

Prediction of Shrinkage of Individual Parameters Using the Bayesian Information Matrix in Non-Linear Mixed Effect Models with Evaluation in Pharmacokinetics

François Pierre Combes • Sylvie Retout • Nicolas Frey • France Mentré

Received: 5 October 2012 / Accepted: 12 May 2013 / Published online: 7 June 2013
© Springer Science+Business Media New York 2013

ABSTRACT

Purpose When information is sparse, individual parameters derived from a non-linear mixed effects model analysis can shrink to the mean. The objective of this work was to predict individual parameter shrinkage from the Bayesian information matrix (M_{BF}). We 1) Propose and evaluate an approximation of M_{BF} by First-Order linearization (FO), 2) Explore by simulations the relationship between shrinkage and precision of estimates and 3) Evaluate prediction of shrinkage and individual parameter precision.

Methods We approximated M_{BF} using FO. From the shrinkage formula in linear mixed effects models, we derived the predicted shrinkage from M_{BF} . Shrinkage values were generated for parameters of two pharmacokinetic models by varying the structure and the magnitude of the random effect and residual error models as well as the design. We then evaluated the approximation of M_{BF} FO and compared it to Monte-Carlo (MC) simulations. We finally compared expected and observed shrinkage as well as the predicted and estimated Standard Errors (SE) of individual parameters.

Results M_{BF} FO was similar to M_{BF} MC. Predicted and observed shrinkages were close. Predicted and estimated SE were similar.

Conclusions M_{BF} FO enables prediction of shrinkage and SE of individual parameters. It can be used for design optimization.

KEY WORDS bayesian fisher information matrix • non-linear mixed effects model • optimal design • pharmacokinetics

ABBREVIATIONS

CL	Clearance
CV	Coefficient of Variation
FO	First-Order linearization
IU	International Units
K_m	Michaelis-Menten constant
MAP	Maximum A Posteriori
M_{BF}	Bayesian Fisher Information Matrix
MC	Monte-Carlo
M_F	Fisher Information Matrix
M_{IBF}	Individual Bayesian Fisher Information Matrix
M_{IF}	Individual Fisher Information Matrix
ML	Maximum Likelihood
MONOLIX	MOdèles NON Linéaires à effets miXtes
NLMEM	Non-Linear Mixed Effects Models
PK	Pharmacokinetics
Q	Intercompartmental clearance
RSE	Relative Standard Errors
SE	Standard Errors
Sh	Shrinkage
V	Volume of distribution

Electronic supplementary material The online version of this article (doi:10.1007/s11095-013-1079-3) contains supplementary material, which is available to authorized users.

F. P. Combes (✉) • F. Mentré
University Paris Diderot, Sorbonne Paris Cité
INSERM, UMR 738 I6, rue Henri Huchard
F-75018 Paris, France
e-mail: francois.combes@inserm.fr

F. P. Combes • S. Retout • N. Frey
Pharma Research and Early Development, Clinical Pharmacology
F. Hoffmann-La Roche Ltd, Basel, Switzerland

F. P. Combes • S. Retout
Institut Roche de Recherche et Médecine Translationnelle
Boulogne-Billancourt, France

V_1 - V_2	Volume of distribution of the central and peripheral compartment
V_m	Maximum elimination rate

INTRODUCTION

Pharmacokinetic (PK) data are usually analyzed using Non-Linear Mixed Effect Models (NLMEM) originally proposed by Sheiner *et al.* in the 70's (1). NLMEM describe the PK behavior of a drug in a population of subjects through estimation of PK population parameters and their between-subject variability using Maximum Likelihood (ML) approach in software such as NONMEM (2) and MONOLIX (3). Once population parameters are estimated, individual parameters can then be derived by a Bayesian approach as, for instance, the Maximum *A Posteriori* (MAP), by incorporating the information obtained from the population in the individual estimation. High precision in individual estimates is mandatory. Indeed, these parameters are extensively used in drug development to investigate any influence of the PK on the pharmacodynamic responses, using the sequential PKPD approach. Individual PK parameters can also be a basis for investigating any influence of physiological and demographic covariates on drug concentration and for determining whether that influence implies a dose adjustment in a subset of the population (*eg.* impact of cytochrome P450 genotypes on clearance), or in clinical practice to individually adapt the dose and dosing regimen *via* early individual parameter estimation and predictions (*eg.* therapeutic drug monitoring).

The informativeness of a dataset in parameter estimation is a function of the number of subjects and of the number and timing of the samples (4). Population analysis allows the use of sparse sampling in large clinical trials, where few samples are collected in each subject. Each sampling time must be chosen carefully to be informative. To evaluate and optimize a population design, two approaches have been proposed. The first approach is based on clinical trial simulation and estimation, leading to a very time-consuming approach that limits the evaluation to a few designs and makes optimization difficult (5). As an alternative, methods based on a mathematical derivation of the Fisher information Matrix (M_F) have been developed (6). The inverse of the M_F is used to derive the expected precision of estimation of the parameters for a given design, M_F being the lower bound of the variance-covariance matrix of any unbiased estimator of the parameters (7). Design optimization relies on the use of a criterion, such as the D-optimality criterion, which consists in minimizing the determinant of the inverse of M_F over the space of possible designs, which is equivalent to reducing the expected variance-covariance matrix. This methodology was originally

developed in classic non-linear regression and was first extended to NLMEM by Mentré *et al.* (8) and Retout *et al.* (9). Its use has been facilitated by the implementation of the Individual M_F (M_{IF}) and Population M_F (M_{PF}) expressions in PFIM, an R function for individual and population design evaluation and optimization (10). This approach to population design evaluation is now used in other software as an alternative to PFIM, like PopED (11), PopDes (12), POPT (13), and has been extended to adaptive or robust optimal design in pharmacokinetic studies (14,15).

Although those software programs are very appealing for deriving sparse sampling design associated with good precision of population parameter estimates, they do not take into account the precision of the individual estimates by MAP. Although Merlé *et al.* (17) proposed an approach to predict the precision of the individual parameter estimates based on an approximation of the Bayesian information Matrix (M_{BF}) in NLMEM using a First Order linearization (FO), this has, to our knowledge, never been implemented in such software. For example, the current version of PFIM predicts the SE of individual estimates, but only for standard non-linear regression, assuming all variances of random effects to be null (16). The bias in the individual estimates that can be induced by the low number of samples per subject is also not taken into account in those optimal design methodologies. Indeed, for a rich design, the *a posteriori* distribution of each random effect is very narrow and centered around the true value (18,19). However, in the case of a sparse design, with less information per subject, the *a posteriori* distribution is more spread out, and the mean value is close to 0. This problem of shrinkage (Sh) is well described in linear mixed-effects methodology (20). In NLMEM, the variations in the random effects distribution and their consequences for model diagnostics have been highlighted recently by Savic *et al.* (21).

The main objective of this work was to predict shrinkage associated with a design using Bayesian methodology. We first approximated M_{BF} using FO of the model proposed by Merlé *et al.* in 1995 (17). From previous work by Fedorov (20) in linear mixed effects modeling methodology, we derived a predicted shrinkage from M_{BF} FO. Shrinkage and SE of random effects were then predicted and compared with observed shrinkage and SE estimated using simulation studies.

We used two different PK models as examples. The first is a simple one compartment PK model (22), where different variability and uncertainty levels are considered in order to evaluate the results in various situations. The second is a two-compartment PK model with an approximated target-mediated drug disposition drawn from real clinical studies (23) and supposed to mimic the complexity of some models commonly used in drug development. Simulated data from those PK models were analyzed using both NONMEM and MONOLIX.

MATERIALS AND METHODS

Models and Notations

For a given individual, let y be the vector of n observed concentrations at n sampling times $\xi = \{t_1, \dots, t_n\}$ and f be the known function describing the PK model. So, y can be described by $y = f(\theta, \xi) + \varepsilon$ where θ is the p -vector of individual PK parameters, $\theta = (\theta_1, \dots, \theta_p)^T$.

ε is the random error, following a normal distribution with zero mean and variance $\Sigma(\theta, \xi)$. $\Sigma(\theta, \xi)$ is defined with an additive part σ_{inter} and a proportional part σ_{slope} with $\Sigma(\theta, \xi) = \text{diag}((\sigma_{inter} + \sigma_{slope} f(\theta, \xi))^2)$.

Considering N individuals i following the same PK model structure, individual parameters θ_i are defined as $\theta_i = g(\mu, \eta_i)$ where μ is the vector of p fixed effects and η_i is the vector of individual random effects $\eta_i = (\eta_{i1}, \dots, \eta_{ip})^T$. Random effects η_i follow a normal distribution with zero-mean and variance Ω . Here we assumed that Ω is a diagonal $p \times p$ matrix of variances of random effects $\Omega = \text{diag}(\omega_1^2, \dots, \omega_p^2)$. We considered two models for $g(\mu, \eta_i)$: additive, where $g(\mu, \eta_i) = \mu + \eta_i$, or exponential, where $g(\mu, \eta_i) = \mu e^{\eta_i}$. The vector Ψ of population parameters is composed of $\{\mu_1, \dots, \mu_p, \omega_1^2, \dots, \omega_p^2, \sigma_{inter}, \sigma_{slope}\}$. As in this work we focused on individual parameter's estimation, Ψ was supposed to be known or previously estimated. By this approach we sought to avoid any bias or distribution discrepancies in parameter distribution to only validate M_{BF} approximation.

Individual Parameter Estimation

From a vector of observed concentrations y , individual parameters can be estimated by standard individual non-linear regression. This approach does not take into account previous information. On the contrary, in NLMEM, individual parameters are usually estimated by a Bayesian approach, such as the MAP. As μ is known, estimating η is similar to estimating θ . More precisely, the MAP estimate $\hat{\eta}$ of η is given by:

$$\hat{\eta} = \underset{\eta}{\text{argmax}} \left(p(\eta|y) \right) \quad (1)$$

It can be also written, using Bayes theorem,

$$\begin{aligned} \hat{\eta} &= \underset{\eta}{\text{argmax}} \left(\frac{p(y|\eta) \times p(\eta)}{p(y)} \right) \\ &= \underset{\eta}{\text{argmax}} \left(\log(p(y|\eta)) + \log(p(\eta)) \right) \end{aligned} \quad (2)$$

as $p(y)$ is independent of η .

Bayesian Fisher Information Matrix

For individual estimation, in a non-linear model, M_{IF} for parameters θ is:

$$M_{IF}(\theta, \xi) = -E_y \left(\frac{\partial^2 \ln(p(y|\theta))}{\partial \theta \partial \theta^T} \right) \quad (3)$$

For homoscedastic error, M_{IF} is expressed as:

$$M_{IF}(\theta, \xi) = F(\theta, \xi)^T \Sigma(\theta, \xi)^{-1} F(\theta, \xi) \quad (4)$$

where $F(\theta, \xi) = \frac{\partial f(\theta, \xi)}{\partial \theta}$.

For Bayesian estimation of the random effects, M_{BF} is defined by (22):

$$\begin{aligned} M_{BF}(\xi) &= -E_\eta \left(\frac{\partial^2 \log(p(\eta|y))}{\partial \eta \partial \eta^T} \right) \\ &= -E_\eta \left(E_{y|\eta} \left(\frac{\partial^2 \log(p(y|\eta))}{\partial \eta \partial \eta^T} \right) \right) - E_\eta \left(\frac{\partial^2 \log(p(\eta))}{\partial \eta \partial \eta^T} \right) \end{aligned} \quad (5)$$

In Eq. (5), M_{BF} is a sum of 2 matrices respectively representing individual and prior information. As $p(\eta)$ is normal, $-E_\eta \left(\frac{\partial^2 \log(p(\eta))}{\partial \eta \partial \eta^T} \right) = \Omega^{-1}$. Also, noting the link between the first part of the equation and M_{IF} , we have:

$$M_{BF}(\xi) = E_\eta(M_{IF}(g(\mu, \eta), \xi)) + \Omega^{-1} \quad (6)$$

We used two methods to compute M_{BF} (17). An evaluation by Monte-Carlo simulation (M_{BF} MC), where a large number of η_i are simulated from $\mathcal{N}(0, \Omega)$ and the expectation in Eq. (5) is computed as the empirical mean. M_{BF} can also be evaluated using FO linearization of the model around the random effect in the expression of the likelihood. It can be shown that M_{BF} is then approximated by:

$$M_{BF}(\xi) = M^T F(\mu, \xi)^T \Sigma(\mu, \xi)^{-1} F(\mu, \xi) M + \Omega^{-1} \quad (7)$$

where M is an identity matrix (I) for additive random effects, and $M = \text{diag}(\mu_1, \dots, \mu_p)$ for exponential random effects.

M_{BF} is the expected information over all η . We can also compute the Bayesian information Matrix for a given individual (M_{IBF}) which depends on its random effect. Using FO, M_{IBF} is approximated by:

$$M_{IBF}(\eta, \xi) = \Theta^T F(g(\mu, \eta), \xi)^T \Sigma(g(\mu, \eta), \xi)^{-1} F(g(\mu, \eta), \xi) \Theta + \Omega^{-1} \quad (8)$$

where Θ is an identity matrix for additive random effects, and $\Theta = \text{diag}(\theta_1, \dots, \theta_p)$ for exponential random effects.

Predicted Shrinkage

Shrinkage in linear mixed effect models is well established (24) and Fedorov in 2011 (20) proposed the matrix $I-W$ for its quantification. Assuming that the linear model is defined as $y = F(\xi)\theta + \varepsilon$, then, for individual estimation

$$\hat{\theta}_{ML} = M_{IF}(\xi)^{-1} F(\xi)^T y \quad (9)$$

where $M_{IF}(\xi) = F(\xi)^T \Sigma(\xi)^{-1} F(\xi)$ (25). In both (20) and (24), it is shown that the MAP estimate of θ can be expressed as

$$\hat{\theta}_{MAP} = (M_{IF}(\xi) + \Omega^{-1})^{-1} (M_{IF}(\xi) \hat{\theta}_{ML} + \Omega^{-1} \mu) \quad (10)$$

By setting

$$W(\xi) = I - (M_{IF}(\xi) + \Omega^{-1})^{-1} \Omega^{-1} \quad (11)$$

Then,

$$I - W(\xi) = (M_{IF}(\xi) + \Omega^{-1})^{-1} \Omega^{-1} \quad (12)$$

Equation (10) becomes

$$\hat{\theta}_{MAP} = (I - W(\xi)) \mu + W(\xi) \hat{\theta}_{ML} \quad (13)$$

W then quantifies the balance between prior and individual information. Indeed, $\hat{\theta}_{MAP} = \hat{\theta}_{ML}$ when $W = I$ and $\hat{\theta}_{MAP} = \mu$ when $W = 0$. We propose to extend this formula in non-linear mixed-effect models where W is computed from M_{BF} as:

$$I - W(\xi) = M_{BF}(\xi)^{-1} \Omega^{-1} \quad (14)$$

As Ω is diagonal, the diagonal elements of $I-W$ are the ratio of the predicted estimated variance of the Bayesian estimation and the *a priori* variance. When M_{BF}^{-1} is close to Ω (no information), $I-W$ is close to I . In this case, shrinkage is expected to be high.

Simulation Study

Example 1

We used a simulated example inspired by similar methodology reported by Mentré *et al.* (22). This theoretical example involved a one-compartment PK model with IV bolus administration. The two parameters are volume of distribution (V) and clearance (CL). For a given dose D , the prediction at time t_j is:

$$f(\theta, t_j) = \frac{D}{\theta_1} \exp\left(-\frac{\theta_2}{\theta_1} t_j\right) \quad (15)$$

with $\theta_1 = V$ and $\theta_2 = CL$. We assumed $\mu_V = 0.2$ IU, $\mu_{CL} = 0.5$ IU and a dose $D = 1$ IU. As this model is theoretical, we did not

consider any specific units. To compute the Fisher matrices in various situations, we considered several scenarios with different types of random effects models (*ie*: additive or exponential), different types of residual error models (*ie*: additive or combined) and different values for the random effects. No correlation between parameters was assumed. All scenarios are given in Table I. Scenarios were named with two letters, the first one representing the type of random effect models and the second one the type of residual error model (*ie*: a for additive, e for exponential, and c for combined). The lower and upper-case letters stand for low variance (Coefficient of Variation (CV) of 20%) and high variance (50% CV) of random effect, respectively. For this example 1, we performed two simulations A and B. Simulation A evaluates the use of M_{BF} in predicting shrinkage for each scenario, the vector Ψ of population parameters being considered as known and fixed at the true simulation value. We defined a set of five possible sampling times: {0.05, 0.15, 0.3, 0.6, 1} and D-optimal designs with 2 to 5 samples were derived considering the first (aa) scenario. These designs are described in Table II. Simulation B compares predicted *versus* observed shrinkages in a more realistic setting, incorporating the discrepancies due to the population parameter estimation step as often encountered in a real estimation context. This simulation B was performed for scenario ec using a combined design with 2 groups of 50 and 150 subjects each containing 5 and 2 samples per subject respectively, with sampling times similar to those in simulation 1 (Table II).

An additional investigation using this example was performed to explore the impact of M_{BF} on design optimization and is presented in the Supplementary Materials I. We compared the optimal design obtained with M_{BF} and M_{PF} and computed the loss of efficiency associated with each one.

Example 2

The second example is a more complex PK model, adapted from the population pharmacokinetic model of an anti-interleukin-6 receptor antibody used as rheumatoid arthritis treatment (23). The PK model, established using a large

Table I Population Parameter Values for Each Scenario of Example 1

Scenario	aa	ac	ea	ec	Ea	Ec
Random effects						
Form	Add	Add	Exp	Exp	Exp	Exp
ω_V (CV%)	20	20	20	20	50	50
ω_{CL} (CV%)	20	20	20	20	50	50
Residual error						
σ_{inter}	0.15	0.15	0.15	0.15	0.15	0.15
σ_{slope} (CV%)	0	15	0	15	0	15

Add additive; Exp Exponential

Table II Sampling Times of Example 1

Number of samples	Sampling times
5	0.05, 0.15, 0.3, 0.6, 1
4	0.05, 0.3, 0.6, 1
3	0.05, 0.3, 0.6
2	0.05, 0.3

phase III database, is a two compartment model with parallel first-order (linear) and Michaelis-Menten (non-linear) elimination kinetics. With this model, the drug amounts in the central (A_1) and peripheral (A_2) compartments were described by the following differential equations:

$$\frac{dA_1}{dt} = -k_{10}A_1 - k_{12}A_1 + k_{21}A_2 - \frac{C_1 V_m}{K_m + C_1} \quad (16)$$

$$\frac{dA_2}{dt} = k_{12}A_1 - k_{21}A_2 \quad (17)$$

where k_{10} , k_{12} and k_{21} are rate constants of drug exchanges between the different compartments: $k_{10} = \frac{CL}{V_1}$, $k_{12} = \frac{Q}{V_1}$ and $k_{21} = \frac{Q}{V_2}$; C_1 is the concentration in the central compartment: $C_1 = \frac{A_1}{V_1}$, where CL denotes the clearance, V_1 is the volume of distribution of the central compartment, V_2 is the volume of distribution of the peripheral compartment, V_m is the maximum elimination rate, K_m is the Michaelis constant, and Q is the intercompartmental clearance.

The population parameter values are detailed in Table III. A combined error model and exponential random effects were used. As the FO approximation is appropriate only with small variances, for simplicity, we rounded off fixed effect values and assumed 30% CV for between-subject variability on CL , V_1 , V_2 and V_m . A model with high variance, as in the basic model developed by N. Frey *et al.* (only full model with covariate described in (23)), and a model with intermediate variabilities (80%CV on V_2) were also simulated without the covariate model and the results are given in the Supplementary Materials II (Table SII and Figure S1). Six repeated one-hour infusions every 4 weeks at the dose of 8 mg/kg were assumed. Four different designs were simulated, involving 10, 9, 4 or 2 samples per subject, at the same sampling times, with 1,800 subjects. The design with 10 samples was created by collecting 3 samples after the first dose (end of infusion, 10 and 20 days), 3 pre-doses samples before the 2nd, 3rd and 5th doses, and then 4 samples after the last dose (end of infusion, 20, 28 and 56 days after the last dose). The last sample (56 days after the last dose) was removed for the 9-samples design. Using the same sampling allocations, a peak and trough design was applied after the first dose for the 2 samples per subject

Table III Population Parameter Values Used in Example 2

Parameter	Simulation value
μ_{CL} (l/d)	0.3
μ_{V_1} (l)	4
μ_Q (l/d)	0.15
μ_{V_2} (l)	3
μ_{V_m} (mg/d)	6
μ_{K_m} (μ g/ml)	2
Random effects (CV%)	
ω_{CL}	30
ω_{V_1}	30
ω_{V_2}	30
ω_{V_m}	30
Residual error	
σ_{inter} (μ g/ml)	0.5
σ_{slope} (%)	30

design. From this last design, 2 sampling allocations after the last dose (end of dose and 56 days after the last dose) were added to create a 4 samples per subject design. Sample allocations are detailed in Table IV.

Software and Analysis

For example 1, datasets were simulated using R 2.14. NONMEM 7.0 and MONOLIX 4.0 were used to estimate individual random effects and their precision of estimation (estimated SE) given the model. We used two software programs as they are both extensively used and considered as references in drug development for population analysis using NLMEM. For simulation A, one dataset of 1,000 subjects was simulated per scenario and per design. Population parameters estimation was skipped using specific options in both software programs and only individual parameters were estimated as MAP. In NONMEM, population parameters were specified to their true values by using the option MAXEVAL = 0 (number of objective function evaluations) (26), and individual parameters were estimated

Table IV Sampling Times of Example 2

Number of samples	Sampling times (days)
10	0.06, 10, 20, 27.99, 82.02, 111.99, 140.06, 168, 196
9	0.06, 10, 20, 27.99, 82.02, 111.99, 140.06, 168, 160
4	0.06, 27.99, 140.06, 196
2	0.06, 27.99

using the first order conditional estimation with interaction method (FOCEI). In MONOLIX, the number of chains and the number of iterations (K_1 and K_2) were specified respectively at 5 chains, 800 and 300 iterations with fixed population parameters. For simulation B, we simulated 1,000 datasets and estimated both population and individual parameters using the SAEM algorithm implemented in MONOLIX.

In a first step, using the individual random effects estimates given by NONMEM, we investigated the influence of shrinkage on individual estimates and on estimated random effects distribution. Then, in order to evaluate the expression of M_{BF} with the FO approximation, we compared Relative Standard Errors (RSE) computed with M_{BF} FO to M_{BF} MC using the 1,000 simulated subjects. We also compared the quantity of information provided by the 3 considered regressions (individual with M_{IF} , population with Ω and individual Bayesian with M_{BF}). Ω defined the variance of the estimation yielded by Bayesian estimation without any individual information, and is therefore an upper bound of the variance of the individual Bayesian estimation. Predicted shrinkage computed with M_{BF} FO was compared with observed shrinkage (Sh) on random effects estimated with NONMEM and MONOLIX without population estimation, and MONOLIX with population estimation. Observed shrinkage for parameter k was defined as (21):

$$Sh_k = 1 - \frac{\text{var}(\hat{\eta}_{i,k})}{\omega_k^2} \quad (18)$$

$\hat{\eta}_{i,k}$ being the estimated value of η , and $\text{var}(\hat{\eta}_{i,k})$ the empirical variance, estimated by NONMEM or MONOLIX. Note that in the case of very low shrinkage Eq. (18) can produce negative values which were set to 0. Although Savic *et al.* expressed shrinkage as the ratio of standard deviations, the ratio of variances, which is more usual in biostatistics, was used here. All thresholds and values extracted from Savic *et al.* were converted to ratio of variances in the present paper. Finally, using simulation results, we compared each estimated individual SE with the predicted SE obtained from the diagonal of M_{BF}^{-1} .

Datasets for example 2 were simulated with NONMEM 7.0, and the same options as for example 1 (simulation A) were used for estimation with NONMEM 7.0 and MONOLIX 4.0. We performed two steps similar to those in the first example: we compared predicted *vs* observed shrinkage and individual SE. M_{BF} , M_{IF} and shrinkage computations were performed with R 2.14. This methodology should be implemented in a future version of PFIM. Meanwhile, the R code used to compute M_{BF} , SE, RSE and predicted shrinkage is given in Supplementary Materials III.

RESULTS

Example 1

Shrinkage Impact on Individual Estimates

The simulated scenarios generated observed shrinkage values that ranged from 0% to 69% for CL and from 0 to 52% for V. All the shrinkage values are given in Table V for each scenario and for each number of samples per subject for the two software programs. The highest shrinkage was obtained for 2 samples per subject with a low variance for CL and V and a combined error model. As NONMEM and MONOLIX provide similar results, and as the observations for volume are similar to those for clearance, only NONMEM results for clearance are shown and discussed in the rest of this section. Figure 1 presents individual parameter estimated with NONMEM *vs* simulated parameters along with observed shrinkage for each scenario and for each number of samples. This figure also shows the distribution of estimated and simulated random effects. With small shrinkage ($\approx 7\%$ for clearance, scenario ea with 5 samples), the estimated η distribution (in red) is close to the simulated η distribution (in black) and the simulated and estimated individual parameters are close to the identity line. With 2 samples per subject with the same scenario, the shrinkage increases to 20%. The simulated and estimated η distributions remain relatively similar, as do the simulated and estimated individual parameters. Changing the form of the random effects from exponential to additive with the same variance (20% CV) had no major effect on shrinkage or parameter estimation. Changing the residual error from additive to combined for the same scenario increased the shrinkage to 69%. For this high shrinkage, the estimated η distribution clearly shrank towards zero, and the estimated individuals parameters deviated from the simulated values. When the CV of exponential random effects was increased to 50% CV, individual random effects were very well estimated whatever the design used. Thus, shrinkage for a combined error model and 2 samples per subject decreased to 33%.

Evaluation of M_{BF} FO Approximation

Predicted Relative Standard Errors (RSE) calculated from M_{IF}^{-1} , M_{BF}^{-1} with FO or MC, and Ω are reported for each scenario in Fig. 2. FO approximated values from M_{BF}^{-1} were very close to those using Monte-Carlo simulations, thus validating the FO approximation in computing M_{BF}^{-1} . Similar values were found for low between-subject variability, and only slight differences were observed, without exceeding 5% for higher variances of random effects (scenarios Ea, Ec). Note that as expected, predicted RSE increased with decrease of information per subject. RSE approximations from M_{BF}^{-1} FO were always below the value of Ω

Table V Observed Shrinkage Values with NONMEM and MONOLIX of Example I

Scenario	Number of samples per patient	Volume		Clearance	
		NONMEM	MONOLIX	NONMEM	MONOLIX
aa	2	1.77	3.39	17.41	17.31
	3	1.55	3.08	8.62	8.64
	4	0.92	2.45	6.22	6.24
	5	0.10	0.10	5.76	5.76
	5	0.10	0.10	5.76	5.76
ac	2	48.95	48.95	66.56	66.56
	3	46.93	46.92	36.90	36.90
	4	45.14	45.14	30.00	30.00
	5	38.35	38.34	28.39	28.39
ea	2	1.91	1.91	20.11	20.11
	3	1.19	1.19	11.42	11.42
	4	0.77	0.77	9.05	9.05
	5	0.54	0.54	7.11	7.11
ec	2	51.54	51.54	68.79	68.79
	3	48.85	48.85	41.31	41.31
	4	44.89	44.89	33.53	33.53
	5	35.91	35.91	31.51	31.51
Ea	2	1.50	1.30	7.55	7.59
	3	1.02	0.82	1.47	1.54
	4	0.49	0.28	0.00	0.00
	5	0.00	0.00	0.00	0.00
Ec	2	16.31	16.31	32.46	32.46
	3	18.32	18.32	21.35	21.35
	4	17.77	17.77	10.14	10.14
	5	15.10	15.10	9.46	9.46

and M_{IF}^{-1} RSE and those computed by M_{BF}^{-1} MC were always smaller than or very close to Ω and M_{IF}^{-1} . When fewer samples per subject were collected, RSE predicted by M_{BF}^{-1} increased less than when predicted by M_{IF}^{-1} , showing the advantages of such a Bayesian approach in individual parameter estimates.

Optimal Design

In Supplementary Materials I, we compared the criteria from M_{BF} and M_{PF} for 3420 designs. We found that the optimal designs for Sh-optimality and population-optimality differed, and were $\{0.05, 0.75, 0.8\}$ and $\{0.05, 0.35, 1\}$, respectively. Also, for each criterion (M_{BF} or M_{PF}) there was a loss of efficiency when using the design optimal for the other criteria (see Table SI).

Shrinkage Prediction

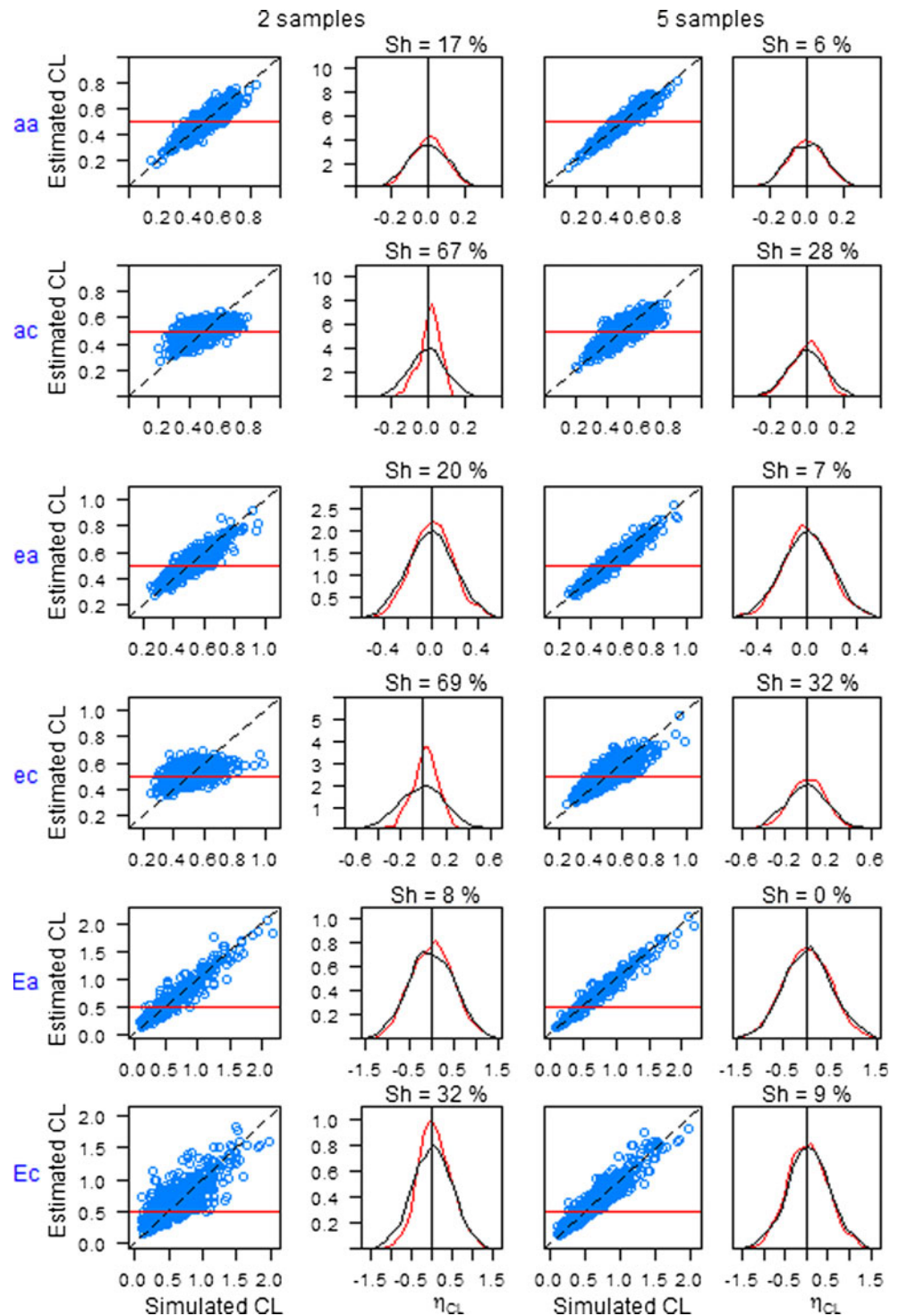
Figure 3 shows predicted *vs* observed shrinkage without population estimation with NONMEM and MONOLIX (simulation A - 2 left columns) and with population estimation with MONOLIX (simulation B - right column) for CL and V. For

simulation A, all the points were close to the identity line for CL and V and for both software programs, indicating that the computation of $I-W$ proposed in Eq. (14) using M_{BF}^{-1} FO predicts shrinkage very well across a large range of shrinkage values when Ψ is known. When including the estimation of the population parameters in the comparison, simulation B, the predicted shrinkage value obtained for each parameter corresponded to the median of the distribution of the observed shrinkages. The greatest difference between the observed and predicted values did not exceed 20%. This validates the use of M_{BF}^{-1} FO to predict shrinkage, even in a real context, including estimation of the population parameters.

SE of Random Effects

SE from M_{IBF}^{-1} were computed and then compared with estimated SE of individual random effects using NONMEM and MONOLIX. The results are shown in Fig. 4. As expected, SE decreased when the number of samples for each subject increased, for the four methods. Adding a proportional part in the residual error model led to an increase in median SE. In the same way, changing random effects from additive to exponential or increasing the value of their variance from

Fig. 1 Example 1: Estimated vs simulated clearance (CL) and comparison of simulated and estimated distribution of random effects by scenario for 2 (left columns) and 5 samples (right columns) per subject. Indicated Sh values are the observed Sh. For the scatterplots, the horizontal red line represents the population value (μ) and the dotted black line is the identity line. For η distributions, the red and black curves are respectively the estimated η distribution and the simulated distribution. The black vertical line represents the population mean of η , 0.



20 to 50% CV led to an increase in SE. The M_{BF}^{-1} FO predicted value was always near the median value computed with M_{IBF}^{-1} . Similar values were observed with NONMEM and MONOLIX. Predicted SE from M_{IBF}^{-1} were very close to the estimated SE for all scenarios. In the case of a combined error model, MONOLIX systematically estimated a wider distribution of the SE of volume.

Example 2

Shrinkage Prediction

Figure 5 shows predicted vs observed shrinkage for each parameter of the complex model and each design. The dots were close to the identity line whatever the number of

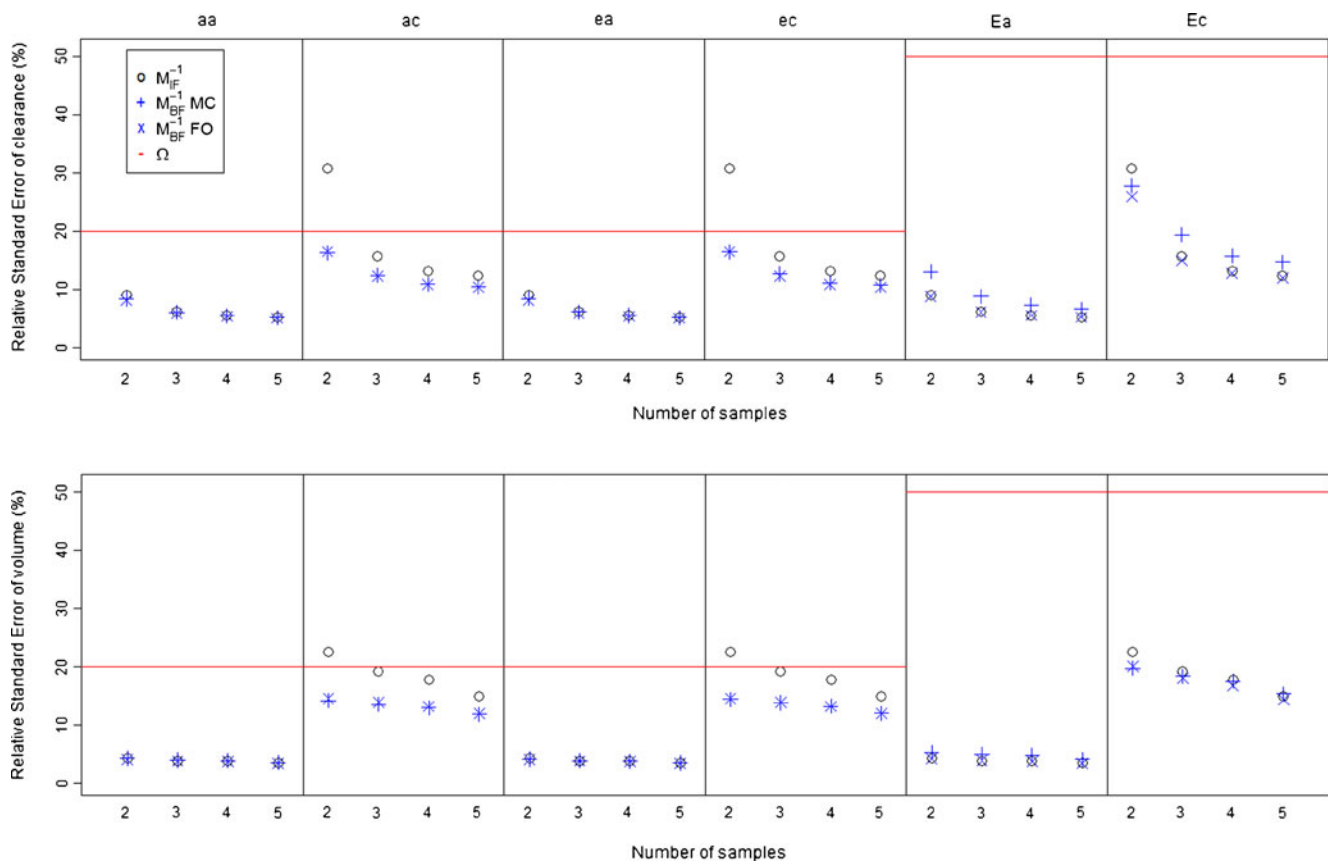


Fig. 2 Example 1: relative standard errors of volume and clearance computed from M_{IF}^{-1} (circle), M_{BF}^{-1} with FO (blue diagonal cross “x”) or MC (blue vertical cross “+”), along with SD computed from Ω (similar to RSE in the absence of any individual data—red horizontal line).

samples per subject and the PK parameter. There was slightly better agreement between the predicted shrinkage and the shrinkage computed using MONOLIX individual estimates than with the NONMEM estimates, especially for CL for the 2 designs with the smaller number of samples. In section II of the Supplementary Material we report two simulations for which real variability values of the basic model as well as intermediate ones (see Table S1) were used to compare predicted *versus* observed shrinkage. These simulations showed some discrepancies between predicted and observed shrinkage, especially for V_2 , and these discrepancies increased with variability (see Fig. S1).

SE of Random Effects

Figure 6 shows the comparison of predicted *vs* observed SE. As expected, SE decreased when the number of samples per subject increased. Whatever the design and the parameter, M_{IBF}^{-1} predictions were close to the values estimated with NONMEM and MONOLIX. SE computed by M_{BF}^{-1} FO were highly consistent with the M_{IBF}^{-1} predictions.

DISCUSSION

We proposed a method to predict shrinkage using an approximation of M_{BF} in NLMEM based on a FO linearization and we evaluated it for a large range of shrinkage values and from a simple linear PK model to a more complex non-linear one. The approximated M_{BF} was very close to the one obtained by Monte-Carlo simulations and the proposed prediction method was shown to give a very good prediction of observed shrinkage. Our investigations also highlight that shrinkage is influenced by design (*eg.* number of samples per subject and sample allocations), but also by the ratio between residual error and random effect variances, a small ratio resulting in high shrinkage.

To the best of our knowledge, this is the first time that shrinkage has been predicted using M_{BF} in NLMEM. The main publications regarding Bayesian design evaluation and optimization in NLMEM were written by Merlé *et al.* (17) and Mentré *et al.* (22) in the 90's. The first paper focuses on defining and simplifying the computation of three Bayesian design criteria for several designs and PK models: the determinant of M_{BF} , the determinant of the pre-posterior covariance matrix and the expected information provided

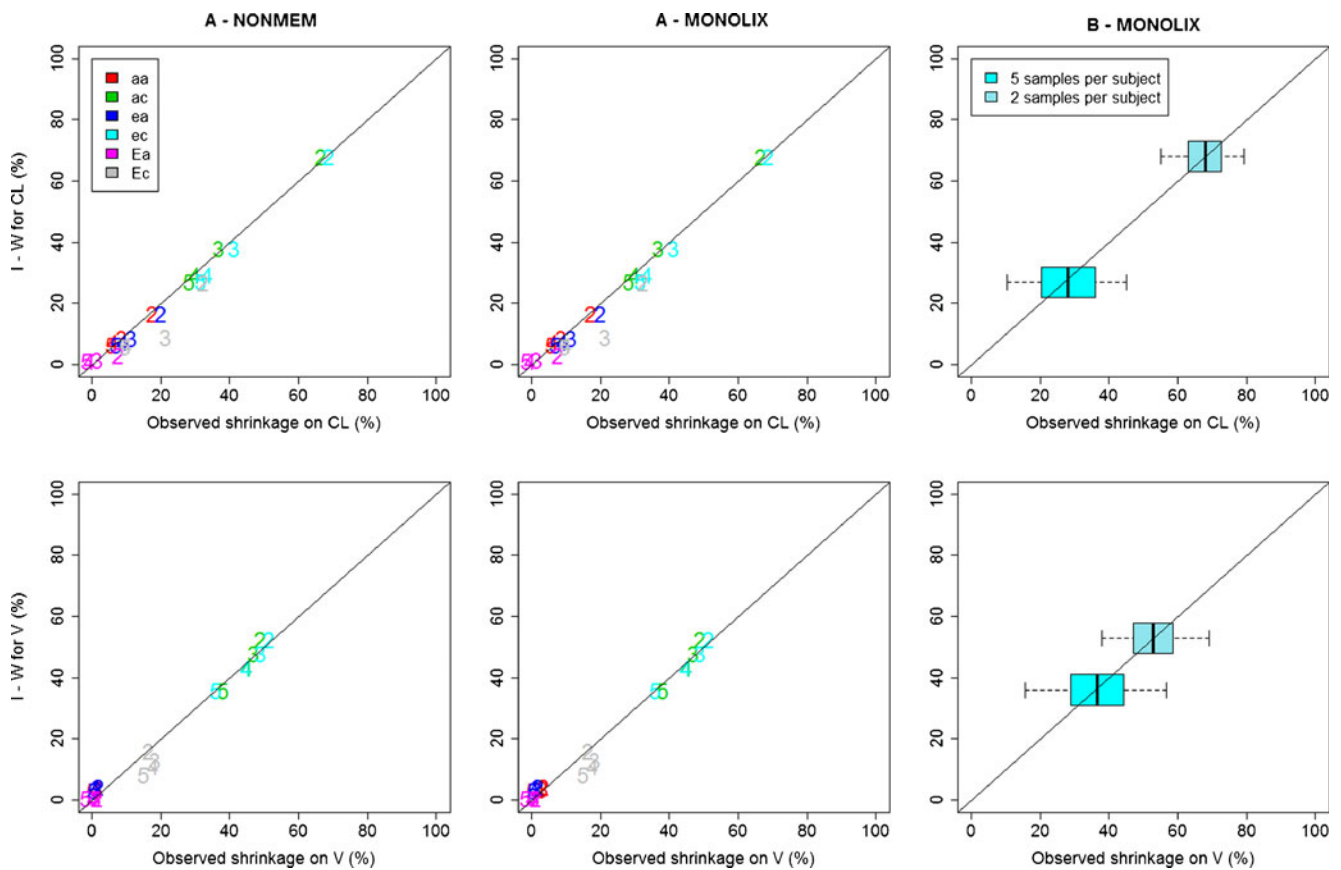


Fig. 3 Example 1: predicted shrinkage value vs observed shrinkage values for simulation A (2 left columns) and simulation B (right column). Each number defines the number of samples per subject. The black line represents the identity line. Boxplot for simulation B represents the 1,000 observed shrinkage values for each design.

by an experiment. These criteria indicate the relevance of computing the matrices using a linearization of the model around the mean of the prior distribution. The second paper focuses on the use of the determinant of M_{BF} to optimize the design for Bayesian individual estimates. Several population and Bayesian designs are evaluated and compared in this study, from which the authors concluded that there is a difference between the optimal Bayesian and population designs, even if optimal Bayesian designs sometime involve the same sampling times than the optimal population design. In the present work, we confirmed that optimal Bayesian and population designs are different for a given model.

With the proposed approach in this publication, shrinkage optimization can be included in addition to optimization of the SE of the population parameters. This optimization requires the use of compound criteria and evaluation of the impact of the weight of each criterion on the design and on the population and individual parameter estimates (27).

This work also evaluated the expression and the use of individual M_{BF} to predict SE of the individual parameter

estimates. In all investigated cases the predictions were very close to the individual SE generated by NONMEM and MONOLIX. In some cases, the ranges of the individual SE estimated by MONOLIX were found to be higher than in NONMEM. This was especially the case for the volume of distribution of the more complex PK model. This may be due to the fact that MONOLIX computes the variance of the conditional distribution without linearization.

Although the predicted and observed shrinkages were similar, we observed minor discrepancies for example 2. For 2 or 4 samples per subject, predicted shrinkages on CL and V_1 were below the observed values. This was not the case when individual parameters were estimated by MONOLIX. Considering M_{IBF} computation of SE in this more complex model, predictions were close to the estimated results using NONMEM for all parameters except V_1 . SE distributions estimated using MONOLIX were wider than those with NONMEM, with a similar median. We hypothesized earlier in this paper that these differences stem from the different SE computation methods used in NONMEM and MONOLIX, but further investigation is warranted.

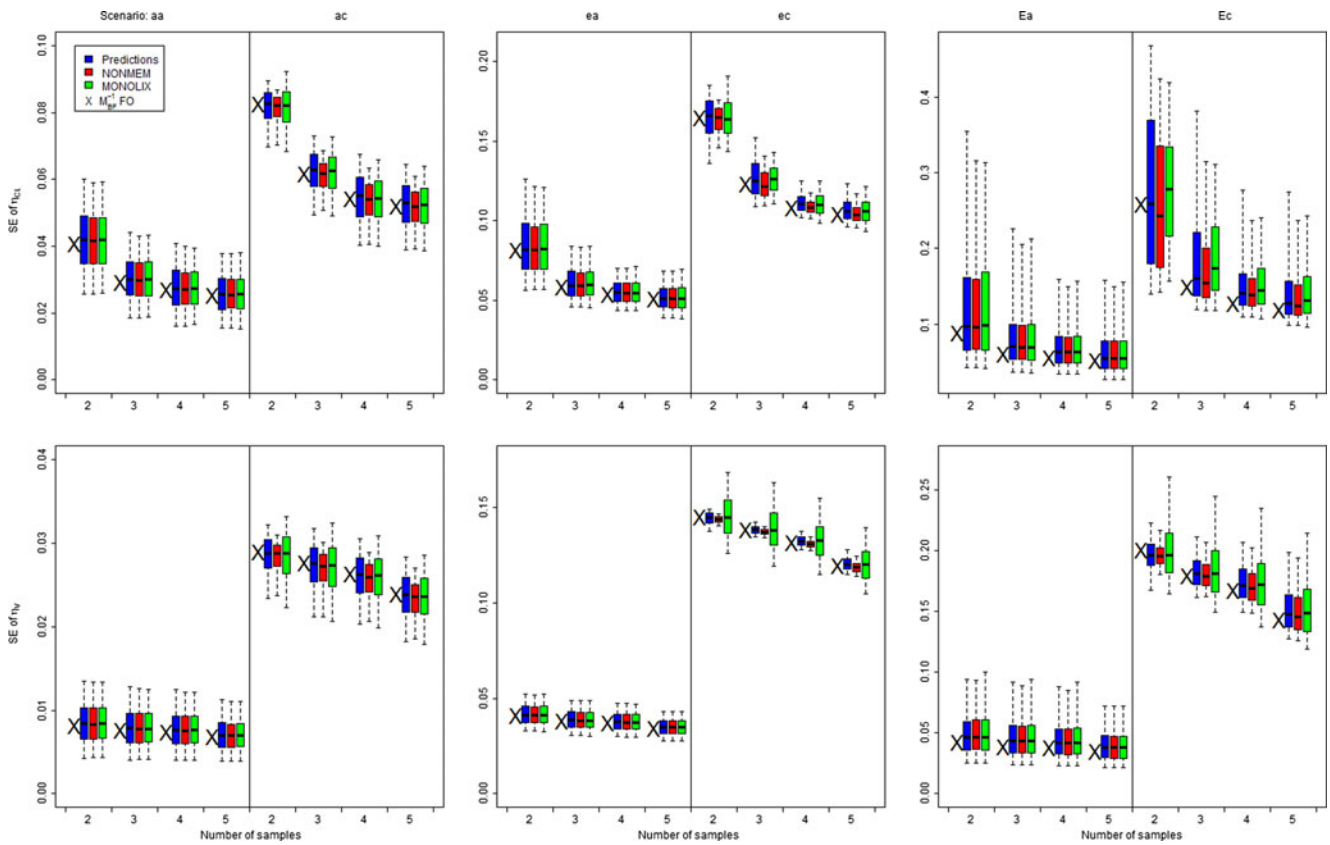


Fig. 4 Example 1: Predicted and estimated individual SE of clearance and volume. Blue boxplots are predicted SE by M_{BF}^{-1} , red ones are SE estimated by NONMEM and green by MONOLIX. Black diagonal cross “X” is the value predicted by M_{BF}^{-1} using FO approximation.

In this paper, we made the assumption of a diagonal Ω matrix for the sake of simplicity. The same formula would hold for a non-diagonal matrix.

Prediction of shrinkage was evaluated for example 2, changing the variances of the random effects to smaller and more reasonable value of 30% compared with the true base model developed by Frey *et al.*, which estimated 110% variability on V_2 . However, although the FO approximation used

to compute M_{BF} performs very well for reasonable variability, increase in variability worsened prediction, especially for very large variabilities as shown in the Supplementary Materials. This is certainly a limitation of the FO approximation, which is known to be appropriate only if the variances of the random effects are relatively small (28,29).

In example 1 - simulation A and example 2, population parameters were set to their true value. With this method,

Fig. 5 Example 2: predicted shrinkage vs observed shrinkage according to the number of samples per subject (2, 4, 9 or R for 10). The black line represents the identity line.

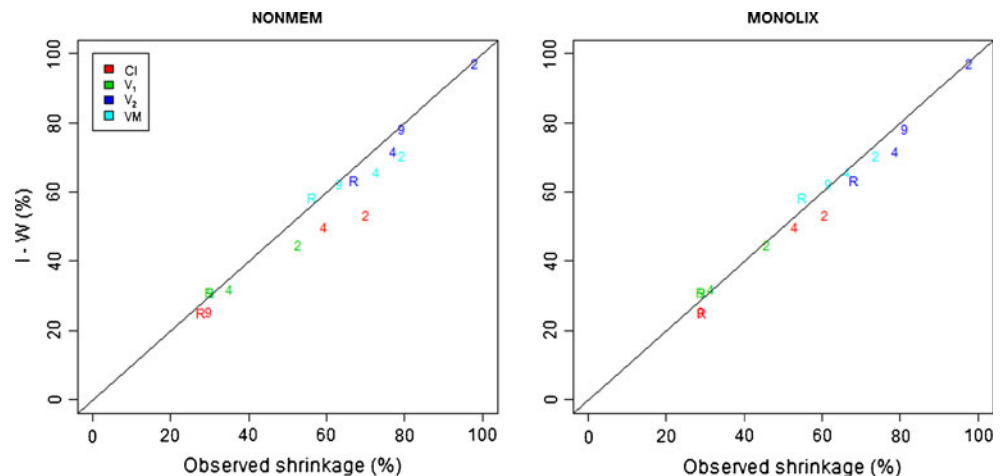
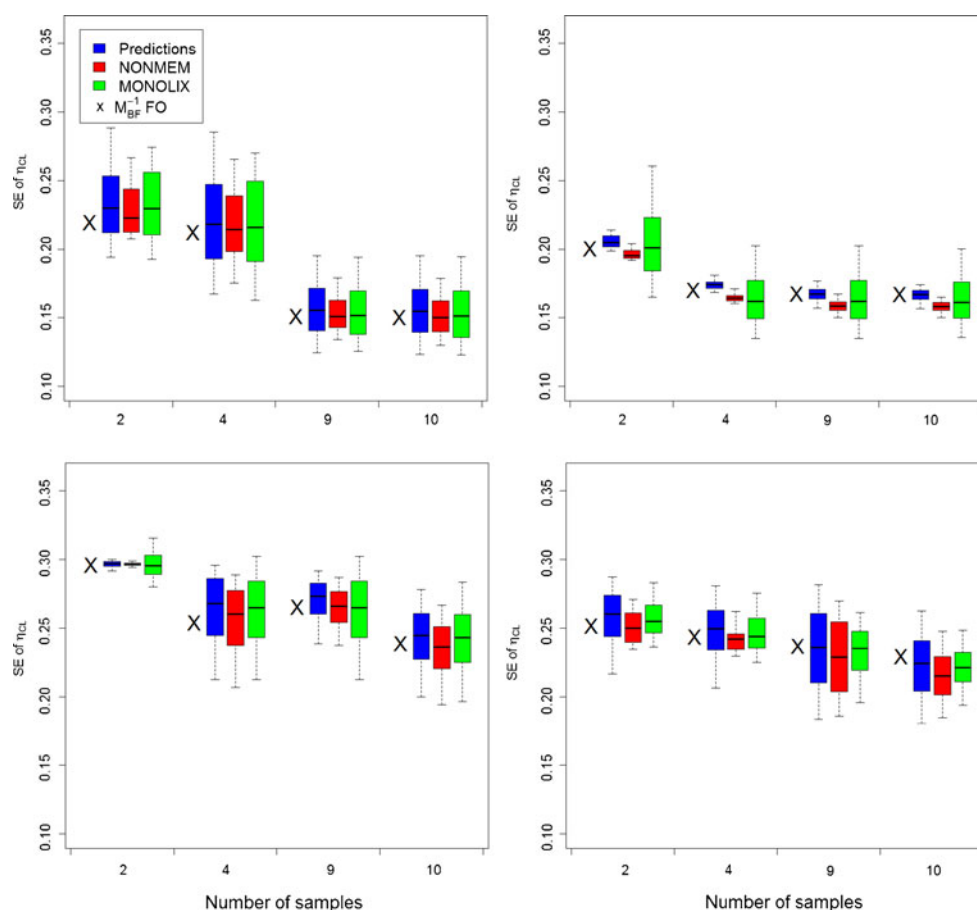


Fig. 6 Example 2: predicted and estimated SE of CL, V_1 , V_2 and V_m . Blue boxplots are predicted SE by M_{BF}^{-1} , red ones are SE estimated by NONMEM and green by MONOLIX. Black diagonal cross "X" is the value predicted by M_{BF}^{-1} using FO approximation.



we wanted to avoid any bias or distribution discrepancies in parameters distribution. Indeed, the goal of this simulation was to evaluate the prediction of shrinkage and individual SE for some given estimated individual parameters, with the smallest possible external influence such as population parameters estimation. However, to explore what could happen in a real estimation context, we also explored the deviation between predicted shrinkage computed for true simulation values and the observed shrinkage obtained after a first step in estimation of the population parameters and a second Bayesian step in estimation of the individual parameters. This exploration showed a good agreement between predictions and observations. Nevertheless, we acknowledge that the example used was rather simple with a fully informative design (combined sparse and rich elementary design with 120 and 50 subjects), and that further investigations with more complex models as often encountered in drug development associated with rather sparse designs should be performed.

Lastly, Savic *et al.* (21) showed that high shrinkage can induce or hide a correlation between parameters and covariates. A hypothesis of no correlation could therefore be falsely accepted when this correlation does not truly exist. One perspective would be to link the shrinkage with the power of the test used to detect that covariate. That could

minimize the risk of falsely detected (or not detected) influence of some covariate due to uninformative design. Future work should then focus on the impact of the precision of estimates on covariate selection and on the link between shrinkage and the power of the test for covariate selection.

CONCLUSION

We introduce a methodology based on the Fisher information Matrix to predict accurately the shrinkage associated with a given design in non-linear mixed-effect models. Bayesian evaluation will soon be implemented in PFIM to allow an easy prediction of shrinkage when planning pharmacokinetic studies. It opens up perspectives for better design and more relevant estimates of the individual pharmacokinetic parameters from population analysis.

ACKNOWLEDGMENTS AND DISCLOSURES

During this work, the PhD of F. Combes was sponsored by a Convention Industrielle de Formation par la Recherche (CIFRE) from the French government and the Institut Roche de Recherche et Médecine Translationnelle. The

authors thank IFR02 and Dr. Hervé Le Nagard for the use of the Centre de Biomodélisation.

REFERENCES

- Pillai GC, Mentré F, Steimer J-L. Non-linear mixed effects modeling - from methodology and software development to driving implementation in drug development science. *J Pharmacokinet Pharmacodyn*. 2005;32(2):161–83.
- Beal S, Sheiner LB, Boeckmann A, Bauer RJ. NONMEM User's guides. (1989–2009), Icon development Solutions, Ellicott City, USA 2009.
- www.lixoft.com.
- Al-Banna MK, Kelman AW. Experimental design and efficient parameter estimation in population pharmacokinetics. *J Pharmacokinet Biopharm*. 1990;18(4):347–60.
- Holford NH, Kimko HC, Monteleone JP, Peck CC. Simulation of clinical trials. *Annu Rev Pharmacol Toxicol*. 2000;40:209–34.
- Atkinson AC, Donev AN. Optimum experimental designs. Oxford: Oxford University Press; 1992.
- Fisher RA. The logic of inductive inference. *J R Stat Soc*. 1935;98(1):39–54.
- Mentré F, Mallet A, Baccar D. Optimal design in random-effects regression models. *Biometrika*. 1997;84(2):429–42.
- Retout S, Mentré F, Bruno R. Fisher information matrix for non-linear mixed-effects models: evaluation and application for optimal design of enoxaparin population pharmacokinetics. *Stat Med*. 2002;21(18):2623–39.
- Bazzoli C, Retout S, Mentré F. Design evaluation and optimisation in multiple response non-linear mixed effect models: PFIM 3.0. *Comput Methods Prog Biomed*. 2010;98(1):55–65.
- Nyberg J, Ueckert S, Strömberg EA, Hennig S, Karlsson MO, Hooker AC. PopED: an extended, parallelized, nonlinear mixed effects models optimal design tool. *Comput Methods Prog Biomed*. 2012;108(2):789–805.
- Gueorguieva I, Ogungbenro K, Graham G, Glatt S, Aarons L. A program for individual and population optimal design for univariate and multivariate response pharmacokinetic-pharmacodynamic models. *Comput Methods Prog Biomed*. 2007;86(1):51–61.
- <http://www.winopt.com/>
- Foo LK, Duffull S. Adaptive optimal design for bridging studies with an application to population pharmacokinetic studies. *Pharm Res*. 2012;29(6):1530–43.
- Foo LK, Duffull S. Methods of robust design of non-linear models with an application to pharmacokinetics. *J Biopharm Stat*. 2010;20(4):886–902.
- Retout S, Mentré F. Optimization of individual and population designs using Splus. *J Pharmacokinet Pharmacodyn*. 2003;30(6):417–43.
- Merlé Y, Mentré F. Bayesian design criteria: computation, comparison, and application to a pharmacokinetic and a pharmacodynamic model. *J Pharmacokinet Biopharm*. 1995;23(1):101–25.
- Sheiner LB, Beal S, Rosenberg B, Marathe VV. Forecasting individual pharmacokinetics. *Clin Pharmacol Ther*. 1979;26(3):294–305.
- Sheiner LB, Beal SL. Bayesian individualization of pharmacokinetics: simple implementation and comparison with non-Bayesian methods. *J Pharm Sci*. 1982;71(12):1344–8.
- Fedorov F. Mixed models: design of experiments. Presented at Design and analysis of experiments, Cambridge, UK. 11th August 2011 (<http://www.newton.ac.uk/programmes/DAE/seminars/081109301.html>).
- Savic RM, Karlsson MO. Importance of shrinkage in empirical bayes estimates for diagnostics: problems and solutions. *AAPS J*. 2009;11(3):558–69.
- Mentré F, Burtin P, Merlé Y, Van Bree J, Mallet A, Steimer J-L. Sparse-sampling optimal designs in pharmacokinetics and toxicokinetics. *Drug Inf J*. 1995;29(3):997–1019.
- Frey N, Grange S, Woodworth T. Population pharmacokinetic analysis of tocilizumab in subjects with rheumatoid arthritis. *J Clin Pharmacol*. 2010;50(7):754–66.
- Verbeke G, Molenberghs G. Linear mixed models for longitudinal data. New York: Springer; 2000.
- Pilz J. Bayesian estimation and experimental design in linear regression models. Leipzig: Teubner-Texte; 1983.
- Boeckmann AJ, Sheiner LB, Beal SL. NONMEM users guide – Part V introductory guide. Ellicott City: Icon development Solutions; 2011.
- McGree JM, Eccleston JA, Duffull SB. Compound optimal design criteria for nonlinear models. *J Biopharm Stat*. 2008;18(4):646–61.
- Jones B, Wang J. Constructing optimal designs for fitting pharmacokinetic models. *Stat Comput*. 1999;9(3):209–18.
- Jones B, Wang J, Jarvis P, Byrom W. Design of cross-over trials for pharmacokinetic studies. *J Stat Plan Infer*. 1999;78(1–2):307–16.