Comparison of model-based tests and selection strategies to detect genetic polymorphisms influencing pharmacokinetic parameters.

Julie Bertrand, Emmanuelle Comets, France Mentre

Comparison of model-based tests and selection strategies to detect genetic polymorphisms influencing pharmacokinetic parameters

Bertrand Julie [*] **, Comets Emmanuelle , Mentre France**

Modèles et méthodes de l'évaluation thérapeutique des maladies chroniques INSERM : U738, Université Denis Diderot - Paris VII, FR

* Correspondence should be adressed to: Julie Bertrand <julie.bertrand@inserm.fr>

Abstract

We evaluate by simulation three model-based methods to test the influence of a single nucleotide polymorphism on a pharmacokinetic parameter of a drug: ANOVA on the empirical Bayes estimates of the individual parameters, likelihood ratio test between models with and without genetic covariate, and Wald tests on the parameters of the model with covariate. Analyses are performed using the FO and FOCE method implemented in the NONMEM software. We compare several approaches for model selection based on tests and global criteria. We illustrate the results with pharmacokinetic data on indinavir from HIV-positive patients included in COPHAR 2-ANRS 111 to study the gene effect prospectively. Only the tests based on the EBE obtain an empirical type I error close to the expected 5%. The approximation made with the FO algorithm results in a significant inflation of the type I error of the LRT and Wald tests.

INTRODUCTION

Pharmacokinetics studies the time course of a drug in the body (Gabrielsson and Weiner, 1999). The variability in the pharmacokinetics of drugs when administered to different subjects is often important and should be studied to improve the use of drugs, avoid toxic events and allow individualization of therapy. The contribution of genetic factors to this variability is potentially important (Licinio and Wong, 2002). Some genes have already been the subject of much attention, for example the ABCB1 gene coding for the P glycoprotein (P-gP) found on the main exchange barriers (Marzolini et al., 2003). The involvement of this protein in drug absorption processes has been demonstrated directly in animals and indirectly in humans. For example, co-administration of paclitaxel and cyclosporine, respectively a substrate and an inhibitor of P-gP, has been shown to increase the bioavailability of paclitaxel (Meerum et al., 1999). More recently, Yamaguchi and colleagues demonstrated the impact of ABCB1 polymorphism on paclitaxel pharmacokinetics in patients with ovarian cancer (Yamaguchi et al., 2006). The ABCB1 gene is composed of 209 kb (Bodor et al., 2005), and to date 28 polymorphisms concerning a nucleotide modification have been described, including several with potential clinical impact (Sakaeda et al., 2002). In the present study, we consider data from patients treated with indinavir in the COPHAR2-ANRS11 study investigating the benefit of early therapeutic drug monitoring in anti-retroviral therapy. Since the ABCB1 genetic polymorphism was found to have an influence on the pharmacokinetics of protease inhibitors (Fellay et al., 2002), patients were genotyped for two exons (exon 21 and 26) of the ABCB1 gene.

The influence of genetic polymorphism on concentration data is usually analysed using non-compartmental analysis (NCA). The area under the concentration versus time curve as well as model-independent parameters are calculated for each individual concentration profile using, for example, the log-trapezoidal method (Gabrielsson and Weiner, 1999) and individual pharmacokinetic (or PK) parameters are then compared between the different genotype groups using ANOVA (Inomata et al., 2005). More sophisticated approaches using nonlinear PK models in individual regression (Min et al., 2004) or mixed effects models (Taguchi et al., 2006) have also been applied to genetic data in PK studies. Nonlinear mixed effects models (NLMEM) require fewer blood samples in each patient and can be important in special populations such as patients with acute disease or neonates, for whom extensive sampling is obviously impractical.

Various methods can be used to include pharmacogenetic information in NLMEM (Comets et al., 2007). Empirical Bayes estimates (EBE) of the individual parameters can be computed from the fit of a model with no covariate and compared between the different genotype groups with an ANOVA to test the genetic polymorphism effect (Henningsson et al., 2005). Another approach is to perform Wald tests on the estimates of the gene effect coefficients in the covariate model (Kerbusch et al., 2003), while stepwise model building is frequently based on the comparison of models with and without gene effect using the likelihood ratio test (LRT) (Mamiya et al., 2000). The first purpose of this work is to assess the statistical properties of these three different strategies to test for a gene effect through a simulation study. The setting for the simulation study is based on the COPHAR2-ANRS11 study. We have also taken the impact of estimation methods into account, comparing results from the FO and FOCE methods implemented in the NONMEM (nonlinear mixed effect model) software (Sheiner and Beal, 1998). The type I error of the tests is evaluated using simulations under the null hypothesis H_0 of

no gene effect for two designs (40 patients as in the study, 200 patients to examine the influence of sample size), while the power is compared using simulations under an alternative hypothesis H_1.

In the same simulation study, we also examine model selection strategies, in which the aim is to choose the best covariate model for the gene effect. We compare test-based strategies with general selection criteria that have been proposed, like the Bayes information criterion (BIC) (Schwartz, 1978) and the Akaike information criterion (AIC) (Akaike, 1974). According to the parsimony principle, these criteria penalize the log-likelihood by the number of model parameters which limits overfitting. Alternative criteria derived from the AIC have been developed in order to deal with small samples or to improve AIC asymptotic consistency. The properties of these criteria have been studied in generalized linear models (Burnham and Anderson, 2002), like the trend to select an over-parameterized model with AIC and under-parameterized model with BIC. AIC and BIC have been widely used in population PK and PK/PD studies during the past decade. In the present paper, we compare through a simulation study their properties for nonlinear mixed effects model selection.

In this paper, we first introduce the model, the notations, the different tests and model building strategies under study and the estimation methods. Then we describe the case study and the simulation study based on its design. Next we present the results of the simulation study and the application to the real data. Finally we discuss our findings.

METHODS

Model and notations

Let the function f denote the PK model, which depends nonlinearly on its parameters θ. The concentration $y_{i,j}$ at time $t_{i,j}$ for subject i =1,…,N and measurement j = 1,…,n_i is given by:

$$y_{i,j} = f(t_{i,j}, \theta_i) + \epsilon_{i,j}$$

where θ_i is the vector of P PK parameters for the ith individual and $\epsilon_{i,j}$ the residual error, assumed to be normally distributed with zero mean and variance $\sigma_{i,j}^2$. Here we assume a proportional error model:

$$\sigma_{i,j}^2 = \sigma^2 f(t_{i,j}, \theta_i)^2$$

We assume a multivariate log-normal distribution for the vector of individual parameters, θ_i are then expressed as:

$$\theta_i = \mu \, e^{b_i}$$

with μ the vector of fixed effects and b_i the vector of random effects. The random effects b_i are assumed to be independent of $\epsilon_{i,j}$ and normally distributed with zero mean and variance matrix Ω. In this work, we used a diagonal Ω with interindividual variance of the pth parameter ω_p^2, however an unspecified positive definite matrix can be assumed.

For simplicity, let us assume that our aim is to detect the effect of a single nucleotide polymorphism (SNP) on one PK parameter, for instance the pth component of θ; θ^p. Let C denote the wild-type allele and T the mutant. An individual can be one of three genotypes (eg. CC for the wild-type homozygotes, CT or TT). Let G_i denote the genotype for subject i, and let $\beta(G_i) = \beta_0$, β_1 or β_2 for G_i = CC, CT or TT, respectively. We write the model for the genetic polymorphism effect in subject i as:

$$\theta_i^p = \mu^p \beta(G_i) e^{b_i^p}$$

We assume that CC is the reference class so that β_0 = 1.

Let model M_{base} be the model in the absence of genetic polymorphism effect: {$\beta_0 = \beta_1 = \beta_2 = 1$}, (CC = CT = TT). There are three models including the gene effect as a covariate: a complete model where the mean of the parameter is different in the three groups, M_{mult}: { $\beta_0 = 1, \beta_1 \neq \beta_2 \neq 1$}, (CC ≠ CT ≠ TT) and two reduced models, $M_{recessive}$: {$\beta_0 = \beta_1 = 1, \beta_2 \neq 1$}, (CC = CT ≠ TT) and $M_{dominant}$: {$\beta_0 = 1, \beta_1 = \beta_2 \neq 1$}, (CC ≠ CT = TT).

In the following, L corresponds to the model log-likelihood, P_{pop} represents the number of population model parameters (mean PK parameters, covariate coefficients, variances and error model parameters), N the sample size and n_{tot} the total number of observations. The standard error and covariance of the parameter estimates are abbreviated by SE and cov, respectively.

Tests for genetic polymorphism effect

In this section, we examine different tests based on NLMEM that can be used to test the existence of a genetic polymorphism effect on one parameter of a PK population model.

Analysis of variance (ANOVA)

Data are analysed with a model with no covariates (M_{base}) and Empirical Bayes estimates (EBE) of the individual PK parameters are computed. A one-way analysis of variance is used to detect differences in these EBE between the different genetic groups.

Wald test

The data are analysed using the complete model M_{mult}, and the significance of the parameters is assessed using a global Wald test:

$$W = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}^T \Sigma^{-1} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

where Σ represents the variance covariance matrix of parameters β_1 and β_2. The statistic W is compared to the critical value of a χ^2 with two degrees of freedom.

Likelihood Ratio Test (LRT)

The third test relies on the comparison between the model with no covariate effect M_{base} and the complete model M_{mult}. These two nested models are compared with the LRT. The test statistic $S_{LR} = -2(L_{base} - L_{mult})$ is compared with a χ^2 with two degrees of freedom where L_{base} and L_{mult} are the log-likelihood of M_{base} and M_{mult} respectively. The two degrees of freedom correspond to the difference in the number of population parameters between the two models.

Strategies for model building

In this section, we examine a second aspect, which is model selection. The different strategies under study provide decision rules for covariate inclusion in order to get the best model. Figure 1 represents in a diagram the decision path of the three strategies using tests. When performing multiple tests, a correction which preserves the false discovery rate is used (Benjamini and Hochberg, 1995).

Selection based on EBE

If the test that compares means of EBE obtained with M_{base} between CC, CT and TT using an ANOVA is significant, the means of the EBE obtained with M_{base} are compared between the CC and the CT, on one hand, and between the CT and the TT, on the other hand, using t-tests. If none or one test is significant, the one with the lower p-value leads to model selection, or else M_{mult} is selected. For example, if the test comparing the CC to the CT has the lower p-value, $M_{dominant}$ is chosen.

If the global test is non-significant, the means of the EBE obtained with M_{base} are compared using t-tests opposing: i) the CC to the CT and TT put together, and ii) the TT to the CT and CC put together. If one or both tests are significant, the model corresponding to the test with the lower p-value is selected, else M_{base} is selected. For example if the test comparing the CC to the group formed by the CT and the TT has the lower p-value, $M_{dominant}$ is chosen.

Selection based on Wald test

If the global Wald test comparing W to a χ^2 with two degrees of freedom is significant, two Wald tests on coefficients estimated in M_{mult} are then realized; comparing both $\frac{\beta_1^2}{SE^2(\beta_1)}$ and $\frac{(\beta_1 - \beta_2)^2}{SE^2(\beta_1) + SE^2(\beta_2) - 2cov(\beta_1, \beta_2)}$ with a χ^2 with one degree of freedom. If none or one test is significant, the one with the lower p-value leads to model selection, or else M_{mult} is kept. As an example, if $\frac{\beta_1^2}{SE^2(\beta_1)}$ has the lower p-value, $M_{dominant}$ is chosen.

If the global test is non-significant, data are analysed using the two models with the gene effect in two classes and Wald tests are performed on the coefficients estimated in $M_{recessive}$ and $M_{dominant}$. The two statistics $\frac{(\beta_{rec} - 1)^2}{SE^2(\beta_{rec})}$ and $\frac{(\beta_{dom} - 1)^2}{SE^2(\beta_{dom})}$ are compared with a χ^2 with one degree of freedom. If one or both tests are significant, the model is chosen based on the test with the lower p-value, or else M_{base} is conserved. By way of example, if $\frac{(\beta_{dom} - 1)^2}{SE^2(\beta_{dom})}$ has the lower p-value, $M_{dominant}$ is selected.

Selection based on LRT

If the LRT comparing M_{base} with M_{mult} is significant, then two LRT are performed comparing M_{mult} with the nested models $M_{dominant}$ and $M_{recessive}$. If none or one test is significant, the test with the higher p-value leads the model choice, or else M_{mult} is selected. As an example, if the test comparing M_{mult} with $M_{dominant}$ obtains the higher p-value, $M_{dominant}$ is selected.

If the global test is non-significant, M_{base} is compared with $M_{dominant}$ and $M_{recessive}$ using LRT. Then, if one or both tests are significant, the model is chosen based on the test with the lower p-value, or else M_{base} is selected. For example, if the test comparing M_{base} to $M_{dominant}$ obtains the lower p-value, $M_{dominant}$ is selected.

Selection based on information criteria

As an alternative, criteria such as AIC or BIC can be used for model selection. Here, the model with the lowest criterion is chosen. Both of these criteria balance the log-likelihood with the number of parameters in the population model. According the parsimony principle, a simpler model is preferred for equivalent information gain. The AIC is written as:

$$AIC = -2L + 2P_{pop}$$

The expression of the BIC involves the total number of observations:

$$BIC = -2L + P_{pop} \log(n_{tot})$$

The BIC is related to the Bayes factor which is a measure of the strength of evidence in favour of a given model and can thus be used to quantify model uncertainty.

Other criteria have been defined, derived from those above: the corrected AIC (AICc) defined by Akaike for small samples i.e. when $n_{tot}/P_{pop} < 40$ (Sugiura, 1978) and the consistent AIC (CAIC) (Bozdogan, 1987). They are derived from AIC and involve the total number of observations:

$$AICc = AIC + \frac{2P_{pop}(P_{pop}+1)}{n_{tot} - P_{pop} - 1}$$
$$CAIC = -2L + P_{pop}(\log(n_{tot}) + 1)$$

Another formulation for BIC has also been proposed, BICc, where n_{tot} is replaced by the number of subjects (Raftery, 1995):

$$BICc = -2L + P_{pop} \log(N)$$

Estimation methods

In NLMEM, the parameters and their standard error are mainly estimated using maximum likelihood. However the likelihood function for these models is expressed as an integral and has no analytical solution. Specific algorithms have therefore been proposed to perform the maximization. The approach most frequently used relies on first-order approximations of the likelihood function. This approach has been implemented in the NONMEM software version V which is the most frequently used software in PK/PD analyses. In NONMEM two approximations can be used: the first-order method (FO), relying on a first-order linearization of the likelihood function around $b_i = 0$, and the first-order conditional estimation method (FOCE), relying on a first-order linearization of the likelihood function around the estimates of the individual random effects. Although the FO method suffers from bias and lack of precision it is still frequently used as it is faster and has less convergence problems than FOCE. In this paper, we compare both estimation methods. More precisely, we use the FOCE with interaction method, allowing possible interaction between inter- and intra-variability.

Real data and simulation study

Real data

We illustrate the different approaches on data from a PK sub-study of the COPHAR 2-ANRS 111 study, a multicentre non-comparative pilot trial of early therapeutic drug monitoring in HIV positive patients naïve of treatment. The objective of the trial was to assess the benefit of a pharmacological intervention after measurement of trough plasma concentrations of protease inhibitors (Mentré et al., 2005). We focus on the PK sub-study from the group of patients receiving indinavir boosted with ritonavir. Patients were genotyped for the ABCB1 exons 21 and 26 to investigate genetic polymorphism impact on pharmacokinetics of the protease inhibitors, which are well-known substrates of the P-gp (Fellay et al., 2002).

Forty-two patients were included, one patient withdrew from the study and one switched to another protease inhibitor during the first week of treatment. We therefore obtained PK data for 40 patients (27 men, 13 women) with an average age of 36.5 years. PK profiles were determined at 1, 3, 6 and 12 h after administration of the drug at a date two weeks after the treatment onset. One patient had missing information for the two genotypes, and ABCB1 exon 26 genotype was missing in two other patients.

Simulated data

The design used in the simulation mimics that of the application data set. We simulate PK studies of N = 40 (equivalent to the application sample size) and N = 200 patients with four samples (1, 3, 6, 12 h after dose) at steady state. The same bid doses of 400 mg for indinavir and 100 mg for ritonavir are assumed for all patients. The concentrations are simulated using the steady-state one-compartment model with first-order absorption and elimination that was used to model the indinavir concentrations in the COPHAR 1-ANRS 102 study (Brendel et al., 2005):

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

where F represents the bioavailability, k_a the absorption rate, k the elimination rate, V the volume of distribution and τ = 12 h the time between two doses. We use the following population parameters: V/F = 102 L with interindividual variability ω_v = 41.3%, k_a = 1.4 h^{-1} with ω_{k_a} = 113%, k = 0.2 h^{-1} with ω_k = 26.4% obtained with a preliminary analysis of the real data using the FO method in NONMEM and a residual error of 20%. The measurements below the quantification limit (BQL) are treated in the analysis using a standard approach: the first value in a series of BQL was set to LOQ/2 and the remaining values were discarded (Beal, 2005).

Under H$_0$, 1000 data sets are simulated with a design of N = 40 patients and 1000 with a design of N = 200 patients. We simulate a combination of SNP on two exons located on the same chromosome. The bioavailability F is assumed to depend on the diplotype. Because the model is parameterized as V/F, k_a and k, this is equivalent to assuming that V/F depends on the diplotype. The distribution of the exons mimics that of exon 26 and exon 21 of the ABCB1 gene as reported by Sakaeda and colleagues (Sakaeda et al., 2002) – we note C and G respectively the wild-type allele for the 2 exons and T the mutant allele. With those properties, for data sets with 40 patients we expect on average 9 individuals with a CC genotype, 18 with CT genotype and 13 with TT genotype for exon 26.

Under the alternative hypothesis, we assume the following effect of the two polymorphisms (modified from equation (4)), where G$_{26i}$ denotes the genotype for the exon 26 and G$_{21i}$ the genotype for the exon 21:

$$(V/F)_i = V/F \, \beta(G_{26i}) \, \delta(G_{21i}) \, e^{b_i}$$

where $\beta(G_{26i})$ is β_0, β_1 or β_2 if = CC, CT or TT as previously and $\delta(G_{21i})$ is δ_0, δ_1 or δ_2 if G$_{21i}$ = GG, GT or TT. Under H$_1$ we set $\beta(G_{26})$ = 1, 1.2, 1.6 and $\delta(G_{21})$ = 1, 1.1, 1.3. These values were chosen to provide a good power for the detection of one SNP effect in the context of the simulations while remaining consistent with results found in the literature concerning the effect of ABCB1 polymorphism. With those genetic coefficient effects and the distribution from the literature, we simulated 100 data sets with the N = 40 design. When we computed the (V/F)$_i$ EBE from M$_{base}$ and performed bilateral t-tests to compare the wild homozygotes and the mutant homozygotes for the exon 26, we obtained a power of 80% (Machin et al., 1997).

The simulation of each data set is performed as follows. The set of possible genotypes for exon 26 and exon 21 is S={CC-GG, CC-GT, CC-TT, CT-GG, CT-GT, CT-TT, TT-GG, TT-GT, TT-TT} with corresponding simulated frequencies f={0.2, 0.02, 0.02, 0.05, 0.38, 0.05, 0.04, 0.04, 0.2}. For each individual the genotypes are drawn from this distribution. Under H$_1$, both genotypes condition the value of the fixed effects for V/F according to equation 11. Then we simulate a random effect vector b$_i$ from a normal distribution N(0, Ω), yielding the individual parameters vector θ_i according to equation 3. The concentrations are computed using these parameters. Finally we add a residual error, generated from a normal distribution N(0, σ^2 f(θ,t$_{i,j}$)), to each predicted concentration to obtain the simulated concentration.

Simulations were performed using the statistical software R (R Development Core Team 2006) running under Linux (Red Hat 9.0).

Evaluation of the tests for genetic polymorphism effect

In the first step, each of the three tests presented above is applied to detect the effect of the exon 26 polymorphism on the 1000 data sets simulated under H$_0$ for each design (40 and 200 patients). Tests are performed on estimations obtained for each data set using FO and FOCE in NONMEM. The type I error for each analysis is defined as the percentage of data sets where the corresponding test was significant. The expected prediction interval with 1000 simulations and a value of 5% is [3.7; 6.3]. To ensure a type I error of 5%, we define a correction threshold as the 5th percentile of the distribution of the p-values of the test under H$_0$.

In the second step, for the design with 40 patients, the same tests are performed using the 1000 data sets simulated under H$_1$. The power is defined as the percentage of data sets where the corresponding test was significant. We use the corrected threshold to compute the corrected powers. This allows comparison of the different tests even if the type I error is different from 5%.

Data sets with a group defined by the genotype for exon 26 with less than two patients are discarded from the analysis. It should be emphasized that the number of simulated data sets for which the tests can be applied are often less than 1000. Indeed, the algorithms FO and FOCE used in the NONMEM software are sometimes unable to converge on a data set. In that case, estimates are not available and no tests can be applied. Similarly, even when the linearization algorithms achieve convergence, the variance covariance matrix is not always available. More precisely, to perform the ANOVA on the EBE only the convergence of model M$_{base}$ is required with or without the SE estimates. The Wald test requires the parameter estimation error to compute the test statistics, thus not only the convergence of model M$_{mult}$ is required but the variance covariance matrix must also be available. The convergence of both M$_{base}$ and M$_{mult}$ is necessary to apply the LRT but the covariance step is not needed to succeed. The three tests were also evaluated on a subset of data sets fulfilling all the conditions listed in this paragraph using FO or FOCE.

Evaluation of the strategies for model building

The different model building strategies described above based either on tests or selection criteria are evaluated on data sets with the N = 40 design. Results are reported as the percentage of data sets for which each model of the exon 26 polymorphism effect is selected, and this with simulation under H_0 or under H_1. The correct model is M_{base} under H_0 and M_{mult} under H_1. The results of the strategies obtained by simulation under H_0 are not used to modify the strategies under H_1 as there are no simple correction methods for over-selecting covariate models under H_0.

As for the evaluation of tests, the analyses are not always performed on the 1000 data sets. The conditions to perform model building following an ANOVA are identical to those described for the test. To choose the best model using Wald tests, the convergence and the covariance step are required for the three models including the gene effect. For the selection with the LRT, convergence of the four models is required and for the criteria based selection, at least one model must have achieved convergence. The different strategies were also evaluated on samples satisfying all conditions with FO and FOCE.

Application to real data

The indinavir concentrations are analysed with the same PK model as in the simulation study. The estimation method is chosen based on performances in the simulation study. Both the exons 26 and 21 of the ABCB1 gene are investigated using the three tests and the eight model building strategies.

RESULTS

Type I error and power of the detection tests

For the design with N = 40, three simulated data sets (one under H_0 and 2 under H_1) are discarded from the analysis due to a group of less than two patients with the TT genotype. Also, FOCE encounters many more convergence problems than FO; K, the number of data sets among the 1000 on which the test is performed, is always lower with FOCE.

Type I error estimates of the three tests performed are shown in Table 1. For the design with N = 40, the ANOVA type I errors do not significantly differ from 5% with both estimation algorithms. The LRT type I error estimates show a slight significant increase with FOCE and increase by ten times using FO. For the Wald test, there is a rather important significant increase for FOCE and again inflation by four times with FO. For the design with 200 patients, the LRT attain a type I error non-significantly different from 5% and the Wald test still has an estimate slightly superior to the nominal level with FOCE. The large increase of the type I error remains the same with FO for the LRT and the Wald test.

Estimates are given of the power of the tests using both estimation algorithms for the ANOVA, the Wald test and the LRT, for the design with 40 patients, in Table 2. For each test and each estimation method, the corrected power is computed using the corresponding empirical threshold in order to maintain a type I error of 5%. The corrected power for LRT and Wald tests based on the FO method were low, but the inflated type I error in these situations already show that this estimation method is poor. The powers for the ANOVA (using FO or FOCE) and the LRT for FOCE are around 70%, but the power of the Wald approach for FOCE is much lower (25%). We explored the FOCE outputs to understand this last result. In fact, with the FOCE algorithm in NONMEM, we observe correlations between the estimates of the gene effect coefficients and their estimation errors. This relationship leads to decreased values of the Wald statistic and therefore reduces the power to detect a genetic polymorphism effect. The same results were obtained considering only the subset of data sets fulfilling all convergence conditions for both FO and FOCE.

Model selection strategies

The results of the model selection strategies applied to the data sets simulated under H_0 with the design of 40 patients are presented in Table 3. Again, the performances of the FO algorithm are unsatisfactory except for the strategies based on EBE. With FOCE, both model selection strategies based on EBE and LRT select the correct model M_{base} in about 90% of instances whereas the Wald approach selects a model with a gene effect in around 15% of the data sets. Using selection criteria, the AIC and AICc obtain the worst performances, selecting a model with a gene effect in 57.8% and 51.7% of the data sets respectively. BICc shows performance close to that of the approach based on Wald tests, while CAIC (4.7%) and BIC (6.7%) select a model with a gene effect in about 5% of the data sets.

The results on data sets with N = 40 simulated under H_1 are given in Table 4. The results of the strategies using the LRT, Wald tests or a criterion with the FO estimation method are not presented because of the poor properties under H_0. A model with a gene effect is selected for about 70% of the data sets using CAIC, for about 80% of the data sets using the different model-building strategies based on tests or BIC, for about 90% of the data sets using BICc, and for about 99% of the data sets using AIC and AICc. However, for the latter, the percentage of data sets where M_{base} is not selected under H_0 is greater than 50%. Another noticeable result is that the model used to simulate the data M_{mult} is seldom selected compared with the intermediate model $M_{recessive}$, which is chosen in 30 to 50% of the data sets

using the different methods. The simulated value (β_1 is low compared with (β_2, therefore if the model-building strategies succeed in selecting a model with a gene effect, it is not always the correct one. The performances are similar using the sample of data sets for which all conditions were put together with FO and FOCE.

Figure 2 represents a summary of these results, placing each strategy on a bi-dimensional plan; the ability to select a model with a gene effect under H_1 versus the trend not to select M_{base} under H_0. The CAIC is the test which has the best properties under H_0, but the BIC more often selects a model with covariate under H_1 while it is only slightly less conservative under H_0. We also note that the LRT, the BIC and the ANOVA are clustered together, and thus offer similar compromises.

Application

Figure 3 represents a spaghetti plot of concentrations versus time for the 37 patients from the indinavir arm of the COPHAR2 study, sorted by genotype classes for exon 26. Concentrations show an important interindividual variability and only three patients were mutant homozygotes for this polymorphism.

As the simulation study has shown poor performances with the FO algorithm, we estimated the model parameters using the FOCE with interaction algorithm. Therefore, the estimates are different from the simulated values obtained in the preliminary analysis. In addition, the estimate of the interindividual variability for k was very small and we fixed it to 0 (no variability). The model with no covariate had an absorption constant of 0.8 h^{-1} with an important interindividual variability of 70.3%, an elimination constant of 0.2 h^{-1} and a volume of distribution of 99.3 L with an interindividual variability of 47%. All the estimation errors were below 20% for the fixed effects and below 40% for the variances.

For the influence of the ABCB1 exon 26 on the indinavir volume of distribution, the ANOVA and the LRT were non-significant (P = 0.7 and 0.2 respectively). The global Wald test obtained a p-value of 0.02, however the corrected threshold defined in the simulation study for this test is $7.5.10^{-3}$. The model with no covariate was chosen using the selection strategies based on the EBE, or on the LRT, using CAIC, BIC and BICc, whereas the strategy based on Wald tests, AIC and AICc selected $M_{recessive}$. No influence of the ABCB1 exon 21 on indinavir volume of distribution was detected using any of the tests or selection strategies under study.

DISCUSSION

In this work, we evaluate several statistical tests and model selection strategies using nonlinear mixed effects models to analyse the impact of a genetic polymorphism on one PK parameter through simulation. We also study the impact of the estimation algorithms in NONMEM, comparing the two first-order approximations most widely used FO and FOCE, which linearize the model function respectively around the random effects equal to 0 and around the individual estimates of the random effects. Although the FO method has been shown to suffer from various problems, it is still used because the FOCE algorithm is known for numerical difficulties and for its slowness. Using the FO algorithm we observe unsatisfactory performances for all the tests and model selection strategies with the exception of methods based on the EBE. The linearization of the likelihood function around the fixed effects leads to type I error inflation (Comets and Mentré, 2001, Wählby et al., 2001, Panhard and Mentré, 2005).

With FOCE, there is a significant increase in the type I error of the LRT and the Wald approach with a design including 40 patients. This increase has already been described for the LRT (Comets and Mentré, 2001, Wählby et al., 2001) and has been also shown for the Wald test (Panhard and Mentré, 2005). The design with 200 patients is closer to asymptotic conditions and shows as expected a correct type I error. Performing simulations under H_0 can be used to correct the threshold for the test under H_1 as we did in the present study. We observe a power around 70% for the tests using ANOVA and LRT which is close to the power of 80% expected from the simulation settings. Of course if the study was designed specifically to detect a gene effect the sample size could be increased to ensure a higher (more ethical) power. Our objective here was to compare different methods and we used the data from the COPHAR2-ANRS 111 trial to provide the settings for the PK simulations, as we have analysed these data thereafter. The simulated effect of gene was chosen to be consistent with the literature, though in the COPHAR2 trial the genotype distribution is slightly unbalanced. The reduced power of the Wald approach and the unsatisfactory efficacy of the model selection strategy based on the Wald tests could result partly from a wrong estimation of the standard errors due to the log-likelihood function linearization. Indeed, we observe with FOCE that in the simulations the estimation errors are highly correlated to their estimates. Because the Wald statistic is based on the ratio of the estimates to the estimation error, this could explain the poor performance of the Wald tests under H_1. Finally, FOCE met with convergence problems: on M_0 and $M_{CCvsCTvsTT}$ which involved no mathematical complexities for N = 40 under H_1, 35 (3.5%) and 48 (4.8%) runs did not achieve convergence, either for numerical reasons or because they had to be terminated and we could not obtain an estimate of the variance-covariance matrix in 32 (3.2%) and 38 (3.8%) runs, respectively. This could have been improved partly by the use of different initial conditions, nonetheless the results were identical using the sample of data sets fulfilling convergence conditions for the three tests using both estimation methods. Among the different implementations of both algorithms available, we chose to use NONMEM as it is the most popular tool in the pharmaceutical industry. Our results can be extended to the FO method implemented in SAS as it computes the same likelihood function

up to a constant (Wang, 2007), although it is not the same algorithm (Roe, 1997). Similarly, FOCE with interaction and the nlme method with the varConstPower option implemented in Splus should give the same results, save for the lower accuracy of the Splus approximation (Girard and Mentré, 2005, Wang, 2007). The simulations and analyses in the present paper have been performed using version 5.1 of NONMEM. A new version of NONMEM, version 6, has been released in December 2006, and was implemented in our department after the major part of the analyses had already been run, so that we kept NONMEM 5.1 for this study. The FOCE routine in NONMEM 6 has been rewritten and should provide more stable runs, which may reduce the convergence issues we have found in the present study. However, in our preliminary results with using NONMEM 6 on a subset of the simulated datasets, an increased number of runs failed the covariance-step. Rather than reporting partial results in the present study, we will investigate this matter in a subsequent work.

Model selection strategies based on tests have, to a certain extent, a high rate of false inclusion under H_0 (over 10%) which could result from the unconnected multiple model comparisons. Further there is no simple way to correct the model selection under H_1 by taking into account the simulations under H_0. Consequently, we have to be cautious about assessing the performance under H_1 when the behaviour is poor under H_0.

With respect to selection based on criteria, except for the performances of the BIC and the BICc not showing any trend to conservatism, under H_0 our results agree with the literature. The acceptable performance of the CAIC has to be noted, as well as the very poor performance of AIC and AICc. Under H_1, there is a satisfactory weak selection of M_{base} with a rather important representation of $M_{recessive}$. Indeed, in our simulation conditions, $M_{recessive}$ and M_{mult} are close and the power to detect a difference between these two models is much lower than the power to detect a difference between M_{base} and M_{mult}. As a side-note, in the Bayesian literature, it is usual also to consider not only the best model (i.e. the model with the lowest criterion), but also models close to the best model (Raftery, 1995). However in our study, the simulation model is rarely close in this sense to the best model.

Finally, in our study, for the design with 40 patients, the ANOVA on the EBE is the only test that maintained a 5% type I error, as well as a good power. It should be noted that we simulated a sufficient number of samples per patient with respect to the number of model parameters. In sparse sample situations, regression to the mean is known to occur with EBE (Panhard and Mentré, 2005), which could result in lower power to detect differences between genotype groups. We plan to test this hypothesis in a subsequent study with a sparse design including only two samples per patient.

We also confirm that both AIC and AICc should not be used for model building but further studies are required to provide recommendations on the other selection strategies. Regarding the estimation methods, if FO can still be used in covariate screening on the EBE, one should avoid performing model building with the LRT or Wald tests based on results from this algorithm.

We illustrate the different approaches using data from the indinavir PK sub-study of the COPHAR2 ANRS-111 trial. The PK model has already been described but no gene effect has been investigated using a population approach to date. The estimated parameters are in accordance with estimations obtained in other studies (Csajka et al., 2004, Goujard et al., 2005, Kappelhoff et al., 2005). The Wald test is the only test to detect an influence of ABCB1 exon 26 on the volume of distribution. Considering the corrected threshold provided by the simulation study, we can probably ascribe this discrepant result to the inflated type I error. Similarly, it is the three strategies with the highest percentage of data sets where M_{base} is not selected under H_0 that select a model with the ABCB1 exon 26 polymorphism as covariate. The polymorphism on the exon 26 of the ABCB1 gene has been shown to impact on plasma concentrations of nelfinavir, another protease inhibitor (Fellay et al., 2002). However, this work agrees with another study (Verstuyft et al., 2005), where no effect of the polymorphisms from ABCB1 exon 26 and 21 on indinavir bioavailability was found.

Another extension of this work would be to simulate under H_1 various levels of the gene effects, which would provide more information on the relationship between the strength of the genetic polymorphism effect and the power. Designing an optimal sampling schedule for testing a gene effect with a given power is also an interesting challenge and the extension of the PFIM software for design optimization in the case of models with covariates could be used (Retout and Mentre, 2003, Retout et al., 2007). Moreover in the genetic framework one should keep in mind the complex pathway leading from DNA to metabolic activity, which is usually controlled by more than one exon. Another perspective would therefore be to analyse the influence of the haplotypes (Innocentia et al., 2005), since such a classification seems to be more relevant at the DNA level.

Pharmacogenetic studies using NLMEM have many advantages as fewer samples are required to estimate parameters with a biological meaning. The current literature presents a wide array of methods for covariate selection using NLMEM. We show in this study that methods using EBE are not only efficient in data exploration but also in model selection on data sets with enough samples per patient. We also emphasize that using estimation algorithms based on likelihood linearization, LRT type I error is inflated, thus one has to perform simulations or work with large data sets. Finally, the problems of the FO and FOCE algorithm in terms of convergence and bias are an incentive to use more recent estimation methods (Samson et al., 2007).

Ackowledgements:

References:

- Akaike H 1974; A new look at the statistical model identification. IEEE Trans Automat Contr. 19: 716- 723
- Beal SL 2005; Conditioning on certain random events associated with statistical variability in PK/PD. J Pharmacokinet Pharmacodyn. 32: 213- 243
- Benjamini Y , Hochberg Y 1995; Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Roy Stat Soc B Met. 57: 289- 300
- Bodor M , Kelly EJ , Ho RJ 2005; Characterization of the human MDR1 gene. AAPS J. 7: E1- E5
- Bozdogan H 1987; Model selection and Akaike's information criteria (AIC): the general theory and its analytical extensions. Psychometrika. 52: 345- 370
- Brendel K , Legrand M , Taburet AM , Baron G , Goujard C , Mentre F 2005; Population pharmacokinetic analysis of indinavir in HIV-infected patient treated with a stable antiretroviral therapy. Fundam Clin Pharmacol. 19: 373- 83
- Burnham KP , Anderson DR 2002; Model Selection and Multimodel. A practical Information-Theoric approach. NY Springer-Verlag;
- Comets E , Mentré F 2001; Evaluation of tests based on individual vs population modelling to compare dissolution curves. J Biopharm Stat. 11: 107- 123
- Comets E , Verstuyft C , Lavielle M , Jaillon P , Becquemont L , Mentré F 2007; Modelling the influence of MDR1 polymorphism on digoxin pharmacokinetic parameters. Eur J Pharmacol. 63: 437- 449
- Csajka C , Marzolini C , Fattinger K , Decosterd LA , Telenti A , Biollaz J , Buclin T 2004; Population pharmacokinetics of indinavir in patients infected with human immunodeficiency virus. Antimicrob Agents Chemother. 48: 3226- 32
- Fellay J , Marzolini C , Meaden E , Back D , Buclin T , Chave J 2002; Response to antiretroviral treatment in HIV-1-infected individuals with allelic variants of the multidrug resistance transporter gene MDR1: a pharmacogenetic study. Lancet. 359: 30- 36
- Gabrielsson J , Weiner D 1999; Pharmacokinetic and Pharmacodynamic Data Analysis: Concepts and Applications. Stockholm Apotekarsocieteten;
- Girard P , Mentré F 2005; A comparison of estimation methods in nonlinear mixed effects models using a blind analysis. PAGE. 14:
- Goujard C , Legrand M , Panhard X , Diquet B , Duval X , Peytavin G , Vincent I , Katlama C , Leport C , Bonnet B , Salmon-Ceron D , Mentre F , Taburet AM 2005; High variability of indinavir and nelfinavir pharmacokinetics in HIV-infected patients with a sustained virological response on highly active antiretroviral therapy. Clin Pharmacokinet. 44: 1267- 78
- Henningsson A , Marsh S , Loos WJ , Karlsson MO , Garsa A , Mross K , Mielke S , Vigano L , Locatelli A , Verweij J , Sparreboom A , McLeod HL 2005; Association of CYP2C8, CYP3A4, CYP3A5, and ABCB1 polymorphisms with the pharmacokinetics of paclitaxel. Clin Cancer Res. 11: 8097- 8104
- Innocentia F , Liua W , Chenb P , Desaia AA , Dasb S , Rataina MJ 2005; Haplotypes of variants in the UDP-glucuronosyltransferase 1A9 and 1A1 genes. Pharmacogenet Genomics. 15: 295- 301
- Inomata S , Nagashima A , Itagaki F , Homma M , Nishimura M , Osaka Y , Okuyama K , Tanaka E , Nakamura T , Kohda Y , Naito S , Miyabe M , Toyooka H 2005; CYP2C19 genotype affects diazepam pharmacokinetics and emergence from general anesthesia. Clin Pharmacol Ther. 78: 647- 655
- Kappelhoff BS , Huitema AD , Sankatsing SU , Meenhorst PL , Van Gorp EC , Mulder JW , Prins JM , Beijnen JH 2005; Population pharmacokinetics of indinavir alone and in combination with ritonavir in HIV-1-infected patients. Br J Clin Pharmacol. 60: 276- 86
- Kerbusch T , Wählby U , Milligan P , Karlsson M 2003; Population pharmacokinetic modelling of darifenacin and its hydroxylated metabolite using pooled data, incorporating saturable first pass metabolism, CYP2D6 genotype and formulation-dependant bio availability. Br J Clin Pharmacol. 56: 639- 652
- Licinio J , Wong ML 2002; Pharmacogenomics The search for individualized therapy. Weinheim Wiley-VCH;
- Machin D , Campbell M , Fayers P , Pinol A 1997; Sample size tables for clinical studies. Oxford Blackwell Science;
- Mamiya K , Hadama A , Yukawa E , Ieiri I , Otsubo K , Ninomiya H , Tashiro N , Higuchi S 2000; CYP2C9 polymorphism effect on phenobarbiton kinetics in Japanese patients with epilepsy: analysis by population phamacokinetics. Eur J Clin Pharmacol. 55: 821- 825
- Marzolini C , Paus E , Buclin T , Kim RB 2003; Polymorphisms in Human MDR1 (P-glycoprotein): Recent advances and clinical relevance. Clin Pharmacol Ther. 75: 13- 33
- Meerum TJ , Malingre M , Beijnen J , Ten B , Huinink W , Rosing H , Koopman F 1999; Co-administration of cyclosporine A enables oral therapy with paclitaxel. Clin Cancer Res. 5: 3379- 84
- Mentré F , Duval X , Rey E , Peytavin G , Auleley S , Legrand M , Biour M , Bouxin A , Goujard C , Taburet A , Katlama C , Diquet B , Lascoux C , Tréluyer JM , Salmon-Céron D 2005; Prospective trial to evaluate how therapeutic drug monitoring of protease inhibitors increases virologic success and tolerance of HAART (COPHAR 2 - ANRS 111 trial). 12th Conference on Retroviruses and Opportunistic Infections Boston, USA
- Min DI , Ellingrod VL , Marsh S , McLeod H 2004; CYP3A5 polymorphism and the ethnic differences in cyclosporine pharmacokinetics in healthy subjects. Ther Drug Monit . 26: 524- 528
- Panhard X , Mentré F 2005; Evaluation by simulation of tests based on non-linear mixed-effects models in pharmacokinetic interaction and bioequivalence cross-over trials. Stat Med. 24: 1509- 1524
- R Development Core Team 2006; R: a language and environment for statistical computing. Vienna, Austria R Foundation for Statistical Computing;
- Raftery AE 1995; Bayesian Model Selection in Social Research (with discussion). Sociol Methodol. 111- 195
- Retout S , Comets E , Samson A , Mentre F 2007; Design in nonlinear mixed effects models: Optimization using the Fedorov-Wynn algorithm and power of the Wald test for binary covariates. Stat Med.
- Retout S , Mentre F 2003; Optimization of individual and population designs using Splus. J Pharmacokinet Pharmacodyn. 30: 417- 43
- Roe DJ 1997; Comparison of population pharmacokinetic modeling methods using simulated data: results from the Population Modeling Workgroup. Stat Med. 16: 1241- 57
- Sakaeda T , Nakamura T , Okumura K 2002; MDR1 Genotype-Related Pharmacokinetics and Pharmacodynamics. Biol Pharm Bull. 25: 1391- 1400
- Samson A , Lavielle M , Mentré F 2007; The SAEM algorithm for group comparison test in longitudinal data analysis based in nonlinear mixed-effects model. Stat Med.
- Schwartz G 1978; Estimating the dimension of a model. Ann Stat. 6: 461- 464
- Sheiner L , Beal S 1998; NONMEM Version 5.1. San Francisco University of California, NONMEM Project Group;
- Sugiura N 1978; Further analysis of the data by Akaike's information criterion and the finite corrections. Communication in Statistics Theory and methods. 7: 13- 26
- Taguchi M , Fujiki A , Iwamoto J , Inoue H , Tahara K , Saigusa K , Horiuchi I , Oshima Y , Hashimoto Y 2006; Nonlinear mixed effects model analysis of the pharmacokinetics of routinely administered bepridil in japanese patients with arrhythmias. Biol Pharm Bull. 29: 517- 521
- Verstuyft C , Marcellin R , Morand-Joubert L , Launay O , Brendel K , Mentré F , Peytavin G , Gérard L , Becquemont L , Aboulker JP 2005; Absence of association between MDR1 genetic polymorphisms, indinavir pharmacokinetics and response to HAART. AIDS. 19: 2127- 31
- Wählby U , Jonsson EN , Karlsson MO 2001; Assessment of actual significance levels for covariate effects in NONMEM. J Pharmacokinet Pharmacodyn. 28: 231- 252
- Wang Y 2007; Derivation of various NONMEM estimation methods. J Pharmacokinet Pharmacodyn. 34: 575- 93
- Yamaguchi H , Hishinuma T , Endo N , Tsukamoto H , Kishikawa Y , Sato M , Murai Y , Hiratsuka M , Ito K , Okamura C , Yaegashi N , Suzuki N , Tomioka Y , Goto J 2006; Genetic variation in ABCB1 influences paclitaxel pharmacokinetics in Japanese patients with ovarian cancer. Int J Gynecol Cancer. 16: 979- 985
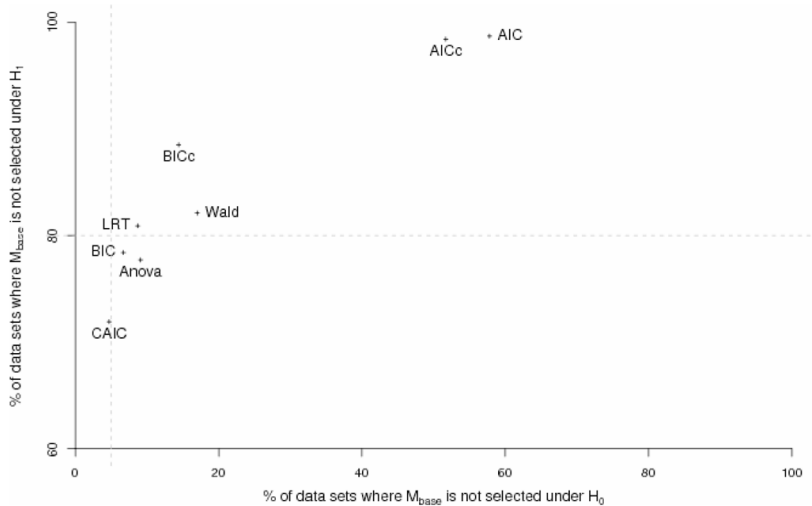
Figure 1

Decision path used to choose the best model for the selection strategies based on tests.
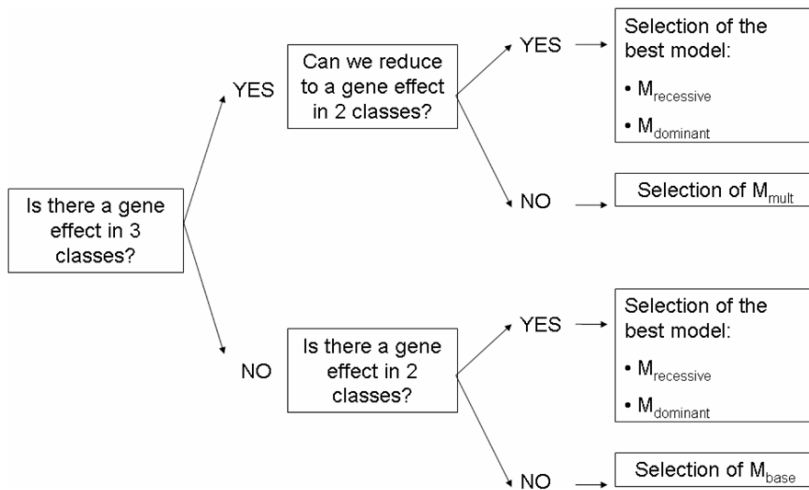
Figure 2

Percentage of data sets simulated under H_1 where the model with no gene effect (M_{base}) is not selected versus the same percentage under H_0 for the eight model selection strategies using FOCE for N = 40. The vertical line corresponds to a value of 5% and the horizontal line to a value of 80%.

Figure 3

Indinavir concentrations (ng/mL) at steady state collected in the COPHAR2 ANRS-111 trial versus time, sorted by ABCB1 exon 26 genotypes. The plain lines correspond to a dose of 400 mg indinavir, the dashed lines correspond to 600 mg and the dotted lines correspond to 800 mg, all with a dose of 100 mg ritonavir bid.
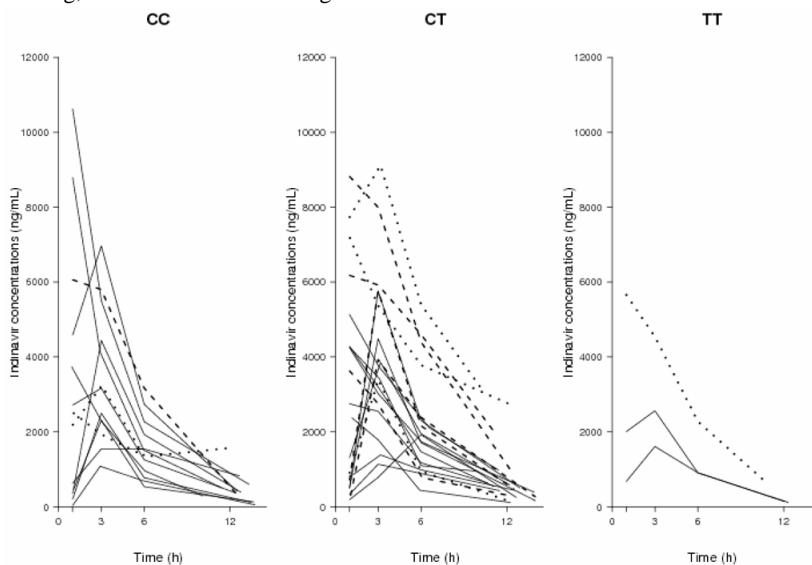
Table 1

Type I error for each test and for each algorithm.

Test	Algorithm	N = 40		N = 200	
		K	Type I error (%)	K	Type I error (%)
ANOVA	FO	997	5.9	1000	4.4
	FOCE	986	5.6	982	5.1
Wald	FO	975	23.4[*]	984	10.3[*]
	FOCE	924	11.7[*]	860	6.5[*]
LRT	FO	989	46.9[*]	977	54.0[*]
	FOCE	964	7.9[*]	956	5.0

K is the number of data sets on which the test could be performed.

[*] Estimate significantly different from 5%.

Table 2

Power for each test for N = 40.

Test	Algorithm	K	Power (%)	Corrected power (%)
ANOVA	FO	995	69.8	66.5
	FOCE	968	71.2	69.3
Wald	FO	974	61.5	7.7
	FOCE	905	57.2	24.7
LRT	FO	958	90.2	48.7
	FOCE	947	78.7	71.0

K is the number of data sets on which the test could be performed.

The corrected power was obtained using the fifth percentile of the empirical distribution of the test statistic under H_0 as the cut-off value for the test.

Table 3

Percentage of data sets simulated under H_0 with a design of $N = 40$ for which each model is selected.

Method	Algorithm	K	Model[1]			
			M_{base}	$M_{recessive}$	$M_{dominant}$	M_{mult}
ANOVA	FO	997	91.6	4.1	3.9	0.4
	FOCE	986	90.9	3.8	4.7	0.6
Wald	FO	947	68.4	11.1	16.3	4.2
	FOCE	876	83.0	5.8	9.8	1.4
LRT	FO	976	50.3	18.7	17.0	14.0
	FOCE	951	91.3	4.0	3.5	1.2
AIC	FO	999	14.9	23.2	22.4	39.5
	FOCE	970	42.2	22.4	21.3	14.1
AICc	FO	999	17.4	24.4	32.2	35.0
	FOCE	970	48.3	20.7	20.1	9.9
CAIC	FO	999	63.0	16.6	14.5	5.9
	FOCE	970	95.3	2.1	2.5	0.1
BIC	FO	999	55.7	19.6	16.6	8.1
	FOCE	970	93.3	3.1	3.0	0.6
BICc	FO	999	44.6	21.7	20.0	13.7
	FOCE	970	85.6	7.0	6.0	1.2

K is the number of data sets on which the test could be performed.

[1] M_{base}: $\{\beta_0 = \beta_1 = \beta_2 = 1\}$ (CC=CT=TT) model with no gene effect.
$M_{recessive}$: $\{\beta_0 = \beta_1 = 1, \beta_2 \neq 1\}$ (CC = CT ≠ TT), reduced model.
$M_{dominant}$: $\{\beta_0 = 1, \beta_1 = \beta_2 \neq 1\}$ (CC ≠ CT = TT), reduced model.
M_{mult}: $\{\beta_0 = 1, \beta_1 \neq \beta_2 \neq 1\}$ (CC ≠ CT ≠ TT), complete model.

Table 4

Percentage of data sets simulated under H_1 with N = 40 for which each model is selected.

Method	Algorithm	K	Model[1]			
			M_{base}	$M_{recessive}$	$M_{dominant}$	M_{mult}
ANOVA	FO	995	25.5	44.8	21.0	8.7
	FOCE	968	22.3	43.3	25.5	8.9
Wald	FOCE	878	17.9	41.7	32.7	7.7
LRT	FOCE	923	19.1	47.3	20.5	13.1
AIC	FOCE	962	1.3	31.1	13.1	54.5
AICc	FOCE	962	1.6	35.1	14.7	48.6
CAIC	FOCE	962	28.1	48.5	19.9	3.5
BIC	FOCE	962	21.6	50.3	21.7	6.4
BICc	FOCE	962	11.5	51.7	33.1	13.7

K is the number of data sets on which the test could be performed.

[1] M_{base}: {$\beta_0 = \beta_1 = \beta_2 = 1$} (CC=CT=TT) model with no gene effect.
$M_{recessive}$: {$\beta_0 = \beta_1 = 1, \beta_2 \neq 1$} (CC = CT ≠ TT), reduced model.
$M_{dominant}$: {$\beta_0 = 1, \beta_1 = \beta_2 \neq 1$} (CC ≠ CT = TT), reduced model.
M_{mult}: {$\beta_0 = 1, \beta_1 \neq \beta_2 \neq 1$} (CC ≠ CT ≠ TT), complete model.
Results obtained with FO are not presented for these strategies because of their poor performance under H0 (table 3).