

Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.



This article was published in an Elsevier journal. The attached copy is furnished to the author for non-commercial research and education use, including for instruction at the author's institution, sharing with colleagues and providing to institution administration.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



ELSEVIER

Available online at www.sciencedirect.com

Journal of Statistical Planning and Inference 138 (2008) 169–183

journal of
statistical planning
and inferencewww.elsevier.com/locate/jspi

Comparison of designs for multivariate generalized linear models

S. Mukhopadhyay*, A.I. Khuri

Department of Statistics, University of Florida, P.O. Box 118545, Gainesville, FL 32611-8545, USA

Available online 16 May 2007

Abstract

The purpose of this paper is to discuss response surface designs for multivariate generalized linear models (GLMs). Such models are considered whenever several response variables can be measured for each setting of a group of control variables, and the response variables are adequately represented by GLMs. The mean-squared error of prediction (MSEP) matrix is used to assess the quality of prediction associated with a given design. The MSEP incorporates both the prediction variance and the prediction bias, which results from using maximum likelihood estimates of the parameters of the fitted linear predictor. For a given design, quantiles of a scalar-valued function of the MSEP are obtained within a certain region of interest. The quantiles depend on the unknown parameters of the linear predictor. The dispersion of these quantiles over the space of the unknown parameters is determined and then depicted by the so-called quantile dispersion graphs. An application of the proposed methodology is presented using the special case of the bivariate binary distribution.

© 2007 Elsevier B.V. All rights reserved.

Keywords: Bivariate binary distribution; Mean-squared error of prediction; Prediction bias; Prediction variance; Quantile dispersion graphs; Quantiles of the mean-squared error of prediction; Response surface design

1. Introduction

One of the objectives of response surface methodology (RSM) is the proper selection of a design for the fitting of a postulated model that relates a response of interest, y , to a set of control variables. A description of the various design criteria used in RSM can be found in the books by Box and Draper (1987), Myers and Montgomery (1995), and Khuri and Cornell (1996). These criteria were initially developed for standard linear models (mainly first-degree or second-degree polynomials), where the responses are assumed to be continuous (quite often, normally distributed) with uncorrelated errors and homogeneous variances. In many experimental situations, however, such assumptions do not hold. For example, the functional relationship between y and the control variables may be nonlinear, as is the case with models that depict the growth of a particular organism (growth models). In other situations, the response data may be discrete, possibly correlated, and/or exhibit heterogeneous variances. This is quite common in, for example, clinical studies and dose-response experiments. For such data, the use of generalized linear models (GLMs) would be more appropriate. A classic book on GLMs is the one by McCullagh and Nelder (1989) (see also Dobson, 2002; McCulloch and Searle, 2001; Myers et al., 2002).

* Corresponding author. Tel.: +1 706 721 4454; fax: +1 706 721 6294.

E-mail addresses: siulimukherjee@gmail.com (S. Mukhopadhyay), ufakhuri@stat.ufl.edu (A.I. Khuri).

1.1. The design dependence problem

The design criteria that are commonly used in linear models, such as A -, D -, E -, and G -optimality, continue to be used in nonlinear models and GLMs. However, unlike designs for linear models, those for the latter models depend on the unknown parameters of the fitted model. In this case, the construction of a design would unfortunately require some knowledge of the parameters. Common approaches to solving this design dependence problem include:

- (a) The specification of initial values, or best “guesses”, of the unknown parameters, and the subsequent determination of the so-called locally optimal designs, which were first suggested in the seminal paper by [Box and Lucas \(1959\)](#). More recent references concerning designs for the logistic regression model include those by [Sitter and Wu \(1993\)](#) and [Mathew and Sinha \(2001\)](#).
- (b) The sequential approach whereby updated estimates of the parameters are obtained in successive stages of the design. Sequential designs for GLMs were proposed by [Wu \(1985\)](#), [Sitter and Forbes \(1997\)](#), and [Sitter and Wu \(1999\)](#). [Huang and Wong \(1998\)](#) presented a sequential procedure for constructing multiple-objective locally optimal designs for nonlinear models. A review of the sequential approach was recently given in [Khuri et al. \(2006\)](#).
- (c) The Bayesian approach where a prior distribution is assumed on the parameters, which is then incorporated into an appropriate design criterion by integrating it over the prior distribution. One of the early papers on this approach is the one by [Zacks \(1977\)](#) who adopted a Bayesian D -optimality criterion. This corresponds to maximizing the average over the prior distribution of the logarithm of the determinant of the information matrix. Another criterion, namely, the Bayesian A -optimality, corresponds to minimizing the average over the prior distribution of the trace (or weighted trace) of the inverse of the information matrix.

In general, Bayesian designs are found by using numerical methods, except for designs for one-parameter nonlinear models when the prior has two points of support, which were derived analytically by [Haines \(1995\)](#). Designs for a family of exponential models were presented by [Dette and Sperlich \(1994\)](#) and [Mukhopadhyay and Haines \(1995\)](#) (see [Atkinson and Haines, 1996](#)). [Chaloner and Larntz \(1989\)](#) discussed the general theory of optimal Bayesian design and applied it to logistic regression. [Atkinson et al. \(1993\)](#) developed Bayesian designs that are D - and D_s (optimal for a subset of parameters)-optimal for a compartmental model. This is a particular nonlinear model of the exponential type that has biological applications. [Chaloner and Verdinelli \(1995\)](#) and [Dasgupta \(1996\)](#) presented an excellent overview of Bayesian design ideas and their applications. [Ford et al. \(1989\)](#) summarized work in optimal designs for nonlinear models, and addressed the problem of dependence of the design on the unknown parameters of the model. A recent review of the Bayesian approach was given in [Khuri et al. \(2006\)](#).

1.2. Designs for multiresponse models

For the most part, design criteria for linear as well as nonlinear models, including GLMs, pertain to models for a single response. There are, however, many experimental situations where several responses can be observed for each setting of a group of control variables. The response data may be correlated within experimental runs but are usually assumed to be independent among runs.

One of the earliest articles on multiresponse design is the one by [Draper and Hunter \(1966\)](#) for nonlinear multiresponse models. They developed a criterion for selecting additional experimental runs after a certain number of runs have already been chosen. They used a Bayesian approach whereby a locally uniform prior distribution is assumed for the parameter vector, θ , in the multiresponse model. The posterior density of θ , obtained after the additional runs have been taken, is then maximized with respect to θ and the additional experimental runs. The variance–covariance matrix, Σ , for the responses (within an experimental run) is assumed known. [Box and Draper \(1972\)](#) extended [Draper and Hunter's \(1966\)](#) criterion by considering designs divided into blocks with different variance–covariance structures, which may be unknown.

One of the most commonly used design criteria for nonlinear multiresponse models is that of D -optimality, according to which a design is selected that minimizes the generalized variance (volume of the joint confidence ellipsoid) of the estimated model parameters. [Hatzis and Larntz \(1992\)](#) constructed locally D -optimal designs for a nonlinear multiresponse model used in describing the behavior of a biological system. [Fedorov \(1972\)](#) introduced a sequential

procedure for the generation of a D -optimal design for a linear multiresponse model, but assumed knowledge of Σ . Wijesinha and Khuri (1987) proposed a modification of Fedorov's procedure by using an estimate of Σ at each step of the sequential process. D -optimal designs for linear multiresponse models were more recently discussed by Krafft and Schaefer (1992), Bischoff (1993), Chang (1994), and Imhof (2000). A detailed review of basic methods used in the design and analysis of multiresponse experiments was given by Khuri (1996).

When the responses are adequately represented using GLMs, we obtain the so-called *multivariate GLMs*. Very little is known about designs for such models. Heise and Myers (1996) developed designs for the bivariate logistic regression model. Their design criterion addressed the minimization of the prediction variance integrated over the experimental region. Zocchi and Atkinson (1999) considered optimal designs for multinomial logistic models.

The purpose of this paper is to present a graphical technique based on the so-called *quantile dispersion graphs* (QDGs) for the comparison of designs for multivariate GLMs. This technique was initially introduced in Robinson and Khuri (2003) for the comparison of designs for a single logistic regression model. It provides an alternative approach to the aforementioned design dependence problem for nonlinear models and GLMs. This approach is based on studying the distribution of the mean-squared error of prediction (MSEP) through its quantiles over the experimental region. The MSEP incorporates the prediction variance and the prediction bias associated with the fitted model. The quantiles of the MSEP depend on the unknown parameters of the model, which are assumed to belong to a specified parameter space. The QDGs consist of plots of the maxima and the minima, over the parameter space, of the quantiles of the MSEP, which are obtained on concentric surfaces within the experimental region. These plots provide a comprehensive assessment of the quality of prediction afforded by a given design. They also depict the dependence of the design on the model's unknown parameters. More recently, Khuri and Mukhopadhyay (2006) adopted the same technique to compare designs for log-linear models representing Poisson-distributed data. However, both articles (the ones by Robinson and Khuri, 2003; Khuri and Mukhopadhyay, 2006) consider GLMs with only one single response. This paper extends the use of QDGs to multivariate GLMs.

2. Multivariate GLMs

In analogy to the univariate case, multivariate GLMs require the specification of the following three components:

- (1) The data set under consideration consists of n independent q -dimensional random variables, $\mathbf{y}_1, \dots, \mathbf{y}_n$. The distribution of \mathbf{y}_j ($j = 1, \dots, n$) belongs to the exponential family with the density function

$$\delta(\mathbf{y}_j | \boldsymbol{\theta}_j, \phi) = \exp[\phi\{\mathbf{y}'_j \boldsymbol{\theta}_j - b(\boldsymbol{\theta}_j)\} + c(\mathbf{y}_j, \phi)], \quad j = 1, \dots, n, \quad (2.1)$$

where $b(\cdot)$ and $c(\cdot)$ are known scalar functions, and ϕ is a dispersion parameter, possibly unknown. The vector $\boldsymbol{\theta}_j$, $j = 1, \dots, n$, consists of q elements. The mean vector, $\boldsymbol{\mu}_j$, and variance–covariance matrix, $\boldsymbol{\Sigma}_j$, of \mathbf{y}_j are

$$\boldsymbol{\mu}_j = \frac{\partial b(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j}, \quad \boldsymbol{\Sigma}_j = \frac{1}{\phi} \frac{\partial^2 b(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j \partial \boldsymbol{\theta}'_j}, \quad (2.2)$$

respectively.

- (2) A q -dimensional *linear predictor*, $\boldsymbol{\eta}$, which is a polynomial function in k control variables, x_1, \dots, x_k , and is of the form

$$\boldsymbol{\eta}(\mathbf{x}) = \mathbf{Z}'(\mathbf{x})\boldsymbol{\beta}, \quad (2.3)$$

where $\mathbf{x} = (x_1, \dots, x_k)'$, $\mathbf{Z}(\mathbf{x}) = \bigoplus_{i=1}^q \mathbf{f}_i(\mathbf{x})$, $\mathbf{f}_i(\mathbf{x})$ is a known vector function of \mathbf{x} , $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \dots, \boldsymbol{\beta}'_q)'$ is a p -dimensional vector of unknown parameters, and $\boldsymbol{\beta}_i = (\beta_{i1}, \dots, \beta_{ip_i})'$ is a vector of unknown parameters for the i th response of order $p_i \times 1$ ($i = 1, \dots, q$), where $\sum_{i=1}^q p_i = p$. For simplicity, we shall drop the use of two subscripts and instead refer to all the elements of $\boldsymbol{\beta}$ as $(\beta_1, \dots, \beta_p)'$, where the first p_1 elements come from $\boldsymbol{\beta}_1$, the next p_2 elements from $\boldsymbol{\beta}_2$, and so on.

- (3) A link function $\mathbf{g} : R^q \rightarrow R^q$, which relates the linear predictor $\boldsymbol{\eta}(\mathbf{x}) = [\eta_1(\mathbf{x}), \dots, \eta_q(\mathbf{x})]'$ to the mean response $\boldsymbol{\mu}(\mathbf{x}) = [\mu_1(\mathbf{x}), \dots, \mu_q(\mathbf{x})]'$ at a given point \mathbf{x} in a region of interest, R ,

$$\boldsymbol{\eta}(\mathbf{x}) = \mathbf{g}[\boldsymbol{\mu}(\mathbf{x})]. \quad (2.4)$$

If the inverse of \mathbf{g} , denoted by \mathbf{h} , exists, where $\mathbf{h} : R^q \rightarrow R^q$, then

$$\boldsymbol{\mu}(\mathbf{x}) = \mathbf{h}[\boldsymbol{\eta}(\mathbf{x})] = \mathbf{h}[\mathbf{Z}'(\mathbf{x})\boldsymbol{\beta}]. \tag{2.5}$$

2.1. Estimating the mean response

The maximum likelihood estimate (MLE) of $\boldsymbol{\beta}$ can be obtained by using an iterated weighted least-squares procedure (see Fahrmeir and Tutz, 2001, p. 106). An estimate of $\boldsymbol{\eta}(\mathbf{x})$ is then given by

$$\hat{\boldsymbol{\eta}}(\mathbf{x}) = \mathbf{Z}'(\mathbf{x})\hat{\boldsymbol{\beta}}, \quad j = 1, \dots, n, \tag{2.6}$$

and the corresponding estimate of the mean, $\boldsymbol{\mu}(\mathbf{x})$, is

$$\hat{\boldsymbol{\mu}}(\mathbf{x}) = \mathbf{h}[\mathbf{Z}'(\mathbf{x})\hat{\boldsymbol{\beta}}], \quad j = 1, \dots, n, \tag{2.7}$$

where $\hat{\boldsymbol{\beta}}$ is the MLE of $\boldsymbol{\beta}$.

2.2. Prediction variance

The variance–covariance matrix of $\hat{\boldsymbol{\beta}}$ is approximately given by (see Fahrmeir and Tutz, 2001, p. 106)

$$\text{Var}(\hat{\boldsymbol{\beta}}) \doteq (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}, \tag{2.8}$$

where $\mathbf{X} = [\mathbf{Z}(\mathbf{x}_1), \dots, \mathbf{Z}(\mathbf{x}_n)]'$ and \mathbf{W} is a block-diagonal matrix of the form, $\mathbf{W} = \text{diag}[\mathbf{W}_1, \dots, \mathbf{W}_n]$, $\mathbf{W}_j = \left(\frac{\partial \boldsymbol{\eta}_j}{\partial \boldsymbol{\mu}'_j} \boldsymbol{\Sigma}_j \frac{\partial \boldsymbol{\eta}_j}{\partial \boldsymbol{\mu}_j} \right)^{-1}$, $j = 1, \dots, n$. Here, $\boldsymbol{\eta}_j = \boldsymbol{\eta}(\mathbf{x}_j)$ ($j = 1, \dots, n$) is the linear predictor evaluated at $\mathbf{x}_j = (x_{j1}, \dots, x_{jk})'$, $\boldsymbol{\mu}_j = \boldsymbol{\mu}(\mathbf{x}_j)$ is the mean response at \mathbf{x}_j , and $\frac{\partial \boldsymbol{\eta}_j}{\partial \boldsymbol{\mu}'_j}$ is the first-order partial derivative matrix of $\boldsymbol{\eta}(\mathbf{x})$ with respect to $\boldsymbol{\mu}'(\mathbf{x})$ evaluated at \mathbf{x}_j ($j = 1, \dots, n$).

The distribution of the MLE, $\hat{\boldsymbol{\beta}}$, is asymptotically normal with mean $\boldsymbol{\beta}$ and a variance–covariance matrix $(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}$ (see Fahrmeir and Tutz, 2001, p. 106) for large n . Based on Wald's (1943) results, we approximately have

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'[\hat{\text{Var}}(\hat{\boldsymbol{\beta}})]^{-1}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim \chi_p^2, \tag{2.9}$$

where p is the total number of parameters and $\hat{\text{Var}}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\hat{\mathbf{W}}\mathbf{X})^{-1}$; $\hat{\mathbf{W}}$ is an estimate of \mathbf{W} using maximum likelihood estimation. Thus an approximate $100(1 - \alpha)\%$ confidence region for $\boldsymbol{\beta}$ is given by

$$\mathcal{C} = \{\boldsymbol{\gamma} : (\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma})'[(\mathbf{X}'\hat{\mathbf{W}}\mathbf{X})](\hat{\boldsymbol{\beta}} - \boldsymbol{\gamma}) \leq \chi_{\alpha, p}^2\}. \tag{2.10}$$

From (2.6) and (2.8) we approximately get

$$\text{Var}[\hat{\boldsymbol{\eta}}(\mathbf{x})] = \mathbf{Z}'(\mathbf{x})\text{Var}(\hat{\boldsymbol{\beta}})\mathbf{Z}(\mathbf{x}) \doteq \mathbf{Z}'(\mathbf{x})(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{Z}(\mathbf{x}). \tag{2.11}$$

A first-order multivariable Taylor series expansion of $\hat{\boldsymbol{\mu}}(\mathbf{x}) = \mathbf{h}[\hat{\boldsymbol{\eta}}(\mathbf{x})]$ around $\boldsymbol{\eta}(\mathbf{x})$ gives

$$\hat{\boldsymbol{\mu}}(\mathbf{x}) \doteq \mathbf{h}[\boldsymbol{\eta}(\mathbf{x})] + \frac{\partial \mathbf{h}[\boldsymbol{\eta}(\mathbf{x})]}{\partial \boldsymbol{\eta}'(\mathbf{x})}[\hat{\boldsymbol{\eta}}(\mathbf{x}) - \boldsymbol{\eta}(\mathbf{x})]. \tag{2.12}$$

Replacing $\mathbf{h}[\boldsymbol{\eta}(\mathbf{x})]$ with $\boldsymbol{\mu}(\mathbf{x})$ in the above expansion results in

$$\hat{\boldsymbol{\mu}}(\mathbf{x}) \doteq \boldsymbol{\mu}(\mathbf{x}) + \frac{\partial \boldsymbol{\mu}(\mathbf{x})}{\partial \boldsymbol{\eta}'(\mathbf{x})}[\hat{\boldsymbol{\eta}}(\mathbf{x}) - \boldsymbol{\eta}(\mathbf{x})]. \tag{2.13}$$

Thus, $\text{Var}[\hat{\boldsymbol{\mu}}(\mathbf{x})]$ is approximately given by

$$\text{Var}[\hat{\boldsymbol{\mu}}(\mathbf{x})] \doteq \frac{\partial \boldsymbol{\mu}(\mathbf{x})}{\partial \boldsymbol{\eta}'(\mathbf{x})} \text{Var}[\hat{\boldsymbol{\eta}}(\mathbf{x})] \frac{\partial \boldsymbol{\mu}'(\mathbf{x})}{\partial \boldsymbol{\eta}(\mathbf{x})}.$$

We refer to $\text{Var}[\hat{\mu}(\mathbf{x})]$ as the prediction variance matrix. Using the expression of $\text{Var}[\hat{\eta}(\mathbf{x})]$ from (2.11), $\text{Var}[\hat{\mu}(\mathbf{x})]$ can be approximately written as

$$\text{Var}[\hat{\mu}(\mathbf{x})] \doteq \frac{\partial \boldsymbol{\mu}(\mathbf{x})}{\partial \boldsymbol{\eta}'(\mathbf{x})} [\mathbf{Z}'(\mathbf{x})(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{Z}(\mathbf{x})] \frac{\partial \boldsymbol{\mu}'(\mathbf{x})}{\partial \boldsymbol{\eta}(\mathbf{x})}. \tag{2.14}$$

2.3. Prediction bias

In this section, we present an expression for the bias associated with estimating $\boldsymbol{\mu}(\mathbf{x})$, using maximum likelihood, in a general multivariate setting.

For a likelihood function with a single parameter to be estimated, the bias to order $\frac{1}{n}$ (n is the sample size) of its MLE was given by Bartlett (1953). A generalization of the bias formula for a likelihood function with several parameters was established by Cox and Snell (1968). Using Cox and Snell's (1968) results, Cordeiro and McCullagh (1991) developed a first-order approximation of the bias in the special case of GLMs. Mukhopadhyay and Khuri (2005) provide a further extension of the bias expression in case of multivariate GLMs where several response variables are considered. They derived the following expression for $\text{Bias}(\hat{\boldsymbol{\beta}})$ correct up to order $\frac{1}{n}$:

$$\begin{pmatrix} -\frac{\phi}{2} \sum_{j=1}^n \left[\sum_{s,t=1}^{p.} I^{ts} \left(\begin{pmatrix} \mathbf{f}'_1(\mathbf{x}_j) \frac{\partial \boldsymbol{\beta}_1}{\partial \beta_s} \\ \vdots \\ \mathbf{f}'_q(\mathbf{x}_j) \frac{\partial \boldsymbol{\beta}_q}{\partial \beta_s} \end{pmatrix} \mathbf{A}_{jt} - \begin{pmatrix} \mathbf{f}'_1(\mathbf{x}_j) \frac{\partial \boldsymbol{\beta}_1}{\partial \beta_t} \\ \vdots \\ \mathbf{f}'_q(\mathbf{x}_j) \frac{\partial \boldsymbol{\beta}_q}{\partial \beta_t} \end{pmatrix} \mathbf{B}_{js} \right) \right] \left[\sum_{r=1}^{p.} I^{1r} \begin{pmatrix} \mathbf{f}'_1(\mathbf{x}_j) \frac{\partial \boldsymbol{\beta}_1}{\partial \beta_r} \\ \vdots \\ \mathbf{f}'_q(\mathbf{x}_j) \frac{\partial \boldsymbol{\beta}_q}{\partial \beta_r} \end{pmatrix} \right] \\ -\frac{\phi}{2} \sum_{j=1}^n \left[\sum_{s,t=1}^{p.} I^{ts} \left(\begin{pmatrix} \mathbf{f}'_1(\mathbf{x}_j) \frac{\partial \boldsymbol{\beta}_1}{\partial \beta_s} \\ \vdots \\ \mathbf{f}'_q(\mathbf{x}_j) \frac{\partial \boldsymbol{\beta}_q}{\partial \beta_s} \end{pmatrix} \mathbf{A}_{jt} - \begin{pmatrix} \mathbf{f}'_1(\mathbf{x}_j) \frac{\partial \boldsymbol{\beta}_1}{\partial \beta_t} \\ \vdots \\ \mathbf{f}'_q(\mathbf{x}_j) \frac{\partial \boldsymbol{\beta}_q}{\partial \beta_t} \end{pmatrix} \mathbf{B}_{js} \right) \right] \left[\sum_{r=1}^{p.} I^{p.r} \begin{pmatrix} \mathbf{f}'_1(\mathbf{x}_j) \frac{\partial \boldsymbol{\beta}_1}{\partial \beta_r} \\ \vdots \\ \mathbf{f}'_q(\mathbf{x}_j) \frac{\partial \boldsymbol{\beta}_q}{\partial \beta_r} \end{pmatrix} \right] \end{pmatrix}, \tag{2.15}$$

where I^{uv} is the (u, v) th $(u, v = 1, \dots, p.)$ element of the inverse of Fisher information matrix,

$$I(\boldsymbol{\beta}) = -E \left[\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right].$$

Here, l is the log-likelihood function obtained using (2.1), that is,

$$l = \sum_{j=1}^n [\phi \{ \mathbf{y}'_j \boldsymbol{\theta}_j - b(\boldsymbol{\theta}_j) \} + c(\mathbf{y}_j, \phi)],$$

$\frac{\partial \boldsymbol{\beta}_i}{\partial \beta_d}$ ($i = 1, \dots, q; d = 1, \dots, p.)$ is the vector of first-order partial derivatives of $\boldsymbol{\beta}_i$ with respect to β_d . By \mathbf{A}_{jt} ($j = 1, \dots, n; t = 1, \dots, p.)$ and \mathbf{B}_{js} ($s = 1, \dots, p.)$ we denote the $q \times q$ matrices:

$$\mathbf{A}_{jt} = \left\{ \frac{\partial}{\partial \beta_t} \left(\frac{\partial \boldsymbol{\mu}'_j}{\partial \boldsymbol{\eta}_j} \right) \right\} \frac{\partial \boldsymbol{\theta}_j}{\partial \boldsymbol{\mu}'_j} \frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\eta}'_j} + \frac{\partial \boldsymbol{\mu}'_j}{\partial \boldsymbol{\eta}_j} \left\{ \frac{\partial}{\partial \beta_t} \left(\frac{\partial \boldsymbol{\theta}_j}{\partial \boldsymbol{\mu}'_j} \right) \right\} \frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\eta}'_j} + \frac{\partial \boldsymbol{\mu}'_j}{\partial \boldsymbol{\eta}_j} \frac{\partial \boldsymbol{\theta}_j}{\partial \boldsymbol{\mu}'_j} \left\{ \frac{\partial}{\partial \beta_t} \left(\frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\eta}'_j} \right) \right\},$$

$$\mathbf{B}_{js} = \frac{\partial \boldsymbol{\mu}'_j}{\partial \boldsymbol{\eta}_j} \left\{ \frac{\partial}{\partial \beta_s} \left(\frac{\partial \boldsymbol{\theta}_j}{\partial \boldsymbol{\mu}'_j} \right) \right\} \frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\eta}'_j} + \frac{\partial \boldsymbol{\mu}'_j}{\partial \boldsymbol{\eta}_j} \frac{\partial \boldsymbol{\theta}_j}{\partial \boldsymbol{\mu}'_j} \left\{ \frac{\partial}{\partial \beta_s} \left(\frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\eta}'_j} \right) \right\}.$$

In the above expressions, $\frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\eta}'_j}$ denotes the matrix of first-order partial derivative of $\boldsymbol{\mu}(\mathbf{x})$ with respect to $\boldsymbol{\eta}'(\mathbf{x})$ evaluated at \mathbf{x}_j ($j = 1, \dots, n$), $\frac{\partial \boldsymbol{\theta}_j}{\partial \boldsymbol{\mu}'_j}$ denotes the matrix of first-order partial derivative of $\boldsymbol{\theta}_j$ with respect to $\boldsymbol{\mu}'_j$, and $\frac{\partial}{\partial \beta_d} \left(\frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\eta}'_j} \right)$

and $\frac{\partial}{\partial \beta_d} \left(\frac{\partial \theta_j}{\partial \mu_j} \right)$ ($d = 1, \dots, p.$) are the first-order partial derivatives of $\frac{\partial \mu_j}{\partial \eta_j}$ and $\frac{\partial \theta_j}{\partial \mu_j}$, respectively, taken with respect to β_d .

Using (2.6) and (2.15) jointly, the bias of $\hat{\eta}(\mathbf{x})$ is given by

$$\text{Bias}[\hat{\eta}(\mathbf{x})] = \mathbf{Z}'(\mathbf{x}) \text{Bias}[\hat{\beta}]. \tag{2.16}$$

From the expansion of $\hat{\mu}(\mathbf{x}) = \mathbf{h}[\hat{\eta}(\mathbf{x})]$ in a neighborhood of $\eta(\mathbf{x})$ by a first-order approximation of Taylor series as in (2.13), we have

$$\hat{\mu}(\mathbf{x}) - \mu(\mathbf{x}) \doteq \frac{\partial \mu(\mathbf{x})}{\partial \eta'(\mathbf{x})} [\hat{\eta}(\mathbf{x}) - \eta(\mathbf{x})].$$

Taking expectation on both sides of the above equation, we get an approximate expression for the prediction bias, that is,

$$\text{Bias}[\hat{\mu}(\mathbf{x})] \doteq \frac{\partial \mu(\mathbf{x})}{\partial \eta'(\mathbf{x})} \text{Bias}[\hat{\eta}(\mathbf{x})] = \frac{\partial \mu(\mathbf{x})}{\partial \eta'(\mathbf{x})} \mathbf{Z}'(\mathbf{x}) \text{Bias}[\hat{\beta}]. \tag{2.17}$$

2.4. Mean-squared error of prediction

The mean-squared error of $\hat{\mu}(\mathbf{x})$, also known as the MSEP, is given by

$$\begin{aligned} \text{MSE}[\hat{\mu}(\mathbf{x})] &= E[(\hat{\mu}(\mathbf{x}) - \mu(\mathbf{x}))(\hat{\mu}(\mathbf{x}) - \mu(\mathbf{x}))'] \\ &= \text{Var}[\hat{\mu}(\mathbf{x})] + \{\text{Bias}[\hat{\mu}(\mathbf{x})]\}\{\text{Bias}[\hat{\mu}(\mathbf{x})]\}'. \end{aligned} \tag{2.18}$$

Using formulae (2.14) and (2.17) we get an approximate expression for the MSEP in the multivariate case.

3. Criterion for comparing designs

In the univariate case, when comparing two designs, say D_1 and D_2 , design D_1 is deemed better than D_2 , in terms of prediction capability, if the MSEP of D_1 is smaller than the MSEP of D_2 . In the multivariate case with several response variables, the MSEP is a matrix. A design comparison can then be based on a scalar-valued function of the MSEP. One possible choice is the largest eigenvalue of the MSEP matrix. Obviously, small values of such a function are desirable. Other possible choices include the trace and the determinant of the MSEP. In this paper, we consider the largest eigenvalue of the MSEP as a criterion for comparing designs for multivariate GLMs. Such a criterion function is denoted by EMSEP.

3.1. Design dependence on β

One major problem with the use of the MSEP is that it depends on β , the vector of unknown parameters of the fitted linear predictor model. We are then faced with a dilemma, since the purpose of a design is to estimate the unknown parameters using the data generated by the design in the experiment. To actually construct the best design, however, one needs to know the true values of the parameters. Other design optimality criteria such as A -, D -, E -optimality, which are variance-based, also suffer from the same problem. Common approaches to solving this *design dependence problem* were summarized in Section 1.1. In this paper, we extend the use of the *QDGs* approach, mentioned earlier in Section 1.2, to multivariate GLMs.

4. Quantile dispersion graphs

The scalar-valued function, EMSEP, in Section 3 is defined over a region of interest, R , which is a subset of the k -dimensional Euclidean space. Its value at a point \mathbf{x} in R depends on the chosen design, denoted by D , and the vector of unknown parameters, β . Let us therefore denote such a value by $\tau_D(\mathbf{x}, \beta)$. We assume that β belongs to a specified parameter space denoted by \mathcal{C} .

In order to study the effect of the design D on EMSEP, we examine the values of $\tau_D(\mathbf{x}, \boldsymbol{\beta})$ within the region R for a given $\boldsymbol{\beta}$ in \mathcal{C} . In particular, we consider its values on concentric surfaces, denoted by R_ν , which are obtained by reducing the boundary of R using a shrinkage factor, ν . Then, for a given design D and a given $\boldsymbol{\beta}$ in \mathcal{C} , the values of $\tau_D(\mathbf{x}, \boldsymbol{\beta})$, for all \mathbf{x} on R_ν , form a data set. Let $Q_D(p, \boldsymbol{\beta}, \nu)$ denote the p th quantile of such a data set, $0 \leq p \leq 1$. These quantiles provide a description of the distribution of $\tau_D(\mathbf{x}, \boldsymbol{\beta})$ for values of \mathbf{x} on R_ν . By varying ν we can cover the entire region R . Small values of $\tau_D(\mathbf{x}, \boldsymbol{\beta})$ throughout R are obviously desirable.

Recall that any design criterion for GLMs suffers from the problem of the dependence of the design on $\boldsymbol{\beta}$. To address this problem, we can vary $\boldsymbol{\beta}$ within the set \mathcal{C} and then examine the corresponding values of $Q_D(p, \boldsymbol{\beta}, \nu)$ for a given design D and given p and ν . The set \mathcal{C} is chosen to be the $(1 - \alpha)100\%$ confidence region on $\boldsymbol{\beta}$ given by formula (2.10). An “initial” data set that may be available on the responses can be used to construct such a confidence region. Thus the dependence of the quantiles of $\tau_D(\mathbf{x}, \boldsymbol{\beta})$ on $\boldsymbol{\beta}$, for \mathbf{x} on R_ν , can be investigated by computing $Q_D(p, \boldsymbol{\beta}, \nu)$ for several values of $\boldsymbol{\beta}$ that form a grid, C , of points randomly chosen from \mathcal{C} . The minimum and maximum values of $Q_D(p, \boldsymbol{\beta}, \nu)$ are subsequently obtained over the values of $\boldsymbol{\beta}$ in C , for fixed p and ν , and for a given design D . We can therefore compute the following functions:

$$Q_D^{\min}(p, \nu) = \min_{\boldsymbol{\beta} \in C} \{Q_D(p, \boldsymbol{\beta}, \nu)\}, \quad (4.1)$$

$$Q_D^{\max}(p, \nu) = \max_{\boldsymbol{\beta} \in C} \{Q_D(p, \boldsymbol{\beta}, \nu)\}. \quad (4.2)$$

For a fixed ν , plotting these values against p results in the QDGs of the EMSEP over the region R_ν . By repeating the same process for several selected values of ν we obtain plots that portray the prediction capability associated with the design D throughout the region R .

Small and close values of the minimum and maximum quantiles in (4.1) and (4.2) of a design D over the range of p ($0 \leq p \leq 1$) are highly desirable. Suppose that it is desired to compare two designs D_1 and D_2 . If the maximum quantile values of design D_1 for a given ν are smaller than those of D_2 , then design D_1 is said to have a better prediction capability than D_2 on R_ν . Also, the closeness of the minimum and maximum quantiles for a design D indicates that the design is robust to changes in the values of the parameter vector, $\boldsymbol{\beta}$.

4.1. Advantages of the QDGs

We list below some of the advantages of the QDGs approach over the other approaches mentioned in Section 1.1:

- (i) The performance of a design can be evaluated throughout the region R . Other design criteria base the evaluation of a design on a single measure, such as D -efficiency, but do not consider the quality of prediction throughout R .
- (ii) Unlike other variance-based design criteria, such as D -optimality or G -optimality, estimation bias is taken into account in the evaluation of a given design.
- (iii) The dependence of a design on the model's unknown parameters is clearly depicted by the QDGs throughout the region R .
- (iv) Plotting the quantiles of a scalar-valued function of the MSEP against p easily permits the consideration of models with several control variables. By contrast, the construction of, for example, locally D -optimal designs for GLMs with two or more control variables is “difficult and cumbersome” (see Atkinson and Haines, 1996, pp. 461–467; Myers, 1999, p. 40). The majority of research in the design area has therefore been confined to the case of a single control variable (see Myers et al., 2002, p. 239).

5. Bivariate binary distribution

In many experimental situations, several responses may be observed for the same subject. For example, in a drug testing experiment, in addition to the standard binary response of success or failure of the drug, some measure of the side effects of the drug may be of interest. This results in two responses, efficacy and toxicity of the drug. The efficacy response is 1 if the drug used has the desired therapeutic effect; toxicity response is 1 if the drug causes unwanted side effects. These responses are often modelled separately, as in Peruca and Pisani (1989), with the assumption that they are uncorrelated. However, since the two responses come from the same subject, it is prudent to allow for correlation in the responses, that is, to consider them as a bivariate response.

5.1. Bivariate binary distribution with a logit link

Consider a bivariate binary response situation in which m_j subjects, or experimental units, are tested at the j th run (j th level of \mathbf{x}), $j = 1, \dots, n$. The measurement taken from the w th subject at the j th run for response l is denoted by y_{jwl} ($j = 1, \dots, n$; $w = 1, \dots, m_j$; $l = 1, 2$). The responses y_{jwl} are 1 or 0 depending on whether a particular outcome, labelled as “success”, is observed or not. Let $\mathbf{y}_{jw} = (y_{jw1}, y_{jw2})'$ denote the vector of bivariate responses from the w th subject at the j th experimental run. If \mathbf{y}_{jw} , for $w = 1, \dots, m_j$, are independent then the p.m.f. (probability mass function) of $\mathbf{y}_j = (\mathbf{y}'_{j1}, \dots, \mathbf{y}'_{jm_j})'$ is given by

$$\delta_j(\mathbf{y}_j) \propto \pi_{j1}^{\sum_{w=1}^{m_j} y_{jw1} y_{jw2}} \pi_{j2}^{\sum_{w=1}^{m_j} y_{jw1} (1 - y_{jw2})} \pi_{j3}^{\sum_{w=1}^{m_j} (1 - y_{jw1}) y_{jw2}} \pi_{j4}^{\sum_{w=1}^{m_j} (1 - y_{jw1}) (1 - y_{jw2})}, \tag{5.1}$$

where the four probabilities $\pi_{j1}, \pi_{j2}, \pi_{j3}$ and π_{j4} for the j th level of \mathbf{x} can be expressed as $\pi_{j1} = P(y_{jw1} = 1, y_{jw2} = 1)$, $\pi_{j2} = P(y_{jw1} = 1, y_{jw2} = 0)$, $\pi_{j3} = P(y_{jw1} = 0, y_{jw2} = 1)$, $\pi_{j4} = P(y_{jw1} = 0, y_{jw2} = 0)$ (these probabilities are the same for all $w = 1, \dots, m_j$). Note that $\pi_{j1} + \pi_{j2} + \pi_{j3} + \pi_{j4} = 1, j = 1, \dots, n$.

Following a suggestion by Cox (1972, p. 115) and McCullagh and Nelder (1989, p. 226), we can write (5.1) as

$$\delta_j(\mathbf{z}_j) \propto \pi_{j1}^{z_{j1}} \pi_{j2}^{z_{j2}} \pi_{j3}^{z_{j3}} \left[1 - \sum_{i=1}^3 \pi_{ji} \right]^{(m_j - \sum_{i=1}^3 z_{ji})}, \quad j = 1, \dots, n, \tag{5.2}$$

where $z_{j1} = \sum_{w=1}^{m_j} y_{jw1} y_{jw2}$, $z_{j2} = \sum_{w=1}^{m_j} y_{jw1} (1 - y_{jw2})$, $z_{j3} = \sum_{w=1}^{m_j} (1 - y_{jw1}) y_{jw2}$ and $m_j - \sum_{i=1}^3 z_{ji} = \sum_{w=1}^{m_j} (1 - y_{jw1}) (1 - y_{jw2})$. We note from the form of the p.m.f in (5.2) (see Casella and Berger, 2002, p. 180; Fahrmeir and Tutz, 2001, p. 70, formula 3.12) that the response $\mathbf{z}_j = (z_{j1}, z_{j2}, z_{j3})'$ follows the multinomial distribution with parameters m_j and $\boldsymbol{\pi}_j = (\pi_{j1}, \pi_{j2}, \pi_{j3})'$, $j = 1, \dots, n$.

Let us write (5.2) as

$$\delta_j(\mathbf{z}_j) \propto \exp \left[z_{j1} \log \pi_{j1} + z_{j2} \log \pi_{j2} + z_{j3} \log \pi_{j3} + \left(m_j - \sum_{i=1}^3 z_{ji} \right) \log \left(1 - \sum_{i=1}^3 \pi_{ji} \right) \right]. \tag{5.3}$$

Comparing (5.3) with (2.1) we note that $\phi = 1$, $\theta_{ji} = \log \left(\frac{\pi_{ji}}{1 - \sum_{i=1}^3 \pi_{ji}} \right)$ is the i th element of $\boldsymbol{\theta}_j$ ($i = 1, 2, 3$), $b(\boldsymbol{\theta}_j) = -m_j \log(1 - \sum_{i=1}^3 \pi_{ji})$, and $c(\mathbf{y}_j, \phi) = 0$. The mean of \mathbf{z}_j is $\boldsymbol{\mu}_j = (\mu_{j1}, \mu_{j2}, \mu_{j3})'$, where $\mu_{j1} = m_j \pi_{j1}$, $\mu_{j2} = m_j \pi_{j2}$, $\mu_{j3} = m_j \pi_{j3}$, and the variance–covariance matrix of \mathbf{z}_j is $\boldsymbol{\Sigma}_j$, where

$$\boldsymbol{\Sigma}_j = m_j \begin{pmatrix} \pi_{j1}(1 - \pi_{j1}) & -\pi_{j1}\pi_{j2} & -\pi_{j1}\pi_{j3} \\ -\pi_{j2}\pi_{j1} & \pi_{j2}(1 - \pi_{j2}) & -\pi_{j2}\pi_{j3} \\ -\pi_{j3}\pi_{j1} & -\pi_{j3}\pi_{j2} & \pi_{j3}(1 - \pi_{j3}) \end{pmatrix}, \quad j = 1, \dots, n. \tag{5.4}$$

Let \mathbf{x} be any point in the region of interest, R . The mean response at \mathbf{x} will be denoted by $\boldsymbol{\pi}(\mathbf{x})$, where $\boldsymbol{\pi}(\mathbf{x}) = [\pi_1(\mathbf{x}), \pi_2(\mathbf{x}), \pi_3(\mathbf{x})]'$ with $\pi_1(\mathbf{x}), \pi_2(\mathbf{x})$ and $\pi_3(\mathbf{x})$ being probabilities similar to the elements of $\boldsymbol{\pi}_j$, but are evaluated at \mathbf{x} . Hence, $\boldsymbol{\mu}(\mathbf{x}) = \boldsymbol{\pi}(\mathbf{x})$. Note, however, that at the j th run, $\boldsymbol{\mu}(\mathbf{x}_j) = \boldsymbol{\mu}_j$, since $\boldsymbol{\pi}(\mathbf{x}_j) = \boldsymbol{\pi}_j$ ($j = 1, \dots, n$). Our objective here is to find a form for \mathbf{g} , the link function, such that

$$\boldsymbol{\eta}(\mathbf{x}) = \mathbf{g}[\boldsymbol{\mu}(\mathbf{x})].$$

From Section 2, $\mathbf{x} = (x_1, \dots, x_k)'$ is a vector of control variables, $\mathbf{Z}'(\mathbf{x}) = \bigoplus_{i=1}^3 \mathbf{f}'_i(\mathbf{x})$, $\mathbf{f}_i(\mathbf{x})$ is a known vector function, and $\boldsymbol{\beta} = [\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2, \boldsymbol{\beta}'_3]'$ is a p -dimensional vector of unknown parameters. The $\boldsymbol{\beta}_i$'s ($i = 1, 2, 3$) correspond to the three responses (outcomes in the multinomial distribution) and are of order $p_i \times 1$, where $\sum_{i=1}^3 p_i = p$.

The corresponding link function used here is (see Fahrmeir and Tutz, 2001, p. 73; Agresti, 2002, pp. 267–274)

$$\eta_i(\mathbf{x}) = \log \left(\frac{\pi_i(\mathbf{x})}{1 - \sum_{l=1}^3 \pi_l(\mathbf{x})} \right) = \mathbf{f}'_i(\mathbf{x}) \boldsymbol{\beta}_i, \quad i = 1, 2, 3. \tag{5.5}$$

Note that $\eta_i(\mathbf{x}_j) = \eta_{ji} = \theta_{ji}$ ($i = 1, 2, 3; j = 1, \dots, n$).

Hence,

$$\hat{\pi}_i(\mathbf{x}) = \frac{\exp[\mathbf{f}'_i(\mathbf{x})\hat{\beta}_i]}{1 + \sum_{l=1}^3 \exp[\mathbf{f}'_l(\mathbf{x})\hat{\beta}_l]}, \quad i = 1, 2, 3,$$

where $\hat{\beta}_i$ is the MLE of β_i for $i = 1, 2, 3$.

Using formulae (2.14) and (2.17), it can be shown (see Appendix A) that the MSEF at \mathbf{x} is approximately given by

$$\text{MSE}[\hat{\mu}(\mathbf{x})] \doteq \frac{\partial \mu(\mathbf{x})}{\partial \boldsymbol{\eta}'(\mathbf{x})} [\mathbf{Z}'(\mathbf{x})(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{Z}(\mathbf{x})] \frac{\partial \mu'(\mathbf{x})}{\partial \boldsymbol{\eta}'(\mathbf{x})} + \left[\frac{\partial \mu(\mathbf{x})}{\partial \boldsymbol{\eta}'(\mathbf{x})} \mathbf{Z}'(\mathbf{x}) \text{Bias}(\hat{\boldsymbol{\beta}}) \right] \left[\frac{\partial \mu(\mathbf{x})}{\partial \boldsymbol{\eta}'(\mathbf{x})} \mathbf{Z}'(\mathbf{x}) \text{Bias}(\hat{\boldsymbol{\beta}}) \right]', \quad (5.6)$$

where \mathbf{X} , \mathbf{Z} , and $\text{Bias}(\hat{\boldsymbol{\beta}})$ are as defined in Section 2, and the first-order partial derivative of μ with respect to $\boldsymbol{\eta}'$ evaluated at \mathbf{x} is

$$\frac{\partial \mu(\mathbf{x})}{\partial \boldsymbol{\eta}'(\mathbf{x})} = \begin{pmatrix} \pi_1(\mathbf{x})[1 - \pi_1(\mathbf{x})] & -\pi_1(\mathbf{x})\pi_2(\mathbf{x}) & -\pi_1(\mathbf{x})\pi_3(\mathbf{x}) \\ -\pi_2(\mathbf{x})\pi_1(\mathbf{x}) & \pi_2(\mathbf{x})[1 - \pi_2(\mathbf{x})] & -\pi_2(\mathbf{x})\pi_3(\mathbf{x}) \\ -\pi_3(\mathbf{x})\pi_1(\mathbf{x}) & -\pi_3(\mathbf{x})\pi_2(\mathbf{x}) & \pi_3(\mathbf{x})[1 - \pi_3(\mathbf{x})] \end{pmatrix},$$

and $\mathbf{W} = \text{diag}[\mathbf{W}_1, \dots, \mathbf{W}_n]$ with $\mathbf{W}_j = \left(\frac{\partial \eta_j}{\partial \mu'_j} \boldsymbol{\Sigma}_j \frac{\partial \eta'_j}{\partial \mu_j} \right)^{-1}$, $j = 1, \dots, n$, and

$$\frac{\partial \eta_j}{\partial \mu'_j} = \frac{1}{m_j} \begin{pmatrix} \frac{(\pi_{j1} + \pi_{j4})}{\pi_{j1}\pi_{j4}} & \frac{1}{\pi_{j4}} & \frac{1}{\pi_{j4}} \\ \frac{1}{\pi_{j4}} & \frac{(\pi_{j2} + \pi_{j4})}{\pi_{j2}\pi_{j4}} & \frac{1}{\pi_{j4}} \\ \frac{1}{\pi_{j4}} & \frac{1}{\pi_{j4}} & \frac{(\pi_{j3} + \pi_{j4})}{\pi_{j3}\pi_{j4}} \end{pmatrix}.$$

Note that

$$\pi_{ji} = \frac{\exp(\eta_{ji})}{1 + \sum_{l=1}^3 \exp(\eta_{jl})} = \frac{\exp[\mathbf{f}'_i(\mathbf{x}_j)\beta_i]}{1 + \sum_{l=1}^3 \exp[\mathbf{f}'_l(\mathbf{x}_j)\beta_l]}, \quad i = 1, 2, 3; \quad j = 1, \dots, n.$$

It should also be noted here that the MSEF is a 3×3 matrix and depends on $\boldsymbol{\beta}$.

6. Numerical example

The data set considered here is taken from Gennings et al. (1994, pp. 429–451). In a combination drug therapy study on male mice, the pain relieving (analgesic) ability of two drugs, namely, Δ^9 -tetrahydro-cannabinol (Δ^9 -THC) and morphine sulfate is studied. Though both drugs are analgesic (i.e. provide pain relief), they are also associated with adverse side-effects. The two responses are y_1 , pain relief, and y_2 , side-effect. The response y_1 takes the value 1 if a mouse takes more than 8 s to flick its tail when placed under a heat lamp, and $y_1 = 0$, otherwise. Response $y_1 = 1$ is considered good, because the mouse does not feel pain when placed under a heat lamp for at least 8 s due to the pain relieving (analgesic) ability of the drugs. The side effect response, y_2 , was determined by recording the rectal temperature of the mouse after 60 min following drug administration. This response is equal to 1 when the rectal temperature of the mouse drops below 35°C (resulting in hypothermia) after the drug administration, and is equal to zero, otherwise. Hypothermia ($y_2 = 1$) is a toxic or harmful side-effect. Thus each of the response variables has two levels and all four combinations are possible. One aim of the investigation was to study how the associated probabilities concerning pain relief and hypothermia are related to dose levels of the two drugs.

For the pain relief and side-effect responses, 35 groups of mice (six animals per group) from a 5×7 factorial experiment were randomly assigned to receive the treatments, where a treatment consists of a single injection of one of the possible combinations of morphine sulfate (0, 2, 4, 6, 8 mg/kg) and Δ^9 -THC (0, 0.5, 1.0, 2.5, 5.0, 10.0, 15.0 mg/kg). Thus we have 35 runs with six experimental units (mice) in each run. The design and the resulting data are presented in Table 1.

Table 1
Experimental design D_1 (5×7 factorial) and response values

Morphine sulfate x_1	Δ^9 -THC x_2	Pain relief		No pain relief		m
		Side-effect	No side effect	Side-effect	No side-effect	
0	0	0	0	0	6	6
0	0.5	0	3	0	3	6
0	1	0	0	0	6	6
0	2.5	0	1	0	5	6
0	5	1	4	0	1	6
0	10	5	0	1	0	6
0	15	2	4	0	0	6
2	0	0	2	0	4	6
2	0.5	0	5	0	1	6
2	1	0	2	0	4	6
2	2.5	0	2	0	4	6
2	5	1	4	0	1	6
2	10	1	5	0	0	6
2	15	3	3	0	0	6
4	0	0	3	0	3	6
4	0.5	0	5	0	1	6
4	1	0	6	0	0	6
4	2.5	1	5	0	0	6
4	5	2	3	1	0	6
4	10	3	3	0	0	6
4	15	3	3	0	0	6
6	0	0	5	0	1	6
6	0.5	0	6	0	0	6
6	1	0	6	0	0	6
6	2.5	0	6	0	0	6
6	5	0	6	0	0	6
6	10	5	1	0	0	6
6	15	6	0	0	0	6
8	0	0	6	0	0	6
8	0.5	0	6	0	0	6
8	1	0	6	0	0	6
8	2.5	0	6	0	0	6
8	5	0	6	0	0	6
8	10	1	5	0	0	6
8	15	2	4	0	0	6

Model (2.3) was fitted to the data in Table 1 using the following first-degree models for the linear predictors

$$\begin{aligned}
 \eta_1(\mathbf{x}) &= \beta_1 + \beta_2x_1 + \beta_3x_2, \\
 \eta_2(\mathbf{x}) &= \beta_4 + \beta_5x_1 + \beta_6x_2, \\
 \eta_3(\mathbf{x}) &= \beta_7 + \beta_8x_1 + \beta_9x_3.
 \end{aligned}
 \tag{6.1}$$

Here, x_1 and x_2 represent the dose levels of the drugs, morphine sulfate and Δ^9 -THC, respectively. The link functions are the same as those described in formula (5.5).

The experimental region, R , is rectangular in shape with $R : \{0 \leq x_1 \leq 8, 0 \leq x_2 \leq 15\}$. The parameter estimates and their standard errors for the above models, obtained by using PROC LOGISTIC in SAS (SAS Institute Inc., 2000), are shown in Table 2.

Let us now refer to the design in Table 1 as D_1 . We compare this design with another design, D_2 . The additional design differs in design settings from D_1 but has the same number of experimental runs (=35) and the same number of mice per run (=6). Design D_2 is a 3^2 factorial with the center point (4, 7.5) replicated three times and all other points replicated four times. The listing of the two designs is given in Table 3. Fig. 1 shows the design points of D_1 and D_2 .

Table 2
Maximum likelihood estimates and standard errors for the parameters in models (6.1)

Parameter	Estimate	Std. error	P-value
β_1	-5.2899	0.7767	<0.0001
β_2	0.7346	0.1651	<0.0001
β_3	0.8355	0.1593	<0.0001
β_4	-2.3166	0.5232	<0.0001
β_5	0.8349	0.1457	<0.0001
β_6	0.5652	0.1523	0.0002
β_7	-5.6437	1.5662	0.0003
β_8	0.4188	0.3512	0.2331
β_9	0.6619	0.2054	0.0013

Scaled deviance = 78.0672, DF = 96.

Table 3
Designs D_1 (5×7 factorial) and D_2 (3^2 factorial)

D_1				D_2			
x_1	x_2	x_1	x_2	x_1	x_2	x_1	x_2
0	0	4	10	0	0	8	15
0	0.5	4	15	0	0	4	15
0	1	6	0	0	0	4	15
0	2.5	6	0.5	0	0	4	15
0	5	6	1	4	0	4	15
0	10	6	2.5	4	0	0	15
0	15	6	5	4	0	0	15
2	0	6	10	4	0	0	15
2	0.5	6	15	8	0	0	15
2	1	8	0	8	0	0	7.5
2	2.5	8	0.5	8	0	0	7.5
2	5	8	1	8	0	0	7.5
2	10	8	2.5	8	7.5	0	7.5
2	15	8	5	8	7.5	4	7.5
4	0	8	10	8	7.5	4	7.5
4	0.5	8	15	8	7.5	4	7.5
4	1			8	15		
4	2.5			8	15		
4	5			8	15		

For each design in Table 3 we study the distribution of the EMSEP on each of several concentric rectangles, R_ν , which are obtained by a reduction of the boundary of R using a shrinkage factor ν , $0.5 < \nu \leq 1$. Thus, R_ν is determined by the inequalities

$$a_i + (1 - \nu)(b_i - a_i) \leq x_i \leq b_i - (1 - \nu)(b_i - a_i), \quad i = 1, 2,$$

where a_i and b_i are the bounds on x_i in R ($i = 1, 2$), that is, $a_1 = 0$, $a_2 = 0$, $b_1 = 8$, $b_2 = 15$. Four or five values of ν are typically used in the calculation of the quantiles.

To investigate the dependence of EMSEP on β we consider \mathcal{C} to be the 95% confidence region on β (see formula (2.10)) and C to be a set of 500 points randomly chosen from \mathcal{C} . For each design and a selected value of β in C , quantiles of the distribution of the EMSEP are obtained for $\mathbf{x} \in R_\nu$. The number of points chosen on each R_ν was 500, consisting of 125 points on each side. The quantiles are calculated for $p = 0(0.05)1$ (p from 0 to 1 in steps of 0.05). The procedure is repeated for all the values of β in C . Then, $Q_D^{\min}(p)$ and $Q_D^{\max}(p)$ are calculated using formulae (4.1) and (4.2). The R software (Version 2.0.1) was used to carry out the numerical investigations and draw the QDGs.

To compare the two designs, we examine the corresponding QDGs shown in Fig. 2. For $\nu = 1$, we note that the

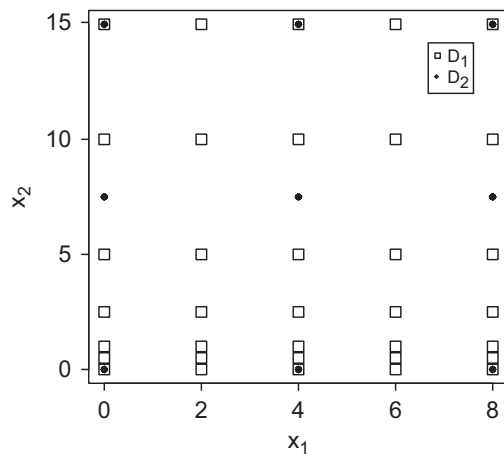


Fig. 1. Designs D_1 (5×7 factorial) and D_2 (3^2 factorial).

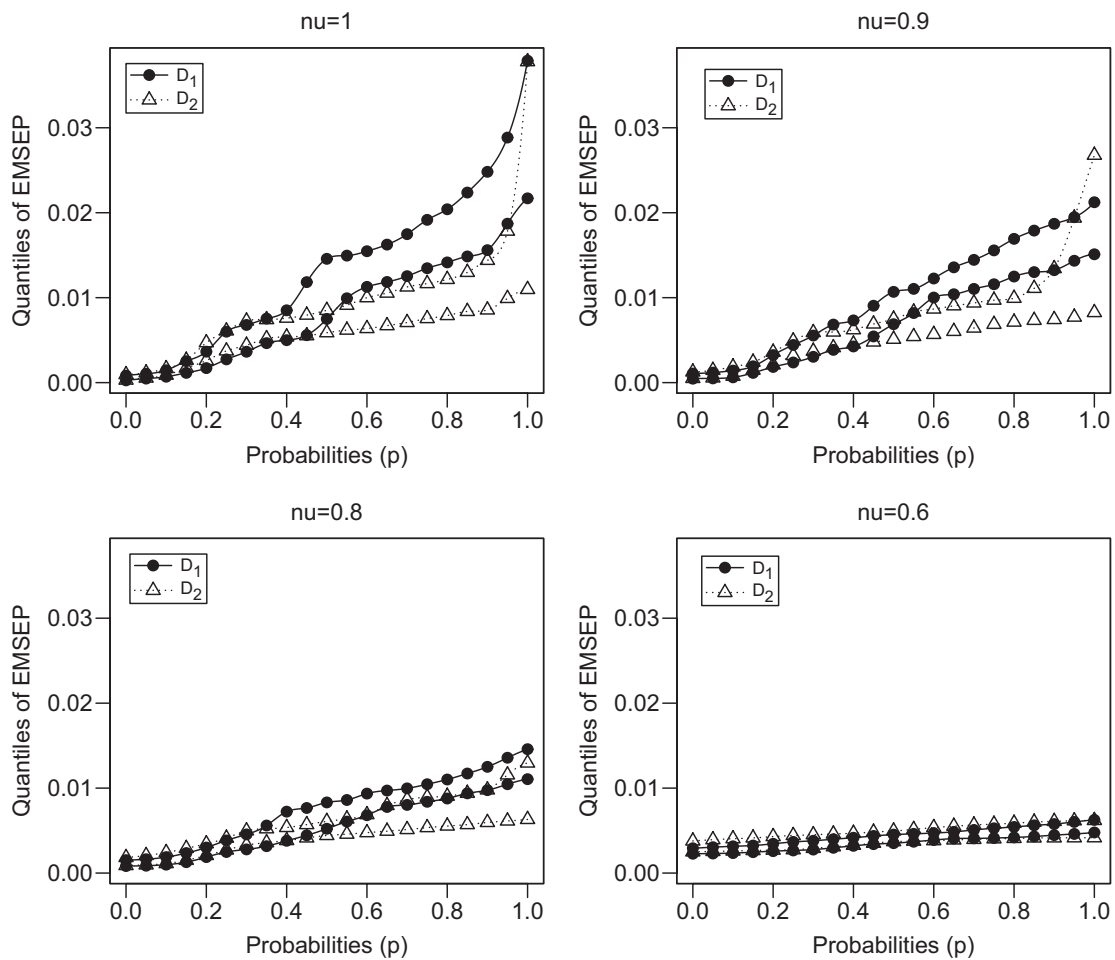


Fig. 2. Comparison of the QDGs for designs D_1 (5×7 factorial) and D_2 (3^2 factorial) given in Table 3, for $p = 0(0.05)1$ and $\nu = 1, 0.9, 0.8, 0.6$.

maximum quantiles of D_1 are above those of D_2 for most values of p , indicating that D_2 has better prediction capability than D_1 . The values of $Q_D^{\max}(p, \nu)$ and $Q_D^{\min}(p, \nu)$ for both designs are close to each other for most of the values of p indicating that they are robust to changes in the parameter values. As ν decreases (that is, as we get closer to the center of R), there is a decrease in the distance between the maximum quantiles of D_1 and D_2 indicating that they

have similar prediction capabilities near the center of R . So, overall, D_2 appears to be the better design, in terms of prediction capability, for almost all values of p and v .

7. Concluding remarks

The proposed graphical procedure for comparing designs for multivariate GLMs can be used in a variety of experimental situations. For example, in the case of the bivariate binary distribution, designs can be compared using different settings of the control variables and/or different numbers of experimental units for the various runs. Furthermore, we can consider situations involving more than one control variable. In this case, quantiles of the EMSEP can be considered on concentric surfaces within the region of interest, as was demonstrated in Section 6.

Appendix A. Derivation of formula (5.6)

In Section 5.1, $\boldsymbol{\mu}(\mathbf{x}) = \boldsymbol{\pi}(\mathbf{x})$, $\phi = 1$,

$$\theta_{ji} = \log \left(\frac{\pi_{ji}}{1 - \sum_{l=1}^3 \pi_{jl}} \right), \quad i = 1, 2, 3; \quad j = 1, \dots, n, \tag{A.1}$$

and

$$\eta_{ji} = \log \left(\frac{\pi_{ji}}{1 - \sum_{l=1}^3 \pi_{jl}} \right), \quad i = 1, 2, 3; \quad j = 1, \dots, n. \tag{A.2}$$

Note that $\boldsymbol{\theta}_j = \boldsymbol{\eta}_j$ ($j = 1, \dots, n$), where $\boldsymbol{\theta}_j = (\theta_{j1}, \theta_{j2}, \theta_{j3})'$ and $\boldsymbol{\eta}_j = (\eta_{j1}, \eta_{j2}, \eta_{j3})'$, thus $\frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\eta}'_j} = \frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\theta}'_j}$. But, $\boldsymbol{\Sigma}_j = \frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\theta}'_j}$, hence $\frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\eta}'_j} = \boldsymbol{\Sigma}_j$. Using the expression in (5.4), we can write

$$\frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\eta}'_j} = m_j \begin{pmatrix} \pi_{j1}(1 - \pi_{j1}) & -\pi_{j1}\pi_{j2} & -\pi_{j1}\pi_{j3} \\ -\pi_{j2}\pi_{j1} & \pi_{j2}(1 - \pi_{j2}) & -\pi_{j2}\pi_{j3} \\ -\pi_{j3}\pi_{j1} & -\pi_{j3}\pi_{j2} & \pi_{j3}(1 - \pi_{j3}) \end{pmatrix}, \quad j = 1, \dots, n. \tag{A.3}$$

Recall from Section 2.2 that $\mathbf{W} = \text{diag}[\mathbf{W}_1, \dots, \mathbf{W}_n]$, where $\mathbf{W}_j = \left(\frac{\partial \boldsymbol{\eta}_j}{\partial \boldsymbol{\mu}'_j} \boldsymbol{\Sigma}_j \frac{\partial \boldsymbol{\eta}'_j}{\partial \boldsymbol{\mu}_j} \right)^{-1}$, $j = 1, \dots, n$. Taking into consideration the well-known property of Jacobian matrices, $\left[\frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\eta}'_j} \right] \left[\frac{\partial \boldsymbol{\eta}_j}{\partial \boldsymbol{\mu}'_j} \right] = \mathbf{I}$, where \mathbf{I} is the 3×3 identity matrix, we can find $\frac{\partial \boldsymbol{\eta}_j}{\partial \boldsymbol{\mu}'_j}$ by simply inverting $\frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\eta}'_j}$ in (A.3). Thus,

$$\frac{\partial \boldsymbol{\eta}_j}{\partial \boldsymbol{\mu}'_j} = \frac{1}{m_j} \begin{pmatrix} \frac{(\pi_{j1} + \pi_{j4})}{\pi_{j1}\pi_{j4}} & \frac{1}{\pi_{j4}} & \frac{1}{\pi_{j4}} \\ \frac{1}{\pi_{j4}} & \frac{(\pi_{j2} + \pi_{j4})}{\pi_{j2}\pi_{j4}} & \frac{1}{\pi_{j4}} \\ \frac{1}{\pi_{j4}} & \frac{1}{\pi_{j4}} & \frac{(\pi_{j3} + \pi_{j4})}{\pi_{j3}\pi_{j4}} \end{pmatrix}, \quad j = 1, \dots, n. \tag{A.4}$$

Using $\boldsymbol{\Sigma}_j$ from formula (5.4) and $\frac{\partial \boldsymbol{\eta}_j}{\partial \boldsymbol{\mu}'_j}$ from (A.4), we get,

$$\mathbf{W}_j = m_j \begin{pmatrix} \pi_{j1}(1 - \pi_{j1}) & -\pi_{j1}\pi_{j2} & -\pi_{j1}\pi_{j3} \\ -\pi_{j2}\pi_{j1} & \pi_{j2}(1 - \pi_{j2}) & -\pi_{j2}\pi_{j3} \\ -\pi_{j3}\pi_{j1} & -\pi_{j3}\pi_{j2} & \pi_{j3}(1 - \pi_{j3}) \end{pmatrix}, \quad j = 1, \dots, n. \tag{A.5}$$

Now, the MSEP at \mathbf{x} is

$$\text{MSE}[\hat{\boldsymbol{\mu}}(\mathbf{x})] = \text{Var}[\hat{\boldsymbol{\mu}}(\mathbf{x})] + \{\text{Bias}[\hat{\boldsymbol{\mu}}(\mathbf{x})]\}\{\text{Bias}[\hat{\boldsymbol{\mu}}(\mathbf{x})]\}'. \tag{A.6}$$

From (2.14) and (A.5) we have

$$\text{Var}[\hat{\boldsymbol{\mu}}(\mathbf{x})] \doteq \frac{\partial \boldsymbol{\mu}(\mathbf{x})}{\partial \boldsymbol{\eta}'(\mathbf{x})} [\mathbf{Z}'(\mathbf{x})(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{Z}(\mathbf{x})] \frac{\partial \boldsymbol{\mu}'(\mathbf{x})}{\partial \boldsymbol{\eta}(\mathbf{x})}, \tag{A.7}$$

where from (A.3) $\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\eta}}$ evaluated at \mathbf{x} is given by

$$\frac{\partial \boldsymbol{\mu}(\mathbf{x})}{\partial \boldsymbol{\eta}'(\mathbf{x})} = \begin{pmatrix} \pi_1(\mathbf{x})[1 - \pi_1(\mathbf{x})] & -\pi_1(\mathbf{x})\pi_2(\mathbf{x}) & -\pi_1(\mathbf{x})\pi_3(\mathbf{x}) \\ -\pi_2(\mathbf{x})\pi_1(\mathbf{x}) & \pi_2(\mathbf{x})[1 - \pi_2(\mathbf{x})] & -\pi_2(\mathbf{x})\pi_3(\mathbf{x}) \\ -\pi_3(\mathbf{x})\pi_1(\mathbf{x}) & -\pi_3(\mathbf{x})\pi_2(\mathbf{x}) & \pi_3(\mathbf{x})[1 - \pi_3(\mathbf{x})] \end{pmatrix}. \tag{A.8}$$

The next step is to determine the different components of Bias($\hat{\boldsymbol{\beta}}$), as given in formula (2.15), for a bivariate binary distribution. We already have the matrices $\frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\eta}'_j}$ and $\frac{\partial \boldsymbol{\eta}_j}{\partial \boldsymbol{\mu}'_j}$. Using $\boldsymbol{\eta}_j = \boldsymbol{\theta}_j$, we get $\frac{\partial \boldsymbol{\theta}_j}{\partial \boldsymbol{\mu}'_j}$.

The only unknown components now left in the bias formula in (2.15) are $\frac{\partial}{\partial \beta_d} \left(\frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\eta}'_j} \right)$ and $\frac{\partial}{\partial \beta_d} \left(\frac{\partial \boldsymbol{\theta}_j}{\partial \boldsymbol{\mu}'_j} \right)$ ($d=1, \dots, p$), the first-order partial derivatives of $\frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\eta}'_j}$ and $\frac{\partial \boldsymbol{\theta}_j}{\partial \boldsymbol{\mu}'_j}$ taken with respect to β_d , respectively. Starting with $\frac{\partial}{\partial \beta_d} \left(\frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\eta}'_j} \right)$, each element of $\frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\eta}'_j}$ is differentiated with respect to β_d ($d=1, \dots, p$). Suppose that the (i, k) th element of $\frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\eta}'_j}$ is denoted by c_{jik} , we can then proceed as follows:

$$\frac{\partial c_{jik}}{\partial \beta_d} = \frac{\partial c_{jik}}{\partial \boldsymbol{\mu}'_j} \frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\eta}'_j} \frac{\partial \boldsymbol{\eta}_j}{\partial \beta_d}. \tag{A.9}$$

A similar technique is used to compute $\frac{\partial}{\partial \beta_d} \left(\frac{\partial \boldsymbol{\theta}_j}{\partial \boldsymbol{\mu}'_j} \right)$ ($d=1, \dots, p$).

To complete our task, we put the expressions for $\frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\eta}'_j}$, $\frac{\partial \boldsymbol{\theta}_j}{\partial \boldsymbol{\mu}'_j}$, $\frac{\partial}{\partial \beta_d} \left(\frac{\partial \boldsymbol{\mu}_j}{\partial \boldsymbol{\eta}'_j} \right)$, and $\frac{\partial}{\partial \beta_d} \left(\frac{\partial \boldsymbol{\theta}_j}{\partial \boldsymbol{\mu}'_j} \right)$ ($d=1, \dots, p$) in formula (2.15). This gives an approximate expression for Bias($\hat{\boldsymbol{\beta}}$) for the bivariate binary distribution. From (2.17) we have

$$\text{Bias}[\hat{\boldsymbol{\mu}}(\mathbf{x})] \doteq \frac{\partial \boldsymbol{\mu}(\mathbf{x})}{\partial \boldsymbol{\eta}'(\mathbf{x})} \mathbf{Z}'(\mathbf{x}) \text{Bias}[\hat{\boldsymbol{\beta}}]. \tag{A.10}$$

Using the expression for $\frac{\partial \boldsymbol{\mu}(\mathbf{x})}{\partial \boldsymbol{\eta}'(\mathbf{x})}$ in (A.8), and the one for Bias($\hat{\boldsymbol{\beta}}$), we get the prediction bias.

Substituting (A.7) and (A.10) in (A.6), we obtain formula (5.6).

References

Agresti, A., 2002. *Categorical Data Analysis*. second ed. Wiley, New York.

Atkinson, A.C., Haines, L.M., 1996. Designs for nonlinear and generalized linear models. In: Ghosh, S., Rao, C.R. (Eds.), *Handbook of Statistics*, vol. 13. Elsevier Science B.V., Amsterdam, pp. 437–475.

Atkinson, A.C., Chaloner, K., Herzberg, A.M., Juritz, J., 1993. Optimum experimental designs for properties of a compartmental model. *Biometrics* 49, 325–337.

Bartlett, M.S., 1953. Approximate confidence intervals II. *Biometrika* 40, 306–317.

Bischoff, W., 1993. On D -optimal designs for linear models under correlated observations with an application to a linear model with multiple response. *J. Statist. Plann. Inference* 37, 69–80.

Box, G.E.P., Draper, N.R., 1987. *Empirical Model-Building and Response Surfaces*. Wiley, New York.

Box, G.E.P., Lucas, H.L., 1959. Design of experiments in nonlinear situations. *Biometrika* 46, 77–90.

Box, M.J., Draper, N.R., 1972. Estimation and design criteria for multiresponse non-linear models with non-homogeneous variance. *J. Roy. Statist. Soc. Ser. C* 21, 13–24.

Casella, G., Berger, R.L., 2002. *Statistical Inference*. second ed. Duxbury, Pacific Grove, CA.

Chaloner, K., Lantz, K., 1989. Optimal Bayesian design applied to logistic regression experiments. *J. Statist. Plann. Inference* 21, 191–208.

Chaloner, K., Verdinelli, I., 1995. Bayesian experimental design: a review. *Statist. Sci.* 10, 273–304.

Chang, S.I., 1994. Some properties of multiresponse D -optimal designs. *J. Math. Anal. Appl.* 184, 256–262.

Cordeiro, G.M., McCullagh, P., 1991. Bias correction in generalized linear models. *J. Roy. Statist. Soc. Ser. B* 53, 629–643.

Cox, D.R., 1972. The analysis of multivariate binary data. *Appl. Statist.* 21, 113–120.

- Cox, D.R., Snell, E.J., 1968. A general definition of residuals. *J. Roy. Statist. Soc. Ser. B* 30, 248–275.
- Dasgupta, A., 1996. Review of optimal Bayes designs. In: Ghosh, S., Rao, C.R. (Eds.), *Handbook of Statistics*, vol. 13. Elsevier Science B.V., Amsterdam, pp. 1099–1142.
- Dette, H., Sperlich, S., 1994. A note on Bayesian D -optimal designs for generalization of the simple exponential growth model. *South African Statist. J.* 28, 103–117.
- Dobson, A.J., 2002. *An introduction to Generalized Linear Models*. second ed. Chapman & Hall/CRC, London, Boca Raton, FL.
- Draper, N.R., Hunter, W.G., 1966. Design of experiments for parameter estimation in multiresponse situations. *Biometrika* 53, 525–533.
- Fahrmeir, L., Tutz, G., 2001. *Multivariate Statistical Modelling Based on Generalized Linear Models*. second ed. Springer, New York.
- Fedorov, V.V., 1972. *Theory of Optimal Experiments*. Academic Press, New York.
- Ford, I., Titterton, D.M., Kitsos, C.P., 1989. Recent advances in nonlinear experimental design. *Technometrics* 31, 49–60.
- Gennings, C., Carter Jr., W.H., Martin, B.R., 1994. Drug interactions between morphine and marijuana. In: Lange, N., Ryan, L., Billard, L., Brillinger, D., Conquest, L., Greenhouse, J. (Eds.), *Case Studies in Biometry*. Wiley, New York, pp. 429–451.
- Haines, L.M., 1995. A geometrical approach to optimal Bayesian designs for one-parameter models. *J. Roy. Statist. Soc. Ser. B* 57, 575–598.
- Hatzis, C., Larntz, K., 1992. Optimal design in nonlinear multiresponse estimation: Poisson model for filter feeding. *Biometrics* 48, 1235–1248.
- Heise, M.A., Myers, R.H., 1996. Optimal designs for bivariate logistic regression. *Biometrics* 52, 613–624.
- Huang, Y.C., Wong, W.K., 1998. Sequential construction of multiple-objective optimal designs. *Biometrics* 54, 1388–1397.
- Imhof, L., 2000. Optimum designs for a multiresponse regression model. *J. Multivariate Anal.* 72, 120–131.
- Khuri, A.I., 1996. Multiresponse surface methodology. In: Ghosh, S., Rao, C.R. (Eds.), *Handbook of Statistics*, vol.13. Elsevier Science B.V., Amsterdam, pp. 377–406.
- Khuri, A.I., Cornell, J.A., 1996. *Response Surfaces*. second ed. Dekker, New York.
- Khuri, A.I., Mukhopadhyay, S., 2006. GLM designs: the dependence on unknown parameters dilemma. In: Khuri, A.I. (Ed.), *Response Surface Methodology and Related Topics*. World Scientific, Singapore, pp. 203–223.
- Khuri, A.I., Mukherjee, B., Sinha, B.K., Ghosh, M., 2006. Design issues for generalized linear models: a review. *Statist. Sci.*, 21, 376–399.
- Krafft, O., Schaefer, M., 1992. D -optimal designs for a multivariate regression model. *J. Multivariate Anal.* 42, 130–140.
- Mathew, T., Sinha, B.K., 2001. Optimal designs for binary data under logistic regression. *J. Statist. Plann. Inference* 93, 295–307.
- McCullagh, P., Nelder, J.A., 1989. *Generalized Linear Models*. second ed. Chapman & Hall, London.
- McCulloch, C.E., Searle, S.R., 2001. *Generalized, Linear, and Mixed Models*. Wiley, New York.
- Mukhopadhyay, S., Haines, L.M., 1995. Bayesian D -optimal designs for the exponential growth model. *J. Statist. Plann. Inference* 44, 385–397.
- Mukhopadhyay, S., Khuri, A.I., 2005. Bias in multivariate generalized linear models. Technical Report 2005-029, Department of Statistics, University of Florida.
- Myers, R.H., 1999. Response surface methodology-current status and future directions. *J. Qual. Tech.* 31, 30–44.
- Myers, R.H., Montgomery, D.C., 1995. *Response Surface Methodology*. Wiley, New York.
- Myers, R.H., Montgomery, D.C., Vining, G.G., 2002. *Generalized Linear Models with Applications in Engineering and the Sciences*. Wiley, New York.
- Peruca, E., Pisani, F., 1989. Dose-response relationship and therapeutic drug monitoring. In: Lasagna, L., Erill, S., Naranjo, C.A. (Eds.), *Dose-Response Relationships in Clinical Pharmacology*. Elsevier Science B.V., Amsterdam, pp. 201–214.
- Robinson, K.S., Khuri, A.I., 2003. Quantile dispersion graphs for evaluating and comparing designs for logistic regression models. *Comput. Statist. Data Anal.* 43, 47–62.
- SAS Institute Inc., 2000. *SAS Online Doc*, Version 8. Cary, NC.
- Sitter, R.R., Forbes, B., 1997. Optimal two-stage designs for binary response experiments. *Statist. Sinica* 7, 941–956.
- Sitter, R.R., Wu, C.F.J., 1993. Optimal designs for binary response experiments: Fieller, D, and A criteria. *Scan. J. Statist.* 20, 329–342.
- Sitter, R.R., Wu, C.F.J., 1999. Two-stage design of quantal response studies. *Biometrics* 55, 396–402.
- Wald, A., 1943. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Amer. Math. Soc.* 54, 426–482.
- Wijesinha, M.C., Khuri, A.I., 1987. The sequential generation of multiresponse D -optimal designs when the variance-covariance matrix is not known. *Comm. Statist. Simulation Comput. B* 16, 239–259.
- Wu, C.F.J., 1985. Efficient sequential designs with binary data. *J. Amer. Statist. Assoc.* 80, 974–984.
- Zacks, S., 1977. Problems and approaches in design of experiments for estimation and testing in non-linear models. In: Krishnaiah, P.R. (Ed.), *Multivariate Analysis*, vol. 4. North-Holland, Amsterdam, pp. 209–223.
- Zocchi, S.S., Atkinson, A.C., 1999. Optimum experimental designs for multinomial logistic models. *Biometrics* 55, 437–444.