Journal of Statistical Planning and Inference 210 (2021) 20–41

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Robust designs for generalized linear mixed models with possible model misspecification

Xiaojian Xu [a,*], Sanjoy K. Sinha [b]

[a] Department of Mathematics and Statistics, Brock University, Canada
[b] School of Mathematics and Statistics, Carleton University, Canada

ARTICLE INFO

Article history:
Received 24 March 2019
Received in revised form 12 April 2020
Accepted 12 April 2020
Available online 21 April 2020

Keywords:
Count data
Integrated mean squared error
Logistic regression
Mixed model
Robust design

ABSTRACT

We study robust designs for generalized linear mixed models (GLMMs) with protections against possible departures from underlying model assumptions. Among various types of model departures, an imprecision in the assumed linear predictor or the link function has a great impact on predicting the conditional mean response function in a GLMM. We develop methods for constructing adaptive sequential designs when the fitted mean response or the link function is possibly of an incorrect parametric form. We adopt the maximum likelihood method for estimating the parameters in GLMMs and investigate both I-optimal and D-optimal design criteria for the construction of robust sequential designs. To study the empirical properties of these sequential designs, we ran a series of simulations using both logistic and Poisson mixed models. As indicated in the simulation results, the I-optimal design generally outperforms the D-optimal design for all scenarios considered. Both designs are more efficient than the conventionally used uniform design and the classical D-optimal design obtained under the assumption that the fitted models are correctly specified. The proposed designs are also illustrated in an example using actual data from a dose–response experiment.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

In this article, we discuss the construction of model-robust designs for generalized linear mixed models (GLMMs). The majority of literature on model-robust designs are for linear models and a few are for generalized linear models (GLMs), but little work has been done on model-robust designs for GLMMs. A possible reason for the dearth of work in this area may be explained by (i) the complexity of the design problem even when the assumed model is exactly correct and (ii) the computational difficulty involved in design problems for GLMMs.

With an awareness of the possibility that the assumed generalized linear model (GLM) might not be exactly correct, Adewale and Wiens (2009) developed criteria to generate robust designs for logistic regression models with possible inaccuracy in assumed linear predictors. In addition, Adewale and Xu (2010) discussed the construction of robust designs, in general, for generalized linear models with protections against not only possible misspecification in assumed linear predictors but also inadequately assumed link function and possible overdispersion.

Robust designs for GLMs were also investigated by Woods et al. (2006), and Li and Wiens (2011). Woods et al. (2006) proposed a method for finding exact designs for experiments with generalized linear models, which uses a design criterion that is robust to uncertainty in the link function, the linear predictor, or the regression parameters. Li and Wiens (2011)

* Corresponding author.
E-mail addresses: xxu@brocku.ca (X. Xu), sinha@math.carleton.ca (S.K. Sinha).

https://doi.org/10.1016/j.jspi.2020.04.006
0378-3758/© 2020 Elsevier B.V. All rights reserved.

developed experimental designs for dose–response studies, which are robust against possibly misspecified link functions. The proposed design minimizes the maximum mean squared error of the estimated dose required to attain a response in certain percentage of the target population.

With the notion that the assumed model is correctly specified, Sinha and Xu (2011) constructed D-optimal sequential designs for GLMMs, while Sinha and Xu (2016) further provided practical methods of design construction for GLMMs which can overcome the computational intensiveness and difficulty involved in constructing optimal designs for the GLMM. In this paper, we aim to investigate optimal and robust sequential designs for GLMMs under the consideration of possible misspecifications in the assumed parametric forms. There are several scenarios in which GLMMs can be potentially misspecified. These scenarios include GLMMs with (M1) misspecified linear predictors, (M2) imprecisions in assumed link functions, (M3) overdispersion and (M4) misspecified distributions for random effects. Any combination of departures from (M1)–(M4) may also occur. Our goal is to consider a design process which would be robust against any of the aforementioned departures from the model assumptions. From our investigation, it appears that both (M1) and (M2) have a greater impact on either predicting the mean response or estimating the fixed effects parameters in GLMMs, while (M3) and (M4) have more impacts on the estimation of variance components. A common goal of optimal designs for GLMMs is the prediction or estimation of the conditional mean response structure rather than the variability across clusters. Therefore, we will focus on robust designs for GLMMs which can address (M1) and (M2) types of model departures.

For (M1), either the covariates included in the systematic component or the linear predictor of a GLMM may not reflect the influence of covariates correctly. This situation often occurs due to the use of an incorrectly specified functional form of the covariates in the model or an omission of essential covariates from the regression model. For instance, Heagerty and Kurland (2001) investigated the impact of model violations on the estimates of regression coefficients in GLMMs. Their study indicates that there can be substantial bias in the conditionally specified regression estimators that may result from using a misspecified random intercept model, where the true random effects distribution depends on measured covariates or when there are autoregressive random effects. For (M2), the link function adopted in the assumed GLMM model is often an approximation to the truth. For example, we often use the logit link in a binary regression model, which is the canonical link for the binomial distribution, when, in fact, the complementary log–log link or the probit link may be more appropriate. In an investigation of designs for nonlinear models, Sinha and Wiens (2002, p.602) have asserted that "although the theoretical response functions are very similar in shape, and possibly indistinguishable if noisy data must be relied upon, the appropriate designs can be quite dissimilar". Similarly, since the link function determines the response function in a GLMM, the misspecification of the link function may also have an impact on the resulting optimal designs.

Here we develop our robust methods for constructing adaptive sequential designs in GLMMs by addressing possible imprecisions in the assumed linear predictors or link functions, where the maximum likelihood method is adopted for fitting GLMMs. The rest of this paper is organized as follows. Section 2 introduces the model and notation, and reviews the maximum likelihood (ML), score function and Fisher information for fitting a GLMM and for finding the asymptotic variance–covariance matrix of the ML estimators. Section 3 discusses how the two types of departures (M1 and M2) are related and what their impacts are on the estimates of the parameters in GLMMs or on the prediction of the overall mean response function. Section 4 presents our proposed I-optimal and D-optimal designs and provides a sequential procedure for finding robust designs in GLMMs. Section 5 explores the distributions of the optimal sequential design points attained by our proposed methods using Monte Carlo simulations and compares the performance of the I- and D-optimal designs with other competitors. Section 6 illustrates our proposed method using an example of a real dataset from a dose–response experiment. Section 7 offers some concluding remarks. Derivations are provided in the Appendix.

2. Model and notation

Usual measures of performance of a design for a GLM or a GLMM depend on parameters being estimated. The most common treatments for such dependency issue in a GLM are the methods of minimax (see Sitter, 1992; King and Wong, 2000), Bayesian (see Chaloner and Verdinelli, 1995; Zhang and Ye, 2014; and most recently Maram and Jafari, 2016), and sequential designs (see Atkinson et al., 2007, § 17.7; Dror and Steinberg, 2008). In this paper, we adopt an adaptive sequential approach to choose design points for the GLMM so that a measure of performance, which is evaluated at estimates obtained from given data, can be maximized.

To define the GLMM, we consider a linear regression function (linear predictor) of a control covariate vector $\mathbf{x}$ in the form

$$\eta_i = \mathbf{r}^t(\mathbf{x}_i)\boldsymbol{\beta} + \mathbf{z}_i^t\mathbf{u},$$

where $\mathbf{r}(\mathbf{x}_i)$ is a p-dimensional vector of regressors for the fixed effects and $\mathbf{z}_i$ is a q-dimensional vector of regressors for the random effects. Suppose conditional on the vector of random effects $\mathbf{u}$, the elements of the observed response vector $\mathbf{y} = (y_1, \ldots, y_n)^t$ are independent and follow a distribution in the exponential family:

$$f_{y_i|u}(y|\mathbf{u}, \boldsymbol{\beta}, \phi) = \exp\left\{\frac{y\theta_i - b(\theta_i)}{a(\phi)} + c(y, \phi)\right\}, \tag{1}$$

for some known functions a, b and c, where θ_i is related to the linear predictor η_i through a known link function, and ϕ is a dispersion parameter. Note that for a canonical link as we consider in this paper, $\theta_i = \eta_i$.

The vector of random effects $\mathbf{u}$ is assumed to follow a distribution:

$$\mathbf{u} \sim f_u(\mathbf{u}|\boldsymbol{\alpha}), \tag{2}$$

depending on a vector of variance components $\boldsymbol{\alpha}$. A link function g relates the linear predictor η_i in (1) to the conditional mean response $E_{y|u}(y_i|\mathbf{u}) = \mu(\mathbf{x}_i, \boldsymbol{\beta}, \mathbf{u}) = \mu(\mathbf{r}^t(\mathbf{x}_i)\boldsymbol{\beta} + \mathbf{z}_i^t\mathbf{u})$ such that

$$\eta_i = g(\mu(\mathbf{x}_i, \boldsymbol{\beta}, \mathbf{u})),$$

where g is a monotonic and differentiable function. When g is the identity function and the response has the normal distribution, we obtain the special class of linear mixed models.

From (1) and (2), the marginal likelihood function can be defined as

$$L(\boldsymbol{\beta}, \phi, \boldsymbol{\alpha}|\mathbf{y}) = \int f_{y|u}(\mathbf{y}|\mathbf{u}, \boldsymbol{\beta}) f_u(\mathbf{u}|\boldsymbol{\alpha}) d\mathbf{u} = \int \prod_{i=1}^{n} f_{y_i|u}(y_i|\mathbf{u}, \boldsymbol{\beta}, \phi) f_u(\mathbf{u}|\boldsymbol{\alpha}) d\mathbf{u}. \tag{3}$$

The ML estimators of the parameters $\boldsymbol{\beta}$, ϕ and $\boldsymbol{\alpha}$ can be obtained by maximizing this likelihood function using suitable numerical techniques.

For simplicity, we consider $\phi = 1$, as this is the case for both binary and Poisson regression models. We note that when the marginal distribution of $\mathbf{y}$ can be defined as a mixture as in (3), the classical ML estimating equations for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ take the form:

$$E_{u|y}\left[\frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \middle| \mathbf{y} \right] = \mathbf{0}, \tag{4}$$

and

$$E_{u|y}\left[\frac{\partial \log f_u(\mathbf{u}|\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \middle| \mathbf{y} \right] = \mathbf{0}, \tag{5}$$

respectively, where $E_{u|y}$ denotes the expectation with respect to the conditional distribution of $\mathbf{u}$ given $\mathbf{y}$ (see McCulloch and Searle, 2001 for details). The ML estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ can be obtained by solving Eqs. (4) and (5) using some iterative method, such as the Newton–Raphson method as described in McCulloch (1997).

The observed Fisher information can be obtained in a matrix form as

$$\mathbf{I}_o(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \begin{bmatrix} \mathbf{I}_{o11}(\boldsymbol{\beta}, \boldsymbol{\alpha}) & \mathbf{I}_{o12}(\boldsymbol{\beta}, \boldsymbol{\alpha}) \\ \mathbf{I}_{o21}(\boldsymbol{\beta}, \boldsymbol{\alpha}) & \mathbf{I}_{o22}(\boldsymbol{\beta}, \boldsymbol{\alpha}) \end{bmatrix}, \tag{6}$$

where the components of the information matrix are given by

$$\begin{aligned} -\mathbf{I}_{o11}(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \frac{\partial^2 \log L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^t} \\ &= E_{u|y}\left[\frac{\partial^2 \log f_{y|u}(\mathbf{y}|\mathbf{u}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^t} \middle| \mathbf{y} \right] \\ &\quad + E_{u|y}\left[\frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^t} \middle| \mathbf{y} \right] \\ &\quad - E_{u|y}\left[\frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \middle| \mathbf{y} \right] E_{u|y}\left[\frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}^t} \middle| \mathbf{y} \right], \end{aligned} \tag{7}$$

$$\begin{aligned} -\mathbf{I}_{o22}(\boldsymbol{\beta}, \boldsymbol{\alpha}) &= \frac{\partial^2 \log L}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^t} \\ &= E_{u|y}\left[\frac{\partial^2 \log f_u(\mathbf{u}|\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha} \partial \boldsymbol{\alpha}^t} \middle| \mathbf{y} \right] \\ &\quad + E_{u|y}\left[\frac{\partial \log f_u(\mathbf{u}|\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \frac{\partial \log f_u(\mathbf{u}|\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^t} \middle| \mathbf{y} \right] \\ &\quad - E_{u|y}\left[\frac{\partial \log f_u(\mathbf{u}|\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} \middle| \mathbf{y} \right] E_{u|y}\left[\frac{\partial \log f_u(\mathbf{u}|\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^t} \middle| \mathbf{y} \right], \end{aligned} \tag{8}$$

and

$$-\mathbf{I}_{o12}(\boldsymbol{\beta},\boldsymbol{\alpha}) = -\mathbf{I}_{o21}^{t}(\boldsymbol{\beta},\boldsymbol{\alpha}) = \frac{\partial^2 \log L}{\partial\boldsymbol{\beta}\partial\boldsymbol{\alpha}^t}$$
$$= E_{u|y}\left[\frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\,\frac{\partial \log f_u(\mathbf{u}|\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}^t}\middle|\mathbf{y}\right]$$
$$- E_{u|y}\left[\frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\middle|\mathbf{y}\right]E_{u|y}\left[\frac{\partial \log f_u(\mathbf{u}|\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}^t}\middle|\mathbf{y}\right]. \tag{9}$$

For the exponential family (1) with a canonical link, we can show that

$$E_{u|y}\left[\frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\middle|\mathbf{y}\right] = \sum_{i=1}^{n}\{y_i - E_{u|y}(\mu(\mathbf{x}_i,\boldsymbol{\beta},\mathbf{u})|\mathbf{y})\}\mathbf{r}(\mathbf{x}_i), \tag{10}$$

and

$$E_{u|y}\left[\frac{\partial^2 \log f_{y|u}(\mathbf{y}|\mathbf{u},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^t}\middle|\mathbf{y}\right] = -\sum_{i=1}^{n}E_{u|y}[w(\mathbf{x}_i,\boldsymbol{\beta},\mathbf{u})|\mathbf{y}]\mathbf{r}(\mathbf{x}_i)\mathbf{r}^t(\mathbf{x}_i), \tag{11}$$

where $w(\mathbf{x}_i,\boldsymbol{\beta},\mathbf{u}) = var(y_i|\mathbf{u})$. The expected Fisher information can be obtained by taking the marginal expectations of the expressions (7)–(9) with respect to the response vector $\mathbf{y}$. After simplification, we can show that

$$E_y\left[-\frac{\partial^2 \log L}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^t}\right] = E_y\left[E_{u|y}\left[\frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\middle|\mathbf{y}\right]E_{u|y}\left[\frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^t}\middle|\mathbf{y}\right]\right], \tag{12}$$

$$E_y\left[-\frac{\partial^2 \log L}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}^t}\right] = E_y\left[E_{u|y}\left[\frac{\partial \log f_u(\mathbf{u}|\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}}\middle|\mathbf{y}\right]E_{u|y}\left[\frac{\partial \log f_u(\mathbf{u}|\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}^t}\middle|\mathbf{y}\right]\right], \tag{13}$$

and

$$E_y\left[-\frac{\partial^2 \log L}{\partial\boldsymbol{\beta}\partial\boldsymbol{\alpha}^t}\right] = E_y\left[E_{u|y}\left[\frac{\partial \log f_{y|u}(\mathbf{y}|\mathbf{u},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\middle|\mathbf{y}\right]E_{u|y}\left[\frac{\partial \log f_u(\mathbf{u}|\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}^t}\middle|\mathbf{y}\right]\right]. \tag{14}$$

Let the model parameters be $\boldsymbol{\theta} = (\boldsymbol{\beta}^t, \boldsymbol{\alpha}^t)^t$. Then the expected full Fisher information becomes $E_y[-\partial^2 \log L/(\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^t)]$.

In this paper, our goal of the study is to determine if there is a significant trend in the mean response with respect to the covariates rather than the variability of such trend across the clusters. This implies that we are looking for the best designs which can produce the most accuracy in estimating $\boldsymbol{\beta}$ rather than $\boldsymbol{\alpha}$.

3. Possible departures from model assumptions

The major contribution of the paper is the development of an optimal design strategy by taking possible model departures into account at the experimental design stage. We consider a situation where the experimenter seeks protection against possible misspecifications in the assumed model form. As discussed in Section 1, the following two types of departures from an assumed GLMM an experimenter should be aware of, and they can be noted as follows:

- (M1) *Misspecified linear predictor:* When the true linear predictor is $\eta = \mathbf{r}^t(\mathbf{x})\boldsymbol{\beta} + \mathbf{z}^t\mathbf{u} + h(\mathbf{x})$, with a contamination function $h(\mathbf{x})$ accounting for possible additional effects of covariates.
- (M2) *Imprecisions in the assumed link function:* When the true link function is different from the assumed one.

3.1. Misspecified linear predictor (M1)

Formally, when (M1) appears in a GLMM, the true but unknown model belongs to a class of alternative models with the conditional mean

$$E_{y|u}(y_i|\mathbf{u}) = \mu^u\left(\eta_{T,i}|\mathbf{u}\right),\ \eta_{T,i} = \mathbf{r}^t(\mathbf{x}_i)\boldsymbol{\beta} + \mathbf{z}_i^t\mathbf{u} + h(\mathbf{x}_i), \tag{15}$$

assuming $h(\mathbf{x}_i)$ is a small contaminant, $\mathbf{u}\sim f_u(\mathbf{u}|\boldsymbol{\alpha})$ with $\boldsymbol{\alpha}$ being independent of $\mathbf{x}_i$, and $E_u(\mathbf{u}) = \mathbf{0}$. However, the experimenter imprecisely fits a GLMM with the conditional mean

$$\mu^u(\eta_i|\mathbf{u}) = \mu\left(\mathbf{r}^t(\mathbf{x}_i)\boldsymbol{\beta} + \mathbf{z}_i^t\mathbf{u}\right). \tag{16}$$

As a result, the working likelihood function is adopted as in (3).

Let $\mu(\mathbf{x}_i, \boldsymbol{\beta}) = E_u(\mu^u(\eta_i|\mathbf{u})) = \int_{\mathbf{u}} \mu\left(\mathbf{r}^t(\mathbf{x}_i)\boldsymbol{\beta} + \mathbf{z}_i^t\mathbf{u}\right) f_u(\mathbf{u}|\boldsymbol{\alpha})d\mathbf{u}$ be the fitted marginal mean response. Then the true marginal mean response can be expressed by

$$\begin{aligned} E_y(y_i|\mathbf{x}_i) &= E_u\left(\mu^u\left(\eta_{T,i}|\mathbf{u}\right)\right) = \int \mu\left(\mathbf{r}^t(\mathbf{x}_i)\boldsymbol{\beta} + \mathbf{z}_i^t\mathbf{u} + h(\mathbf{x}_i)\right) f_u(\mathbf{u}|\boldsymbol{\alpha})d\mathbf{u} \\ &= \int \left[\mu(\mathbf{r}^t(\mathbf{x}_i)\boldsymbol{\beta} + \mathbf{z}_i^t\mathbf{u}) + \mu'(\mathbf{r}^t(\mathbf{x}_i)\boldsymbol{\beta} + \mathbf{z}_i^t\mathbf{u})h(\mathbf{x}_i) + o(h(\mathbf{x}_i))\right] f_u(\mathbf{u}|\boldsymbol{\alpha})d\mathbf{u} \\ &= \mu(\mathbf{x}_i, \boldsymbol{\beta}) + d(\mathbf{x}_i, \boldsymbol{\beta}), \end{aligned}$$

where

$$\begin{aligned} d(\mathbf{x}_i, \boldsymbol{\beta}) &= \int \left[\mu'(\mathbf{r}^t(\mathbf{x}_i)\boldsymbol{\beta} + \mathbf{z}_i^t\mathbf{u})h(\mathbf{x}_i) + o(h(\mathbf{x}_i))\right] f_u(\mathbf{u}|\boldsymbol{\alpha})d\mathbf{u} \\ &= E_y(y_i|\mathbf{x}_i) - \mu(\mathbf{x}_i, \boldsymbol{\beta}). \end{aligned}$$

We also define the "true" $\boldsymbol{\beta}_0$ through minimization of integrated squared discrepancy

$$\boldsymbol{\beta}_0 = \arg\min \int_S d^2(\mathbf{x}, \boldsymbol{\beta})d\mathbf{x}. \tag{17}$$

This implies that

$$\int_S \mathbf{q}(\mathbf{x}, \boldsymbol{\beta}_0)d(\mathbf{x}, \boldsymbol{\beta}_0)d\mathbf{x} = \mathbf{0},$$

with $\mathbf{q}(\mathbf{x}, \boldsymbol{\beta}) = \partial\mu(\mathbf{x}, \boldsymbol{\beta})/\partial\boldsymbol{\beta}$. Then the true marginal mean is given by

$$E_y(y_i|\mathbf{x}_i) = \mu(\mathbf{x}_i, \boldsymbol{\beta}_0) + d(\mathbf{x}_i, \boldsymbol{\beta}_0).$$

3.2. Misspecified link function (M2)

3.2.1. Link function misspecification in a logistic mixed model

Uncertainty in the choice of a link function employed in a GLM or a GLMM often occurs (see discussion in Pregibon, 1980, and Ponce de Leon and Atkinson, 1993, for instance). The optimal design problems for GLMs with misspecified link functions have been discussed previously. For example, Biedermann et al. (2006) introduce a robust approach in which the experimenter considers a finite set of plausible link functions with the uncertainties in the suitability of each of them quantified with known probabilities. The probabilities reflect the preference that the experimenter attaches to each link function. The resulting robust criterion is a weighted average of the respective criterion corresponding to each link function. Later, Adewale and Xu (2010) take a different approach given that the experimenter intends to fit the logistic model (a binomial model with the logit link). However, the robust designs are proposed to protect against the possibility of the imprecision in the logit link specified, within a parametric link function family. In the present paper, we adopt a similar approach for GLMMs.

Assume that both the true but unknown link function and the fitted logit link function belong to a generalized family defined by

$$g\left(\mu^B, \lambda\right) = \log\left[\left\{\left(\frac{1}{1-\mu^B}\right)^{\lambda} - 1\right\} \Big/ \lambda\right], (\lambda \geq 0), \tag{18}$$

with a link parameter λ. This family encompasses the logit and the complementary log–log links as special cases. The logit link corresponds to $\lambda = 1$ while the complementary log–log corresponds to $\lim_{\lambda\to 0}$. Although some link functions such as the probit link may not be exactly expressed as a member of the family (18), these can be approximated to a particular member of the family. For instance, the probit link function is quite close to the member with $\lambda = 1/3$. In reality, the true link function corresponds to an unknown value of λ that might be different from the fitted one, such as the logit link with $\lambda = 1$.

In generalized linear mixed modeling, the link function connects the systematic component (the linear predictor) of the model to the mean response via

$$\eta = g\left(\mu^B, \lambda\right),$$

where η is the linear predictor representing the mixed effects in the model on a linear scale. The reasonable choice of link functions is suggested by the distribution of the response. For example, $g\left(\mu^B, \lambda\right)$ is suggested as the logit function when the response takes the logistic distribution function, and as the complementary log–log link when it takes the extreme value distribution. We considered the following Taylor's expansion of (18) about the parameter value $\lambda = 1$ corresponding to the logit link that the experimenter contemplates fitting. That is,

$$\eta = \log\left(\frac{\mu^B}{1-\mu^B}\right) + g^*(\mathbf{x}; \lambda),$$

with $\mu^B = E_{y|u}(y|\mathbf{u})$ and

$$g^*(\mathbf{x};\lambda) = \left.\frac{\partial g}{\partial \lambda}\right|_{\lambda=1}(\lambda-1) + o(\lambda-1). \tag{19}$$

Thus the link misspecification problem may be cast as a linear predictor misspecification problem. The true mean response $\mu_i^B = E_{y|u}(y_i|\mathbf{u})$ can be expressed as a function of the linear predictor

$$\eta_i = \mathbf{r}^t(\mathbf{x}_i)\boldsymbol{\beta} + \mathbf{z}_i^t\mathbf{u} + g^*(\mathbf{x}_i,\lambda),$$

where the contamination function is given by

$$h(\mathbf{x}_i,\lambda) = \log\left(\frac{\mu^B}{1-\mu^B}\right) - \left(\mathbf{r}^t(\mathbf{x}_i)\boldsymbol{\beta} + \mathbf{z}_i^t\mathbf{u}\right) = -g^*(\mathbf{x}_i,\lambda). \tag{20}$$

3.2.2. Link function misspecification in a Poisson mixed model

Similarly, in designing an experiment for a Poisson mixed model with a possibly misspecified link function, we may employ a link function class where both the true link function and the fitted canonical link function are included. Consider the parameterized family of link functions defined by

$$g\left(\mu^P,\lambda\right) = \begin{cases} \left(\mu^P\right)^\lambda, & \lambda \neq 0 \\ \log\mu^P, & \lambda = 0 \end{cases},$$

where λ is the link parameter. The log link corresponds to $\lambda = 0$. The strategy is to linearize this generalized link function by expanding it around $\lambda = 0$. Then the true linear predictor can be written as

$$\eta_i = \mathbf{r}^t(\mathbf{x}_i)\boldsymbol{\beta} + \mathbf{z}_i^t\mathbf{u} + h(\mathbf{x}_i,\lambda),$$

with the contamination function $h(\mathbf{x},\lambda)$ being

$$h(\mathbf{x},\lambda) = -\left.\frac{\partial g(\mathbf{x},\lambda)}{\partial\lambda}\right|_{\lambda=0}\lambda + o(\lambda) \approx -\lambda\log\mu^P. \tag{21}$$

As discussed above, we note that the link function misspecification in a GLMM can be treated as a linear predictor misspecification problem. This aligns with the work of Pregibon (1980) indicating that the wrong link function is a systematic misspecification of the model. Therefore, an (M2) problem can be treated as an (M1) problem.

3.3. General results

This subsection provides general results for the asymptotic bias and asymptotic variance of the maximum likelihood estimators of the fixed effects parameters in approximately specified models with awareness of any possible M1 and/or M2 departures.

From (3), the marginal log-likelihood function l can be defined as

$$\begin{aligned} l(\boldsymbol{\beta}) &= \log\int\left[\prod_{i=1}^n f_{y_i|u}(y_i|\mathbf{u},\boldsymbol{\beta},\phi)\right]f_u(\mathbf{u}|\boldsymbol{\alpha})d\mathbf{u} \\ &= \log\int\left[\prod_{i=1}^n \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i,\phi)\right\}\right]f_u(\mathbf{u}|\boldsymbol{\alpha})d\mathbf{u}. \end{aligned} \tag{22}$$

Letting $l_i^u = (y_i\theta_i - b(\theta_i))/a(\phi) + c(y_i,\phi)$, we employ the chain rule to obtain the first derivative of the log-likelihood l, given by

$$\begin{aligned} \frac{\partial l(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} &= \int\left[\frac{\partial}{\partial\boldsymbol{\beta}}f_{y|u}(\mathbf{y}|\mathbf{u},\boldsymbol{\beta})\right]f_u(\mathbf{u}|\boldsymbol{\alpha})d\mathbf{u}\Big/ f_y(\mathbf{y}) \\ &= \int\left[\frac{\partial}{\partial\boldsymbol{\beta}}\log f_{y|u}(\mathbf{y}|\mathbf{u},\boldsymbol{\beta})\right]f_{u|y}(\mathbf{u}|\mathbf{y})d\mathbf{u} \\ &= E_{u|y}\left(\frac{\partial}{\partial\boldsymbol{\beta}}\sum_{i=1}^n\log f_{y_i|u}(y_i|\mathbf{u},\boldsymbol{\beta})\middle|\mathbf{y}\right) \\ &= \sum_{i=1}^n E_{u|y}\left(\frac{\partial l_i^u}{\partial\boldsymbol{\beta}}\middle|\mathbf{y}\right) \\ &= \sum_{i=1}^n E_{u|y}\left(\frac{\partial l_i^u}{\partial\theta_i}\frac{\partial\theta_i}{d\mu_i^u}\frac{\partial\mu_i^u}{\partial\eta_i}\frac{\partial\eta_i}{\partial\boldsymbol{\beta}}\middle|\mathbf{y}\right), \end{aligned}$$

where $\partial l_i^u/\partial\theta_i = (y_i - b'(\theta_i))/a(\phi)$. Using $\mu_i^u = b'(\theta_i)$ and $var(y_i|\mathbf{u}) = a(\phi)\, b''(\theta_i)$, we have

$$\frac{\partial l_i^u}{\partial \theta_i} = (y_i - \mu_i^u)/a(\phi),$$
$$\frac{\partial \mu_i^u}{d\theta_i} = b''(\theta_i) = var(y_i|\mathbf{u})/a(\phi), \text{ and}$$
$$\frac{\partial \eta_i}{\partial \boldsymbol{\beta}} = \mathbf{r}(\mathbf{x}_i) \text{ since } \eta_i = \mathbf{r}^t(\mathbf{x}_i)\boldsymbol{\beta} + \mathbf{z}_i^t\mathbf{u}, \text{ as assumed.}$$

Thus

$$l'(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} E_{u|y}\left(\frac{y_i - \mu_i^u}{var(y_i|\mathbf{u})} \frac{\partial \mu_i^u}{\partial \eta_i} \mathbf{r}(\mathbf{x}_i) \middle| \mathbf{y} \right),$$

and

$$l''(\boldsymbol{\beta}) = \frac{\partial^2 l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^t} = \sum_{i=1}^{n} E_{u|y}\left(\frac{\partial^2 l_i^u}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^t} \middle| \mathbf{y} \right) + E_{u|y}\left(\sum_{i=1}^{n} \frac{\partial l_i^u}{\partial \boldsymbol{\beta}} \sum_{i=1}^{n} \frac{\partial l_i^u}{\partial \boldsymbol{\beta}^t} \middle| \mathbf{y} \right)$$
$$- E_{u|y}\left(\sum_{i=1}^{n} \frac{\partial l_i^u}{\partial \boldsymbol{\beta}} \middle| \mathbf{y} \right) E_{u|y}\left(\sum_{i=1}^{n} \frac{\partial l_i^u}{\partial \boldsymbol{\beta}^t} \middle| \mathbf{y} \right),$$

where

$$\frac{\partial^2 l_i^u}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^t} = \frac{\partial \mu_i^u/\partial \eta_i}{var(y_i|\mathbf{u})} \mathbf{r}(\mathbf{x}_i) \frac{\partial (y_i - \mu_i^u)}{\partial \boldsymbol{\beta}^t} + (y_i - \mu_i^u)\, \mathbf{r}(\mathbf{x}_i) \frac{\partial}{\partial \boldsymbol{\beta}^t} \left\{ \frac{\partial \mu_i^u/\partial \eta_i}{var(y_i|\mathbf{u})} \right\}.$$

We also have

$$\frac{\partial (y_i - \mu_i^u)}{\partial \boldsymbol{\beta}^t} = -\frac{\partial \mu_i^u}{\partial \eta_i} \mathbf{r}^t(\mathbf{x}_i),$$

and for a canonical model (which often the fitted model is),

$$\theta_i = \eta_i, \text{ and } \frac{\partial}{\partial \boldsymbol{\beta}^t} \left\{ \frac{\partial \mu_i^u/\partial \eta_i}{var(y_i)} \right\} = \mathbf{0}.$$

Thus we have

$$\frac{\partial l_i^u}{\partial \boldsymbol{\beta}} = \left(\frac{y_i - \mu_i^u}{a(\phi)} \right) \mathbf{r}(\mathbf{x}_i),$$

and the second derivative for the possibly misspecified canonical model is

$$\frac{\partial^2 l_i^u}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^t} = -\left[\frac{\left(\partial \mu_i^u/\partial \eta_i\right)^2}{var(y_i)} \right] \mathbf{r}(\mathbf{x}_i)\mathbf{r}^t(\mathbf{x}_i) = -\left[\frac{\partial \mu_i^u/\partial \eta_i}{a(\phi)} \right] \mathbf{r}(\mathbf{x}_i)\mathbf{r}^t(\mathbf{x}_i).$$

With a sample size of n, the following result can be established for a general GLMM in the form of (15) with possible (M1) and/or (M2) types of model misspecifications:

Theorem 1. *For a GLMM with a canonical link function, in the form of (15) with possible model misspecification, when the experimenter fits model (16) imprecisely, and assuming that the observed responses are independent and follow a distribution with a density of the form (1), the maximum likelihood estimator $\hat{\boldsymbol{\beta}}_n$ of the model parameter vector $\boldsymbol{\beta}_0$ from the misspecified model has the following properties: $\sqrt{n}\left(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\right)$ has an asymptotic multivariate normal distribution, and the asymptotic bias and asymptotic variance–covariance matrix of $\hat{\boldsymbol{\beta}}_n$ are given by*

$$bias\left(\hat{\boldsymbol{\beta}}\right) = E\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) = \mathbf{H}_n^{-1}(\boldsymbol{\beta}_0)\mathbf{b}_n(\boldsymbol{\beta}_0) + o\left(n^{-1/2}\right), \tag{23}$$

$$var\left(\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)\right) = \mathbf{H}_n^{-1}(\boldsymbol{\beta}_0) + o(1), \tag{24}$$

respectively, for functions $\mathbf{b}_n(\boldsymbol{\beta}_0) = (1/n)E_y[l'(\boldsymbol{\beta}_0)]$ and $\mathbf{H}_n(\boldsymbol{\beta}_0) = -(1/n)E_y[l''(\boldsymbol{\beta}_0)]$.

The proof of the theorem is given in the Appendix.

4. Optimal sequential design

As concluded in the previous section, the link function misspecification problems (M2) can be treated as a linear predictor misspecification (M1) problem. Therefore, starting this section, we will focus on robust designs for GLMMs with protections against possible imprecisions in the assumed linear predictors: Scenario (M1).

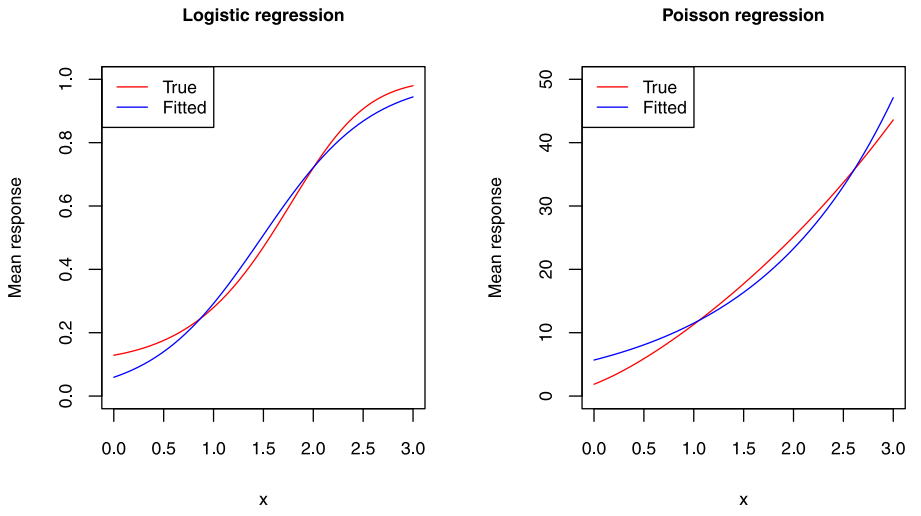

Fig. 1. True versus fitted models. Left panel – (logistic model) plot of true marginal mean response $\mu^*(x, \boldsymbol{\delta}) = E_u[\exp(\delta_0+\delta_1 x+\delta_2 x^2+u)/(1+\exp(\delta_0+\delta_1 x+\delta_2 x^2+u))]$ with $\boldsymbol{\delta} = (\delta_0, \delta_1, \delta_2)^t = (-2, 0.5, 0.5)^t$ and fitted marginal mean response $\mu(x, \boldsymbol{\beta}_0) = E_u[\exp(\beta_{00}+\beta_{10}x+u)/(1+\exp(\beta_{00}+\beta_{10}x+u))]$ with $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{10})^t = (-2.87, 1.94)^t$. Right panel – (Poisson model) plot of true marginal mean response $\mu^*(x, \boldsymbol{\delta}) = E_u[\exp(\delta_0+\delta_1 x+\delta_2 x/(1+x)+u)]$ with $\boldsymbol{\delta} = (\delta_0, \delta_1, \delta_2)^t = (0.5, 0.3, 3)^t$ and fitted marginal mean response $\mu(x, \boldsymbol{\beta}_0) = E_u[\exp(\beta_{00} + \beta_{10}x + u)]$ with $\boldsymbol{\beta}_0 = (\beta_{00}, \beta_{10})^t = (1.613, 0.705)^t$. Parameters $\boldsymbol{\beta}_0$ are obtained by minimizing the integrated squared distance $d(\boldsymbol{\beta}) = \int_0^3 \{\mu^*(x, \boldsymbol{\delta}) - \mu(x, \boldsymbol{\beta})\}^2 dx$ with respect to $\boldsymbol{\beta}$ over design space $x \in [0, 3]$.

4.1. Examples

We first take two most commonly used GLMMs to illustrate the impact of misspecified linear predictors. These two specific examples will be used through the present and next sections.

As the first example, we suppose that an experimenter considers fitting a binary logistic mixed model with a single covariate x and a random intercept u, given by

$$\text{logit}\{\mu_L^u(x, u, \boldsymbol{\beta})\} = \beta_0 + \beta_1 x + u, \tag{25}$$

over the design space $x \in [0, 3]$. However, the true model is given by

$$\text{logit}\{\mu_L^{*u}(u, x, \boldsymbol{\delta})\} = \delta_0 + \delta_1 x + \delta_2 x^2 + u,$$

where $u \sim N(0, 0.25)$. Fig. 1 (left) graphs a plot of two marginal means $E_u[\mu_L^{*u}(u, x, \boldsymbol{\delta})]$ with parameters $\boldsymbol{\delta} = (-2, 0.5, 0.5)^t$ and $E_u[\mu_L^u(x, u, \boldsymbol{\beta}_0)]$ with parameters $\boldsymbol{\beta}_0 = (-2.87, 1.94)^t$ obtained by minimizing the squared distance between two marginal response functions as defined in (17), namely,

$$\begin{aligned}\boldsymbol{\beta}_0 &= \arg\min_{\boldsymbol{\beta}} d(x, \boldsymbol{\beta}) = \arg\min_{\boldsymbol{\beta}} \int_0^3 \left\{E_u(\mu_L^{*u}(u, x, \boldsymbol{\delta})) - E_u(\mu_L^u(x, u, \boldsymbol{\beta}))\right\}^2 dx \\ &= (-2.87, 1.94)^t.\end{aligned}$$

As the second example, we consider a Poisson mixed regression model

$$\log\{\mu_P^u(x, u, \boldsymbol{\beta})\} = \beta_0 + \beta_1 x + u,$$

over the design space $x \in [0, 3]$, whereas the true mean response is given by

$$\log\{\mu_P^{*u}(u, x, \boldsymbol{\delta})\} = \delta_0 + \delta_1 x + \delta_2 \left(\frac{x}{1+x}\right) + u,$$

where $u \sim N(0, 0.25)$. Fig. 1 (right) shows a plot of the marginal means $E_u[\mu_P^{*u}(u, x, \boldsymbol{\delta})]$ with parameters $\boldsymbol{\delta} = (0.5, 0.3, 3)^t$ and $\mathrm{E}_u[\mu_P^u(x, u, \boldsymbol{\beta}_0)]$ with parameters $\boldsymbol{\beta}_0 = (1.613, 0.705)^t$ whose values also minimize the overall squared discrepancy $d(x, \boldsymbol{\beta})$.

As noted in this graph, particularly the one with logistic regression (on the left), the most obvious discrepancy $d(x, \boldsymbol{\beta})$ appears in two ending regions, around $[0, 0.5]$ and $[2.0, 3.0]$, as well as the middle area around 1.5. We would expect the robust design with protection on such discrepancy due to the misspecification in assumed linear predictor used for modeling the marginal mean. On the other hand, the plot for Poisson regression (on the right), the most discrepancy $d(x, \boldsymbol{\beta})$ appears in two regions, around $[0, 0.5]$, and $[1.5, 2.5]$. We would expect that the robust design can help with the protection against the discrepancy in these areas due to the same reason.

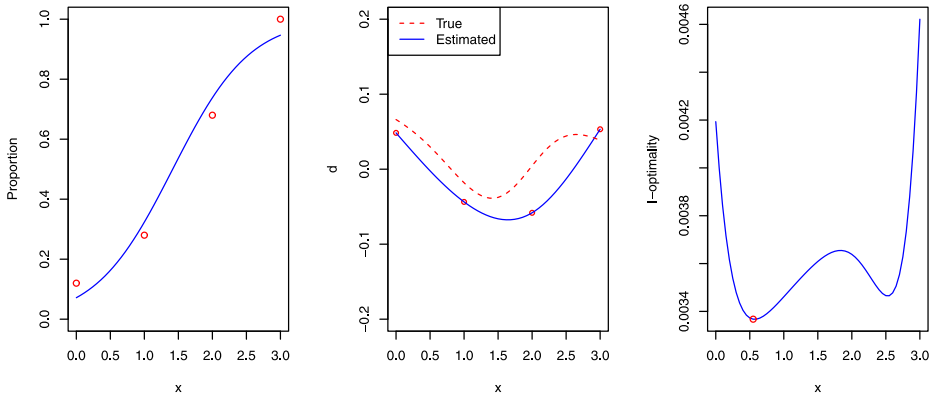

Fig. 2. True and estimated discrepancies for a sample dataset from a binary mixed model. Left panel — scatter plot of observed proportions of successes at design points $x = 0, 1, 2, 3$ along with a fitted logistic model. Middle panel — plot of true and estimated discrepancies $d(x, \boldsymbol{\beta})$; dashed line represents the true discrepancy function, red circles represent observed discrepancies at four design points and solid line an estimated smooth discrepancy function. Right panel — plot of I-optimality loss function over design space $x \in [0, 3]$; minimum loss is located around $x = 0.5$ by a red circle indicating an I-optimal design point.

For illustration, we generate 200 observations from $N = 50$ subjects, with $n_i = 4$ observations from the ith subject ($i = 1, \ldots, N$) obtained at four equally spaced design points $x_i = 0, 1, 2, 3$ over the design space [0, 3] using the logistic model

$$\text{logit}\{\mu^{*u}(u, x, \boldsymbol{\delta})\} = -2 + 0.5x + 0.5x^2 + u, \;\; u \sim N(0, 0.25).$$

However, model (25) is fitted using the maximum likelihood method. Fig. 2 exhibits the true and estimated discrepancies for fitting the incorrect binary mixed model (25). The left panel shows the scatter plot of the observed proportions of successes at the four design points $x = 0, 1, 2, 3$ along with the fitted logistic model. The middle panel shows the plot of true and estimated discrepancies $d(x, \boldsymbol{\beta})$, where the dashed line represents the true discrepancy function, red circles represent observed discrepancies at four design points and solid line an estimated smooth discrepancy function. We investigate the empirical properties of the smoothed residuals in approximating $d(x, \boldsymbol{\beta})$ shown in Fig. 2. Here the values of

$$\int_0^3 \mathbf{q}(x, \hat{\boldsymbol{\beta}})\hat{d}(x, \hat{\boldsymbol{\beta}})dx$$

appear to be $(-0.018, -0.028)$, which are close to zero as expected. For larger samples (not shown here), the values were found to be even closer to zero. The right panel of Fig. 2 shows the plot of I-optimality loss function, as defined in Eq. (27), over the design space $x \in [0, 3]$, where the minimum loss is located around $x = 0.5$ by a red circle indicating an I-optimal design point.

4.2. Design criteria and sequential design procedure

The classical optimal design strategies are often developed by maximizing a function of the Fisher information matrix. However, when the fitted model is misspecified, as investigated in Section 2, the ML estimators of the model parameters are generally biased. Consequently, the optimal design criteria should be based on the asymptotic integrated mean squared error (IMSE) of prediction defined by

$$\begin{aligned} &\int_S E\left[\left\{\mu\left(\mathbf{x}, \hat{\boldsymbol{\beta}}_n\right) - E(\boldsymbol{y}|\mathbf{x})\right\}^2\right] d\mathbf{x} \qquad (26)\\ &= \int_S E\left[\left\{(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)^t \mathbf{q}(\mathbf{x}, \boldsymbol{\beta}_0) - d(\mathbf{x}, \boldsymbol{\beta}_0)\right\}^2\right] d\mathbf{x} + o\left(n^{-1/2}\right)\\ &\approx \text{trace}\{\text{MSE}_n(\boldsymbol{\beta}_0)\mathbf{A}(\boldsymbol{\beta}_0)\} + \int_S d^2(\mathbf{x}, \boldsymbol{\beta}_0)d\mathbf{x}, \end{aligned}$$

where MSE_n is the MSE of $\hat{\boldsymbol{\beta}}_n$ and $\mathbf{A}(\boldsymbol{\beta}) = \int_S \mathbf{q}(\mathbf{x}, \boldsymbol{\beta})\mathbf{q}^t(\mathbf{x}, \boldsymbol{\beta})d\mathbf{x}$.

We consider two design criteria, I-optimality and D-optimality, where the I-optimal design obtains design points by minimizing the loss function

$$\mathcal{L}_I = \text{trace}\{\text{MSE}_n(\boldsymbol{\beta}_0)\mathbf{A}(\boldsymbol{\beta}_0)\}, \qquad (27)$$

and the D-optimal design obtains design points by minimizing the loss function

$$\mathcal{L}_D = \left[\det\{\mathrm{MSE}_n(\boldsymbol{\beta}_0)\}\right]^{1/p}, \tag{28}$$

where

$$\mathrm{MSE}_n(\boldsymbol{\beta}) = (1/n)\mathbf{H}_n^{-1}(\boldsymbol{\beta}) + \mathbf{H}_n^{-1}(\boldsymbol{\beta})\mathbf{b}_n(\boldsymbol{\beta})\mathbf{b}_n^t(\boldsymbol{\beta})\mathbf{H}_n^{-1}(\boldsymbol{\beta}),$$

with

$$\mathbf{b}_n(\boldsymbol{\beta}) = (1/n)\sum_{i=1}^{n} d(\mathbf{x}_i, \boldsymbol{\beta})\mathbf{r}(\mathbf{x}_i),$$

and

$$\mathbf{H}_n(\boldsymbol{\beta}) = (1/n)E_y\left[\left\{\sum_{i=1}^{n}(y_i - E_{u|y}(\mu_i^u|\mathbf{y}))\mathbf{r}(\mathbf{x}_i)\right\}\left\{\sum_{i=1}^{n}(y_i - E_{u|y}(\mu_i^u|\mathbf{y}))\mathbf{r}(\mathbf{x}_i)\right\}^t\right].$$

For choosing the optimal designs, the calculation of $\mathbf{H}_n(\boldsymbol{\beta})$ requires intensive computation, as it involves numerical integration at each design point in the design space. To reduce the computational burden, we further approximate the function $\mathbf{H}_n(\boldsymbol{\beta})$ by

$$\tilde{\mathbf{H}}_n(\boldsymbol{\beta}) = (1/n)E_y\left[\left\{\sum_{i=1}^{n}(y_i - E_u(\mu_i^u))\mathbf{r}(\mathbf{x}_i)\right\}\left\{\sum_{i=1}^{n}(y_i - E_u(\mu_i^u))\mathbf{r}(\mathbf{x}_i)\right\}^t\right].$$

In our experience with the empirical study, the optimal design criterion does not heavily rely on the conditional expectation $E_{u|y}(\mu_i^u|y)$ as used in the function $H_n(\boldsymbol{\beta})$ above. The I-optimal loss functions based on $H_n(\boldsymbol{\beta})$ and $\tilde{H}_n(\boldsymbol{\beta})$ are similar in shape and both provide design points that are generally close to each other. We should also point out that we adopt this approximation only at the design stage of the analysis for computational simplicity. However, the variance of the ML estimators is obtained from the exact Fisher information derived from the full likelihood function.

Finally, the discrepancy function $d(\mathbf{x}_i, \boldsymbol{\beta})$ is approximated by a nonparametric smoothing technique similar to Sinha and Wiens (2002). Specifically, assuming that multiple responses $\{y_{ij}, j = 1, \ldots, n_i\}$ are obtained at the ith distinct design point $\mathbf{x}_i$ $(i = 1, 2, \ldots)$, so that $n = \sum_i n_i$, the discrepancies $d(\mathbf{x}_i, \boldsymbol{\beta})$ are first estimated by the means of the "residuals" $r_{ij} = y_{ij} - \mu(\mathbf{x}_i, \hat{\boldsymbol{\beta}}), j = 1, \ldots, n_i$. These means are then smoothed to get estimates of the discrepancy $d(\mathbf{x}, \boldsymbol{\beta})$ at any design point $\mathbf{x}$.

We adopt a sequential approach to the design problem assuming that the experimenter initially obtains data from a group of individuals measured at n_0 pre-selected design points $\mathbf{x}_i\,(i = 1, \ldots, n_0)$. Subsequent data are then observed at n_1 new design points $\mathbf{x}_i\,(i = n_0 + 1, n_0 + 2, \ldots, n_0 + n_1)$ determined sequentially by minimizing either (27) or (28) in order to obtain the I-optimal or D-optimal sequential designs, respectively. Then the resulting designs shall be robust against possible model departures. Our proposed algorithm for constructing such robust I-optimal or D-optimal sequential designs can be described as follows:

1. Given the initial data $\{(y_i, \mathbf{x}_i);\ i = 1, \ldots, n_0\}$, find the ML estimates $\hat{\boldsymbol{\theta}}_0 = (\hat{\boldsymbol{\beta}}_0^t, \hat{\boldsymbol{\alpha}}_0^t)^t$ of $(\boldsymbol{\beta}^t, \boldsymbol{\alpha}^t)^t$ by solving the estimating Eqs. (4) and (5) simultaneously.
2. Based on these initial estimates and a candidate design point $\mathbf{x}$, calculate the design criteria $\mathcal{L}_I$ or $\mathcal{L}_D$ as

$$\mathcal{L}_I = \mathrm{trace}\{\mathrm{MSE}_{n_0+1}(\hat{\boldsymbol{\beta}}_0)\mathbf{A}(\hat{\boldsymbol{\beta}}_0)\}, \tag{29}$$

or

$$\mathcal{L}_D = \left[\det\{\mathrm{MSE}_{n_0+1}(\hat{\boldsymbol{\beta}}_0)\}\right]^{1/p}. \tag{30}$$

3. Determine a new design point $\mathbf{x}_{n_0+1}^*$ by minimizing (29) or (30):

$$\mathbf{x}_{n_0+1}^* = \arg\min_{\mathbf{x}}\left[\mathrm{trace}\{\mathrm{MSE}_{n_0+1}(\hat{\boldsymbol{\beta}}_0)\mathbf{A}(\hat{\boldsymbol{\beta}}_0)\}\right] \tag{31}$$

or

$$\mathbf{x}_{n_0+1}^* = \arg\min_{\mathbf{x}}\left\{\left[\det\{\mathrm{MSE}_{n_0+1}(\hat{\boldsymbol{\beta}}_0)\}\right]^{1/p}\right\}. \tag{32}$$

4. Update the parameter estimates $\hat{\boldsymbol{\theta}}$ using the augmented data with the observations obtained at the new design point $\mathbf{x}_{n_0+1}^*$, and then select the next sequential design point based on the new set of estimates, and so on.
5. Repeat Steps 2–4 to choose n_1 design points, $\mathbf{x}_{n_0+1}^*, \ldots, \mathbf{x}_{n_0+n_1}^*$, sequentially using this algorithm.

An important issue of a sequential design procedure is to set up a stopping rule. The number of new design points should be determined in order to obtain reliable estimates of the model parameters. In reality, it often depends on both the experimenter's point of view on estimation precision and available experimental resource. If the experimental resource is not limited, then the experimenter may continue the process of generating observations based on the sequentially chosen optimal design points until satisfactory estimates of the model parameters are attained. A workable rule is to continue sampling until the variances of the ML estimates fall below a predetermined threshold.

5. Simulation study

In this section, we present results from a series of simulations in order to explore the properties of the ML estimators obtained under the sequential design schemes proposed earlier. We also carry out a comparative study based on the following three classes of designs:

- **Class I:** Proposed I-optimal or robust D-optimal designs, where initial data at n_0 locations are augmented by an additional k observations at each of n_1 new locations chosen sequentially by numerically minimizing $\mathcal{L}_I$ or $\mathcal{L}_D$.
- **Class II:** Conventionally used uniform designs, where initial data at n_0 locations are augmented by an additional k observations at each of n_1 new locations. In this case, the experimental points are uniformly distributed throughout the design space, where a total of $n_0 + n_1$ locations are equally spaced over $[0, 3]$, and for smaller values of $n_0 + n_1$ these may form a subset of the locations. Thus these design points are sequential but nonadaptive.
- **Class III:** Classical optimal design for a GLMM without consideration of model departures. It obtains classical D-optimal designs sequentially using the same procedure as described in the previous section. So these designs are sequential and adaptive, but are obtained by maximizing the determinant of $E_y[-\partial^2 \log L/(\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^t)]$ instead of minimizing $\mathcal{L}_I$ or $\mathcal{L}_D$.

In the next two subsections, we study empirical properties of the aforementioned three classes of designs using both binary logistic and Poisson mixed models.

5.1. Simulation results for logistic mixed models

Here initial data were generated from the "true" binary mixed model

$$\begin{aligned} y_{ij}|u_i &\sim \text{ind. Bernoulli}(p^*_{ij}),\\ \text{logit}(p^*_{ij}) &= \delta_0 + \delta_1 x_j + \delta_2 x_j^2 + u_i,\\ u_i &\sim \text{ind. } N(0, \sigma_u^2), \end{aligned} \tag{33}$$

for $i = 1, \ldots, k, j = 1, \ldots, n_0$ with $n_0 = 4$ initial design points $\{x_j : 0, 1, 2, 3\}$. The model parameters were fixed at $(\delta_0, \delta_1, \delta_2) = (-2, \delta_1, 0.5)$ and $\sigma_u^2 = 0.25$. The fitted model assumed a working binary mixed model $\text{logit}(p_{ij}) = \beta_0 + \beta_1 x_j + u_i$ with $u_i \sim$ ind. $N(0, \sigma_u^2)$. In this case, the marginal likelihood function for $(\boldsymbol{\beta}, \sigma_u^2)$ takes the form

$$L(\boldsymbol{\beta}, \sigma_u^2|\mathbf{y}) = \prod_{i=1}^{k} \int \prod_{j=1}^{n_0} f_{y_{ij}|u_i}(y_{ij}|u_i, \boldsymbol{\beta}) f_{u_i}(u_i|\sigma_u^2) du_i.$$

For calculating the design criteria $\mathcal{L}_I$ and $\mathcal{L}_D$ in Eqs. (27)–(28), the functions $\mathbf{b}_n(\boldsymbol{\beta})$ and $\mathbf{H}_n(\boldsymbol{\beta})$ may be expressed as

$$\mathbf{b}_n(\boldsymbol{\beta}) = (k/n) \sum_{j=1}^{n_0} d(\mathbf{x}_j, \boldsymbol{\beta}) \mathbf{x}_j,$$

and

$$\mathbf{H}_n(\boldsymbol{\beta}) = (1/n) \sum_{i=1}^{k} E_y \left[\left\{ \sum_{j=1}^{n_0} (y_{ij} - E_{u|y}(\mu_{ij}^u|\mathbf{y}_i)) \mathbf{x}_j \right\} \left\{ \sum_{j=1}^{n_0} (y_{ij} - E_{u|y}(\mu_{ij}^u|\mathbf{y}_i)) \mathbf{x}_j \right\}^t \right],$$

with $n = n_0 k$, $n_0 = 4$, $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_0})^t$, $\mathbf{x}_j = (1, x_j)^t$, and $\mu_{ij}^u = p_{ij}$.

We found four new design points $x_j (j = 5, \ldots, 8)$ in $x \in [0, 3]$ by each of the three design schemes I–III described earlier. Fig. 3 exhibits histograms of the sequentially chosen I-optimal design points $x_j (j = 5, \ldots, 8)$ for 1000 replicates of datasets obtained from the above working binary mixed model with the number of clusters $k = 25$ and parameter $\delta_1 = 0.5$. We repeat the above for other combinations of $(k, \delta_1) = (50, 0.5), (25, 0.2)$, and $(50, 0.2)$. The results are shown in the Appendix in Figs. B.1–B.3.

From the histogram plots, we note that the modes of the first four I-optimal sequential design points often alternately appear at the regions where the least model discrepancies appear (see the plot on the left in Fig. 1, when $\delta_1 = 0.5$), even more so for a larger sample size with $k = 50$. As expected, the I-optimal design appears to provide protections against the

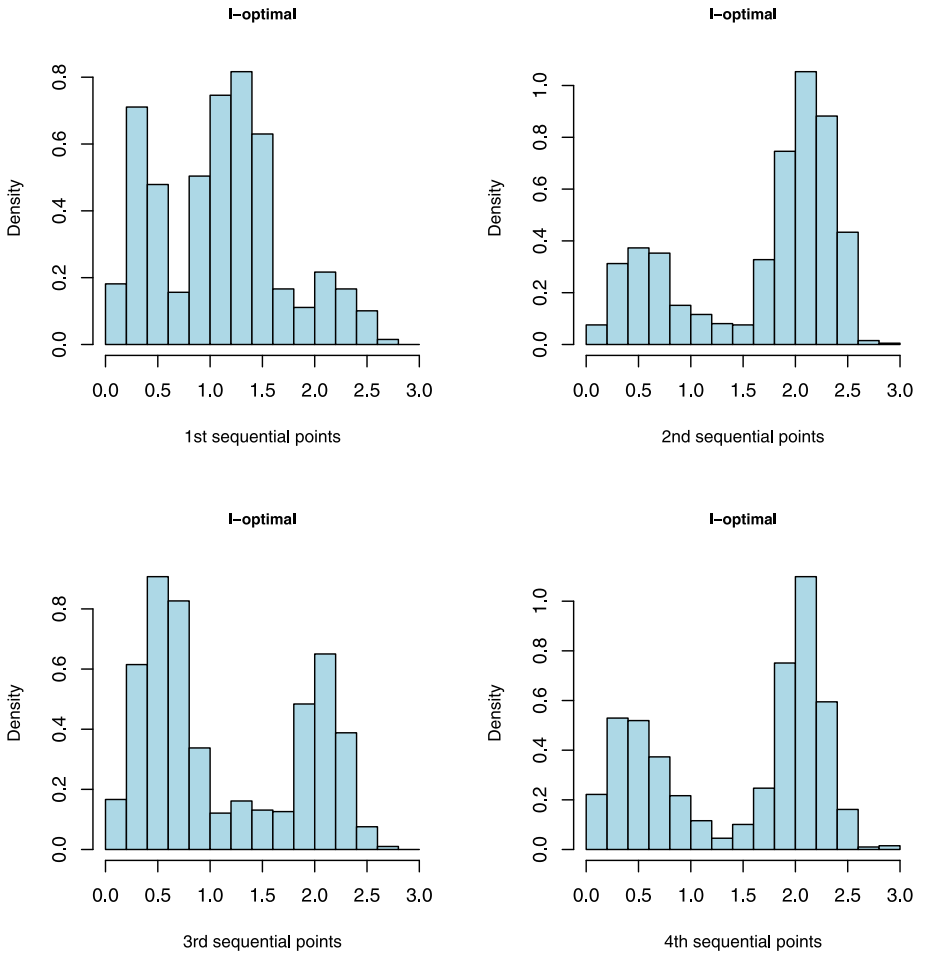

Fig. 3. Histograms of four sequentially chosen I-optimal design points for 1000 replicates of datasets obtained from a binary mixed model. Initial data were generated from the true logistic model $\text{logit}(p_{ij}^*) = \delta_0 + \delta_1 x_j + \delta_2 x_j^2 + u_i, (i = 1, \ldots, k, j = 1, \ldots, n_0)$, where the fitted model assumes a simple binary model $\text{logit}(p_{ij}) = \beta_0 + \beta_1 x_j + u_i$ with $u_i \sim$ ind. $N(0, \sigma_u^2)$. Parameters were fixed at $(\delta_0, \delta_1, \delta_2) = (-2, \mathbf{0.5}, 0.5)$ and $\sigma_u^2 = 0.25$. Number of clusters $k = \mathbf{25}$.

model discrepancies considered in the simulations. The histogram plots also indicate that: (i) the distributions of these four sequential points are complementary to each other in a consecutive way; (ii) the patterns of the distributions are quite similar for different values of k but with the same value of δ_1 and the distribution is less uniform with a larger value of k; (iii) in contrast, the distribution patterns are quite different for the different values of δ_1 even with the same value of k and the distribution with $\delta_1 = 0.2$ is much less uniform than the corresponding distribution with $\delta_1 = 0.5$; and (iv) for all the cases considered above, the very first sequential points selected appear to be more uniform than their followers.

To compare the performance of the three classes of designs, we draw scatter plots of $\sqrt{n \times \text{IMSE}}$ versus n for $n = n_0 k, (n_0 + 1)k, \ldots, (n_0 + n_1)k$. Figs. 4 and 5 exhibit plots of adjusted empirical values of $\sqrt{n \times \text{IMSE}}$ against n for 1000 replicates of datasets obtained from the above binary mixed model. The IMSEs were rescaled by dividing them with the overall mean IMSE for all designs. The plots are shown for $\delta_1 = 0.5$ and $\delta_1 = 0.2$ in Figs. 4 and 5, respectively, with $k = 25$ on the left and $k = 50$ on the right. We note that for all the cases considered in our simulations, the proposed I-optimal robust designs outperform very much all their competitors and for most cases their efficiency gains grow higher as more new sequentially selected points are added in.

It is also interesting to note that for a larger sample size with $k = 50$ the naive D-optimal design that ignores the model discrepancy appears to perform better than the "robust" D-optimal design. This seemingly unexpected pattern is difficult to interpret, as it appears only in the case of binary data. Our proposed I-optimal robust design, however, outperforms all its competitors for all sample sizes considered.

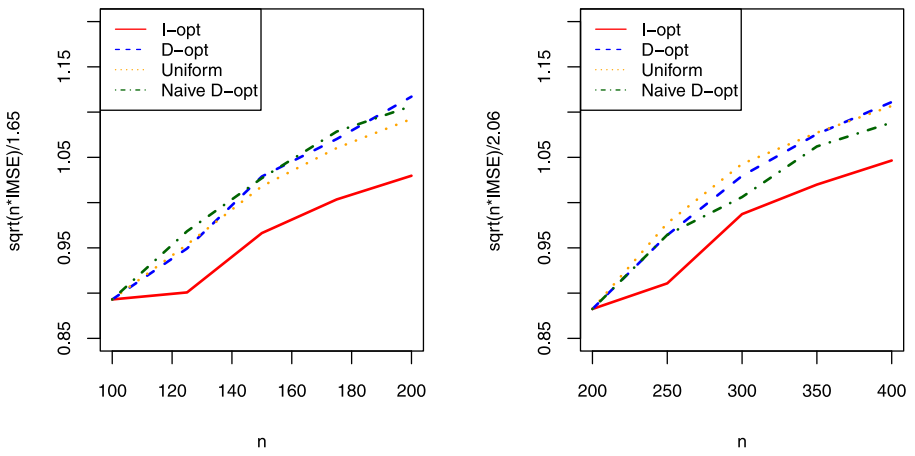

Fig. 4. Plots of adjusted empirical values of $\sqrt{n \times \text{ IMSE}}$ versus n for 1000 replicates of datasets from binary mixed models, where IMSE $= \int_S E_y[(\mu(\mathbf{x}, \hat{\boldsymbol{\beta}}_n) - E_y(\mathbf{y}|\mathbf{x}))^2]d\mathbf{x}$ and values of n were chosen as $n = n_0k, (n_0+1)k, \ldots, (n_0+n_1)k$ with $n_0 = 4$, $n_1 = 4$. Initial data were generated from the true logistic model $\text{logit}(p_{ij}^*) = \delta_0 + \delta_1 x_j + \delta_2 x_j^2 + u_i, (i = 1, \ldots, k, j = 1, \ldots, n_0)$, where the fitted model assumes a simple binary model $\text{logit}(p_{ij}) = \beta_0 + \beta_1 x_j + u_i$ with $u_i \sim$ ind. $N(0, \sigma_u^2)$. Parameters were fixed at $(\delta_0, \delta_1, \delta_2) = (-2, \mathbf{0.5}, 0.5)$ and $\sigma_u^2 = 0.25$. Results are shown for naive D-optimal, uniform, I-optimal and D-optimal designs. Left panel — number of clusters $k = 25$. Right panel — number of clusters $k = 50$.

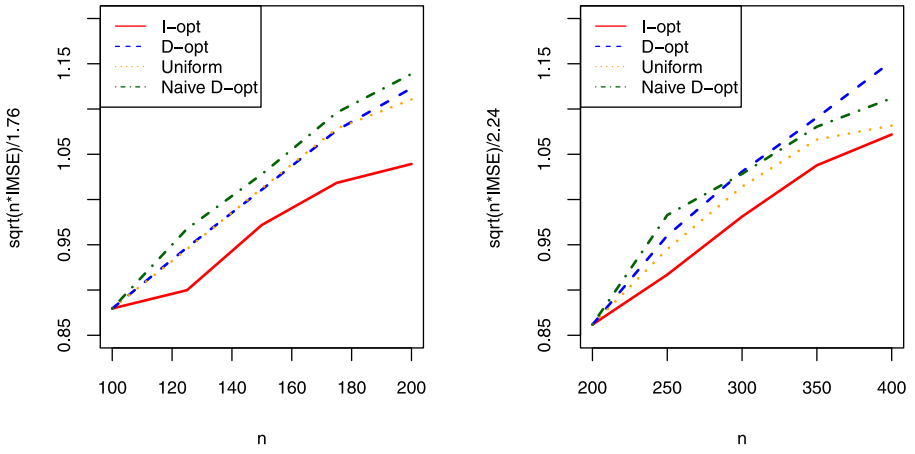

Fig. 5. Plots of adjusted empirical values of $\sqrt{n \times \text{ IMSE}}$ versus n for 1000 replicates of datasets from binary mixed models, where IMSE $= \int_S E_y[(\mu(\mathbf{x}, \hat{\boldsymbol{\beta}}_n) - E_y(\mathbf{y}|\mathbf{x}))^2]d\mathbf{x}$ and values of n were chosen as $n = n_0k, (n_0+1)k, \ldots, (n_0+n_1)k$ with $n_0 = 4$, $n_1 = 4$. Initial data were generated from the true logistic model $\text{logit}(p_{ij}^*) = \delta_0 + \delta_1 x_j + \delta_2 x_j^2 + u_i, (i = 1, \ldots, k, j = 1, \ldots, n_0)$, where the fitted model assumes a simple binary model $\text{logit}(p_{ij}) = \beta_0 + \beta_1 x_j + u_i$ with $u_i \sim$ ind. $N(0, \sigma_u^2)$. Parameters were fixed at $(\delta_0, \delta_1, \delta_2) = (-2, \mathbf{0.2}, 0.5)$ and $\sigma_u^2 = 0.25$. Results are shown for naive D-optimal, uniform, I-optimal and D-optimal designs. Left panel — number of clusters $k = 25$. Right panel — number of clusters $k = 50$.

5.2. Simulation results for Poisson mixed models

Here initial data were generated from the "true" Poisson mixed model

$$\begin{aligned} y_{ij}|u_i &\sim \text{ind. Poisson}(\lambda_{ij}^*), \\ \log(\lambda_{ij}^*) &= \delta_0 + \delta_1 x_j + \delta_2 \left(\frac{x_j}{1+x_j}\right) + u_i, \\ u_i &\sim \text{ind. } N(0, \sigma_u^2), \end{aligned} \tag{34}$$

for $i = 1, \ldots, k, j = 1, \ldots, n_0$ with $n_0 = 4$ initial design points $\{x_j : 0, 1, 2, 3\}$. The model parameters were fixed at $(\delta_0, \delta_1, \delta_2) = (0.5, \delta_1, 3)$ and $\sigma_u^2 = 0.25$. The fitted model assumes a working Poisson mixed model $\log(\lambda_{ij}) = \beta_0 + \beta_1 x_j + u_i$ with $u_i \sim$ ind. $N(0, \sigma_u^2)$. As before, we found four new design points $x_j\,(j = 5, \ldots, 8)$ in $x \in [0, 3]$ by each of the three design schemes. Fig. 6 exhibits histograms of four sequentially chosen I-optimal design points $x_j\,(j = 5, \ldots, 8)$ for 1000 replicates of datasets obtained from the above working Poisson mixed model with the number of clusters $k = 25$ and

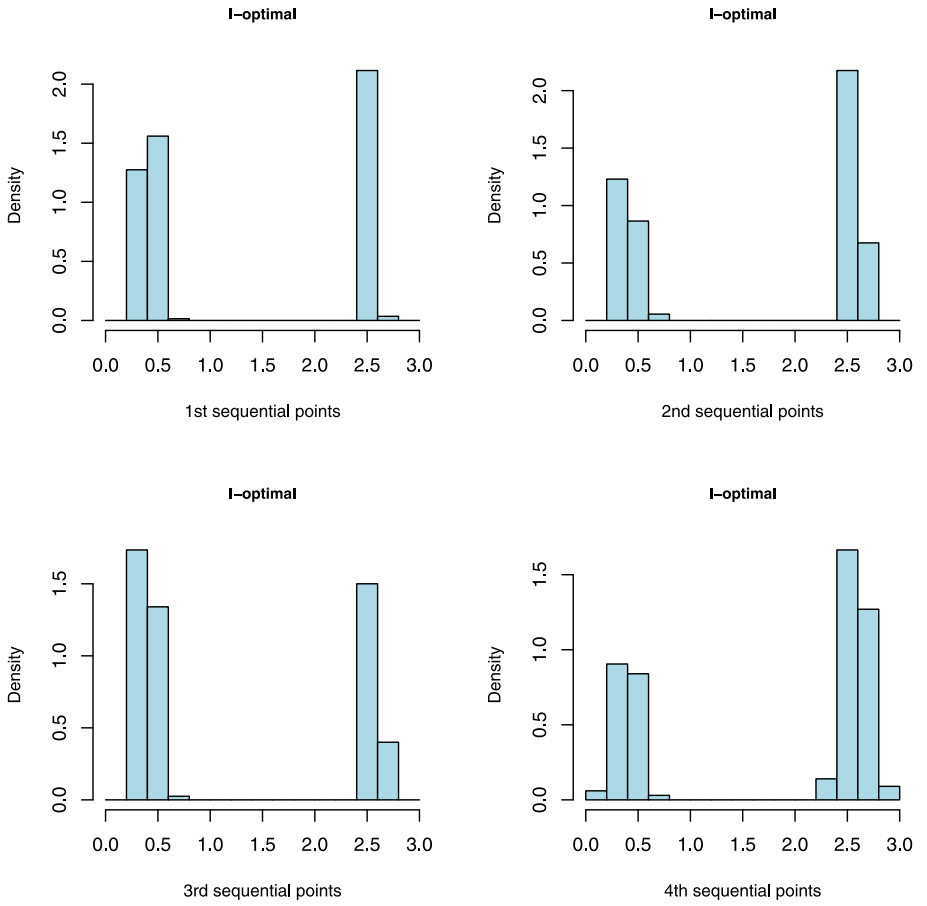

Fig. 6. Histograms of four sequentially chosen I-optimal design points for 1000 replicates of datasets obtained from a Poisson mixed model. Initial data were generated from the true Poisson model $\log(\lambda_{ij}^*) = \delta_0 + \delta_1 x_j + \delta_2 x_j/(1+x_j) + u_i, (i = 1, \ldots, k, j = 1, \ldots, n_0)$, where the fitted model assumes a simple Poisson model $\log(\lambda_{ij}) = \beta_0 + \beta_1 x_j + u_i$ with $u_i \sim$ ind. $N(0, \sigma_u^2)$. Parameters were fixed at $(\delta_0, \ \delta_1, \delta_2) = (0.5, \mathbf{0.1}, 3)$ and $\sigma_u^2 = 0.25$. Number of clusters $k = \mathbf{25}$.

parameter $\delta_1 = 0.1$. We repeat the above for other combinations of $(k, \delta_1) = (50, 0.1)$, $(25, 0.3)$, and $(50, 0.3)$. The results are shown in the Appendix in Figs. B.4–B.6.

The histogram plots reveal that the modes of the first four I-optimal design sequential points often alternately appear at the regions where the least model discrepancies appear (see the plot on the right in Fig. 1, when $\delta_1 = 0.3$), even more so for a larger number of clusters, $k = 50$. Also, as expected, the proposed robust design appears to provide protections against the model discrepancies considered in the simulation study. In addition, similar to the logistic model, we also note that the distributions of the four sequential design points are still complementary to each other in a consecutive way, but the patterns of the distributions are now quite similar for different values of k and δ_1. In comparison, the resulting design points distributions are much less uniform than those for the logistic cases, and the protection areas of the robust designs for a Poisson mixed model do not seem as accurate as that for a logistic mixed model.

We also compare the performance of the three classes of designs for the Poisson mixed model. Figs. 7 and 8 exhibit plots of adjusted empirical values of $\sqrt{n \times \text{IMSE}}$ against n for 1000 replicates of datasets. The values of the IMSEs were rescaled by dividing them with their overall mean IMSE from all designs. The plots are shown for $\delta_1 = 0.1$ and $\delta_1 = 0.3$ in Figs. 7 and 8, respectively, with $k = 25$ on the left and $k = 50$ on the right. We note that (i) for all cases, both I-optimal and D-optimal robust designs outperform the classical D-optimal designs where the misspecification in the model was not taken into account; (ii) for most cases, both our proposed I-optimal and D-optimal robust designs outperform the uniform designs; (iii) the performances of the I-optimal and D-optimal designs are quite similar; and (iv) the relative efficiency gains of I-optimal and D-optimal robust designs over the classical D-optimal designs grow higher as more new sequentially selected points are added in, whereas the gains relative to the uniform designs grow less.

6. An example: Quinoline data

Breslow (1984) presented and analyzed data from an Ames Salmonella reverse mutagenicity assay. The data (shown in Table 1) were obtained from a test with Salmonella strain TA98 and induced rat liver homogenate S9. Dimethyl sulfoxide

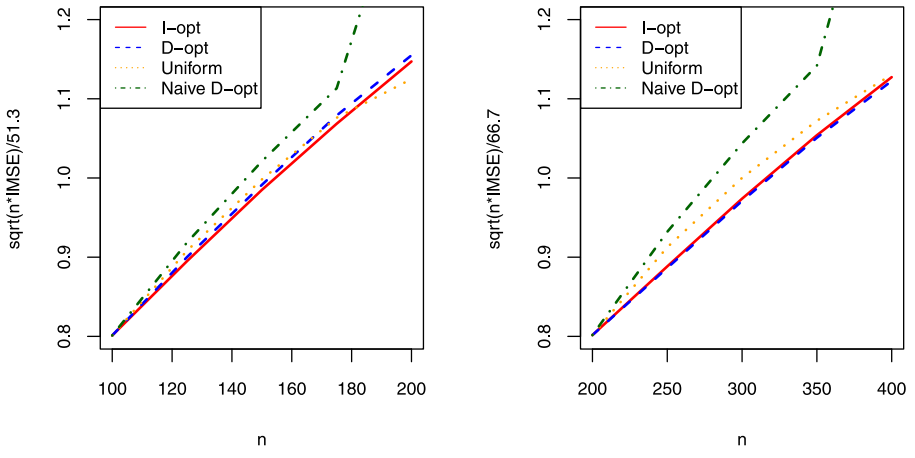

Fig. 7. Plots of adjusted empirical values of $\sqrt{n \times \text{IMSE}}$ versus n for 1000 replicates of datasets from Poisson mixed models, where IMSE $= \int_S E_y[(\mu(\mathbf{x}, \hat{\boldsymbol{\beta}}_n) - E_y(\mathbf{y}|\mathbf{x}))^2]d\mathbf{x}$ and values of n were chosen as $n = n_0k, (n_0 + 1)k, \ldots, (n_0 + n_1)k$ with $n_0 = 4$, $n_1 = 4$. Initial data were generated from the true Poisson model $\log(\lambda^*_{ij}) = \delta_0 + \delta_1 x_j + \delta_2 x_j/(1 + x_j) + u_i, (i = 1, \ldots, k, j = 1, \ldots, n_0)$, where the fitted model assumes a simple Poisson model $\log(\lambda_{ij}) = \beta_0 + \beta_1 x_j + u_i$ with $u_i \sim$ ind. $N(0, \sigma_u^2)$. Parameters were fixed at $(\delta_0, \delta_1, \delta_2) = (0.5, \mathbf{0.1}, 3)$ and $\sigma_u^2 = 0.25$. Results are shown for naive D-optimal, uniform, I-optimal and D-optimal designs. Left panel — number of clusters $k = 25$. Right panel — number of clusters $k = 50$.

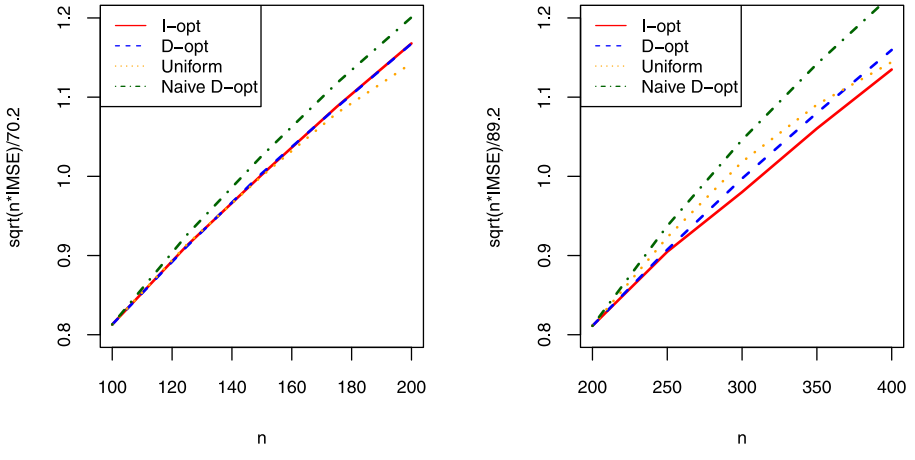

Fig. 8. Plots of adjusted empirical values of $\sqrt{n \times \text{IMSE}}$ versus n for 1000 replicates of datasets from Poisson mixed models, where IMSE $= \int_S E_y[(\mu(\mathbf{x}, \hat{\boldsymbol{\beta}}_n) - E_y(\mathbf{y}|\mathbf{x}))^2]d\mathbf{x}$ and values of n were chosen as $n = n_0k, (n_0 + 1)k, \ldots, (n_0 + n_1)k$ with $n_0 = 4$, $n_1 = 4$. Initial data were generated from the true Poisson model $\log(\lambda^*_{ij}) = \delta_0 + \delta_1 x_j + \delta_2 x_j/(1 + x_j) + u_i, (i = 1, \ldots, k, j = 1, \ldots, n_0)$, where the fitted model assumes a simple Poisson model $\log(\lambda_{ij}) = \beta_0 + \beta_1 x_j + u_i$ with $u_i \sim$ ind. $N(0, \sigma_u^2)$. Parameters were fixed at $(\delta_0, \delta_1, \delta_2) = (0.5, \mathbf{0.3}, 3)$ and $\sigma_u^2 = 0.25$. Results are shown for naive D-optimal, uniform, I-optimal and D-optimal designs. Left panel — number of clusters $k = 25$. Right panel — number of clusters $k = 50$.

Table 1
Quinoline data[a].

Revertant colonies with increasing quinoline					
0	10	33	100	333	1000
15	16	16	27	33	20
21	18	26	41	38	27
29	21	33	60	41	42

[a]Source: Breslow (1984).

was the solvent control, where each of the six doses was replicated three times. The number of revertant colonies was observed on each of three replicate plates tested at each of six dose levels of quinoline.

We revisit the quinoline data and consider analyzing them using a count response variable y_{ij} representing the number of revertant colonies with an associated covariate $x_j = \text{dose}_j/100$. We consider a working Poisson mixed model for the

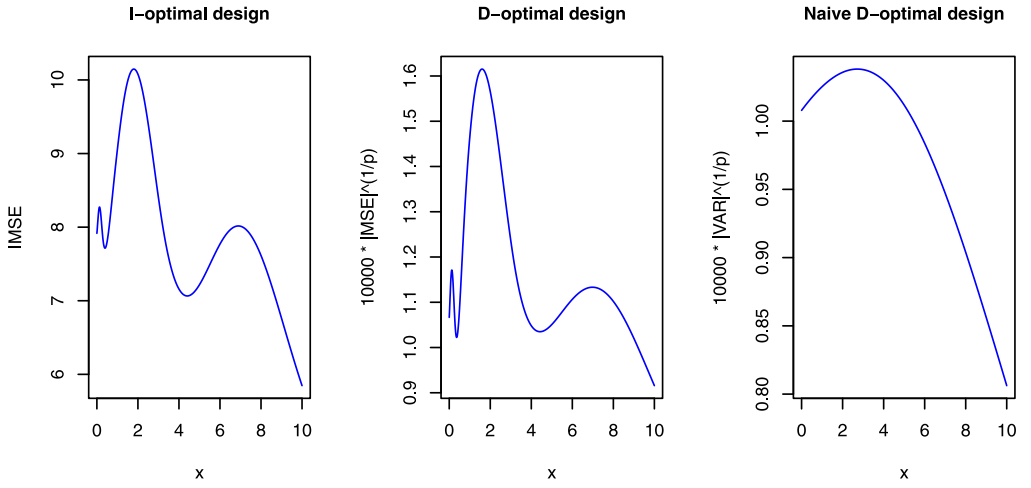

Fig. 9. Plots of IMSE loss functions over the design space $x \in [0, 10]$. Left panel: I-optimal loss function; Middle panel: D-optimal loss function; Right panel: Naive D-optimal loss function. All loss functions show a minimum loss at the highest dose level $x = 10$.

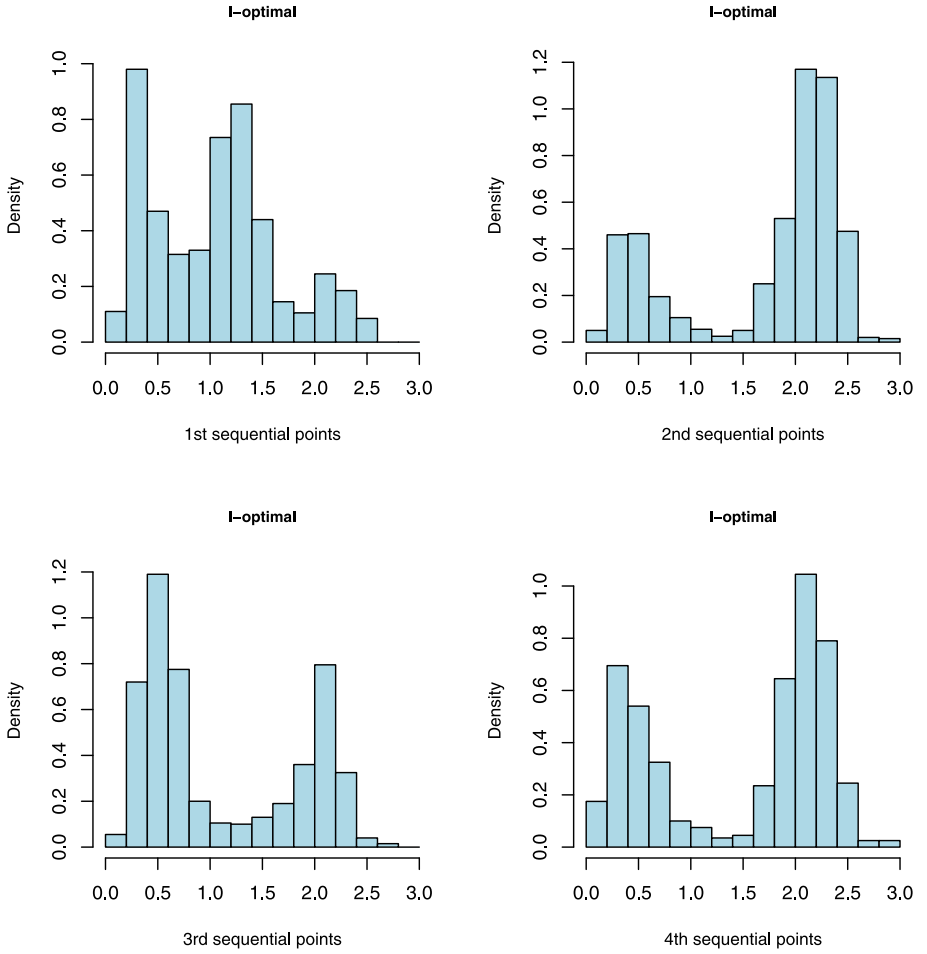

Fig. B.1. Histograms of four sequentially chosen I-optimal design points for 1000 replicates of datasets obtained from a binary mixed model. Initial data were generated from the true logistic model $\text{logit}(p_{ij}^*) = \delta_0 + \delta_1 x_j + \delta_2 x_j^2 + u_i, (i = 1, \ldots, k, j = 1, \ldots, n_0)$, where the fitted model assumes a simple binary model $\text{logit}(p_{ij}) = \beta_0 + \beta_1 x_j + u_i$ with $u_i \sim$ ind. $N(0, \sigma_u^2)$. Parameters were fixed at $(\delta_0, \delta_1, \delta_2) = (-2, \mathbf{0.5}, 0.5)$ and $\sigma_u^2 = 0.25$. Number of clusters $k = \mathbf{50}$.

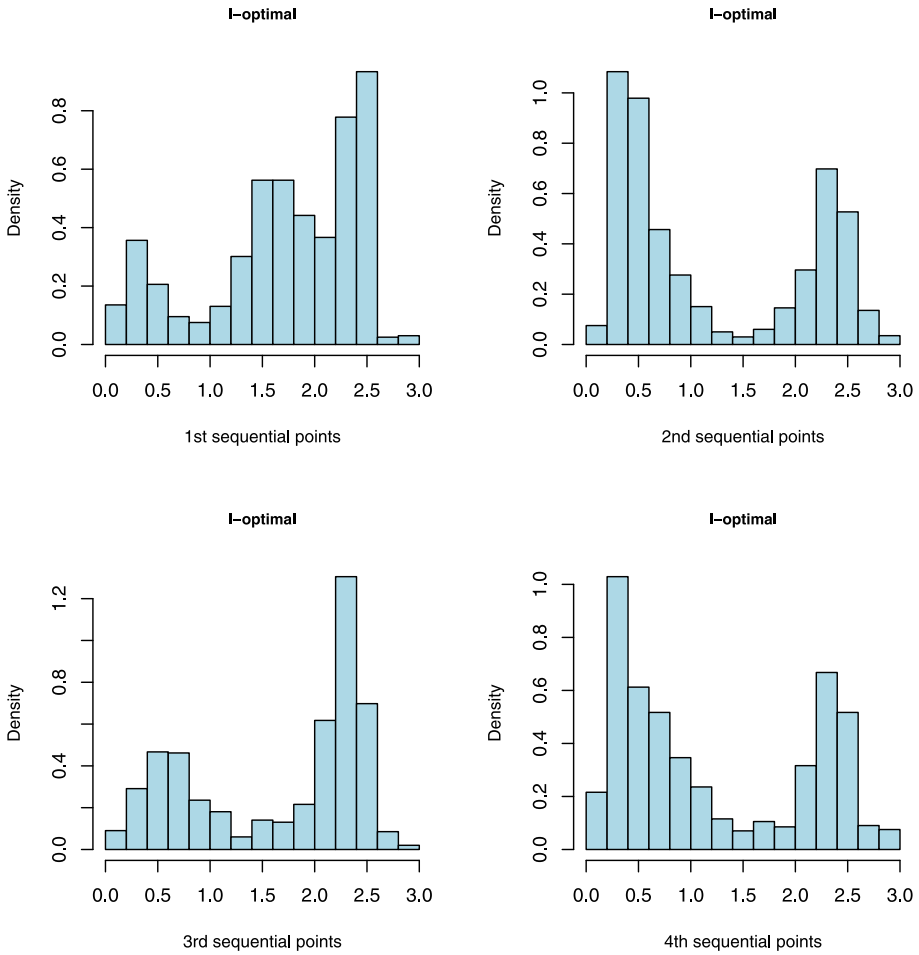

Fig. B.2. Histograms of four sequentially chosen I-optimal design points for 1000 replicates of datasets obtained from a binary mixed model. Initial data were generated from the true logistic model $\text{logit}(p^*_{ij}) = \delta_0 + \delta_1 x_j + \delta_2 x_j^2 + u_i, (i = 1, \ldots, k, j = 1, \ldots, n_0)$, where the fitted model assumes a simple binary model $\text{logit}(p_{ij}) = \beta_0 + \beta_1 x_j + u_i$ with $u_i \sim$ ind. $N(0, \sigma_u^2)$. Parameters were fixed at $(\delta_0, \delta_1, \delta_2) = (-2, \mathbf{0.2}, 0.5)$ and $\sigma_u^2 = 0.25$. Number of clusters $k = \mathbf{25}$.

conditional mean response $E(y_{ij}|u_{ij})$ in the form

$$\begin{aligned} y_{ij}|u_{ij} &\sim \text{ind. Poisson}(\lambda_{ij}), i = 1, \ldots, k, j = 1, \ldots, n_0, \\ \log(\lambda_{ij}) &= \beta_0 + \beta_1 x_j + u_{ij}, \\ u_{ij} &\sim \text{ind. } N(0, \sigma_u^2), \end{aligned} \tag{35}$$

for $k = 3$ and $n_0 = 6$. With the initial data taken as in Table 1 and with awareness of possible inaccuracy in the assumed model, we aim to optimally choose a new design point (dose level) with 3 replications using our proposed robust sequential approach within the design space $x \in [0, 10]$. We note that here the model setting is slightly different than those used in the simulations, where the whole set of clusters was replicated at a given design point. In this example, the given dataset contains no clusters but replicates of observations at different dose levels. Both D-optimal and I-optimal robust sequential designs choose the maximum dose $x = 10$ as the next sequential dose level, as can be seen from Fig. 9 depicting the plot of IMSE loss function over the design space $x \in [0, 10]$. Both loss functions give minimum values at the maximum dose level. It is also interesting to note that although the naive D-optimal design chooses the optimal dose $x = 10$, but the shape of the loss function is different than the I-optimal and D-optimal loss functions.

7. Discussion and conclusion

A primary focus of regression analysis is often on response estimation or prediction. In the presence of uncertainty in the fitted response model as considered in the paper, a robust design strategy should focus on the minimization of the "average prediction error" over a given design space when the marginal mean response $E(y|\mathbf{x})$ is predicted by

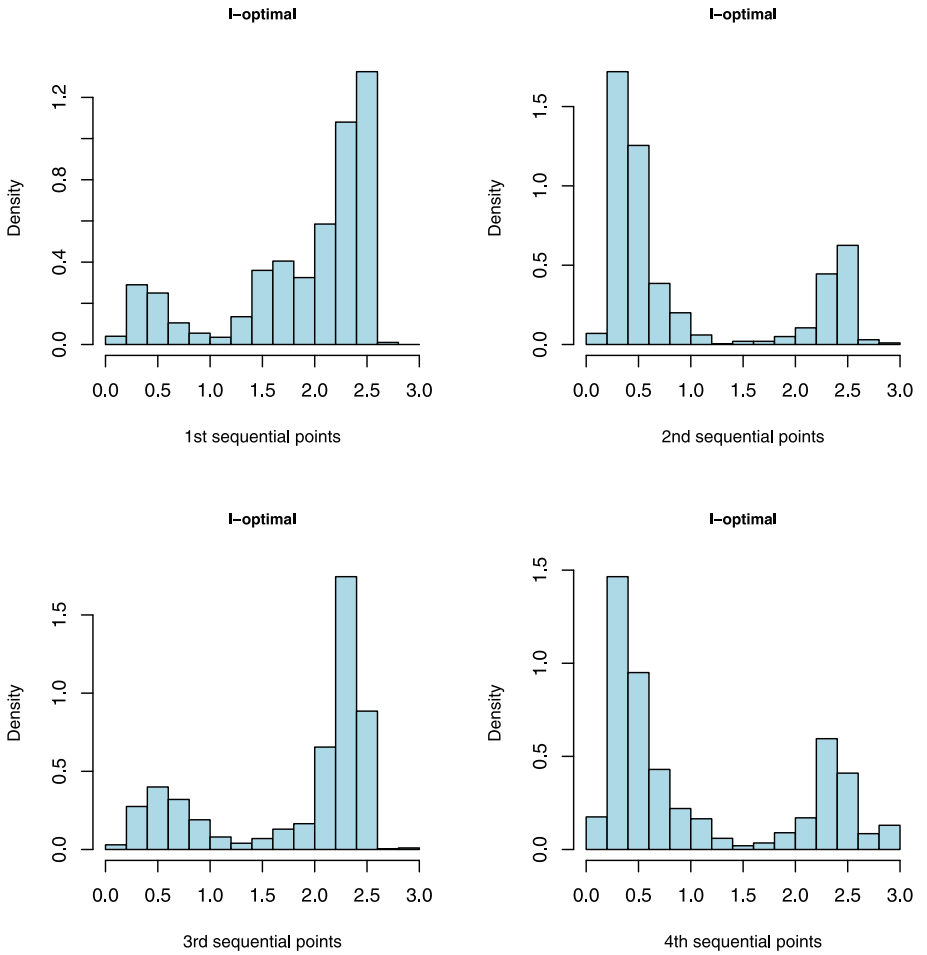

Fig. B.3. Histograms of four sequentially chosen I-optimal design points for 1000 replicates of datasets obtained from a binary mixed model. Initial data were generated from the true logistic model $\text{logit}(p_{ij}^*) = \delta_0 + \delta_1 x_j + \delta_2 x_j^2 + u_i, (i = 1, \ldots, k, j = 1, \ldots, n_0)$, where the fitted model assumes a simple binary model $\text{logit}(p_{ij}) = \beta_0 + \beta_1 x_j + u_i$ with $u_i \sim$ ind. $N(0, \sigma_u^2)$. Parameters were fixed at $(\delta_0, \delta_1, \delta_2) = (-2, \mathbf{0.2}, 0.5)$ and $\sigma_u^2 = 0.25$. Number of clusters $k = \mathbf{50}$.

$\mu(\mathbf{x}, \hat{\boldsymbol{\beta}})$. So we use IMSE as the measure of performance for comparing different designs. It is important to note that the proposed robust I-optimal design criterion is "optimal" under misspecified models. In the case of correctly specified models, however, the D-optimal design still performs better than the I-optimal design as per the IMSE measure, as indicated by Sinha and Wiens (2002).

In our algorithm for the design construction as described in Section 4.2, we point out that the initial data are augmented at each of the new design points chosen sequentially, and parameters are re-estimated for the augmented data after finding each sequential design point. In our simulations, we considered drawing only 4 sequential design points to reduce the computational burden involving a large number of data sets. In practice, for a given set of initial data, we may continue drawing new sequential design points until satisfactory estimates of the model parameters are found.

For both binary and Poisson mixed models as considered in our empirical study, the extent of the model misspecification is not very large, as illustrated in Fig. 1. Standard goodness-of-fit or lack of fit tests often fail to identify such model discrepancies especially when the sample size is small. As noted by Sinha and Wiens (2002), in robust analysis a common strategy is to perform both the robust analysis and a classical analysis that is valid only for the fitted model. A general agreement between these analyses may be taken as evidence in favor of the fitted model. When there is disagreement in these analyses, we can perform model diagnostics to determine where and how the disagreement occurs. The discrepancies used in the estimation of MSE may be thought of as indicators of model misspecification. At each stage of the sequential design, we are able to assess the extent of model misspecification from the estimated discrepancies. In case of correctly specified models, the estimated discrepancies should be close to zero. Large estimated discrepancies would lead us to think of an alternative model. However, here our goal is to adopt a robust design that leads to a minimum loss when the fitted model is approximately correct.

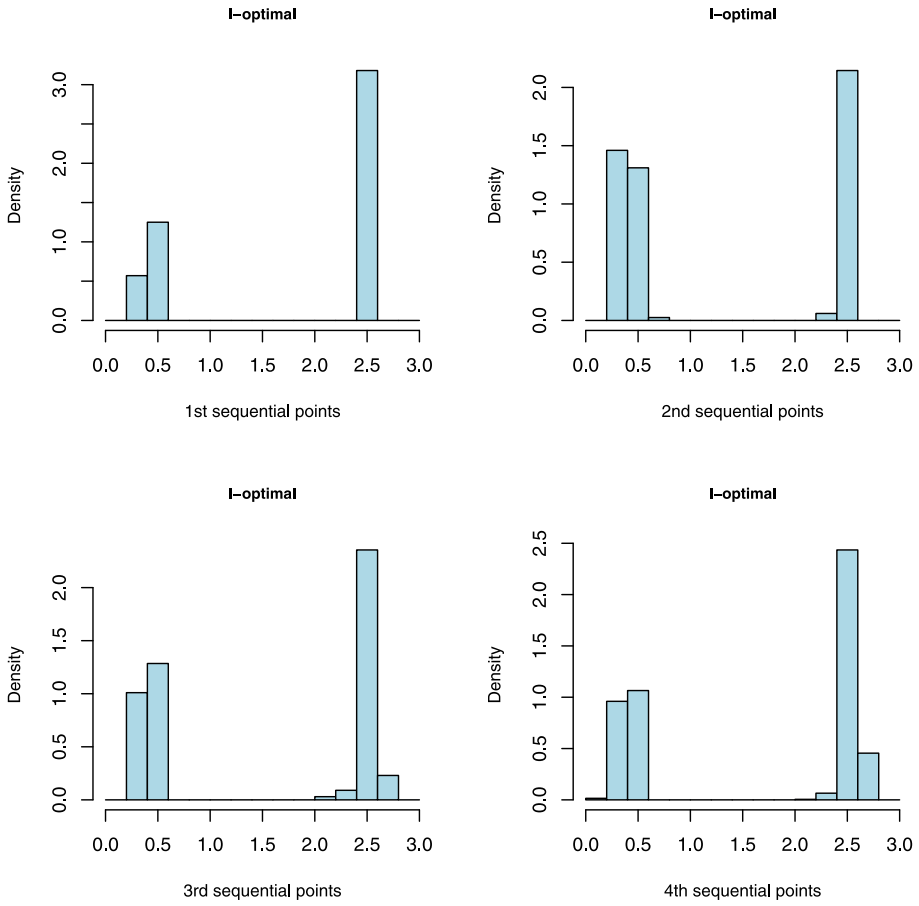

Fig. B.4. Histograms of four sequentially chosen I-optimal design points for 1000 replicates of datasets obtained from a Poisson mixed model. Initial data were generated from the true Poisson model $\log(\lambda_{ij}^{*}) = \delta_0 + \delta_1 x_j + \delta_2 x_j/(1+x_j) + u_i, (i = 1, \ldots, k, j = 1, \ldots, n_0)$, where the fitted model assumes a simple Poisson model $\log(\lambda_{ij}) = \beta_0 + \beta_1 x_j + u_i$ with $u_i \sim$ ind. $N(0, \sigma_u^2)$. Parameters were fixed at $(\delta_0,\ \delta_1, \delta_2) = (0.5, \mathbf{0.1}, 3)$ and $\sigma_u^2 = 0.25$. Number of clusters $k = \mathbf{50}$.

We have studied the performance of the D- and I-optimal sequential designs for analyzing generalized linear mixed models with possible misspecification of the mean response function. Both design criteria require intensive computation. Some approximate method has been investigated to reduce the computational difficulties. We have carried out simulations for both binary and Poisson mixed models. The simulation results have shown that the I-optimal design is generally more efficient than the D-optimal design for all scenarios under misspecified logistic models. Also, the I-optimal design appears to be as efficient as or more efficient than the D-optimal design under misspecified Poisson regression models. The efficiencies of the proposed robust designs, their corresponding classical optimal designs obtained without awareness of model misspecifications, and the sequential uniform designs have also been compared through the simulations. The results have demonstrated that one could attain considerable gain in efficiency from the maximum likelihood estimators when data are augmented with the proposed I-optimal robust sequential design scheme rather than the classical D-optimal designs and the conventionally used uniform designs.

We also investigated the "classical" or naive I-optimal design without regard for the bias or robustness when analyzing binary and Poisson mixed models. Our study (not shown here) suggests that the naive I-optimal design is generally less efficient than the naive D-optimal design under misspecified response models.

In our empirical study with simulations, the results of the Poisson mixed model appeared to be more clear cut as compared to the binary mixed model. This is probably due to the fact that for finite samples, the ML estimators of the Poisson regression model for count data are generally more stable than those of the logistic regression model for binary data in terms of the variability in estimation.

In this paper, we have explored the sequential design schemes in the case of a one-dimensional design space. The proposed I-optimal and D-optimal designs require intensive computation involving integrations with respect to the random effects. To reduce the computational burden, we used a slightly modified version of the two design criteria, as described in Section 4.2. This approximation drastically reduced the computation time (roughly by 90%) when conducting

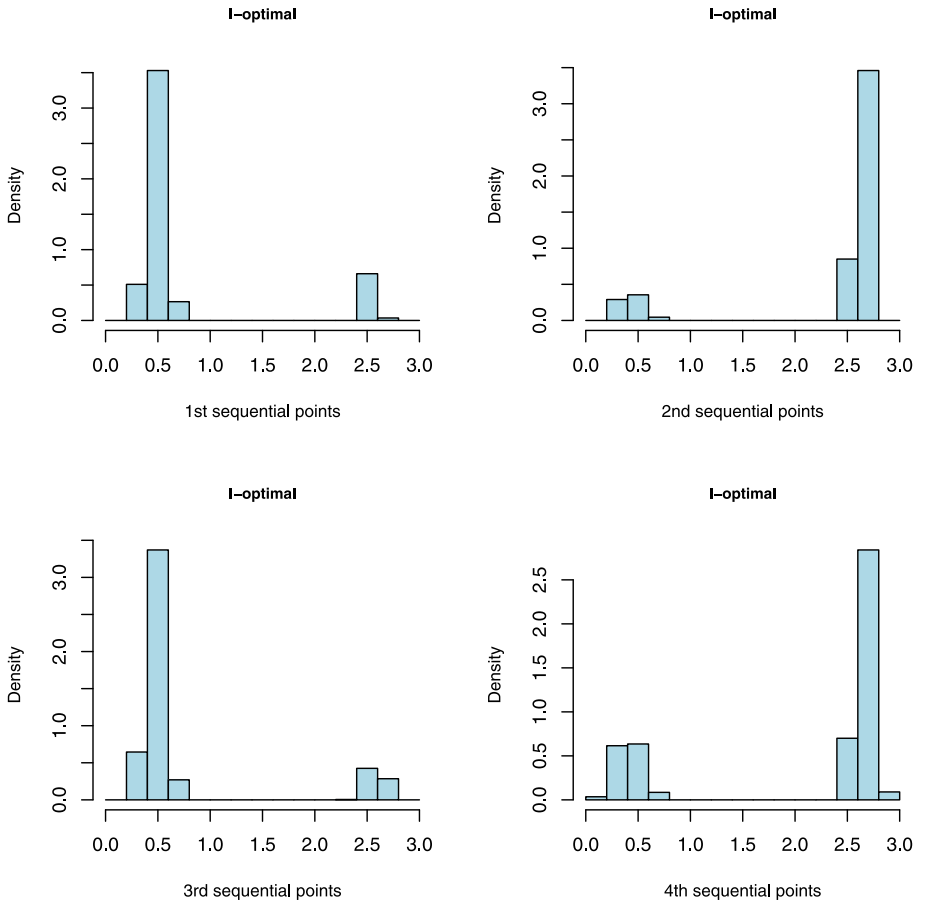

Fig. B.5. Histograms of four sequentially chosen I-optimal design points for 1000 replicates of datasets obtained from a Poisson mixed model. Initial data were generated from the true Poisson model $\log(\lambda_{ij}^{*}) = \delta_0 + \delta_1 x_j + \delta_2 x_j/(1+x_j) + u_i, (i = 1, \ldots, k, j = 1, \ldots, n_0)$, where the fitted model assumes a simple Poisson model $\log(\lambda_{ij}) = \beta_0 + \beta_1 x_j + u_i$ with $u_i \sim$ ind. $N(0, \sigma_u^2)$. Parameters were fixed at $(\delta_0,\ \delta_1, \delta_2) = (0.5, \mathbf{0.3}, 3)$ and $\sigma_u^2 = 0.25$. Number of clusters $k = \mathbf{25}$.

our simulation study. The proposed design scheme can be extended for choosing multidimensional sequential designs in a similar manner. The choice of multidimensional design naturally would require more computation, as it involves the calculation of the Fisher information at each possible location in the multidimensional design space.

CRediT authorship contribution statement

Xiaojian Xu: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Validation, Visualization, Writing - original draft. **Sanjoy K. Sinha:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Validation, Visualization.

Acknowledgment

The research of both authors is supported by grants from the Natural Sciences and Engineering Research Council of Canada.

Appendix A. Proof of Theorem 1

Under conditions as in Fahrmeir (1990), the maximum likelihood estimator exists and is consistent for GLMs, and the score function is $o_p\left(n^{-1/2}\right)$. This conclusion applies to the maximum likelihood estimators obtained by the marginal log-likelihood function l as defined in (22) for GLMMs specified by (1) and (2). Recall that in the case of the GLMM, the score function is given by

$$l'(\boldsymbol{\beta}) = \frac{\partial l(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n} E_{u|y}\left(\frac{y_i - \mu_i^u}{var(y_i|\mathbf{u})} \frac{\partial \mu_i^u}{\partial \eta_i} \middle| \mathbf{y} \right) \mathbf{r}(\mathbf{x}_i).$$

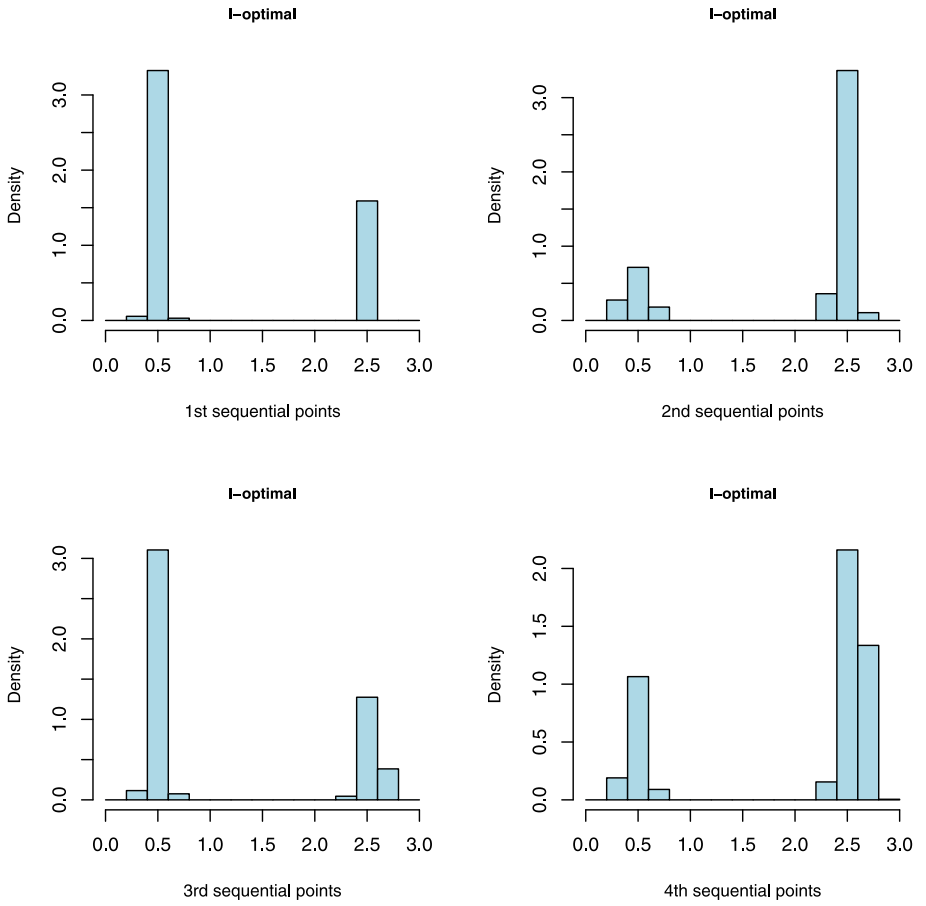

Fig. B.6. Histograms of four sequentially chosen I-optimal design points for 1000 replicates of datasets obtained from a Poisson mixed model. Initial data were generated from the true Poisson model $\log(\lambda_{ij}^*) = \delta_0 + \delta_1 x_j + \delta_2 x_j/(1+x_j) + u_i, (i = 1, \ldots, k, j = 1, \ldots, n_0)$, where the fitted model assumes a simple Poisson model $\log(\lambda_{ij}) = \beta_0 + \beta_1 x_j + u_i$ with $u_i \sim$ ind. $N(0, \sigma_u^2)$. Parameters were fixed at $(\delta_0,\ \delta_1, \delta_2) = (0.5, \mathbf{0.3}, 3)$ and $\sigma_u^2 = 0.25$. Number of clusters $k = \mathbf{50}$.

An expansion of $l_j'(\boldsymbol{\beta}) = \partial l(\boldsymbol{\beta})/\partial\beta_j$ around $\boldsymbol{\beta}_0$ gives

$$l_j'(\boldsymbol{\beta}) = l_j'(\boldsymbol{\beta}_0) + \sum_k (\beta_k - \beta_{0,k})\, l_{jk}''(\boldsymbol{\beta}_0) + \frac{1}{2}\sum_k\sum_l (\beta_k - \beta_{0,k})(\beta_l - \beta_{0,l})\, l_{jkl}'''(\boldsymbol{\beta}_*),$$

where $l_j'(\boldsymbol{\beta}_0)$ is the value of $l_j'(\boldsymbol{\beta})$ evaluated at $\boldsymbol{\beta}_0$,

$$l_{jk}''(\boldsymbol{\beta}) = \frac{\partial^2 l(\boldsymbol{\beta})}{\partial\beta_j\partial\beta_k},$$
$$l_{jkl}'''(\boldsymbol{\beta}) = \frac{\partial^3 l(\boldsymbol{\beta})}{\partial\beta_j\partial\beta_k\partial\beta_l},$$

β_j and $\beta_{0,j}$ are the jth terms of the vectors $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_0$, respectively, and $\boldsymbol{\beta}_*$ is a point on the line segment connecting $\boldsymbol{\beta}$ and $\boldsymbol{\beta}_0$. If we replace $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}$, the above expansion gives

$$\sqrt{n}\sum_k \left(\hat{\beta}_k - \beta_{0,k}\right)\left[\frac{1}{n} l_{jk}''(\boldsymbol{\beta}_0) + \frac{1}{2n}\sum_l \left(\hat{\beta}_l - \beta_{0,l}\right) l_{jkl}'''(\boldsymbol{\beta}_*)\right] = -\frac{1}{\sqrt{n}} l_j'(\boldsymbol{\beta}_0).$$

For the exponential family, the third derivatives $l_{jkl}'''(\boldsymbol{\beta}_*)$ are bounded (Kredler, 1986). Using the consistency of $\hat{\boldsymbol{\beta}}$, we have

$$\frac{1}{n} l_{jk}''(\boldsymbol{\beta}_0) + \frac{1}{2n}\sum_l \left(\hat{\beta}_l - \beta_{0,l}\right) l_{jkl}'''(\boldsymbol{\beta}_*) \xrightarrow{p} -H_{jk},$$

where H_{jk} is the (j,k)th element of the matrix $\mathbf{H}_n(\boldsymbol{\beta}_0) = -(1/n)E_y[l''(\boldsymbol{\beta}_0)]$. Thus the limiting distribution of $\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)$ is that of the solution of the equations

$$\sum_k H_{jk}\sqrt{n}\left(\hat{\beta}_k - \beta_{0,k}\right) = \frac{1}{\sqrt{n}} l'_j(\boldsymbol{\beta}_0),$$

i.e., the limiting distribution of $\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right)$ is given by $\mathbf{H}_n^{-1}(\boldsymbol{\beta}_0)(1/\sqrt{n})l'(\boldsymbol{\beta}_0)$. Using the central limit theorem, we can argue that $(1/\sqrt{n})l'(\boldsymbol{\beta}_0)$ has an approximate multivariate normal distribution with the mean

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} E_y\left[E_{u|y}\left(\frac{y_i - \mu_i^u}{var(y_i|\mathbf{u})}\frac{\partial \mu_i^u}{\partial \eta_i}\middle|\mathbf{y}\right)\right]\mathbf{r}(\mathbf{x}_i) = \sqrt{n}\mathbf{b}_n\left(\boldsymbol{\beta}_0\right),$$

and variance–covariance matrix $\mathbf{H}_n(\boldsymbol{\beta}_0)$. It follows that

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) \sim AN\left(\sqrt{n}\mathbf{H}_n^{-1}(\boldsymbol{\beta}_0)\mathbf{b}_n(\boldsymbol{\beta}_0), \mathbf{H}_n^{-1}(\boldsymbol{\beta}_0)\right).$$

Appendix B. Additional plots from simulations: Figs. B.1–B.6

See Figs. B.1–B.6.

References

Adewale, A.J., Wiens, D.P., 2009. Robust designs for misspecified logistic models. J. Statist. Plann. Inference 139, 3–15.

Adewale, A.J., Xu, X., 2010. Robust designs for generalized linear models with possible overdispersion and misspecified link functions. Comput. Statist. Data Anal. 54, 875–890.

Atkinson, A.C., Donev, A.N., Tobias, R.D., 2007. Optimum Experimental Designs, with SAS. Oxford University Press, Oxford.

Biedermann, S., Dette, H., Pepelyshev, A., 2006. Some robust design strategies for percentile estimation in binary response models. Canad. J. Statist. 34, 603–622.

Breslow, N.E., 1984. Extra-Poisson variation in log-linear models. J. R. Stat. Soc. Ser. C. Appl. Stat. 33, 38–44.

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

Dror, H.A., Steinberg, D.M., 2008. Sequential experimental designs for generalized linear models. J. Amer. Statist. Assoc. 92, 162–170.

Fahrmeir, L., 1990. Maximum likelihood estimation in misspecified generalized linear models. Statistics 21, 487–502.

Heagerty, P.J., Kurland, B.F., 2001. Misspecified maximum likelihood estimates and generalised linear mixed models. Biometrika 88, 973–985.

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

Kredler, C., 1986. Behaviour of third order terms in quadratic approximations of LR-statistics in multivariate generalized linear models. Ann. Statist. 14, 326–335.

Ponce de Leon, A.M., Atkinson, A.C., 1993. Designing optimal experiments for the choice of link function for a binary data model. In: Muller, W.G., Wynn, H.P., Zhigljavsky, A.A. (Eds.), Model- Oriented Data Analysis. Physica-Verlag, Heidelberg, pp. 25–36.

Li, P., Wiens, D.P., 2011. Robustness of design in dose–response studies. J. R. Stat. Soc. Ser. B Stat. Methodol. 73, 215–238.

Maram, P.P., Jafari, H., 2016. Bayesian D-optimal design for logistic regression model with exponential distribution for random intercept. J. Comm. Statist. - Theory Methods 43, 1234–1247.

McCulloch, C.E., 1997. Maximum likelihood algorithms for generalized linear mixed models. J. Amer. Statist. Assoc. 103, 288–298.

McCulloch, C.E., Searle, S.R., 2001. Generalized, Linear, and Mixed Models. John Wiley, New York.

Pregibon, D., 1980. Goodness of link selection in generalized linear models. Appl. Stat. 29, 15–24.

Sinha, S.K., Wiens, D.P., 2002. Robust sequential designs for nonlinear regression. Canad. J. Statist. 30, 601–618.

Sinha, S.K., Xu, X., 2011. Sequential optimal designs for generalized linear mixed models. J. Statist. Plann. Inference 141, 1394–1402.

Sinha, S.K., Xu, X., 2016. Sequential designs for repeated measures experiments. J. Stat. Theory Pract. 10, 497–514.

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

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

Zhang, Y., Ye, K., 2014. Bayesian d-optimal designs for Poisson regression models. Communications in Statistics – Theory and Methods 43, 1234–1247.