



HAL
open science

Robust and Adaptive Two-stage Designs in Nonlinear Mixed Effect Models

Lucie Fayette, Romain Leroux, France Mentré, Jérémy Seurat

► **To cite this version:**

Lucie Fayette, Romain Leroux, France Mentré, Jérémy Seurat. Robust and Adaptive Two-stage Designs in Nonlinear Mixed Effect Models. *AAPS Journal*, 2023, 25 (4), pp.71. 10.1208/s12248-023-00810-9 . inserm-04398222

HAL Id: inserm-04398222

<https://inserm.hal.science/inserm-04398222>

Submitted on 16 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL
open science

Robust and Adaptive Two-stage Designs in Nonlinear Mixed Effect Models

Lucie Fayette, Romain Leroux, France Mentré, Jérémy Seurat

► **To cite this version:**

Lucie Fayette, Romain Leroux, France Mentré, Jérémy Seurat. Robust and Adaptive Two-stage Designs in Nonlinear Mixed Effect Models. *AAPS Journal*, 2023, 25 (4), pp.71. 10.1208/s12248-023-00810-9 . inserm-04398222

HAL Id: inserm-04398222

<https://inserm.hal.science/inserm-04398222>

Submitted on 16 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Robust and adaptive two-stage designs in nonlinear mixed effect models

Lucie Fayette^{1,2}, Romain Leroux¹, France Mentré¹
and Jérémy Seurat^{1*}

¹Université Paris Cité and Université Sorbonne Paris Nord,
Inserm, IAME, F-75018 Paris, France.

²École des Ponts, UGE, Champs-sur-Marne, France.

*Corresponding author(s). E-mail(s): jeremy.seurat@inserm.fr;

Abstract

To get informative studies for nonlinear mixed effect models (NLMEM), design optimization can be performed based on Fisher Information Matrix (FIM) using the D-criterion. Its computation requires knowledge about models and parameters, which are often prior guesses. Thus, adaptive designs composed of several stages may be used. Robust approach can also be used to account for various candidate models. *In the* estimation step of a *given* stage, model selection (MS) or model averaging (MA) can be performed. In this work we propose a new two-stage adaptive design strategy, based on the robust expected FIM and MA over several candidate models. The methodology is applied to a clinical trial simulation in ophthalmology to optimize doses and time measurements. A set of dose-response candidate models is defined, and one-stage designs are compared to two-stage 50/50 designs (i.e., each stage performed with half of the available subjects), using either local optimal design or robust design, and performing analysis with one model, MS or MA. Performing a two-stage design with MS at the interim analysis can correct the choice of a wrong model for designing the first stage. Overall, starting from a robust design (1- or 2-stage) is valuable and leads to reasonable bias and precision. The proposed robust adaptive design strategy is a new tool to design longitudinal studies that could be used in different therapeutic areas.

Keywords: Adaptive Design, Non Linear Mixed Effect Model, Fisher Information Matrix, Model Averaging, Optimal design

Introduction

In pharmacometrics, the science of quantitative pharmacology, nonlinear mixed effect models (NLMEM) are the most appropriate tool, as they can exploit the richness of longitudinal pharmacokinetics (PK) and/or pharmacodynamics (PD) data, even if few samples are collected from each individual [1–3]. The use of NLMEM (i.e. the population approach) allows the estimation of mean parameters, inter and intra-individual variabilities as well as covariate effects.

Before a study is carried out, it is possible to evaluate a design to know whether or not enough information will be collected to meet its objectives. Indeed, poor designs lead to poor precision of parameter estimates, and therefore to inconclusive studies [4]. Moreover, algorithms can be used to optimize a design, accounting for the imposed practical constraints. *A design in NLMEM, called a population design, is composed of a number of subjects and a corresponding set of elementary designs.* An elementary design represents the design of an individual. The elementary design elements *include the dosing regimen and the number and allocation of PK and PD measurement times* [5]. In practice groups of individuals with identical elementary design are defined. *Then, two approaches have been proposed to evaluate and optimize designs.* The first approach, based on clinical trial simulation (CTS), is very time-consuming and therefore *limits the number of evaluated population designs*[6]. Alternatively, the Fisher information matrix (FIM) can be used, *its inverse being defined as the lower bound of the variance–covariance matrix of any unbiased estimated parameters.* Several methods exist to evaluate the FIM, the most commonly used is based on first order (FO) linearization of the model around the expectation of random effects [7]. *The expected relative standard errors (RSE) can then be calculated from FIM, as well as the widely used D-criterion which is based on the FIM determinant. The latter accounts for the precision of estimation of all model parameters. It is therefore maximized to find the optimal design.*

To evaluate or optimize a design, knowledge is needed about *the model and its parameters*, which are however often prior guesses at this step. *There are several possibilities to deal with these situations.* Among them, adaptive designs (ADs) are a promising tool, as they offer flexibility compared to traditional one-stage studies [8]. These designs are composed of several stages. The first stage of an AD is defined from what is known before the study, as for a non-adaptive design. The data are then analyzed at the end of the first stage (i.e. after the inclusion of the first cohort). The design can be refined for the next stages of the study depending on the results obtained in the previous stages. The model and parameters uncertainties are therefore reduced after each stage. *Interestingly, AD also allows to halt the study if an interim analysis reaches a stopping criterion.* (efficacy or futility) [9]. Several studies have been published in the field of AD for NLMEM, proposing to use the FIM [10] or CTS [11] to choose the design of each step. It was shown that performing

a two-stage design instead of one-stage is relevant in pharmacometrics examples, especially if the total sample size is halved and allocated to each stage (50/50) [10, 12]. *However, the gain in efficiency* generated by the realization of more than two stages does not seem to compensate the cost and the clinical constraints of such studies [12, 13]. *Moreover, interim analyses inflate the type I error if no control method is used* [14]. *Therefore in our work, only designs with a reduced number of stages are considered and compared: one-stage and two-stage 50/50 designs.*

Another possibility is to use robust design methods that account for model and/or parameter uncertainty [15]. It has been *shown* that accounting for model uncertainty could be *more relevant than for parameter uncertainty* [16]. In this approach, several candidate models with a corresponding weight are set to calculate the robust FIM-based optimality criterion (i.e. to find the robust optimal design [17]).

At estimation step, pharmacometricians classically perform model selection (MS) by using the model with the lowest likelihood-based criterion (e.g. Akaike information criterion, AIC or Bayesian information criterion, BIC). Alternatively, *one can perform model averaging (MA) between different candidate models by weighting appropriately each of them (e.g. with AIC)*. MA leads to better predictive performances than MS, especially in case of sparse data [18, 19].

This work aims to propose a method combining the approaches of model-based AD and robust design. It also aims to compare *MS and MA* at estimation step, and to combine these methods with AD and robust designs. Based on a disease progression example with dose and measurement *time* optimization, simulations are performed to assess which approach or combination of approaches is *the* most efficient to estimate the dose-response relationship.

In material and methods, the notations and methods are detailed. Then, an application example is shown. Finally, results are presented, then discussed.

Material and Methods

In this section we present notations and methods on study design as well as mixed effects models and the Fisher information matrix. After that, optimality criteria for a design with a given model and with model uncertainty are shown. Then, the robust adaptive design strategy and the optimization procedure are detailed.

Design

The population design is denoted $\Xi = \{N, (\xi_1, \dots, \xi_N)\}$, with N the number of individuals, and (ξ_1, \dots, ξ_N) the set of elementary designs. ξ_i is the elementary design of the i -th subject and we denote $\xi_i = (d_i, (t_{i1}, \dots, t_{in_i}))$, where d_i is the dose received, t_{ij} the j -th sampling time, and n_i the total number of samples for the subject i .

Let Q denote the total number of possible *unique* elementary designs ξ_q ,

indexed by $q = 1, \dots, Q$. This design space is predefined through doses and time measurements constraints. Denoting n_q the number of *subjects* whose individual design is ξ_q , the population design is: $\Xi = \{(\xi_1, n_1), \dots, (\xi_Q, n_Q)\}$.

The population design can also be expressed with the weights associated to these possible elementary designs across the population. Let $0 \leq \alpha_q = \frac{n_q}{N} \leq 1$ denote the weight given to the individual design ξ_q in the population design Ξ , such that $\sum_q \alpha_q = 1$. Thus we have $\Xi = \{N, \Xi_\alpha\}$, where $\Xi_\alpha = \begin{pmatrix} \xi_1 & \dots & \xi_Q \\ \alpha_1 & \dots & \alpha_Q \end{pmatrix}$.

In the framework of K-stage design, K population designs Ξ_1, \dots, Ξ_K are defined for groups of N_1, \dots, N_K individuals, such that $\sum_k N_k = N$.

Statistical models

In this work M candidate nonlinear mixed effects models $m = 1, \dots, M$ are considered to describe observation y_i with design ξ_i . They all have the following form:

$$y_i = f_m(\theta_i, \xi_i) + \varepsilon_i \quad (1)$$

where $\theta_i = g(\mu_m, b_i)$ is the vector of individual parameters, depending on μ_m , the vector of fixed effects and b_i , the vector of random effects ($b_i \sim \mathcal{N}(0, \Omega_m)$), and ε_i is the vector of residual errors, assuming independent ε_i for each subject, conditionally to random effects. It follows a normal distribution of mean 0 and variance Σ_i ($\varepsilon_i \sim \mathcal{N}(0, \Sigma_i)$), where Σ_i is a diagonal matrix of size n_i defined as: $\Sigma_i = \text{diag}(\sigma_{inter,m} + \sigma_{slope,m} \times f_m(\theta_i, \xi_i))^2$, i.e. a combined error model. If $\sigma_{slope,m} = 0$, the error model is additive with a standard error of $\sigma_{inter,m}$, whereas if $\sigma_{inter,m} = 0$, the error is proportional to the model with a standard error of $\sigma_{slope,m}$. The vector of population parameters, of length P_m , is denoted ψ_m with $\psi_m^T = (\mu_m^T, \Omega_{m,u}^T, \sigma_{inter,m}, \sigma_{slope,m})$, with $\Omega_{m,u}$ the vector containing the unique elements of Ω_m .

Fisher Information Matrix

Given a model m and its parameters values ψ_m and assuming that the N individuals are independent, the Fisher Information Matrix for the population design ξ is the weighted sum of individual FIMs:

$$\mathcal{M}_F(\psi_m, \Xi) = N \sum_{q=1}^Q \alpha_q \mathcal{M}_F(\psi_m, \xi_q) \quad (2)$$

For individual i , the elementary FIM $\mathcal{M}_F(\psi_m, \xi_i)$ for the elementary design ξ_i is defined as:

$$\mathcal{M}_F(\psi_m, \xi_i) = \mathbb{E} \left[\frac{\partial \log L(\psi_m, y_i)}{\partial \psi_m} \frac{\partial \log L(\psi_m, y_i)^T}{\partial \psi_m} \right] \quad (3)$$

where $\log L(\psi_m, y_i)$ is the log-likelihood of the vector of observations for individual i and parameters ψ_m , given by:

$$\log L(\psi_m, y_i) = \log \left(\int_{\mathbb{R}} p(y_i|b_i, \psi_m)p(b_i|\psi_m)db_i \right) \quad (4)$$

with $p(y_i|b_i, \psi_m)$ the probability density function of y_i given parameters ψ_m and random effects b_i , and with $p(b_i|\psi_m)$ the probability density function of b_i given ψ_m .

Optimality criteria

The D-criterion ϕ_D is a widely used optimality criterion in [the](#) field of optimal design. For a given model m , it writes:

$$\phi_{D,m}(\Xi) = \text{Det}(\mathcal{M}_F(\psi_m, \Xi))^{1/P_m} = \text{Det} \left(N \sum_{q=1}^Q \alpha_q \mathcal{M}_F(\psi_m, \xi_q) \right)^{1/P_m} \quad (5)$$

with $P_m = \dim(\psi_m)$ and $\text{Det}(A)$ the determinant of the matrix A .

The relative D-efficiency $E_{D,m}(\Xi)$ of a design Ξ with respect to a reference design Ξ^* is computed as:

$$E_{D,m}(\Xi) = \frac{\phi_{D,m}(\Xi)}{\phi_{D,m}(\Xi^*)} \quad (6)$$

The locally optimal design for a model m according to the D-criterion will be denoted $\Xi_{LOD,m}$ in the following.

In the context of robust design, the aim is to account for the model uncertainty, and therefore to compute an optimality criterion based on the theory of compound optimality[20]. Assuming that a set of M models, denoted \mathcal{M} , is available, each model being associated [with](#) a weight $0 \leq w_m \leq 1$ such

that $\sum_{m=1}^M w_m = 1$, the CD-criterion is defined as the weighted product of the D-criteria from the different models:

$$\phi_{CD}(\Xi) = \prod_{m=1}^M (\phi_{D,m}(\Xi))^{w_m} \quad (7)$$

The CD-optimal design, known as the robust design, will be denoted $\Xi_{CD} = \Xi_{Robust}$ in the following.

FIM based Adaptive Designs

In the general setting of robust and adaptive design with K stages, the first stage of the study is designed and conducted according to previous information

and accounting for model uncertainty. Then each following stage is planned *by* leveraging the data obtained in the previous stages. Indeed, the FIM at stage k depends on the design of the k^{th} stage, but also on those of all previous stages, from 1 to $k - 1$. Thus, for a given model m (index m is omitted in the following equations for the sake of simplicity), the total FIM $\mathcal{M}_F^{(k)}(\hat{\psi}^{(k-1)}, \Xi)$ at stage k , for a design Ξ and given estimated parameters $\hat{\psi}^{(k-1)}$ from stage $k - 1$, writes:

$$\begin{aligned} \mathcal{M}_F^{(k)}(\hat{\psi}^{(k-1)}, \Xi) &= \mathcal{M}_F(\hat{\psi}^{(k-1)}, \Xi_{CD,1}) + \mathcal{M}_F(\hat{\psi}^{(k-1)}, \Xi_{CD,2}) \\ &+ \dots + \mathcal{M}_F(\hat{\psi}^{(k-1)}, \Xi_{CD,k-1}) \\ &+ \mathcal{M}_F(\hat{\psi}^{(k-1)}, \Xi) \end{aligned} \quad (8)$$

where the $\Xi_{CD,j}$ is the Ξ_{Robust} computed at stage j .

Two different strategies can be chosen to conduct the design optimization at stage k . First, the model whose AIC is the lowest at stage $k - 1$ may be selected, and the k^{th} stage design determined as the locally optimal design for this selected model, using its estimated parameters at stage $k - 1$. Otherwise, the robust design can be determined for the stage k , *using the candidate models with weights updated according to Eq. 9*.

$$w_m^{(k)} = \frac{e^{-\frac{AIC_m^{(k-1)}}{2}}}{\sum_{m'=1}^M e^{-\frac{AIC_{m'}^{(k-1)}}{2}}} \quad (9)$$

where $w_m^{(k)}$ is the weight of the m^{th} candidate model at stage k , with $AIC_m^{(k-1)}$, the AIC of the m^{th} *candidate model* at stage $k - 1$. *To design the first stage, as no data has been collected and fitted yet, the same weight is assigned for each model (total uncertainty) i.e. $w_m^{(1)} = 1/M$.*

Workflow for robust adaptive two-stage design is given on Figure 1.

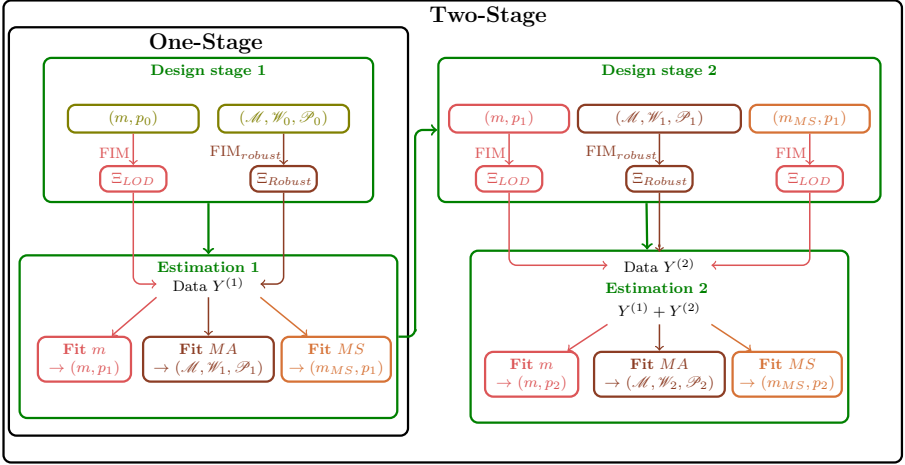


Fig. 1 Design and Estimation Options: One-stage or Two-Stage, Locally Optimal or Robust, fitting with one model, MA or MS
 (m, p) refers to a model m and an associated set of parameters p , $(\mathcal{M}, \mathcal{W}_k, \mathcal{P}_k)$ is a collection of models \mathcal{M} with their respective estimated parameters \mathcal{P}_k and weights \mathcal{W}_k after a given stage k - Ξ_{LOD} is a Locally Optimal Design, while Ξ_{Robust} is a Robust Design - MA and MS respectively stand for Model Averaging and Model Selection (m_{MS} refers to the selected model)

Design optimization procedure

The multiplicative algorithm, *with* either D-criterion for local optimal designs, or CD-criterion for robust designs, is used to find optimal designs. More details on *the* computation are given in Appendix A. Designs returned by this algorithm present groups with non-integer number of subjects. Therefore we defined the following rounding procedure.

Let $\Xi = \{(\xi_1, W_1), \dots, (\xi_{Q'}, W_{Q'})\}$ denote the design returned by the multiplicative algorithm, Q' being the number of individual designs. In particular, we have $\sum_{q'=1}^{Q'} W_{q'} = N$, with N the total number of subjects at the considered stage. Our aim is to round $W_{q'}$ to $n_{q'} = \alpha_{q'} N$ to obtain a design $\Xi_{rounded}$:

$$\Xi_{rounded} = \{N, \Xi_{rounded, \alpha}\} \text{ with } \Xi_{rounded, \alpha} = \begin{pmatrix} \xi_1 & \dots & \xi_{Q'} \\ \alpha_1 & \dots & \alpha_{Q'} \end{pmatrix} \quad (10)$$

such that for all $q' = 1, \dots, Q'$, the number of subjects having the individual design $\xi_{q'}$ is an integer, i.e. $n_{q'} \in \mathbb{N}$, and $\sum_{q'=1}^{Q'} n_{q'} = N$.

To be more realistic, only groups with at least n_{min} subjects are kept, where n_{min} is chosen according to the context. After discarding the smallest arms, the remainders' weights are normalized. Arm sizes are rounded to *integers* using an efficient rounding method, based on multipliers for apportionment [21].

Evaluation by simulation

Our clinical trial simulation is based on the one proposed by Buatois et al. (see [18]) which is itself inspired by a study on a monoclonal antibody indicated in the treatment of wet age-related macular degeneration (wet-AMD). This disease leads to a reduction in visual acuity (VA). The latter is the number of letters the patient can successively read on a given table.

Models

As in [18], the disease progression model for individuals suffering from wet-AMD is supposed to be known and is modelled as follows:

$$f(\theta, \xi) = VA(\theta, d, t) = VA_0 + (1 - e^{-kt})(E(d) - \beta \times VA_0) \quad (11)$$

with VA_0 the visual acuity at baseline, k the rate of the exponential decrease in VA to the asymptote $VA_0(1 - \beta)$ without treatment, and $E(d)$ the dose effect relationship.

In the following, the "true" model for $E(d)$ is the Emax model, given by Eq. 12c, meaning that it is used to simulate the data. Three additional models are also considered as candidate models for the investigation: Linear model (Eq. 12a), Log-linear model (Eq. 12b) and Sigmoid model (Eq. 12d).

$$\text{Linear: } E(d) = \alpha d \quad (12a)$$

$$\text{Log-linear: } E(d) = \alpha \log(d + 1) \quad (12b)$$

$$\text{Emax: } E(d) = \frac{E_{max}d}{ED_{50} + d} \quad (12c)$$

$$\text{Sigmoid: } E(d) = \frac{E_{max}d^\gamma}{ED_{50}^\gamma + d^\gamma}, \text{ with } \gamma = 2 \text{ (fixed)} \quad (12d)$$

The vector of individual parameters θ_i is thus different in the four models: for Linear and Log-linear models, $\theta_i = (VA_{0i}, k_i, \beta_i, \alpha_i)$, while for Emax and Sigmoid models, $\theta_i = (VA_{0i}, k_i, \beta_i, E_{maxi}, ED_{50i})$.

The distributions followed by the individual parameters are given in Table 1. In addition, we assumed an additive error model: $\varepsilon_i \sim \mathcal{N}(0, \sigma_{inter}^2)$. For the simulation, we took $\sigma_{inter}^2 = 28 \text{ letters}^2$.

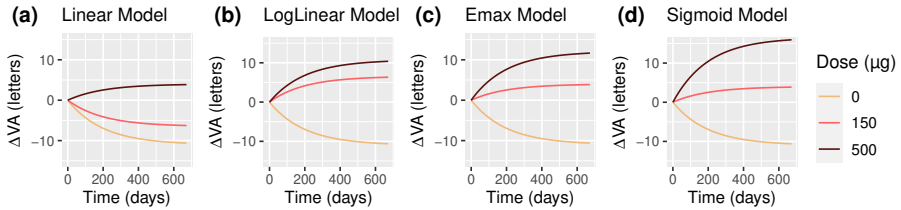
For all models, the population values of the dose-effect parameters are chosen to reach the same asymptotic VA for dose $d_{max} = 1000 \mu g$.

Table 1 Parameter values and distribution of the different candidate models

	VA_0 (letters)	k (day^{-1})	β	α	E_{max} (letters)	ED_{50} (letters)
Distribution	log-normal	log-normal	log-normal	normal	normal	log-normal
Linear	$\mu = 55$	$\mu = 0.005$	$\mu = 0.2$	$\mu = 0.03$ $\omega^2 = 0.0009$		
Log-linear				$\mu = 3.5$ $\omega^2 = 1.5$		
Emax Sigmoid	$\omega^2 = 0.07$	$\omega^2 = 0.5$	$\omega^2 = 1$			$\mu = 30$ $\omega^2 = 150$ $\mu = 150$ $\omega^2 = 0.49$

Note: μ refers to the fixed effects, and ω^2 to the variance of the random effects

Using the median parameters (i.e. the fixed effects μ), the change from baseline in visual acuity ΔVA ($\Delta VA(d, t, \mu) = VA(d, t, \mu) - VA(d, t = 0, \mu)$) between 0 and 672 days is represented on the Figure 2.

**Fig. 2** Evolution of the change from baseline in visual acuity (ΔVA) according to the fixed effect, for the candidate models, with 0, 150 and 500 μg doses

Scenarios

The design is based on the scenario described in [17]: the duration of the trial is 24 months, with 26 measures of visual acuity per individual (day 0, day 7 and every 28 days from the 28th to the 672nd day) and a total of $N = 300$ individuals.

Design optimization is performed on the tested doses and on the time points measurements. Each possible elementary design is composed of two fixed times, at day 0 and 672, and two different measurements optimized among the proposed times $\{7; 28k, k = 1, \dots, 23\}$. Each individual is allocated to a dose group, receiving at day 0 a dose among $\{0, 25, 50, 100, 150, 300, 500\}$ μg . Thus, there is $\binom{24}{2} \times 7 = 1932$ possible elementary designs. Designs are rounded with $n_{min} = 5$.

Both one-stage and two-stage designs are explored.

One-stage design

Different one-stage designs are compared, based on three approaches (see Figure 1, block "One-Stage") for the design optimization: locally optimized

designs for Linear model (i.e. the simplest model) and Emax model, respectively $\Xi_{LOD,Linear}$ and $\Xi_{LOD,Emax}$, and robust design (i.e. accounting for the uncertainty on the model) Ξ_{Robust} .

Then, different models are considered for the estimation step, in order to highlight the effect of the design *when the model is accurate or misspecified*. For the three initial designs, parameters are estimated using Emax model, MA approach and MS approach. For $\Xi_{LOD,Linear}$ an additional case is considered by estimating parameters using the Linear model. The best case is supposed to be $\Xi_{LOD,Emax}$, followed by a fit using Emax model, as it is the true model, meaning the one used to simulate the data.

For MA approach, fits are performed with the four candidate models, which are then weighted according to their AIC as given in Eq. 9. For MS approach, the model with *the* lowest AIC is selected.

Two-stage design

Two-stage design strategies are illustrated on Figure 1 and explained in *material and methods. Each stage contains half of the subjects*, i.e. $N_1 = N_2 = N/2 = 150$. This analysis aims to evaluate the performance of the adaptive procedure combined with Ξ_{Robust} . The different options come after a first stage similar to the scenarios explored in the previous section with one-stage designs. As previously, we consider $\Xi_{LOD,Linear}$, $\Xi_{LOD,Emax}$ and Ξ_{Robust} as three possible first stage *designs*. Then, the second stage is performed using either a local optimal design for Emax when the parameters are estimated with the Emax model, a robust design based on MA or a local optimal design for the model selected after the first stage's estimation step. For the $\Xi_{LOD,Linear}$ design at the first stage, we also consider the case in which the second stage is designed again using a local optimal design for the Linear model.

Implementation and simulation settings

All the design optimizations are performed using the multiplicative algorithm implemented in PFIM 5.0 (beta version) [22]. An extension of this package is implemented for the CD-criterion and for K -stages designs.

Estimations are performed using the Stochastic Approximation Expectation-Maximization (SAEM) algorithm implemented in saemix 3.0 [23]. Running default parameters are kept (300 iterations for the exploratory phase and 100 for the smoothing one). For the estimation step in one-stage setting and in the first stage in two-stage setting, the SAEM algorithm is initialized with the values reported in Table 1.

Data simulation is performed using R version 4.1.0 and the package mlxR version 4.2.0.

One-stage design settings

For each explored design, the Emax model (Eq. 12c) and the associated values presented in Table 1 are used to simulate $S = 100$ datasets with $N = 300$ subjects. Then, fits are performed on each dataset using the different strategies and comparison quantities are computed.

Two-stage design settings

First $S_1 = 100$ datasets $Y_{s_1}^{(1)}$ are simulated with $N_1 = 150$ and fitted with the different first stage strategies. In each scenario, for each dataset, results obtained are used to design the second stage. This means that the FIM for second stage design is computed using the estimated parameters on $Y_{s_1}^{(1)}$, the empirical FIM from the first stage and the expected FIM for second stage. In case of a robust design scenario, the AIC weights are updated according to the first stage results. Then, a dataset $Y_{s_1}^{(2)}$ is simulated following this second stage design. *Parameters are re-estimated*, using both $Y_{s_1}^{(1)}$ and $Y_{s_1}^{(2)}$ and comparison quantities are computed.

Performance Evaluation

The different design strategies are compared according to several criteria. As proposed by Buatois et al. [18], we use the Minimum Effective Dose (MED) which is the minimum dose at which a *clinically* relevant drug effect is achieved, *here defined as an increase of the median ΔVA at the end of treatment of at least 15 letters, to evaluate predictive performances. Derived from the true model and the population parameters given in Table 1, MED is equal to 250 μg .*

The individual visual acuity change from baseline at the end of the trial ($t = t_{EOT}$) is denoted ΔVA . Monte-Carlo (MC) simulations are used to estimate the distribution of the individual ΔVA at t_{EOT} for the different doses and models.

For a candidate model m , a given dose d and a simulation s , the estimated probability density evolution from baseline is:

$$p_{m,s}(d, \Delta VA) = p_{m,s}(f_m(d, t_{EOT}, \theta) - f_m(d, 0, \theta)), \quad \theta \sim p(\hat{\psi}_{m,s}) \quad (13)$$

In model averaging approach we use :

$$p_{CD,s} = \sum_{m=1}^M w_m p_{m,s}(d, \Delta VA) \quad (14)$$

As the drug is supposed to have a relevant effect if the median ΔVA has increased of at least 15 letters at the end of the treatment, the MED is defined as:

$$MED_{m,s} = \underset{d}{\operatorname{argmin}} \left[\operatorname{median} \left((p_{m,s}(d, \Delta VA)) \right) - \operatorname{median} \left((p_{m,s}(0, \Delta VA)) \right) \right] \geq 15 \quad (15)$$

The design strategies are then compared in terms of MED estimation, using the Relative Bias (*RB*) and the Relative Root Mean Square Error (*RRMSE*). Formula for these assessment quantities are given in Eq. 16, with x^* referring to the true value of the parameter and \hat{x}_s its estimation in the s -th simulation.

$$\text{Relative Estimation Error: } REE(\hat{x}_s) = \frac{\hat{x}_s - x^*}{x^*} \times 100 \quad (16a)$$

$$RB(\hat{x}_s) = \frac{1}{S} \sum_{s=1}^S REE(\hat{x}_s) \quad (16b)$$

$$RRMSE(\hat{x}_s) = \sqrt{\frac{1}{S} \sum_{s=1}^S REE(\hat{x}_s)^2} \quad (16c)$$

The impact of *the* design in the frequency of selection of candidate models in MS, and the resulting ability to select the true model are explored, as the weights distribution in MA.

Results

One-stage design

The robust design Ξ_{Robust} and the locally optimal designs (LOD) $\Xi_{LOD,m}$ for each model m for the one-stage designs are reported in Table 2. Overall, the designs are similar in terms of sampling times. The informative sampling times are located between 140 and 196 days, and close to the end of the study, between 588 and 644 days. An early measuring time, 7 days after the beginning of the treatment, is also optimal excepted for the Log-linear model. Between the different models, the most critical design elements are the doses. In case of Linear $\Xi_{LOD,Linear}$ or Log-linear $\Xi_{LOD,Log-linear}$ models, all the individuals are allocated to extreme doses (0 and 500 μg) whereas optimal designs for Emax $\Xi_{LOD,Emax}$ and Sigmoid models $\Xi_{LOD,Sigmoid}$ also include intermediate doses such as 100 or 150 μg , as expected. In a coherent manner, the robust design Ξ_{Robust} over these four candidate models involves individuals with an intermediate dose, in smaller numbers than for $\Xi_{LOD,Emax}$ and $\Xi_{LOD,Sigmoid}$ (57 vs. 92 and 83, respectively).

Table 2 Optimal one-stage designs, $N = 300$.

Design and arms	Dose (μg)	Sampling times (days)	Subjects
Ξ_{Robust}			
arm 1	0	(0, 168, 196, 672)	126
arm 2	100	(0, 7, 616, 672)	6
arm 3	100	(0, 7, 644, 672)	51
arm 4	500	(0, 168, 196, 672)	117
$\Xi_{LOD,Emax}$			
arm 1	0	(0, 168, 196, 672)	99
arm 2	100	(0, 7, 616, 672)	13
arm 3	100	(0, 7, 644, 672)	79
arm 4	500	(0, 168, 196, 672)	109
$\Xi_{LOD,Linear}$			
arm 1	0	(0, 168, 196, 672)	197
arm 2	500	(0, 7, 588, 672)	10
arm 3	500	(0, 7, 616, 672)	27
arm 4	500	(0, 7, 644, 672)	66
$\Xi_{LOD,Log-linear}$			
arm 1	0	(0, 140, 196, 672)	78
arm 2	0	(0, 168, 644, 672)	65
arm 3	500	(0, 168, 196, 672)	157
$\Xi_{LOD,Sigmoid}$			
arm 1	0	(0, 7, 644, 672)	9
arm 2	0	(0, 168, 644, 672)	72
arm 3	150	(0, 7, 644, 672)	83
arm 4	500	(0, 168, 196, 672)	136

Ξ_{Robust} is the robust design accounting for model uncertainty over the four candidate models, $\Xi_{LOD,m}$ is the local optimal design for the model m .

Table 3 Relative D-efficiencies of optimal one-stage designs under the different candidate models, $N = 300$.

Model \ Design	Ξ_{Robust}	$\Xi_{LOD,Emax}$	$\Xi_{LOD,Linear}$	$\Xi_{LOD,Log-linear}$	$\Xi_{LOD,Sigmoid}$
Emax	0.98	1	0	0.0023	0.95
Linear	0.92	0.91	1	0.96	0.86
Log-linear	0.96	0.94	0.97	1	0.92
Sigmoid	0.96	0.96	0.0017	0	1
Mean	0.96	0.95	0.49	0.49	0.93

Ξ_{Robust} is the robust design accounting for model uncertainty over the four candidate models, $\Xi_{LOD,m}$ is the local optimal design for the model m .

These LOD are compared based on their relative D-efficiencies reported in Table 3. For a given model, the corresponding optimal design is more efficient than the Ξ_{Robust} , but the latter is almost always better than the designs determined as the optimal one for another model. $\Xi_{LOD,Emax}$ and $\Xi_{LOD,Sigmoid}$ have overall good efficiencies while it can be noticed that $\Xi_{LOD,Linear}$ and $\Xi_{LOD,Log-linear}$ lead to efficiencies close to 0 when the model is Emax or Sigmoid. This is explained by the absence of intermediate doses in these designs, which makes impossible the estimation of ED_{50} . Using a robust design allows to get around this issue. In addition, the efficiency of the Ξ_{Robust} is the largest in average.

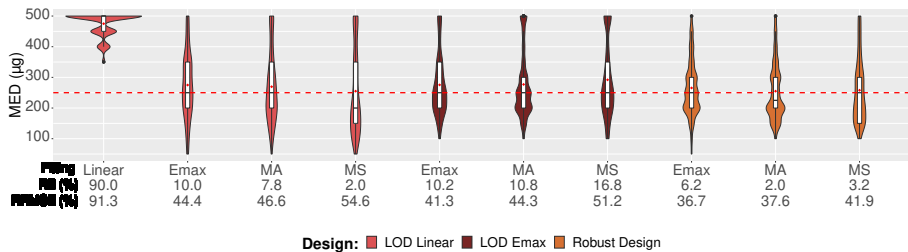


Fig. 3 Violins and boxplots for the distribution of the predicted Minimal Effective Dose (MED) for one-stage designs.

100 Simulations, $N = 300$. The upper whiskers extend from the hinge to the largest value below $1.5 \times$ the interquartile range, and the lower whiskers extend from the hinge to the smallest value above $1.5 \times$ the interquartile range. Red diamonds represent the means. The red dashed line is the true MED.

LOD: Local Optimal Design, MA: Model Averaging, MS: Model Selection, RB: Relative Bias, RRMSE: Relative Root Mean Square Error

From simulated datasets and for different fitting methods presented in *scenarios subsection*, the minimum effective dose (MED) is calculated and shown in Figure 3. As expected, $\Xi_{LOD,Linear}$ - Fit Linear is the worst performer ($RB = 90\%$ and $RRMSE = 91.3\%$). Nevertheless, fitting with the true model or using model averaging or, to a lesser extent, model selection gives acceptable results. As the Emax model is the true model used to simulate the data, $\Xi_{LOD,Emax}$ scenarios provide good performances. MS analysis leads to a *higher risk of misidentifying* the MED than MA. Similar or better performances are obtained with the Ξ_{Robust} (lowest RB , between 2.0% with MA to 6.2% when fitting with the Emax model, and lowest $RRMSE$, between 36.7% when fitting with the Emax model and 41.9% with MS). Importantly, choosing a local design for a misspecified model such as $\Xi_{LOD,Linear}$ gives less good results than choosing a robust design Ξ_{Robust} , whether fitting by MA ($RRMSE$ of 46.6% vs. 37.6%, respectively) or by MS ($RRMSE$ of 54.6% vs. 42.0%, respectively).

MS and MA weights distributions for the different designs are shown on Figure 4. After $\Xi_{LOD,Linear}$, Log-linear model is much more selected than the other ones: 78% of the cases while only respectively 12% for Emax, 7% for

Sigmoid and 3% for Linear. Consistently, the most weighted model in MA is Log-linear, with an average weight of 0.65, then comes Emax (0.17), then Sigmoid (0.15) and finally Linear (0.04). After $\Xi_{LOD,Emax}$, model selection is more balanced between Log-linear, Emax (both 38%) and Sigmoid (24%). Linear model is never selected. In the same way, MA weights are more balanced between Emax (0.37), Log-linear (0.34), and Sigmoid (0.29). After Ξ_{Robust} , Log-linear is once again the most selected (50%), while Emax and Sigmoid are almost equally selected (respectively, 26% and 24%). As for $\Xi_{LOD,Emax}$, Linear is never selected in MS approach, and the weight *of* this model in MA is null. Once again Log-linear model *is the most weighted* (0.44), then comes Emax (0.30) and Sigmoid (0.27). Thus, design impacts the ability to select the true model (Emax), or at least one close to the true one (Sigmoid), in MS and in the weights distribution of MA. In this framework, Ξ_{Robust} performs better than $\Xi_{LOD,Linear}$ and its results are even close to those obtained with $\Xi_{LOD,Emax}$ (true model). The prevalence of Log-linear model in MS and MA may be explained by the similar trends between Emax and Log-linear models (see Figure 2) coupled with the smaller number of parameters to be estimated in Log-linear model. Thus, because of the penalty term in AIC computation, with equal likelihood, Log-linear model is preferred.

Another important result from these simulations is that the MA method outperforms MS, especially in the case of local designs (*RRMSE* of 46.6% vs. 54.6% for $\Xi_{LOD,Linear}$ and *RRMSE* of 44.3% vs. 51.2% for $\Xi_{LOD,Emax}$). Indeed, this sparse design setting (only 4 measuring times per individual) leads to the selection of a model which is not the true one (Emax) in most of the cases, even with the $\Xi_{LOD,Emax}$ (Emax model selected in only 38% of the simulations).

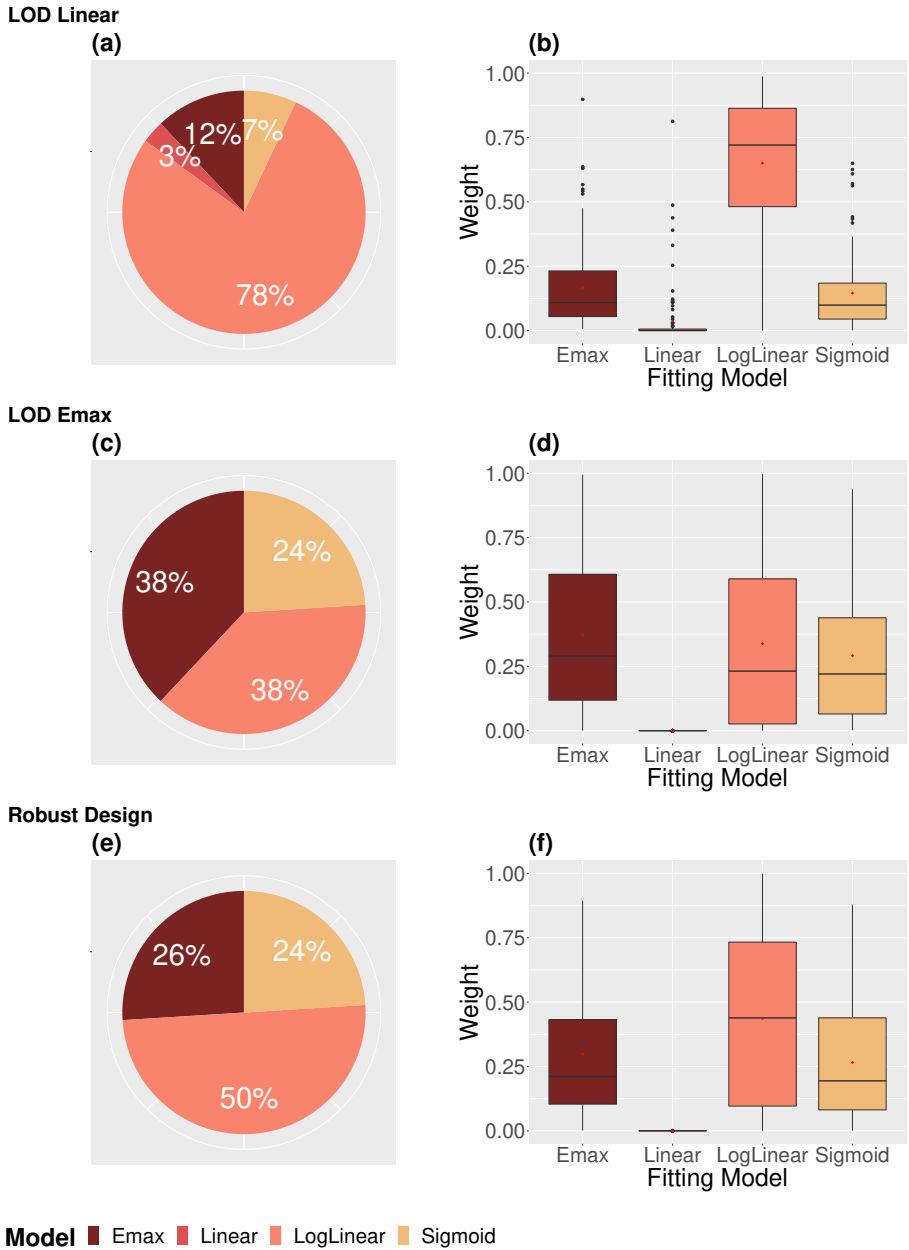
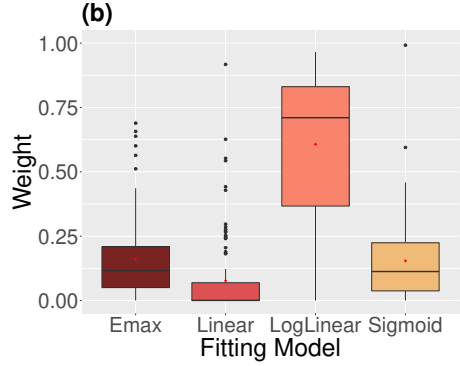
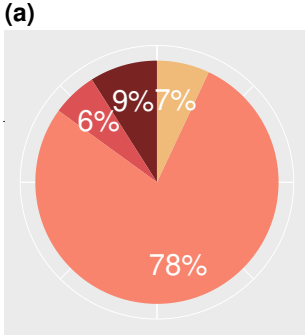


Fig. 4 Selected Proportions for MS (left) and AIC Weights distribution w (right) for MA in different One-Stage Designs: (a) and (b): LOD (Local Optimal Design) for the Linear model, (c) and (d): LOD for the Emax model, (e) and (f): Robust Design 100 Simulations, $N = 300$. The upper whiskers extend from the hinge to the largest value below $1.5 \times$ the interquartile range, and the lower whiskers extend from the hinge to the smallest value above $1.5 \times$ the interquartile range. Red diamonds represent the means.

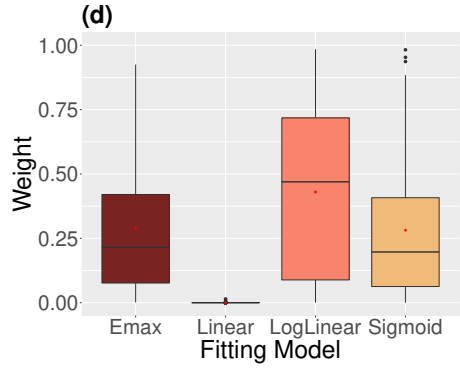
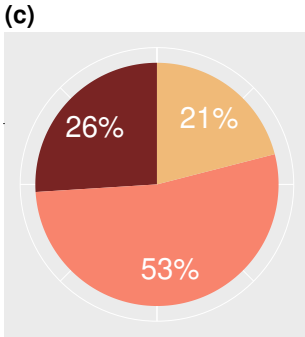
Two-stage design

Optimal designs obtained with the multiplicative algorithm for the first stage with $N_1 = N/2$ are presented in Appendix B. They are similar as the ones previously obtained for N , but with only approximately half as many patients in each arm. The small differences come from the constraint we set of having at least 5 patients per arm. This explains why the Ξ_{Robust} only has 3 arms for the first stage while it used to have 4 in the one-stage design. Therefore, the relative efficiencies have the same hierarchy *as the one found* with a single stage.

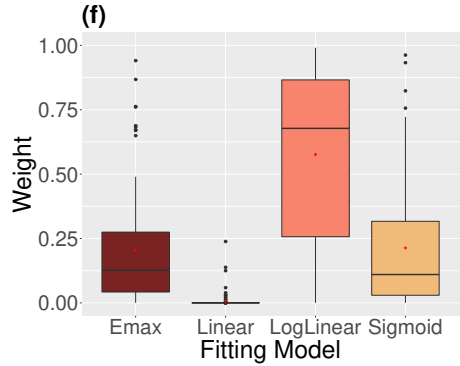
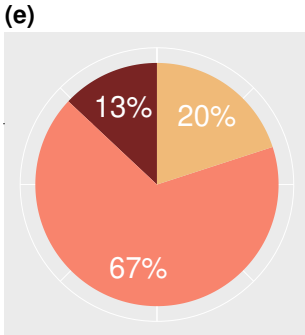
First Stage Design: LOD Linear, N=150



First Stage Design: LOD Emax, N=150



First Stage Design: Robust Design, N=150



Model ■ Emax ■ Linear ■ LogLinear ■ Sigmoid

Fig. 5 Selected Proportions for MS (left) and AIC weights distribution (right) after different First Stage Designs: (a) and (b): LOD (Local Optimal Design) for the Linear model, (c) and (d): LOD for the Emax model, (e) and (f): Robust Design
 100 Simulations, N = 150. The upper whiskers extend from the hinge to the largest value below $1.5 \times$ the interquartile range, and the lower whiskers extend from the hinge to the smallest value above $1.5 \times$ the interquartile range. Red diamonds represent the means.

The Figure 5 shows that *whatever the design used, the Log-linear model is always the most selected after the first stage in the MS approach* (78% after $\Xi_{LOD,Linear}$, 53% after $\Xi_{LOD,Emax}$ and 67% after Ξ_{Robust}). It means that the second stage will be conducted using its optimal design, according to the new estimated parameters. After $\Xi_{LOD,Emax}$ the second *most* selected model is Emax (26%), then comes Sigmoid (21%), while Linear is never selected. After Ξ_{Robust} , Emax and Sigmoid are respectively selected in 13% and 20% of the cases. After $\Xi_{LOD,Linear}$, Emax, Sigmoid and Linear are selected almost equally (respectively, 9%, 7% and 6%).

Overall these results are not as good as those obtained with the one-stage designs as the Emax model is less selected. Especially it is selected in only 13% of the simulations for Ξ_{Robust} while it used to be 26%. This may be explained by the smaller amount of subjects, and thus of data, leading to select a simpler model. Indeed, the Log-linear model has one less parameter than the Emax and Sigmoid models.

Consistently, the Log-linear model is also always the most weighted one (0.61 after $\Xi_{LOD,Linear}$, 0.43 after $\Xi_{LOD,Emax}$ and 0.53 after Ξ_{Robust}), according to AIC, in the MA approaches (see Figure 5). The most balanced weights are obtained after $\Xi_{LOD,Emax}$ with 0.29 for Emax, 0.43 for Log-linear and 0.28 for Sigmoid. Interestingly, the robust design leads to a weight of 0 for the Linear model, which is the *most* distant model from the true one, in more than 90% of the simulations.

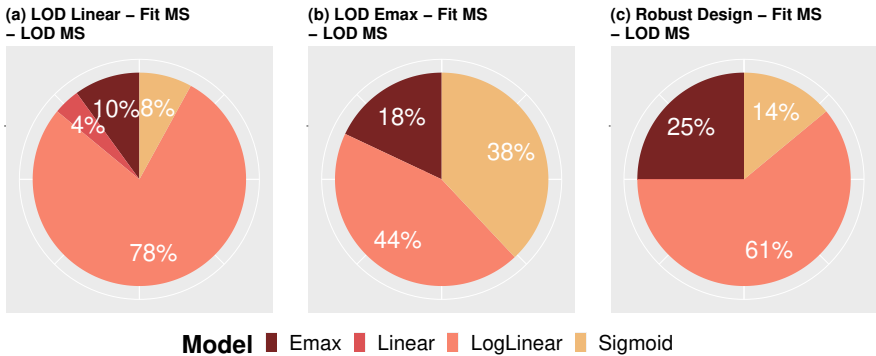


Fig. 6 Selected Proportions for Model Selection (MS) after different first stage and a second stage based on the model selected after the first stage
100 Simulations, $N_1 = N_2 = 150$, LOD: Local Optimal Design

The Figure 6 shows the proportion of selection in the MS approach for the second stage depending of the design used in first stage, the selected model being the one used to compute the MED. Once again we can see that the

Log-linear model is the most selected, but having more data (completing the second stage) *leads to select slightly more frequently the right model* compared to after the first stage, even if the design is often optimized for a wrong model. For example, starting with Ξ_{Robust} , the Emax model is selected in 13% of *the* simulations after the first stage vs. 25% after the second stage.

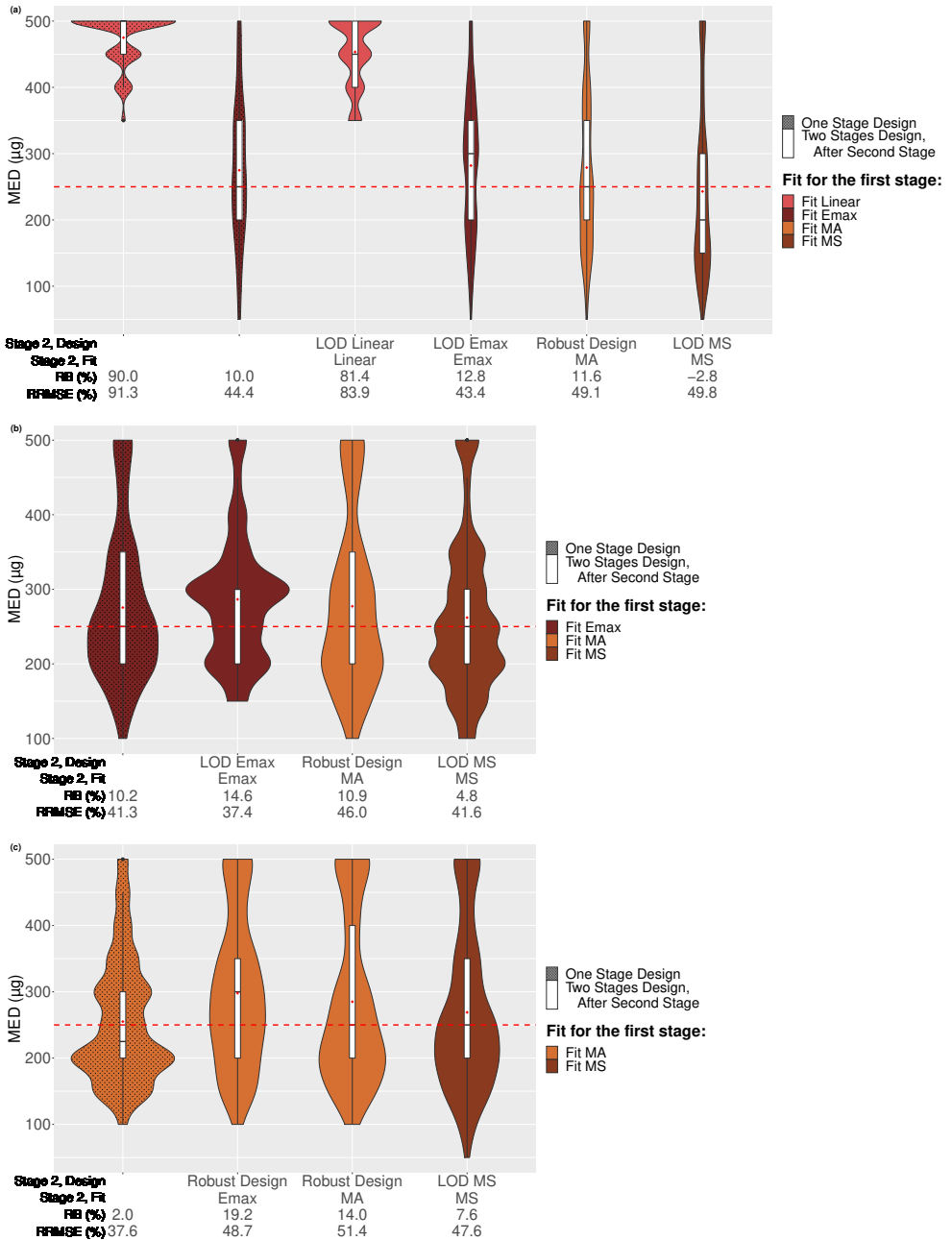


Fig. 7 Violins and boxplots for the distribution of the predicted Minimal Effective Dose (MED) in μg for different approaches: (a) after $\Xi_{LOD,Linear}$, (b) after $\Xi_{LOD,Emax}$, (c) after Ξ_{Robust} first stage

100 Simulations, $N = 300$. The upper whiskers extend from the hinge to the largest value below $1.5 \times$ the interquartile range, and the lower whiskers extend from the hinge to the smallest value above $1.5 \times$ the interquartile range. Red diamonds represent the means. The red line is the true MED.

Ξ_{Robust} is the robust design accounting for model uncertainty over the four candidate models, $\Xi_{LOD,m}$ is the local optimal design for the model m .

LOD: Local Optimal Design, MA: Model Averaging, MS: Model Selection, RB: Relative Bias, RRMSE: Relative Root Mean Square Error

From simulated datasets and for different two-stage design approaches presented *in scenarios subsection*, the minimum effective dose (MED) is calculated and shown in Figure 7.

Approaches derived from a $\Xi_{LOD,Linear}$ first stage design are presented on panel (a). As seen previously, this design followed by a fit using Linear model performs poorly. The same design but with a fit using the true model leads to a reasonable result ($RB = 10\%$ and $RRMSE = 44.4\%$). A two-stage design, even with $\Xi_{LOD,Linear}$ - Fit Linear approach is slightly better than just a one-stage design in linear design, both RB and $RRMSE$ are smaller (respectively $RB = 81.4\%$ and $RRMSE = 83.9\%$ for two-stage design, and $RB = 90.0\%$ and $RRMSE = 91.3\%$ for the one-stage design), *but this approach is still performing poorly*. The two-stage design Fit Emax - $\Xi_{LOD,Emax}$ - Fit Emax does not perform better than the one-stage one fitting with Emax (respectively, $RB = 12.8\%$, $RRMSE = 43.4\%$ and $RB = 10\%$, $RRMSE = 44.4\%$). This may be explained by error propagation: the SAEM algorithm is initialized at the beginning of the first stage with the true values when fitting with Emax, while at the second stage, estimates from the first stage are used as starting points. This may cause an overly optimistic evaluation when fitting with Emax model in one-stage design. Using MA at the first stage, then using Ξ_{Robust} for the second stage and MA again performs reasonably, RB is of 11.6% and $RRMSE$ is 49.1% . Selecting a model after a first stage to design the second stage (Fit MS - $\Xi_{LOD,MS}$ - Fit MS) gives here the best results in terms of RB ($RB = -2.8\%$). In addition, the $RRMSE$ is smaller than the one obtained with one-stage design (see Figure 3) $\Xi_{LOD,Linear}$ followed by MS (49.8% vs. 54.6%) while RB are close (-2.8% vs. 2%). These results show that even if the model is highly misspecified when designing the first stage, using MA - Ξ_{Robust} or MS - $\Xi_{LOD,MS}$ in a second stage allows to reach good results.

As expected, starting with $\Xi_{LOD,Emax}$, i.e. the design of the first stage under the true model, leads to good performances (Figure 7, panel (b)). The expected best scenario was the two-stage with Fit Emax - $\Xi_{LOD,Emax}$ - Fit Emax. It has the lowest $RRMSE$ (37.4%) but it *is more biased* ($RB = 14.6$) than other approaches. This result is consistent with one-stage design approaches, in which $\Xi_{LOD,Emax}$ led to more bias than Ξ_{Robust} . The robust two-stage approach using MA for the two fits performs equivalently to the one-stage design followed by a fit using Emax model (respectively, $RB = 10.9\%$, $RRMSE = 46.0\%$ and $RB = 10.2\%$, $RRMSE = 41.3\%$). On this example the two-stage design Fit MS - $\Xi_{LOD,MS}$ - Fit MS has again the best performances ($RB = 4.8\%$ and $RRMSE = 41.6\%$).

Overall, starting with Ξ_{Robust} leads to good performances (Figure 7, panel (c)). The Fit MA - Ξ_{Robust} - Fit Emax is more biased than *the one ending by MA*, despite Emax being the true model (respectively, $RB = 19.2\%$, $RRMSE = 48.7\%$ vs. $RB = 14\%$, $RRMSE = 51.4\%$). Here, the Robust Adaptive Design using MS performs slightly better than the one using MA (respectively, $RB = 7.6\%$, $RRMSE = 47.6\%$ vs. $RB = 14\%$, $RRMSE = 51.4\%$).

Overall, MS at both stages is a decent strategy. One can prefer to start with a robust design at the first stage rather than a *possibly* wrong local design. For example, although the metrics indicate similar bias and precision in $\Xi_{LOD,Linear}$ - Fit MS - $\Xi_{LOD,MS}$ - Fit MS and Ξ_{Robust} - Fit MS - $\Xi_{LOD,MS}$ - Fit MS, the median MED corresponds to the true MED in the second strategy which is not the case when the first stage is $\Xi_{LOD,Linear}$.

Importantly, in our example, one-stage robust design coupled with MA ($RB = 2\%$ and $RRMSE = 37.6\%$) is performing better than any two-stage strategy starting with a robust design at first stage.

Discussion

In this work, we propose a new two-stage adaptive design strategy in longitudinal studies, based on the robust expected FIM and MA over several candidate models. The concepts and methodology of this approach are explained in *material and methods*. Then, this strategy is applied to a clinical dose-response design of a treatment in ophthalmology, with several candidate models. This application includes simulations to evaluate the relevance of *the* approach and to compare it with other one- or two-stage designs and analysis (e.g. MS or MA) approaches.

Within one-stage, our work confirms even more conclusively than in previous works [17], that a robust design should be chosen, and local designs that lead to poor D-efficiencies with more complex models should be avoided. By simulation, we also demonstrated that local design optimization can have severe impacts at estimation step. It increases the risk of selecting a model distant from the true model, or, when performing MA, of obtaining higher weights in the most distant models (in our example, the linear model). Thus, it leads to higher bias than robust design and imprecision on parameters and therefore on outcomes such as MED in our example, which would lead to a suboptimal dose selection. Among two-stage designs, performing MS or MA at the end of the first stage to design the second stage is as expected better than focusing on the same wrong model at both stages. These two methods lead to reasonable bias and precision of estimation. If one *focuses* on local designs and model selection methods, performing a two-stage design, with a model selection at *the choice of a wrong model when designing the first stage*. The most robust procedure (robust design and MA at both stages) did not show a noteworthy interest in our example. On the other hand it seems valuable to start from a robust design at the first stage in order to collect information that could satisfy the different candidate models *and* then to perform MS after first stage to optimize the design of the second stage. Notice that this result *could be a consequence of the low number* of measurements per individual in our example, *which* may lead to high weights on simpler (and in our case, wrong) models.

A limit of our work is that the RB on MED estimation could be around 10%, even in an ideal scenario. This can also be explained *by our relatively sparse design* compared to the number of parameters to be estimated and to

the work of Buatois on the same models (4 vs. 26 measurement times, respectively) [18]. Here we decided to focus on MED, its RB and RRMSE but other metrics such as the clinically relevant drug effect (CRE) can also be used [18]. *Moreover, it would be of interest to consider more dose/response relationships (e.g. quadratic, exponential or logistic) in the set of candidate models [24, 25].* Overall, it would be of interest to apply the proposed approaches to other examples with different number of measurements, different candidate models or another type of study to investigate which design and/or which analysis method is the most suitable depending on the situation.

In our work, we searched local D-optimal *designs* or robust CD-optimal *designs*, as we aimed to optimize *the precision of the estimates* of all the model parameters. Other options can be of interest, when focusing on a subset of parameters, like maximizing c- or D_S - criteria, or even better to combine those to D-optimality to avoid sacrificing other parameters, i.e. cD- [26] or DD_S - criteria, respectively (CcD- or CDD_S - in case of robust design [17]). In other studies, it could be a good option to optimize design accounting for parameter uncertainty [15], which can also be done on top of accounting for model uncertainty [16]. In our article, we used the multiplicative algorithm implemented in PFIM5 R package [22], to optimize designs in a discrete space. Another perspective could be to optimize robust adaptive design with algorithms working *with* a continuous design space, such as particle swarm optimization [27] or PGBO [28], which are also implemented in PFIM5.

Conclusion

In this work, we propose a new approach to conduct robust and adaptive designs in longitudinal data analyses such as pharmacometrics studies. This approach *has* been applied to a dose-response study example and shows the relevance of adaptive two-stages design and of robust design accounting for model uncertainty. A perspective is to investigate the relevance of this approach in other examples, which would be eased by the PFIM R package.

Funding. The authors declare no specific sponsorship or funding related to this project.

Author contribution. Conceptualization: FM, JS Analysis: LF, RL Supervision: FM, JS Writing—original draft: LF, JS Writing—review and editing: LF, FM, JS

Acknowledgments. The authors thank Lionel de la Tribouille for the use of the CATIBioMed calculus facility.

Declarations

Conflict of interest. The authors declare no competing interests.

Appendix A Multiplicative Algorithm for Robust Approach

The multiplicative algorithm is an iterative algorithm which can be used to compute the weights α_q to attribute to the individual designs in order to optimize the D-criterion or the CD-criterion in case of Ξ_{Robust} . [29]
The derivative of these criteria in case of one-stage and K -stage designs, which are required for the optimization process, are computed in the following.

CD-derivative for one-stage design

In the multiplicative algorithm, an expression for $\frac{\partial \phi_D(\Xi)}{\partial \alpha_q}$ is needed. Therefore let denote:

$$\phi_D(\Xi) = Det(A(\alpha_q))^{1/P} \quad \text{with} \quad \mathcal{M}_F(\Xi) = A(\alpha_q) = N \sum_{q'=1}^Q \alpha_{q'} \mathcal{M}_F(\xi_{q'}) \quad (\text{A1})$$

We have,

$$\frac{\partial \phi_D(\Xi)}{\partial \alpha_q} = \frac{1}{P} \left(\frac{d}{d\alpha_q} Det(A(\alpha_q)) \right) Det(A(\alpha_q))^{1/P-1} \quad (\text{A2})$$

According to Jacobi's formula,

$$\frac{d}{d\alpha_q} Det(A(\alpha_q)) = tr \left[A(\alpha_q)^{-1} \frac{dA(\alpha_q)}{d\alpha_q} \right] Det(A(\alpha_q)) \quad (\text{A3})$$

and we have: $\frac{dA(\alpha_q)}{d\alpha_q} = N \mathcal{M}_F(\xi_q)$, which leads to:

$$\frac{d}{d\alpha_q} Det(A(\alpha_q)) = N tr [\mathcal{M}_F(\Xi)^{-1} \mathcal{M}_F(\xi_q)] Det(\mathcal{M}_F(\Xi)) \quad (\text{A4})$$

Thus,

$$\begin{aligned} \frac{\partial \phi_D(\Xi)}{\partial \alpha_q} &= \frac{N}{P} \left(\frac{d}{d\alpha_q} Det(A(\alpha_q)) \right) Det(A(\alpha_q))^{1/P-1} \\ &= \frac{N}{P} tr [\mathcal{M}_F(\Xi)^{-1} \mathcal{M}_F(\xi_q)] Det(\mathcal{M}_F(\Xi))^{1/P} \end{aligned} \quad (\text{A5})$$

For the robust approach, we need expressions for derivatives of $\phi_{CD}(\Xi) = \prod_{m=1}^M (\phi_{D,m}(\Xi))^{w_m}$.

Recall that

$$\frac{d}{dx} \prod_{i=1}^n f_i(x) = \sum_{i=1}^n \left(\frac{d}{dx} f_i(x) \prod_{j \neq i} f_j(x) \right) \quad (\text{A6})$$

Thus,

$$\frac{\partial \phi_{CD}(\Xi)}{\partial \alpha_q} = \sum_{m=1}^M \left(w_m \frac{\partial \phi_{D,m}(\Xi)}{\partial \alpha_q} (\phi_{D,m}(\Xi))^{w_m-1} \prod_{m' \neq m} (\phi_{D,m'}(\Xi))^{w_{m'}} \right) \quad (\text{A7})$$

which writes:

$$\frac{\partial \phi_{CD}(\Xi)}{\partial \alpha_q} = \phi_{CD}(\Xi) \sum_{m=1}^M \left(w_m \frac{1}{\phi_{D,m}(\Xi)} \frac{\partial \phi_{D,m}(\Xi)}{\partial \alpha_q} \right) \quad (\text{A8})$$

The robust version was implemented using PFIM and the new MultiplicativeAlgorithmRobust class.

CD-derivative for K -stage design

In the general setting of robust and adaptive design with K stages, the first stage of the study is designed and conducted according to previous information and accounting for model uncertainty. Then each following stage is planned leveraging the data obtained in the previous stages. Indeed the FIM at stage k depends on the design of the k^{th} stage, but also on those of all previous stages, from 1 to $k-1$. Thus, for a given model m (index m is omitted in the following equations for the sake of simplicity), the total FIM $\mathcal{M}_F^{(k)}(\hat{\psi}^{(k-1)}, \Xi)$ at stage k , for a design Ξ and given estimated parameters $\hat{\psi}^{(k-1)}$ from stage $k-1$, writes:

$$\begin{aligned} \mathcal{M}_F^{(k)}(\hat{\psi}^{(k-1)}, \Xi) &= \mathcal{M}_F(\hat{\psi}^{(k-1)}, \Xi_{CD,1}) + \mathcal{M}_F(\hat{\psi}^{(k-1)}, \Xi_{CD,2}) \\ &\quad + \dots + \mathcal{M}_F(\hat{\psi}^{(k-1)}, \Xi_{CD,k-1}) \\ &\quad + \mathcal{M}_F(\hat{\psi}^{(k-1)}, \Xi) \end{aligned} \quad (\text{A9})$$

where $\Xi_{CD,j}$ is the Ξ_{Robust} computed at stage j .

Thus, only $\mathcal{M}_F(\hat{\psi}^{(k-1)}, \Xi)$ depends on Ξ , the design on which optimization is performed. Therefore, at stage k , let denote the previous FIM:

$$\begin{aligned} \mathcal{M}_F^{(k),prev}(\hat{\psi}^{(k-1)}) &= \mathcal{M}_F(\hat{\psi}^{(k-1)}, \Xi_{CD,1}) + \mathcal{M}_F(\hat{\psi}^{(k-1)}, \Xi_{CD,2}) \\ &\quad + \dots + \mathcal{M}_F(\hat{\psi}^{(k-1)}, \Xi_{CD,k-1}) \end{aligned} \quad (\text{A10})$$

which leads to write:

$$\mathcal{M}_F^{(k)}(\hat{\psi}^{(k-1)}, \Xi) = \mathcal{M}_F^{(k),prev}(\hat{\psi}^{(k-1)}) + \mathcal{M}_F(\hat{\psi}^{(k-1)}, \Xi)$$

$$= \mathcal{M}_F^{(k),prev}(\hat{\psi}^{(k-1)}) + N_k \sum_{q=1}^Q \alpha_q \mathcal{M}_F(\hat{\psi}^{(k-1)}, \xi_q) \quad (\text{A11})$$

As previously and with $A(\alpha_q) = \mathcal{M}_F^{(k)}(\hat{\psi}^{(k-1)}, \Xi)$, we have:

$$\frac{\partial \phi_D(\Xi)}{\partial \alpha_q} = \frac{1}{P} \left(\frac{d}{d\alpha_q} \text{Det}(A(\alpha_q)) \right) \text{Det}(A(\alpha_q))^{1/P-1} \quad (\text{A12})$$

which writes:

$$\frac{\partial \phi_D(\Xi)}{\partial \alpha_q} = \frac{N_k}{P} \text{tr} \left[\mathcal{M}_F^{(k)}(\hat{\psi}^{(k-1)}, \Xi)^{-1} \mathcal{M}_F(\xi_q) \right] \text{Det}(\mathcal{M}_F^{(k)}(\hat{\psi}^{(k-1)}, \Xi))^{1/P} \quad (\text{A13})$$

Thus, in robust approach, the derivative of ϕ_{CD} keeps the same expression:

$$\frac{\partial \phi_{CD}(\Xi)}{\partial \alpha_q} = \phi_{CD}(\Xi) \sum_{m=1}^M \left(w_m \frac{1}{\phi_{D,m}(\Xi)} \frac{\partial \phi_{D,m}(\Xi)}{\partial \alpha_q} \right) \quad (\text{A14})$$

Appendix B Two-stage design: designs for the first stage and relative efficiency

Table B1 Optimal First Stage Designs, $N_1 = 150$.

Design and arms	Dose (μg)	Sampling times (days)	Subjects
Ξ_{Robust}			
arm 1	0	(0, 168, 196, 672)	65
arm 2	100	(0, 7, 644, 672)	26
arm 3	500	(0, 168, 196, 672)	59
$\Xi_{LOD,Emax}$			
arm 1	0	(0, 168, 196, 672)	49
arm 2	100	(0, 7, 616, 672)	7
arm 3	100	(0, 7, 644, 672)	39
arm 4	500	(0, 168, 196, 672)	55
$\Xi_{LOD,Linear}$			
arm 1	0	(0, 168, 196, 672)	102
arm 2	500	(0, 7, 616, 672)	14
arm 3	500	(0, 7, 644, 672)	34
$\Xi_{LOD,Log-linear}$			
arm 1	0	(0, 168, 196, 672)	39
arm 2	0	(0, 168, 644, 672)	32
arm 3	500	(0, 168, 196, 672)	79
$\Xi_{LOD,Sigmoid}$			
arm 1	0	(0, 168, 644, 672)	37
arm 2	150	(0, 7, 644, 672)	43
arm 3	500	(0, 168, 196, 672)	70

Ξ_{Robust} is the robust design accounting for model uncertainty over the four candidate models, $\Xi_{LOD,m}$ is the local optimal design for the model m .

Table B2 Relative D-Efficiencies between designs, $N_1 = 150$.

Design Model	Ξ_{Robust}	$\Xi_{LOD,Emax}$	$\Xi_{LOD,Linear}$	$\Xi_{LOD,Log-linear}$	$\Xi_{LOD,Sigmoid}$
Emax	0.98	1	0	0	0.95
Linear	0.92	0.86	1	0.92	0.75
Log-linear	0.95	0.89	0.91	1	0.89
Sigmoid	0.95	0.97	3.0e-05	0	1
Mean	0.95	0.93	0.48	0.48	0.90

Ξ_{Robust} is the robust design accounting for model uncertainty over the four candidate models, $\Xi_{LOD,m}$ is the local optimal design for the model m .

References

- [1] Mould DR, Upton RN. Basic concepts in population modeling, simulation, and model-based drug development. *CPT: Pharmacometrics & Systems Pharmacology*. 2012;1:e6.
- [2] Mould DR, Upton RN. Basic concepts in population modeling, simulation, and model-based drug development - Part 2: Introduction to Pharmacokinetic Modeling Methods. *CPT: Pharmacometrics & Systems Pharmacology*. 2013;2:e38.
- [3] Mould DR, Upton RN. Basic concepts in population modeling, simulation, and model-based drug development - Part 3: Introduction to Pharmacodynamic Modeling Methods. *CPT: Pharmacometrics & Systems Pharmacology*. 2014;3:e88.
- [4] Jonsson EN, Wade JR, Karlsson MO. Comparison of some practical sampling strategies for population pharmacokinetic studies. *Journal of Pharmacokinetics and Biopharmaceutics*. 1996;24(2):245–263.
- [5] Ogungbenro K, Gueorguieva I, Majid O, Graham G, Aarons L. Optimal design for multiresponse pharmacokinetic-pharmacodynamic models—dealing with unbalanced designs. *Journal of Pharmacokinetics and Pharmacodynamics*. 2007;34(3):313–331.
- [6] Sheiner LB, Hashimoto Y, Beal SL. A simulation study comparing designs for dose ranging. *Statistics in Medicine*. 1991;10(3):303–321.
- [7] Mentré F, Mallet A, Baccar D. Optimal design in random effect regression models. *Biometrika*. 1997;84:429–442.
- [8] Thorlund K, Haggstrom J, Park JJ, Mills EJ. Key design considerations for adaptive clinical trials: a primer for clinicians. *British Medical Journal*. 2018;360:k698.
- [9] Pallmann P, Bedding AW, Choodari-Oskooei B, Dimairo M, Flight L, Hampson LV, et al. Adaptive designs in clinical trials: why use them, and how to run and report them. *BMC Medicine*. 2018;16(1):1–15.
- [10] Dumont C, Chenel M, Mentré F. Two-stage adaptive designs in nonlinear mixed effects models: application to pharmacokinetics in children. *Communications in Statistics-Simulation and Computation*. 2016;45(5):1511–1525.
- [11] Pierrillas PB, Fouliard S, Chenel M, Hooker AC, Friberg LF, Karlsson MO. Model-based adaptive optimal design (MBAOD) improves combination dose finding designs: an example in oncology. *The AAPS Journal*.

- 2018;20(2):1–11.
- [12] Lestini G, Dumont C, Mentré F. 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*. 2015;32(10):3159–3169.
 - [13] Chen TT. Optimal three-stage designs for phase II cancer clinical trials. *Statistics in Medicine*. 1997;16(23):2701–2711.
 - [14] Jahn-Eimermacher A, Hommel G. Performance of adaptive sample size adjustment with respect to stopping criteria and time of interim analysis. *Statistics in Medicine*. 2007;26(7):1450–1461.
 - [15] Foo LK, McGree J, Eccleston J, Duffull S. Comparison of robust criteria for D-optimal designs. *Journal of Biopharmaceutical Statistics*. 2012;22(6):1193–1205.
 - [16] Loingeville F, Nguyen TT, Riviere MK, Mentré F. Robust designs in longitudinal studies accounting for parameter and model uncertainties—application to count data. *Journal of Biopharmaceutical Statistics*. 2020;30(1):31–45.
 - [17] Seurat J, Nguyen TT, Mentré F. Robust designs accounting for model uncertainty in longitudinal studies with binary outcomes. *Statistical Methods in Medical Research*. 2020;29(3):934–952.
 - [18] Buatois S, Ueckert S, Frey N, Retout S, Mentré F. Comparison of model averaging and model selection in dose finding trials analyzed by nonlinear mixed effect models. *The AAPS Journal*. 2018;20(3):1–9.
 - [19] Aoki Y, Röshammar D, Hamrén B, Hooker AC. Model selection and averaging of nonlinear mixed-effect models for robust phase III dose selection. *Journal of Pharmacokinetics and Pharmacodynamics*. 2017;44(6):581–597.
 - [20] Atkinson A, Donev A, Tobias R. Optimum experimental designs, with SAS. vol. 34. OUP Oxford; 2007.
 - [21] Pukelsheim F, Rieder S. Efficient rounding of approximate designs. *Biometrika*. 1992;79(4):763–770.
 - [22] Leroux R, Seurat J, Le Nagard H, Mentré F, on behalf of the PFIM group. Design evaluation and optimisation in nonlinear mixed effects models with the R package PFIM. *PAGE* 30. 2022;Abstr 10183.

- [23] Comets E, Lavenu A, Lavielle M. Parameter estimation in nonlinear mixed effect models using saemix, an R implementation of the SAEM algorithm. *Journal of Statistical Software*. 2017;80:1–41.
- [24] Bretz F, Pinheiro JC, Branson M. Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics*. 2005;61(3):738–748.
- [25] Pinheiro J, Bornkamp B, Glimm E, Bretz F. Model-based dose finding under model uncertainty using general parametric models. *Statistics in medicine*. 2014;33(10):1646–1661.
- [26] Holland-Letz T. On the combination of c-and D-optimal designs: General approaches and applications in dose-response studies. *Biometrics*. 2017;73(1):206–213.
- [27] Wong WK, Chen RB, Huang CC, Wang W. A modified particle swarm optimization technique for finding optimal designs for mixture models. *PLoS One*. 2015;10(6):e0124720.
- [28] Le Nagard H, Chao L, Tenaillon O. The emergence of complexity and restricted pleiotropy in adapting networks. *BMC Evolutionary Biology*. 2011;11(1):1–15.
- [29] Seurat J, Tang Y, Mentré F, Nguyen TT. Finding optimal design in nonlinear mixed effect models using multiplicative algorithms. *Computer Methods and Programs in Biomedicine*. 2021;207:106–126.