Some Alternatives to Asymptotic Tests for the Analysis of Pharmacogenetic Data using Nonlinear Mixed Effects Models

Julie Bertrand[1,*], Emmanuelle Comets[1], Marylore Chenel[2] and France Mentré[1]

[1] UMR 738, INSERM, Université Paris Diderot, Paris F-75018, France

[2] Institut de Recherches Internationales Servier, Courbevoie F-92400, France

email: julie.bertrand@inserm.fr

Summary: Nonlinear mixed effects models allow investigating individual differences in drug concentration profiles (pharmacokinetics) and responses. Pharmacogenetics focusses on the genetic component of this variability. Two tests often used to detect a gene effect on a pharmacokinetic parameter are i) the Wald test, assessing whether estimates for the gene effect are significantly different from 0 and ii) the likelihood ratio test comparing models with and without the genetic effect. Because those asymptotic tests show inflated type I error on small sample size and/or with unevenly distributed genotypes, we develop two alternatives and evaluate them by means of a simulation study. First, we assess the performance of the permutation test using the Wald and the likelihood ratio statistics. Second, for the Wald test we propose the use of the F-distribution with four different values for the denominator degrees of freedom. We also explore the influence of the estimation algorithm using both the First-Order Conditional Estimation with interaction linearisation-based algorithm and the Stochastic Approximation Expectation Maximization algorithm. We apply these methods to the analysis of the pharmacogenetics of indinavir in HIV patients recruited in the COPHAR2-ANRS 111 trial. Results of the simulation study show that the permutation test seems appropriate but at the cost of an additional **computational** burden. One of the four F-distribution based approaches provides a correct type I error estimate for the Wald test and should be further investigated.

Key words: F-distribution based approach; First-order approximation; Nonlinear mixed effects models; Permutation tests; Pharmacogenetics.

1. Introduction

Pharmacokinetics (PK) studies the time course of a drug in the body. Nonlinear mixed effects models (NLMEM) in the analysis of PK data **allow integrating** the knowledge accumulated on drug absorption, distribution, metabolism and elimination (ADME), to quantify the inter-individual variability with fewer samples per patient than standard non compartmental approaches. Pharmacogenetics (PG) studies the relationship between this inter-individual variability and variations in the DNA sequence of proteins involved in the ADME mechanisms of the drug.

In previous simulation studies, we have shown that a correction for type I error inflation was required for the two most commonly used tests for differences between groups in NLMEM, the Wald and the likelihood ratio tests (LRT), in data sets of small sample size and/or with unevenly distributed genotypes (Bertrand et al., 2008, 2009b). Also in the context of testing parametric against semi-parametric nonlinear mixed models, Wu and Zhang (2002) have shown that asymptotic distributions cannot be used **for the Wald test and the LRT** when the normality assumption of the random effects is violated**, though the LRT was much less affected**.

The aim of the present work is to investigate two alternatives to these asymptotic tests to detect a gene effect in PG studies: permutation and F-distribution based tests. Permutation testing is a way of determining whether the null hypothesis of randomness is reasonable, i.e. whether the pattern present in the data **could have happen by chance alone** (Good, 1994; Manly, 1998). This alternative requires fewer assumptions than a correction based on simulations from the model under the null hypothesis, but is seldom used in NLMEM (Ding and Wu, 2001). The second alternative is to correct the Wald test for the underestimation of the **variance of parameter estimates** in small sample size. We investigate four such

corrections based on F-distributions with various values of denominator degrees of freedom (df) which have been proposed in different settings.

In the present work, we study the **type I error and power of these two alternatives** through simulations. The permutation tests are evaluated for both the Wald and the LRT statistics whereas the F-distributions are used to correct the Wald test only. **The same simulation setting as in the simulation studies mentioned before (Bertrand et al., 2008, 2009b) was used here to enable a comparison with the results of the simulation-based correction studied in these works.** As an illustration, we present an analysis of the indinavir concentration-time profiles collected in the COPHAR2 trial.

As in NLMEM the integral in the likelihood has no analytical form, estimation algorithms use model linearisation, such as the first order (FO) and first order conditional estimation with interaction (FOCE-I) (Lindstrom and Bates, 1990) methods implemented in the NONMEM and the Phoenix NLME software. Others consider numerical approximations of the likelihood such as the Laplacian or Adaptive Gaussian quadrature algorithms (AGQ) or more recently stochastic approximation of the EM algorithm (SAEM) (Deylon et al., 1999) as implemented in the MONOLIX software. In Bertrand et al. (2008, 2009b), we have shown that the bias in the FO algorithm leads to a very large inflation of both asymptotic tests type I error. Thus, to account for the influence of the estimation algorithm in the present work we use FOCE-I and SAEM.

In Section 2, we present the model and how the likelihood and the estimation variance matrix are obtained. Then, we introduce the usual asymptotic tests and both investigated alternatives in Section 3. In Sections 4 and 5 we describe the real data, the simulation study and the evaluation protocol. The results of the evaluation and the illustration are presented in Section 6. We finally discuss our findings and conclusions in Sections 7 and 8.

2. Evaluation of the likelihood and the estimation variance matrix in NLMEM

2.1 *Models and notation*

To describe the n_i-dimensional vector of concentrations $\boldsymbol{y_i}$ of a subject $i = 1, ..., N$, we use a pharmacokinetic function f **which is** nonlinear in its parameters $\boldsymbol{\phi_i}$:

$$\boldsymbol{y_i} = f(\boldsymbol{X_i}; \boldsymbol{\phi_i}) + \boldsymbol{\epsilon_i} \ , \tag{1}$$

where $\boldsymbol{X_i}$ is the within-subject design vector and $\boldsymbol{\phi_i} = \boldsymbol{\mu} + \boldsymbol{A_i}\boldsymbol{\beta} + \boldsymbol{B}\boldsymbol{\eta_i}$ is the subject p-**dimensional** vector of parameters, $\boldsymbol{A_i}$ is the $(p \times k)$-**dimensional** covariate matrix modelling the relationship between the covariates and $\boldsymbol{\phi_i}$ with $\boldsymbol{\beta}$ the corresponding k-coefficient effects vector. Because the covariate coefficients enter the model nonlinearly through the structural parameter associated to the covariate, the matrix-based notation peculiar to linear mixed effect models can no longer be used and therefore $\boldsymbol{\theta}$, the $(p + k)$-**dimensional** fixed effects vector is decomposed into $[\boldsymbol{\mu}' \quad \boldsymbol{\beta}']'$. $\boldsymbol{B}$ is a $(p \times q)$-**dimensional** design matrix, permitting some components of $\boldsymbol{\phi_i}$ to have no associated random effect when $p > q$. $\boldsymbol{\eta_i}$ is the random effect q-**dimensional** vector which follows a Gaussian distribution with null mean and variance-covariance $(q \times q)$-**dimensional** matrix $\boldsymbol{\Omega}$. $\boldsymbol{\epsilon_i}$ is the residual error n_i-**dimensional** vector which follows a Gaussian distribution with null mean and variance-covariance $(n_i \times n_i)$-**dimensional** matrix $\boldsymbol{\Sigma_i}(\boldsymbol{X_i}; \boldsymbol{\phi_i}, \boldsymbol{\gamma})$ whose diagonal terms are equal to $g(\boldsymbol{X_i}; \boldsymbol{\phi_i}, \boldsymbol{\gamma})^2$ and off-diagonal terms are equal to 0.

To obtain a combination of constant and proportional error models, $g(\boldsymbol{X_i}; \boldsymbol{\phi_i}, \boldsymbol{\gamma})$ can be set equal to $a + bf(\boldsymbol{X_i}; \boldsymbol{\phi_i})$ with $\boldsymbol{\gamma} = [a \quad b]'$ the vector of the error model parameters. Define the l-**dimensional** vector of variance parameters as $\boldsymbol{\lambda} = [\text{Vech}(\boldsymbol{\Omega})' \quad \boldsymbol{\gamma}']'$ where the operator Vech(.) creates a column vector from the matrix $\boldsymbol{\Omega}$ by stacking its lower diagonal elements below one another. Finally, define the $(p+k+l)$-**dimensional** vector of all model parameters as $\boldsymbol{\Psi} = [\boldsymbol{\theta}' \quad \boldsymbol{\lambda}']'$.

2.2 *Evaluation of the likelihood*

Because of the nonlinearity of the regression function in the random effects, the likelihood of NLMEM cannot be expressed in a closed form. Indeed, for the subject i, the marginal loglikelihood $L_i(\boldsymbol{y_i};\boldsymbol{\Psi})$ of $\boldsymbol{\Psi}$ for the data $\boldsymbol{y_i}$ is given by

$$\begin{aligned} L_i(\boldsymbol{y_i};\boldsymbol{\Psi}) =& \log\left(\int p(\boldsymbol{y_i};\boldsymbol{\phi_i},\boldsymbol{\Psi})d\boldsymbol{\phi_i}\right) \\ =& \log\left(\int p(\boldsymbol{y_i}|\boldsymbol{\phi_i},\boldsymbol{\Psi})p(\boldsymbol{\phi_i};\boldsymbol{\Psi})d\boldsymbol{\phi_i}\right) \end{aligned} \tag{2}$$

with $p(\boldsymbol{y_i}|\boldsymbol{\phi_i},\boldsymbol{\Psi})$ **the density of the observations conditioned on the random effects and the model parameters**, $p(\boldsymbol{\phi_i};\boldsymbol{\Psi})$ the density of the individual parameters and $p(\boldsymbol{y_i};\boldsymbol{\phi_i},\boldsymbol{\Psi})$ the likelihood of the complete data, that is,

$$\begin{aligned} -2\,\log\, p(\boldsymbol{y_i};\boldsymbol{\phi_i},\boldsymbol{\Psi}) =& n_i\,\log(2\pi) + \log(|\boldsymbol{\Sigma_i}(\boldsymbol{X_i};\boldsymbol{\phi_i},\boldsymbol{\gamma})|) \\ &+(\boldsymbol{y_i} - f(\boldsymbol{X_i};\boldsymbol{\phi_i}))'\boldsymbol{\Sigma_i}(\boldsymbol{X_i};\boldsymbol{\phi_i},\boldsymbol{\gamma})^{-1}(\boldsymbol{y_i} - f(\boldsymbol{X_i};\boldsymbol{\phi_i})) + q\,\log(2\pi) \\ &+\log(|\boldsymbol{\Omega}^*|)(\boldsymbol{\phi_i} - (\boldsymbol{\mu} + \boldsymbol{A_i}\boldsymbol{\beta}))'\boldsymbol{\Omega}^{*-1}(\boldsymbol{\phi_i} - (\boldsymbol{\mu} + \boldsymbol{A_i}\boldsymbol{\beta})) \end{aligned}$$

where $\boldsymbol{\Omega}^* = \boldsymbol{B}'\boldsymbol{\Omega}\boldsymbol{B}$.

Different approaches have been proposed to estimate $L_i(\boldsymbol{y_i};\boldsymbol{\Psi})$. Linearisation based methods consist of taking a Taylor expansion of the model function f and approximating (2) by the loglikelihood of a linear mixed effect model:

$$-2\,L_i(\boldsymbol{y_i};\boldsymbol{\Psi}) = n_i\,\log(2\pi) + \log(|\boldsymbol{V_i}|) + (\boldsymbol{y_i} - \boldsymbol{E_i})'\boldsymbol{V_i}^{-1}(\boldsymbol{y_i} - \boldsymbol{E_i})$$

where $\boldsymbol{E_i}$ and $\boldsymbol{V_i}$ are respectively the marginal expectation and variance of the vector $\boldsymbol{y_i}$.

Another approach is to use an importance sampling procedure (Robert and Casella, 1983) to compute an estimate $L_i(\boldsymbol{y_i};\boldsymbol{\Psi})_T$ of the observed loglikelihood, such as

$$L_i(\boldsymbol{y_i};\boldsymbol{\Psi})_T = \log\left(\frac{1}{T}\sum_{t=1}^{T} p(\boldsymbol{y_i}|\boldsymbol{\phi_i}^{(t)},\boldsymbol{\Psi})\frac{p(\boldsymbol{\phi_i}^{(t)};\boldsymbol{\Psi})}{h(\boldsymbol{\phi_i}^{(t)})}\right)$$

where the $\boldsymbol{\phi_i}^{(t)}$ are sampled from an instrumental distribution $h(\boldsymbol{\phi_i}^{(t)})$ which is chosen to minimize the variance of the estimate $L_i(\boldsymbol{y_i};\boldsymbol{\Psi})_T$. In the MONOLIX software where such a procedure is used, the instrumental function is a non-centred Student distribution (Samson et al., 2007; Lavielle, 2008).

2.3 *Evaluation of the estimation variance matrix*

The estimation variance matrix is composed of

$$\mathrm{Var}(\boldsymbol{\Psi}) = \begin{pmatrix} \mathrm{Var}(\boldsymbol{\theta}) & \mathrm{Var}(\boldsymbol{\theta}, \boldsymbol{\lambda}) \\ \mathrm{Var}(\boldsymbol{\theta}, \boldsymbol{\lambda})' & \mathrm{Var}(\boldsymbol{\lambda}) \end{pmatrix}$$

where $\mathrm{Var}(\boldsymbol{\theta})$ is the $((p+k) \times (p+k))$-estimation variance matrix for the fixed effects, $\mathrm{Var}(\boldsymbol{\lambda})$ is the $(l \times l)$-estimation variance matrix for the variance components, and $\mathrm{Var}(\boldsymbol{\theta}, \boldsymbol{\lambda})$ is the $((p+k) \times l)$-estimation covariance matrix between the fixed effects and the variance components. Based on the Cramer-Rao inequality, the inverse of the Fisher estimation matrix $\boldsymbol{M_F}$ is the lower bound of the variance covariance matrix of any unbiased estimators of the parameters. In the framework of normal theory maximum likelihood, $\boldsymbol{M_F}$ is computed as the Hessian of the loglikelihood in all the model parameters, and thus

$$\boldsymbol{M_F} = \sum_{i=1}^{N} \frac{-\partial^2 L_i(\boldsymbol{y_i}; \boldsymbol{\Psi})}{\partial \boldsymbol{\Psi} \partial \boldsymbol{\Psi}'}$$

where N is the total number of subjects. Thus, the (asymptotic) estimate of $\mathrm{Var}(\boldsymbol{\Psi})$ is ${\boldsymbol{M_F}}^{-1}$.

Here as well, linearisation of the model has been proposed to derive an approximate expression of $\boldsymbol{M_F}$. The FOCE-I approach linearises the model around the individual random effect estimates $\hat{\boldsymbol{\eta}}_{\boldsymbol{i}}$ (Pinheiro and Bates, 2000).

$$\boldsymbol{y_i} = f(\boldsymbol{X_i}; \tilde{\boldsymbol{\phi}}_{\boldsymbol{i}}) + \left. \frac{\partial f(\boldsymbol{X_i}; \boldsymbol{\phi_i})}{\partial \boldsymbol{\eta_i}} \right|'_{\tilde{\boldsymbol{\phi}}_{\boldsymbol{i}}} (\boldsymbol{\eta_i} - \hat{\boldsymbol{\eta}}_{\boldsymbol{i}}) + \boldsymbol{\epsilon_i}$$

with $\tilde{\boldsymbol{\phi}}_{\boldsymbol{i}} = \boldsymbol{\mu} + \boldsymbol{A_i}\boldsymbol{\beta} + \hat{\boldsymbol{\eta}}_{\boldsymbol{i}}$. The approximate marginal expectation $\boldsymbol{E_i}$ and variance $\boldsymbol{V_i}$ of the vector $\boldsymbol{y_i}$ are then given by

$$\begin{aligned} \boldsymbol{E_i} &= f(\boldsymbol{X_i}; \tilde{\boldsymbol{\phi}}_{\boldsymbol{i}}) - \left. \frac{\partial f(\boldsymbol{X_i}; \boldsymbol{\phi_i})}{\partial \boldsymbol{\eta_i}} \right|'_{\tilde{\boldsymbol{\phi}}_{\boldsymbol{i}}} \hat{\boldsymbol{\eta}}_{\boldsymbol{i}} \\ \boldsymbol{V_i} &= \left. \frac{\partial f(\boldsymbol{X_i}; \boldsymbol{\phi_i})}{\partial \boldsymbol{\eta_i}} \right|'_{\tilde{\boldsymbol{\phi}}_{\boldsymbol{i}}} \boldsymbol{\Omega}^* \left. \frac{\partial f(\boldsymbol{X_i}; \boldsymbol{\phi_i})}{\partial \boldsymbol{\eta_i}} \right|_{\tilde{\boldsymbol{\phi}}_{\boldsymbol{i}}} + \boldsymbol{\Sigma_i}(\boldsymbol{X_i}; \tilde{\boldsymbol{\phi}}_{\boldsymbol{i}}, \boldsymbol{\gamma}) \end{aligned} \tag{3}$$

Alternatively, a Taylor expansion of the model function f can also be performed around the individual parameter estimates $\hat{\boldsymbol{\phi}}_{\boldsymbol{i}} = \hat{\boldsymbol{\mu}} + \boldsymbol{A_i}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\eta}}_{\boldsymbol{i}}$, which provides the following

approximation. Then, (1) is rewritten as

$$\boldsymbol{y_i} = f(\boldsymbol{X_i}; \hat{\boldsymbol{\phi}}_{\boldsymbol{i}}) + \left.\frac{\partial f(\boldsymbol{X_i}; \boldsymbol{\phi_i})}{\partial \boldsymbol{\phi_i}}\right|'_{\hat{\boldsymbol{\phi}}_{\boldsymbol{i}}} (\boldsymbol{\phi_i} - \hat{\boldsymbol{\phi}}_{\boldsymbol{i}}) + \boldsymbol{\epsilon_i}$$

which is reformulated for sake of simplicity:

$$\tilde{\boldsymbol{y}}_{\boldsymbol{i}} = \left.\frac{\partial f(\boldsymbol{X_i}; \boldsymbol{\phi_i})}{\partial \boldsymbol{\phi_i}}\right|'_{\hat{\boldsymbol{\phi}}_{\boldsymbol{i}}} \boldsymbol{\phi_i} + \boldsymbol{\epsilon_i}$$

with $\tilde{\boldsymbol{y}}_{\boldsymbol{i}} = \boldsymbol{y_i} - f(\boldsymbol{X_i}; \hat{\boldsymbol{\phi}}_{\boldsymbol{i}}) + \left.\frac{\partial f(\boldsymbol{X_i}; \boldsymbol{\phi_i})}{\partial \boldsymbol{\phi_i}}\right|'_{\hat{\boldsymbol{\phi}}_{\boldsymbol{i}}} \hat{\boldsymbol{\phi}}_{\boldsymbol{i}}$. Then, the approximate marginal expectation $\boldsymbol{E_i}$ and variance $\boldsymbol{V_i}$ of the vector $\tilde{\boldsymbol{y}}_{\boldsymbol{i}}$ are given by

$$\begin{aligned} \boldsymbol{E_i} &= \left.\frac{\partial f(\boldsymbol{X_i}; \boldsymbol{\phi_i})}{\partial \boldsymbol{\phi_i}}\right|'_{\hat{\boldsymbol{\phi}}_{\boldsymbol{i}}} (\mu + \boldsymbol{A_i}\beta) \\ \boldsymbol{V_i} &= \left.\frac{\partial f(\boldsymbol{X_i}; \boldsymbol{\phi_i})}{\partial \boldsymbol{\phi_i}}\right|'_{\hat{\boldsymbol{\phi}}_{\boldsymbol{i}}} \boldsymbol{\Omega}^* \left.\frac{\partial f(\boldsymbol{X_i}; \boldsymbol{\phi_i})}{\partial \boldsymbol{\phi_i}}\right|_{\hat{\boldsymbol{\phi}}_{\boldsymbol{i}}} + \boldsymbol{\Sigma_i}(\boldsymbol{X_i}, \hat{\boldsymbol{\phi}}_{\boldsymbol{i}}, \boldsymbol{\gamma}) \end{aligned} \tag{4}$$

This latter approach is implemented in the MONOLIX software.

As we need its expression in the following, let us write here the estimation variance for the fixed effects. Using a first order linearisation of the model where $\boldsymbol{E_i}$ and $\boldsymbol{V_i}$ are given in either (3) or (4), we have $\mathrm{Var}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = 0$ and thus

$$\mathrm{Var}(\boldsymbol{\theta}) = \left(\sum_{i=1}^{N} \frac{\partial \boldsymbol{E_i}'}{\partial \boldsymbol{\theta}} \boldsymbol{V_i}^{-1} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}}\right)^{-1} \tag{5}$$

3. Asymptotic tests and alternatives

3.1 *Asymptotic tests*

In asymptotic conditions, testing the null hypothesis $H_0 : \boldsymbol{C\theta} = 0$ based on the estimates of the fixed effects $\hat{\boldsymbol{\theta}}$ can be carried out with the usual Wald test, comparing the Wald statistic to the critical value of a χ^2 distribution. Here, $\boldsymbol{C}$ is a $((p+k) \times U)$-contrast matrix and U the number of contrasts one wishes to test. If H_0 is true, then:

$$W = (\boldsymbol{C}\hat{\boldsymbol{\theta}})'(\boldsymbol{C}'\mathrm{Var}(\boldsymbol{\theta})\boldsymbol{C})^{-1}(\boldsymbol{C}\hat{\boldsymbol{\theta}}) \sim \chi^2_U \tag{6}$$

H_0 can also be tested using the likelihood ratio test (LRT) that compares the models with $\boldsymbol{C\theta} = 0$ (reduced) with the full model where $\boldsymbol{\theta}$ is estimated. If H_0 is true, then:

$$S = -2 \times (L_{reduced} - L_{full}) \sim \chi^2_U$$

where $L_{reduced}$ and L_{full} are the loglikelihood of the two models.

3.2 *Permutation based alternative*

To perform permutation tests, R data sets are generated by permuting the rows of the covariate matrix from the original data set. For a given test, the statistic Q^{obs} is estimated from the original data and the statistic Q^{perm} is estimated from each of the R data sets. Thus, $r = 1, ..., R$ values Q^{perm_r} are obtained which constitute a distribution of the statistic under the null hypothesis of no covariate effect. The permutation p-value is the proportion: $(\text{Card}(Q^{perm_r} \geqslant Q^{obs}) + 1)/(R + 1)$, where the operator Card(.) counts the realisations of $Q^{perm_r} \geqslant Q^{obs}$.

3.3 *F-distribution based alternative*

Comparing W/U to an F-distribution with an infinite denominator df is equivalent to using the classical Wald test. However, in biology studies, sample size is often small to moderate, thus approximate F-distributions with numerator df= U and non infinite denominator df have been proposed to correct for the departure from the asymptotic. In this section,we consider four different F-distribution based approaches.

The first approach derives from decomposition of the degrees of freedom in the ANOVA (Pinheiro and Bates, 2000) (=DF_{PB}) with a denominator df equal to $\sum_{i=1}^{N} n_i - (N+p+k-U)$ and is implemented in the nlme function in R. In the second approach which is implemented in the NLMIXED Procedure in SAS, the computation of the degrees of freedom is based on the number of random effects (Wolfinger, 2000) (=DF_{W}) with a denominator df equal to $N - q$. The third approach comes from the multivariate nonlinear models (MNLM) framework (Gallant, 1975) (=DF_{G}). In his study on nonlinear regressions contemporaneously but not serially correlated, Gallant (1975) observed that estimation variances were underestimated and thus recommended **the** multiplication of $\text{Var}(\boldsymbol{\theta})$ by the factor $N/(N-p)$ and using a denominator

df equal to $N - p$. In classical non linear regression, the downward bias of the maximum likelihood variance estimator is corrected using a factor of $n/(n-p)$, with n being the number of observations and p being the number of model parameters (Huet et al., 1996). Here, Gallant (1975) proposed substituting N and not $\sum_{i=1}^{N} n_i$ for n because MNLM, as a population-averaged approach, focusses on the marginal expectation of the response variable (Vonesh and Chinchilli, 1997).

The fourth approach is an extension to NLMEM of the method developed by Fai and Cornelius (1996) (=$\mathrm{DF_{FC}}$) that is implemented in the DDFM=SATTERTH option of the MIXED Procedure in SAS. Fai and Cornelius (1996) proposed using the spectral decomposition of $\boldsymbol{C}'\mathrm{Var}(\boldsymbol{\theta})\boldsymbol{C}$ in (6) to decompose W into a sum of U squared Student statistics with **of** df

$$\upsilon_u = \frac{2(c_u'\mathrm{Var}(\boldsymbol{\theta})c_u)^2}{\mathrm{Var}(c_u'\mathrm{Var}(\boldsymbol{\theta})c_u)}$$

with $u = 1, .., U$ **the decomposition index and** where $\mathrm{Var}(c_u'\mathrm{Var}(\boldsymbol{\theta})c_u)$ can be approximated by the delta method using the estimates of $\mathrm{Var}(\boldsymbol{\lambda})$ and the estimates of the model parameters:

$$\mathrm{Var}(c_u'\mathrm{Var}(\boldsymbol{\theta})c_u) \approx \left(\frac{\partial c_u'\mathrm{Var}(\boldsymbol{\theta})c_u}{\partial \boldsymbol{\lambda}}\right)' \mathrm{Var}(\boldsymbol{\lambda}) \left(\frac{\partial c_u'\mathrm{Var}(\boldsymbol{\theta})c_u}{\partial \boldsymbol{\lambda}}\right) \tag{7}$$

When $U > 1$, Fai and Cornelius (1996) showed that W/U follows an approximate F-distribution with denominator df:

$$\upsilon = \frac{2\sum_{u=1}^{U} \frac{\upsilon_u}{\upsilon_u - 2}}{\sum_{u=1}^{U} \frac{\upsilon_u}{\upsilon_u - 2} - U}$$

As, in NLMEM, $\mathrm{Var}(\boldsymbol{\theta})$ includes derivatives, we use the following property of matrix derivatives $\frac{dA^{-1}}{dx} = -A^{-1}\frac{dA}{dx}A^{-1}$ to take the derivative of $c_u'\mathrm{Var}(\boldsymbol{\theta})c_u$ in (7) with respect to each element l of the vector $\boldsymbol{\lambda}$:

$$\frac{\partial c_u'\mathrm{Var}(\boldsymbol{\theta})c_u}{\partial \boldsymbol{\lambda}_l} = c_u' \left(-\mathrm{Var}(\boldsymbol{\theta}) \frac{\partial \left(\sum_{i=1}^{N} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}}' \boldsymbol{V_i}^{-1} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}}\right)}{\partial \boldsymbol{\lambda}_l} \mathrm{Var}(\boldsymbol{\theta})\right) c_u \tag{8}$$

We also use this property to take the derivative of $\sum_{i=1}^{N} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}}' \boldsymbol{V_i}^{-1} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}}$ in (8) with respect to each element l of $\boldsymbol{\lambda}$, so we have, for each element q of $Vech(\boldsymbol{\Omega})$,

$$\frac{\partial \left(\sum_{i=1}^{N} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}}' \boldsymbol{V_i}^{-1} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}} \right)}{\partial Vech(\boldsymbol{\Omega})_q} = \sum_{i=1}^{N} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}}' \left(-\boldsymbol{V_i}^{-1} \left. \frac{\partial f(\boldsymbol{X_i}; \boldsymbol{\phi_i})}{\partial \boldsymbol{\phi_i}} \right|_{\hat{\boldsymbol{\phi}}_i}' \left. \frac{\partial f(\boldsymbol{X_i}; \boldsymbol{\phi_i})}{\partial \boldsymbol{\phi_i}} \right|_{\hat{\boldsymbol{\phi}}_i} \boldsymbol{V_i}^{-1} \right) \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}}$$

and for the error model parameters,

$$\frac{\partial \left(\sum_{i=1}^{N} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}}' \boldsymbol{V_i}^{-1} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}} \right)}{\partial a^2} = \sum_{i=1}^{N} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}}' \left(-\boldsymbol{V_i}^{-1} diag \left(1 + \frac{b f(\boldsymbol{X_i}; \hat{\boldsymbol{\phi}}_i)}{a} \right) \boldsymbol{V_i}^{-1} \right) \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}}$$

$$\frac{\partial \left(\sum_{i=1}^{N} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}}' \boldsymbol{V_i}^{-1} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}} \right)}{\partial b^2} = \sum_{i=1}^{N} \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}}' \left(-\boldsymbol{V_i}^{-1} diag \left(\frac{a f(\boldsymbol{X_i}; \hat{\boldsymbol{\phi}}_i)}{b} + f(\boldsymbol{X_i}; \hat{\boldsymbol{\phi}}_i)^2 \right) \boldsymbol{V_i}^{-1} \right) \frac{\partial \boldsymbol{E_i}}{\partial \boldsymbol{\theta}}$$

In these calculations, we linearise the model around the individual parameters estimates as described in Section 2.3.

4. Real Data and Simulation Study

4.1 *Real data*

As in our previous papers (Bertrand et al., 2008, 2009b), we illustrate the different approaches with data from a PK substudy of the COPHAR2-ANRS 111 study, a multicentre non-comparative pilot trial of early therapeutic drug monitoring in HIV-positive patients naïve of treatment. We focus on the PK sub-study from the group of patients receiving indinavir doses of 400, 600 and 800 mg boosted with ritonavir at a dose of 100 mg twice a day. Patients were genotyped for the exons 21 and 26 of the ABCB1 gene which code for the P-glycoprotein, **and for** polymorphisms on gene coding for proteins involved in the metabolism of indinavir; the *CYP* 3A4*1B, *CYP* 3A5*3 and *6 polymorphisms.

The **indinavir pharmacokinetic** profiles were determined at 1, 3, 6, and 12h following drug administration, 2 weeks after the treatment onset **(Fig. 1)**.

[Figure 1 about here.]

4.2 *Simulation setting*

An extended description of the simulations can be found in Bertrand et al. (2008) and is briefly summarised below.

Parameters from the simulated NLMEM were set based on a preliminary analysis of the indinavir data without covariates. The concentrations at time t were simulated using a one compartment model at steady state ($\tau = 12\ h$) with first order absorption ($k_a = 1.4\ h^{-1}$), first order elimination ($k = 0.2\ h^{-1}$) and apparent volume of distribution ($V/F = 102\ L$).

$$f(t; k_a, k, V/F) = \frac{D}{V/F}\frac{k_a}{k_a - k}\left(\frac{e^{-kt}}{1 - e^{-k\tau}} - \frac{e^{-k_a t}}{1 - e^{-k_a \tau}}\right)$$

We use a diagonal $\mathbf{\Omega}$ matrix and a proportional error model ($g(t; \boldsymbol{\phi_i}, \boldsymbol{\gamma}) = bf(t, \boldsymbol{\phi_i})$) setting the dose (D) to 400 mg. The model parameters are expressed in term of natural logarithms to achieve the positivity requirement of pharmacokinetic parameters, $\boldsymbol{\mu} = [log(k_a) \quad log(k) \quad log(V/F)]'$ with random effect standard deviations set to 1.13, 0.41 and 0.26, respectively and the coefficient of variation for the residual error set to 20% (b=0.2).

We simulate a diplotype of SNP_1 and SNP_2, with a distribution mimicking that of exons 26 and 21 of the ABCB1 gene Sakaeda et al. (2002); we obtained frequencies for the rare homozygotes, heterozygotes and common homozygotes of respectively 24%, 48% and 28% for SNP_1 and 29%, 44% and 27% for SNP_2. As, in the intestine, the P-glycoprotein restricts drug entry into the body, we consider an effect on the drug bioavailability (F) through the volume of distribution V/F.

The genetic coefficient values are chosen to be consistent with results found in the literature for ABCB1 polymorphisms on drugs disposition (Marzolini et al., 2003) and provide clinically relevant effect, with V/F and CL/F ($=k \times V/F$) increasing from 105.4 to 200.5 L and 21.1 to 40.1 L/h respectively between common and rare homozygotes for SNP_1.

In this work, the tests under study assess only the effect of SNP_1 on V/F even if we simulated diplotypes to allow for population genetic factors such as linkage disequilibrium.

In the model including the covariate, the effect coefficient vector is $\boldsymbol{\beta} = [\beta_1 \quad \beta_2]'$. As we test for the effect of the SNP as a whole, we have $U = 2$ and $\boldsymbol{C} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}'$.

We simulated 200 data sets with $N = 40$ subjects and $n = 4$ samples as in the COPHAR2 trial under both the null hypothesis of no genetic effect (H_0) and the alternative hypothesis of a genetic effect (H_1) (Fig. 2).

[Figure 2 about here.]

5. Evaluation protocol

5.1 *Simulation study*

In the present research we used the FOCE-I and SAEM algorithms implemented in the softwares NONMEM version 7 and MONOLIX version 2.1 respectively to estimate the parameters, the likelihood and the estimation variance matrix. In NONMEM, the starting values were set to the simulated ones + 10% except for the standard deviation of k_a which was set to 0.7. In MONOLIX, the starting values were similar for the fixed effects but the random effect standard deviations were set to 1.

For the likelihood estimation in MONOLIX, the number of iterations for the importance sampling T was set to 10000.

The alternatives proposed in the present work are evaluated in terms of type I error and power. The type I error estimate is the percentage of the 200 data sets simulated under H_0 where the test is significant and the power estimate is the same percentage but for the 200 data sets simulated under H_1. The prediction interval around 5% for 200 data sets is [2%; 8%]. The estimates of type I error and power obtained with permutations or F-distributions are compared to values obtained using the asymptotic tests (theoretical threshold of p-value=0.05) and using a simulation based correction (threshold built from simulations under the model with $\boldsymbol{C\theta} = 0$).

As described in Section 3, the type I error of the asymptotic tests is obtained using a threshold of 5.99 which is the 95^{th} percentile of a χ^2 distribution with 2 degrees of freedom. The type I error for the simulation based correction is obtained using a threshold of 10.18 and 7.05 with FOCE-I and 7.80 and 6.96 with SAEM for the Wald test and the LRT, respectively. These values are the 95^{th} percentiles of the distributions obtained for both the Wald test and the LRT statistics on 1000 data sets simulated under the null hypothesis in Bertrand et al. (2008) using the same setting as described in Section 4 of the present paper.

We also evaluate the p-value distribution for both tests on the 200 data sets simulated under H_0. If the test statistic does follow its reference distribution, these p-values should follow an uniform distribution on the interval $[0, 1]$. Thus, we computed the Kolmogorov distance between the uniform distribution on the interval $[0, 1]$ and the p-value distributions observed for the Wald test and the LRT under their asymptotic form and using the proposed alternatives with both estimation algorithms (Ruckdeschel et al., 2006). For a sample of 200 values, the Kolmogorov-Smirnov (K-S) test of equality of one-dimensional probability distributions is significant at a level of 5% when the Kolmogorov distance is below $\frac{1.358}{\sqrt{200}} = 0.096$.

For the permutation test, the number R of permutations is set to 1000.

The fourth approach based on F-distribution (DF_{FC}) is evaluated only with SAEM in MONOLIX. Our implementation of this denominator df calculation is nested within the MONOLIX code, following the computation of the estimation variance matrix. We did not program the extension with the FOCE-I algorithm as we had no easy access to the NONMEM code dedicated to the estimation variance matrix.

5.2 *Application to real data*

The indinavir concentrations were analysed with the same PK model as in the simulation study, using the SAEM algorithm. The details of the covariate model building strategy are

described in Bertrand et al. (2009a). Briefly, we tested for demographic and biological covariates along with the genetic polymorphisms. First, a screening on the estimated individual parameters was performed, followed by a forward selection based on LRT. Then, the p-value of the covariates remaining in the final model were assessed using the asymptotic LRT as well as the four F-distribution based Wald tests for illustrative purpose.

6. Results

6.1 *Simulation study*

As shown on Table 1, SAEM achieved convergence on all data sets and always provided the estimation variance matrix, whereas FOCE-I failed to converge and/or compute the estimation variance matrix on few data sets.

[Table 1 about here.]

The thresholds used for the asymptotic test, the simulation based correction, and the first three F-distribution based approaches, are the same for the 200 data sets under both hypotheses. For the permutation test and the $\mathrm{DF}_{\mathrm{FC}}$ approach, the threshold comes from, respectively, the permutation based distribution and the estimated denominator df, which are both specific to the data set.

With SAEM, the type I error for the asymptotic Wald test and LRT are significantly inflated. With FOCE-I, the type I error for the asymptotic Wald test is significantly inflated whereas the asymptotic LRT estimate is at the upper limit of the prediction interval around 5%. Of note, the Wald test and LRT asymptotic estimates on the 1000 data sets used to derive the simulation-based threshold are 8.9% and 7.6% with SAEM and 12.2% and 7.4% with FOCE-I. For the permutation test, the type I error estimates of the Wald test and the LRT are non-significantly different from the nominal level of 5%, with both algorithms. Of note with FOCE-I, the maximal value observed over the 195 thresholds is four times larger

than that obtained with SAEM for the Wald test. Using the simulation-based correction or the permutation test, the corrected power estimates for the Wald test are much lower using FOCE-I than SAEM

[Figure 3 about here.]

As displayed on Fig. 3, all the Kolmogorov distance estimates are below the K-S threshold for the permutation test and the simulation based correction. With SAEM and FOCE-I, the Kolmogorov distance estimate of the $\mathrm{DF_G}$ approach is similar to that obtained using permutation test or simulations based correction with both estimation algorithms, although the type I error estimate is about 10% with FOCE-I and not significantly different from the nominal level of 5% with SAEM. For the asymptotic test and the $\mathrm{DF_{PB}}$ approach, the Kolmogorov distance and type I error estimates are close and significantly inflated. Indeed, the large denominator df of the $\mathrm{DF_{PB}}$ approach leads to a threshold quite close to that of a χ^2 divided by 2. The $\mathrm{DF_{FC}}$ and the $\mathrm{DF_W}$ methods also have close and significantly inflated estimates of Kolmogorov distance and type I error, though lower than the asymptotic test. The range of denominator df obtained with the $\mathrm{DF_{FC}}$ method is quite narrow and happens to include that of the $\mathrm{DF_W}$ approach.

Fig. 4 shows that for the Wald test with SAEM, one should use a denominator df of 10 to obtain a type I error of exactly 5%, and between 6 and 27 for an estimate non significantly different from 5%.

[Figure 4 about here.]

6.2 *Real data*

As described in Bertrand et al. (2009a), the indinavir concentration time data are adequately fitted by a one compartment model parametrised in first-order absorption rate, apparent volume of distribution and apparent elimination clearance (Cl/F). The model with no co-

variate has an absorption constant of 1.3 h^{-1} with an important inter-individual variability of 118%, an elimination clearance of 21.9 $L.h^{-1}$ and a volume of distribution of 93.9 L with inter-individual variabilities of 34.4% and 19.3%, respectively. The standard deviations of the random effect are expressed as the coefficient of variation of the pharmacokinetic parameters, as in the model the latter are coded in natural logarithms. Those estimates are in accordance with the literature on indinavir given in combination with ritonavir. The coefficient of variation for the residual error is 44.5%. All of the estimation relative standard errors (RSE) were below 25% with the exception of k_a and V/F (around 30% and 60%, respectively).

After ascending selection using the asymptotic LRT, only an effect of the *CYP* 3A4*1B polymorphism effect on k_a (p-value=0.02) and an age effect on Cl/F (p-value=0.03) remain in the model. The p-values of the permutation tests are 0.04 and 0.1 for the *CYP* 3A4*1B polymorphism and age effects, respectively. The corresponding p-values are 0.014 and 0.043 using the DF_{PB} approach, 0.018 and 0.048 using the DF_W approach, 0.023 and 0.057 using the DF_G approach, and 0.02 and 0.047 using the DF_{FC} approach.

Only the effect of the *CYP* 3A4*1B polymorphism remains in the final model, as the effect of age is discarded based on the p-value estimates from the permutation test and the DF_G approach.

7. Discussion

Several studies have evaluated by simulation the performance of tests for discrete covariate on continuous responses using NLMEM with various designs and estimation methods (see the review in the discussion of Bertrand et al. (2009b)). Permutation tests and F-distribution based tests have been compared in the linear mixed effects framework (Lin and Heagerty, 2004; Routledge, 1997). In particular Routledge (1997) concluded that when unable to guarantee, in advance, the reliability of p-values based on the F-distribution one should

use a permutation test. However, as far as we know, such an evaluation has not yet been published in the specific context of NLMEM.

Both **simulation and permutation based-**approaches are relatively time consuming with computing time estimates which are about 3.4 h and 24 min compared to 19.2 s and 12.4 s for the asymptotic Wald test with SAEM and FOCE-I, respectively. The permutation test is found to be a robust alternative to correction based on simulations under H_0 from a model, with type I error close to 5%. With regard to the main assumption of permutation tests, our simulation setting ensures exchangeability of the observations as the genetic covariate only affected the fixed effects. The external validity of a permutation test is often questioned as it is by construction 'sample dependent'. Manly (1998) argues however that it is equivalent to question the representativeness of the study sample which also conditions the external validity of classical tests.

Because the covariates enter the model in a linear fashion, we also consider using a F-distribution for the Wald test, a widely used method in LMEM (Verbeke and Molenberghs, 1997). The df of the $\mathbf{DF_{PB}}$ approach which is derived from the multivariate ANOVA differs greatly from the df used within the mixed effect framework. Indeed, in the multivariate ANOVA, the subject concentrations are considered as independent variables under the influence of the genotype and the multivariate statistic is based on the comparison of the between and within variance matrices whose covariance terms inform on the correlation between the variables. Thus in this approach, n variables are considered in N subjects leading to an $N \times n$ sample size. Then, one degree of freedom is subtracted per parameter estimates minus the number of contrast tested. On the other hand in the mixed effects model, the independent variable is the subject with N the sample size. For $\mathbf{DF_W}$ and $\mathbf{DF_G}$, one degree of freedom is subtracted per

random effect variance and fixed effect respectively. For $\mathbf{DF_{FC}}$, the information available in the data for the parameter estimation is quantified through the variance of parameter estimates rather than by subtracting degrees of freedom to the sample size. Here, all these methods however fail to correct for the Wald type I error inflation, as the denominator df estimates for the 200 data sets are too large. Actually, the optimal denominator df is even smaller than the number of independent observations to compensate for the downward bias of the variance of the estimates, as shown by the performance of the $\mathbf{DF_G}$ method.

Indeed, the only method based on an F-distribution that corrects for the type I error inflation of the Wald test is the $\mathrm{DF_G}$ method with SAEM. With the FOCE-I algorithm, the Kolmogorov distance suggested an improvement using the $\mathrm{DF_G}$ method despite the fact that we observed an inflation of the type I error, indicating that the correction may be useful even in this case. To go past the simple design under study here, we considered two other designs from a previous study evaluating the Wald test and the LRT (Bertrand et al., 2009b): a design optimised using the PFIM software (Retout et al., 2007) including $N = 80$ subjects sorted in 4 groups with $n = 2$ samples and a combined design with $N = 20$ subjects having $n = 4$ samples plus $N = 80$ subjects with only one trough concentration ($n = 1$). The type I error estimates using the classical Wald test and the $\mathrm{DF_G}$ method are 8.7% and 5.7%, respectively for the optimal design and 8.4% and 4.5%, respectively for the combined design. These results show that the $\mathrm{DF_G}$ method also corrects the type I error inflation observed in these designs. However, it does so not only through approximating the distribution for the Wald test but essentially through inflating the estimation variance for the effect coefficient by the factor $N/(N - p)$. This factor, which brings Bessel's correction for the unbiased estimator of the variance to mind, was recommended by Gallant (1975) based only on a Monte Carlo study in the multivariate nonlinear model framework. Thus, caution should be

used and further investigation is needed as to designs involving different number of samples per subject and in other simulation settings.

One limitation of the present research is that we consider the method developed by Fai and Cornelius (1996) and not the method developed by Kenward and Roger (1997). The latter method is indeed based on restricted maximum likelihood (REML) which is not commonly used in NLMEM. Further, Faes et al. (2009) have recently revisited the effective sample size in correlated data and provided an approximate degrees of freedom method. They used REML estimation, thus we did not assess their approach. Meza et al. (2007) have developed a REML extension of the SAEM algorithm not implemented in MONOLIX yet. A perspective of the present work would therefore be to assess the Kenward and Roger (1997) and Faes et al. (2009) methods using this extension. **Of note,** a Bayesian framework would be appealing, not the least because it would circumvent the issue of the type I error by summarizing the available information in a posterior distribution instead of a p-value.

8. Conclusion

While this research confirms the feasibility of permutation tests in pharmacogenetic studies for both the LRT and the Wald test, further investigations in other simulation settings are required to recommend the F-distribution based approach proposed in multivariate nonlinear models for the Wald test. As permutation does only slightly better than the asymptotic alternatives for an additional computing burden, a sensible course of action could be to use it only when decisions based respectively on the asymptotic test and the Gallant alternative are discordant.

Acknowledgements

The authors would like to thank IFR02 and Hervé Le Nagard for the use of the Centre de Biomodélisation as well as Prof. Marc Lavielle for the valuable help he provided in using MONOLIX.

During this research, Julie Bertrand was supported by a grant from the Institut de Recherches Internationales Servier (France).

References

Bertrand, J., Comets, E., Laffont, C., Chenel, M., and Mentré, F. (2009b). Pharmacogenetics and population pharmacokinetics: impact of the design on three tests using the SAEM algorithm. *Journal of Pharmacokinetics and Pharmacodynamics* **36,** 317–339.

Bertrand, J., Comets, E., and Mentré, F. (2008). Comparison of model-based tests and selection strategies to detect genetic polymorphisms influencing pharmacokinetic parameters. *Journal of Biopharmaceutical Statistics* **18,** 1084–1102.

Bertrand, J., Tréluyer, J., Panhard, X., Tran, A., Rey, E., Salmon-Céron, D., Duval, X., Mentré, F., and the COPHAR2-ANRS 111 study group. (2009a). Influence of pharmacogenetics on indinavir disposition and short-term response in HIV patients initiating HAART. *European Journal of Clinical Pharmacology* **65,** 667–678.

Deylon, B., Lavielle, M., and Moulines, E. (1999). Convergence of a stochastic approximation version of EM algorithm. *Annals of Statistics* **27,** 94–128.

Ding, A. A. and Wu, H. (2001). Assessing antiviral potency of anti-HIV therapies in vivo by comparing viral decay rates in viral dynamic models. *Biostatistics* **2,** 13–29.

Faes, C., Molenberghs, G., Aerts, M., Verbeke, G., and Kenward, M. G. (2009). The effective sample size and an alternative small-sample degrees-of-freedom method. *American Statistician* **63,** 389–399.

Fai, A. and Cornelius, P. (1996). Approximate F-tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments. *Journal of Statistical Computing and Simulation* **54,** 363–378.

Gallant, A. R. (1975). Seemingly unrelated nonlinear regressions. *Journal of Econometrics* **3,** 35–50.

Good, P. (1994). *Permutation tests: a practical guide to resampling methods for testing hypotheses*. New York: Springer.

Huet, S., Bouvier, A., Gruet, M. A., and Jolivet, E. (1996). *Statistical Tools for Nonlinear Regression: A Practical Guide with S-PLUS Examples*. New York: Springer.

Kenward, M. G. and Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* **53,** 983–997.

Lavielle, M. (2008). *MONOLIX (MOdèles NOn LInéaires à effets miXtes)*. MONOLIX group, Orsay, France.

Lin, D. Y. and Heagerty, P. J., editors (2004). *Proceedings of the Second Seattle Symposium in Biostatistics: Analysis of Correlated Data*. New York: Springer.

Lindstrom, M. and Bates, D. (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics* **46,** 673–687.

Manly, B. (1998). *Randomization, Bootstrap and Monte Carlo Methods in Biology*. Texts in statistical science. London: Chapman & Hall, 2 edition.

Marzolini, C., Paus, E., Buclin, T., and Kim, R. B. (2003). Polymorphisms in human MDR1 (P-glycoprotein): recent advances and clinical relevance. *Clinical Pharmacology and Therapeutics* **75,** 13–33.

Meza, C., Jaffrézic, F., and Foulley, J. L. (2007). REML estimation of variance parameters in nonlinear mixed effects models using the SAEM algorithm. *Biometrical Journal* **49,** 876–888.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed Effects Models in S and S-Plus*. New York: Springer.

Retout, S., Comets, E., Le Nagard, H., Bazzoli, C., and Mentré, F. (2007). *PFIM Interface 2.1*. UMR738, INSERM, Université Paris 7, Paris, France.

Robert, C. P. and Casella, G. (1983). *Monte Carlo Statistical Methods*. New York: Springer.

Routledge, R. D. (1997). P-values from permutation and F-tests. *Computational Statistics & Data Analysis* **24,** 379–386.

Ruckdeschel, P., Kohl, M., Stabla, T., and Camphausen, F. (2006). S4 classes for distributions. *R News* **6,** 2–6.

Sakaeda, T., Nakamura, T., and Okumura, K. (2002). MDR1 genotype-related pharmacokinetics and pharmacodynamics. *Biological & Pharmaceutical Bulletin* **25,** 1391–1400.

Samson, A., Lavielle, M., and Mentré, F. (2007). The SAEM algorithm for group comparison tests in longitudinal data analysis based on non-linear mixed-effects model. *Statistics in Medicine* **26,** 4860–4875.

Verbeke, G. and Molenberghs, G. (1997). *Linear Mixed Models in Practice. A SAS-Oriented Approach*. New York: Springer.

Vonesh, E. F. and Chinchilli, V. M. (1997). *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. New York: Marcel Dekker.

Wolfinger, R. D. (2000). Fitting nonlinear mixed models with the new NLMIXED procedure. Technical report, SAS Institute Inc., Cary, NC, USA. Paper 287.

Wu, H. and Zhang, J. T. (2002). The study of long-term HIV dynamics using semi-parametric non-linear mixed-effects models. *Stat Med* **21,** 3655–3675.

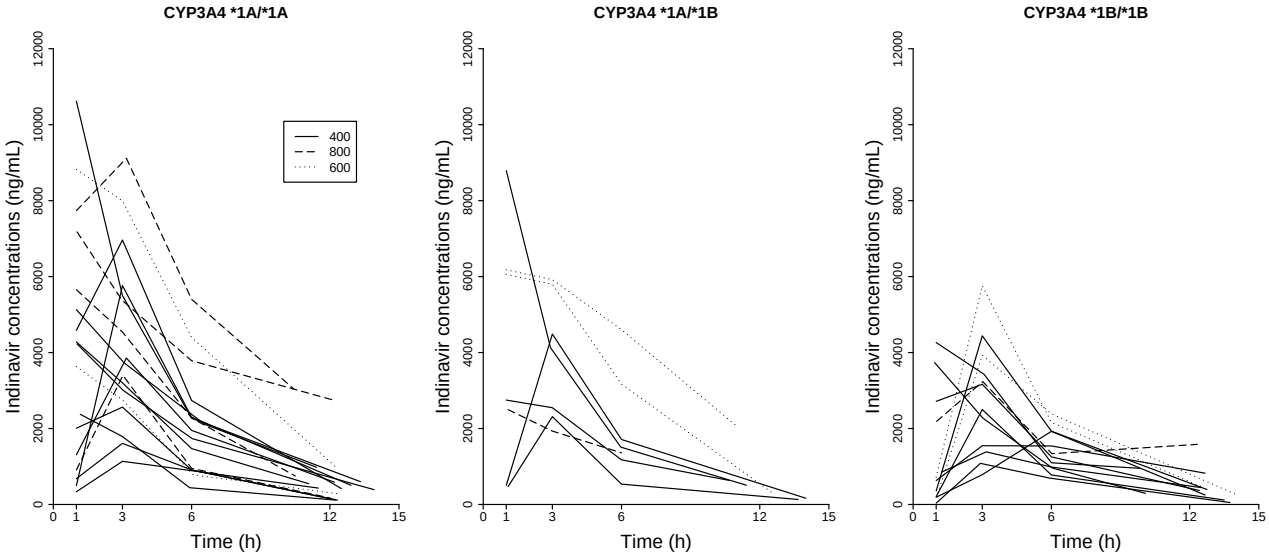

Figure 1. Indinavir pharmacokinetic concentrations versus time profiles for the 40 patients included in the COPHAR2-ANRS 111 study sorted by dose and genotype for the *CYP* 3A4*1B polymorphism. The plain, dashed and dotted lines represent patients with a 400, 600 and 800 mg dose, respectively.

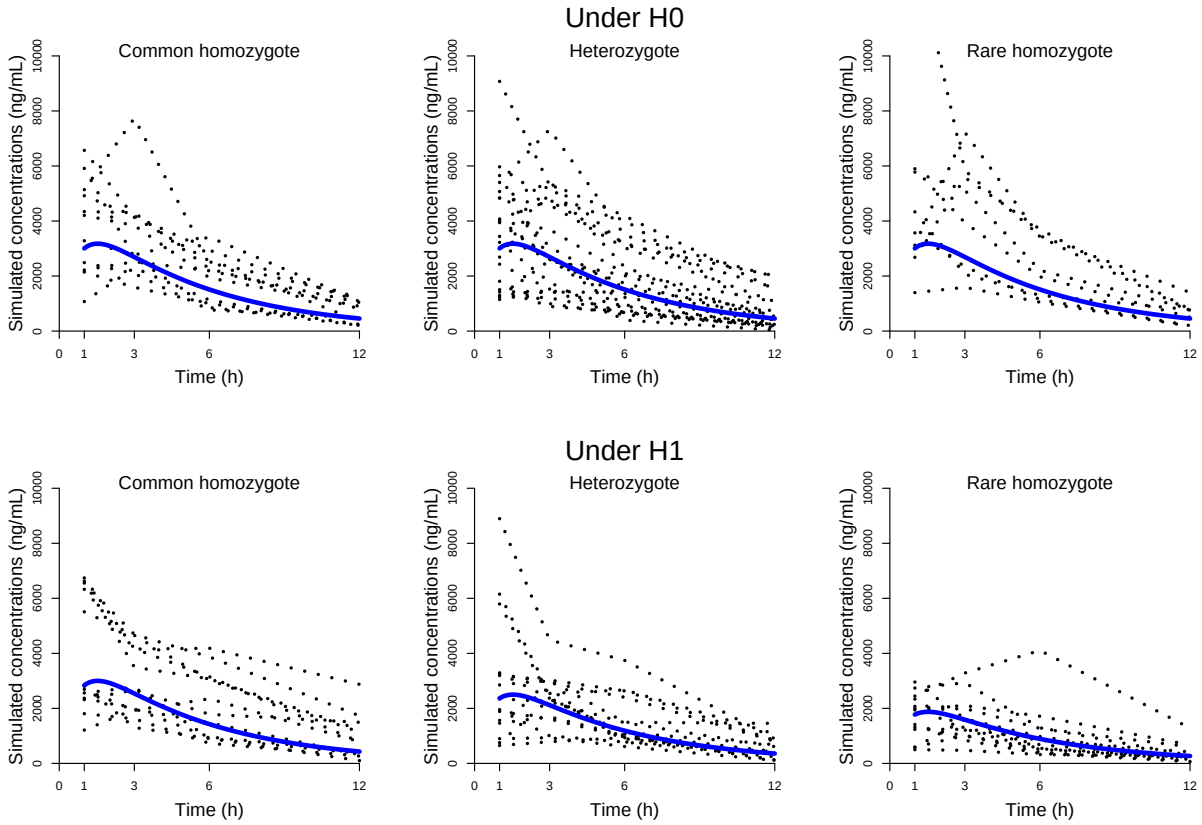

Figure 2. Concentrations (ng/mL) simulated for a representative data set under H_0 (top) and a representative one under H_1 (bottom). The subjects are sorted by genotype for the exon SNP_1: common homozygotes (left), heterozygotes (center) and rare homozygotes (right). The profile for the mean parameters is represented by a thick line. Individual concentration curves are represented by dotted lines.

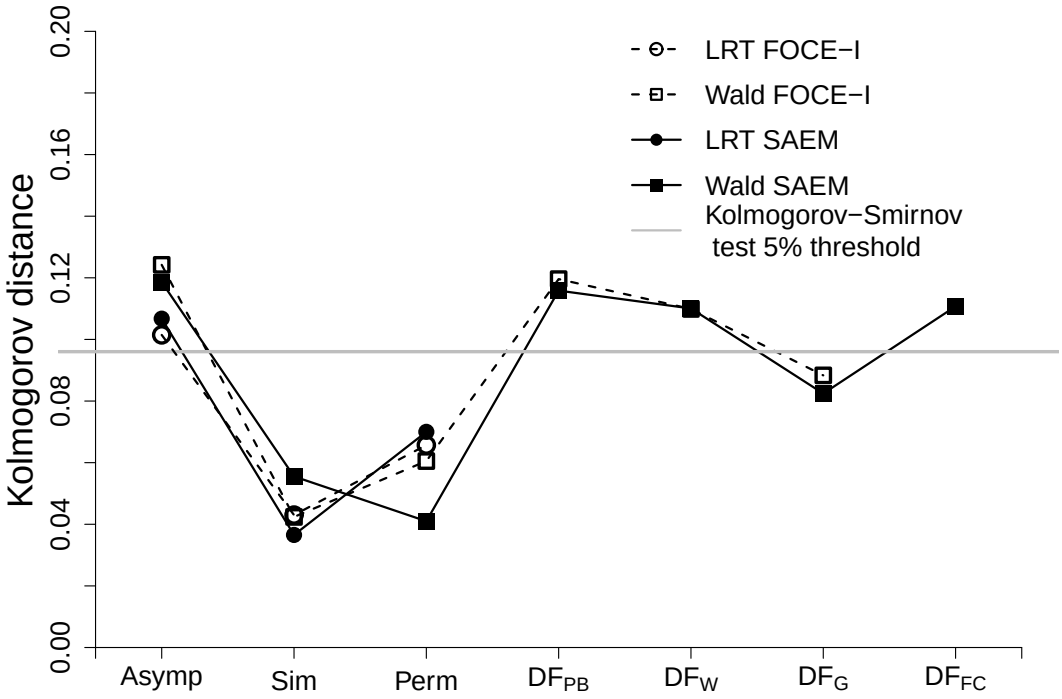

Figure 3. Kolmogorov distance between the uniform distribution on the interval $[0, 1]$ and the p-value distributions obtained on 200 simulations under H_0 for the Wald test and the LRT under their asymptotic form and using the proposed alternatives with FOCE-I and SAEM. The horizontal gray line is the corresponding threshold of the Kolmogorov-Smirnov test for 200 data sets.

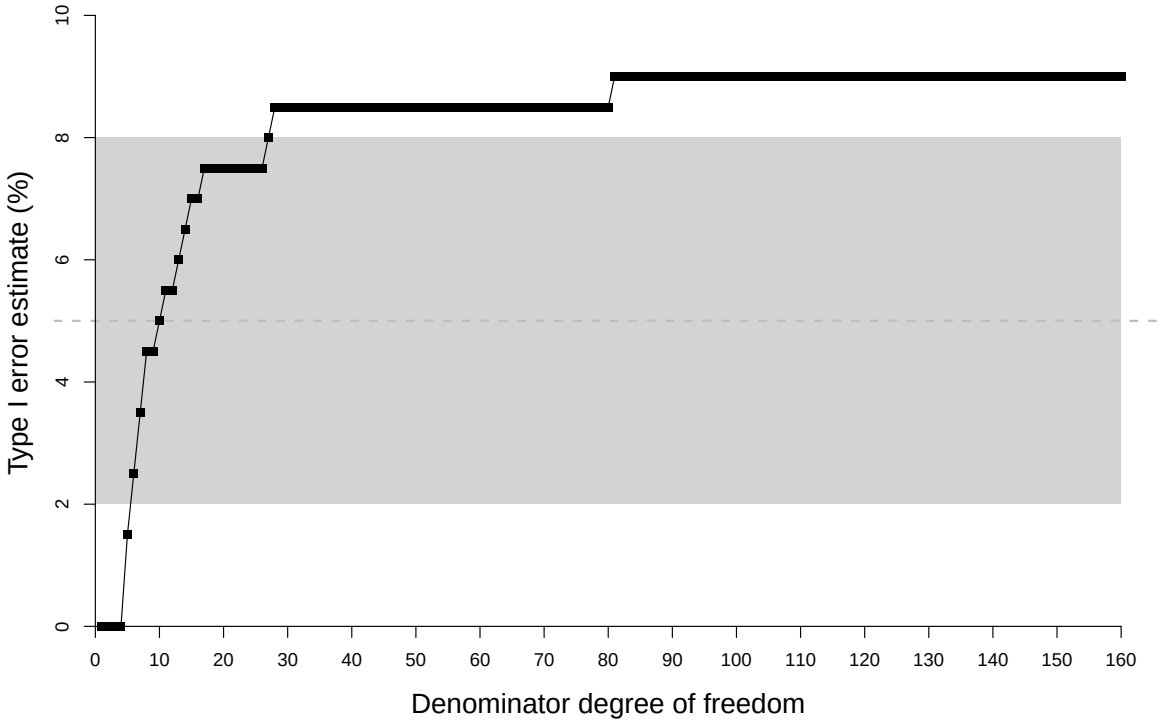

Figure 4. Type I error estimates (%) of the Wald test on the 200 data sets simulated under H_0 versus the corresponding denominator degree of freedom with SAEM. The gray dotted line and shaded area represent the nominal level of 5% and its 95% prediction interval.

Table 1
Threshold, denominator degrees of freedom (df), type I error, and power estimates (%) of the Wald test and the LRT for different approaches, using both FOCE-I and SAEM

Algorithm	Test	Approach	denominator df	Threshold	T	Type I error	T	Power
FOCE	Wald	asymptotic	-	5.99*	199	13.1	199	53.8
		simulation	-	10.18*	199	4.5	199	18.6
		permutation	-	9.09[6.9;34.65]*†	195	4.6	197	27.4
		DF_{PB}	117	3.07	199	12.1	199	52.3
		DF_W	37	3.25	199	11.1	199	48.2
		DF_G	37	3.25	199	10.1	199	44.2
	LRT	asymptotic	-	5.99*	200	8.0	199	76.4
		simulation	-	7.05*	200	5.5	199	69.8
		permutation	-	6.57[5.75;7.63]*†	198	6.5	199	72.4
SAEM	Wald	asymptotic	-	5.99*	200	10.0	200	78.0
		simulation	-	7.80*	200	5.5	200	71.0
		permutation	-	7.33[6.33;8.46]*†	200	6.0	200	73.0
		DF_{PB}	117	3.07	200	9.0	200	78.0
		DF_W	37	3.25	200	8.5	200	77.0
		DF_G	37	3.25	200	7.5	200	75.0
		DF_{FC}	39.8[36.3;43.8]†	3.23[3.21;3.26]†	200	8.5	200	77.0
	LRT	asymptotic	-	5.99*	200	9.0	200	77.0
		simulation	-	6.89*	200	6.5	200	72.0
		permutation	-	6.73[4.45-8.11]*†	200	7.0	200	72.0

T = number of data sets on which the test could be performed.

Predicted interval for a nominal level of 5% = $[2.0 - 8.0]$.

* Threshold for the distribution of W, not W/2.

† The threshold for the permutation test and both the threshold and the denominator df for the DF_{FC} approach are displayed as median [range] over the T data sets simulated under H_0.