I_m$  the  $m \times m$  identity matrix and  $1_m$  an  $m$ -vector of ones.

### 3.3. Binary response: adjustment for attenuation of parameters

Use of marginal quasilielihood for the logistic random effects model results in the assumption that the marginal mean has the form  $E(y_{ij}) \approx g^{-1}\{f^T(x_{ij})\beta\}$  (Breslow & Clayton, 1993). Zeger et al. (1988) showed that a better approximation to the marginal mean is given by a logistic relationship with attenuated coefficients,

$$E(y_{ij}) \approx g^{-1} \left[ f^T(x_{ij})\beta / \{1 + c^2 z(x_{ij})^T G z(x_{ij})\}^{1/2} \right], \quad (7)$$

where  $c = 16\sqrt{3}/(15\pi)$ . For a random intercept model,

$$E(y_{ij}) \approx g^{-1}\{f^T(x_{ij})\beta_{\text{att}}\}, \quad \beta_{\text{att}} = \beta (1 + c^2 \sigma^2)^{-1/2}. \quad (8)$$

This suggests that for the logistic random intercept model, more efficient designs might be obtained by adjusting the parameter values to better approximate the marginal mean using (8). Explicitly, we define the adjusted marginal quasilielihood information matrix by

$$M_{\beta}^{\text{adj}}(\xi, \theta) = M_{\beta}^{\text{marg}}(\xi, \theta_{\text{adj}}), \quad \theta_{\text{adj}} = (\beta_{\text{att}}^T, \sigma^2)^T,$$

where  $M_{\beta}^{\text{marg}}$  is the information matrix for  $\beta$  under marginal quasilielihood. In models other than the random intercept the attenuation factor depends on  $x$ , so a constant adjustment cannot be applied for every design point. However, one possibility for a similar approximation may be to apply quasilielihood or generalized estimating equations using (7) as the marginal mean.

To extend the methods of Woods & Van de Ven (2011), we also take account of parameter attenuation by forming the adjusted generalized estimating equation approximation,

$$M_{\beta}^{\text{adj. gen.}}(\xi, \theta, \rho) = M_{\beta}^{\text{gen.}}(\xi, \beta_{\text{att}}, \rho). \quad (9)$$

Here we either choose a value of  $\rho$  following the guidelines laid out, for estimation, by Chaganty & Joe (2004), or treat  $\rho$  as a tuning parameter, i.e., we choose the value of  $\rho$  such that the corresponding  $D$ -optimal design using (9) maximizes  $|M_{\beta}|$  approximated via naïve outcome enumeration.

#### 4. LOGISTIC RANDOM INTERCEPT MODEL

##### 4.1. Asymptotic outcome enumeration

For the logistic random intercept model, the important case of large  $\sigma^2$  results in substantial block-to-block variability and poses a more difficult design problem. In this case, responses in the same block are strongly dependent, and the adjusted marginal quasilielihood and adjusted generalized estimating equation designs may perform quite poorly, see § 5.2. Moreover, for large  $\sigma^2$  naïve outcome enumeration becomes even more computationally expensive, as more quadrature points are required to maintain accuracy in the approximation of the integrals in (5). In this section, we develop asymptotic,  $\sigma^2 \rightarrow \infty$ , expressions for  $P(Y | \theta, \zeta)$  and its derivatives which are combined with (6) to provide a new, direct approximation to the information matrix for large finite  $\sigma^2$ . The additional approximation enables selection of efficient designs for large  $\sigma^2$  at low computational cost compared to naïve outcome enumeration, see § 5.2. Our main results are in Theorems 1–3; first we define some necessary assumptions and notation.

For fixed values of the conditional parameters, the ‘marginal effects’ in  $\beta_{\text{att}}$  attenuate to zero as  $\sigma^2 \rightarrow \infty$ . In order to approximate the more interesting and realistic case where both  $\sigma^2$  is large and there are nonzero marginal effects, we assume the following asymptotic conditions:

*Assumption 1.*  $\beta_{\text{att}} = \beta / (1 + c^2 \sigma^2)^{1/2}$  is fixed;

*Assumption 2.* for each  $j$ , either  $\eta_j^* = f^T(x_j)\beta_{\text{att}}$  is fixed or there exists  $l \neq j$  such that  $\eta_l^*$  is fixed and  $\eta_l^* - \eta_j^* = o(\sigma^{-1})$ .

In order to meet these conditions, we allow the  $x_j$  to vary with  $\sigma^2$ . A simple asymptotic approximation to the information matrix could be derived by treating all  $\eta_j^*$  as distinct and fixed as  $\sigma^2 \rightarrow \infty$ . However, such an approximation would be very poor for designs with  $\eta_l^* \approx \eta_j^*$  for some  $l \neq j$ . Our asymptotic framework allows consideration of the case where there is near-replication of linear predictor values in a block.

Assumptions 1 and 2 allow the partition of  $\mathcal{S} = \{1, \dots, m\}$  as  $\mathcal{N}(j) \cup \mathcal{Z}(j) \cup \mathcal{P}(j)$  for each  $j = 1, \dots, m$ , where  $\mathcal{N}(j) = \{l : \eta_l - \eta_j \rightarrow -\infty\}$ ,  $\mathcal{Z}(j) = \{l : \eta_l - \eta_j \rightarrow 0\}$ , and  $\mathcal{P}(j) = \{l : \eta_l - \eta_j \rightarrow \infty\}$ . Intuitively,  $\mathcal{N}(j)$ ,  $\mathcal{Z}(j)$ ,  $\mathcal{P}(j)$  are the respective sets of indices of linear predictors less than, similar to, and greater than  $\eta_j$ . The limiting expressions we develop for  $P(Y | \theta, \zeta)$  and  $\partial P(Y | \theta, \zeta) / \partial \eta_j$  depend on which elements of  $\mathcal{S}$  belong to  $\mathcal{N}$ ,  $\mathcal{Z}$  and  $\mathcal{P}$ .

It will be useful to identify some particular classes of outcomes.

**DEFINITION 1.** Outcome  $Y = (y_1, \dots, y_m)^T$  is increasing, within the block, if there exists  $j' \in \mathcal{S}$  such that  $y_l = 0$  when  $\eta_l - \eta_{j'} < 0$  and  $y_l = 1$  when  $\eta_l - \eta_{j'} > 0$ .



DEFINITION 2. Outcome  $Y$  is quasi-increasing if there exists  $j' \in \mathcal{S}$  such that  $\mathcal{N}(j') \subseteq \mathcal{S}_0$  and  $\mathcal{P}(j') \subseteq \mathcal{S}_1$ , where  $\mathcal{S}_0 = \{j : y_j = 0\}$  and  $\mathcal{S}_1 = \{j : y_j = 1\}$ , or, equivalently, if  $\{\mathcal{S}_0 \cap \mathcal{P}(j')\} \cup \{\mathcal{S}_1 \cap \mathcal{N}(j')\} = \emptyset$ .

Any outcome that is increasing with the same  $j'$  for all  $\sigma^2$  is clearly also quasi-increasing.

We now make a further assumption necessary for our theorems.

Assumption 3. There exist  $A_j, B_j > 0$  such that  $|\eta_l - \eta_j| > \sigma A_j$  for  $l \in \{\mathcal{S}_0 \cap \mathcal{N}(j)\} \cup \{\mathcal{S}_1 \cap \mathcal{P}(j)\}$  and  $|\eta_l - \eta_j| > \sigma B_j$  for all  $l \in \{\mathcal{S}_1 \cap \mathcal{N}(j)\} \cup \{\mathcal{S}_0 \cap \mathcal{P}(j)\}$ .

This condition holds for large  $\sigma^2$  by Assumptions 1 and 2; it implies that pairs of predictors that diverge asymptotically are at least  $\min_{j=1, \dots, m} \{A_j, B_j\} \sigma$  apart.

THEOREM 1. Suppose that the outcome is quasi-increasing. Then there exists  $j' \in \mathcal{S}$  such that  $\{\mathcal{S}_0 \cap \mathcal{P}(j')\} \cup \{\mathcal{S}_1 \cap \mathcal{N}(j')\} = \emptyset$ , and:

(i) if  $|\mathcal{S}_0 \cap \mathcal{Z}(j')| = 0$  or  $|\mathcal{S}_1 \cap \mathcal{Z}(j')| = 0$ , the outcome is increasing and, as  $\sigma^2 \rightarrow \infty$ ,

$$P(Y | \theta, \zeta) = \max \left[ 0, \Phi \left\{ -\max_{j \in \mathcal{S}_0} (\eta_j / \sigma) \right\} - \Phi \left\{ -\min_{j \in \mathcal{S}_1} (\eta_j / \sigma) \right\} \right] + O(\sigma^{-1});$$

(ii) if  $|\mathcal{S}_0 \cap \mathcal{Z}(j')| \geq 1$  and  $|\mathcal{S}_1 \cap \mathcal{Z}(j')| \geq 1$ , then as  $\sigma^2 \rightarrow \infty$ ,

$$P(Y | \theta, \zeta) = \frac{\phi(\eta_{j'} / \sigma)}{\sigma} \int_{-\infty}^{\infty} \{1 - h(t)\}^{|\mathcal{S}_0 \cap \mathcal{Z}(j')|} h(t)^{|\mathcal{S}_1 \cap \mathcal{Z}(j')|} dt \\ + \sum_{l \in \mathcal{Z}(j')} O(\Delta_{lj'} / \sigma) + O(\sigma^{-2}),$$

where  $\Delta_{lj} = \eta_l - \eta_j$ . The integral has value 1 when  $|\mathcal{Z}(j')| = 2$ .

THEOREM 2. (i) Assume  $j \in \mathcal{S}$  is such that  $\mathcal{N}(j) \subseteq \mathcal{S}_0$  and  $\mathcal{P}(j) \subseteq \mathcal{S}_1$ , which implies that the outcome is quasi-increasing. Then, for arbitrary  $\epsilon > 0$ ,

$$(2y_j - 1) \frac{\partial P(Y | \theta, \zeta)}{\partial \eta_j} = \frac{1}{\sigma} \phi \left( \frac{-\eta_j}{\sigma} \right) \left\{ C_{I(j), J(j)}^{(1)} + \sum_{l \in \mathcal{S}_1 \cap \mathcal{Z}(j) \setminus \{j\}} \Delta_{lj} C_{I(j)-1, J(j)}^{(3)} \right. \\ \left. - \sum_{l \in \mathcal{S}_0 \cap \mathcal{Z}(j) \setminus \{j\}} \Delta_{lj} C_{I(j), J(j)-1}^{(3)} \right\} + \frac{1}{\sigma^2} \phi' \left( \frac{-\eta_j}{\sigma} \right) C_{I(j), J(j)}^{(2)} \\ + \sum_{l \in \mathcal{Z}(j)} O(\Delta_{lj}^2 / \sigma) + O(\sigma^{-3}) + O \left[ \frac{1}{\sigma} e^{-\sigma A_j / \{(1+\epsilon)m\}} \right],$$

where  $I(j) = |\mathcal{S}_1 \cap \mathcal{Z}(j) \setminus \{j\}|$ ,  $J(j) = |\mathcal{S}_0 \cap \mathcal{Z}(j) \setminus \{j\}|$  and, for integers  $I, J \geq 0$ ,

$$C_{I, J}^{(1)} = \int_{-\infty}^{\infty} h'(t) h(t)^I \{1 - h(t)\}^J dt, \quad C_{I, J}^{(2)} = \int_{-\infty}^{\infty} t h'(t) h(t)^I \{1 - h(t)\}^J dt, \\ C_{I, J}^{(3)} = \int_{-\infty}^{\infty} \{h'(t)\}^2 h(t)^I \{1 - h(t)\}^J dt.$$

(ii) If  $j$  is such that  $\mathcal{N}(j) \not\subseteq \mathcal{S}_0$  or  $\mathcal{P}(j) \not\subseteq \mathcal{S}_1$ , then the derivative satisfies

$$\frac{\partial P(Y|\theta, \zeta)}{\partial \eta_j} = O \left\{ \frac{1}{\sigma} e^{-\sigma B_j/(1+\epsilon)} \right\}. \quad (10)$$

Moreover, if the outcome is neither increasing nor quasi-increasing, then (10) holds for all  $j$ .

**THEOREM 3.** *Quasi-increasing outcomes contribute terms of order  $O(\sigma^{-1})$  or  $O(\sigma^{-2})$  to the information matrix in (6). Outcomes that are not quasi-increasing contribute terms of order  $O \left\{ \sigma^{-1} e^{-\sigma \min_j B_j/(1+\epsilon)} \right\}$ , which are asymptotically negligible.*

Proofs are in the Supplementary Material. From Theorem 3, an asymptotic approximation to the information matrix in (6) need only include contributions from increasing and quasi-increasing outcomes. The importance of quasi-increasing outcomes seems difficult to capture using approximations, such as those in §§ 3.1–3.3, that only incorporate the first- and second-order moments of the joint distribution of the responses.

We combine the asymptotic approximations to  $P(Y|\theta, \zeta)$  and  $\partial P(Y|\theta, \zeta)/\partial \eta_j$ , from Theorems 1 and 2 respectively, with (6) to provide an asymptotic outcome-enumeration approximation to the information matrix. The only additional requirement is that for the  $j$ th treatment in the  $i$ th support block ( $i = 1, \dots, b$ ;  $j = 1, \dots, m$ ), we must define a suitable partition of the indices  $\{1, \dots, m\}$  into sets  $\mathcal{N}_i(j)$ ,  $\mathcal{Z}_i(j)$ , and  $\mathcal{P}_i(j)$ . This partition should be such that linear predictor  $\eta_{ij'}$ , relative to  $\sigma$ , is close to, less than, or greater than  $\eta_{ij}$  for  $j' \in \mathcal{Z}_i(j)$ ,  $j' \in \mathcal{N}_i(j)$  or  $j' \in \mathcal{P}_i(j)$  respectively. We propose to form these partitions automatically using the heuristic algorithm given in Appendix 1. As presently implemented, the objective function corresponding to the asymptotic approximation is discontinuous; nonetheless, the resulting designs typically have high efficiencies relative to designs from the naïve outcome-enumeration approximation, competitive with those from the other methods. In certain circumstances, discussed in § 5.2, this asymptotic approximation outperforms other methods.

When  $\sigma^2$  is large, the recovery of inter-block information that occurs when using a mixed model for analysis is important for parameter estimation. For large  $\sigma^2$ , separation of outcomes (Albert & Anderson, 1984) occurs within all blocks with high probability, in which case the parameters of the corresponding fixed block effects model are not estimable. Despite this, efficient parameter estimation is still possible under the mixed model without substantially increasing the number of blocks,  $n$ . For further details, see the Supplementary Material.

#### 4.2. Interpolated outcome enumeration

In this section, we discuss a more direct numerical approximation for  $M_\beta(\xi, \theta)$  under the logistic random intercept model. For this model, the likelihood depends on the regression parameters  $\beta$  only through the vector  $\eta$ . Let  $P_Y(\eta, \sigma^2) = P(Y|\zeta, \theta)$  and define

$$\mathcal{Q}(\eta, \sigma^2) = \sum_{Y \in \{0,1\}^m} \frac{1}{P_Y} \left( \frac{\partial P_Y}{\partial \eta} \right) \left( \frac{\partial P_Y}{\partial \eta} \right)^T,$$

so that, by (6),  $M_\beta(\zeta, \theta) = F^T \mathcal{Q} F$ .

An interpolated outcome-enumeration approximation to  $M_\beta(\zeta, \theta)$  can be developed by surrogate modelling of the matrix-valued function  $\mathcal{Q}$ . The idea is to compute the values of the function  $\mathcal{Q}$  at a collection of training points, and interpolate these data to predict the value of  $\mathcal{Q}$  at new sites  $(\eta, \sigma^2)$ . Interpolating  $\mathcal{Q}$  as a function of  $\eta$  is particularly computationally efficient for finding Bayesian designs, as the same interpolator can be used for any value of  $\beta$ .



Surrogate modelling is widely applied in computer experiments on expensive-to-evaluate computational models for complex phenomena (Santner et al., 2003). For computer experiments, Gaussian process modelling is well-established as a surrogate; it can be used with training sets not arranged in a regular grid and is straightforward to apply to multi-dimensional problems. For block size  $m = 2$ , it is faster to use bilinear or bicubic interpolation and a regular grid.

## 5. EXAMPLES FOR BINARY RESPONSE

### 5.1. Preliminaries for the examples

In §§ 5.2 and 5.3,  $D$ -optimal designs are found, compared and assessed for blocks of size  $m = 4$  and a binary response logistic random intercept model with two variables and the linear predictor

$$v(x; u, \beta) = \beta_0 + \beta_1 x^{(1)} + \beta_2 x^{(2)} + u, \quad u \sim N(0, \sigma^2), \quad (11)$$

where  $x = (x^{(1)}, x^{(2)})^\top \in [-1, 1]^2$ . In § 5.4,  $D$ -optimal designs are found for a logistic random intercept model with four factors and eight fixed parameters.

In §§ 5.2 and 5.4, we find locally  $D$ -optimal designs for various parameter scenarios by approximating the information matrix using adjusted generalized estimating equations and adjusted marginal quasilielihood. In § 5.2 we also find locally optimal designs using unadjusted generalized estimating equations assuming  $\beta^* = \beta$  and, for large  $\sigma^2$ , asymptotic outcome enumeration. In these sections we find it advantageous to specify parameter scenarios on the scale of the marginal effects,  $\beta_{\text{att}}$ , to facilitate performance comparisons across different values of  $\sigma^2$ . This set-up mimics strong information being available for the marginal effects, and uncertainty in the strength of dependence. In § 5.3, we find Bayesian  $D$ -optimal designs, with the prior information specified on the conditional parameters, as no comparisons are made across different values of  $\sigma^2$ ; we set  $\beta_0 \sim U[-0.5, 0.5]$ ,  $\beta_1 \sim U[3, 5]$ ,  $\beta_2 \sim U[0, 10]$ , and  $\sigma^2 = 5$ . Thus, there is substantial uncertainty about  $\beta_2$ , and moderate block-to-block variability. Here, we approximate the information matrix using the adjusted marginal quasilielihood, adjusted generalized estimating equations, and interpolated outcome-enumeration methods. For all of our examples, efficiencies of optimal designs found using the different approximations are calculated relative to  $D$ -optimal designs found using the naïve outcome-enumeration approximation.

For all approximations, we use the Broyden–Fletcher–Goldfarb–Shanno algorithm (Nocedal & Wright, 1999, pp. 136–43) to obtain optimal, or near-optimal, designs numerically; that is, to obtain optimal or highly efficient combinations of  $\zeta_k$ ,  $w_k$  and  $b$ . Multiple random starts of the algorithm are used to attempt to identify a global optimum of the objective function. Convergence is assessed via comparison of the optima obtained from the different starts, and was considered satisfactory for the examples presented here. We assess performance of the obtained designs using local efficiency,  $\text{eff}(\xi; \theta) = \{|M_\beta(\xi, \theta)| / \sup_{\xi'} |M_\beta(\xi', \theta)|\}^{1/p}$ .

### 5.2. Example 1: locally optimal designs

The purpose of this example is two-fold. Firstly, we wish to illustrate the performance of the methods for different  $\sigma^2$ . Secondly, we demonstrate circumstances under which the resulting designs are robust to a reasonable range of values assumed for  $\sigma^2$ .

Table 1 gives the efficiencies of optimal designs from the different approximations relative to an optimal design found using the naïve outcome-enumeration approximation. The unadjusted generalized estimating equation approach is by far the worst method, with efficiencies frequently less than 90%. In most cases, the other closed-form approximations are competitive with

Table 1. *Example 1: computed efficiencies of locally D-optimal designs from different methods*

$\beta_{\text{att}}^T$	Design	$\sigma^2$					
		1	2	5	10	20	50
(0, 1, 1)	Unadj. gen.	96.3–100.0	94.5–100.1	82.9–99.6	78.8–94.3	71.8–87.1	59.9–76.6
	Adj. marg.	100	100	100	100	100	100
	Adj. gen.	99.7–100.0	99.7–100.1	99.2–100.0	98.8–100.0	98.5–100.0	98.1–100.0
	Asymp. enum.					100.0	94.8
(0, 3, 2)	Unadj. gen.	86.2–97.3	84.5–93.2	79.1–85.3	74.7–79.0	70.9–73.4	63.3–67.7
	Adj. marg.	99.9	99.9	100.0	99.9	99.4	95.2
	Adj. gen.	85.3–99.8	85.6–99.6	86.3–99.5	87.2–99.7	87.7–100.0	83.9–98.5
	Asymp. enum.					96.4	97.4
(0, 5, 10) $\times (1 + 5c^2)^{-\frac{1}{2}}$	Unadj. gen.	82.3–96.1	79.8–91.9	70.1–84.4	65.2–78.4	64.7–72.8	52.2–67.0
	Adj. marg.	99.9	99.9	100.0	99.8	99.7	99.8
	Adj. gen.	83.9–99.1	84.1–98.7	84.8–98.9	85.6–99.4	86.0–99.5	83.9–99.1
	Asymp. enum.					94.8	96.1
(1, 2, 3)†	Unadj. gen.	84.3–96.8	83.2–94.5	75.2–88.5	70.6–78.7	65.7–78.9	58.9–73.9
	Adj. marg.	100.4	99.1	96.6	92.1	84.8	73.5(*)
	Adj. gen.	84.1–99.5	85.0–99.0	86.6–99.2	87.6–98.9	86.7–97.2	77.6–91.7(*)
	Asymp. enum.					93.5	98.3
(1, 4, 4)	Unadj. gen.	81.7–96.9	80.7–94.1	76.0–86.3	70.8–79.8	65.0–73.1	58.3–64.6
	Adj. marg.	100.0	100.0	99.9	99.4	98.2	97.2
	Adj. gen.	80.3–99.4	81.0–99.1	81.9–99.3	82.5–99.7	82.5–98.9	82.5–97.4
	Asymp. enum.					97.1	98.2
(1, 3, 3)	Unadj. gen.	82.2–97.1	80.6–93.5	76.4–86.4	71.0–79.7	65.0–72.7	57.1–63.1
	Adj. marg.	99.7	99.5	100.0	99.3	97.9	95.1
	Adj. gen.	79.6–99.3	80.5–98.6	83.1–99.4	84.9–99.7	85.9–98.8	84.6–95.6
	Asymp. enum.					97.1	98.2
(1, 2, 2)	Unadj. gen.	82.1–97.9	81.9–93.4	78.6–88.2	73.4–78.6	67.3–69.5	58.4–64.3
	Adj. marg.	100.6	100.3	100.2	98.5	96.1	92.6
	Adj. gen.	83.3–100.1	84.6–99.3	87.4–99.8	89.3–99.6	90.1–98.0	88.5–95.4
	Asymp. enum.					95.1	97.9
(2, 1, 3)†	Unadj. gen.	84.3–96.8	83.2–94.5	78.5–89.2	73.4–85.9	62.8–76.7	56.4–64.1
	Adj. marg.	99.9	99.1	96.6	92.1	84.8	78.0(*)
	Adj. gen.	83.6–99.5	84.5–99.0	86.1–99.1	86.8–98.9	85.8–97.0	77.4–90.9(*)
	Asymp. enum.					95.0	98.0

Unadj. gen., Unadjusted generalized estimating equations; Adj. marg., adjusted marginal quaslikelihood; Adj. gen., adjusted generalized estimating equations; Asymp. enum., asymptotic outcome enumeration. Reported efficiencies for generalized estimating equation methods are for  $\rho = 0.1, 0.15, 0.2, \dots, 0.7$ . Symbols † and (\*) indicate parameter values for which the adjusted marginal modelling approximations give particularly inefficient designs.

naïve outcome enumeration. The performance of the adjusted generalized estimating equation approach depends critically on the choice of  $\rho$ , treated here as a tuning parameter.

We observed two cases for which the adjusted marginal and adjusted generalized estimating equation methods performed poorly. For  $\beta_{\text{att}} = (1, 2, 3)^T$  and  $(2, 1, 3)^T$ , with  $\sigma^2 = 50$ , the design efficiencies from the former two methods were below 92%. These cases are unusual in that, for all  $\sigma^2 > 1$ , the two marginal approximations selected designs that replicate treatments within at least one of their blocks. This appears inefficient: the designs from both the naïve and asymptotic outcome-enumeration approximations do not feature within-block replication, and the latter design is at least 98% efficient. The only other case where this replication occurred in the

Table 2. *Example 1: computational expense for locally D-optimal designs*

Method	Time per parameter vector (processor-seconds)
Naïve outcome enumeration ( $\sigma^2 = 50$ )	$3 \times 10^5$
Naïve outcome enumeration ( $\sigma^2 = 1$ )	$5 \times 10^4$
Asymptotic outcome enumeration	$7 \times 10^3$
Adjusted generalized estimating equations*	$1 \times 10^3$ ( $8 \times 10^3$ )
Adjusted marginal quasilielihood	$4 \times 10^2$

\*Figure in brackets is indicative of time in processor-seconds when design re-use is impossible.

marginal approximation designs for large  $\sigma^2$  was  $\beta_{\text{att}} = (1, 2, 2)^T$ , where the efficiency was again relatively low. Our theoretical results in § 4.1 suggest that marginal methods may poorly approximate the information matrix for designs featuring within-block replication when  $\sigma^2$  is large. Thus we would recommend some caution when  $\sigma^2$  is large and use of the marginal approximations yields designs featuring within-block replication of treatments. For such designs, the error from these approximations may be large. Additionally, the small  $u$  Taylor approximations underlying the covariance approximation in the adjusted marginal quasilielihood method cannot be expected to be accurate when  $\sigma^2$  is large and large random effects are anticipated.

Table 2 gives the average total processor time for each method, as recorded in a high performance parallel computing environment with 12 2.4 GHz cores per node. The times given are per parameter vector for 100 random starts of the optimization algorithm. Naïve outcome enumeration is the most expensive method followed by asymptotic outcome enumeration, adjusted generalized estimating equations and adjusted marginal quasilielihood. The computational expense of the adjusted generalized estimating equations method depends on the structure of the problem. Here, there are many parameter scenarios with the same values of  $\beta_{\text{att}}$ , which allows re-use of adjusted generalized estimating equation designs for a given  $\beta_{\text{att}}$  for various  $\sigma^2$ . If re-use were not possible, then the time per design would be higher: an indicative figure is given in parentheses. The time to obtain a design for given  $\rho$  is comparable with that from adjusted marginal quasilielihood.

For moderate dependence, i.e.,  $\sigma^2 \leq 10$ , choosing a single value of  $\sigma^2 = 5$  appears to be very robust; for all  $\beta_{\text{att}}$  considered, the naïve outcome-enumeration design with  $\sigma^2 = 5$  has a calculated efficiency of at least 99.1% for  $\sigma^2 = 1, 2, 10$ . Assuming a single value of  $\sigma^2 = 1$  is less robust, though still reasonable, the worst case is when  $\beta_{\text{att}} = (1, 2, 2)^T$  and the true  $\sigma^2 = 10$ ; the efficiency of the resulting design is 97.1%. However, if the dependence is actually strong then the above designs may perform comparatively poorly; when  $\beta_{\text{att}} = (0, 3, 2)^T$ ,  $\sigma^2 = 50$ , the design obtained assuming  $\sigma^2 = 5$  has a calculated  $D$ -efficiency of 93.6%. This robustness of an optimal design to a wide range of assumed values of  $\sigma^2$  is a consequence of specifying the parameters on the marginal scale.

### 5.3. *Example 2: Bayesian optimal designs*

We choose the design  $\xi$  to maximize  $\psi(\xi)$  from § 2.3 and approximate the integral in the objective function by averaging over a Latin hypercube sample of 50 values of  $\beta$  from  $[-0.5, 0.5] \times [3, 5] \times [0, 10]$ . As the value of  $\sigma^2$  is assumed known, for the interpolated outcome-enumeration approximation we only need build a surrogate model of  $Q$  as a function of  $\eta$ .

Bayesian  $D$ -optimal designs were computed for each of the different approximations using 1000 random starts; the support blocks of the designs are shown in Fig. 1, with corresponding weights given in Table 3. A single value  $\rho = 0.6$ , corresponding to fairly strong correlation, was

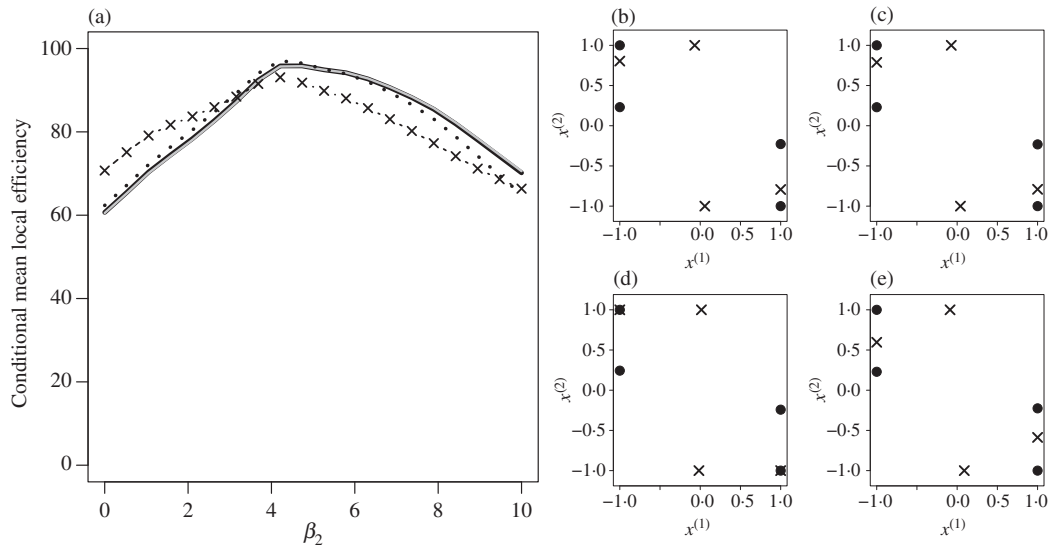


Fig. 1. Example 2: conditional mean efficiencies  $E\{\text{eff}(\xi; \theta) | \beta_2\}$  (a) and support blocks (b–e) of Bayesian  $D$ -optimal designs from maximum likelihood via naïve outcome enumeration (black line, (b)), maximum likelihood via interpolated outcome enumeration (solid grey line, (c)), adjusted marginal quasiliikelihood (dotted black line, (d)), and adjusted generalized estimating equations (dashed and crossed black line, (e)). Treatments with the same plotting character are in the same block.

Table 3. Example 2: details of Bayesian designs

Design method	Block weights		Bayes efficiency	Time (processor-seconds)
	●	×		
Likelihood, naïve outcome enumeration	0.744	0.256	100.00	$1.65 \times 10^7$
Likelihood, interpolated outcome enumeration	0.749	0.251	99.96	$5.19 \times 10^6$
Adjusted marginal quasiliikelihood	0.748	0.252	99.79	$1.80 \times 10^5$
Adjusted estimating equations	0.466	0.534	97.94	$2.20 \times 10^5$

● and × correspond to symbols in Fig. 1(b)–(e). The Bayes efficiency of  $\xi$  is  $\exp[\{\psi(\xi) - \sup_{\xi'} \psi(\xi')\}/p]$ .

used in the adjusted generalized estimating equations approximation. For each method, the best of the 1000 designs generated was selected with respect to naïve outcome enumeration.

All of the designs contain multiple support blocks due to the degree of uncertainty in the parameters and the small block size. Locally  $D$ -optimal designs were also found for each of the 50 sampled parameter vectors under the naïve outcome-enumeration approximation, and the local efficiency of each Bayesian design was calculated relative to each of these 50 designs. Then, Gaussian process emulators were constructed for the efficiency profile of each Bayesian design. Figure 1 shows the dependence of the efficiency on  $\beta_2$ , via approximations of  $E\{\text{eff}(\xi; \theta) | \beta_2\}$  obtained from the efficiency profile emulators. The performance of all of the Bayesian designs varied little according to the value of  $\beta_0$  or  $\beta_1$ , with  $E\{\text{eff}(\xi; \theta) | \beta_0\}$  and  $E\{\text{eff}(\xi; \theta) | \beta_1\}$  changing by fewer than 4 percentage points over the ranges of  $\beta_0$  and  $\beta_1$  respectively. The conditional mean efficiency of the design from the adjusted generalized estimated equations approach is clearly quite different, as a function of  $\beta_2$ , from the mean local efficiencies from the other methods. The designs from all of the approximations appear similar to the naïve outcome-enumeration design, compare Figs. 1(b)–(e).

To train the interpolated outcome-enumeration approximation of  $\mathcal{Q}$ , a random Latin hypercube sample of 10 000  $\eta$  vectors was drawn from  $[-20, 20]^4$ , and the matrix  $\mathcal{Q}$  evaluated for each

vector. The second-order, compactly-supported Wendland covariance function was used, with range parameter chosen as 15 to make the predictions appear reasonably smooth and accurate. Independent Gaussian process models were fitted to the  $m^2$  entries of  $\mathcal{Q}$ . The use of a compactly-supported covariance function is advantageous here due to the large number of training points; it enables inversion of the covariance matrix in a reasonable time, and permits relatively fast predictions from the fitted model. For finding Bayesian designs, the interpolation method required around 3.2 times less computational effort than naïve outcome enumeration for this example, see Table 3. If more quadrature points were used to approximate the prior distribution, or if an adequate emulator could be found using fewer training points, then the advantage of using interpolation to approximate the objective function would be greater. For 200 quadrature points, with the same training set, objective function evaluation using interpolation is approximately 6 times faster than naïve outcome enumeration. The advantage will also be more pronounced for larger  $\sigma^2$ . The closed-form approximations using a single  $\rho$  are approximately two orders of magnitude faster than naïve outcome enumeration.

#### 5.4. Example 3: locally optimal designs, four factors

We investigated locally optimal designs with  $x = (x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)})^T \in [-1, 1]^4$ , and

$$\begin{aligned} v(x; u, \beta) = & \beta_0 + \beta_1 x^{(1)} + \beta_2 x^{(2)} + \beta_3 x^{(3)} + \beta_4 x^{(4)} \\ & + \beta_{12} x^{(1)} x^{(2)} + \beta_{13} x^{(1)} x^{(3)} + \beta_{14} x^{(1)} x^{(4)} + u, \quad u \sim N(0, \sigma^2), \end{aligned}$$

with  $\beta_{\text{att}} = (2, 3, 0, 3, 0, 0, -2, 0)^T$ ,  $(1, 2, 1, -3, -1, 0.25, -0.5, 3)^T$ ,  $(0, 1, 1, 1, 1, 0.5, 0.5, 0.5)^T$ , and  $\sigma^2 = 1, 2, 5$ . Designs were found using the naïve outcome-enumeration, adjusted marginal quasilielihood and adjusted generalized estimating equation approximations with 100, 1000 and 1000 random starts respectively, and  $\rho = 0.3, 0.5, 0.6$ . In all cases, the marginal approximations required less computational effort despite the more thorough search, yielding designs with at least 99.5% efficiency relative to the design from the naïve outcome-enumeration approximation.

## 6. POISSON RESPONSE

### 6.1. Approach

In this section we use marginal quasilielihood approximation to find  $D$ -optimal designs for a Poisson model with random intercept. We compare the designs to those of Niaparast (2009), who investigated design for this model using a direct quasilielihood approximation to the information matrix, and also to the designs from the analytical results of Russell et al. (2009) for the Poisson model with no random effects. The conditional distribution of the response is assumed to be Poisson, with link function  $g(\mu) = \log(\mu)$ . In the random intercept model,  $u \sim N(0, \sigma^2)$  is a scalar, and  $v(x; u, \beta) = f^T(x)\beta + u$ .

Quasilielihood estimating equations may be formed given a parametric specification of the marginal mean and variance of the response. Niaparast (2009) obtained a covariance matrix for the resulting parameter estimators using the actual marginal mean and variance for the Poisson random intercept model, which are analytically tractable. We shall refer to this as the direct approach. Although in general there may not be a quasilielihood function corresponding to these estimating equations (McCullagh & Nelder, 1989, Ch.9), the approach of Niaparast could be viewed as an application of generalized estimating equations (Liang & Zeger, 1986) with a working correlation structure calculated from the full probability model.



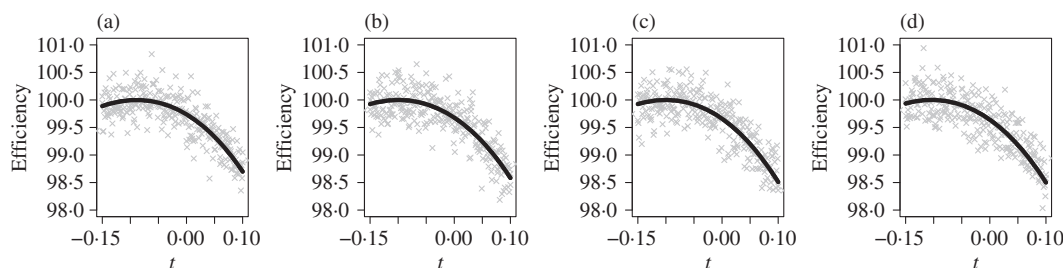


Fig. 2. Approximate efficiency of the design  $\zeta(t)$  obtained using Monte Carlo approximation and nonparametric smoothing, for the following values of  $\sigma^2$ : (a) 0.01, (b) 0.025, (c) 0.05, (d) 0.1.

## 6.2. Comparison of designs, $m = 3$

Locally  $D$ -optimal designs for the Poisson random intercept model were computed by numerically optimizing the determinant of the information matrix under the marginal quasiliquelihood and direct, quasiliquelihood, approximations. The linear predictor structure (11) was assumed, with conditional parameter values  $(\beta_0, \beta_1, \beta_2) = (3, 1, 2)$ , together with several values for  $\sigma^2$ .

For  $\sigma^2 = 0$ , the designs found numerically coincided with those anticipated by the theoretical results of Russell et al. (2009) for models with no random effects. For  $\sigma^2 = 0.01, 0.025, 0.05, 0.1$ , each of the designs contains a single support block ( $b = 1, w_1 = 1$ ) of the form  $\zeta(t) = \{(1, 1)^T, (-1, 1)^T, (1, t)^T\}$ , with  $t = -0.083, -0.091, -0.095, -0.096$  respectively. The designs from the two methods agree to three decimal places.

We assess the efficiency, for maximum likelihood estimation, of designs resulting from the choice of  $t \in [-1, 1]$  by using Monte Carlo integration to approximate  $M_\beta\{\zeta(t); \theta\}$  and nonparametric smoothing to obtain a surrogate,  $\tilde{\Psi}(t)$ , for  $\Psi(t) = |M_\beta\{\zeta(t); \theta\}|$  (see also Müller & Parmigiani, 1995). Let  $t^* = \max_{t' \in [-1, 1]} \tilde{\Psi}(t')$ . Figure 2 shows the approximate efficiency in the neighbourhood of the optimal  $t$ , obtained from  $\text{eff}(t) = \{\tilde{\Psi}(t)/\tilde{\Psi}(t^*)\}^{1/p}$ , together with estimates of  $\{\Psi(t)/\Psi(t^*)\}^{1/p}$  each using  $10^5$  Monte Carlo samples. The total processor time for the Monte Carlo computations was approximately  $1.5 \times 10^5$  s, using a sixteen-core 2.6 GHz node. The results indicate that, for all values of  $\sigma^2$  considered here, both the marginal and direct quasiliquelihood designs have an efficiency around 100%, and also any choice of  $t$  in  $[-0.15, 0]$  will be very highly efficient.

The direct, quasiliquelihood, approach for the Poisson response is similar to the adjusted marginal quasiliquelihood or adjusted generalized estimating equations methods for a binary response, in the sense that it accounts for the form of the marginal mean. Theoretically, it has the advantage of not relying on Taylor series approximations. Unlike for binary data, where an unadjusted marginal method is poor, for a Poisson response the unadjusted method has virtually identical performance to the direct method. This is perhaps to be expected if we consider that the normal approximations to the response distribution used in the marginal approximation are much more accurate for Poisson than binary responses. There is essentially no computational advantage to the Taylor-series based approximation and so we would recommend the direct approach as a default first choice.

The values of  $\sigma^2$  used here are much smaller than those used for binary response models in § 5. For a Poisson response,  $\sigma^2$  is chosen to give a plausible range of marginal overdispersion over  $\mathcal{X}$ , e.g., approximately 1.10–45.6 when  $\sigma^2 = 0.1$ , and plausible correlation between responses from units in the same block receiving the same treatment.



## 7. DISCUSSION

For the logistic random intercept model, use of a correction for the marginal attenuation of the parameters yields much improved designs; in our examples, designs using this idea often performed on a par with those from naïve outcome enumeration. Further investigations, including simulations to assess small sample properties, are available in the first author's PhD thesis.

Tekle et al. (2008) employed an information matrix approximation derived from penalized quasilielihood (Breslow & Clayton, 1993). Their approach requires predictions of the random effects, which they approximated at the design stage using Monte Carlo simulation. The resulting approximation is computationally intensive and is not suitable for routine use on more complex problems. Hence, we choose not to pursue this.

Avenues for future research include developing the necessary methodology to extend the adjusted closed-form approximations to find designs for models with more complex random effects, and extension of the asymptotic results in § 4.1 to other link functions for binary response.

## ACKNOWLEDGEMENT

The authors thank T. H. Waterhouse, Eli Lilly, for helpful discussions, and the referees for useful comments. This work was supported by the UK Engineering and Physical Sciences Research Council, a PhD studentship and Doctoral Prize for the first author, and a Fellowship for the second author. The authors acknowledge the use of the Iridis computational cluster at the University of Southampton.

## SUPPLEMENTARY MATERIAL

Supplementary material available at *Biometrika* online includes proofs of the theoretical results in § 4.1, and additional results for large  $\sigma^2$ .

## APPENDIX

### *An algorithm for determining membership of sets $\mathcal{N}(j)$ , $\mathcal{Z}(j)$ , $\mathcal{P}(j)$*

To obtain an asymptotic approximation that performs reasonably for a broad choice of designs, a decision is required on which  $\eta_j$  should be considered close; that is, for given  $\zeta$  and  $j$ , which indices should we treat as belonging to  $\mathcal{Z}(j)$  in order to apply Theorems 1–3? Below we give the algorithm used in Example 1. The algorithm approximates exponentially decaying error terms as zero.

For calculation of the derivatives, the algorithm iteratively augments  $\mathcal{Z}(j)$  with the index,  $l$ , of the next closest predictor to  $\eta_j$  if two conditions, labelled (A) and (B) in the algorithm below, are satisfied. Condition (A) concerns the coefficient of  $\phi(-\eta_j/\sigma)/\sigma$  in the expression of Theorem 2, which is an approximation to an integral of the form  $\int_{-\infty}^{\infty} h'(t)f_2(t, \sigma^2) dt$ , see equation (A3) in the Supplementary Material. The value of this integral decreases as the set  $\mathcal{Z}(j)$  is augmented. Condition (B), concerning the same coefficient, is a heuristic that prevents the application of a Taylor approximation when  $\Delta_{lj}$  is too large, see (A4) in the Supplementary Material.

For the probability calculation, we use the expression from part (i) of Theorem 1 unless  $\max_{j \in \mathcal{S}_0} \{\eta_j\}$  and  $\min_{j \in \mathcal{S}_1} \{\eta_j\}$  are less than 1 apart, in which case we take  $j' = \arg \max_{j \in \mathcal{S}_0} \{\eta_j\}$ ,  $l' = \arg \min_{j \in \mathcal{S}_1} \{\eta_j\}$ ,  $\mathcal{Z}(j') = \{j', l'\}$  and use the expression in part (ii) of Theorem 1. The cut-off distance of  $\gamma = 1$  is chosen because at this point the probabilities in parts (i) and (ii) should be similar, since  $\Phi(-\eta_j/\sigma) - \Phi(-\eta_l/\sigma) \approx (\eta_l - \eta_j)\phi(-\eta_j/\sigma)/\sigma = \phi(-\eta_j/\sigma)/\sigma$ .

*Algorithm A1.* For each possible outcome  $Y$ , approximate its contribution,  $P(Y)^{-1} \{ \partial P(Y)/\partial \eta \} \{ \partial P(Y)/\partial \eta \}^T$ , to the information matrix in (6) using Theorems 1–3 to approximate  $P(Y)$  and  $\partial P(Y)/\partial \eta$ , and add it to the total. To compute  $P(Y)$ :

Compute  $\lambda_0 = \max_{j \in S_0} \{\eta_j\}$  and  $\lambda_1 = \min_{j \in S_1} \{\eta_j\}$   
 If  $\lambda_1 \geq \lambda_0 + \gamma$ :  
   Set  $P(Y) \leftarrow \Phi(\lambda_1/\sigma) - \Phi(\lambda_0/\sigma)$ , using Theorem 1(i)  
 If  $|\lambda_1 - \lambda_0| \leq \gamma$ :  
   Set  $P(Y) \leftarrow \phi(\lambda_1/\sigma)/\sigma$ , using Theorem 1(ii)  
 If  $\lambda_1 \leq \lambda_0 - \gamma$ , set  $1/P(Y) \leftarrow 0$ , and do not compute  $\partial P(Y)/\partial \eta_j$   
   i.e., do not include a contribution from this outcome in the information matrix approximation.

To compute  $\partial P(Y)/\partial \eta_j$ :

Declare  $\mathcal{Z}(j) = \{j\}$   
 Set  $C_4 = 1$   
 Propose augmenting  $\mathcal{Z}(j)$  to  $\mathcal{Z}'(j) = \{j, \arg \min_{l \neq j} |\eta_j - \eta_l|\}$   
 Iterate until STOP. Given current proposal  $\mathcal{Z}'(j)$ :  
   Calculate  $I(j)$ ,  $J(j)$  for  $\mathcal{Z}'(j)$ , refer to as  $I'$ ,  $J'$  respectively  
   Set  $C'_4 \leftarrow C_{I',J'}^{(1)} + C_{I'-1,J'}^{(3)} \sum_{l \in S_1 \cap \mathcal{Z}'(j) \setminus \{j\}} \Delta l_j - C_{I',J'-1}^{(3)} \sum_{l \in S_0 \cap \mathcal{Z}'(j) \setminus \{j\}} \Delta l_j$   
   If (A)  $0 \leq C'_4 \leq C_4$  and (B)  $|C'_4 - C_{I',J'}^{(1)}| \leq |C_4 - C_{I',J'}^{(1)}|$ , accept proposal  
   Update  $C_4 \leftarrow C'_4$ ,  $\mathcal{Z}(j) \leftarrow \mathcal{Z}'(j)$   
   If did not accept proposal in previous step, then STOP  
   Otherwise make new proposal,  $\mathcal{Z}'(j) \leftarrow \mathcal{Z}(j) \cup \arg \min_{l \notin \mathcal{Z}(j)} \{|\eta_l - \eta_j|\}$   
 Set  $\mathcal{N}(j) \leftarrow \{l : \eta_l < \eta_{l'}, \text{ for all } l' \in \mathcal{Z}(j)\}$   
 Set  $\mathcal{P}(j) \leftarrow \{l : \eta_l > \eta_{l'}, \text{ for all } l' \in \mathcal{Z}(j)\}$   
 If  $\{S_1 \cap \mathcal{N}(j)\} \cup \{S_0 \cap \mathcal{P}(j)\} = \emptyset$ :  
   Deem the outcome as quasi-increasing  
   Set  $\partial P(Y)/\partial \eta_j \leftarrow (2y_j - 1) \max \left\{ 0, \phi(-\eta_j/\sigma) C_4/\sigma + \phi'(-\eta_j/\sigma) C_{I(j),J(j)}^{(2)}/\sigma^2 \right\}$ ,  
   using Theorem 2(i)  
 Else set  $\partial P(Y)/\partial \eta_j \leftarrow 0$ , using Theorem 2(ii).

## REFERENCES

- ALBERT, A. & ANDERSON, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika* **71**, 1–10.
- BRESLOW, N. E. & CLAYTON, D. G. (1993). Approximate inference in generalized linear mixed models. *J. Am. Statist. Assoc.* **88**, 9–25.
- CHAGANTY, N. R. & JOE, H. (2004). Efficiency of generalized estimating equations for binary responses. *J. R. Statist. Soc. B* **66**, 851–60.
- CHALONER, K. & LARNTZ, K. (1989). Optimal Bayesian design applied to logistic regression experiments. *J. Statist. Plan. Infer.* **21**, 191–208.
- CHENG, C. S. (1995). Optimal regression designs under random block-effects models. *Statist. Sinica* **5**, 485–97.
- GOLDSTEIN, H. & RASBASH, J. (1996). Improved approximations for multilevel models with binary responses. *J. R. Statist. Soc. A* **159**, 505–13.
- GOOS, P. & VANDEBROEK, M. (2001).  $D$ -optimal response surface designs in the presence of random block effects. *Comp. Statist. Data Anal.* **37**, 433–53.
- JOE, H. (1997). *Multivariate Models and Dependence Concepts*. Boca Raton: Chapman and Hall/CRC.
- KHURI, A. I., MUKHERJEE, B., SINHA, B. K. & GHOSH, M. (2006). Design issues for generalized linear models: A review. *Statist. Sci.* **21**, 376–99.
- LIANG, K. Y. & ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.
- MCCULLAGH, P. & NELDER, J. A. (1989). *Generalized Linear Models*, 2nd ed. London: Chapman and Hall.

- MOERBEEK, M. & MAAS, C. J. M. (2005). Optimal experimental designs for multilevel logistic models with two binary predictors. *Commun. Statist. A* **34**, 1151–67.
- MÜLLER, P. & PARMIGIANI, G. (1995). Optimal design via curve fitting of Monte Carlo experiments. *J. Am. Statist. Assoc.* **90**, 1322–30.
- NIAPARAST, M. (2009). On optimal design for a Poisson regression model with random intercept. *Statist. Prob. Lett.* **79**, 741–7.
- NIAPARAST, M. & SCHWABE, R. (2013). Optimal design for quasi-likelihood estimation in Poisson regression with random coefficients. *J. Statist. Plan. Infer.* **143**, 296–306.
- NOCEDAL, J. & WRIGHT, S. J. (1999). *Numerical Optimization*. New York: Springer.
- RETOUT, S. & MENTRÉ, F. (2003). Further developments of the Fisher information matrix in nonlinear mixed effects models with evaluation in population pharmacokinetics. *J. Biopharma. Statist.* **13**, 209–27.
- RUSSELL, K. G., WOODS, D. C., LEWIS, S. M. & ECCLESTON, J. A. (2009). D-optimal designs for Poisson regression models. *Statist. Sinica* **19**, 721–30.
- SANTNER, T. J., WILLIAMS, B. J. & NOTZ, W. I. (2003). *The Design and Analysis of Computer Experiments*. New York: Springer-Verlag.
- SILVEY, S. D. (1980). *Optimal Design*. London: Chapman and Hall.
- TEKLE, F. B., TAN, F. E. S. & BERGER, M. P. F. (2008). Maximin D-optimal designs for binary longitudinal responses. *Comp. Statist. Data Anal.* **52**, 5253–62.
- WOODS, D. C. & VAN DE VEN, P. (2011). Block designs for experiments with correlated non-normal response. *Technometrics* **53**, 173–82.
- 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**, 284–92.
- YANG, M., ZHANG, B. & HUANG, S. (2011). Optimal designs for generalized linear models with multiple design variables. *Statist. Sinica* **21**, 1415–30.
- ZEGER, S. L., LIANG, K. Y. & ALBERT, P. S. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics* **44**, 1049–60.

[Received March 2013. Revised January 2015]