Statistics and Computing (2023) 33:114
https://doi.org/10.1007/s11222-023-10279-3

ORIGINAL PAPER

Optimal designs for generalized linear mixed models based on the penalized quasi-likelihood method

Yao Shi[1] · Wanchunzi Yu[2] · John Stufken[3]

Received: 10 December 2022 / Accepted: 3 July 2023

Abstract
While generalized linear mixed models are useful, optimal design questions for such models are challenging due to complexity of the information matrices. For longitudinal data, after comparing three approximations for the information matrices, we propose an approximation based on the penalized quasi-likelihood method. We evaluate this approximation for logistic mixed models with time as the single predictor variable. Assuming that the experimenter controls at which time observations are to be made, the approximation is used to identify locally optimal designs based on the commonly used A- and D-optimality criteria. The method can also be used for models with random block effects. Locally optimal designs found by a Particle Swarm Optimization algorithm are presented and discussed. As an illustration, optimal designs are derived for a study on self-reported disability in older women. Finally, we also study the robustness of the locally optimal designs to mis-specification of the covariance matrix for the random effects.

Keywords Locally optimal design · Longitudinal study · Logistic model · Penalized quasi-likelihood · Robustness

1 Introduction

Many longitudinal studies are designed to investigate a characteristic of a subject, with the characteristic being measured repeatedly over time. If the response to measure this characteristic is categorical, then generalized linear mixed models (GLMMs) may be the appropriate choice. For instance, for a binary response variable, logistic and probit mixed-effects models are frequently used, while Poisson mixed-effects models can be used for a count variable.

At which time points should measurements be made? Answers to this optimal design question are important since it can improve the precision of parameter estimations. Optimal designs for GLMMs are studied in the literature, such as optimal designs for Poisson mixed-effects models (Niaparast 2009), for logistic mixed-effects models under a maximin D-optimality criterion (Tekle et al. 2008), and under Bayesian D-optimality (Abebe et al. 2014). Waite and Woods (2015) proposed locally D-optimal and Bayesian D-optimal designs for GLMMs with random intercept in a block design, through marginal quasi-likelihood (MQL) and an outcome-enumeration method. Ueckert and Mentré (2017) applied Monte Carlo and adaptive Gaussian quadrature method to approximate the Fisher information matrix under discrete mixed effect models, which is then proved to be better than MQL. Sequential D-optimal designs for GLMMs were investigated by Sinha and Xu (2011). Robustness analysis to incorrect specification of the covariance matrix of random effects also matters, because this matrix is generally unknown at the design stage.

The reason why approximations are needed is the unavailability of a closed-form expression for the information matrix under a GLMM. This is also true for many other mixed effect longitudinal models, where responses from the same subject are correlated, like discrete-time survival models with random effects. Zhou et al. (2021) applied the MQL method in

✉ Yao Shi
yshi97@asu.edu

Wanchunzi Yu
wyu@bridgew.edu

John Stufken
jstufken@gmu.edu

[1] School of Mathematics and Statistics, Qingdao University, 308 Ningxia Road, Qingdao 266071, Shandong, China

[2] Department of Mathematics, Bridgewater State University, 131 Summer Street, Bridgewater, MA 02325, USA

[3] Department of Statistics, George Mason University, 4400 University Drive, Fairfax, VA 22030, USA

Published online: 30 July 2023

approximating the information matrix under such a model and derived optimal designs.

Section 2 presents a GLMM that, while not of the form in Molenberghs and Verbeke (2005), is known as a random coefficients model, such as a random intercept or random slope model (Schmelter 2007). We illustrate the penalized quasi-likelihood (PQL) approximation to information matrices for GLMMs. In Sect. 3, for a logistic model with a two-parameter linear predictor, we assess the accuracy of this approximation by comparison to exact information matrices. In Sect. 4, using the approximation, locally D- and A-optimal designs are found for different local parameter settings under this logistic model. We study a real-life example on self-reported disability among older women (Carrière and Bouyer 2002) to identify optimal designs, and then discuss robustness of locally optimal designs in Sect. 5. Section 6 presents a summary and discussion of future work.

2 Theoretical framework

The general form for a generalized linear model (GLM) with a link function $\eta(\cdot)$ is

$$\mu = E(y) = \eta^{-1}(\mathbf{f}(x)^T\boldsymbol{\beta}). \tag{1}$$

When using the maximum likelihood approach for estimating the model parameters, as is commonly the case because of its asymptotic efficiency, the asymptotic variance-covariance matrix of the estimators is given by the inverse of the Fisher information matrix. In view of this, the Fisher information matrix is of great interest for comparing different designs. One would like to maximize some function of this information matrix. With a canonical link function and under the assumption that responses are independent, for an exact design $\xi = \{(x_l), l = 1, \ldots, k\}$, the information matrix for $\boldsymbol{\beta}$ is $M(\xi) = \mathbf{F}^T\mathbf{V}\mathbf{F}$, where $\mathbf{F} = (\mathbf{f}(x_1), \ldots, \mathbf{f}(x_k))^T$, and $\mathbf{V} = \text{diag}(\text{var}(y_1), \ldots, \text{var}(y_k))$. For a longitudinal study with N subjects, denote $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^T$ as the response vector for subject i, $i = 1, \ldots, N$, using design $\xi_i = \{(x_{il}), l = 1, \ldots, n_i\}$, where n_i is the number of measurements on subject i and $x_{i1}, \ldots, x_{in_i}$ are the time points at which observations are made for that subject. The total number of observations over all subjects is then $K = \sum_{i=1}^{N} n_i$. Extending the GLM to a GLMM, the conditional mean for y_{ij}, $j = 1, \ldots, n_i$, is

$$\mu_{ij}^{\mathbf{b}_i} = E(y_{ij}|\mathbf{b}_i) = \eta^{-1}(\mathbf{f}(x_{ij})^T\mathbf{b}_i), \tag{2}$$

where $\mathbf{f}(x_{ij})$ and $\mathbf{b}_i$ are vectors of length q, $\mathbf{b}_i = (b_{i0}, b_{i1}, \ldots, b_{i,q-1})^T = \boldsymbol{\beta} + \boldsymbol{\alpha}_i$ consists of the subject-independent fixed-effects vector $\boldsymbol{\beta}$ and the random vector $\boldsymbol{\alpha}_i \sim N_q(0, \boldsymbol{\Sigma})$, i.e., $\mathbf{b}_i \sim N_q(\boldsymbol{\beta}, \boldsymbol{\Sigma})$. The covariance matrix $\boldsymbol{\Sigma}$ can be singular, allowing some effects to be fixed effects. The conditional variance of y_{ij} given $\mathbf{b}_i$ is $v_{ij}^{\mathbf{b}_i} = \phi a v(\mu_{ij}^{\mathbf{b}_i})$, where $v(\cdot)$ is a known function depending on $\eta(\cdot)$, a is a known constant, and ϕ is a dispersion parameter. As an example, for a binary response with the logistic link,

$$\begin{aligned}\eta[E(y_{ij}|\mathbf{b}_i)] &= \eta\left(P(y_{ij} = 1 \mid \mathbf{b}_i)\right)\\ &= \log\left(\frac{P(y_{ij} = 1 \mid \mathbf{b}_i)}{1 - P(y_{ij} = 1 \mid \mathbf{b}_i)}\right),\end{aligned} \tag{3}$$

with the corresponding conditional variance being

$$v_{ij}^{\mathbf{b}_i} = \text{var}(y_{ij}|\mathbf{b}_i) = \frac{\exp(\mathbf{f}(x_{ij})^T\mathbf{b}_i)}{\{1 + \exp(\mathbf{f}(x_{ij})^T\mathbf{b}_i)\}^2}. \tag{4}$$

We mainly focus on this logistic link in simulations, but the methodology is suitable for other links as well.

Model (2) can also be used for experiments that fit a GLM (with fixed effects) but that are run in a block design with random block effects (Waite and Woods 2015). Thus, N now denotes the number of blocks, and the matrix $\boldsymbol{\Sigma}$ is a singular matrix with only the first element being nonzero. In contrast to a longitudinal study, where the x_{il}'s denote time points, in this block design setting, the values of $x_{i1}, \ldots, x_{in_i}$ need not all be distinct. For now, we refrain from requiring that all values must be different, but we will return to this at the end of Sect. 4. For simplicity, we will continue to refer to subjects rather than blocks.

For estimating $\boldsymbol{\beta}$, we seek a design $\xi = (\xi_1, \ldots, \xi_N)$ that is locally optimal for this goal. The likelihood function for $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ is given by

$$\begin{aligned}L_{Total}(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\mathbf{y}) &= \prod_{i=1}^{N} L_i(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\mathbf{y}_i)\\ &= \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} h(y_{ij}|\mathbf{b}_i = \boldsymbol{\beta} + \boldsymbol{\alpha}_i) g(\boldsymbol{\alpha}_i|\boldsymbol{\Sigma}) d\boldsymbol{\alpha}_i,\end{aligned}$$

where $h(y_{ij}|\mathbf{b}_i)$ is the conditional density function of y_{ij} and $g(\boldsymbol{\alpha}_i|\boldsymbol{\Sigma})$ is the density function of random effects $\boldsymbol{\alpha}_i$ (Molenberghs and Verbeke 2005). For simplicity, we will just write $L_i(\boldsymbol{\beta}, \boldsymbol{\Sigma})$ for $L_i(\boldsymbol{\beta}, \boldsymbol{\Sigma}|\mathbf{y}_i)$. An approximation to, or a surrogate for, the information matrix for $\boldsymbol{\beta}$ can be obtained by approximating and inverting $\text{cov}(\hat{\boldsymbol{\beta}})$ based on an estimation method for $\boldsymbol{\beta}$. In general, the parameter estimation for Model (2) is difficult, and there are no analytical expressions for $\text{cov}(\hat{\boldsymbol{\beta}})$. Niaparast (2009) discussed the optimal design based on quasi-likelihood estimation under a Poisson regression model with random intercept. Niaparast and Schwabe (2013) and Niaparast et al. (2023) extended the results to random slopes. Other methods for approximating $\text{cov}(\hat{\boldsymbol{\beta}})$ include generalized estimating equations (GEE, Liang and

Zeger 1986, cf. Atkinson and Woods 2015), PQL (Breslow and Clayton 1993), and MQL (Breslow and Clayton 1993; cf. Atkinson and Woods 2015). Ueckert and Mentré (2017) also applied adaptive Gaussian quadrature to approximate the information matrix directly. Mielke (2012) studied multiple approximations of the Fisher information under nonlinear mixed models.

The complicated form of the likelihood function makes it impossible to find closed-form expressions for the maximum likelihood estimators (MLEs) of the parameters in Model (2). Nonetheless, given a design and data, the MLEs can be obtained numerically. The Fisher information matrix depends on unknown parameters, but even with intelligent guesses for those parameters (e.g., based on prior experience or a pilot study), it must be evaluated numerically due to its complicated nature. Doing this once is fine, but when searching for an optimal or efficient design, this must be done for many designs, and becomes computationally inefficient. In order to avoid this, we seek an approximation of the Fisher information that is sufficiently simple so that it can be used in a design search. Besides simplicity of this approximation, we would also like it to be, approximately, consistent in ranking designs with the ranking that would have been obtained if we had been able to use the Fisher information to do this. One such simpler alternative comes from PQL estimation, as studied in Breslow and Clayton (1993). They defined an integrated quasi-likelihood function for estimating $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, and then, using the Laplace method for integral approximation and Green (1987)'s result for a likelihood function with a penalty term, simplified it to obtain the PQL approximation. Like MLE estimators, PQL estimators of $\boldsymbol{\beta}$ are consistent and asymptotically normal, with a variance-covariance matrix that can be viewed as an approximation of the inverse of the Fisher information matrix. Motivated by the results in Sect. 3, which suggest that this is a good approximation, our optimal design search will be based on PQL estimation rather than MLE estimation. The designs found in this way should however also perform well with MLE estimation.

For PQL estimation of $\boldsymbol{\beta}$, the approximate conditional variance-covariance matrix given the random effects $\mathbf{b}$, $\text{cov}(\hat{\boldsymbol{\beta}} \mid \mathbf{b})$, under Model (2) is given by (Breslow and Clayton 1993; Tekle et al. 2008)

$$\text{cov}(\hat{\boldsymbol{\beta}} \mid \mathbf{b}) \approx \left(\sum_{i=1}^{N} \mathbf{F}_i^T \mathbf{U}_i^{-1} \mathbf{F}_i \right)^{-1}, \tag{5}$$

where $\mathbf{b} = (\mathbf{b}_1^T, \ldots, \mathbf{b}_N^T)^T$, $\mathbf{U}_i = \mathbf{V}_i^{-1} + \mathbf{F}_i \boldsymbol{\Sigma} \mathbf{F}_i^T$ is a weight matrix for the ith subject, $\mathbf{F}_i$ is the design matrix for design ξ_i, i.e., $\mathbf{F}_i = \mathbf{F}_i(\xi_i) = (\mathbf{f}(x_{i1}), \ldots, \mathbf{f}(x_{in_i}))^T$. Observations from different subjects are assumed to be independent, while observations from the same subject are considered to be conditionally independent given the random effects. Here, we also assume a canonical link function, although, as discussed by Breslow and Clayton (1993), PQL estimation remains valid for other link functions. With these assumptions, $\mathbf{V}_i = \mathbf{V}_i(\xi_i) = \text{diag}(v_{i1}^{\mathbf{b}_i}, \ldots, v_{in_i}^{\mathbf{b}_i})$. Design specification requires choices for the number of distinct covariate values x_{ij} for each subject, the values of these covariates, and the number of measurements at each of these values. Since $\text{cov}(\hat{\boldsymbol{\beta}})$ in (5) depends on the unknown $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, we will substitute values based on prior knowledge. This will eventually result in locally optimal designs. An approximation of the conditional Fisher information, denoted by $\mathfrak{M}_{Total}(\xi \mid \boldsymbol{b})$, is obtained by taking the inverse of the matrix in Equation (5). Taking its expectation with respect to $\mathbf{b}$,

$$E_{\mathbf{b}}(\mathfrak{M}_{Total}(\xi \mid \mathbf{b})) = \sum_{i=1}^{N} E_{\mathbf{b}}(\mathbf{F}_i^T (\mathbf{V}_i^{-1} + \mathbf{F}_i \boldsymbol{\Sigma} \mathbf{F}_i^T)^{-1} \mathbf{F}_i) \tag{6}$$

gives the expectation of this PQL-based approximation to the information matrix. It becomes the information matrix for the corresponding fixed-effects model when $\boldsymbol{\Sigma}$ goes to 0.

Since some subjects may receive the same design, it is convenient to change the notation for a design from $\xi = (\xi_1, \ldots, \xi_N)$ to $\xi = \{(\xi_p, m_p), p = 1, \ldots, N_s\}$, where $\xi_p = \{(x_{pl}), l = 1, \ldots, n_p\}$, $N_s \leq N$ is the number of distinct designs ξ_p, m_p is the number of subjects receiving ξ_p, and $\sum_p m_p = N$. At this stage we also switch from exact designs to approximate designs, so that we can ignore N. We represent the design by $\xi = \{(\xi_p, w_p), p = 1, \ldots, N_s\}$, where $\sum_p w_p = 1$. If we start from an exact design for N subjects in which sequence ξ_p is used m_p times, then it is converted to an approximate design by taking $w_p = m_p/N$. The design ξ is also known as the population design, while the ξ_p's, which remain exact, are known as individual designs. One could use the same individual design for each subject, i.e., $N_s = 1$. But as noted by Schmelter (2007), for exact individual designs, optimal population designs may use more than a single individual design. The number of measurements on a subject, n_p, is in practice often the same for all subjects, and we will assume from now on that $n_p = n$. Based on (6), in the approximate design setting we define

$$\begin{aligned} \mathfrak{M}^{PQL}(\xi) &= \mathfrak{M}^{PQL}(\xi \mid \boldsymbol{\beta}, \boldsymbol{\Sigma}) \\ &= \sum_{p=1}^{N_s} w_p E_{\mathbf{b}}(\mathbf{F}_p^T (\mathbf{V}_p^{-1} + \mathbf{F}_p \boldsymbol{\Sigma} \mathbf{F}_p^T)^{-1} \mathbf{F}_p) \end{aligned} \tag{7}$$

as an approximation to the information matrix for $\boldsymbol{\beta}$ under population design ξ. Each expectation in (7) can be approximated by an average based on a representative sample from $N(\boldsymbol{\beta}, \boldsymbol{\Sigma})$, say $\{\mathbf{b}_p^{i_s}\}_{i_s=1}^{S}$, $p = 1, \ldots, N_s$.

For a balance between running time and accuracy, a judicious selection of the samples $\{\mathbf{b}_p^{i_s}\}_{i_s=1}^{S}$ is important. Tekle et al. (2008) suggested using $S = 500$ with random samples. We will instead use the support points method introduced by Mak and Joseph (2018), which provides stable estimates of the expectations with a much smaller sample size. We will return to this in Sect. 3.

Computations for the expression in (7) can be further simplified by the following lemma.

Lemma 1 (Miller 1981) *Let A and $A+B$ be invertible matrices, with the rank of B equal to $r > 0$. Let $B = B_1 + \cdots + B_r$, where each B_i has rank 1 and, for $k = 1, \ldots, r$, each $C_{k+1} = A + \sum_{i=1}^{k} B_i$ is invertible. Setting $C_1 = A$, then*

$$C_{k+1}^{-1} = C_k^{-1} - g_k C_k^{-1} B_k C_k^{-1},$$

where $g_k = \frac{1}{1+tr(C_k^{-1}B_k)}$. In particular,

$$(A+B)^{-1} = C_r^{-1} - g_r C_r^{-1} B_r C_r^{-1}.$$

To apply Lemma 1, if $\boldsymbol{\Sigma}$ is full rank, we simply partition $\boldsymbol{\Sigma}$ into a sum of $\boldsymbol{\Sigma}_i$'s, where $\boldsymbol{\Sigma}_i$ is the matrix obtained from $\boldsymbol{\Sigma}$ by replacing all its elements by 0 except for those in the ith row. Then, the inverse of $\mathbf{V}_p^{-1} + \mathbf{F}_p \boldsymbol{\Sigma} \mathbf{F}_p^T$ in (7) can be computed by multiplication and summation of matrices, which is much faster than computing an inverse. For the case where the rank $\boldsymbol{\Sigma}$ is $R(\boldsymbol{\Sigma}) < q$, we can do an eigen decomposition to get $R(\boldsymbol{\Sigma})$ matrices and apply the Lemma.

The contribution to the likelihood function by a single subject can also be obtained by enumerating all possible outcomes. For the logistic link in Model (2), the likelihood function for a subject, say i, who is assigned to individual design ξ_p is

$$L_i(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \int \prod_{l=1}^{n} \frac{\exp[\mathbf{f}(x_{pl})^T \mathbf{b}_i]^{y_{il}}}{1+\exp[\mathbf{f}(x_{pl})^T \mathbf{b}_i]} \Phi(\mathbf{b}_i; \boldsymbol{\beta}, \boldsymbol{\Sigma}) d\mathbf{b}_i.$$

This corresponds to a contribution to the information matrix given by

$$\begin{aligned} &\mathfrak{I}_i(\xi_p | \boldsymbol{\beta}, \boldsymbol{\Sigma}) \\ &= E_{\mathbf{y}_i}\left(\frac{\partial \log L_i(\boldsymbol{\beta}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\beta}} \frac{\partial \log L_i(\boldsymbol{\beta}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\beta}'} \right) \\ &= \sum_{\mathbf{y}_i \in \{0,1\}^n} \frac{1}{L_i(\boldsymbol{\beta}, \boldsymbol{\Sigma})} \frac{\partial L_i(\boldsymbol{\beta}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\beta}} \frac{\partial L_i(\boldsymbol{\beta}, \boldsymbol{\Sigma})}{\partial \boldsymbol{\beta}'}. \end{aligned} \tag{8}$$

Using numerical integration, this expression can serve as a benchmark for other approximations of the information matrix. We will pursue this in Sect. 3.

3 Comparison of approximations

Without a closed-form expression for an information matrix, finding a design that optimizes a function of the information matrix is computationally challenging. If we can use a reliable and computationally simpler approximation or surrogate of the information matrix for comparing different designs, then this could be used for finding optimal designs. We conduct a comparison using the D- and A-optimality criteria to study different approximations. The D-optimality criterion maximizes $\phi_D(\xi) = det(\mathfrak{M})^{1/q}$, while the A-optimality criterion maximizes $1/tr(\mathfrak{M}^{-1})$, where $\mathfrak{M}$ is the Fisher information matrix or an approximation of it. The two optimality criteria focus on slightly different aspects of the design. For a fixed confidence level, a D-optimal design minimizes the expected volume of a simultaneous confidence ellipsoid for the parameters, while an A-optimal design minimizes the sum of the variances of the parameter estimators. The PQL- and MQL-based approximations that we use in the comparison are shown in Table 1. As explained in Sect. 2, to reduce computing time for the PQL-based approximation, we explore the use of support points introduced by Mak and Joseph (2018) to reduce the sample size S for evaluating the expectations. We also consider the PQL-based approximation using a larger sample size without support points to show that using support points is a good choice. We use a fixed test set of designs and multiple parameter settings for the logistic model with a single covariate and random coefficients.

The choices for $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ are shown in Table 2. The values for $\boldsymbol{\beta}$ correspond to different scenarios of the complete class results for fixed effects models in Yang and Stufken (2009). For $\boldsymbol{\Sigma}$, we use a fixed correlation $\rho = .5$ and three choices for the diagonals (I, II, and III in Table 2) that reflect different levels of uncertainty for the random intercept and slope.

All designs in the test set of designs consist of a single sequence, i.e., $N_s = 1$. Based on additivity of the information, a good approximation for $N_s = 1$ implies a good approximation for moderately larger values of N_s. We select [1, 6] as the design region, and include every design that is supported on exactly two points from {1, 2, 3, 4, 5, 6} and a total number of measurements of $n = 10$ in the test set. Thus, the test set consists of $9 \times \binom{6}{2} = 135$ designs.

For evaluating the performance of the approximations, correctly identifying the best designs is more important than matching values of the optimality criteria. We focus therefore on how well the approximations perform in distinguishing between better and worse designs. To do this, we identify the B best designs in the test set based on the exact information matrix, and rank the 135 designs for each of the approximations. For each approximation, we then compute the proportion of pairs of designs with precisely one from the $B = 5$ best designs for which the approximation ranks the two designs correctly. Thinking of this as a classification

Table 1 Candidates in the comparison of the approximations

Candidates	Description
PQL	PQL approximation in (7) using random sample ($S = 1000$)
PQL SP	PQL approximation in (7) using support points ($S = 100$)
MQL	MQL approximation cf. Breslow and Clayton (1993)
adj. MQL	MQL approximation adjusted by Zeger et al. (1988)
Exact Info.	Exact Information defined in (8) as a benchmark

Table 2 Parameter choices for comparing the approximations

Choices for $\boldsymbol{\beta}$	$(1, -1)$	$(3, -1)$	$(7/2, -1)$	$(4, -1)$	$(6, -1)$
Diagonals of $\boldsymbol{\Sigma}$	I: (1.7145, 1.05)		II: (6,.3)		III: (.3, 6)

problem (classifying a design as a better design or not), this measure corresponds to the area under the curve (AUC) of the receiver operating characteristic (ROC) curve (cf. Provost and Fawcett 1999). The results for $B = 5$ are shown in Table 3. The PQL approximation achieves a very high AUC in all cases, also when using the faster approach with support points (PQL SP). The MQL approximation is relatively poor, while the adjusted MQL works well except for Type I and $\boldsymbol{\beta} = (6, -1)'$. Similar conclusions hold for other values of B.

Computing the exact information matrix for a single design takes about 800 CPU seconds in Mathematica on a 3.9 GHz Intel CPU. This would be too slow in a search algorithm for optimal designs. In comparison, the PQL approximation takes about 0.16 CPU seconds without support points and about 0.011 CPU seconds with support points on the same platform.

Overall, the PQL approximation performs well in ordering the designs correctly and is fast, especially when using support points.

4 Optimal designs

In this section, we use the PQL approximation to the information matrix and Particle Swarm Optimization (PSO) to search for locally A- or D-optimal designs for the mixed-effects logistic model. PSO is a widely used meta-heuristic algorithm in optimal design searching (Zhou et al. 2021). To illustrate the results, we use the same settings for $\boldsymbol{\beta}$ as in Table 2. We take $\boldsymbol{\Sigma} = r \cdot \boldsymbol{\Sigma}_1$ as a diagonal matrix, where $r \in \{5, 7, 10, 15, 25\}$ and $\boldsymbol{\Sigma}_1 = \text{diag}(\sigma_1^2, \sigma_2^2)$ with $(\sigma_1^2, \sigma_2^2) = (.1143, .07), (.4, .02),$ or $(.02, .4)$. We refer to these three choices for (σ_1^2, σ_2^2) as type I, II and III, respectively. For fixed r, the generalized variance, $\det(r\boldsymbol{\Sigma}_1)$, is approximately the same for all three types. The design region is set to be $[1, 6]$.

Due to correlated observations, neither the information matrix nor its approximations are additive for observations on the same subject. Additivity does hold when combining information from different subjects. Therefore, using an approximate design approach, we use a population design $\xi = \{(\xi_p, w_p), p = 1, \ldots, N_s\}$, where $w_p > 0$, $\sum_p w_p = 1$, and each individual design ξ_p is an n-point design for a fixed value of n. For a fixed value of N_s, there are $N_s(n + 1)$ unknowns when trying to determine the best design. These unknowns correspond to the n design points in each sequence ξ_p and the corresponding weight w_p. There are constraints on each element of this vector since the design points must be in the design region and the weights must be nonnegative and sum to 1. Since N_s is unknown, a search may be needed for different values of N_s, although a single search can suffice if N_s is taken to be large enough so that some sequences receive weight 0 for the best design. For a given N_s, each vector of length $N_s(n + 1)$ corresponds to a particle in the PSO algorithm. A swarm is a collection of multiple particles that, through communication with each other and certain rules for particle movement, explore this $N_s(n + 1)$-dimensional constraint space for finding a design that maximizes an objective function corresponding to, in our case, A-optimality or D-optimality. The algorithm terminates when the best value for the objective function does not change after a sufficient number of iterations. We do not insist that the n points in the design region $[1, 6]$ are distinct. In a situation where it is not possible to have repeated values (for example when the design variable denotes the time at which an observation is to be made in a longitudinal study), we can distribute the design points around the repeated value. Alternatively, if there is a specified minimum distance d_0 between any two design points, we can build this constraint into the PSO algorithm. We considered optimal designs for $n = 2$ through 6, and show results for $n = 2$ and 5.

Since we use exact individual designs, the optimization problem is a discrete problem, and we can no longer use a general equivalence theorem to verify optimality of a design. Even though we are not able to guarantee that our designs are indeed optimal, by running the PSO algorithm multiple times with different starting designs and a large number of

Table 3 AUC using the top 5 designs for PQL, PQL with support points, MQL, and adjusted MQL

		$\beta = (1, -1)'$				$\beta = (6, -1)'$			
Σ	Optimality	PQL	PQL SP	MQL	adj. MQL	PQL	PQL SP	MQL	adj. MQL
I	A	.997	.997	.854	.994	1	1	.815	.882
	D	.998	.998	.872	.994	.975	.975	.803	.874
II	A	.997	.997	.854	.997	.998	.998	.815	.989
	D	.997	.997	.826	.997	.994	.997	.832	.988
III	A	.974	.971	.909	.974	.920	.911	.797	.922
	D	.966	.968	.925	.966	.918	.914	.798	.918

iterations, we are confident that the population designs are highly efficient.

In Figs. 1, 2, 3, 4, we show D- and A-optimal population designs for $n = 2$ and 5. In all cases, the design either has $N_s = 1$ or 2. When $N_s = 1$, which occurs most often, we simply present the n points of the corresponding individual design ξ_1. For $n = 5$, repeated points are presented slightly apart so that the number of points at a location is discernible. For $n = 2$ and A-optimality, some cases yield $N_s = 2$. Different plotting symbols (dot and cross) are then used for the two individual designs ξ_1 and ξ_2, and weights are shown for ξ_2 next to one of the design points.

The optimal designs have several noteworthy features. First, for $n = 5$, the number of distinct design points increases with the value of r for both A- and D-optimality. This increase can especially be seen in results for $\boldsymbol{\beta} = (6, -1)'$ in Figs. 2 and 4. For other $\boldsymbol{\beta}$'s in these figures, a larger design region would have exhibited a similar pattern. Also for $n = 5$, for the smallest value of r that is considered here, $r = 5$, most of the optimal designs are 2-point designs, except for some type III cases. It is known that for the limiting case $r \to 0$, which corresponds to a (fixed-effects) GLM, a 2-point design is A-optimal among approximate designs (Yang and Stufken 2009). Optimal approximate designs for GLMs are shown in Table 4 for reference. However, since the individual designs here are exact designs, with a $n = 5$ we cannot match the weights for the A-optimal approximate designs in Table 4 precisely.

Second, for $\boldsymbol{\beta} = (1, -1)'$ or $(6, -1)'$, an optimal design for the GLM includes one of the endpoints of the design region (Yang and Stufken 2009). This need not be the case for the mixed-effects model (see, for example, Fig. 1 with $\boldsymbol{\beta} = (6, -1)'$ for type III and r=10). Third, while the results in Schmelter (2007) do not apply here, optimal population designs still consist often (but not always) of the same individual design for all subjects. Figure 5 shows A-efficiencies of the best one-sequence designs for $n = 2$ as measured by $\frac{\phi_A(\xi)}{\phi_A(\xi^*)}$, where $\phi_A(\xi) = 1/tr((\mathfrak{M}^{PQL}(\xi))^{-1})$ denotes the A-optimality criterion, and ξ^* is an A-optimal design. As illustrated by Fig. 5, the best one-sequence design tends to be highly efficient for virtually all values of r, and tends to become A-optimal for larger values of r. The efficiencies and optimality do however vary with $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$.

These general conclusions extend to other choices for β_1, even though optimal designs show some differences. Results for other choices of β_1 are reported in the Online Resource Supplementary Material.

Finally, for our PSO algorithm, the number of measurements n per subject must be specified. For the optimal designs reported in this paper, we use 20 particles and 500 iterations to get a near optimal design ξ_a. Next, we use a new random start while keeping ξ_a as the initial global optimum, with additional 1500 iterations. The final design may contain replicated points. While not done here, if replicated points are undesirable, we can run PSO with a constraint that enforces a lower bound on the minimum distance, such as d_0, between any pair of design points. Alternatively, we could retain one replicate of the distinct design points, and use PSO to add additional points, enforcing a minimum distance of d_0 between any two points. In our examples, this approach allowed us to replace replicated points with the nearest points with distance $d_0 = .25$ or .5. We will revisit this approach in the following section.

As a special case of model (2), if a random block effects model is of interest as discussed in Sect. 2, we don't need to worry about these replicated points.

5 An application and robustness analysis

For illustration of the methodology, we focus on the French EPIDOS study (Epidétiology de l'ostéoporose), a prospective multi-center study of the risk factors for hip fractures in women who were 75 years or older in 1992-1993. The participating women completed health-related questionnaires annually for six years. Carrière and Bouyer (2002) analyzed the data from Montpellier, one of the 5 participating centers, using a generalized linear mixed model with a logistic link function as in model (2). After testing the significance of random effects, the authors finally determined the covariance structure with random intercept and random slope. Let y_{ij} be the indicator of disability, "needing help to go out-

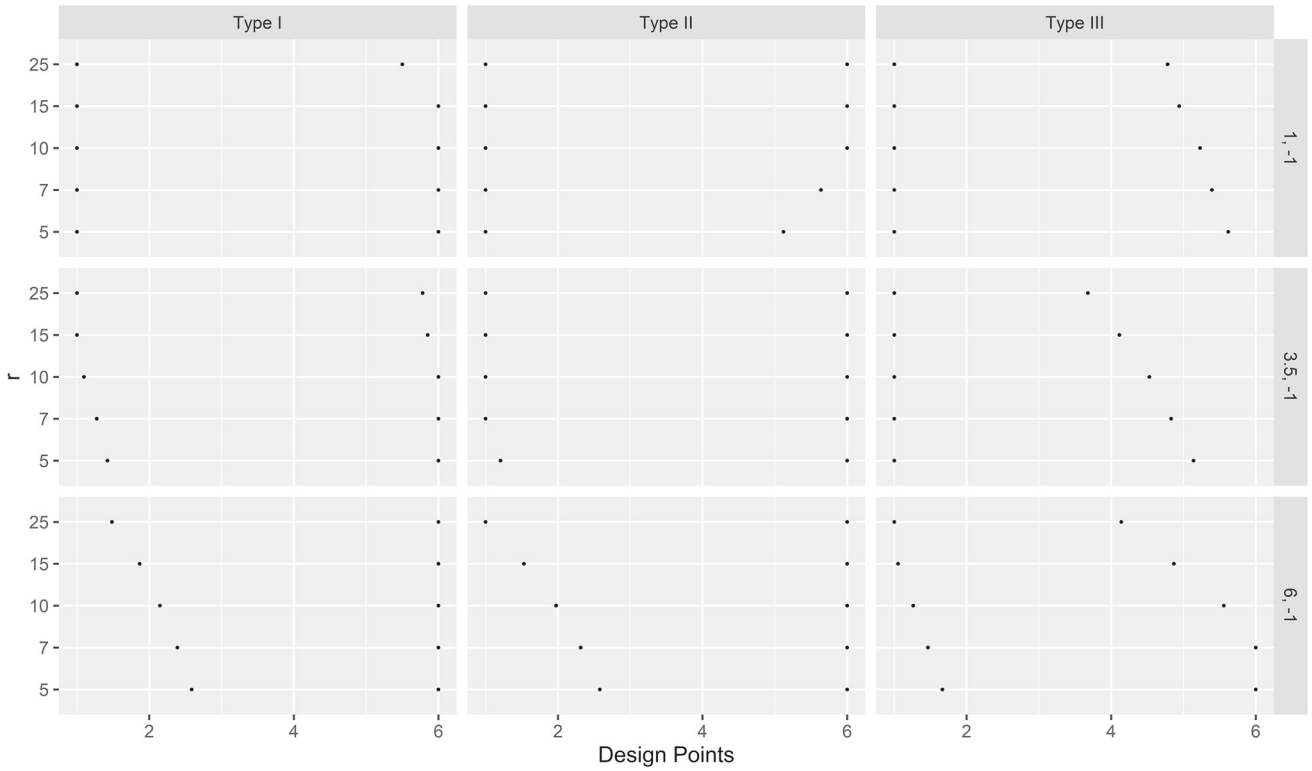

Fig. 1 D-optimal designs for $n = 2$ observations per subject. The covariance type is shown in the top bar and the choice for $\boldsymbol{\beta}$ in the right-hand bar. The value of r for the covariance matrix $\boldsymbol{\Sigma}$ is shown along the vertical axis, and the design region [1, 6] is shown along the horizontal axis

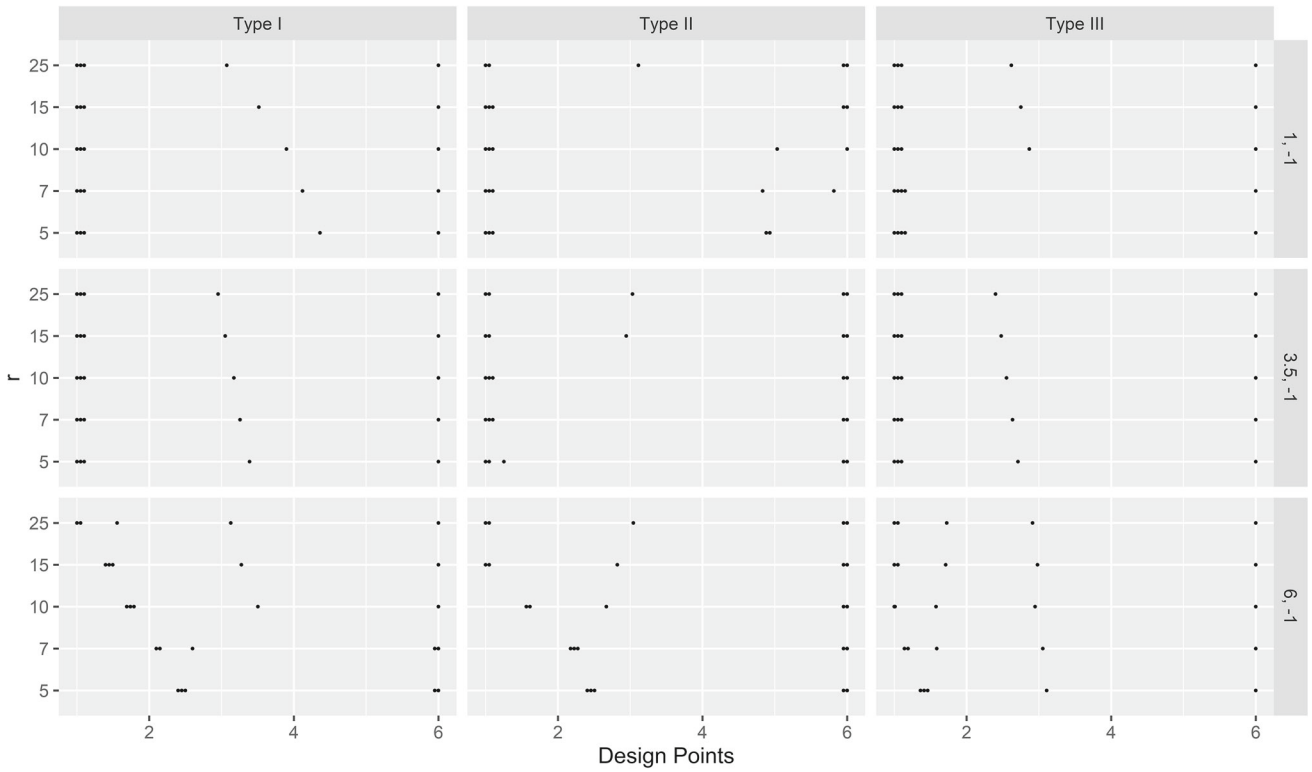

Fig. 2 D-optimal designs for $n = 5$ observations per subject. The covariance type is shown in the top bar and the choice for $\boldsymbol{\beta}$ in the right-hand bar. The value of r for the covariance matrix $\boldsymbol{\Sigma}$ is shown along the vertical axis, and the design region [1, 6] is shown along the horizontal axis

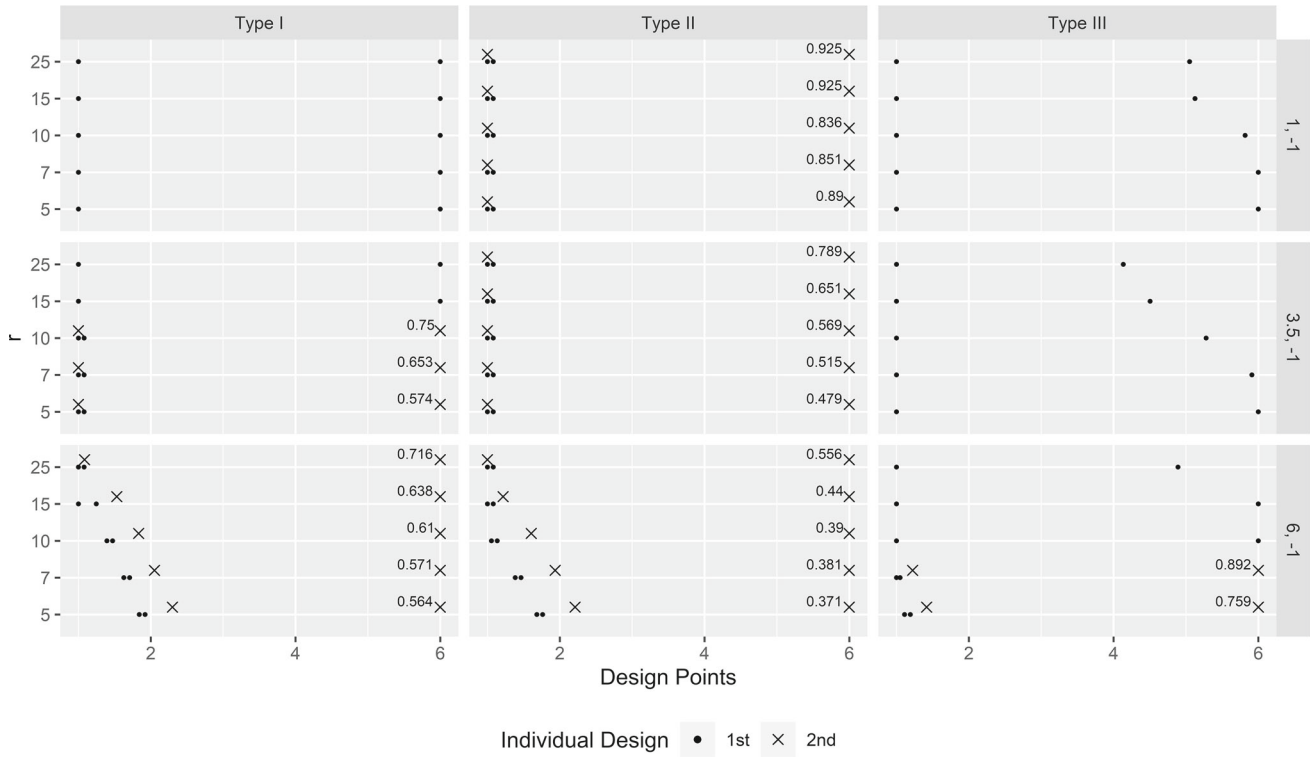

Fig. 3 A-optimal designs with $n = 2$ observations per subject. The covariance type is shown in the top bar and the choice for $\boldsymbol{\beta}$ in the right-hand bar. The value of r for the covariance matrix $\boldsymbol{\Sigma}$ is shown along the vertical axis, and the design region $[1, 6]$ is shown along the horizontal axis. The plots of the second individual design is vertically adjusted to avoid overlapping

doors or home-confined", for woman i and year j ($i = 1, \ldots, 1548$; $j = 1, \ldots, 6$). The proposed logistic mixed-effects model is

$$\begin{aligned} &\text{logit}[E(y_{ij} \mid \mathbf{b}_i)] \\ &= \log\left(\frac{P(y_{ij} = 1 \mid \mathbf{b}_i)}{1 - P(y_{ij} = 1 \mid \mathbf{b}_i)}\right) \\ &= b_{i0} + b_{i1} x_{ij}, \end{aligned} \tag{9}$$

where the design points are $x_{ij} = j$, $j = 1, \ldots, 6$; $\mathbf{b}_i = (b_{i0}, b_{i1})^T \sim N_2(\boldsymbol{\beta}, \boldsymbol{\Sigma})$, $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$, and $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{pmatrix}$. The estimated parameters from this study are displayed in Table 5.

While there may have been practical reasons for using the same equally spaced design for each of the women, one may wonder whether such a design is optimal or efficient. Considering $n = 6$ observations for each woman, using a PSO algorithm, the estimated parameters in Table 5, and design space $[1, 6]$, locally optimal designs for this model are presented in Table 6. Both A- and D-optimal designs are one-sequence designs.

From Table 6, both designs require each woman to complete multiple questionnaires at the first time point, which is clearly infeasible. We could simply replace replicated points by nearest neighbors, with all design points at least separated by a specified distance d_0. Alternatively, as already suggested, we can use PSO again keeping one copy of the replicated point and forcing a distance of d_0 between any two design points. In this example, these two methods provide the same answers. Using half a year, a quarter of a year, and one month for d_0, the obtained designs are ξ_1, ξ_2 and ξ_3, respectively, where

$$\xi_1 = \begin{cases} \{1, 1.5, 2, 2.5, 3, 6\}, & \text{A-optimality} \\ \{1, 1.5, 2, 5, 5.5, 6\}, & \text{D-optimality,} \end{cases}$$

$$\xi_2 = \begin{cases} \{1, 1.25, 1.5, 1.75, 2, 6\}, & \text{A-optimality} \\ \{1, 1.25, 1.5, 5.5, 5.75, 6\}, & \text{D-optimality,} \end{cases}$$

and

$$\xi_3 = \begin{cases} \{1, 1.08, 1.17, 1.25, 1.33, 6\}, & \text{A-optimality} \\ \{1, 1.08, 1.17, 5.83, 5.92, 6\}, & \text{D-optimality.} \end{cases}$$

Efficiencies of these designs and the design $\xi_0 = \{1, 2, 3, 4, 5, 6\}$ used in the study relative to an optimal design ξ^* are computed as the ratio $\phi(\xi)/\phi(\xi^*)$, where $\phi(\xi) = \phi_A(\xi) = 1/tr((\mathfrak{M}^{PQL}(\xi))^{-1})$ or $\phi(\xi) = \phi_D(\xi) = det(\mathfrak{M}^{PQL}(\xi))^{1/q}$

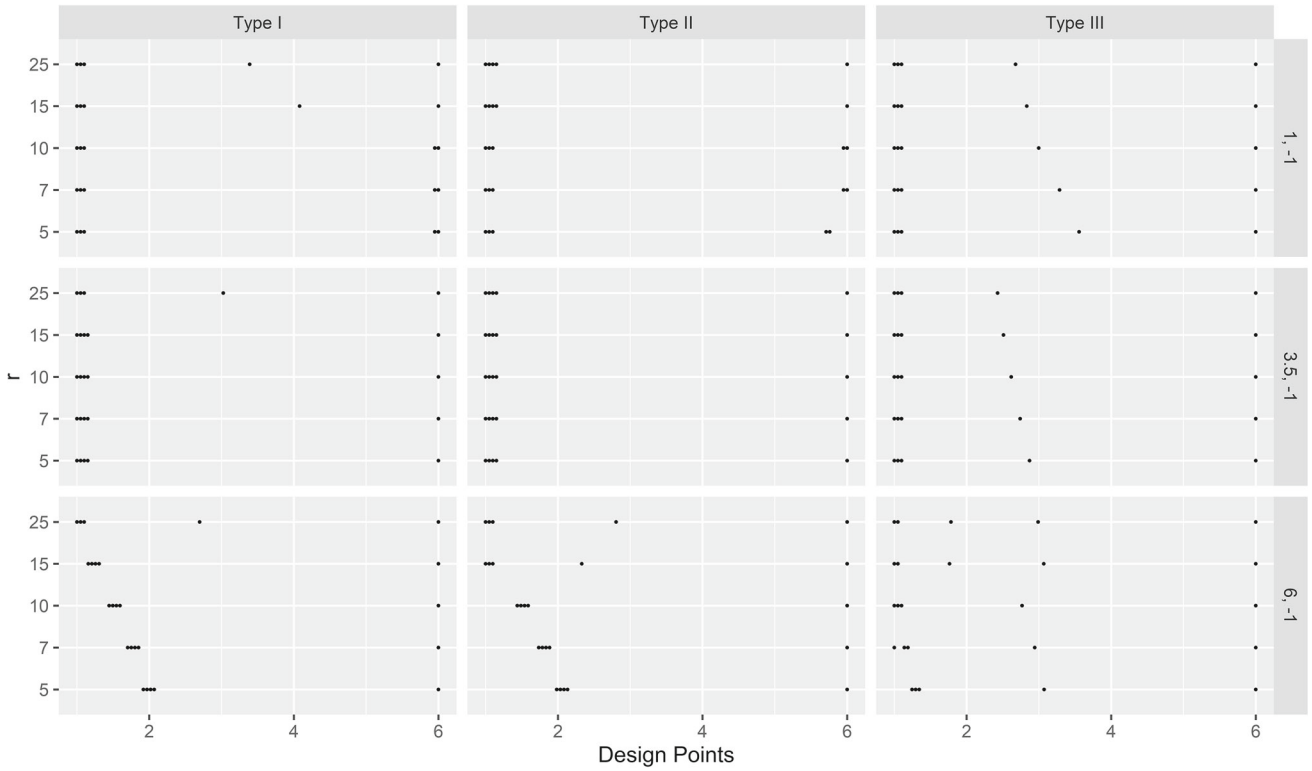

Fig. 4 A-optimal designs with $n = 5$ observations per subject. The covariance type is shown in the top bar and the choice for $\boldsymbol{\beta}$ in the right-hand bar. The value of r for the covariance matrix $\boldsymbol{\Sigma}$ is shown along the vertical axis, and the design region [1, 6] is shown along the horizontal axis

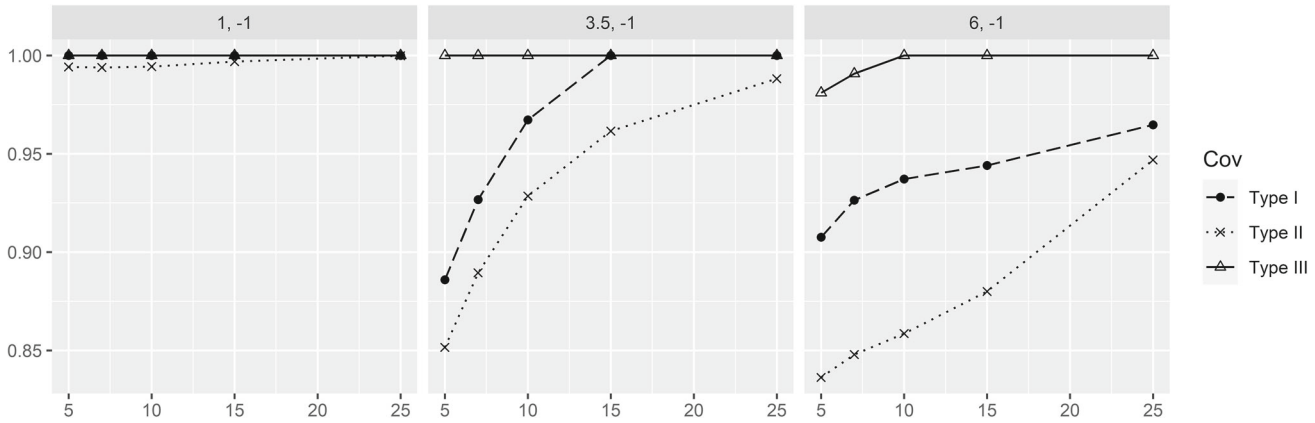

Fig. 5 A-efficiencies of the best one-sequence designs for $n = 2$ for different choices of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$. The axes shows the values of r (horizontal) and the efficiency (vertical). The top bar in each panel shows $\boldsymbol{\beta}$, and different line types correspond to the three choices for $\boldsymbol{\Sigma}$ that were also used in Figs. 1, 2, 3, 4

Table 4 D- and A-optimal designs for the fixed effects GLM for three choices of the regression parameters

Optimality	$(1, -1)$	$(3.5, -1)$	$(6, -1)$
D	(1,3.399;.5,.5)	(1.957,5.043;.5,.5)	(3.601,6;.5,.5)
A	(1,4.009;.553,.447)	(1.211,5.789;.789,.211)	(2.920,6;.828,.172)

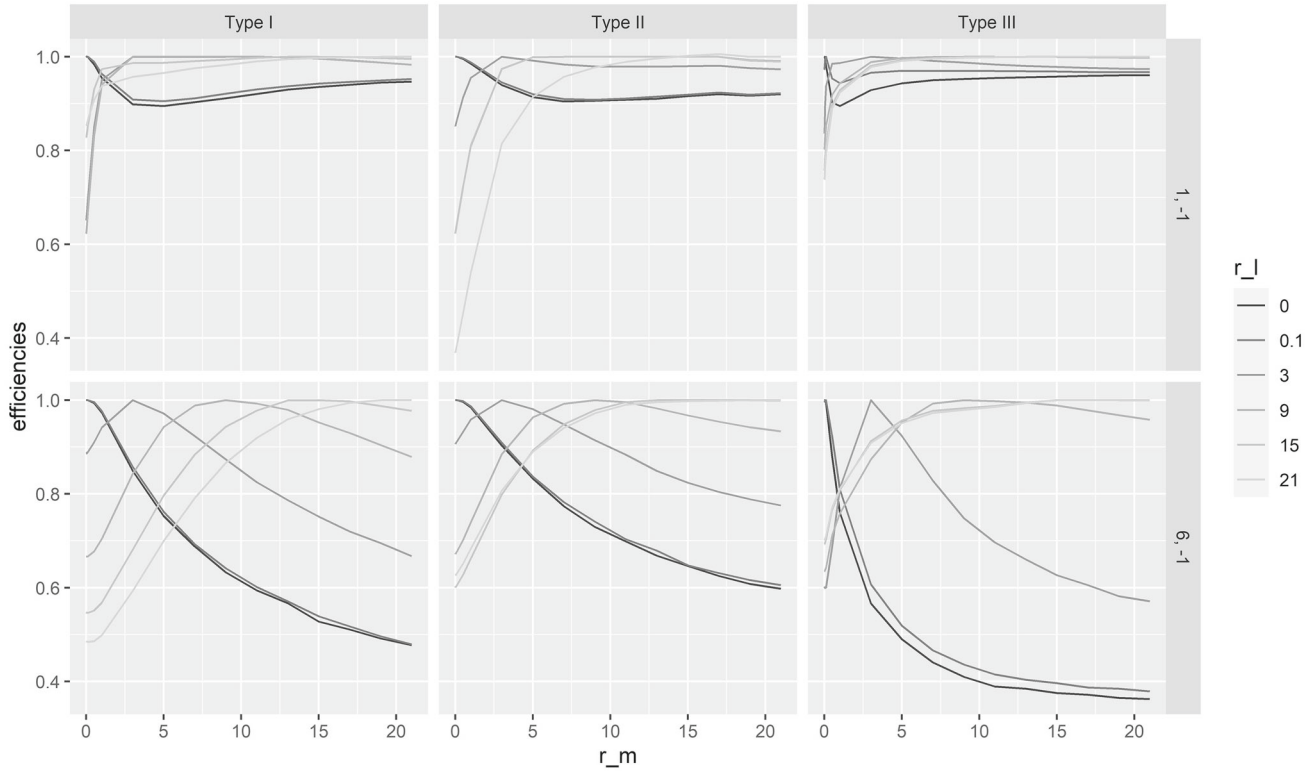

Fig. 6 Robustness study: Efficiencies of A-optimal Designs, $n = 5$. The covariance type is shown in the top bar and the choice for $\boldsymbol{\beta}$ in the right hand bar. Value of r_m is along the horizontal axis, which corresponds to the true covariance matrix, and values of r_l are represented by the six different lines, each corresponds to the optimal design under such r_l. The efficiencies of these designs are shown along the vertical axis

Table 5 Parameter estimates for French EPIDOS study

Parameter	β_0	β_1	σ_{11}	σ_{22}
Estimate	−3.61	0.17	7.25	0.18

Table 6 A- and D-optimal designs ξ^* for self-reported disability study

Optimality	x_1	x_2	x_3	x_4	x_5	x_6
A	1	1	1	1	1	6
D	1	1	1	6	6	6

for A- and D-optimality, respectively, and ξ^* is an A-optimal or D-optimal design. The results are shown in Table 7.

Table 7 indicates that designs remain highly efficient if d_0 is not large. There is more loss of efficiency under A-optimality because the proposed replacement moves us further from the A-optimal design in Table 6, which places more emphasis on one of the endpoints.

Locally optimal designs depend on "guessed" parameters, and poor guesses may lead to poor designs. We are particularly interested in robustness to mis-specification of the variance-covariance matrix $\boldsymbol{\Sigma}$. With the notation from Sect. 4, let ξ_r denote a locally optimal design under A- or D-optimality for $\boldsymbol{\Sigma}_r = r \cdot \boldsymbol{\Sigma}_1$ and a given vector $\boldsymbol{\beta}$. The A- and D-efficiencies e^A_{lm} and e^D_{lm} for design ξ_{r_l} relative to design ξ_{r_m}, where $r_l, r_m \in \{0, .01, .05, .1, .5, 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21\}$, were computed for $\boldsymbol{\beta} = (1, -1)'$ or $(6, -1)'$ and $\boldsymbol{\Sigma}_{r_m}$. Results for $n = 5$ are shown in Figs. 6 (A-optimality) and 7 (D-optimality), but, for clarity of the figures, only for $r_l = 0, .1, 3, 9, 15, 21$. Notice that $e_{ll} = 1$ for all l. From the figures we see that when r_m (the true r) is 0, a larger guessed value r_l can cause significant loss in efficiency. This says that using an optimal design for a GLMM when the true model is a GLM can result in an inefficient design. For $\boldsymbol{\beta} = (1, -1)'$, if we use an optimal design for the GLM, from the line for $r_l = 0$, we can see that it performs well for any r_m, with the efficiency even increasing for larger r_m. For $\boldsymbol{\beta} = (6, -1)'$, both underestimation and overestimation of the true r_m can

Table 7 Efficiencies (in percentage) of designs in practice

Optimality	ξ_0	ξ_1	ξ_2	ξ_3
A	82.57	92.60	97.03	99.25
D	94.22	98.12	98.98	99.62

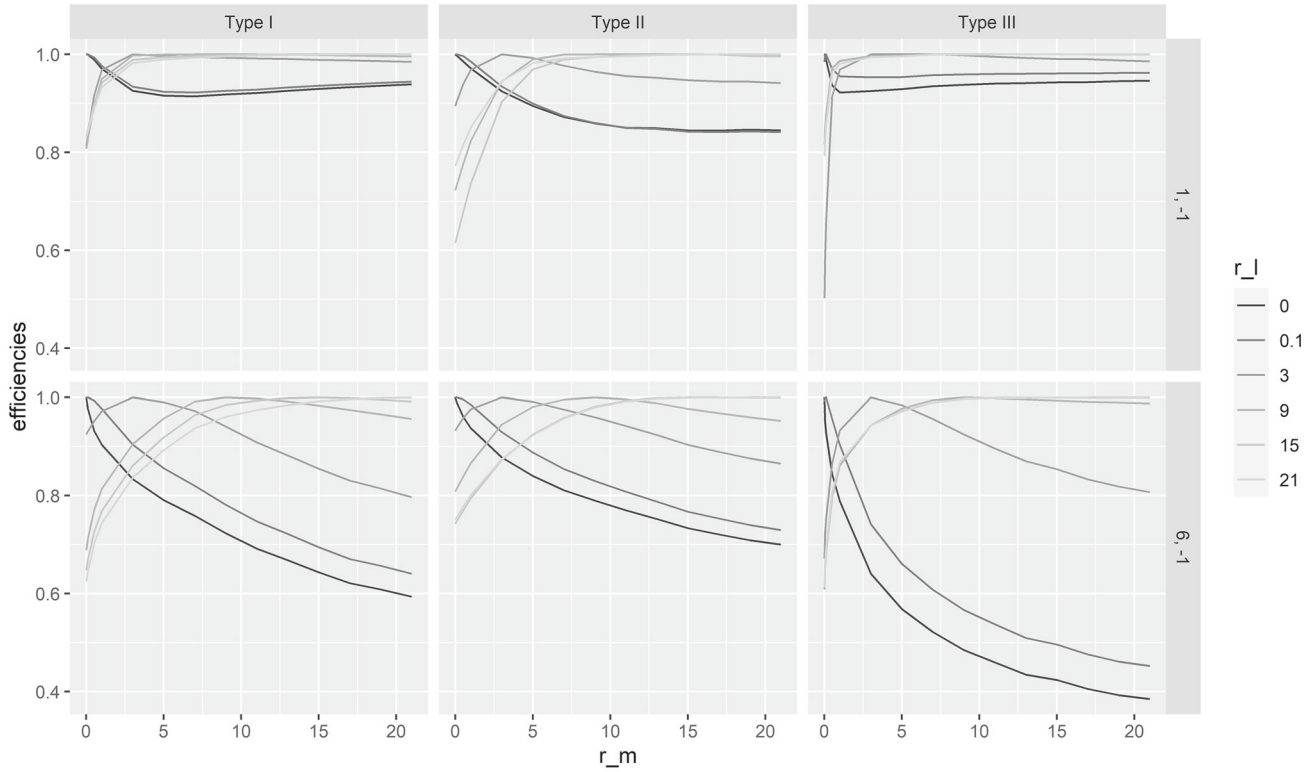

Fig. 7 Robustness study: Efficiencies of D-optimal Designs, $n = 5$. The covariance type is shown in the top bar and the choice for $\boldsymbol{\beta}$ in the righthand bar. Value of r_m is along the horizontal axis, which corresponds to the true covariance matrix, and values of r_l are represented by the six different lines, each corresponds to the optimal design under such r_l. The efficiencies of these designs are shown along the vertical axis

result in designs that are not very efficient. In Fig. 6, some of the lines reach values slightly higher than 1, such as in the panel for Type II and $\boldsymbol{\beta} = (1, -1)'$. This is due to PSO finding highly efficient designs, but not necessarily optimal designs.

Mis-specification of $\boldsymbol{\Sigma}$ can also be studied using Wishart distributions. We can set a true $\boldsymbol{\Sigma} = \text{diag}(1.7145, 1.05)$, say, then generate $\mathbf{G}$ from the Wishart distribution $W_2(\boldsymbol{\Sigma}, df)$, and use $\mathbf{G}/df$ as a mis-specified guess. We do this for df =2, 4, 7 and 15 and for $\boldsymbol{\beta} = (1, -1)'$ and $(6, -1)'$. Note that the coefficient of variation for the diagonals of $\mathbf{G}/df$ is $\sqrt{2/df}$, so that it is smaller when r is larger. For each df, 100 $\mathbf{G}$'s are generated, and the corresponding locally optimal designs are found. The efficiencies of these locally optimal designs relative to the locally optimal design for the true $\boldsymbol{\Sigma}$ are shown in Fig. 8. By examining these box plots, we can conclude that the designs are quite robust against the mis-specification. When the coefficient of variation is largest, corresponding to the case where $r = 2$, the efficiencies tend to be lower, but remain high with very few exceptions, both for A- and D-optimality. The efficiencies are slightly lower for $\boldsymbol{\beta} = (6, -1)'$ than for $\boldsymbol{\beta} = (1, -1)'$.

6 Summary and discussion

The primary objective of this paper is to identify optimal designs for GLMMs, focusing primarily on the logistic link. As closed-form expressions for the information matrix for GLMMs are not available, we propose using the PQL-based approximation, which performs better than the MQL- and adjusted MQL-based approximations. Computationally, We evaluate the PQL-based approximation using support points, which reduces computing time and enables us to use a PSO algorithm for finding optimal population designs. Through this approach, we observe characteristics of the optimal designs and their relationship to optimal designs for corresponding fixed-effects models. For the cases studied, optimal population designs for GLMMs often use a single individual design for all subjects, but not always. Also, in part because

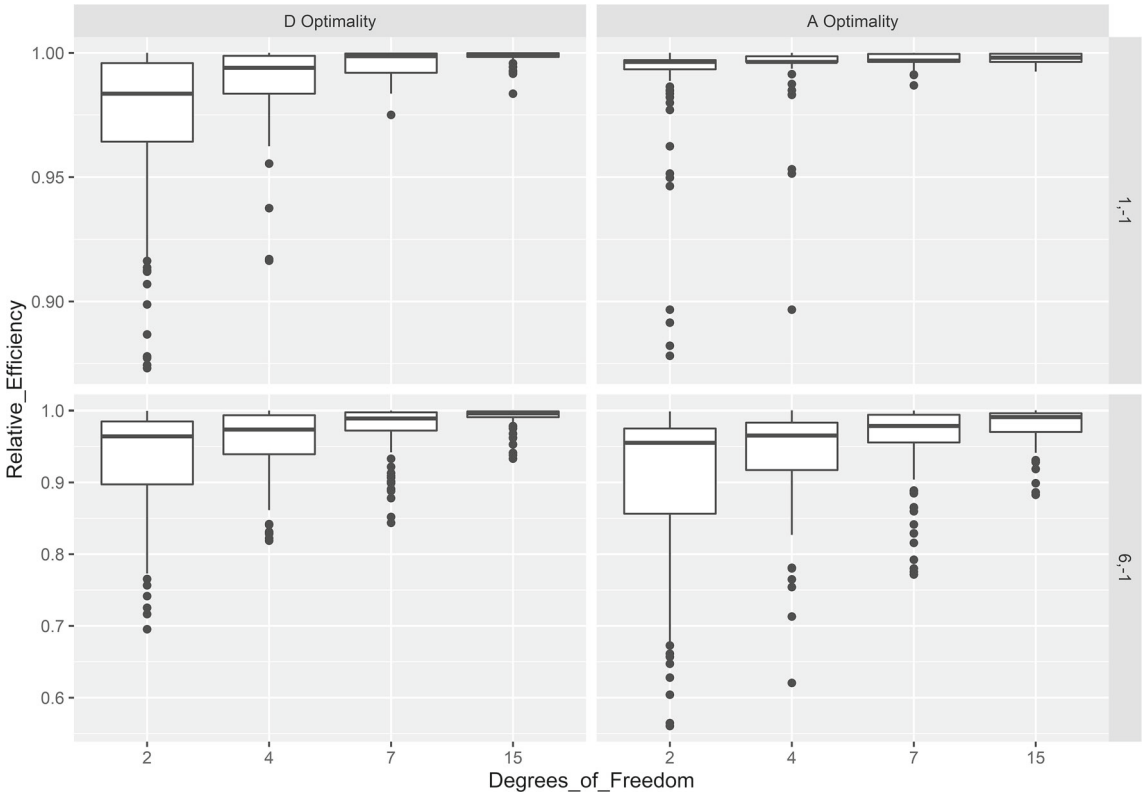

Fig. 8 Efficiencies of D- and A-optimal Designs with mis-specified $\boldsymbol{\Sigma}$ from Wishart distributions. The optimality criterion is shown on the top bar, and the choice for $\boldsymbol{\beta}$ is in the righthand bar. The degrees of freedom in the Wishart distribution are shown along the horizontal axis, and efficiency is along the vertical axis

individual designs are treated as exact designs, the individual designs may have more support points than optimal approximate designs for fixed-effects.

For future study, besides making the methodology available for other link functions, we plan to explore use of the PQL-based approximation for other optimality criteria, including Bayesian or minimax criteria. We also plan to investigate the use of more complex linear predictors, such as quadratic predictors or predictors with multiple variables, which will be more challenging.

Online Resource

The file Supplementary Material contains optimal designs for $n = 2$ and $n = 5$ when the slope parameter is $\beta_1 = -2$ and $\beta_1 = 1$ rather than $\beta_1 = -1$ as used in Sect. 4 of the paper.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11222-023-10279-3.

Author Contributions All authors contributed equally to the study that results in this manuscript, and have read and approved this final version.

Funding The work by JS was partially funded by NSF grants DMS-1935729 and DMS-2304767.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

Abebe, H.T., Tan, F.E.S., Breukelen, G.J.P.V., et al.: Robustness of Bayesian D-optimal design for the logistic mixed model against misspecification of autocorrelation. Comput. Stat. **29**(6), 1667–1690 (2014). https://doi.org/10.1007/s00180-014-0512-3

Atkinson, A.C., Woods, D.C.: Designs for generalized linear models. In: Dean, A., Morris, M., Stufken, J., et al. (eds.) Handbook of Design and Analysis of Experiments. Chapman and Hall/CRC, New York (2015)

Breslow, N.E., Clayton, D.G.: Approximate inference in generalized linear mixed models. J. Am. Stat. Assoc. **88**(421), 9–25 (1993). https://doi.org/10.2307/2290687

Carrière, I., Bouyer, J.: Choosing marginal or random-effects models for longitudinal binary responses: application to self-reported disabil-

ity among older persons. BMC Med. Res. Methodol. **2**, 15 (2002). https://doi.org/10.1186/1471-2288-2-15
Green, P.J.: Penalized likelihood for general semi-parametric regression models. Int. Stat. Rev. **55**, 245–259 (1987). https://doi.org/10.2307/1403404
Liang, K.Y., Zeger, S.L.: Longitudinal data analysis using generalized linear models. Biometrika **73**(1), 13–22 (1986). https://doi.org/10.1093/biomet/73.1.13
Mak, S., Joseph, V.R.: Support points. Ann. Stat. **46**, 2562–2592 (2018)
Mielke, T.: Approximations of the Fisher information for the construction of efficient experimental designs in nonlinear mixed effects models. PhD thesis, Faculty of Mathematics, Otto-von-Guericke University, Magdeburg (2012)
Miller, K.S.: On the inverse of the sum of matrices. Math. Mag. **54**(2), 67–72 (1981). https://doi.org/10.2307/2690437
Molenberghs, G., Verbeke, G.: Models for Discrete Longitudinal Data. Springer, New York (2005)
Niaparast, M.: On optimal design for a Poisson regression model with random intercept. Stat. Probab. Lett. **79**(6), 741–747 (2009). https://doi.org/10.1016/j.spl.2008.10.035
Niaparast, M., Schwabe, R.: Optimal design for quasi-likelihood estimation in Poisson regression with random coefficients. J. Stat. Plann. Inference **143**(2), 296–306 (2013). https://doi.org/10.1016/j.jspi.2012.07.009
Niaparast, M., MehrMansour, S., Schwabe, R.: V-optimality of designs in random effects Poisson regression models. Metrika (2023). https://doi.org/10.1007/s00184-023-00896-3
Provost, F., Fawcett, T.: Analysis and visualization of classifier performance: comparison under imprecise class and cost distributions. Proceedings of the Third International Conference on Knowledge Discovery and Data Mining 43-48 (1999)
Schmelter, T.: The optimality of single-group designs for certain mixed models. Metrika **65**(2), 183–193 (2007). https://doi.org/10.1007/s00184-006-0068-5
Sinha, S.K., Xu, X.: Sequential D-optimal designs for generalized linear mixed models. J. Stat. Plann. Inference **141**(4), 1394–1402 (2011). https://doi.org/10.1016/j.jspi.2010.10.008
Tekle, F.B., Tan, F.E., Berger, M.P.: Maximin d-optimal designs for binary longitudinal responses. Comput. Stat. Data Anal. **52**(12), 5253–5262 (2008). https://doi.org/10.1016/j.csda.2008.04.037
Ueckert, S., Mentré, F.: A new method for evaluation of the fisher information matrix for discrete mixed effect models using Monte Carlo sampling and adaptive Gaussian quadrature. Comput. Stat. Data Anal. **111**, 203–219 (2017). https://doi.org/10.1016/j.csda.2016.10.011
Waite, T.W., Woods, D.C.: Designs for generalized linear models with random block effects via information matrix approximations. Biometrika **102**(3), 677–693 (2015). https://doi.org/10.1093/biomet/asv005
Yang, M., Stufken, J.: Support points of locally optimal designs for nonlinear models with two parameters. Ann. Stat. **37**(1), 518–541 (2009). https://doi.org/10.1214/07-AOS560
Zeger, S.L., Liang, K.Y., Albert, P.S.: Models for longitudinal data: a generalized estimating equation approach. Biometrics **44**(4), 1049–1060 (1988). https://doi.org/10.2307/2531734
Zhou, X., Wang, Y., Yue, R.: Optimal designs for discrete-time survival models with random effects. Lifetime Data Anal. **27**, 300–332 (2021). https://doi.org/10.1007/s10985-020-09512-2