Journal of Pharmacokinetics and Pharmacodynamics, Vol. 34, No. 3, June 2007 (© 2007)
DOI: 10.1007/s10928-006-9046-9

Evaluation of Uncertainty Parameters Estimated by Different Population PK Software and Methods

Céline Dartois,[1,2,8] Annabelle Lemenuel-Diot,[3] Christian Laveille,[4] Brigitte Tranchand,[1,2,5] Michel Tod,[1,6,7] and Pascal Girard,[1,2]

Received August 11, 2006—Accepted December 7, 2006—Published Online January 10, 2007

The uncertainty associated with parameter estimations is essential for population model building, evaluation, and simulation. Summarized by the standard error (SE), its estimation is sometimes questionable. Herein, we evaluate SEs provided by different non linear mixed-effect estimation methods associated with their estimation performances. Methods based on maximum likelihood (FO and FOCE in NONMEM^{TM}, `nlme` in Splus^{TM}, and SAEM in MONOLIX) and Bayesian theory (WinBUGS) were evaluated on datasets obtained by simulations of a one-compartment PK model using 9 different designs. Bootstrap techniques were applied to FO, FOCE, and `nlme`*. We compared SE estimations, parameter estimations, convergence, and computation time. Regarding SE estimations, methods provided concordant results for fixed effects. On random effects, SAEM and WinBUGS, tended respectively to under or over-estimate them. With sparse data, FO provided biased estimations of SE and discordant results between bootstrapped and original datasets. Regarding parameter estimations, FO showed a systematic bias on fixed and random effects. WinBUGS provided biased estimations, but only with sparse data. SAEM and WinBUGS converged systematically while FOCE failed in half of the cases. Applying bootstrap with FOCE yielded CPU times too large for routine application and bootstrap with* `nlme` *resulted in frequent crashes. In conclusion, FO provided bias on parameter estimations and on SE estimations of random effects. Methods like FOCE provided unbiased results but convergence was the biggest issue. Bootstrap did not improve SEs for FOCE methods, except when confidence interval of random effects is needed. WinBUGS gave consistent results but required long computation times. SAEM was in-between, showing few under-estimated SE but unbiased parameter estimations.*

KEY WORDS: Uncertainty; Standard error; non-linear mixed effect model; pharmacokinetics; estimation method.

[1]Université de Lyon, Lyon, F-69003, France.
[2]Université Lyon 1, EA3738, CTO, Faculté de Médecine Lyon Sud, Oullins, F-69600, France.
[3]Servier Research Group, Courbevoie, France.
[4]Exprimo NV, Lummen, Belgium.
[5]Centre Léon Bérard, Lyon, F-69008, France.
[6]Hôpital Cochin, Paris, France.
[7]Université de Lyon, Lyon, F-69003, France;ISPB, Université Lyon1, Lyon, F-69008, France.
[8]To whom correspondence should be addressed. E-mail: celine.dartois@adm.univ-lyon1.fr.

INTRODUCTION

Non-linear mixed effect models, also referred to as hierarchical or population non-linear models in clinical pharmacology, have gained broad acceptance as a suitable framework for pharmacokinetics (PK) and pharmacodynamics (PD) analysis, particularly where repeated measurements are performed from a sample of individuals in a population of interest. According to the objectives of this study, these can be performed either for descriptive purposes or for predictive ones. Quantifying and explaining dose-concentration and drug disease models using patho-physiological and genetic characteristics (termed covariates) are the primary objective. Secondary objectives include obtaining correct predictions for choosing first dose in man after animal studies, estimating maximum tolerated dose, and establishing doses for further phase II studies or to be tested in phase III confirmatory studies (1,2). Then, in clinical practice, after drug approval, the models are also used for therapeutic drug monitoring and dosage individualization (3–6).

These models are complex, not only structurally (due to non linearities, multiple outputs, different time scales) but also statistically, with different levels of randomness (between and within patient variabilities) (2,7). Potential imbalance and sparseness of data as well as non-optimal design add other difficulties (8,9). Given that the use of these models may also have large consequences on patients' health and drug development in case of model inadequacy or lack of precision, these models require rigorous validation. This validation, sometimes referred as qualification or evaluation of the model (10–13), can be of basic internal (i.e. goodness of fit, study of uncertainty, sensitivity and robustness), advanced internal (i.e. data splitting, cross-validation, bootstrap and Monte–Carlo simulations), or external validation type (Brendel, PAGE, 2006). One of the main criteria used in the basic internal validation (the most often performed), is the uncertainty associated with parameter estimations (fixed effect parameter uncertainty but also random effect one). This uncertainty is also largely used in model building to detect over parameterization and it is very crucial for simulations which are performed with the model. Ideally, this uncertainty should be quantified by the joint posterior distribution of all parameters, allowing the building of confidence intervals or for performing hypothesis testing. However, most of the time, this uncertainty is simply summarized by standard errors (SE) and correlation coefficients. Those SEs are frequently derived from an approximate of the inverse of Fisher Information Matrix (FIM) (9,14,15), which is sometimes difficult to estimate, particularly when the convergence toward maximum likelihood estimate is weak (16). Even when FIM is obtained, some researchers have

expressed serious doubts about the value of derived SE estimates, especially for random effects. Some of them have alternatively suggested using bootstrap estimates in order to assess parameter precision from the final model (17–20).

In this context, the primary objective of our work consisted of studying SE estimations. We tested performance of different estimation methods based on maximum likelihood (ML) and Bayesian theory. This comparison was performed on simulated datasets of a typical non-linear mixed effect PK model from different optimal designs. As SE estimations are largely influenced by biases of parameter estimates through the FIM, our second objective was also to compare the quality of parameter estimations. Finally, those comparisons were put into perspective with the computation time needed to obtain the SE as well as "convergence" success rates of the different methods. It appears that the methods giving the best description of uncertainty are also the ones requiring the largest amount of computation time. Similarly, as has been experienced in the past, although poorly documented, the approximate likelihood methods are highly sensitive to initial estimates and sometimes, the optimization algorithm that maximizes the likelihood fails to converge.

The methods selected for this comparison were first order (FO), first order conditional estimate (FOCE) implemented in NONMEM^{TM} (21) and first order conditional estimate implemented in `nlme` from Splus^{TM} (22) where an approximation of the likelihood is performed, a method based on "exact" likelihood using stochastic approximation EM algorithm (SAEM) (23,24) and a full Bayesian method provided in WinBUGS (25). We also performed a non-parametric (wild) bootstrap for approximated ML methods (FO, FOCE and `nlme`) (26) but only for comparison of SE estimation.

The paper is organised as follows: in material and methods, we briefly introduce the different estimation methods we compared and present the simulations used to compare these methods. Then, we explain more precisely how we performed estimations and compared methods based on SE and parameter estimations. Result part includes comparisons of SE estimations, then parameter estimations, and finally convergence and computation time.

MATERIALS AND METHODS

Estimation Methods

This section briefly introduces maximum likelihood (ML) and Bayesian estimation methods for non-linear mixed effect models. It puts more

emphasize on SAEM, which is among the most recent ML based methods. The way the core elements of this paper, i.e. the SEs are computed for those different methods, and how their convergence is assessed is described in more details.

Maximum likelihood Estimation Methods

ML corresponds to the set of parameter values that maximizes the probability of observing the data given the model. This likelihood is a multi-dimensional integral which for non-linear mixed effects models such as PK or PD analyses, cannot be evaluated analytically (22,27,28). One solution to this issue is approximating the likelihood by linearization. The historical first approximation is FO (6) but it presents severe limitations. It leads to substantial bias, especially when there is large inter-individual variability. To overcome those limitations, the First Order Conditional Estimation (FOCE) method was developed. It uses the same linearization, but replaces the expected value (zero) of the random effects (difference between typical average and individual patient parameters), by the conditional modes of individual random effects (29). Different versions of FOCE algorithm are implemented in NONMEM^{TM} (21) and `nlme` function in Splus^{TM} and R (22).

An alternative solution to approximate likelihood is to use an exact ML computation and approximate the multiple integral using a stochastic approximation (SAEM) (23). The SAEM procedure is based on Expectation-Maximization (EM) algorithm and consists, at each iteration, in successively simulating the random effects with the conditional distribution (E step), and updating the unknown parameters of the model (M step) (23). When this algorithm is coupled with a Markov Chain Monte–Carlo (MCMC) procedure for the E step, the convergence of the algorithm toward the ML estimate is established (24). This algorithm is available in MONOLIX software at http://www.math.u-psud.fr/~lavielle/monolix.

In ML theory, the expected values of SEs for fixed and random effect parameters are derived from FIM and are more precisely computed as the square root of the diagonal elements of the inverse of this matrix. According to the Rao–Cramer inequality using hypothesis of unbiased estimated parameter, these values are the lower bound of the SEs of parameters estimation (30,31). FIM is a key concept in the theory of statistical inference. Let assume a model with a vector of parameters $\theta = (\theta_1, \ldots, \theta_k)$, and random sample Y with probability density function $p(Y\ |\theta)$, and let $l_Y(\theta)$, $l'_Y(\theta)$, and $l''_Y(\theta)$ be the log-likelihood, first and second derivatives of it, respectively.

The FIM of sample size n is given by the $k \times k$ symmetric matrix, whose ijth element can be computed either by the covariance between first partial derivatives of the log-likelihood, or using the expected values of the second partial derivatives:

$$\mathrm{FIM}(\theta)_{ij} = \mathrm{Cov}_{\theta}\left[\frac{\partial \mathrm{l}_{\mathrm{Y}}(\theta)}{\partial \theta_i}, \frac{\partial \mathrm{l}_{\mathrm{Y}}(\theta)}{\partial \theta_j}\right] \tag{1}$$

$$\mathrm{FIM}(\theta)_{ij} = -E_{\theta}\left[\frac{\partial^2 \mathrm{l}_{\mathrm{Y}}(\theta)}{\partial \theta_i \, \partial \theta_j}\right] \tag{2}$$

Strictly, the latter definition corresponds to the *expected* FIM. If no expectation is made, a data-dependent quantity is obtained that is called the *observed* FIM. Given the expression of the information, the log-likelihood $\mathrm{l}_{\mathrm{Y}}(\theta)$, is needed in order to compute the *observed* FIM. Since the log-likelihood exact form is not available with FO and FOCE methods, it is replaced by the approximate linearized one for computing what is called an *empirical* FIM. In NONMEM$^{\mathrm{TM}}$, there are two matrices corresponding to Eqs. (1) and (2) that are computed in the covariance step as intermediate steps toward computing the covariance matrix. The first matrix, R, corresponds to the Hessian, i.e. second derivatives of log-likelihood evaluated at the final estimate of the model parameter. The second matrix, S, corresponds to the sum of matrices S_i, with one matrix for each individual i. Each matrix S_i is the cross-product $\nabla_i \nabla_i^t$, where ∇_i is the vector of the first derivatives of log-likelihood of the contribution to the objective function from the ith individual (gradient in column), evaluated at the final estimate of the model parameter. As noted in the NONMEM$^{\mathrm{TM}}$ guide: "Under the assumptions that all random effects are normally distributed, the R and S matrices, each divided by the number, N, of individuals in the sample, tend to the same matrix as N increases. In this case the inverse of either matrix serves as an estimate of the true covariance matrix (32). However when normality cannot be assumed, Beal suggests a second way to estimate the asymptotic covariance matrix with the matrix product $R^{-1}SR^{-1}$. This idea, introduced by White in early eighties (33), is the default implementation in NONMEM$^{\mathrm{TM}}$ and is sometimes called the *sandwich estimator*. In `nlme`, empirical information matrix is computed after convergence according to Eq. (2), without the expectation operator, and corresponds to the Hessian (22). SAEM algorithm uses the same theory to derive the variance of the estimates but takes advantage that, thanks to the ML estimator obtained with the method, it is possible to compute simultaneously an estimation of the FIM. It uses the fact that the gradient and the Hessian of the likelihood can be obtained almost

directly from the simulated missing random effects by taking advantage of Louis' missing information principle (34). Once the observed *empirical* FIM has been obtained, software performs the adequate matrix inversion to get the asymptotic variance covariance matrix and derives the SEs from its diagonal. Regardless of the estimation method used, it is worth noting that NONMEM^{TM} provides standard errors, and correlations, for all fixed and random effect parameters, while `nlme` provides them only for fixed effect parameters and SAEM for all parameters except covariances.

In case of approximate likelihood methods and, in certain circumstances where the Hessian and/or the gradient are not computable matrices, software such as NONMEM^{TM} or function `nlme` cannot provide the asymptotic standard error of the estimates. Estimated parameters obtained after convergence can also be biased. In such circumstances, Rao–Cramer inequality, which states the FIM is the lower boundary of the standard errors of parameters estimation, can no longer be applied. In this context, it is possible to use an adaptation of the non-parametric bootstrap technique to population PK–PD models (26). Initially developed for simple (non-linear) statistical models where the experimental unit is a single observation (35), the bootstrap method has been extended to non-linear mixed effect models by (i) taking at random and with replacement complete random effect vectors, corresponding to an individual, then (ii) fitting the non-linear mixed effect model to the bootstrapped dataset and (iii) computing the bootstrap statistic or using the bootstrap distribution (16,26,36). Since this bootstrap technique has never been properly evaluated in terms of convergence properties, it is sometimes called wild bootstrap by analogy with another bootstrap technique used for heteroscedastic variances (37,38). We used this additional technique to estimate the SEs on all replicated simulated datasets from bootstrapped datasets where individuals were resampled with replacement. We finally expressed the SEs as the standard deviation of the bootstrap parameters.

Since all those estimation methods are based either on an optimisation process or on the achievement of some stationary property, we considered different ways for defining a "successful convergence". For NONMEM^{TM}, obtention of formal flag (`MINIMIZATION SUCCESSFUL`) combined with a successful covariance step was considered as a success. For `nlme` successful estimation of both parameter estimates and SEs was considered as a success. For SAEM, convergence was defined graphically when all parameters reached their stationary distributions. Concerning bootstrap applied to NONMEM^{TM} or `nlme`, only a success for optimization step was required.

Bayesian Estimation Method

The hierarchical structure of the non-linear mixed effects model makes it a natural candidate for Bayesian inference (39). For the present work, WinBUGS software, that performs Bayesian inference using Gibbs sampling, was chosen (25). It characterises the Bayesian hierarchical model in three stages: within-subject variability, between-subject variability and prior distribution. In our study, all these stages were defined by assuming a very flat and uninformative prior distribution. The hyperprior distribution of the population parameters was specified in accordance with initial parameter estimates used in the FOCE method. Convergence was assumed by using the method originally proposed by Gelman and Rubin (40) and subsequently modified (41). We simulated three chains starting at "overdispersed" initial values (I_0, $I_0\backslash 2$ or $I_0 \times 2$) and we compared within and between chain variabilities by the Brooks and Gelman ratio. When this value was close to 1 and the trace plots of population parameters were stable, convergence was assumed. For this study 50,000 iterations were required for "convergence" success, but only last 2000 were used for inferences as the posterior distribution. It typically involves all parameters and, therefore, is considered as a joint distribution (2,42–45). SEs were estimated by standard deviation (SD) of relevant parameter posterior distributions.

Simulations

Simulations were performed with population PK database of theophyllin provided in the NONMEM software package (46). A one-compartment model with first-order absorption, first-order elimination and bioavailability set to one describes this dataset. The vector of parameter of the ith individual includes the elimination constant, $k_{\mathrm{ei}}(h^{-1})$, the volume of distribution, $V_{\mathrm{i}}(\mathrm{l})$, and a hybrid parameter $k_{\mathrm{di}}(h^{-1})$ which is derived from the absorption constant ($k_{\mathrm{ai}} = k_{\mathrm{ei}} + k_{\mathrm{di}}, k_{\mathrm{di}} > 0$) in order to avoid flip–flop:

$$C_{ij} = D \cdot V_i \frac{k_{\mathrm{di}} + k_{\mathrm{ei}}}{k_{\mathrm{di}}} \cdot \left(\mathrm{e}^{-k_{\mathrm{ei}} \cdot t_{ij}} - \mathrm{e}^{-(k_{\mathrm{di}} + k_{\mathrm{ei}}) \cdot t_{ij}} \right) + \varepsilon_{ij} \tag{3}$$

where C_{ij} is the concentration of ith individual at time t_{ij}, D the dose and $\varepsilon_{ij} \sim N(0, \sigma^2)$ the random residual variability.

Inter-individual variability (IIV) was assumed to follow a log-normal statistical model:

$$\theta_{\mathrm{i}} = \theta \cdot \mathrm{e}^{\eta_i} \tag{4}$$

with individual random effects η_i resampled from a k-dimensional multivariate normal distribution $MVN_k(0, \Omega)$, with Ω a full covariance matrix ($k \times k$) matrix, including variances $\omega^2_V, \omega^2_{ke}$ and ω^2_{kd}, and covariances $\omega_{Vke}, \omega_{Vkd}$ and ω_{kekd}. There were therefore 10 population parameters which described this pharmacokinetic model: $V, k_e, k_d, \omega^2_V, \omega^2_{ke}, \omega^2_{kd}, \omega_{Vke}$, $\omega_{Vkd}, \omega_{kekd}$ and σ^2.

In order to compare the different estimation methods without the burden of sparse incomplete design, we chose to simulate the data according to optimal designs (31,47). Three different numbers of sampling times ($n = 3$, 6, and 15) were chosen optimally in POPOS^{TM} 1.0 software (14). Each of these sampling schedules was combined with 3 different numbers of subjects, ($N = 30$, 100 or 500) which produced 9 different designs. Each of those designs was simulated 100 times, totalling 900 simulated datasets.

NONMEM^{TM} subroutine ADVAN2 was used to simulate the data (46). Fixed effect simulation parameters were set respectively to 31.71, $0.0873\,h^{-1}$, $1.48\,h^{-1}$ for V, k_e and k_d,. Random effect parameters were 0.0157, 0.016, 0.471 for $\omega^2_V, \omega^2_{ke}, \omega^2_{kd}$, respectively, and 0.0143, 0.0322 and -0.00498 for the three non-diagonal covariance terms $\omega_{Vke}, \omega_{Vkd}$ and ω_{kekd}. Finally σ^2, the additive residual variance, was set to 0.479 mg/l. With those parameters, few ($n < 0.08\%$) concentrations were simulated as negative values and were removed from the simulated data sets.

Comparison

A Splus^{TM} script was written to drive in batch mode all model estimation processes performed in NONMEM^{TM} FO and FOCE (+bootstrap), in `nlme` (+bootstrap), and in WinBUGS. For SAEM, the batch process was performed through a Matlab® script. For NONMEM^{TM}, the estimation model was identical to the simulation one. Convergence assessment was recorded in a variable at each run. For `nlme`, the structural model was identical to the simulation one, but model parameters were log-transformed. It ensured non-negative parameter estimation, since `nlme` does not offer the option to define any parameter boundaries, and implementation of the log normal distribution of random effects η_i. Inter-individual variability was classically implemented: η_i were assumed to follow a k-dimensional multivariate normal distribution $MVN_k(0, \Omega)$, Ω being a general positive-definite and non-diagonal ($k \times k$) matrix. The `nlme` function was launched automatically in S-plus^{TM} and controlled by the function `try( )`, to recover from function crashes and record the convergence success rate. Initial values of population parameters were chosen close to the logarithm of those employed for NONMEM^{TM} FOCE ($\ln(\theta_0)$). Model parameters were also log transformed in SAEM before a

blinded launching. In WinBUGS initial values of log transformed parameters, $\ln(\theta_i)$ and $\ln(\theta)$ were also chosen close to the logarithm of θ_0 used in FOCE. Moreover, the initial value for τ was chosen close to the inverse of σ_0^2 and the one for Ω^{-1} was chosen close to the inverse of Ω_0, σ_0^2 and Ω_0 being initial values used in FOCE. At convergence, this software allowed one to directly obtain inferences of no log-transformed parameters.

Results were automatically processed in Splus^{TM}. For FO, FOCE and WinBUGS, no parameter transformation was needed. For `nlme` and SAEM, logarithms of fixed effects parameterized θ_L were transformed back to express them in the same metric as others methods (θ):

$$\theta \approx \exp(\theta_L) \tag{5}$$

$$SE(\theta) \approx SE(\theta_L) * \exp(\theta_L) \tag{6}$$

SE estimation accuracy was evaluated independently for each method and each design. We used empirical SEs, which were considered as the "reference" SE values (SE_{ref}). They were estimated as the SD of the 100 parameter estimates obtained for each method (FO, FOCE, `nlme`, SAEM and Winbugs) and each design. Estimated SEs (from these methods with or without bootstrap) were then compared to SE_{ref} corresponding by calculating the ratios SE/SE_{ref}. Those ratios were summarized by median, 25th and 75th percentiles and presented as star plots, with one star per combination of design and parameter, with each radius of the star representing one method. Those multidimensional plots allowed for easy comparison of all methods across designs. Estimation dispersion was expressed by inter-quartile ranges across designs and methods. Raw SE estimates, were compared between methods utilizing the same type of star plots described above, with one star per combination of design and parameter.

Parameter estimation accuracy was evaluated by relative biases:

$$bias(\theta) = (\theta - \theta_{true})/(\theta_{true}) \cdot 100 \tag{7}$$

where θ were parameter estimations, and θ_{true} were "true" values of parameters. A boxplot multipanel was chosen to represent the biases and to easily compare them (values of parameter across designs and methods on a same plot). Estimation dispersion was evaluated by computing relative Root Mean Square Error (RMSE):

$$RMSE = 100 * \sqrt{\text{mean}\left((\theta - \theta_{true})^2\right)/\theta_{true}^2} \tag{8}$$

Analyses of parameter estimations were not performed for bootstrap methods as they were only used here, for evaluating their performance in SE estimations.

RESULTS

SE Evaluation

Raw SE estimates were compared across all the designs. Figure 1 illustrates for FOCE method and median SE volume how this estimate decreases when the number of subjects or sampling times per subject increase. Overall, raw SE estimates presented consistent results across all methods for fixed effects (e.g. Fig. 2, Elimination constant). For random effects like inter-individual variability of the absorption hybrid rate constant or inter-individual variability of the volume, with 3 or 6 samples per patient, SE values decreased, as expected, when the number of patients increased, except for FO (Fig. 2). FO method, on original or bootstraped data sets, showed marked different estimates for random effects SEs, which are either smaller or larger than the ones of other methods.

SE_{ref} values have been obtained for all designs, all methods and all parameters but only results for one design, P6S100 (6 points per subject and 100 subjects) are presented in Table I. It shows that all the methods present approximately same efficiency (precision) across the 100 runs, except FO. On random effects, this method presents large differences, compared to the others: SE_{ref} computed from FO estimates are either much larger (interindividual variability of elimination rate constant k) or much smaller (interindividual variability of absorption hybrid rate constant k_d).

Concerning the ratio SE/SE_{ref}, main results can also be characterized by design P6S100. For this design, these ratios for fixed effects

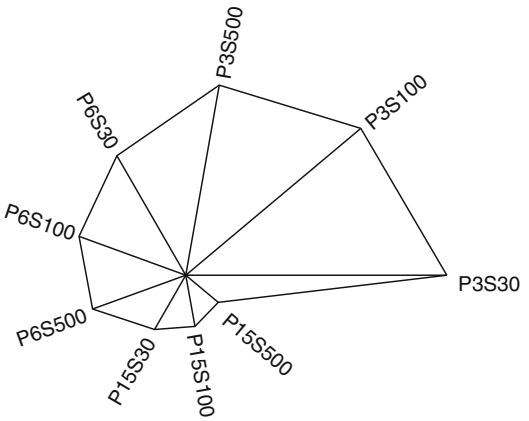

Fig. 1. Stars of median SE estimated for all designs, FOCE method and the volume (number of sampling per subject (P) = 3, 6 or 15 and number of subjects (S) = 30, 100 or 500). The length of each radius is the median of approximately 100 estimations.

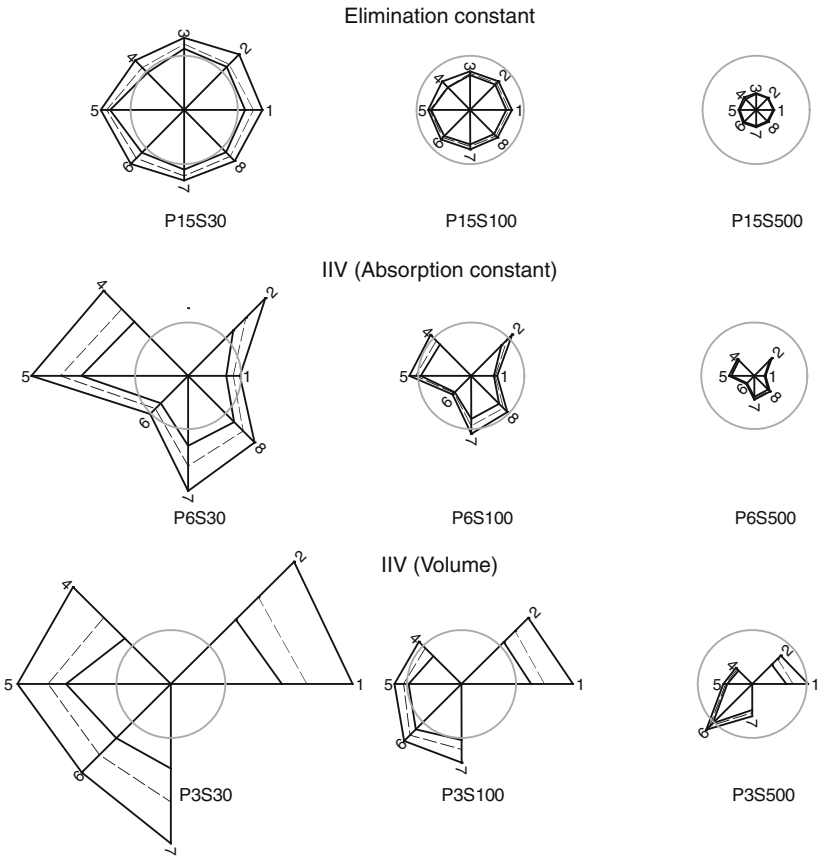

Fig. 2. Stars of raw SE values of 9 different designs (number of sampling per subject (P) = 3, 5, or 15 and number of subjects (S) = 30, 100 or 500) for one fixed effect (Elimination constant) and two random effects (IIV of the absorption hybrid rate constant and the volume). The length of each radius is based on approximately 100 estimations. Medians are linked by a dotted line and quartiles by continuous lines. All methods are represented: 1: FO, 2: FOCE, 3: `nlme`, 4: SAEM, 5: WinBUGS, 6: bootstrap and FO, 7: bootstrap and FOCE , 8: bootstrap and `nlme`. The reference circle is equivalent to estimation CV of 2.5% for Elimination constant, 18% for IIV of absorption hybrid rate constant, and 27% for IIV of volume.

parameters, were consistent across all the methods (in terms of value and variability, Fig. 3) but significantly inferior to one, for absorption hybrid rate constant. SAEM also showed a ratio slightly lower than one. For random effects, ratios were consistent with the value of one for all methods, except for WinBUGS, where it was found larger, and SAEM, (when available) where it was smaller than one (Fig. 3). For random and fixed effects,

Table I. SE_{ref} function of estimation methods on fixed and random effecs for design P6S100[a]

Methods	N	V(l)	$k(h^{-1})$	$k_d(h^{-1})$	$\omega^2 V$	$\omega V - k$
FO	97	0.51	0.0018	0.12	0.004	0.004
FOCE	57	0.55	0.0018	0.13	0.003	0.003
Nlme	93	0.51	0.0020	0.12	0.003	0.003
SAEM	100	0.53	0.0021	0.13	0.003	0.003
Winbugs	100	0.51	0.0019	0.13	0.003	0.002
Methods	N	$\omega V - k_d$	$\omega^2 k$	$\omega k - k_d$	$\omega^2 k_d$	σ^2
FO	97	0.009	0.010	0.014	0.040	0.039
FOCE	57	0.013	0.005	0.016	0.083	0.030
Nlme	93	0.013	0.005	0.015	0.075	0.031
SAEM	100	0.013	0.005	0.017	0.088	0.031
Winbugs	100	0.011	0.004	0.013	0.080	0.031

SE estimations from bootstrap and non-bootstrap methods gave similar results. For covariance parameters, WinBUGS once again showed ratios larger than one. And, finally, all methods provided a ratio close to one for residual variance SE. These results were concordant with those from other designs except for sparse data (3 points per subject) and random effects. In this case, FO presented inconsistent results with overestimated results when the number of subjects was 30 and underestimated results when number of subjects was 500 (results not shown).

According to interquartile ranges and for the design P6S100, WinBUGS presented the less dispersed results (Table II). FO presented non consistent results with low or high interquartile ranges depending of the parameter. This tendency was increased with sparse datasets (results not shown).

Parameter Evaluation

Parameter estimations provided by FO are clearly biased for fixed and random effects (Fig. 4). For random effects and sparse data (sampling times = 3), WinBUGS also presented some biases. Nevertheless, only FO showed a consistent bias across designs and for all random effects (for example, as with $\omega^2 k_d$). RMSE calculations showed that FO afforded the lowest precision (see Table III) whether the parameter was fixed or random as compared with other methods which presented equivalent results.

Convergence and CPU

A summary of the results for convergence and CPU are presented in Table IV. Firstly, `nlme` did not produce any results for sparse designs with

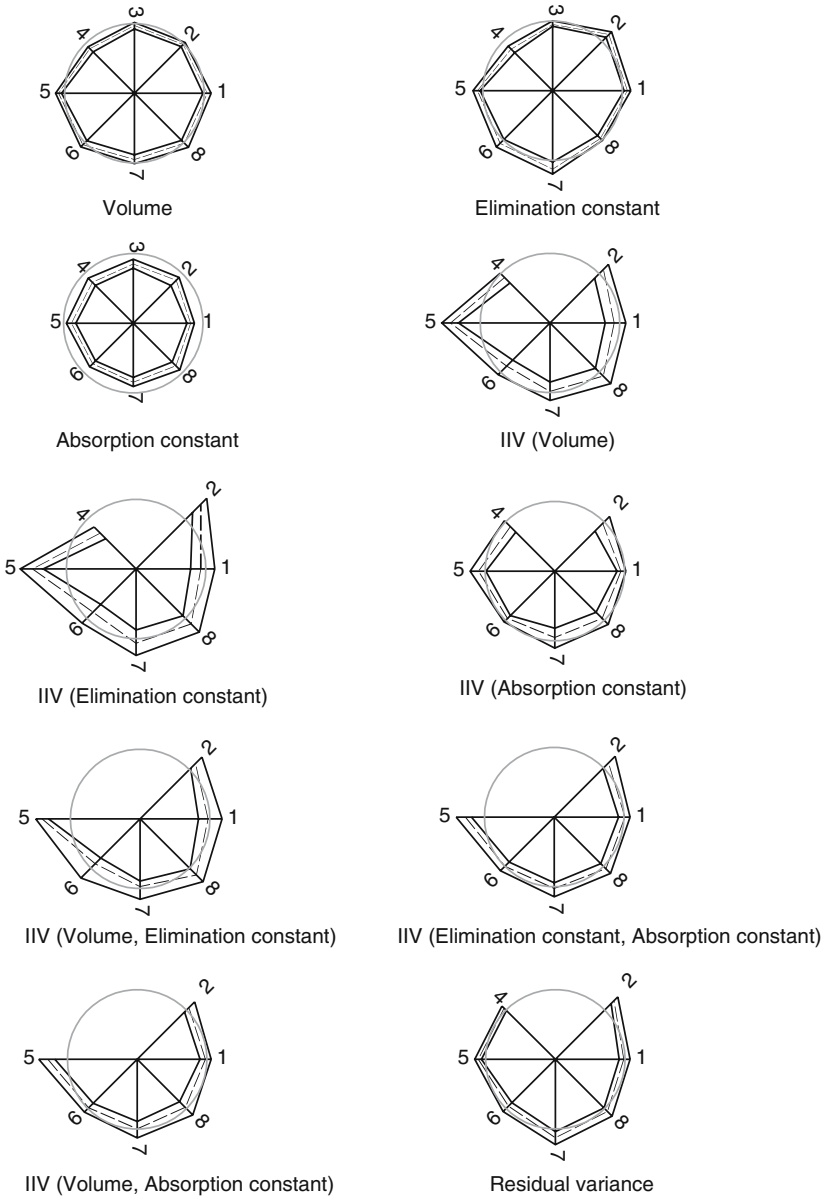

Fig. 3. Stars of ratio SE/SE_{ref} for the design P6S100 (number of sampling per subject (P) = 6 and number of subjects (S) = 100). The length of each radius is based on approximately 100 estimations. Medians are linked by a dotted line and quartiles by continuous lines. All methods are represented: 1: FO, 2: FOCE, 3: `nlme`, 4: SAEM, 5: WinBUGS, 6: bootstrap and FO, 7: bootstrap and FOCE , 8: bootstrap and `nlme`. The reference circle is equivalent to a ratio of 1.

Table II. Dispersion (interquartile range) of estimated SE on fixed and random effects for design P6S100[a]

Methods	N	V(l)	k(h^{-1})	k_d(h^{-1})	$\omega^2 V$	$\omega V - k$
FO	97	0.061	0.00018	0.013	0.00122	0.00151
BOOTFO	96	0.064	0.00028	0.014	0.00121	0.00181
FOCE	57	0.045	0.00019	0.021	0.00095	0.00069
BOOTFOCE	85	0.070	0.00031	0.017	0.00089	0.00078
nlme	93	0.047	0.00016	0.015	NA	NA
BOOTnlme	95	0.071	0.00020	0.019	0.00092	0.00070
SAEM	100	0.057	0.00020	0.018	0.00073	NA
WINBUGS	100	0.044	0.00020	0.018	0.00062	0.00044
Methods	N	$\omega V - k_d$	$\omega^2 k$	$\omega k - k_d$	$\omega^2 k_d$	σ^2
FO	97	0.0015	0.0035	0.0023	0.0051	0.0060
BOOTFO	96	0.0017	0.0038	0.0025	0.0051	0.0070
FOCE	57	0.0027	0.0015	0.0039	0.0243	0.0041
BOOTFOCE	85	0.0031	0.0020	0.0032	0.0236	0.0055
nlme	93	NA	NA	NA	NA	NA
BOOTnlme	95	0.0034	0.0016	0.0029	0.0184	0.0051
SAEM	100	NA	0.0015	NA	0.0212	0.0027
WINBUGS	100	0.0026	0.0014	0.0029	0.0180	0.0030

[a] Results of N estimations.
NA not available.

Table III. Precision (RMSE in %) of estimated fixed and random effects for design P6S100[a]

Methods	N	V(l)	k(h^{-1})	k_d(h^{-1})	$\omega^2 V$	$\omega V - k$	$\omega V - k_d$	ω^2k	$\omega k - k_d$	$\omega^2 k_d$	σ^2
FO	97	6.0	8.1	16	34	55	92	103	542	41	8.2
FOCE	57	1.9	2.4	8.8	22	23	41	36	321	18	7.3
nlme	93	1.6	2.3	8.1	19	19	42	32	333	19	6.8
SAEM	100	1.7	2.3	8.8	21	23	42	39	344	19	7.5
WINBUGS	100	1.7	2.2	9.0	32	22	35	57	268	20	7.9

[a] Results of N estimations.
NA not available.

three samples per subject. For richer designs, the convergence success was achieved for more than 90% of the runs. With respect to NONMEM^{TM} FO, convergence was achieved for almost all runs as compared to FOCE, where it was achieved for only 56% of the runs. Most of the time, the failure was due to the abortion of the covariance step, as shown by the success rate for bootstrap with NONMEM^{TM} FOCE where no covariance step was required, and where the success rate was 85%. FO was the fastest method, followed by `nlme`, SAEM and FOCE, requiring about 10 times

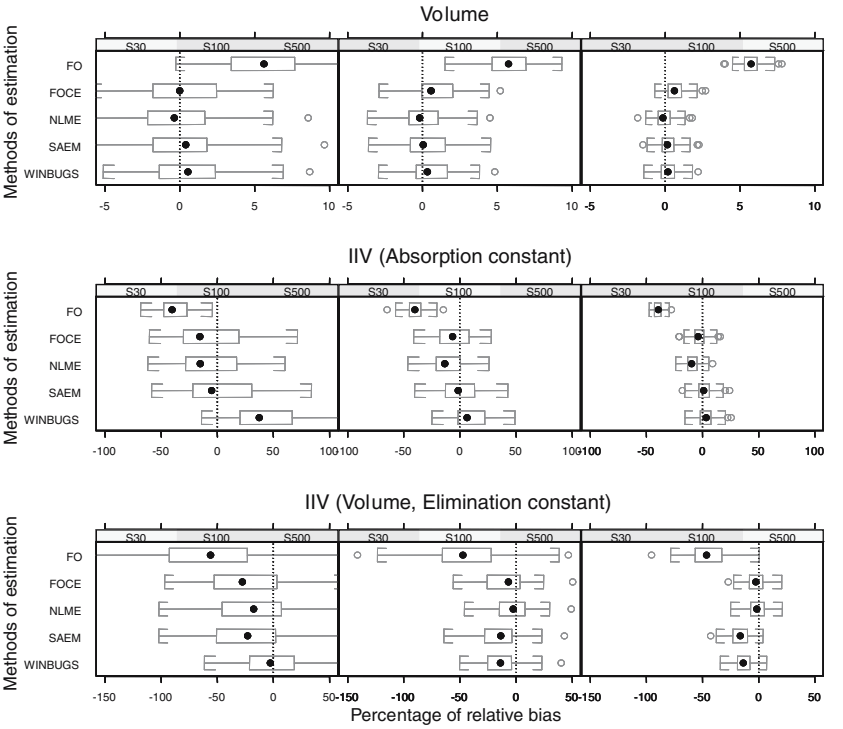

Fig. 4. Box plot of relative biases for 3 parameters and different designs (number of sampling per subject $P = 3$ and number of subjects $S = 30$, 100 or 500): fixed effect parameter volume; random effect parameters variability of absorption hybrid rate constant and covariance between volume and elimination constant.

more than FO (Table IV). WinBUGS was the slowest method, albeit much faster than bootstrap applied to NONMEM^{TM} FOCE or `nlme`.

DISCUSSION

The main purpose of this study was to evaluate SEs estimated by different methods. The secondary purpose was to compare their parameter estimation performances. A simple and usual PK model, reflecting many models used in practice, was chosen. Despite its simplicity, this nonlinear model should allow one to distinguish between methods. All simulated experimental designs were optimized (14) so that SEs should be minimal and algorithms should not have any difficulty of convergence due to identifiability problems. The chosen algorithms to be compared were all

Table IV. Method convergence and computation time

Methods	Mean % of success	Computation time for one dataset[d]
FO	97[a]	1
FOCE	56[a]	10
nlme (S-plus)	92[b]	3
SAEM	100	5
WinBUGS	100	189
BOOTFO	98[c]	96
BOOTFOCE	85[c]	936
BOOTnlme	91[b]	NA

NA not available.
[a] Convergence and $COVcovariance achieved.
[b] Only for designs where sampling time number $\geqslant 6$.
[c] Convergence achieved.
[d] Ratio of (computation time of a method/FO computation time).

"routine" ones, except for the recent stochastic algorithm, SAEM, implemented in Monolix software which is still under developments (24).

At first, we evaluated SE estimation accuracy of the methods. For this purpose, we needed some "reference" value for SEs. POPOS^{TM}, the software used to optimize sampling times, computes SE values, for a given design, based on FIM. This calculation is based on a linearization of the PK model similar to approximate ML methods and might be inaccurate for highly non-linear models (48). Hooker *et al.*, demonstrated that asymptotic FIM may not be reliable to calculate SEs of non-linear mixed effect model parameter estimates, and should only be used to predict trends of SE across different designs (Hooker, PAGE, 2004). Other authors have proposed to compute reference SE from standard deviation of several estimations from a given experimental design (49, 50). In fact, this standard deviation should asymptotically be the true SE if the true analysis model is used and the exact likelihood is computed (51). We simulated therefore 100 datasets for each design. Reference SEs for, each of our tested method, was considered to be the standard deviation for the 100 estimates of this method. It should be emphasized that this is not an absolute reference, but only the reference for a given method.

As exemplified by FOCE method, the decreasing trend of estimated SE with increasing available information was recovered as expected (Fig. 1). Evaluation of method efficiency by the calculation of SE_{ref} allowed to distinguish FO by its incoherent results on random effects. Evaluation of SE estimation accuracy by the ratio SE/SE_{ref} gave consistent results across all designs, for most parameters. For K_d (absorption hybrid rate

constant), the ratio SE/SE_{ref} was estimated systematically to be less than one, which may be linked to the usual difficulty in estimating this parameter (Fig. 3) (52). Two methods gave slightly different results, SAEM and WinBUGS, which showed a trend to respectively under or over-estimate some of the SEs (Fig. 3). SAEM under-estimation has no apparent explanation except that it bases its SE computation on a diagonal inter-individual variability matrix. This work allowed one to highlight that simplification could slightly influence SE estimations. This method is still a work in progress and improvements continue to be made. Concerning WinBUGS, SEs were deduced from the posterior distributions of the parameters. As our model was non-linear in the majority of cases, distributions of random effects were asymmetrical (plots not shown). SEs were also over-estimated since they were considered as the standard deviation of these distributions. For this method and in this context, SE should not be used in place of posterior distribution. Confidence interval (CI) would be more appropriate to take into account this eventual asymmetry as exemplified in Fig. 5 for one replication and one design. As for bootstrap, SE represents a large reduction of information and biased estimation of uncertainty, especially when posterior distributions are not multivariate normal. To illustrate this we present confidence CI calculation for one design, P6S30 and one replication. The design was chosen so that the number of observations would be minimum, with SEs obtainable for all methods, including `nlme`. For FO, FOCE, `nlme` and SAEM, we obtained 95% CI by using the normality assumption. We also performed FO and FOCE estimations on log parameters, to take into account possible asymmetrical distributions of thetas (LOGFO and LOGFOCE). For all bootstraps, 1000 samplings of the dataset were performed, runs were fitted, and the quantiles 2.5% and 97.5% were calculated. For Winbugs, we took quantiles 2.5% and 97.5% of the last 2000 iterations. Although asymmetry of CI was recovered for randon effect parameters (Fig. 5), this tendency to an over-estimation of SE was not systematically observed for bootstrap estimates of SE, and most of the times bootstrap gave very similar CI for fixed effect as compared to the ones of FIM^{-1}.

Concerning SE dispersion, WinBUGS tended to have less dispersed values of SE (Figs. 2, 3 and Table II) regardless of the design, and then appeared to be the most consistent method while FO is inconsistent and seemed to be not reliable. Finally, concerning raw values of SEs, consistent results were observed between the different methods for fixed effects whereas SEs of random effect showed some discrepancies. We already pointed out the SE results for WinBUGS and SAEM. Another method, FO, provided discordant results between estimations of non-bootstrapped data and bootstrapped data when number of subjects increases. This was

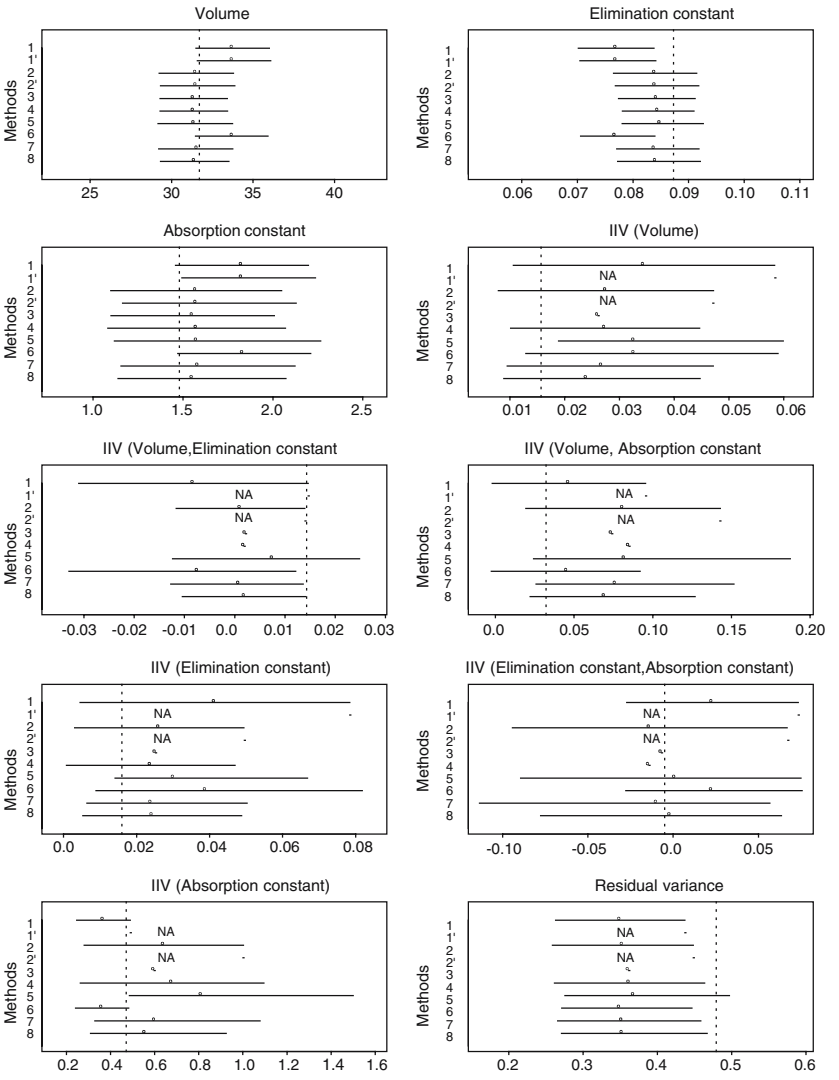

Fig. 5. 95% confidence interval of parameters for one run from design P6S30 (number of sampling per subject P = 3 and number of subjects $S = 30$). FO, FOCE, `nlme` confidence intervals were constructed assuming normal distribution of parameters, why LOGFO and LOGFOCE assumed log-normal distributions. Bootstrap and WinBUGS confidence intervals are based on the 2.5 and 97.5% quantiles of 1000 bootstrapped datasets and last 2000 iterations of Monte–Carlo Markov Chain, respectively. The point on the CI is the median; the dotted line is true value of the parameter. All methods are represented: 1: FO, 1′: LOGFO, 2: FOCE, 2′: LOFOCE, 3: `nlme`, 4: SAEM, 5: WinBUGS, 6: bootstrap and FO, 7: bootstrap and FOCE , 8: bootstrap and `nlme`.

found only for random effects and sparse data, but shows a lack of reliability for FO estimations of SE, perhaps linked to biases on these parameters.

With regard to parameter estimations, FO showed a systematic bias, toward fixed and random effects. As expected, these biases were most significant for random effects and could be as large as 50% for parameters of interest. WinBUGS with sparse data presented some biased estimations which disappeared with larger designs, but FO biases were persistent. In order to check that biases on inter-individual variability observed for WinBUGS were the result of uninformative prior distributions, new estimations were performed on the same datasets, but with informative priors. As expected, this change resulted in a large decrease of those biases, for example, a reduction from 49% to 7% for the absorption hybrid rate constant. Comparison of parameter estimation precision also allowed us to distinguish FO with the lowest precision.

Finally, we compared methods with other criteria like success and computation time. Success was achieved when convergence and covariance estimations were reached as are needed in practice. This convergence estimation was difficult to obtain for FOCE as shown in Table IV (difference between bootstrapped and non bootstrapped data). `nlme` has also some difficulties to converge and appeared to be not very usable in population PK. These convergence issues are surprising given the fact that the simulations were created with optimal population designs. In contrast, SAEM and WinBUGS were fully successful in term of "convergence". Comparison of computation times revealed that `nlme` and FOCE with bootstrapped data took too much time to be applicable in practice, but can be useful in some cases of non-convergence (16). In this case we showed that SE from FIM were in total accordance with SE obtained after bootstrap. The use of WinBUGS appeared possible only for simple model.

The purpose of our study was to compare different estimation methods used in population PK by their SE estimation, their parameter estimation, their computation time and their ability to converge. FO appeared as the less reliable methods bases on its results on SE and on parameters. On SE, this incoherence was observed for random effects, on method efficiency and on raw SE estimations with sparse data, i.e. when this method is the most used. On parameters, this method provided biased results for all parameters. This method presented therefore the highest RMSE. Use of bootstrap and FO with sparse data did not improve the SE estimations. FOCE presented difficulties either to reach convergence and/or estimate covariance matrix; `nlme` did not converge with sparse data and was laborious in other situations. Nevertheless, these methods presented good estimations of SE and parameters. Use of bootstrap did not improve the

SE estimation for fixed effect, but is probably a good solution when one is interested on confidence interval of random effect parameters, which most of the time are not normally distributed. The price to pay is extensive computation time, but also convergence difficulties with bootstrapped datasets. SAEM appeared to be a good compromise in terms of computation time and performance. Its SE estimates were slightly underevaluated, but this bias was identified and subsequently fixed (results not shown). WinBUGS was the most consistent method, with the limit that when implemented in an uninformative manner (as is done in practice), it showed some bias on random effects and required a lot of CPU time.

In conclusion, FO showed bias on parameter estimations and unreliable SE estimations on random effects. Methods like FOCE in NONMEM and `nlme` presented unbiased results and good SEs, but difficult to obtain in term of convergence despite the optimal design that was used. Bootstrap method did not improve SE estimations and appeared only useful in few cases, when SEs are not provided and confidence interval of random effects is needed. WinBUGS provided consistent results but after a long computation time. SAEM, a quickly evolving algorithm, seems to be between these two approaches (ML and Bayesian) with slightly underestimated SEs but unbiased parameter estimations.

ACKNOWLEDGMENTS

The authors would like to thank Nicole Pegon for her technical assistance with implementation of parallel computations and Marc Lavielle (Université Paris Sud, Orsay) who performed all blinded MONOLIX runs and improved the software. This work was sponsored by Servier Research Group. Pascal Girard is funded by INSERM, France.

REFERENCES

1. B. Meibohm and H. Derendorf. Pharmacokinetic/pharmacodynamic studies in drug product development. *J. Pharm. Sci.* **91**:18–31 (2002).
2. L. B. Sheiner and J. L. Steimer. Pharmacokinetic/pharmacodynamic modeling in drug development. *Annu. Rev. Pharmacol. Toxicol.* **40**:67–95 (2000).
3. W. E. Evans, M. V. Relling, J. H. Rodman, W. R. Crom, J. M. Boyett, and C. H. Pui. Conventional compared with individualized chemotherapy for childhood acute lymphoblastic leukemia. *N. Engl. J. Med.* **338**:499–505 (1998).
4. R. W. Jelliffe. Clinical applications of pharmacokinetics and adaptive control. *IEEE Trans. Biomed. Eng.* **34**:624–632 (1987).
5. R. W. Jelliffe, A. Schumitzky, G. M. Van, M. Liu, L. Hu, P. Maire, P. Gomis, X. Barbaut, and B. Tahani. Individualizing drug dosage regimens: roles of population pharmacokinetic and dynamic models, Bayesian fitting, and adaptive control. *Ther. Drug Monit.* **15**:380–393 (1993).

6. L. B. Sheiner, B. Rosenberg, and K. L. Melmon. Modelling of individual pharmacokinetics for computer-aided drug dosage. *Comput. Biomed. Res.* **5**:411–459 (1972).
7. M. O. Karlsson and L. B. Sheiner. The importance of modeling interoccasion variability in population pharmacokinetic analyses. *J. Pharmacokinet. Biopharm.* **21**:735–750 (1993).
8. S. Retout, S. Duffull, and F. Mentré. Development and implementation of the population Fisher information matrix for the evaluation of population pharmacokinetic designs. *Comput. Methods Programs Biomed.* **65**:141–151 (2001).
9. S. Retout and F. Mentré. Further developments of the Fisher information matrix in nonlinear mixed effects models with evaluation in population pharmacokinetics. *J. Biopharm. Stat.* **13**:209–227 (2003).
10. Y. Yano, S. L. Beal, and L. B. Sheiner. Evaluating pharmacokinetic/pharmacodynamic models using the posterior predictive check. *J. Pharmacokinet. Pharmacodyn.* **28**:171–192 (2001).
11. Food and Drug Administration. Guidance for industry – Population pharmacokinetics. Available from: http://www.fda.gov/cder/guidance/1852fnl.pdf. Last Accessed: July 4 2006.
12. F. Mentré and M. E. Ebelin. Validation of population pharmacokinetic/pharmacodynamic analyses: review of proposed approaches. In: *COST B1. European cooperation in the field of scientific and technical research. The population approach: measuring and managing variability in response, concentration and dose*, L. Aarons, L. P. Balant, M. Danhof, M. Gex-Fabry, U. Gundert-Remy, M. Karlsson, F. Mentre, P. Morselli, M. Rowland, J. L. Steimer, S. Vozeh, and F. Rombout, eds. pp. 148–160, Office for official publications of the European Communities, Brussels.
13. E. I. Ette, P. J. Williams, Y. H. Kim, J. R. Lane, M. J. Liu, and E. V. Capparelli. Model appropriateness and population pharmacokinetic modeling. *J. Clin. Pharmacol.* **43**:610–623 (2003).
14. M. Tod and J. M. Rocchisani. Implementation of OSPOP, an algorithm for the estimation of optimal sampling times in pharmacokinetics by the ED, EID and API criteria. *Comput. Methods Programs Biomed.* **50**:13–22 (1996).
15. M. Tod and J. M. Rocchisani. Comparison of ED, EID, and API criteria for the robust optimization of sampling times in pharmacokinetics. *J Pharmacokinet. Biopharm.* **25**:515–537 (1997).
16. I. Matthews, C. Kirkpatrick, and N. Holford. Quantitative justification for target concentration intervention–parameter variability and predictive performance using population pharmacokinetic models for aminoglycosides. *Br. J. Clin. Pharmacol.* **58**:8–19 (2004).
17. P. L. Bonate, A. Craig, P. Gaynon, V. Gandhi, S. Jeha, R. Kadota, G. N. Lam, W. Plunkett, B. Razzouk, M. Rytting, P. Steinherz, and S. Weitman. Population pharmacokinetics of clofarabine, a second-generation nucleoside analog, in pediatric patients with acute leukemia. *J. Clin. Pharmacol.* **44**:1309–1322 (2004).
18. K. Jolling, J. J. Ruixo, A. Hemeryck, V. Piotrovskij, and T. Greway. Population pharmacokinetic analysis of pegylated human erythropoietin in rats. *J. Pharm. Sci.* **93**:3027–3038 (2004).
19. R. Bruno, C. B. Washington, J. F. Lu, G. Lieberman, L. Banken, and P. Klein. Population pharmacokinetics of trastuzumab in patients With HER2+ metastatic breast cancer. *Cancer Chemother. Pharmacol.* (2005).
20. L. Lindbom, P. Pihlgren, and E. N. Jonsson. PsN-Toolkit–a collection of computer intensive statistical methods for non-linear mixed effect modeling using NONMEM. *Comput. Methods Programs Biomed.* **79**:241–257 (2005).
21. S. L. Beal and L. B. Sheiner. *NONMEM User's Guide - Part I. Users Basic Guide*. University of California, San Francisco, 1989.
22. J. C. Pinheiro and D. M. Bates. *Mixed-Effects Models in S and S-PLUS*. Springer-Verlag, New York, 2001.

23. B. Delyon, M. Lavielle, and E. Moulines. Convergence of a stochastic approximation version of the EM algorithm. *Ann. Stat.* **27**:94–128 (1999).
24. E. Kuhn and M. Lavielle. Coupling a stochastic approximation version of EM with a MCMC procedure. *ESAIM P&S*: 115–131 (2004).
25. D. Spiegelhalter, A. Thomas, N. Best, and D. J. Lunn. WinBUGS User Manual Version 1.4. Available from: http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml. Last Accessed: July 4 2006.
26. E. I. Ette. Comparing non-hierarchical models: application to non-linear mixed effects modeling. *Comput. Biol. Med.* **26**:505–512 (1996).
27. M. Davidian and D. M. Giltinan. *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall, London, 1995.
28. M. Davidian and D. M. Giltinan. Nonlinear Models for Repeated Measurement Data: An Overview and Update. *J. Agr. Biol. Envir. St.* **8**:387–419 (2003).
29. M. J. Lindstrom and D. M. Bates. Nonlinear mixed effects models for repeated measures data. *Biometrics* **46**:673–687 (1990).
30. J. I. Myung and D. J. Navarro. Information matrix. In: *Encyclopedia of behavioral Statistics*, B. Everitt and D. Howel, eds., 2004.
31. S. Retout, F. Mentré, and R. Bruno. Fisher information matrix for non-linear mixed-effects models: evaluation and application for optimal design of enoxaparin population pharmacokinetics. *Stat. Med.* **21**:2623–2639 (2002).
32. S. L. Beal and L. B. Sheiner. *NONMEM User's Guide – Part II. Users Supplemental Guide*. University of California, San Francisco, 1988.
33. H. White. Maximum likelihood estimation of misspecified model. *Econometrica* **50**:1–25 (1982).
34. E. Kuhn and M. Lavielle. Maximum likelihood estimation in nonlinear mixed effects model. *Comput. Stat. Data Anal.* **49**:1020–1038 (2005).
35. B. Efron and R. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, New York, 1993.
36. E. I. Ette. Stability and performance of a population pharmacokinetic model. *J. Clin. Pharmacol.* **37**:486–495 (1997).
37. S. Das and A. Krishen. Some bootstrap methods in nonlinear mixed-effect models. *J. Stat. Plan. Inference* **75**:237–245 (1999).
38. C. F. J. Wu. Comment on jackknife, bootstrap and resampling methods in regression analysis. *Ann. Stat.* **14**:1261–1350 (2006).
39. A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis, 2nd ed*. Chapman & Hall, London, 2004.
40. A. Gelman and D. B. Rubin. Inference from iterative simulation using multiple sequences (with discussion). *Stat. Sci.* **7**:457–511 (1992).
41. S. P. Brooks and A. Gelman. Alternative methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.* **7**:434–455 (1998).
42. N. G. Best, K. K. Tan, W. R. Gilks, and D. J. Spiegelhalter. Estimation of population pharmacokinetics using the Gibbs sampler. *J. Pharmacokinet. Biopharm.* **23**:407–435 (1995).
43. D. J. Lunn, N. Best, A. Thomas, J. Wakefield, and D. Spiegelhalter. Bayesian analysis of population PK/PD models: general concepts and software. *J. Pharmacokinet. Pharmacodyn.* **29**:271–307 (2002).
44. R. Pouillot, I. Albert, M. Cornu, and J.-B. Denis. Estimation of uncertainty and variability in bacterial growth using Bayesian inference. Application to *Listeria monocytogenes*. *Int. J. Food. Microbiol.* **81**:87–104 (2003).
45. D. J. Roe. Comparison of population pharmacokinetic modeling methods using simulated data: results from the population modeling workgroup. *Stat. Med.* **16**:1241–1262 (1997).
46. A. J. Boeckmann, L. Sheiner, and S. L. Beal. *NONMEM User's Guide – Part VIII. Help Guide*. University of California, San Francisco, 1996.

47. M. Tod, F. Mentré, Y. Merlé, and A. Mallet. Robust optimal design for the estimation of hyperparameters in population pharmacokinetics. *J. Pharmacokinet. Biopharm.* **26**:689–716 (1998).
48. M. Tod, F. Mentré, Y. Merlé, and A. Mallet. Introduction to POPOS (optimization of sampling times for population parameter estimation in pharmacokinetics). In *POPOS 1.0.* 2000, pp. 4–5.
49. P. O. Gisleskog, M. O. Karlsson, and S. L. Beal. Use of prior information to stabilize a population data analysis. *J. Pharmacokinet. Biopharm.* **29**:473–505 (2005).
50. D. B. White, C. A. Walawander, Y. Tung, and T. H. Grasela. An evaluation of point and interval estimates in population pharmacokinetics using NONMEM analysis. *J. Pharmacokinet. Biopharm.* **19**:87–112 (1991).
51. N. M. Laird and T. A. Louis. Empirical Bayes confidence intervals based on bootstrap samples. *J. Am. Stat. Assoc.* **32**:739–757 (1987).
52. J. R. Wade, A. W. Kelman, C. A. Howie, and B. Whiting. Effect of misspecification of the absorption process on subsequent parameter estimation in population analysis. *J. Pharmacokinet. Biopharm.* **21**:209–222 (1993).