Impact of modelling intra-subject variability on tests based on non-linear mixed-effects models in cross-over pharmacokinetic trials with application to the interaction of tenofovir on atazanavir in HIV patients.

Xavière Panhard, Anne-Marie Taburet, Christophe Piketti, France Mentré

▶ **To cite this version:**

Xavière Panhard, Anne-Marie Taburet, Christophe Piketti, France Mentré. Impact of modelling intra-subject variability on tests based on non-linear mixed-effects models in cross-over pharmacokinetic trials with application to the interaction of tenofovir on atazanavir in HIV patients.. Statistics in Medicine, 2007, 26 (6), pp.1268-84. 10.1002/sim.2622 . inserm-00157143

HAL Id: inserm-00157143

https://inserm.hal.science/inserm-00157143

Submitted on 23 Aug 2011

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

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

HAL author manuscript
Statistics in Medicine 15/03/2007; 26(6): 1268-84

HAL author manuscript inserm-00157143, version 1

Impact of modelling intra-subject variability on tests based on non-linear mixed-effects models in cross-over pharmacokinetic trials with application to the interaction of tenofovir on atazanavir in HIV patients

Xavière Panhard[1*], Anne-Marie Taburet[2], Christophe Piketti[3] and France Mentré[1]

July 2, 2007

[1] INSERM U738, Department of Epidemiology, Biostatistics and Clinical Research, University Hospital Bichat - Claude Bernard, AP-HP, Paris, France

[2] Clinical Pharmacy Department, Bicêtre Hospital, AP-HP, Le Kremlin-Bicêtre, France

[3] Department of Immunology, Georges Pompidou Hospital, AP-HP, Paris ,France

* Corresponding author : INSERM U738, Département d'Epidémiologie, Biostatistique and Recherche Clinique, Groupe Hospitalier Bichat - Claude Bernard, 46, rue Henri Huchard, 75018 Paris, France

Abstract

We evaluated the impact of modelling intra-subject variability on the Likelihood Ratio Test (LRT) and the Wald test based on non-linear mixed effects models in pharmacokinetic interaction and bioequivalence cross-over trials. In a previous study, these tests were found to achieve a good power but an inflated type I error when intra-subject variability was not taken into account. Trials were simulated under H_0 and several H_1 and analysed with the nlme function. Different configurations of the number of patients n and of the number of samples per patient J were evaluated for pharmacokinetic interaction and bioequivalence trials. Assuming

HAL author manuscript inserm-00157143, version 1

intra-subject variability in the model dramatically improved the type I error of both interaction tests. For the Wald test, the type I error decreased from 22%, 14% and 7.7% for the original (n=12, J=10), intermediate (n=24, J=5) and sparse (n=40, J=3) designs, respectively, down to 7.5%, 6.4% and 3.5% when intra-subject variability was modelled. The LRT achieved very similar results. This improvement seemed mostly due to a better estimation of the standard error of the treatment effect, related to a better estimation of the total and residual variability. Power was satisfactory for both tests. For bioequivalence trials, the type I error of the Wald test was 6.4%, 5.7% and 4.2% for the original, intermediate and sparse designs, respectively, when modelling intra-subject variability. We applied the Wald test to the pharmacokinetic interaction of tenofovir on atazanavir, a novel protease inhibitor. A significant decrease of the area under the curve and of the absorption duration of atazanavir were found when patients received tenofovir.

1 Introduction

Pharmacokinetic (PK) interaction and bioequivalence are usually tested in cross-over studies by comparing the $log(AUC)$ and the $log(C_{max})$ of the drug of interest R given alone (or with placebo) versus R given with an additional drug S, drug association that we call the test drug T. The standard method, recommended in guidelines [1, 2, 3], is the comparison of $log(AUC)$ and $log(C_{max})$ using a Student test or a Schuirmann's two-one sided test (TOST) to test interaction or absence of interaction, respectively. In order to take benefit of the knowledge accumulated on the PK of the studied drugs and to decrease the number of samples taken per patient, we proposed in a previous paper [4] to use tests based on non-linear mixed effects models (NLMEM) in cross-over trials, focusing on the comparison of the $log(AUC)$ between treatment groups. We proposed two approaches to test PK interaction, depending on whether concentration data of the two treatment groups are analysed separately or simultaneously. In the first case, we proposed to perform the paired parametric and non parametric tests between the two groups on the individual Empirical Bayes (EB) estimates of $log(AUC)$. In the second case, the likelihood ratio test (LRT) compared the results obtained with and without estimating an interaction effect on $log(AUC)$, and the Wald

HAL author manuscript inserm-00157143, version 1

test was directly performed on the interaction parameter. We adapted these approaches to test PK bioequivalence, except for the LRT. More precisely, we extended the method of the TOST to the EB tests and to the Wald test, the latter using the estimated standard error (SE) of the interaction parameter.

In that previous paper, we evaluated the type I error and power of these tests for both interaction and bioequivalence cross-over trials. The main conclusions were first that EB tests achieved satisfactory type I error and power for comparison tests, but were not of a great value for testing equivalence, and second that the LRT (except for equivalence) and the Wald test suffered from an inflation of their type I error up to 22 %, but that they were the most powerful, even after correction of the corresponding significance threshold.

It should be noted that in this previous study, intra-subject variability (also called inter-occasion variability) was not modelled, since the simulated one was small: 5% for $log(V/F)$ and 10% for $log(k_a)$ and $log(AUC)$. However, the observed inflation of the type I error for the LRT and the Wald test motivated the present study. Our objective was to evaluate the influence of modelling the intra-subject variability on the properties of the LRT and the Wald comparison test (used for PK interaction), and of the Wald equivalence test (used for bioequivalence). Because EB tests were based on a separate analysis of the two treatment groups, therefore not making any assumption on intra-subject variability, which was thus implicitely taken into account, they were not further studied. We applied this approach to the interaction of tenofovir on atazanavir concentrations data obtained in a PK substudy of the Puzzle 2 - ANRS 107 trial. Atazanavir is a novel protease inhibitor (PI) used for the treatment of HIV-1 infection that has recently be approved for patients previously treated with PIs, with a recommended dose of 300 mg QD associated with 100 mg of ritonavir. Its drug-drug interaction profile is not well established.

The statistical model and the tests are described in section 2. The simulation settings used to evaluate the properties of the tests are described in section 3, and the main results of this simulation

HAL author manuscript inserm-00157143, version 1

study are exposed in section 4. In section 5, we applied the Wald test based on NLMEM to PK interaction of tenofovir on atazanavir in HIV patients. The overall results are discussed in section 6.

2 METHODS

2.1 Model and notations

We studied a two period two sequence cross-over PK trial. The two treatments were compared through the ratio of the geometric mean of the area under the curve. We assumed a period effect equal to zero. Since it is not recommended to take a carry-over or a time by period interaction [5], there is no need to take account of the order of administration. The analysis is therefore performed as for a two period one sequence cross-over. Concentrations are noted y_{ijk} where k is the treatment ($k = R, T$), i is the subject ($i = 1, \cdots, n$) and j is the measure performed for that subject at time t_{ijk}, ($j = 1, \cdots, J$). We used identical sampling times for each group and for all subjects so that for all j in $(1, \cdots, J)$, $t_{ijk} = t_j$, ($i = 1, \cdots, n$), ($k = R, T$) and the same dose D for all patients. We assumed that we knew a non-linear function f such as the statistical model linking concentrations to sampling times can be written as:

$$y_{ijk} = f(t_j, \theta_{ik}) + \varepsilon_{ijk} \tag{1}$$

where θ_{ik} is the vector of the parameters of subject i for treatment k and ε_{ijk} is the measurement error. The errors ε_{ijk} given θ_{ik} were assumed independent and normally distributed with a null mean and an heteroscedastic variance σ^2_{ijk}, with:

$$\sigma^2_{ijk} = \sigma^2(a + f(t_j, \theta_{ik}))^b \tag{2}$$

This combined error model (additive and proportional) is commonly used in population pharmacokinetics with b fixed to 2. We made the hypothesis that the individual parameters θ_{ik} are random

HAL author manuscript inserm-00157143, version 1

vectors and that for each treatment k, θ_{ik} can be decomposed as:

$$\theta_{ik} = \mu_k + b_i + c_{ik} \tag{3}$$

where θ_{ik}, μ_k, b_i and c_{ik} are vectors of size p; μ_k is the mean value for treatment k and b_i is vector of the random effect of subject i, which is assumed to be distributed as a normal distribution with zero mean and variance Ψ. The vector of random effect c_{ik} models, for each subject i, the variability between the two periods of treatment, also called intra-subject variability or inter-occasion variability. We assumed that c_{ik} is distributed according to a normal distribution with zero mean and variance Γ. The total variance Ω of the PK parameters is thus decomposed between the inter-subject variance Ψ and the inter-occasion variance Γ. In this study, Ψ and Γ were assumed diagonal. The parameters θ are the logarithm of the usual pharmacokinetic parameters. The standard deviations of the random effects are the square roots, noted ψ_ℓ ($\ell = 1, \cdots, p$) and γ_ℓ ($\ell = 1, \cdots, p$) of the diagonal elements of Ψ and Γ, respectively.

We study tests based on $log(AUC)$ which is therefore a component of θ. In this paper we restricted ourselves to the case were the data are analysed in a global analysis performed by adding a vector of treatment effect noted β on the vector of parameters. We therefore assumed:

$$\mu_T = \mu_R + \beta \tag{4}$$

We note β_{AUC} and $\mu_{k,AUC}$ respectively the component of β and μ_k corresponding to $log(AUC)$.

2.2 Comparison tests

The null hypothesis tested by a comparison test is: $H_0 : \{\mu_{T,AUC} - \mu_{R,AUC} = 0\}$, i.e. $H_0 : \{\beta_{AUC} = 0\}$. The first test based on a global analysis is the likelihood ratio test (LRT). Let M_1 be the model including a treatment effect on all components of θ except AUC, and M_2 the model where the treatment effect on AUC is added. The respective log-likelihoods of M_1 and M_2 are noted L_1 and L_2. Under the null hypothesis $H_0 : \{\beta_{AUC} = 0\}$, $2(L_2 - L_1)$ is asymptotically

HAL author manuscript inserm-00157143, version 1

distributed as a χ^2 with one degree of freedom, which is the assumption we made here. The other global test is the Wald test comparing the treatment effect to zero in the model M_2 [6]. Let $\hat{\beta}_{AUC}$ be the estimator of the treatment effect on $log(AUC)$ in M_2 and $SE_{\beta_{AUC}}$ its standard error. Under H_0, the statistic $t = \hat{\beta}_{AUC}/SE_{\beta_{AUC}}$ is distributed as a Student distribution with $m = n(2J-v)-q$ degrees of freedom, where q is the number of fixed effects estimated in the model and v is equal to 2 if the intra-individual variability is estimated and 1 if not. The null hypothesis is rejected if $|t|$ is greater than the 97.5^{th} quantile of a Student distribution with m degrees of freedom.

2.3 Equivalence tests

The null hypothesis of an equivalence test is: $\{(\mu_{T,AUC}-\mu_{R,AUC}) < -\delta$ or $(\mu_{T,AUC}-\mu_{R,AUC}) > \delta\}$ were δ belongs to $]0;1[$. δ is usually taken equal to 20%, leading to a rounded equivalence interval of $[0.8; 1.25]$ for the ratio of the geometric means of the untransformed AUC, as proposed in the FDA guidelines [3]. H_0 can also be formulated as $\{\beta_{AUC} < -\delta$ or $\beta_{AUC} > \delta\}$.

The test recommended by the FDA guideline leads to the rejection of H_0 if the 90% confidence interval of $(\mu_{T,AUC} - \mu_{R,AUC})$, calculated on non-compartmental $log(AUC)$, belongs to $[-\delta; \delta]$. We did not develop a test based on loglikelihood for the case of equivalence tests, as no simple extension of the LRT for that kind of composite null hypothesis exists. The extension of this approach for the Wald test is to reject H_0 if the 90% confidence interval of $\hat{\beta}_{AUC}$ is included in $[-\delta; \delta]$.

3 EVALUATION OF THE TESTS BY SIMULATION

3.1 Simulation settings

We used the same theophyllin data set than for the previous study to simulate concentration data. A one compartment model with first order absorption and elimination, with a combined error model, was fitted to the data using using Lindström and Bates' algorithm [7], implemented in the nlme function of R 2.0 software [8]. We fixed the dose for all subjects to the rounded median dose (4 mg) of the above study. We used the values of the estimated parameters estimates for

HAL author manuscript inserm-00157143, version 1

the simulations. The estimated values for the fixed effects were $\hat{\mu}_{V/F} = -0.73$, $\hat{\mu}_{k_a} = 0.39$ and $\hat{\mu}_{AUC} = 4.61$. We kept the same value of the mean parameters for treatment R and modify the mean of $log(Cl/F)$ and $log(V/F)$ for treatment T in order to mimic modifications of bioavailability. For inter-subject variabilities, we chose ψ_{k_a}=ψ_{AUC}=0.2 and $\psi_{V/F}$=0.1. The intra-subject variability was chosen equal to half the inter-subject variability, i.e. γ_{k_a}=γ_{AUC}=0.1 and $\gamma_{V/F}$=0.05. The corresponding global standard deviation of the PK parameters were therefore ω_{k_a}=ω_{AUC}=0.224 and $\omega_{V/F}$=0.112. The combined error model was similar to that of the parent study, with σ rounded to 0.1, $a = 1$ and b fixed to 2. Data sets were simulated using the same method than for the previous study, using R 2.0 [4]. Data were simulated for the original design (n=12, J=10), an intermediate design (n=24, J=5) and a sparse design (n=40, J=3). In order to study the evolution of the type I error when n increases for a fixed J, we also simulated datasets with n equal to 24, 40 and 60 with J=10. We also simulated datasets for the original design (n=12, J=10) with no simulated intra-individual variability.

3.2 Evaluation of type I error and power

The LRT (only for comparison tests) and the Wald test were evaluated on trials simulated under the null hypothesis H_0 and under several alternatives H_1, in both interaction and bioequivalence case. All data of both treatment groups of each simulated trial were analysed using two different NLME models: a global model M_1 of both treatment groups with a treatment effect on k_a and V but not on AUC and a global model M_2 with a treatment effect added on AUC. Intra-subject variability was assumed in M_1 and M_2. We also constructed the analog models $M_1^{'}$ and $M_2^{'}$ where no intra-individual variability was assumed. More precisely, for M_2, we estimated the three fixed effects $log(k_a)$, $log(AUC)$ and $log(V/F)$, the three corresponding treatment effects β_{k_a}, β_{AUC} and β_V, the diagonal element of Ψ, i.e. the inter-subject variances, ψ_{k_a}, ψ_{AUC} and ψ_V, the diagonal element of Γ, i.e. the intra-subject variances , γ_{k_a}, γ_{AUC} and γ_V. For M_1, β_{AUC} was not estimated. $M_1^{'}$ and $M_2^{'}$ correspond to M_1 and M_2 with no estimation of γ_{k_a}, γ_{AUC} and $\gamma_{V/F}$. The LRT compares the loglikelihood of M_1 and M_2, i.e. when modelling intra-subject variability, and, for comparison purpose, of $M_1^{'}$ and $M_2^{'}$, i.e. when not modelling intra-subject variability. The Wald test compares

HAL author manuscript inserm-00157143, version 1

the treatment effect on $log(AUC)$ in M_2 (and $M_2^{'}$) to zero.

For comparison tests, 5000 interaction trials were simulated under H_0 in order to evaluate the type I error of each of the two studied tests, leading to a 95% prediction interval of $[4.4\%; 5.6\%]$ around the expected type I error of 5%. The type I error is estimated by the proportion of simulated trials for which H_0 is rejected. The type I error was estimated only for comparison tests for the designs were n equals 24, 40 and 60 with J=10, and for the case where no intra-individual variability was simulated. Several alternative hypothesis were evaluated by multiplying, for treatment T, the fixed effects for both Cl/F and V/F of treatment R by a factor e^{δ}. In this precise case, a modification of Cl/F corresponds to the same modification of AUC since the dose is the same for all patients. Trials are simulated for 4 different values of δ: -0.223,-0.105, 0.105, and 0.223, and for the original design only. These values correpond to a ratio of the geometric means of the untransformed AUC of 0.8, 0.9, 1.11 and 1.25, respectively. For each alternative, the power was estimated from 1000 simulated trials by the proportion of studies for which a significant difference is shown. As in the previous study, we used a correction of the significance threshold for power evaluation [4].

In the case of equivalence, the null hypothesis is composite; it has been shown that the global type I error is the supremum of the type I error over the null space [9]. As previously, we therefore simulated 5000 trials under each unilateral null hypothesis by multiplying the fixed effects for $log(Cl)$ (and therefore $log(AUC)$) and $log(V)$ by $\delta = -0.223$ or $\delta = 0.223$ respectively for treatment T. These values correpond to a ratio of the geometric means of the untransformed AUC of 0.8 and 1.25, respectively. For each case, the unilateral type I error rate of the Wald test was estimated by the proportion of studies for which the 90% confidence interval of $\hat{\beta}_{AUC}$ is comprised in [-0.223; 0.223]. The global type I error rate was estimated by the maximum of these two "unilateral" proportions. To evaluate the power, only for the original design, 1000 trials were simulated for 3 values of δ: -0.105, 0, and 0.105 respectively. These values correpond to a ratio of the geometric means of the untransformed AUC of 0.9, 1 and 1.11, respectively. As before, the estimation of the power was corrected by using for each unilateral null hypothesis the empirical threshold corresponding to a type I error of 5%.

HAL author manuscript inserm-00157143, version 1

3.3 Impact of modelling intra-subject variability on parameters estimation

We evaluated performances of parameter estimation under H_0 with and without modelling intra-subject variability in the case of PK interaction trials, in order to better understand the impact of modelling intra-subject variability. We calculated the bias and root mean square error (RMSE) for the treatment effect on $log(AUC)$, for the inter-subject and intra-subject standard deviations ψ_{AUC} and γ_{AUC} of $log(AUC)$, for the total standard deviation of $log(AUC)$ $\omega_{AUC} = \sqrt{\psi^2_{AUC} + \gamma^2_{AUC}}$, and for the residual error σ for the three studied design and for the two types of analysis (with and without estimation of the intra-subject variability).

We also estimated, for each design and each type of analysis, the empirical standard deviation $SD_{\hat{\beta}_{AUC}}$ on the 5000 estimated β_{AUC}. We considered it as the true standard error of β_{AUC}. For each trial h ($h = 1, \cdots, 5000$), we formed the ratio $r_h = \frac{SE_{\hat{\beta}_{AUC,h}}}{SD_{\hat{\beta}_{AUC}}}$. For each dataset h, the standard error $SE_{\hat{\beta}_{AUC,h}}$ is the diagonal element of the square root of the approximate variance-covariance matrix of the fixed effects provided by nlme [8]. The mean of the distribution of the r_h should be close to 1 if the standard error provided by nlme is correclty estimated.

4 RESULTS

4.1 Comparison tests

The type I error of the Wald test and of the LRT with and without modelling intra-subject variability are displayed in figure 1(A) for the original, intermediate and poor designs and in figure 1(B) for the designs with J fixed to 10. The type I errors of the Wald test on the 5000 replications of interaction cross-over trials were 7.5%, 6.4% and 3.5% in the original, intermediate and sparse designs, respectively, when the intra-subject variability was estimated, which were much closer to 5% than when intra-individual variability was not taken into account, with results of 22%, 14% and 7.7%. Very similar results were observed for the LRT. The type I error was found to be closer

HAL author manuscript inserm-00157143, version 1

to 5% when n increases with J fixed to 10. The estimation was even in the confidence interval of the nominal level of 5% for n equal to 40 and 60 for the Wald test, and for n equal to 60 for the LRT. In the case were no intra-subject variability was simulated, the two tests achieved a too conservative type I error (3.8% and 3.7% for the Wald test and the LRT, respectively) when not modelling intra-subject variability. These type I errors further decreased to 3.5% and 3.2%, respectively, when modelling intra-subject variability.

The powers of the two tests with and without modelling intra-individual variability for the original design are presented on table I. Power was satisfactory for both tests and for the three considered designs for $\delta = \pm 0.223$. The power for $\delta = \pm 0.105$ is around 40% for both tests and both methods. These values of δ do not correspond to realistic alternative hypotheses, but they were tested to observe whether the power curves obtained for the two methods were comparable, which was the case here. However, the power obtained when modelling intra-subject variability was always slightly smaller than that obtained without modelling this variability, even with correction of the significance threshold.

The bias and RMSE for β_{AUC}, ψ_{AUC}, γ_{AUC}, ω_{AUC} and σ for the three studied design and for the two types of analysis (with and without estimation of the intra-subject variability) are displayed on table III. No clear improvement of the bias and RMSE was found when modelling intra-subject variability. The only exception was for the bias on σ and for the bias and RMSE on the variance parameter a, which were smaller for the original and intermediate designs when modelling intra-subject variability. When γ_{AUC} was not estimated, the corresponding bias and RMSE were fixed to the true value of the parameter, i.e. 0.1. For the sparse design, the bias and RMSE on all parameters were very close between the two methods. The boxplot of the estimated γ_{AUC} for the 3 different designs, diplayed in figure 2, shows that the intra-subject variability was in fact often estimated very close to zero for the sparse design. The median of the 5000 estimated γ_{AUC} was indeed equal to 10^{-4}.

The boxplots of the ratios $r_h = \frac{SE_{\hat{\beta}_{AUC,h}}}{SD_{\hat{\beta}_{AUC}}}$ $(h = 1, \cdots, 5000)$ obtained under H_0 in the case of comparison for the three designs are displayed in figure 3.A, together with the corresponding $\hat{\beta}_{AUC}$ (figure 3.B). When modelling intra-subject variability, r was closer to 1 for the original

HAL author manuscript inserm-00157143, version 1

and intermediate designs. The ratio obtained when not modelling intra-subject variability was systematically underestimated for these two designs, the bigger difference being for the original design. As the corresponding β_{AUC} were correctly estimated, this might explain the inflation of the type I error observed for these two designs. For the sparse design, both methods achieved similar results, with a slightly inflated ratio, which might explain why the corresponding type I errors are slighlty inferior to 5%.

4.2 Equivalence tests

The type I error of the Wald test with and without modelling intra-subject variability are displayed in figure 4 for the 3 considered designs for the equivalence tests using the extension of the TOST. The type I errors of the Wald test on the 5000 replications of bioequivalence cross-over trials were found to be 6.4%, 5.8% and 4.7% for the original, intermediate and sparse designs, respectively. These type I errors are also closer to 5% than those obtained without modelling intra-subject variability, i.e. 13.2%, 9.4% and 5.4%. As for comparison tests, the bigger difference between the two methods was found for the original design, followed by the intermediate one. The type I error obtained for the sparse design are very close for the two methods.

The power of the Wald test with and without modelling intra-subject variability for the original design is presented on table II. Power was satisfactory for the three considered alternative hypothesis.

5 APPLICATION TO THE INTERACTION OF TENOFOVIR ON ATAZANAVIR PHARMACOKINETICS

5.1 Study population

The Puzzle 2 - ANRS 107 trial was a randomized open-label, multiple-dose study performed in HIV-infected patients who had failed previous antiretroviral therapy. A pharmacokinetic substudy was performed at the beginning of the trial in 11 patients. All these patients gave written informed consent to participate in the study and the pharmacokinetic study, which was approved by the

HAL author manuscript inserm-00157143, version 1

Institutional Review Board of Hopital Saint-Antoine, Paris VI University. HIV-infected adults were eligible for inclusion if they met the following criteria: no change in antiretroviral treatment within the last month before inclusion, plasma HIV-1 RNA loads greater than 10 000 copies/ml, documented failure of previous treatment with at least two PIs and one non-nucleoside reverse transcriptase inhibitor and the absence of cardiomyopathy or conduction system disease. Patients were randomized to receive for the first two weeks either unchanged treatment with PIs and nucleoside reverse transcriptase inhibitors (NRTIs) (group 1) or unchanged treatment with NRTIs in combination with atazanavir (300 mg QD) and ritonavir (100 mg QD) as a substitute for the failing PI therapy (group 2). For week 3 (day 15) to week 26, patients from either group switched to atazanavir (300 mg QD) and ritonavir (100 mg QD) plus tenofovir disoproxil fumarate (DF) (300 mg QD) and NRTIs selected according to the baseline reverse transcriptase genotype of the HIV isolate infecting each patient.

The pharmacokinetic substudy of ANRS trial 107 was conducted in 11 HIV-infected patients in group 2. The objective of the study was to measure the pharmacokinetic parameters of atazanavir (administered with ritonavir) either before (day 14 [week 2]) or after (day 42 [week 6]) initiation of tenofovir DF in HIV-infected patients in order to detect pharmacokinetic interactions of tenofovir on atazanavir. A significant decrease of the AUC estimated by a non-compartmental approach was found with the addition of tenofovir to atazanavir [10]. One patient did not complete the study because of the occurrence of asymptomatic ventricular bigeminy during sampling for the pharmacokinetic study on day 14. The plasma drug concenytations of this patient were therefore excluded from the analysis. We analysed the data of this substudy with a non-linear mixed effect model, in order to determine which PK parameter is influenced by the addition of tenofovir.

5.2 PK samples and concentration measurement

Blood samples for atazanavir and ritonavir pharmacokinetic assessments were collected on days 14 (period R, week 2) and 42 (period T, week 6) from the 10 patients. Samples were drawn prior to drug intake in the morning and then after drug administration at times of 1, 2, 3, 5, 8, and 24

HAL author manuscript inserm-00157143, version 1

h. The actual times of drug administration and samplings were recorded for samples taken after drug administration, whereas the dosing interval recorded for the sample taken before drug intake was declared by the patient and was therefore subject to an important uncertainty. We chose to exclude it from the analysis.

5.3 Population PK model

We used a one-compartment model with zero-order absorption and first order elimination to describe atazanavir concentrations. The parameters of this model are F, the bioavailability, the volume of distribution of atazanavir V, the absorption duration (T_a) and the elimination rate constant k. The equation of this model after a single administration is:

$$g(\theta, t) = \frac{FD}{T_a V k}(1 - e^{-kh(t)})e^{-k(t-h(t))} \tag{5}$$

where :

$$h(t) = \begin{cases} t & \text{when } t < Ta \\ Ta & \text{when } t \geq Ta \end{cases} \tag{6}$$

From single dose equations, we derived equations at steady state, where we used a dosing interval τ of 24 hours until the day of the PK visit. The equation used to fit the model to the data was therefore:

$$f(\theta, t) = \frac{FD}{T_a V k}\left((1 - e^{-kt})\mathbb{1}_{t<T_a} + \frac{e^{-k\tau \mathbb{1}_{t<T_a}}(1 - e^{-kT_a})e^{-k(t-T_a)}}{(1 - e^{-k\tau})}\right) \tag{7}$$

The model was then reparametrized in T_a, V and AUC using $AUC = \frac{DF}{kV}$. Since atazanavir was orally administered, only T_a, V/F and AUC were identifiable. The vector θ of the PK parameters was composed of $\log(T_a)$, $log(AUC)$ and $log(V/F)$. As for the simulation study, concentrations data were analysed using the nlme funciton of R 2.0. For each PK parameter, we assumed a random effect modelling inter-subject variability, a random effect modelling intra-subject variability and a treatment effect. Regarding the error model, we tested a homoscedastic error model, a constant CV error model and a combined error model.

HAL author manuscript inserm-00157143, version 1

5.4 Results

Concentration data of both visits are displayed in figure 5. The best fit and the smallest Akaïke criterion were obtained for the homoscedastic error model, which was therefore kept for the final analysis. When analysing concentration data using the full model, i.e. with estimation of all treatment effects and all inter and intra-subject variances, the estimated inter-patient variability on $log(V/F)$ was inferior to 0.01. We therefore fixed it to zero, which resulted in a better AIC. In this model, intra-subject variabilities were inferior to 0.01 for $log(AUC)$ and $log(T_a)$. Similarly, the AIC was better when these two variabilities were fixed to zero. No other variance was inferior to 0.01 in the resulting model, which was therefore kept as the final model.

A significant effect of comedication with tenofovir was found on $log(AUC)$ ($p<10^{-4}$) and $log(T_a)$ ($p=0.0019$), resulting in a decrease of 1.46 fold of AUC and an increase by 1.45 fold of T_a, respectively, when patients received TFV. We also derived the 90% confidence interval for the three treatment effects. When taking an equivalence interval of [-0.2 ; 0.2], equivalence between the two treatment periods is rejected for the three PK parameters, even for $log(V/F)$, for which no significant interaction was found. Parameters estimates are displayed in table III with their 90 % confidence intervals. The curves of the population parameters corresponding to the PK of atazanavir with and without tenofovir are overlayed on the plot of concentrations data of figure 5. The goodness-of-fit plots (population and individual predicted concentrations versus observed concentrations; standardized residuals versus predicted concentrations and versus time) were judged satisfactory, and are displayed on figure 6.

6 Discussion

This study confirmed that NLMEM are a good and useful approach to test PK interaction or bioequivalence in cross-over trials. Indeed, the Wald test and the LRT achieved adequate type I error when taking intra-subject variability into account in the model, and also achieved an adequate power.

We observed, for comparison tests, that modelling intra-subject variability prevented the under-

HAL author manuscript inserm-00157143, version 1

estimation of standard error of the treatment effect, essentially for the original and intermediate designs. This might be due to a better estimation of the global variability of $log(AUC)$, and can explain the accuracy of the corresponding type I errors for the Wald test, and also for the LRT, since it behaved almost exactly like the Wald test in this study. For the sparse design, as intra-subject variability was estimated close to 0, both methods achieved very similar parameter estimates, and therefore almost equal type I errors.

As far as regulatory considerations are concerned, the guidelines of the FDA and of the EMEA recommend to test PK interaction and bioequivalence based on both AUC and C_{max}. In this study, we focused on tests based on $log(AUC)$, but the proposed methods can be extended to $log(C_{max})$. If $log(C_{max})$ cannot be taken as a component of θ, the delta method can be used to derive the standard error of $log(C_{max})$, and the corresponding Wald test can therefore be performed.

Our main motivation for this work was to study drug-drug interactions between antiretrovirals used in HIV-infected patients. Some of these interactions are tested in PK trials performed in healthy volunteers. However, an important concern is that the pharmacokinetics of antiretroviral drugs seem to be different in patients, partly because of a modification of absorption characteristics during HIV-infection [11, 12, 13]. The possibility of performing PK interaction and bioequivalence trials in patients, requiring a reduced number of samples per subject, is therefore of a great importance. The application to the test of the interaction of tenofovir on the PK of atazanavir confirmed the impact of tenofovir on the absorption of atazanavir, more precisely a decrease of the AUC and an increase of T_a. The mechanism of this drug interaction has not been completely elucidated yet, but the two major hypothesis are the induction by tenofovir of the human multidrug resistance P-glycoprotein, which plays a key role in reducing the bioavailability of PIs, and/or a physicochemical interaction between atazanavir and tenofovir in the gut [10].

Tests based on NLMEM allow both to test PK interaction or lack of interaction while greatly decreasing the number of samples per patient. This point is of great interest when performing such trials in patients, for instance in HIV patients, as illustrated here, or in special populations (children, older patients). Our next step is the design of such PK interaction studies. Since the expected standard error of the interaction effect can be estimated when taking intra-subject

HAL author manuscript inserm-00157143, version 1

variability into account using an extension of PFIM [14], the power of the interaction or lack of interaction test can be derived and thus the sample size for a given power. We also aim at proposing an adaptation of the LRT based on non-linear mixed effects models for the case of equivalence.

We thank the Puzzle2 - ANRS 107 study team (investigator: Dr Piketti, pharmacology: Dr Taburet, methodology: Dr Aboulker), who allowed us to use atazanavir concentrations data of the PK substudy for modelling purpose.

We thank Pr Stephen Senn, Dr Nicolas Frey and the reviewer for their useful comments.

References

[1] EMEA. Note for guidance on the investigation of bioavaibility and bioequivalence. Technical report, EMEA, 2000.

[2] EMEA. Note for guidance on the investigation of drug interactions. Technical report, EMEA, 1998.

[3] FDA. Statistical approaches to establishing bioequivalence. Technical report, FDA, 2001.

[4] Panhard X, Mentré F. Evaluation by simulation of tests based on non-linear mixed-effects models in pharmacokinetic interaction and bioequivalence cross-over trials. *Stat Med* 2005; **24**(10):1509–1524.

[5] Senn S. *Cross-over trials in clinical research*. John Wiley and Sons, Chichester, 2002.

[6] Kowalski K, Hutmacher M. Efficient screening of covariates in population models using Wald's approximation to the likelihood ratio test. *J Pharmacokinet Pharmacodyn* 2001; **28**(3):253–75.

[7] Lindstrom M, Bates D. Nonlinear mixed effects models for repeated measures data. *Biometrics* 1990; **46**(3):673–87.

[8] Pinheiro J, Bates D. *Mixed-effect models in S and Splus*. Springer-Verlag, New York, 2000.

[9] Berger L, Hsu J. Bioequivalence trials, intersection-union tests and equivalence confidence sets (with discussion). *Statist Sci* 1996; **11**:283–319.

HAL author manuscript inserm-00157143, version 1

[10] Taburet AM, Piketty C, Chazallon C, Vincent I, Gérard L, Calvez V, Clavel F, Aboulker JP, Girard PM. Interactions between atazanavir-ritonavir and tenofovir in heavily pretreated human immunodeficiency virus-infected patients. *Antimicrob Agents Chemother* 2004; **48**(6):2091–6.

[11] Mouly S, Aymard G, Diquet B, Caulin C, Bergmann J. Oral ganciclovir systemic exposure is enhanced in HIV-infected patients with diarrhea and weight loss. *J Acquir Immune Defic Syndr* 2000; **24**(4):344–51.

[12] Mouly S, Aymard G, Tillement J, Caulin C, Bergmann J, Urien S. Increased oral ganciclovir bioavailability in HIV-infected patients with chronic diarrhoea and wasting syndrome–a population pharmacokinetic study. *Br J Clin Pharmacol* 2001; **51**(6):557–65.

[13] Trout H, Mentré F, Panhard X, Kodjo A, Escaut L, Pernet P, Gobert JG, Vittecoq D, Knellwolf AL, Caulin C, Bergmann JF. Enhanced saquinavir exposure in human immunodeficiency virus type 1-infected patients with diarrhea and/or wasting syndrome. *Antimicrob Agents Chemother* 2004; **48**(2):538–45.

[14] Retout S, Mentré F. Further developments of the Fisher information matrix in nonlinear mixed effects models with evaluation in population pharmacokinetics. *J Biopharm Stat* 2003; **13**(2):209–27.

HAL author manuscript inserm-00157143, version 1

Test	Estimation of intra-subject variability	δ -0.223	-0.105	0.105	0.223
LRT	Yes	0.94	0.36	0.39	0.95
	No	0.98	0.40	0.41	0.98
Wald	Yes	0.97	0.38	0.41	0.96
	No	0.98	0.43	0.43	0.98

Table 1: Power of the comparison tests estimated from 1000 simulated interaction trials under several alternative hypothesis ($\delta \neq 0$) with corrected thresholds, for the original design (n=12, J=10).

Test	Estimation of intra-subject variability	δ -0.105	0	0.105
Wald	Yes	0.75	0.97	0.73
	No	0.79	0.99	0.75

Table 2: Power of the Wald test for equivalence estimated from 1000 simulated bioequivalence trials under several alternative hypothesis ($\delta \neq -0.223$ or 0.223) with corrected thresholds, for the original design (n=12, J=10).

HAL author manuscript inserm-00157143, version 1

Design		Estimation of	β_{AUC}		ψ_{AUC}		γ_{AUC}		ω_{AUC}		σ		a	
n	J	intra- individual variability	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
12	10	Yes	-0.000830	0.0514	-0.0200	0.057	-0.018	0.045	-0.019	0.048	-1.87	16.68	0.75	1.35
		No	-0.000637	0.0525	-0.0063	0.049	0.100[a]	0.100[a]	-0.030	0.057	-3.91	16.93	1.18	1.64
24	5	Yes	0.000367	0.0369	-0.0097	0.037	-0.025	0.041	-0.016	0.036	0.87	16.64	0.67	1.13
		No	0.000564	0.0371	-0.0023	0.034	0.100[a]	0.100[a]	-0.026	0.043	-1.72	15.33	1.10	1.42
40	3	Yes	-0.000013	0.0317	-0.015	0.033	-0.041	0.087	-0.035	0.045	3.37	20.24	1.18	1.64
		No	-0.000113	0.0317	-0.015	0.033	0.100[a]	0.100[a]	-0.038	0.048	-4.87	19.96	1.18	1.57

[a] Fixed to the true value of γ_{AUC}, since γ_{AUC} was not estimated.

Table 3: Bias and root mean squared error (RMSE) for treatment effect β_{AUC}, inter- and intra-subject (when applicable) variabilities ψ_{AUC} and γ_{AUC}, total variability ω_{AUC} on $log(AUC)$, residual error σ and variance parameter a from 5000 simulated interaction trials under H_0 for the original (n=12, J=10), intermediate (n=24, J=5) and sparse (n=40, J=3) designs.

HAL author manuscript inserm-00157143, version 1

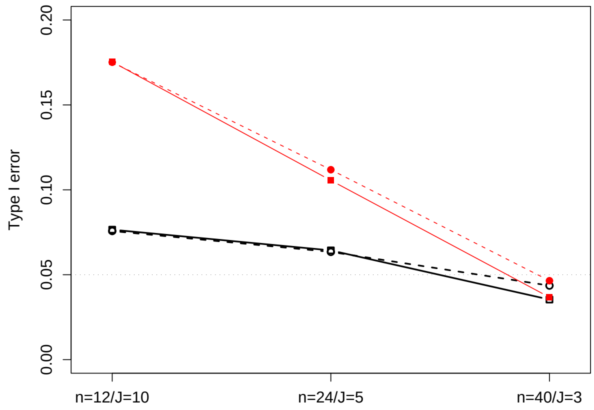
Parameter	Paramater estimates	90% CI
$log(T_a)$	1.32	[1.17;1.48]
β_{T_a}	0.306	[0.133;0.441]
$log(AUC)$	10.7	[10.4;11.0]
β_{AUC}	-0.380	[-0.523;-0.241]
$log(V/F)$	4.01	[3.67;4.34]
$\beta_{V/F}$	0.159	[-0.315;0.632]
ψ_{T_a}	0.217	[0.121;0.369]
ψ_{AUC}	0.492	[0.336;0.749]
$\psi_{V/F}$	0^a	–
γ_{T_a}	0^a	–
γ_{AUC}	0^a	–
$\gamma_{V/F}$	0.535	[0.355;0.775]
σ (ng/mL)	653.5	[567.7;743.0]

[a] Fixed to zero.

Table 4: Population PK parameter estimates of atazanavir and 90% confidence interval (CI) obtained with the final model.

HAL author manuscript inserm-00157143, version 1

(A)



(B)

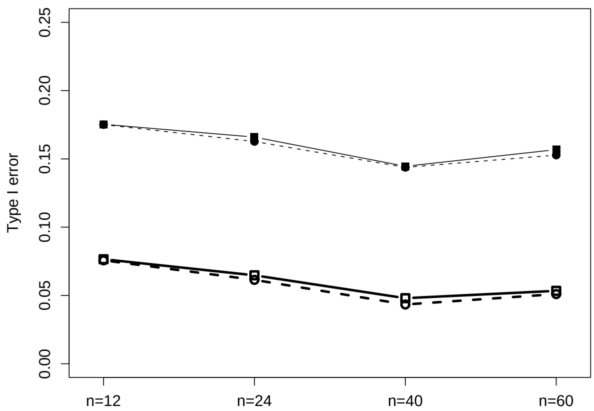


Figure 1: Type I error of comparison tests estimated from 5000 simulated interaction trials under H_0 for different designs with (open symbols) and without (plain symbols) estimation of intra-subject variability (n : number of subjects per treatment group, J : number of samples per subject) for the Wald test (squares) and the LRT (circles)

HAL author manuscript inserm-00157143, version 1

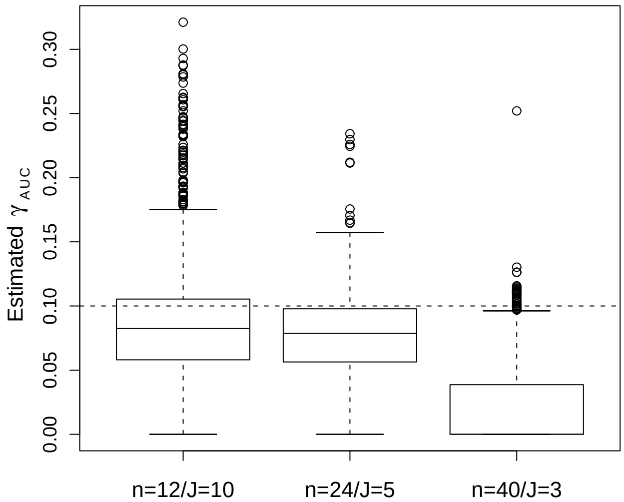
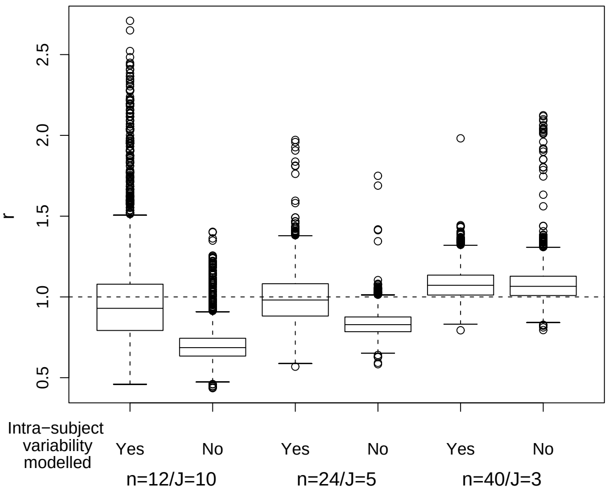


Figure 2: Estimated standard deviation γ_{AUC} of the random effect modelling intra-subject variability for the three designs for the 5000 simulations of PK interaction cross-over trials under H_0.

HAL author manuscript inserm-00157143, version 1

(A)



(B)

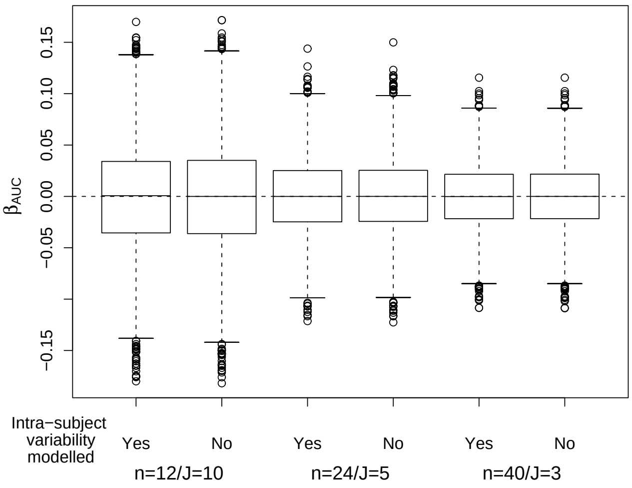


Figure 3: Estimated ratio r (A) and β_{AUC} (B) for the three designs for the 5000 simulations of PK interaction cross-over trials under H_0.

HAL author manuscript inserm-00157143, version 1

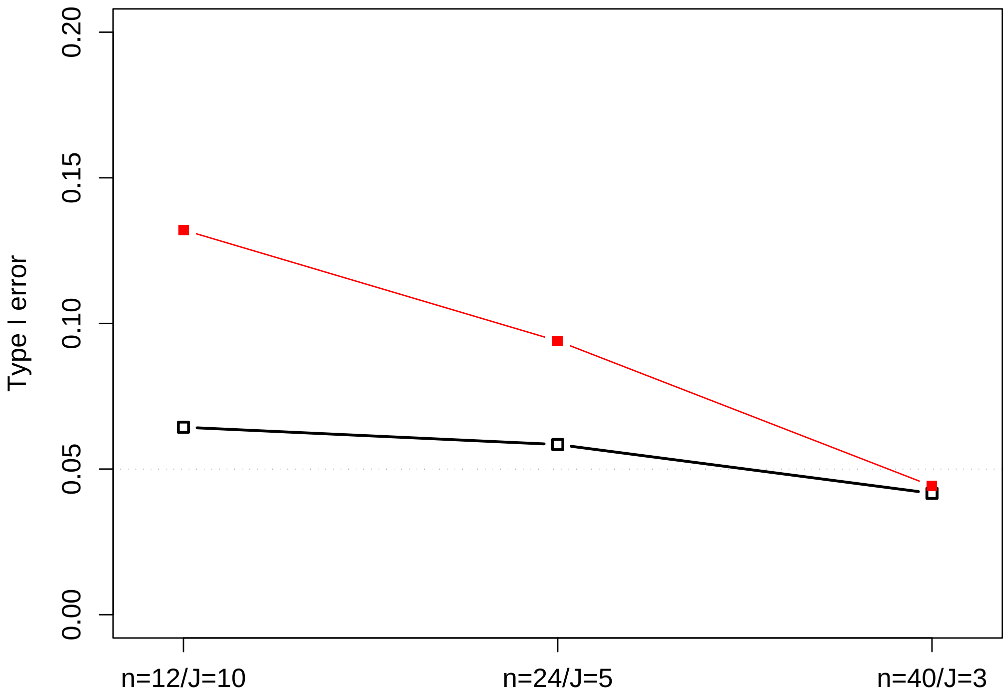


Figure 4: Type I error of the Wald test for equivalence estimated from 5000 simulated bioequivalence trials under H_0 for different designs with (open squares) and without (plain squares) estimation of intra-subject variability (n : number of subjects per treatment group, J : number of samples per subject)

HAL author manuscript inserm-00157143, version 1

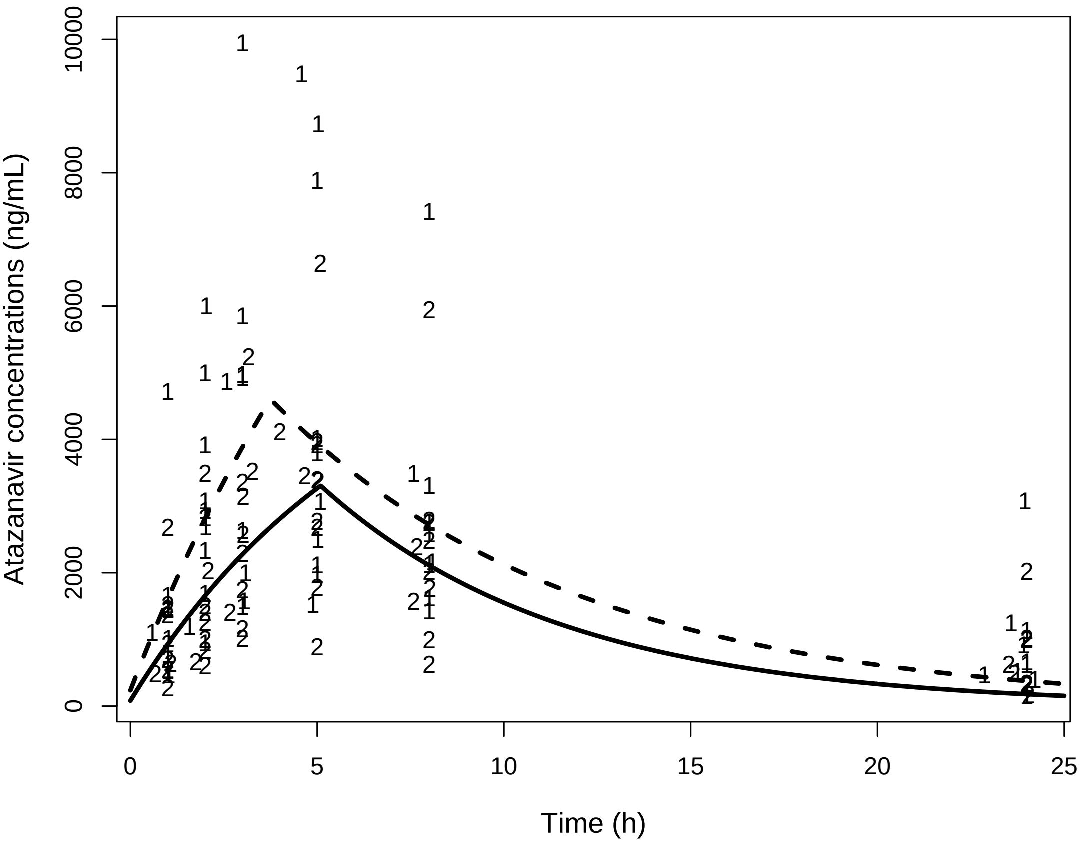


Figure 5: Observed atazanavir concentrations administered without (1) or with (2) tenofovir, and predicted PK profile for the estimated population parameters without (dotted line) or with (full line) tenofovir.

HAL author manuscript inserm-00157143, version 1

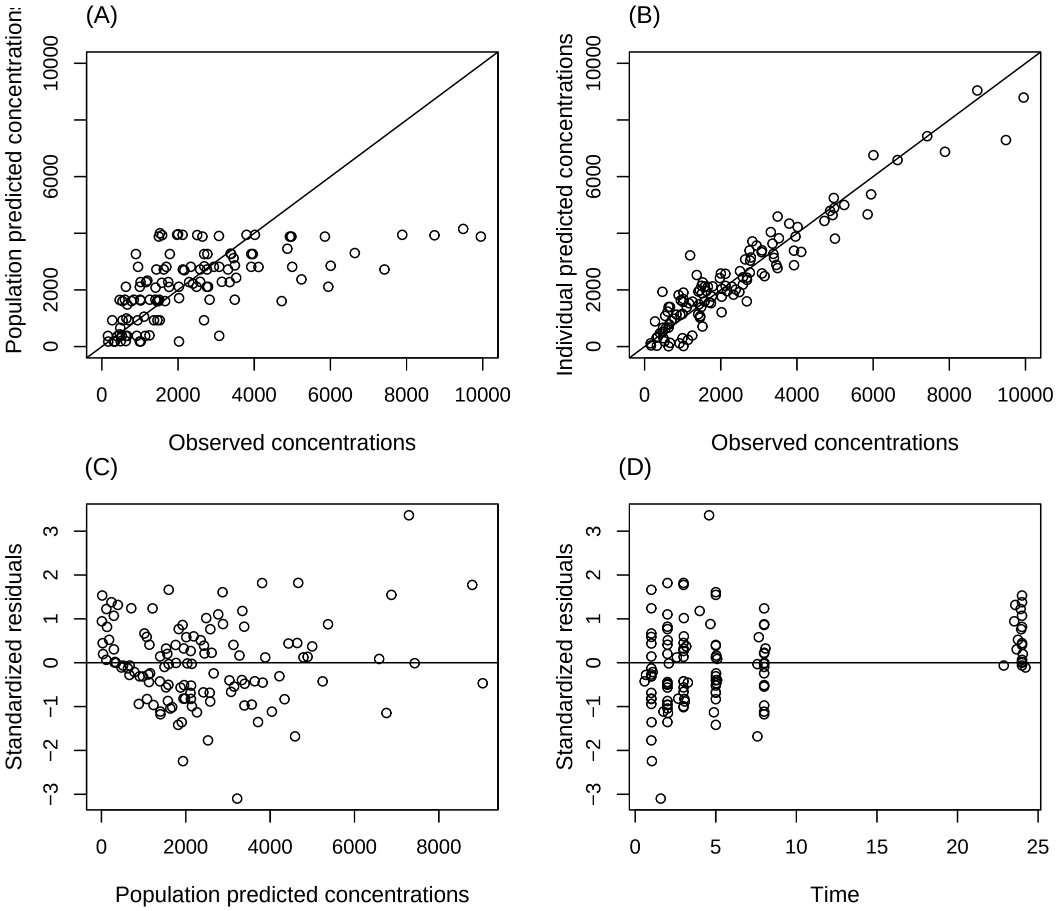


Figure 6: Goodness of fit plots for atazanavir final population PK model: population (A) and individual (B) predicted concentrations (in ng/mL) versus observed concentrations (in ng/mL), standardized residuals versus predicted concentrations (in ng/mL) (C) and versus time (in hours) (D).