Robust designs in longitudinal studies accounting for parameter and model uncertainties -Application to count data

Florence Loingeville, Thu Thuy Nguyen, Marie-Karelle Riviere, France Mentré

Robust designs in longitudinal studies accounting for parameter and model uncertainties - Application to count data

Florence Loingeville[1,2] Thu Thuy Nguyen[1] Marie-Karelle Riviere[3]
France Mentré[1]

[1] INSERM, IAME, UMR 1137, F-75018 Paris, France;
University Paris Diderot, Sorbonne Paris Cité, F-75018 Paris, France

[2] Faculty of Pharmacy, University of Lille, EA 2694, Public Health: Epidemiology and Healthcare quality, F-59000 Lille, France

[3] Statistical Methodology Group, Biostatistics and Programming department, Sanofi-Aventis R&D, Chilly-Mazarin, France

* *Email address: florence.loingeville@univ-lille.fr*

Acknowledgments

This work was supported by the European Union 7th Framework Programme under Grant Agreement n602552. The authors gratefully acknowledge François Cohen and Hervé Le Nagard for their valuable and indispensable help with CATBioMed cluster.

Abstract

Nonlinear mixed effect models (NLMEMs) are widely used for the analysis of longitudinal data. To design these studies, optimal designs based on the expected Fisher information matrix (FIM) can be used. A method evaluating the FIM using Monte-Carlo Hamiltonian Monte-Carlo (MC-HMC) has been proposed and implemented in the R package MIXFIM using Stan. This approach, however, requires *a priori* knowledge of models and parameters, which leads to locally optimal designs. The objective of this work was to extend this MC-HMC-based method to evaluate the FIM in NLMEMs accounting for uncertainty in parameters and in models. When introducing uncertainty in the population parameters, we evaluated the robust FIM as the expectation of the FIM computed by MC-HMC over the distribution of these parameters. Then, the compound D-optimality criterion (CD optimality), corresponding to a weighted product of the D-optimality criteria of several candidate models, was used to find a common CD-optimal design for the set of candidate models. Finally, a compound DE-criterion (CDE optimality), corresponding to a weighted product of the normalized determinants of the robust FIMs of all the candidate models accounting for uncertainty in parameters, was calculated to find the CDE-optimal design which was robust on both parameters and model. These methods were applied in a longitudinal Poisson count model. We assumed prior distributions on the population parameters as well as several candidate models describing the relationship between the logarithm of the event rate parameter and the dose. We found that assuming uncertainty in parameters could lead to different optimal designs, and misspecification of models could induce designs with low efficiencies. The CD- or CDE-optimal designs therefore provided a good compromise for different candidate models. Finally, the proposed approach allows for the first time optimization of designs for repeated discrete data accounting for parameter and model uncertainties.

Key words: Fisher information matrix; Longitudinal count data; Markov Chain

Hamiltonian Monte Carlo; Nonlinear mixed effect models; Optimal design; Robust design.

1 Introduction

In pharmacometrics, longitudinal analysis of clinical data is increasingly performed in studies that provide repeated continuous or discrete data over time on several individuals. In recent years, several methods and software packages have been developed for maximum likelihood estimation of population parameters from longitudinal data using nonlinear mixed effect models (NLMEMs) (Mould and Upton, 2012, 2013; Upton and Mould, 2014). Before modeling, it is crucial to choose an appropriate population design in order to obtain good precision of estimates. Indeed, the informativeness of a dataset for parameter estimation depends on the number of subjects and on the number and timing of the samples. The Fisher information matrix (FIM) can be used in design evaluation and optimization, as its inverse is the lower bound of the variance covariance matrix of any unbiased parameter estimator according to the Cramer-Rao inequality (Atkinson, Donev, and Tobias, 2009; Fedorov and Leonov, 2013; Pronzato and Pázman, 2013).

Before discussing experimental design we briefly consider calculation of the information matrix. The problem is that in NLMEMs the FIM has no analytical form, and its calculation, which requires multiple integrations, can be challenging. Therefore, an expression of the FIM based on first-order linearization (FO) of the model around the expectation of the random effects was proposed by Mentré, Mallet, and Baccar (1997). Although efficient in general (Bazzoli, Retout, and Mentré, 2009), FO has limitations in the case of complex nonlinear models, with large variability (Jones et al., 1999; Nguyen and Mentré, 2014) and in studies with discrete outcomes. FO was extended by Ogungbenro and Aarons (2011) in studies with discrete endpoints, where they developed a method based on generalized estimating equations

and the marginal quasi-likelihood (MQL) approximation. However, this approach requires computation of an analytical form for the partial derivatives of the conditional log-likelihood for each type of data, which is inconvenient in practice. The probabilistic programming language Stan (Stan Development Team, 2016) provides automatically numerical differentiation to evaluate these partial derivatives. New approaches have recently been developed for the computation of the FIM without linearization. Ueckert and Mentré (2016) extended the method proposed by Nguyen and Mentré (2014) based on Monte Carlo - Adaptive Gaussian quadrature (MC-AGQ) and applied it to longitudinal discrete data. This approach performed better than the MQL-based method for the estimation of random effect variances in discrete NLMEMs (Ueckert and Mentré, 2016). A method based on Monte Carlo Hamiltonian Monte Carlo (MC-HMC) has also been developed by Riviere, Ueckert, and Mentré (2016). In this approach, MC is used to compute the integral over the observations and the integral over the random effects is evaluated by HMC. This method yields an asymptotically exact FIM, and is especially suitable for complex NLMEMs presenting a large variability. Its calculation time increases linearly with the number of random parameters for MC-HMC, whereas it increases exponentially for MC-AGQ. This approach is implemented in the R package MIXFIM (Riviere and Mentré, 2015), which uses functions written in the probabilistic programming language Stan, which was developed for Bayesian inference. Nyberg, Karlsson, and Hooker (2009) also proposed an MC-based method to compute the FIM using a second-order approximation of the likelihood and applied it to binary and count responses. However, they used MC sampling also to integrate the likelihood over the random effects, in contrast to HMC performed in Riviere, Ueckert, and Mentré (2016). The latter is expected to be substantially more efficient and could explain why only 200 HMC samples (plus 500 burn-in) were required in Riviere, Ueckert, and Mentré (2016) while Nyberg, Karlsson, and Hooker (2009) reported the use of 1 000 000 MC samples.

Given some FIM, one of the criteria widely used to optimize designs is D-optimality, which consists of maximizing the determinant of the FIM. This approach however requires *a priori* knowledge of the model and its parameters, which can sometimes be obtained from previous experiments, leading to locally optimal designs. Several robust criteria have been proposed in order to optimize designs assuming uncertainty in parameters (Pronzato and Walter, 1985; Walter and Pronzato, 1987; Foo et al., 2012). However, these robust criteria do not account for uncertainty in the model. Uncertainty in the model choice can be taken into account by assuming a set of candidate models when computing the optimality criterion. To do that, we make use of the theory of compound optimal design of Atkinson, Donev, and Tobias (2009), which was used to propose optimal designs for both estimation and model discrimination (Atkinson, 2008) as well as to find a common design for several drugs (Nguyen et al., 2016). Here, we combined criteria for several models in a compound optimality criterion, following the principle of model averaging, which associates the optimality criterion of each model with its weight (Pinheiro et al., 2014).

In this article, we propose methods for optimal design of mixed models for discrete data accounting for parameter and/or model uncertainty. We first extend the method based on MC-HMC to compute the robust FIM by introducing uncertainty in the population parameters. Then, we use the compound D-optimality criterion to optimize designs for several candidate models. Finally, we propose an approach to find an optimal design that is robust with respect to both parameters and model by using the compound DE-criterion for the robust FIM.

We apply the proposed methods for design optimization in models for repeated count responses. In Section 2, we detail the notation and methods, and we introduce the different optimality criteria used to account for uncertainty in parameters, the model, or both. In Section 3, we explain how these methods were applied to design optimization

in a model of repeated count response inspired by Riviere, Ueckert, and Mentré (2016) and Ogungbenro and Aarons (2011). Finally, we discuss the results and perspectives of this work in Section 4.

2 Methods

2.1 Population design

The elementary design ξ_i for the subject i ($i = 1, \ldots, N$) is defined by the number n_i of observations and the design variables $(x_{i1}, \ldots, x_{in_i})$. A population design $\Xi = \{N, (\xi_1, \ldots, \xi_N)\}$ is described by the total number of patients N, and the set of individual elementary designs to be performed on each patient: $(\xi_1, \ldots, \xi_N)$ with a total number of observations $\sum_{i=1}^{N} n_i$.

2.2 Nonlinear mixed effect models for a discrete outcome

In this article, we consider NLMEMs for discrete data where the conditional probability for observation y_{ij} at sample $j = 1, \ldots, n_i$ from patient i can be written as:

$$p(y_{ij}|b_i) = h(y_{ij}, \xi_i, g\{\mu, b_i\}). \tag{1}$$

In this expression, h is a known function describing the probability of y_{ij} given b_i for a vector of design variables ξ_i and a vector of subject-specific parameters modeled through the nonlinear function g. The function g can be expressed as a function of the vector of fixed effects parameters μ and the vector of random effects b_i. The random effects are assumed to follow a multivariate normal distribution with mean zero and covariance matrix Ω, i.e. $b_i \sim \mathcal{N}(0, \Omega)$. We denote by ψ the vector of all model parameters, i.e. $\psi = (\mu^T, \Omega_u^T)^T$, where Ω_u is a vector containing all unique elements of Ω. Observations are usually assumed to be independent, conditional upon the random effect, i.e. the joint conditional probability for

the vector of observations for subject i ($y_i = (y_{i1}, \ldots, y_{in_i})^T$) is equal to:

$$p(y_i|b_i, \psi) = \prod_{j=1}^{n_i} h(y_{ij}, \xi_i, g\{\mu, b_i\}). \tag{2}$$

2.3 Evaluation of the FIM and robust FIM accounting for parameter uncertainty using MC-HMC

Evaluation of the FIM

The FIM $\mathcal{M}(\psi, \Xi)$ for the population design Ξ is the sum of the N elementary FIMs, $\mathcal{M}(\psi, \xi_i)$, so that $\mathcal{M}(\psi, \Xi) = \sum_{i=1}^{N} \mathcal{M}(\psi, \xi_i)$. In this work, we assumed the same elementary design in all patients ($\xi_i = \xi$ for $i = 1, \ldots, N$), then $\Xi = \{\xi; N\}$, and $\mathcal{M}(\psi, \Xi) = N \times \mathcal{M}(\psi, \xi)$. In this expression, $\mathcal{M}(\psi, \xi)$ is the individual FIM of ψ and can be expressed as:

$$\mathcal{M}(\psi^*, \xi) = E_y\left(\frac{\partial \log\{L(y|\psi^*)\}}{\partial \psi} \times \frac{\partial \log\{L(y|\psi^*)\}}{\partial \psi}^T\right), \tag{3}$$

where the likelihood of the observations vector y of an individual i (subscript i will be omitted in the following) is the integral over the random effects of its conditional likelihood, and is therefore given by:

$$L(y, \psi^*) = \int_b p(y|b, \psi^*) p(b|\psi^*) db, \tag{4}$$

with $p(y|b, \psi^*)$ the p.d.f. of y given the random effects b, and $p(b|\psi^*)$ the p.d.f. of b.
The (k, l) term of the FIM can be written as:

$$\begin{aligned} \mathcal{M}(\psi^*, \xi)_{k,l} &= E_y\left(\frac{\partial \log\{L(y, \psi^*)\}}{\partial \psi_k} \frac{\partial \log\{L(y, \psi^*)\}}{\partial \psi_l}\right) \\ &= E_y(D_{k,l}\{y, \psi^*\}). \end{aligned} \tag{5}$$

Riviere, Ueckert, and Mentré (2016) showed that this expectation with respect to the observations y can be estimated using MC, whereas, after calculation, the quantity $D_{k,l}(y, \psi^*)$ can be written as:

$$D_{k,l}(y,\psi^*) = E_{b|y,\psi^*}\left(\frac{\partial\left(\log\{p(y|b,\psi^*)p(b|\psi^*)\}\right)}{\partial\psi_k}\middle| y\right) \times E_{b|y,\psi^*}\left(\frac{\partial\left(\log\{p(y|b,\psi^*)p(b|\psi^*)\}\right)}{\partial\psi_l}\middle| y\right). \quad (6)$$

To evaluate (5), Riviere, Ueckert, and Mentré (2016) (see appendix A of supplementary material) proposed using MC to evaluate the expectation with respect to the observations y, by sampling in the marginal distribution of y. Then, for each sampled vector of observations y, they recommend to use HMC for evaluation of the expectation with respect to b given y in (6).

Evaluation of the robust FIM

Let $p_m(\psi_m)$ denote the distribution of the vector of parameters ψ_m of model M_m. In this paragraph, indices m in ψ and p will be omitted for simplicity. The robust FIM for elementary design ξ can be evaluated by computing the expectation of the FIM over $p(\psi)$:

$$\mathcal{M}_R(\xi) = E_\psi\left(\mathcal{M}\{\psi,\xi\}\right). \quad (7)$$

The evaluation of $\mathcal{M}_R(\xi)$ requires computation of one supplementary integral with respect to $p(\psi)$. Using the (k,l) term, the robust FIM can be expressed as:

$$\mathcal{M}_R(\xi)_{k,l} = E_\psi(E_y\{D_{k,l}(y,\psi)\}), \quad (8)$$

where $D_{k,l}(y,\psi)$ is evaluated by HMC as in (6). In (8), one vector of ψ is first sampled. Then, conditioning on each value of ψ, one vector of observations y is sampled, instead of drawing several vectors of observations y for each value of ψ sampled. It follows that the robust FIM can be estimated as:

$$\tilde{\mathcal{M}}_R(\xi)_{k,l} = \frac{1}{n_R}\sum_{r=1}^{n_R} B^{(1)}_{k,r} B^{(2)}_{l,r}, \tag{9}$$

with

$$\begin{aligned} B^{(1)}_{k,r} &= \frac{1}{n_H}\sum_{h=1}^{n_H} \frac{\partial\left(\log\{p(y_r|b^{(1)}_{h,r},\psi_r)p(b^{(1)}_{h,r}|\psi_r)\}\right)}{\partial\psi_k} \\ B^{(2)}_{l,r} &= \frac{1}{n_H}\sum_{h=1}^{n_H} \frac{\partial\left(\log\{p(y_r|b^{(2)}_{h,r},\psi_r)p(b^{(2)}_{h,r}|\psi_r)\}\right)}{\partial\psi_l}, \end{aligned}$$

where $(\psi_r, y_r)_{r=1,\ldots,n_R}$ is an n_R-sample of the joint distribution of (ψ, y), and $(b^{(1)}_{h,r})_{h=1,\ldots,n_H}$ and $(b^{(2)}_{h,r})_{h=1,\ldots,n_H}$ are $2n_R$ n_H-samples of the conditional p.d.f. of b given (ψ_r, y_r). Two independent samples from the posterior of b given (ψ_r, y_r) are used so that $B^{(1)}_{k,r}$ and $B^{(2)}_{l,r}$ are independent. So, even with a small value of the number of HMC samples n_H, a large value of the number of MC samples n_R will lead to convergence of the FIM estimate (Riviere, Ueckert, and Mentré, 2016).

Finally, in order to be symmetric, $\mathcal{M}_R(\xi)$ is obtained as:

$$\mathcal{M}_R(\xi) = \frac{\tilde{\mathcal{M}}_R(\xi) + \tilde{\mathcal{M}}_R(\xi)^T}{2}.$$

Of note, instead of this final symmetry step, one could alternatively fill up the lower part of the matrix by copying the $(l,k)^{\text{th}}$ to the $(k,l)^{\text{th}}$ term. However, the approach we propose does not take more computing time, as the same samples are used to compute $\tilde{\mathcal{M}}_R$ and $\tilde{\mathcal{M}}_R^T$.

This is expected to provide a more precise estimate of $\mathcal{M}_R$, as it makes use of more samples.

2.4 Optimality criteria

Let ψ_m^* be a given *a priori* population parameter vector for a model M_m, $p_m(\psi_m)$ be the *a priori* distribution of the vector of parameters ψ_m of model M_m, and P_m be the number of population parameters of this model. We propose to use four criteria, summarized in Table 1, for design optimization accounting for uncertainty in parameters and/or model.

To obtain the D-optimal design $\Xi_{D,m}$, the D-optimality criterion, corresponding to the normalized determinant of $\mathcal{M}(\psi_m^*, \Xi)$ weighted by the inverse of the number of parameters, is used (expression (10) in Table 1). When accounting for uncertainty in parameters, the DE-optimality criterion (Foo and Duffull, 2010; Foo et al., 2012; Lestini et al., 2016) is used for design optimization (expression (11) in 1). This criterion can be shown to be unimodal as there is some monotone increasing function of this criterion for which the equivalence theorem holds (Atkinson, Donev, and Tobias, 2009; Kiefer, 1974).

Uncertainty in the model choice is introduced by computing an optimality criterion based on the principle of model averaging. A set of M candidate models is considered. The D-optimal design for each model M_m given population parameter values ψ_m^* (with $m = 1, \ldots, M$), can be obtained using the D-optimality criterion. Then, the common design Ξ_{CD} for the M models given their population parameter values can be optimized using the compound D-criterion (CD-criterion, expression (12) in Table 1) inspired by Atkinson (2008) and Nguyen et al. (2016). The CD-criterion corresponds to a product of the D-criteria for the M models, in which each model M_m is associated with a weight, α_m, quantifying the balance between the M models (with $\sum_{m=1}^{M} \alpha_m = 1$). In order to find the CDE-optimal design that is robust with respect to both parameters and model, a compound DE-criterion (CDE-criterion) combining the determinants of the robust FIMs is computed (expression (13) in Table 1). Assuming a

set of M candidate models, the CDE-criterion is then a weighted product of the DE-criterion evaluated for each model M_m. For a given model M_m, the X-relative efficiency of a design Ξ with respect to the optimal design for this model $\Xi_{X,m}$ can then be calculated as follows:

$$E_{X,m}(\Xi) = \frac{\Phi_{X,m}(\Xi)}{\Phi_{X,m}(\Xi_{X,m})} \text{ for } X = \{D, DE\}.$$

Averaging over M models, the X-relative efficiency of a design Ξ with respect to the optimal design for the M models Ξ_X can then be calculated as follows:

$$E_X(\Xi) = \frac{\Phi_X(\Xi)}{\Phi_X(\Xi_X)} \text{ for } X = \{CD, CDE\},$$

which has the usual interpretation in terms of numbers of observations since $\sum_{m=1}^{M} \alpha_m = 1$.

3 Application to design optimization for count data

In this section, we present how the methods introduced in Section 2 were applied to design optimization for discrete data when introducing parameter and/or model uncertainties. An NLMEM for repeated counts at three dose levels with several replications inspired by Riviere, Ueckert, and Mentré (2016); Ogungbenro and Aarons (2011); Nyberg, Karlsson, and Hooker (2009), was considered. The number of subjects N and the number of observations per subject per dose n_{rep} were adapted in order to have satisfactory Relative Standard Errors (RSE) and D-optimality criterion values for three fixed doses. Even if this example seems theoretical here, it can easily be extended to real trials with count data in various settings (Marques and Loingeville, 2016). We first chose the number of MC and HMC samples to use to evaluate the FIM and the robust FIM. As we aim to optimize a design with three dose levels, denoted as $\xi = \{d_1, d_2, d_3\}$, combinatorial optimization of two non-zero dose

levels (d_2, d_3), in addition to d_1 fixed to 0 (i.e placebo), was performed, accounting or not for uncertainty in parameters and in the model.

3.1 Studied models and parameters

The NLMEM studied describes $n_{rep} = 10$ repeated count data for $N = 60$ patients observed at different dose levels. The probability of each observation y_{ij} for patient i was modeled using a Poisson distribution:

$$P(y_{ij} = k|b_i) = \frac{\lambda_i^k exp(-\lambda_i)}{k!}.$$

In the following, index i was omitted for the sake of simplicity. We studied five candidate models for the relationship between $\log(\lambda)$ and the dose level d, which are the full $\mathrm{I_{max}}$ model (M_1), the linear model (M_2), the log-linear model (M_3), the $\mathrm{I_{max}}$ model (M_4), and the quadratic model (M_5). These are traditionally used candidate models for dose-response relationships, inspired from the literature on model averaging in dose finding studies (Buatois et al., 2018; Pinheiro et al., 2014; Maloney et al., 2013; Wedenberg, 2013) and in the R package MCP-Mod (Bornkamp et al., 2009). These models express $\log(\lambda)$ as a function of the dose level d and are represented in Figure 1:

- $M_1 : \log(\lambda) = \beta_1(1 - \frac{d}{d+\beta_2})$,
- $M_2 : \log(\lambda) = \beta_1(1 - \beta_2 d)$,
- $M_3 : \log(\lambda) = \beta_1(1 - \beta_2 \log\{d + 1\})$,
- $M_4 : \log(\lambda) = \beta_1(1 - \frac{\beta_3 d}{d+\beta_2})$,
- $M_5 : \log(\lambda) = \beta_1(-\beta_2 d^2 + \beta_3 d + 1)$.

Model M_1 and its parameters are inspired by Riviere, Ueckert, and Mentré (2016). Of note, after a re-parameterization and re-scaling, this model is a hyperbolic equation with respect to dose. Parameters β_1 and β_2 follow a log-normal distribution $\beta_p = g(\mu_p, b_p) = \mu_p exp(b_p)$ for $p = (1, 2)$, where b_p follows a normal distribution with mean 0 and variance ω_p^2. Parameter values ψ_1^* are displayed in Table 2. As $\log(\lambda)$ for $d = 0$ should be equal to β_1 for each of the five models, uncertainty was introduced only in the parameters μ_2 and ω_2, in order to have the same mean value of $\log(\lambda)$ at $d = 0$ and $d = 1$ for each of the five models. Moreover, often the placebo rate is well characterized, so no uncertainty might be needed on β_1 in those cases. Log-normal *a priori* distributions, which yield exclusively positive values for μ_2 and ω_2, were assigned to these parameters. These log-normal distributions have expectations equal to the *a priori* values μ_2^* and ω_2^* of model M_1, and coefficients of variation (CV) of 70% in μ_2 and 90% in ω_2. Of note, the relationship between the CV and σ, the standard deviation of the log-normal *a priori* distribution, is $CV = \sqrt{\exp(\sigma^2) - 1}$. The expectations and standard deviations for $p_1(\psi_1)$ were then calculated (Table 2).

Models M_2 to M_5 are alternative models with parameters chosen in order to have the same mean value of $\log(\lambda)$ as for model M_1 at the lower and upper boundaries of the dose level interval, i.e. $d = 0$ and $d = 1$. Parameter values ψ_m^* for model M_m (for $m = 2, \ldots, M$), and parameters of the log-normal distributions $p_m(\psi_m)$ when considering uncertainty in parameters are displayed in Table 2. As for model M_1, we chose a CV of 70% in μ_2 and 90% in ω_2. For models M_4 and M_5, for the sake of simplicity and to have similar number of random effects than in models M_1 to M_3, there was no variability in β_3. The proposed approach could easily be applied to models M_4 and M_5 with variability in both β_2 and β_3.

3.2 Choice of the number of MC and HMC samples

The methods proposed for evaluation of $\mathcal{M}$ and $\mathcal{M}_R$ rely on MC-HMC. Therefore, it is essential to choose carefully the number of MC and HMC samples to use in order to obtain accurate approximations. Riviere, Ueckert, and Mentré (2016) recommended using $n_H = 200$ HMC iterations. Indeed, they noticed that at a fixed number of MC samples, improving the number of HMC iterations does not seem to impact the determinant of the FIM. Thus, for good estimation, $n_H = 200$ HMC iterations seems high enough. To choose an appropriate number of MC samples, we studied the convergence of the D-optimality criteria based on $\mathcal{M}$, and of the DE-optimality criteria based on $\mathcal{M}_R$ as functions of the number of MC samples (denoted as n_R) for the five considered models, while fixing the number of HMC iterations at $n_H = 200$. We used here the elementary design $\xi = \{0, 0.4, 0.7\}$. To evaluate $\mathcal{M}$ and $\mathcal{M}_R$ for model M_m, we used population parameters ψ_m^* and $p_m(\psi_m)$, respectively.

The D- and DE-optimality criteria computed seemed stable enough for 5000 MC samples using $n_H = 200$ (Figure 2). Thus, $n_H = 200$ HMC samples and $n_R = 5000$ MC samples were used in the present work with 500 burn-in samples for evaluation of $\mathcal{M}$ and $\mathcal{M}_R$.

3.3 D- and DE-optimal designs for the five models

Assuming the first dose set to zero ($d_1 = 0$), two doses were optimized among the following values: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1. We did not consider repetition, i.e. the three doses were different. This corresponds to $\binom{10}{2} = 45$ possible elementary designs. We performed design optimization for the five models given the parameter values ψ_m^* using the D-optimality criterion. Then, a robust design was optimized with respect to the population parameter values μ_2 and ω_2 using the DE-optimality criterion and the *a priori* distributions $p_m(\mu_{2,m})$ and $p_m(\omega_{2,m})$ in Table 2.

Figure 3 displays the D- and DE-efficiencies obtained when performing combinatorial

optimization of two doses, accounting or not for uncertainty in population parameters of the five models. As expected, we noticed that assuming or not uncertainty in parameters could lead to different optimal designs, as in the case of M_1 or M_4. However, the DE-efficiency of $\Xi_{D,1}$ on model M_1 is 94.1% whereas the DE-efficiency of $\Xi_{D,4}$ on model M_4 is 84.6% showing that the D-optimal designs were quite robust here. Furthermore, different models led to different optimal designs.

Table 3 reports the D-efficiencies of the five D-optimal designs $\Xi_{D,m}$ when the true model is another one. For instance, when using the D-optimal design for model M_1, $\xi_{D,1} = \{0, 0.4, 0.5\}$, with model M_2, we lost almost 40% of efficiency. Worse still, when using the design $\xi_{D,2} = \{0, 0.9, 1\}$, which is optimal for models M_2 and M_3, with model M_4, we lost almost 70% of efficiency. We thus noted that misspecification of models could lead to important loss of D-efficiency.

The DE-efficiencies of the five DE-optimal designs when the model was or was not correctly pre-specified are displayed in Table 4. We noticed that none of the elementary optimal designs obtained for the M models $\xi_{DE,m}$ were good across all models. On the contrary, we observed notable loss of DE-efficiency in the case of model misspecifications as for D-optimal designs.

Of note, the evaluation of the optimality criterion for one model using one elementary design takes about one hour, on a computer i7-5600 U CPU, with frequency 2.60 GHz, 4 cores, 8 GB of RAM.

3.4 Compound D- and Compound DE-optimal designs

To propose a design that is robust with respect to the model, the CD-criterion was evaluated for the combination of the five models considered. To obtain a robust design with respect to the population parameter values μ_2 and ω_2, as well as to the five models, the CDE-criterion

was evaluated. Here, the same weight ($\alpha_m = 1/5$ for model M_m with $m = 1, \ldots, 5$) was used for each model. The CD- and CDE-efficiencies of every possible design are represented in Figure 4.

The CD-optimal design obtained, $\xi_{CD} = \{0, 0.3, 1\}$, was different from the five D-optimal designs obtained separately for each model. When using the CD-optimal design and when the model is actually one of the five models M_1 to M_5, the D-efficiency is at least 79% (Table 3). Thus, the CD-optimal design performed well in terms of D-efficiency and was robust to the choice of models. The CDE-optimal design obtained was $\xi_{CDE} = \{0, 0.2, 1\}$. This design has DE-efficiencies above 82% when assuming one of the five models, showing that it is robust with respect to parameters and model.

4 Discussion

We have proposed an approach based on MC-HMC for design optimization in NLMEMs, to account for uncertainty in the parameters and in the model. MC-HMC has been shown by clinical trial simulations in Riviere, Ueckert, and Mentré (2016) to be a relevant method of computation for multiple integrations over the distribution of observations and of random effects. Therefore, we used it to compute one supplementary integral in the distribution of the population parameters in order to evaluate the robust FIM with uncertainty in parameters. Optimization of design can then be performed using the DE-criterion. To take into account uncertainty in the model, we used a compound optimality criterion which is a weighted product of the D- or DE-criterion for a set of candidate models.

This approach provides the first application to robust design optimization for repeated discrete data. Several previous works proposed to use D-optimality to find design robust to parameter values for count data. Russell et al. (2009) calculated D-optimal designs in Poisson regression models and used clustering techniques to account for uncertainties in pa-

rameter values. Maloney, Simonsson, and Schaddelee (2013) proposed a method for design optimization for three Poisson dose-response models with random effects. This method relied on D-optimality criterion over a range of different parameter values to introduce robustness with respect to parameter values. However, none of these works accounted for robustness with respect to the model.

In this work, we found that, as expected, accounting or not for uncertainty in parameters may lead to different allocations of optimal doses, with little impact on efficiencies. This can be explained by the fact that the prior distributions were chosen centered around the prior guess in our example. However misspecification of models can lead to D-efficiencies as low as 30% for some models. The CD- or CDE-optimal designs, optimized accross the models, were found to provide a good compromise for different candidate models, with D-efficiencies of at least 80% for each model.

In the proposed example, the DE-optimal doses for the full $\mathrm{I}_{\max}$ model M_1 were close to 0.5, i.e. μ_2, the fixed effect on D50, the dose that is required to get 50% of the maximal effect. However, by increasing the coefficient of variation in μ_2 or by assigning non-centered distributions to population parameters, we would expect to obtain different optimal designs with doses far from the D50.

In the example presented in this papier, we considered a Poisson model. In practice, a greater variability than the one expected using a Poisson model can be observed. The methods we proposed in this paper can address this problem by assuming a negative binomial model to account for overdispersion. The proposed methods can also be applied to continuous data, as well as to other types of discrete data, such as binary or time to event data.

Moreover, in the example presented, design optimization of doses was performed with $N = 60$ subjects and $n_{rep} = 10$ replications per subject per dose. More work on the appro-

priate choice of N and n_{rep} given experimental or cost constraints could be conducted, for example by varying (N, n_{rep}) (Retout et al., 2009).

When considering Bayesian D-optimum designs, the ELD criterion, corresponding to the expectation of the logarithm of the determinant of the FIM, could be used instead of the DE criterion Atkinson, Donev, and Tobias (2009). One could compute this criterion using the MC-HMC approach. However, this would require drawing of several vectors of observations for each vector of population parameters sampled. We should consider this criterion after reduction of the execution time of the proposed method. Indeed, performing combinatorial optimization by evaluating the FIM and the robust FIM for every possible design is time consuming. Here we perform optimization for sparse designs, but if we aim at optimizing richer designs, this would result in more elementary designs to evaluate. An alternative approach would therefore consist in implementing an optimization algorithm as suggested as a possibility in the discussion of Riviere, Ueckert, and Mentré (2016). The particle swarm optimization algorithm (Kim and Li, 2011; Chen et al., 2015) could also be a solution.

Riviere, Ueckert, and Mentré (2016) showed that, based on calculation time, AGQ should be recommended as a fast algorithm for simple discrete models, and HMC for complex models with more parameters and random effects. Therefore, an alternative method would be to use MC-AGQ instead of MC-HMC in the evaluation of the robust FIM. MC-AGQ could be used for binary or time-to-event models with a maximum of two random effect parameters. Gotwalt, Jones, and Steinberg (2009) used an AGQ-based method to integrate over the prior distribution of parameters along with the ELD criterion for robust design optimization in nonlinear models and in generalized linear models. This work could be extended to NLMEM.

Besides, to save execution time, MC in MC-HMC and MC-AGQ could be replaced by a more efficient approach, such as quasi-random sampling or especially quasi-random Monte-Carlo (Pan and Thompson, 2007; Ueckert and Mentré, 2015). Quasi-random sam-

pling avoids the exponential growth of the number of quadrature points as the number of random effects increases, and has a higher accuracy than classic MC methods with the same number of function evaluations.

Furthermore, the methods studied in this paper could be combined with adaptive designs (Lestini, Dumont, and Mentré, 2015; Dumont, Chenel, and Mentré, 2016) to further improve designs in longitudinal studies. At each stage of the study, accumulating information will be used to update knowledge of the parameter distributions and of the candidate models to be taken into account in design optimization.

Finally, the methods we proposed for robust designs accounting for parameter and model uncertainties were implemented in R version 3.2.3, extending the R package MIXFIM (Riviere, Ueckert, and Mentré, 2016). We plan to include these new developments in a new version of this package, as well as in PFIM, which is the software program for designing longitudinal studies developed by our laboratory (Dumont et al., 2018) (www.pfim.biostat.fr).

References

Atkinson, A. C. (2008). DT-optimum designs for model discrimination and parameter estimation. *Journal of Statistical Planning and Inference 138*(1), 56–64.

Atkinson, A. C., A. Donev, and R. Tobias (2009). *Optimum Experimental Designs with SAS.* Oxford, Oxford University Press.

Bazzoli, C., S. Retout, and F. Mentré (2009). Fisher information matrix for nonlinear mixed effects multiple response models: evaluation of the appropriateness of the first order linearization using a pharmacokinetic/pharmacodynamic model. *Statistics in Medicine 28*(14), 1940–1956.

Bornkamp, B., J. Pinheiro, F. Bretz, et al. (2009). MCPMod: An R package for the design and analysis of dose-finding studies. *Journal of Statistical Software 29*(7), 1–23.

Buatois, S., S. Ueckert, N. Frey, S. Retout, and F. Mentré (2018). Comparison of model averaging and model selection in dose finding trials analyzed by nonlinear mixed effect models. *The AAPS Journal 20*(3), 56.

Chen, R.-B., S.-P. Chang, W. Wang, H.-C. Tung, and W. K. Wong (2015). Minimax optimal designs via particle swarm optimization methods. *Statistics and Computing 25*(5), 975–988.

Dumont, C., M. Chenel, and F. Mentré (2016). Two-stage adaptive designs in nonlinear mixed effects models: application to pharmacokinetics in children. *Communications in Statistics-Simulation and Computation 45*(5), 1511–1525.

Dumont, C., G. Lestini, H. Le Nagard, F. Mentré, E. Comets, T. T. Nguyen, and the PFIM group (2018). PFIM 4.0, an extended R program for design evaluation and optimization in nonlinear mixed-effect models. *Computer Methods and Programs in Biomedicine 156*, 217–229.

Fedorov, V. V. and S. L. Leonov (2013). *Optimal design for nonlinear response models*. Boca Raton, Chapman and Hall/CRC Press.

Foo, L.-K. and S. Duffull (2010). Methods of robust design of nonlinear models with an application to pharmacokinetics. *Journal of Biopharmaceutical Statistics 20*(4), 886–902.

Foo, L. K., J. McGree, J. Eccleston, and S. Duffull (2012). Comparison of robust criteria for D-optimal designs. *Journal of Biopharmaceutical Statistics 22*(6), 1193–1205.

Gotwalt, C. M., B. A. Jones, and D. M. Steinberg (2009). Fast computation of designs robust to parameter uncertainty for nonlinear settings. *Technometrics 51*(1), 88–95.

Jones, B., J. Wang, P. Jarvis, and W. Byrom (1999). Design of cross-over trials for pharmacokinetic studies. *Journal of Statistical Planning and Inference 78*(1-2), 307 – 316.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *The Annals of Statistics*, 849–879.

Kim, S. and L. Li (2011). A novel global search algorithm for nonlinear mixed-effects models using particle swarm optimization. *Journal of Pharmacokinetics and Pharmacodynamics 38*(4), 471–495.

Lestini, G., C. Dumont, and F. Mentré (2015). Influence of the size of cohorts in adaptive design for nonlinear mixed effects models: an evaluation by simulation for a pharmacokinetic and pharmacodynamic model for a biomarker in oncology. *Pharmaceutical Research 32*(10), 3159–3169.

Lestini, G., S. Ueckert, and F. Mentré (2016). Robust design in model-based analysis of longitudinal clinical data. *PODE 2016*, Number Abstr 5. Available from: http://www.maths.qmul.ac.uk/ bb/PODE/PODE2016.html.

Maloney, A., U. S. Simonsson, and M. Schaddelee (2013). D optimal designs for three poisson dose–response models. *Journal of Pharmacokinetics and Pharmacodynamics 40*(2), 201–211.

Marques, F. J. and F. Loingeville (2016). Improved near-exact distributions for the product of independent generalized gamma random variables. *Computational Statistics & Data Analysis 102*, 55–66.

Mentré, F., A. Mallet, and D. Baccar (1997). Optimal design in random-effects regression models. *Biometrika 84*(2), 429–442.

Mould, D. and R. Upton (2012). Basic concepts in population modeling, simulation, and model-based drug development. *CPT: Pharmacometrics & Systems Pharmacology 1*(9), 1–14.

Mould, D. and R. Upton (2013). Basic concepts in population modeling, simulation, and modelbased drug development: Part 2: Introduction to pharmacokinetic modeling methods. *CPT: Pharmacometrics & Systems Pharmacology 2*(4), 1–14.

Nguyen, T. T., H. Bénech, M. Delaforge, and N. Lenuzza (2016). Design optimisation for pharmacokinetic modeling of a cocktail of phenotyping drugs. *Pharmaceutical Statistics 15*(2), 165 – 177.

Nguyen, T. T. and F. Mentré (2014). Evaluation of the Fisher information matrix in nonlinear mixed effect models using adaptive Gaussian quadrature. *Computational Statistics & Data Analysis 80*(0), 57 – 69.

Nyberg, J., M. O. Karlsson, and A. C. Hooker (2009). Population optimal experimental design for discrete type data. *PAGE 18*, Number Abstr 1468. Available from: www.page-meeting.org/?abstract=1468.

Ogungbenro, K. and L. Aarons (2011). Population Fisher information matrix and optimal design of discrete data responses in population pharmacodynamic experiments. *Journal of Pharmacokinetics and Pharmacodynamics 38*(4), 449–469.

Pan, J. and R. Thompson (2007). Quasi-Monte Carlo estimation in generalized linear mixed models. *Computational Statistics & Data Analysis 51*(12), 5765–5775.

Pinheiro, J., B. Bornkamp, E. Glimm, and F. Bretz (2014). Model-based dose finding under model uncertainty using general parametric models. *Statistics in Medicine 33*(10), 1646–1661.

Pronzato, L. and A. Pázman (2013). Design of experiments in nonlinear models. *Lecture Notes in Statistics 212*.

Pronzato, L. and É. Walter (1985). Robust experiment design via stochastic approximation. *Mathematical Biosciences 75*(1), 103–120.

Retout, S., E. Comets, C. Bazzoli, and F. Mentré (2009). Design optimization in nonlinear mixed effects models using cost functions: application to a joint model of infliximab and methotrexate pharmacokinetics. *Communications in Statistics - Theory and Methods 38*(18), 3351–3368.

Riviere, M.-K. and F. Mentré (2015). R package MIXFIM, version 1.0. `https://cran.r-project.org/web/packages/MIXFIM/`.

Riviere, M.-K., S. Ueckert, and F. Mentré (2016). An MCMC method for the evaluation of the Fisher information matrix for non-linear mixed effect models. *Biostatistics 17*(4), 737–750.

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

Stan Development Team (2016). RStan: the R interface to stan, version 2.12.0. `http://mc-stan.org/`.

Ueckert, S. and F. Mentré (2015). Computation of the Fisher information matrix for discrete nonlinear mixed effect models. *CMStatistics 2015*, Number Abstr EO0936.

Ueckert, S. and F. Mentré (2016). A new method for evaluation of the Fisher information matrix for discrete mixed effect models using Monte Carlo sampling and adaptive Gaussian quadrature. *Computational Statistics & Data Analysis 111*, 203–219.

Upton, R. and D. Mould (2014). Basic concepts in population modeling, simulation, and model-based drug development: Part 3: Introduction to pharmacodynamic modeling methods. *CPT: Pharmacometrics & Systems Pharmacology 3*(1), 1–16.

Walter, É. and L. Pronzato (1987). Optimal experiment design for nonlinear models subject to large prior uncertainties. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology 253*(3), R530–R534.

Wedenberg, M. (2013). *From cell survival to dose response: modeling biological effects in radiation therapy*. Karolinska Universitetssjukhuset, Dept of Oncology-Pathology, Solna, Sweden.

Table 1: Optimality criteria.

Parameters \ Model	Given model M_m	Set of candidate models M_m $(m = 1, \ldots, M)$
Given parameter values ψ_m^*	**D-optimality** $\Phi_{D,m}(\Xi) = \det(\mathcal{M}(\Psi_m^*, \Xi))^{1/P_m}$ (10)	**Compound-D optimality** $\Phi_{CD}(\Xi) = \prod_{m=1}^{M} \Phi_{D,m}(\Xi)^{\alpha_m}$ (12)
a priori distribution on parameters $p_m(\psi_m)$	**DE-optimality** $\Phi_{DE,m}(\Xi) = \det(\mathcal{M}_R(\Xi))^{1/P_m}$ (11)	**Compound-DE optimality** $\Phi_{CDE}(\Xi) = \prod_{m=1}^{M} \Phi_{DE,m}(\Xi)^{\alpha_m}$ (13)

$\mathcal{M}(\Psi_m^*, \Xi))$ is the Fisher information matrix (FIM) with population parameters Ψ_m^* and design Ξ, and $\mathcal{M}_R(\Xi))$ is the robust FIM with design Ξ. P_m is the number of population parameters of model M_m and α_m is the weight quantifying the balance between the M models ($\sum_{m=1}^{M} \alpha_m = 1$).

Table 2: Population parameter values ψ_m^* and distributions $p_m(\psi_m)$ for model M_m with $m = 1, \ldots, 5$.

	ψ_m^*					$p_m(\psi_m)$				
	μ_1^*	μ_2^*	μ_3^*	ω_1^*	ω_2^*	μ_1	μ_2	μ_3	ω_1	ω_2
M_1	1.00	0.50		0.30	0.30	1.00	$\mathcal{LN}(-0.89, 0.63)$		0.30	$\mathcal{LN}(-1.50, 0.77)$
M_2	1.00	0.67		0.30	0.30	1.00	$\mathcal{LN}(-0.60, 0.63)$		0.30	$\mathcal{LN}(-1.50, 0.77)$
M_3	1.00	0.96		0.30	0.30	1.00	$\mathcal{LN}(-0.24, 0.63)$		0.30	$\mathcal{LN}(-1.50, 0.77)$
M_4	1.00	0.20	0.80	0.30	0.30	1.00	$\mathcal{LN}(-1.81, 0.63)$	0.80	0.30	$\mathcal{LN}(-1.50, 0.77)$
M_5	1.00	0.80	0.13	0.30	0.30	1.00	$\mathcal{LN}(-0.42, 0.63)$	0.13	0.30	$\mathcal{LN}(-1.50, 0.77)$

μ_p is the fixed effect and ω_p the standard deviation of the random effect for parameter β_p; ω_3 is assumed to be 0. Parameters of the log-normal distribution for uncertainty in μ_2 and ω_2 were chosen to have an expectation equal to μ_2^* and ω_2^* respectively and a coefficient of variation equal to 70% and 90% in μ_2 and ω_2 respectively.

Table 3: D-efficiencies of the five D-optimal designs $\Xi_{D,m} = \{N = 60, \xi_{D,m}\}$ for model M_m with $m = 1, \ldots, 5$ when the model is or is not correctly pre-specified. Efficiencies lower than 80% are highlighted in bold. D-efficiencies of the CD-optimal design $\Xi_{CD} = \{N = 60, \xi_{CD}\}$ across all models are indicated in the last line.

	$E_{D,1}(\Xi)$	$E_{D,2}(\Xi)$	$E_{D,3}(\Xi)$	$E_{D,4}(\Xi)$	$E_{D,5}(\Xi)$	$E_{CD}(\Xi)$
$\xi_{D,1} = \{0, 0.4, 0.5\}$	100.0%	**60.8%**	**68.9%**	**50.3%**	**27.7%**	**65.1%**
$\xi_{D,2} = \{0, 0.9, 1\}$	87.0%	100.0%	100.0%	**30.8%**	**67.2%**	82.3%
$\xi_{D,3} = \{0, 0.9, 1\}$	87.0%	100.0%	100%	**30.8%**	**67.2%**	82.3%
$\xi_{D,4} = \{0, 0.2, 1\}$	88.4%	85.7%	85.4%	100.0%	85.6%	98.0%
$\xi_{D,5} = \{0, 0.5, 1\}$	94.6%	89.9%	91.7%	**69.9%**	100.0%	98.5%
$\xi_{CD} = \{0, 0.3, 1\}$	94.1%	88.1%	88.5%	79.7%	93.1%	100.0%

Table 4: DE-efficiencies of the five DE-optimal designs $\Xi_{DE,m} = \{N = 60, \xi_{DE,m}\}$ for model M_m with $m = 1, \ldots, 5$ when the model is or is not correctly pre-specified. Efficiencies lower than 80% are highlighted in bold. DE-efficiencies of the CDE-optimal design $\Xi_{CDE} = \{N = 60, \xi_{CDE}\}$ across all models are indicated in the last line.

	$E_{DE,1}(\Xi)$	$E_{DE,2}(\Xi)$	$E_{DE,3}(\Xi)$	$E_{DE,4}(\Xi)$	$E_{DE,5}(\Xi)$	$E_{CDE}(\Xi)$
$\xi_{DE,1} = \{0, 0.2, 0.4\}$	100.0%	**46.9%**	**56.7%**	**77.5%**	**23.6%**	**63.5%**
$\xi_{DE,2} = \{0, 0.9, 1\}$	**73.3%**	100.0%	100.0%	**43.5%**	87.1%	89.9%
$\xi_{DE,3} = \{0, 0.9, 1\}$	**73.3%**	100.0%	100.0%	**43.5%**	87.1%	89.9%
$\xi_{DE,4} = \{0, 0.1, 0.7\}$	89.1%	**68.1%**	**73.9%**	100.0%	**51.4%**	86.6%
$\xi_{DE,5} = \{0, 0.5, 1\}$	83.1%	87.8%	89.6%	**58.5%**	100.0%	95.8%
$\xi_{CDE} = \{0, 0.2, 1\}$	90.9%	83.8%	83.9%	84.6%	82.8%	100.0%

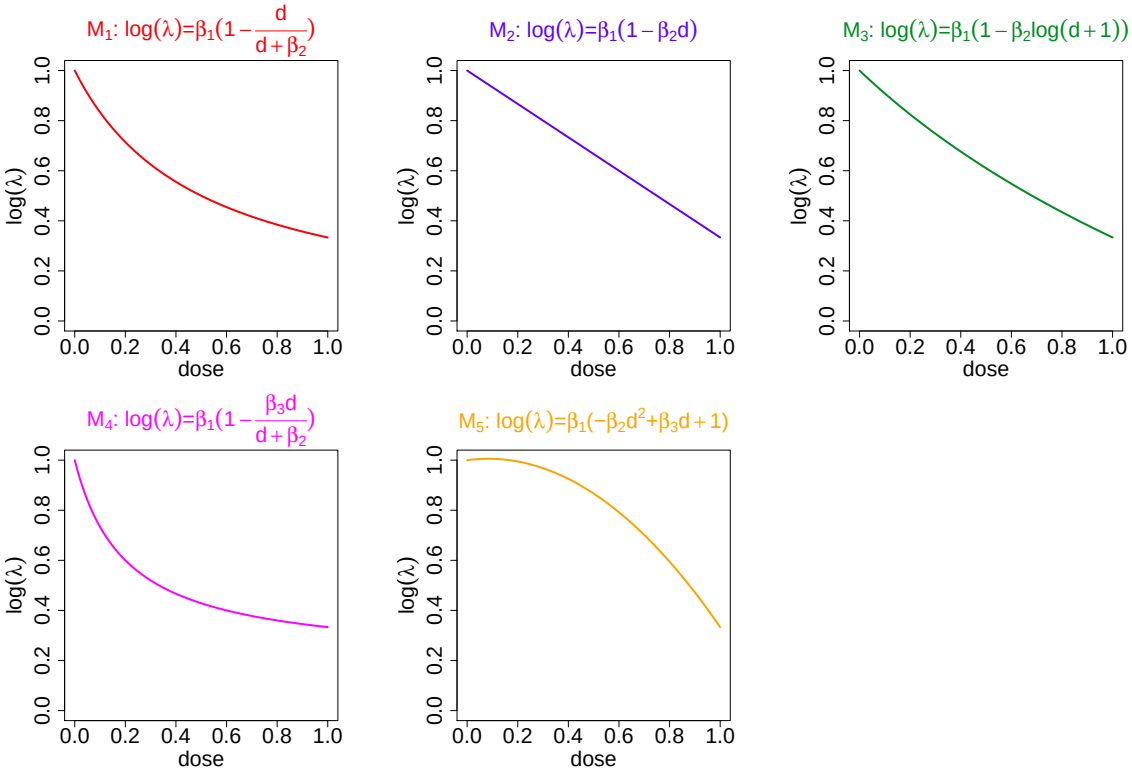


Figure 1: Representation of $\log(\lambda)$, where λ is the parameter of the Poisson model, as a function of the dose level d for model M_m $(m = 1, \ldots, 5)$, with parameters ψ_m^* indicated in Table 2.

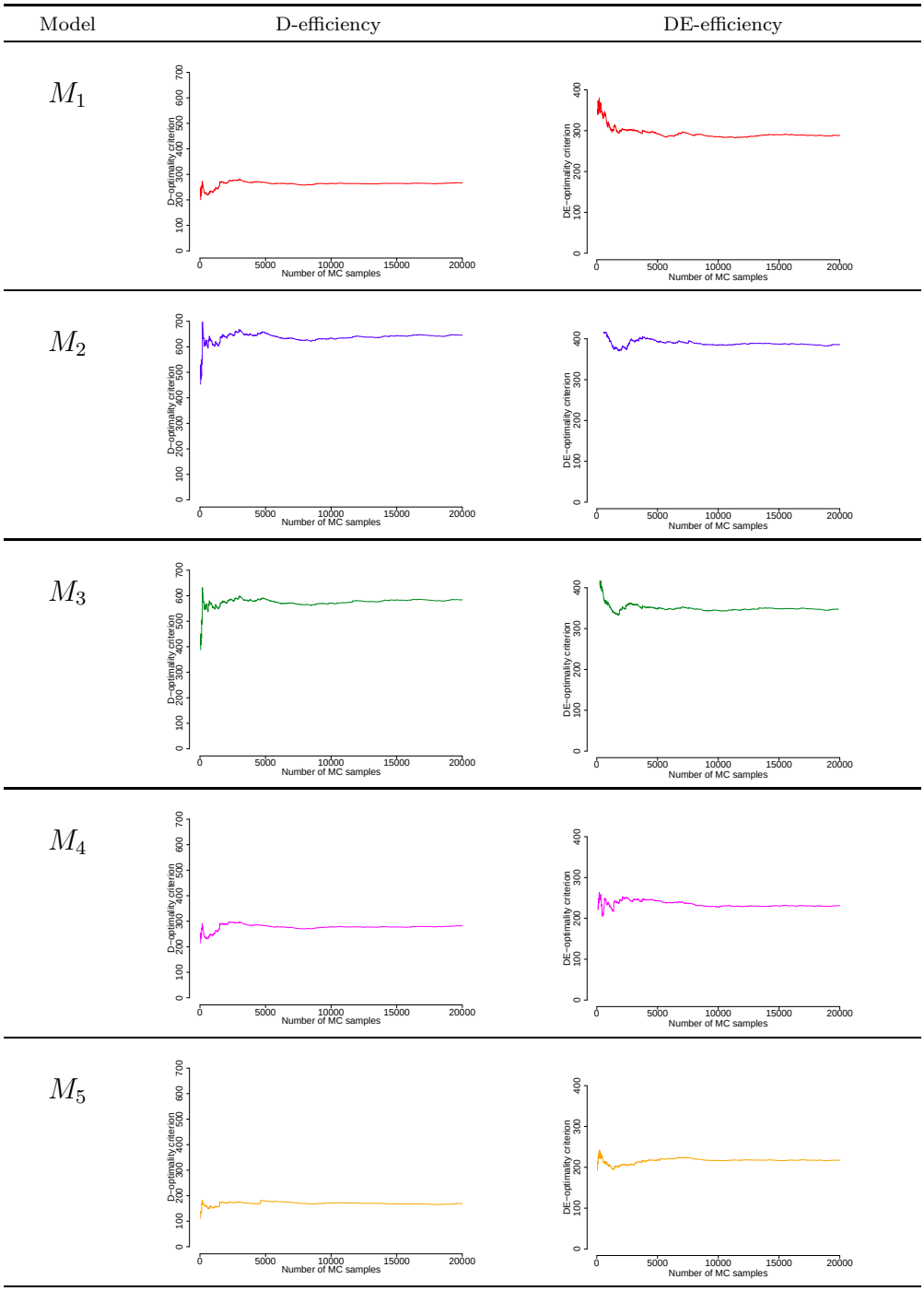


Figure 2: Convergence of the D- (left) and DE- (right) optimality criteria with respect to the number of MC samples for the MC-HMC method, for model M_m ($m = 1, \ldots, 5$) with population parameters ψ_m^* and *a priori* distribution on parameters $p_m(\psi_m)$ respectively. $N = 60$ patients and $n_{rep} = 10$ replications per patient per dose were considered, with elementary design $\xi = \{0, 0.4, 0.7\}$, $n_H = 200$ HMC samples, and 500 burn-in trials.

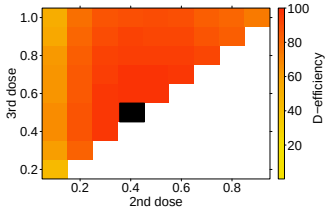
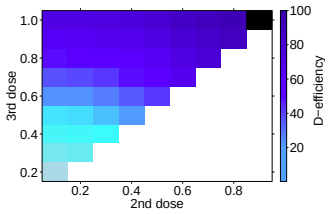
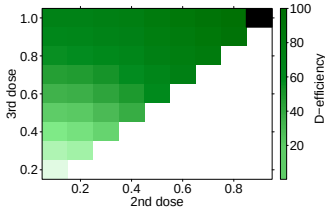
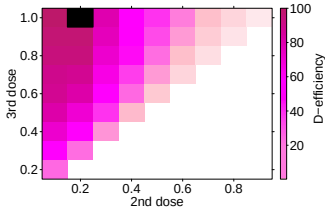
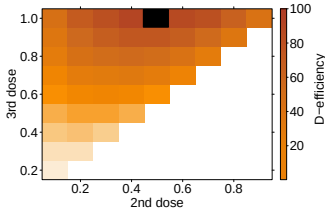
Model	D-efficiency	DE-efficiency
M_1		
	Optimal design: $\xi = \{0, 0.4, 0.5\}$	**Optimal design:** $\xi = \{0, 0.2, 0.4\}$
M_2		
	Optimal design: $\xi = \{0, 0.9, 1\}$	**Optimal design:** $\xi = \{0, 0.9, 1\}$
M_3		
	Optimal design: $\xi = \{0, 0.9, 1\}$	**Optimal design:** $\xi = \{0, 0.9, 1\}$
M_4		
	Optimal design: $\xi = \{0, 0.2, 1\}$	**Optimal design:** $\xi = \{0, 0.1, 0.7\}$
M_5		
	Optimal design: $\xi = \{0, 0.5, 1\}$	**Optimal design:** $\xi = \{0, 0.5, 1\}$

Figure 3: Heatmaps for optimization of doses d_2 and d_3 in Poisson models M_m with $m = 1, \ldots, 5$ for repeated count response without (left) or with uncertainty (right) in population parameters ψ_m as specified in Table 2. The resulting D- or DE-efficiencies (%) using the FIM or robust FIM respectively are reported for all possible designs of two doses from 0.1 to 1 (with a first dose $d_1 = 0$ fixed). The black squares correspond to the D- and DE-optimal designs.

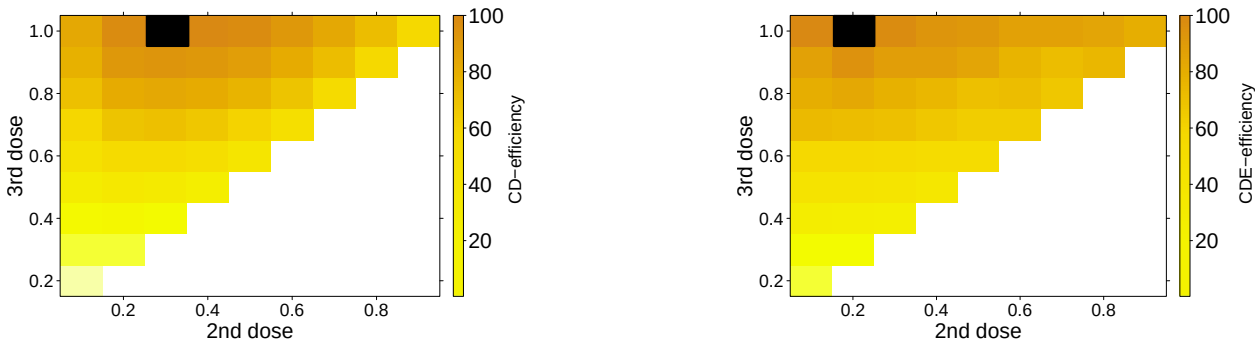


Figure 4: Compound-D-efficiencies (left) and Compound-DE-efficiencies (right) in a combination of the five Poisson models M_1 to M_5, with $N = 60$ patients and $n_{rep} = 10$ replications per patients per dose. The considered CD- and CDE-efficiencies correspond to a balanced combination of the five Poisson models. The black squares correspond to the CD- ($\Xi_{CD} = \{N = 60, \xi_{CD}\}$) and CDE- ($\Xi_{CDE} = \{N = 60, \xi_{CD}\}$) optimal designs.