

# Enhanced Method for Diagnosing Pharmacometric Models: Random Sampling from Conditional Distributions

Marc Lavielle<sup>1</sup> and Benjamin Ribba<sup>2</sup>

<sup>1</sup>Inria Saclay & CMAP, Ecole Polytechnique, University Paris-Saclay, France

<sup>2</sup>Roche Pharma Research and Early Development, Roche Innovation Center  
Basel, Grenzacherstrasse 124, 4070 Basel, Switzerland

September 13, 2016

## Abstract

Purpose: For nonlinear mixed-effects pharmacometric models, diagnostic approaches often rely on individual parameters, also called empirical Bayes estimates (EBEs), estimated through maximizing conditional distributions. When individual data are sparse, the distribution of EBEs can “shrink” towards the same population value, and as a direct consequence, resulting diagnostics can be misleading.

Methods: Instead of maximizing each individual conditional distribution of individual parameters, we propose to randomly sample them in order to obtain values better spread out over the marginal distribution of individual parameters.

Results: We evaluated, through diagnostic plots and statistical tests, hypothesis related to the distribution of the individual parameters and show that the proposed method leads to more reliable results than using the EBEs. In particular, diagnostic plots are more meaningful, the rate of type I error is correctly controlled and its power increases when the degree of misspecification increases. **An application to the warfarin pharmacokinetic data confirms the interest of the approach for practical applications.**

Conclusions: The proposed method should be implemented to complement EBEs-based approach for increasing the performance of model diagnosis.

# 1 Abbreviations section

Abbreviations	Description
EBE	Empirical Bayes estimates
MAP	Maximum a posteriori
MCMC	Markov Chain Monte Carlo
PK	Pharmacokinetics
PD	Pharmacodynamics
PPC	Posterior predictive checks
VPC	Visual predictive checks

## 2 Introduction

Mixed-effects modelling is nowadays established as a gold-standard approach for the analysis of longitudinal pharmacokinetics (PK) and pharmacodynamics (PD) data. These models are widely used for their ability to describe different levels of variability, and in particular inter-individual variability. Usually, a mixed-effect model is composed by two main components: the model for the observations including the structural model and the residual error model; and the model for the individual parameters, including their relationships with the individual covariates as well as the correlation structure of the random effects [1, 2].

Model diagnosing represents a key activity aimed at building confidence around the developed models before using them for any purpose, such as prediction or simulation. Several diagnostic tools already exist for evaluating the structural model and the residual error model; among them the individual fits, the residual-based diagnostic plots and prediction versus observation plots [3, 4, 5]. Visual predictive checks (VPC) and posterior predictive checks (PPC) are also powerful tools based on the posterior predictive distribution for evaluating simultaneously all the features of the model [6, 7].

Herein, we will focus on diagnosing the model for the individual parameters where diagnostics are often performed to check their marginal distributions, to detect some possible relationships between individual parameters and covariates, or some possible correlations between random effects. Corresponding diagnostic plots are usually based on the empirical Bayes estimates (EBEs) of the individual parameters and EBEs of the random effects.

It is known that the use of EBEs for diagnostic plots and statistical tests is efficient with rich data, when a significant amount of information is available in the data for recovering

accurately all the individual parameters. On the contrary, in case of sparse data, tests and plots can be misleading when the estimates of the individual parameters shrink towards the same population values. Diagnostic tools relying on EBEs are therefore not recommended for high shrinkage [8, 9].

The objective of a diagnostic tool is twofold: first is to check if the assumptions made on individual parameters are valid ; then, if some assumptions are rejected, diagnosis tools should give some guidance on how to improve the model. Model diagnostics is therefore used to eliminate model candidates that do not seem capable of reproducing the observed data [4, 5, 2]. In such process of model building, by definition, none of the features of the “final model” should be rejected. As is the usual case in statistics, it is not because this “final” model has not been rejected that it is necessarily the “true” one. All that we can say is that the experimental data does not allow us to reject it. It is merely one of perhaps many models that cannot be rejected.

The objective of this paper is to propose a new approach for diagnosing models using individual conditional distribution and formally compare this method to the EBE-based classical approach through numerical experiments based on simulated data.

There exists few useful methods for statistical testing in mixed-effects models. Several existing test procedures only concern linear mixed-effects models [10, 11, 12] or generalized mixed-effects models [13, 14, 15]. Furthermore, the aim of most of these procedures is to detect possible misspecifications of the random-effects structure. Other specific features of the model are considered by several authors, such as the normality of the random effects [16, 17], or the error distribution [18].

Bootstrap is a popular method for the global validation of a nonlinear mixed-effects model [19]. Even if bootstrapping is an appealing approach, it requires an important computing effort for

validating a single model which needs to be fitted many times. Then, it cannot be used for model building, but only for validating the final model. Another method for a global test and which relies on the use of a random projection technique is described in [20].

We propose a general approach for testing separately several features of a mixed-effects model. The method consists in generating individual parameters and individual random effects using their conditional distributions. Then, the sampled parameters, the sampled random effects and the original observations can be used together for producing diagnostic plots and building statistical tests.

Herein, we use a one compartment PK model for oral administration to illustrate the practical properties of the proposed method. The design is such that a limited information about the individual absorption rate constant  $ka_i$ , for individual  $i$ , can be obtained from the data. We compare then diagnostic plots and statistical tests when parameters are given by the EBEs or by a random sample of the conditional distributions.

## 3 Methods

### 3.1 Empirical Bayes estimates versus random sampling from the conditional distribution

Calculating the EBE of an individual parameter consists in estimating  $\psi_i$  by maximizing the conditional distribution  $p(\psi_i|y_i)$  where  $y_i = (y_{ij}, 1 \leq j \leq n_i)$  is a sequence of observations. This *conditional mode*, also known as the maximum a posteriori (MAP) estimate of  $\psi_i$ , is the

most likely value of the individual parameter  $\psi_i$ , given the observations and a given population distribution  $p(\psi_i)$ .

However, when the data are sparse, individual estimates of a parameter can “shrink” towards the same population value, which is the mode of the population distribution of this parameter. For a parameter  $\psi_i$  which is a function of a random effect  $\eta_i$ , we can quantify this phenomena by defining the so-called  $\eta$ -shrinkage [9] as:

$$\eta\text{-shrinkage} = 1 - \frac{\text{var}(\hat{\eta}_i)}{\omega^2},$$

where  $\text{var}(\hat{\eta}_i)$  is the empirical variance of the  $\hat{\eta}_i$ 's and  $\hat{\eta}_i$  the empirical Bayes estimate of  $\eta_i$  that maximizes  $p(\eta_i|y_i)$ .

Saying that the observations  $y_i$  provide little information about  $\eta_i$  means that  $\hat{\eta}_i$  is close to 0. This results as a high level of shrinkage (close to 1) whenever  $\text{var}(\hat{\eta}_i) \ll \omega^2$ . Estimates of the  $\psi_i$  are therefore biased because they do not correctly reflect the marginal distribution  $p(\psi_i)$ . In particular, their empirical variance is much reduced.

Alternatively, individual parameters  $\psi_i$  can be drawn from the conditional distribution  $p(\psi_i|y_i)$  rather than taking the mode. The resulting estimator is unbiased in the following sense:

$$p(\psi_i) = \mathbb{E}(p(\psi_i|y_i)). \tag{1}$$

This relationship is a fundamental one when considering mixed-effects models. It means that, if we randomly draw a vector  $y_i$  of observations for an individual in a population and then generate a vector  $\psi_i$  using the conditional distribution  $p(\psi_i|y_i)$ , the distribution of  $\psi_i$  is the population distribution  $p(\psi_i)$ . In other words, even if each  $\psi_i$  is randomly generated using

its own conditional distribution, the fact of pooling them allows us to look at them as if they were a sample from  $p(\psi_i)$ . A consequence of this important property is that any diagnostic plot based on such simulated individual parameters can be used confidently.

To conceptually illustrate the difference between EBEs and sampled parameters, we will consider 10 individual parameters  $\psi_1, \dots, \psi_{10}$  and 10 observations  $y_1, \dots, y_{10}$  so that each individual only has one unique observation. The 10 conditional distributions and their modes (i.e. the 10 EBEs) are shown Figure 1(a) while 10 parameters randomly sampled from these distributions are displayed Figure 1(b).

Because of the  $\eta$ -shrinkage (there is only one observation per individual), we can see Figure 1(c) that the empirical distribution of the EBEs is concentrated around the mean  $\psi_{\text{pop}}$  of the population distribution. On the other hand, Figure 1(d) shows that the empirical distribution of the sampled parameters correctly represents this population distribution.



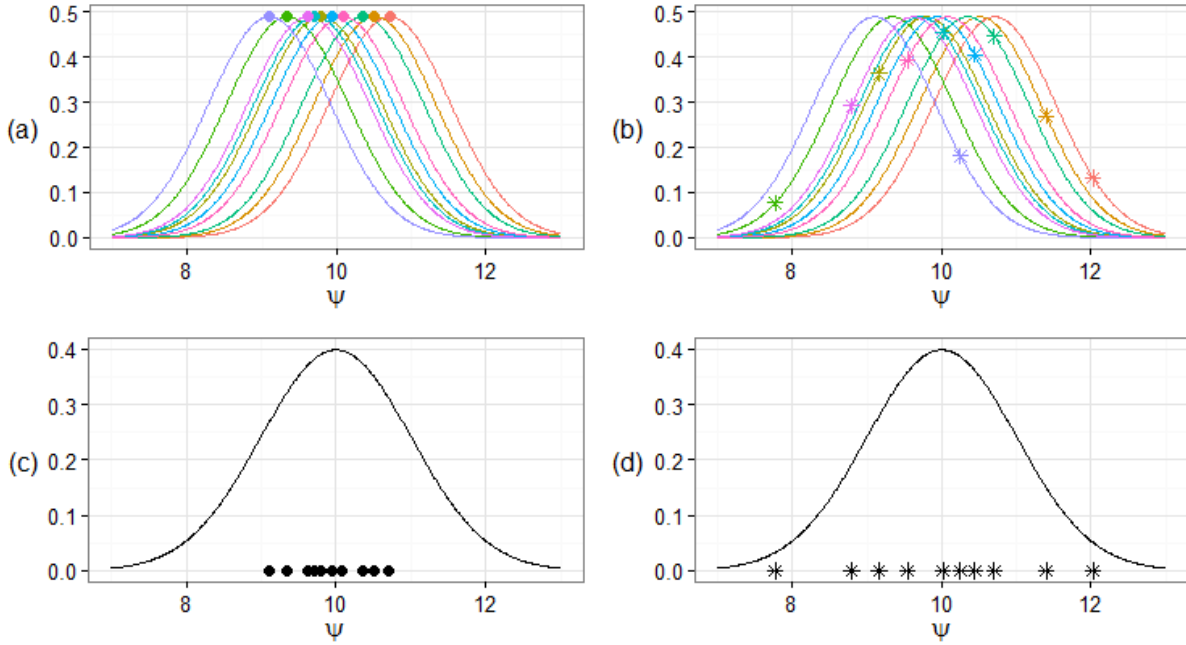


Figure 1: Illustration of how sampling from conditional distributions can describe population distribution compared to EBEs in case of shrinkage. (a-b) Conditional distributions of  $\psi_1, \dots, \psi_{10}$  and (a) the EBEs maximizing these 10 conditional distributions (circles), (b) individual parameters sampled from these 10 conditional distributions (stars); (c-d) Population distribution of  $\psi$  and (c) the EBEs, (d) the sampled parameters. The model used to generate this illustration is:  $y_i = \psi_i + \varepsilon_i$  where  $\psi_i \sim \mathcal{N}(\psi_{\text{pop}}, \omega^2)$  and  $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ . In that case, the conditional distribution of  $\psi_i$  given  $y_i$  is a normal distribution with mean  $\mu_i = V(y_i/\sigma^2 + \psi_{\text{pop}}/\omega^2)$  and variance  $V = \sigma^2\omega^2/(\sigma^2 + \omega^2)$ . We used  $\psi_{\text{pop}} = 10$  and  $\omega^2 = \sigma^2 = 1$  for this numerical example.

### 3.2 Pharmacokinetic model

Throughout the manuscript, we will use a simple and classical PK model to illustrate the proposed approach for model diagnostic and hypothesis testing. The model is a one compartment PK model for single oral administration, with first order absorption and linear elimination:

$$C(t, \psi) = \frac{D ka}{V ka - Cl} (e^{-(Cl/V)t} - e^{-kat}).$$

where  $D$  is the amount of drug administered at time 0. Here, the PK parameters are  $\psi = (ka, V, Cl)$ .

We then assume an exponential error model for the observed concentration:

$$y_{ij} = C(t_{ij}, \psi_i) e^{\varepsilon_{ij}},$$

where  $\varepsilon_{ij} \sim \mathcal{N}(0, a^2)$ . Here,  $y_{ij}$  is the concentration measured for patient  $i$  at time  $t_{ij}$  and  $\psi_i = (ka_i, V_i, Cl_i)$ .

We assume that the individual PK parameters  $ka_i$ ,  $V_i$  and  $Cl_i$  are log-normally distributed. Furthermore, a linear relationship between log-weight and each log-parameter is assumed:

$$\log(ka_i) = \log(ka_{\text{pop}}) + \beta_{ka} \log(w_i/w_{\text{pop}}) + \eta_{ka,i},$$

$$\log(V_i) = \log(V_{\text{pop}}) + \beta_V \log(w_i/w_{\text{pop}}) + \eta_{V,i},$$

$$\log(Cl_i) = \log(Cl_{\text{pop}}) + \beta_{Cl} \log(w_i/w_{\text{pop}}) + \eta_{Cl,i},$$

where  $w_i$  is the weight of patient  $i$  and  $w_{\text{pop}}$  the typical weight in the population.

The random effects are normally distributed:  $\eta_i = (\eta_{ka,i}, \eta_{V,i}, \eta_{Cl,i}) \sim \mathcal{N}(0, \Omega)$ . Variances of the random effects, i.e. the diagonal elements of  $\Omega$ , are  $(\omega_{ka}^2, \omega_V^2, \omega_{Cl}^2)$  and the correlations between random effects are  $(r_{ka,V}, r_{ka,Cl}, r_{V,Cl})$ .

PK data for  $N = 150$  patients were simulated with this model using  $D = 100\text{mg}$  and the following values of the population parameters:  $ka_{\text{pop}} = 1$ ,  $V_{\text{pop}} = 10$ ,  $Cl_{\text{pop}} = 1$ ,  $\omega_{ka} = 0.3$ ,

$\omega_V = 0.2$ ,  $\omega_{Cl} = 0.2$  and  $a = 0.15$ . The volume is function of weight in this example:  $\beta_V = 1$  while  $\beta_{ka} = \beta_{Cl} = 0$ . We furthermore assume that log-volume and log-clearance are positively correlated:  $r_{V,Cl} = 0.6$  while  $r_{ka,V} = r_{ka,Cl} = 0$ .

Individual weights were sampled from a normal distribution with mean  $w_{pop} = 70\text{kg}$  and standard deviation 7kg.

### 3.3 Simulated designs and tested model

We used a design with 3 sampling times per patient: (2h, 4h, 8h, 12h) for  $1 \leq i \leq 50$ ; (4h, 12h, 24h, 48h) for  $51 \leq i \leq 100$  and (1h, 8h, 12h, 24h) for  $101 \leq i \leq 150$ . The simulated concentrations of the 150 individuals are displayed Figure 2.

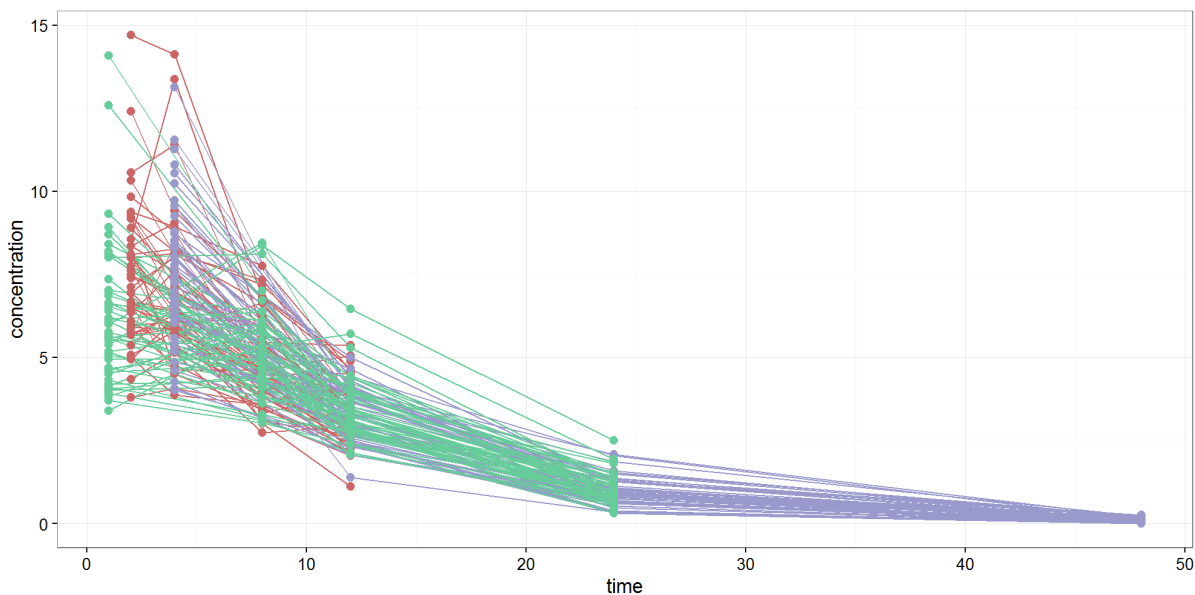


Figure 2: Simulated PK data for 150 patients with three different designs (50 patients per design). Simulated PK data of individuals subjected to design 1 (2h, 4h, 8h, 12h) are depicted in red. Design 2 (1h, 8h, 12h, 24h) in green, and design 3 (4h, 12h, 24h, 48h) in blue.

While the data were simulated with a model where  $\beta_V = 1$  and  $r_{V,Cl} = 0.6$ , we first fitted this data with a “wrong” model, where  $\beta_V = 0$  and  $r_{V,Cl} = 0$ .

### 3.4 Diagnosing tools

The process of model building is an iterative process where, at each iteration, we make some hypotheses about the joint distribution of the individual PK parameters. We then fit this model to the PK data and produce some diagnostic plots. The objective of these plots is to evaluate graphically which of the hypotheses can be considered as valid and which one should be rejected. Then, rejecting some of the hypotheses leads to proposing a new model which in turn needs to be evaluated. Ideally, this process of model building should lead to a final model for which none of the diagnostic plots detect any misspecification.

Here, we will make the following hypotheses:

- The PK parameters are log-normally distributed.
- There is no relationship between the covariate (the weight) and the PK parameters:

$$\beta_{ka} = \beta_V = \beta_{Cl} = 0.$$

- There is no correlation between random effects:

$$r_{ka,V} = r_{ka,Cl} = r_{V,Cl} = 0.$$

We will fit this model to the simulated data, calculate individual parameters by using EBE or by sampling from the conditional distributions and produce the following plots:

- Comparison of the empirical distribution of the  $(\psi_i)$  with their theoretical distribution given by the model. We can for instance compare histograms and probability density functions (pdf).
- Visualization of the possible relationships between covariates and parameters, or between covariates and random effects, through scatter plots.
- Visualization of the possible relationships between random effects through scatter plots.

Using diagnostic plots for model building remains quite empirical. Indeed, there is no well defined decision rule to decide which hypotheses made on the model are incorrect and should be rejected. Some quantitative criteria associated to these plots might be helpful for the modeller to take such decision. In other word, we would like to derive formal statistical tests from diagnostic plots.

Testing our hypotheses about the distribution of the individual PK parameters can be carried out with some standard statistical tests, such as:

- Kolmogorov-Smirnov test for testing the fit of distributions,
- Pearson's test for testing linear correlations between - possibly transformed - covariates and random effects: the test statistic is based on Pearson's correlation coefficient and follows a  $t$ -distribution with  $N - 2$  degrees of freedom if the samples follow independent normal distributions.
- Pearson's test for testing linear correlations between random effects.

The fundamental property (1) ensures that any statistical test based on such sampled individual parameters is *unbiased*: the effective level of the implemented test is precisely the desired

level  $\alpha$ . We can then expect that each of the proposed statistical test will wrongly reject the null hypothesis (i.e. reject the model being evaluated when it is correct) with probability  $\alpha$ .

Controlling the level of each of these tests is important of course, but their role is mainly to detect misspecification in the model. It is therefore essential to also evaluate the power of these tests in order to know which kind of misspecification can be identified with a reasonable probability.

### 3.5 Sampling individual parameters from conditional distributions

In practice, sampling  $\psi_i$  from the conditional distribution  $p(\psi_i|y_i)$  can be done by Markov Chain Monte Carlo (MCMC) [21].

For the numerical experiments presented below, we used the Metropolis-Hastings (MH) algorithm described in [2] and that combines several proposal distributions. In order to get samples of the individual conditional distributions, we run 200 iterations of this algorithm and used 10 independent Markov chains per individual. We then kept the individual parameters obtained from all the chains at the last iteration.

This MH algorithm was initially implemented in Monolix<sup>1</sup> together with the SAEM algorithm used for the estimation of the population parameters [2]. Monolix then returns individual parameters and random effects sampled from the conditional distributions and use them for the diagnostic plots.

---

<sup>1</sup><http://lixoft.com/products/monolix/>

On the other hand, diagnostic plots derived from NONMEM<sup>2</sup> are only based on EBEs (individual parameters and random effects). Nevertheless, since the MCMC algorithm implemented in Monolix is also implemented in NONMEM, it should be possible to also return the sampled individual parameters together with the EBEs.

## 4 Results

### 4.1 Diagnostic plots

Estimated parameters under this model are:  $\hat{k}a_{\text{pop}} = 0.99$ ,  $\hat{V}_{\text{pop}} = 10.2$ ,  $\hat{Cl}_{\text{pop}} = 1.02$ ,  $\hat{\omega}_{ka} = 0.14$ ,  $\hat{\omega}_V = 0.25$ ,  $\hat{\omega}_{Cl} = 0.19$ ,  $\hat{a} = 0.156$ .

Even if the number of subjects is quite large ( $N = 150$ ), the data can be considered sparse (4 sampling points per individual), providing especially a limited information on the absorption process. The  $\eta$ -shrinkage for  $ka$ ,  $V$  and  $Cl$  are respectively 88%, 20% and 20% when it is computed using the EBEs. Then, even if the histograms of the EBEs displayed Figure 3 (top row) look quite different from the log-normal distributions obtained with the estimated population parameters (in solid red lines), we cannot conclude that the population distributions of the individual PK parameters are misspecified.

Figure 3 (middle row) shows that identification of relationships between covariates and individual parameters is much less sensitive to shrinkage: EBEs correctly identify the linear relationship existing between log-weight and log-volume, while the other PK parameters  $ka$  and  $Cl$  do not clearly seem to be function of weight. This good behavior can be explained by

---

<sup>2</sup><https://nonmem.iconplc.com/>

the fact that such relationship is related to the central tendency of the distributions of the PK parameters, which is pretty well approximated by the modes.

The  $\eta$ -shrinkage also strongly impacts the joint distribution of the random effects. We can see Figure 3 (bottom row) that the joint distribution of the estimates of the random effects does not reflect correctly the true distribution. Artificial correlations wrongly appear between all the random effects. This diagnostic plot does not allow to detect that only the correlation between  $\eta_V$  and  $\eta_{Cl}$  is relevant.



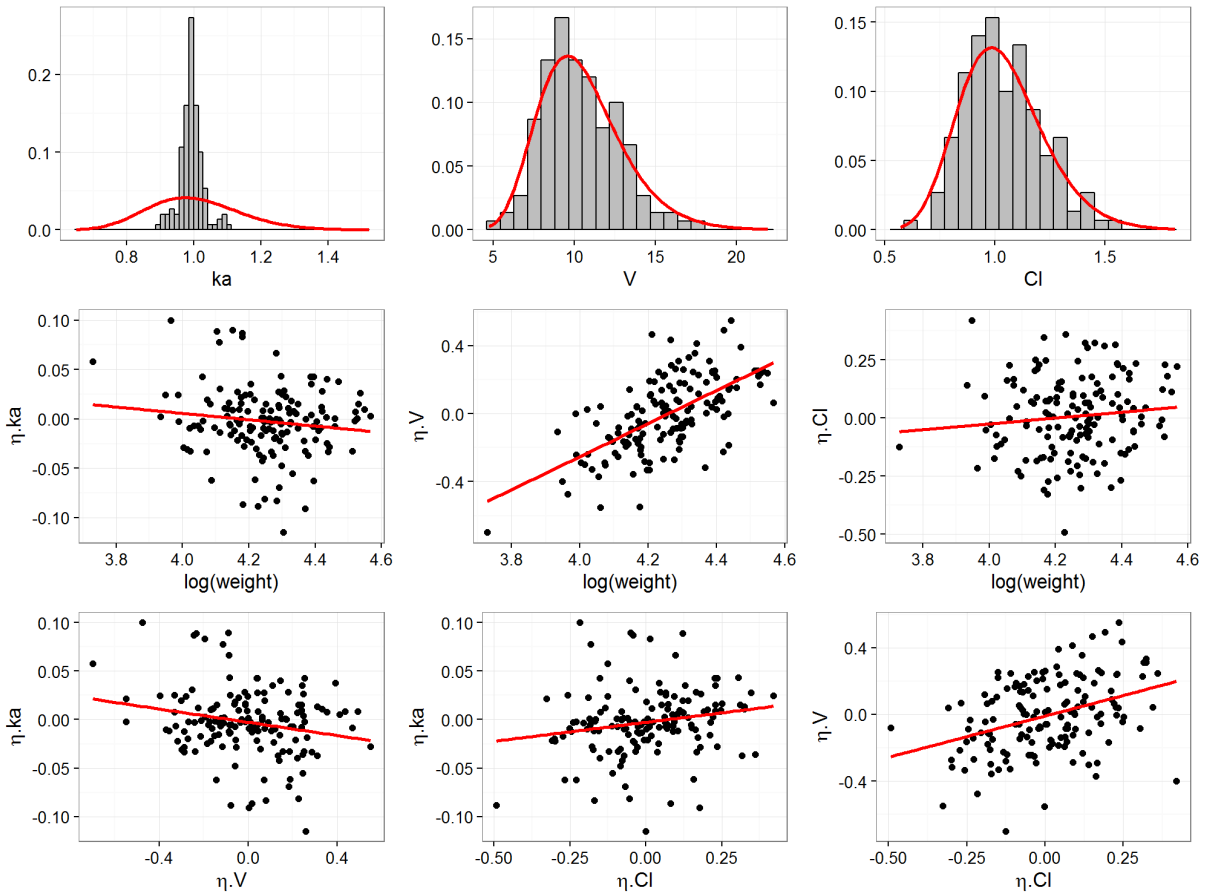


Figure 3: Diagnostic plots with EBEs. Top row: Empirical distributions of the individual parameters maximizing the conditional distributions. The estimated population pdf's are displayed in solid line. Middle row: Relationships between log-weight and random effects maximizing the conditional distributions. Bottom row: Relationships between random effects maximizing the conditional distributions.

On the other hand, creating diagnostic plots based on sampled parameters and sampled random effects allows us to use all these diagnostic plots for decision making. Figure 4 (top row) shows a very nice fit between the empirical distributions of the individual PK parameters and the theoretical pdf's. Based on this plot, we can conclude that there is no reason for rejecting the hypothesis that the three PK parameters are log-normally distributed.

Figure 4 (middle row) shows a correlation between  $\log(w)$  and  $\log(V)$ : based on this plot, we can then reject the hypothesis that  $\beta_V = 0$  while, in the contrary, there is no reason for rejecting the hypothesis that  $\beta_{ka} = \beta_{Cl} = 0$ .

Only a correlation between  $\eta_V$  and  $\eta_{Cl}$  clearly appears Figure 4 (bottom row): we can then reject the hypothesis that  $r_{V,Cl} = 0$ , as it was assumed in the model. On the other hand, there is no reason for rejecting the hypothesis that  $r_{ka,V} = r_{ka,Cl} = 0$ .

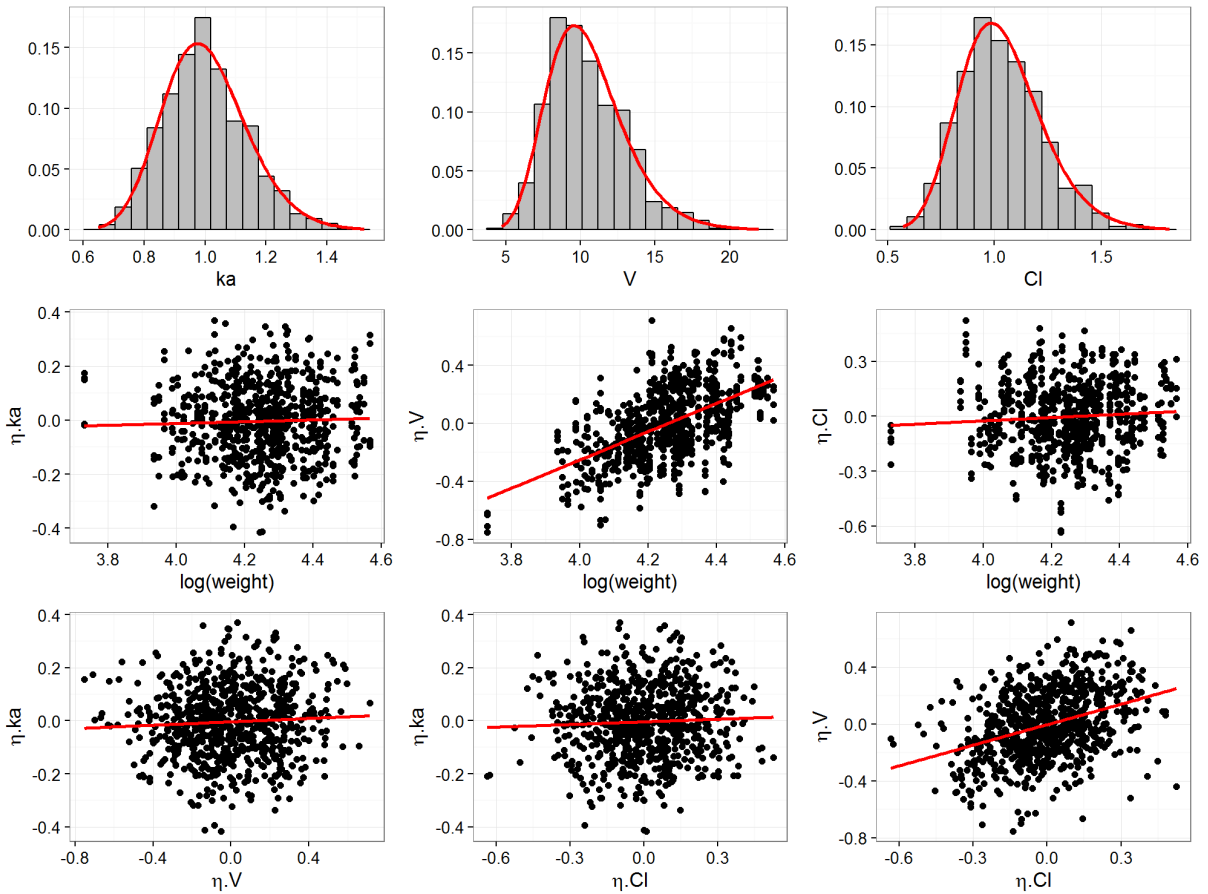


Figure 4: Diagnostic plots with sampled parameters. Top row: Empirical distributions of the individual parameters sampled from the conditional distributions. The estimated population pdf's are displayed in solid line. Middle row: Relationships between  $\log(\text{weight})$  and random effects sampled from the conditional distributions. Bottom row: relationships between random effects sampled from the conditional distributions. For visual purpose, the conditional distributions were sampled five times for each individual resulting in generating 5 times more points than in the previous figure.

## 4.2 Statistical tests

### 4.2.1 Type I error

In the following, we will test formally each of the hypothesis of the model being evaluated and check if the “true” model can be identified.

First, we can test separately if each PK parameter follows a log-normal distribution defined by the estimated parameters. Table 1 gives the p-values of the Kolmogorov-Smirnov tests when either the EBEs or the sampled individual PK parameters are used. Results confirm what could be seen in Figure 3: because of the strong shrinkage for  $ka$ , EBEs of  $ka$  do not follow the estimated population distribution. On the other hand, the tests based on sampled PK parameters are not affected by the  $\eta$ -shrinkage. Then, both the diagnostic plot displayed Figure 4 and the p-values of these three tests can be used with confidence to decide not to reject the hypothesis that the PK parameters are log-normally distributed.

Then, we can test if there exists some linear correlation between the log-weight and the log-parameters. Results presented Table 1 show that EBEs and sampled parameters give similar results for these tests. Both lead to the conclusion that a significant correlation exists between log-weight and log-volume.

Lastly, we can test if there exists some linear correlation between the random effects. Results presented Table 1 confirm what could be seen in Figures 3 and 4: empirical Bayes estimates of the  $\eta$ 's creates some artificial correlations while sampled random effects correctly reproduce the correlation structure of the random effects. Based on this test, we can conclude with high confidence that  $r(\eta_V, \eta_{Cl}) \neq 0$ .

Null hypothesis	p-value	
	EBEs	sampled parameters
$\log(ka) \sim \mathcal{N}(\log(0.99), 0.14^2)$	$< 10^{-10}$	0.58
$\log(V) \sim \mathcal{N}(\log(10.2), 0.25^2)$	0.43	0.71
$\log(Cl) \sim \mathcal{N}(\log(1.02), 0.19^2)$	0.85	0.82
$r(\log(w), \log(ka)) = 0$	0.11	0.66
$r(\log(w), \log(V)) = 0$	$< 10^{-10}$	$< 10^{-10}$
$r(\log(w), \log(Cl)) = 0$	0.19	0.40
$r(\eta_{ka}, \eta_V) = 0$	0.007	0.40
$r(\eta_{ka}, \eta_{Cl}) = 0$	0.02	0.49
$r(\eta_V, \eta_{Cl}) = 0$	$3 \cdot 10^{-6}$	$9 \cdot 10^{-6}$

Table 1: Kolmogorov-Smirnov tests for the probability distributions of the individual PK parameters, the relationships between weight and parameters, and the joint distribution of the random effects.

Controlling the significance level  $\alpha$  of a statistical test means that the effective rate of type I error, i.e. the rate of falsely rejected null hypotheses, is expected to be  $\alpha$ . Here, the probability to reject the null hypothesis cannot be computed in a closed form, but it can be estimated by Monte-Carlo. We have simulated 500 replicates of the data under the null hypothesis and performed each of the proposed statistical tests at levels  $\alpha = 0.05$  and  $\alpha = 0.10$ , for each replicate, and using either the EBEs or the sampled parameters and random effects.

Table 2 provides the rates of falsely rejected null hypotheses for each of these tests. We can see that the significance levels  $\alpha = 0.05$  and  $\alpha = 0.10$  are very well controlled when sampled PK parameters and sampled random effects are used for any of these tests.

On the other hand, a strong bias is observed when the EBEs are used for testing the distribution of the parameters or the correlation between random effects. The rate of type I error of the tests concerning the relationship between weight and PK parameters is more correctly controlled.

Null hypothesis	Rates of falsely rejected null hypotheses			
	EBEs		sampled parameters	
	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$
$\log(ka) \sim \mathcal{N}(\log(1), 0.3^2)$	1	1	0.030	0.090
$\log(V) \sim \mathcal{N}(\log(10), 0.2^2)$	0.158	0.288	0.056	0.120
$\log(Cl) \sim \mathcal{N}(\log(1), 0.2^2)$	0.068	0.142	0.060	0.102
$\beta_{ka} = 0$	0.034	0.092	0.044	0.088
$\beta_V = 0$	0.038	0.060	0.052	0.118
$\beta_{CL} = 0$	0.034	0.068	0.036	0.092
$r_{ka,V} = 0$	0.870	0.912	0.042	0.096
$r_{ka,Cl} = 0$	0.344	0.440	0.046	0.092
$r_{V,Cl} = 0$	0.140	0.234	0.042	0.094

Table 2: Rates of falsely rejected null hypotheses for statistical tests with significance level  $\alpha = 0.05$  using either the EBEs or the sampled individual PK parameters.

#### 4.2.2 Power of the tests

We finally explore how the statistical tests behave under several alternative hypotheses. We have simulated 200 replicates of the same experiment using the previous design and under various parameter scenari. For each of these scenari, the estimated power of the test is the proportion of rejected null hypotheses among the 200 replicates. Only tests of level  $\alpha = 0.05$  have been performed for this power analysis (similar conclusions were obtained with  $\alpha = 0.10$ ).

In the first experiment, the values of population PK parameters used for drawing the individual PK parameters are different from the values defining the null hypothesis, i.e.  $ka_{\text{pop}} = 1$ ,  $V_{\text{pop}} = 10$  and  $Cl_{\text{pop}} = 1$ . Figure 5 top row confirms that tests based on EBEs are biased and should not be used for testing the marginal distribution of the parameters. Indeed, even if they look quite powerful, the type I error is significantly overestimated. The three tests based on sampled parameters are unbiased, even if a misspecification in the distribution of  $ka$  is difficult to detect with this design.

We see Figure 5 bottom row that an effect of moderate size of the covariate (weight) on  $V$  and  $Cl$  is correctly detected using either the EBEs or the sampled PK parameters. On the other hand, the design only allows to detect an important effect on  $ka$ . We also see that, contrary to the previous examples, tests based on EBEs can be used for detecting a relationship between the covariate and an individual parameter. Indeed, these tests seem to be unbiased and slightly more powerful than the tests based on sampled parameters.

We then investigate if linear correlations between PK parameters can be detected. Figure 5 shows that a clear bias is introduced when EBEs are used, while correlation between  $V$  and  $Cl$  is correctly detected with sampled parameters. Because of the  $\eta$ -shrinkage on  $ka$ , only strong correlations (positive or negative) between  $ka$  and  $V$  or  $ka$  and  $Cl$  can be detected with this method.

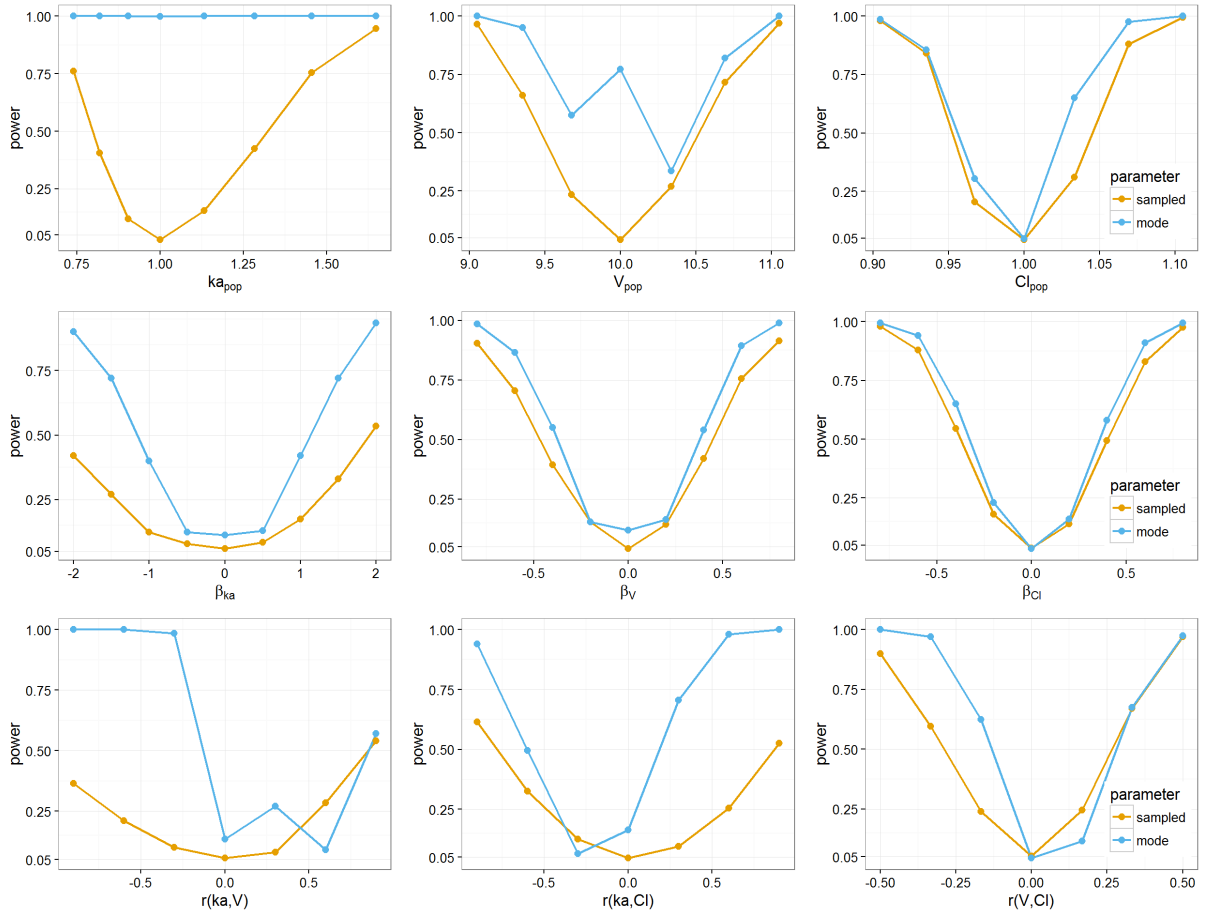


Figure 5: Power of the statistical tests. Top row: testing the probability distributions of the parameters. Middle row: testing a linear correlation between log-weight and log-parameters. Bottom row: testing linear correlation between parameters. Blue and yellow curves show power when using EBEs and sampled parameters respectively.

### 4.3 Application to the warfarin PK data

We will now use the pharmacokinetics of warfarin [22] to illustrate the proposed method. 32 healthy volunteers received a 1.5 mg/kg single oral dose of warfarin, an anticoagulant used in the prevention of thrombosis. Supplemental Figure S1 shows the warfarin plasmatic concentration for these patients measured at different



times.

We will consider a one compartment PK model, assuming a first-order absorption process after a lag-time and a linear elimination:

$$C(t, \psi) = \begin{cases} \frac{Dka}{Vka-Cl} (e^{-Cl/V(t-Tlag)} - e^{-ka(t-Tlag)}) & \text{if } t \geq Tlag \\ 0, & \text{otherwise.} \end{cases}$$

Here,  $\psi = (Tlag, ka, V, Cl)$ .

We assume log-normal distributions for these 4 PK parameters and a diagonal variance-covariance matrix  $\Omega$  for the random effects. The residual error model for the observations is a combined error model of the form  $y_{ij} = C(y_{ij}, \psi_i) + (a + bC(y_{ij}, \psi_i))\varepsilon_{ij}$ .

We used Monolix 2016R1 for fitting this model to the warfarin PK data. The empirical distribution of the EBEs of the random effects is displayed Supplemental Figure S2 (top row) and shows a strong shrinkage for the absorption parameters  $Tlag$  and  $ka$ . Indeed, more than half of the patients have no measurements during the first 24 hours. Then, we merely use the population parameters to predict  $Tlag$  and  $ka$  for these patients. This large shrinkage does not mean that the model is misspecified, but that the data does not allow us to correctly estimate these individual parameters. As a consequence, EBEs cannot be used for diagnosing the model.

On the other hand, individual PK parameters sampled from the conditional distributions can be used with confidence. The distribution of the random effects

displayed Supplemental Figure S2 (bottom row) shows that there is no reason for rejecting the hypothesis of log-normal distributions. A relationship between weight and volume is clearly visible Supplemental Figure S3 as well as a possible relationship between weight and clearance. Lastly, Supplemental Figure S4 identifies a correlation between  $\eta_V$  and  $\eta_{Cl}$ . The hypothesis of independent random effects should also be rejected.

In summary, based on these diagnostic plots and statistical tests, the model used for fitting the warfarin PK data should be rejected. A new model to be tested should integrate a correlation between  $\eta_V$  and  $\eta_{Cl}$ , a relationship between weight and volume, and possibly between weight and clearance.

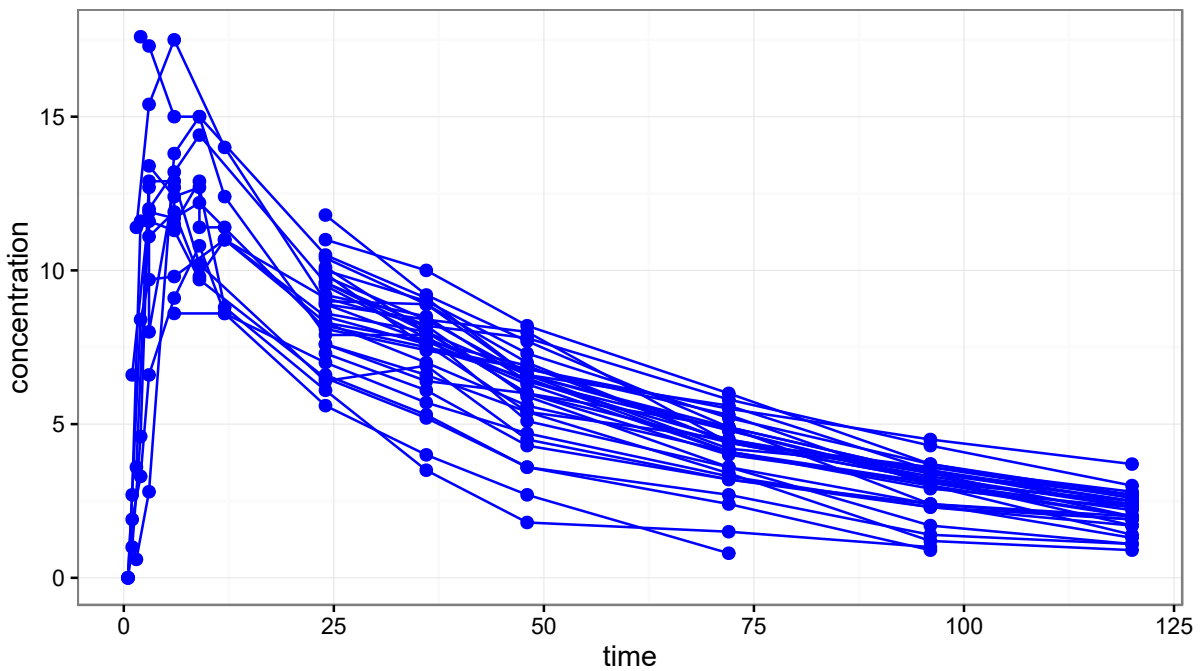


Figure 6: Supplemental Figure S1: Warfarin PK data.

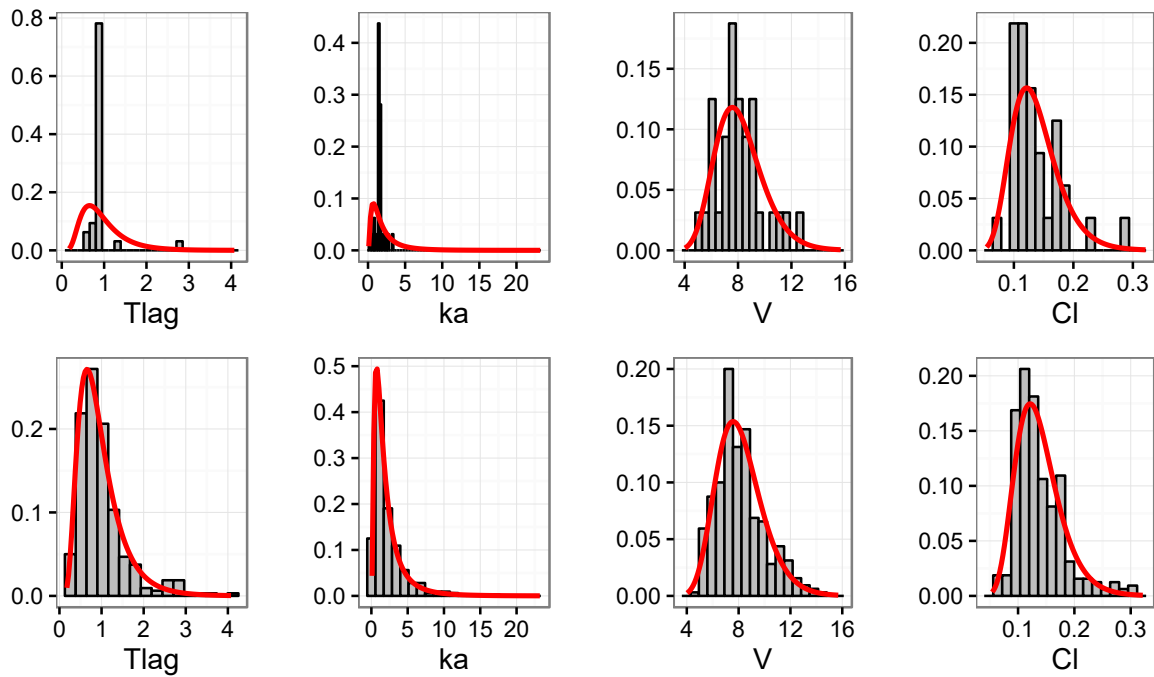


Figure 7: Supplemental Figure S2: Empirical distribution of the individual parameters. The estimated pdf's are displayed in solid line. Top: EBEs, bottom: sampled from the conditional distributions.

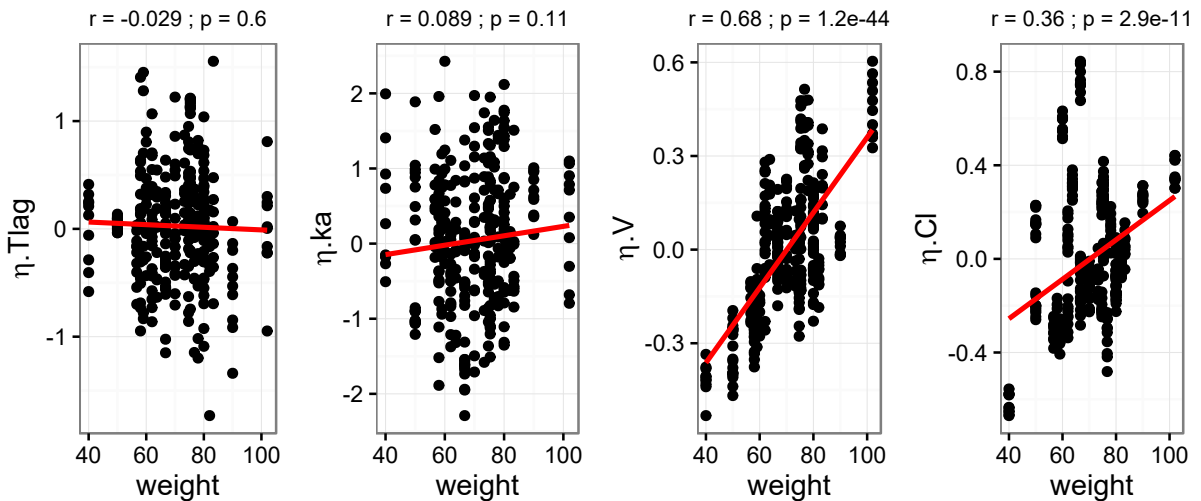


Figure 8: Supplemental Figure S3: Sampled individual parameters versus weight. The correlation coefficient and the p-value of the test  $r = 0$  are displayed for each parameters.

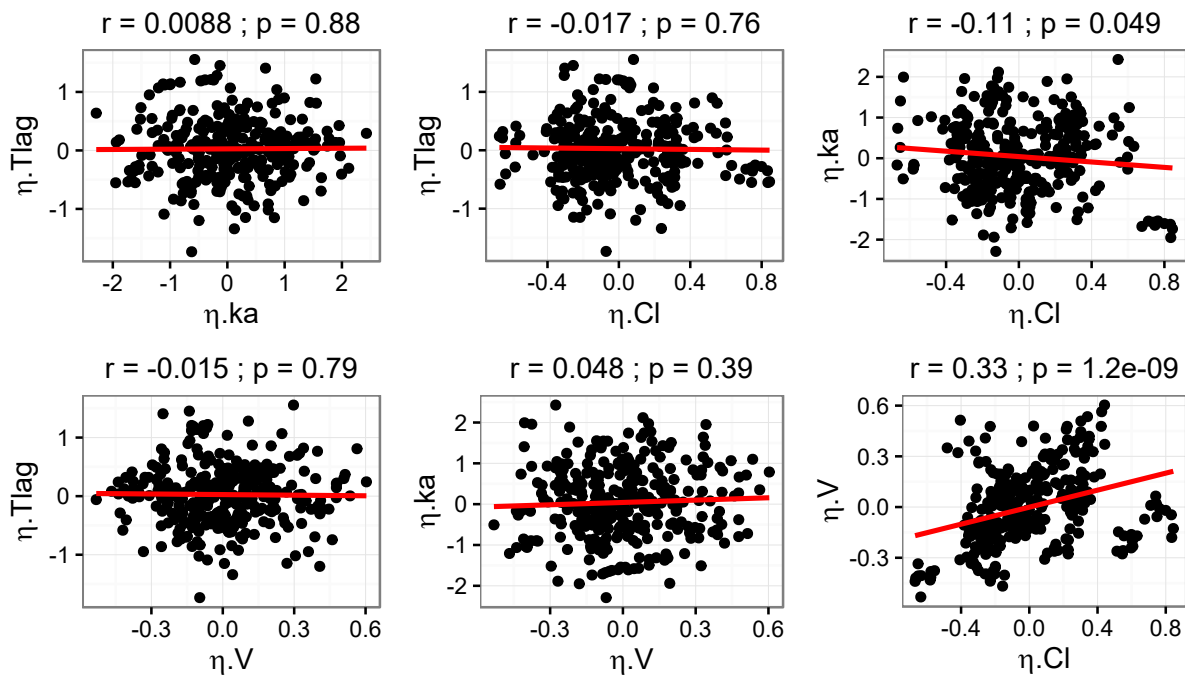


Figure 9: Supplemental Figure S4: Joint distributions of the sampled random effects. The correlation coefficient and the p-value of the test  $r = 0$  are displayed for each pair of parameters.

## 5 Conclusions

In this manuscript, we propose a new method for deriving individual parameters used in a diagnostic perspectives. Instead of using the classical approach of maximizing each conditional distribution, we show that randomly sampling these distribution leads to reliable results and can complement the EBE-based approach widely used. In particular, we show that each proposed test is unbiased, the type I error rate is the desired significance level of the test and the probability to detect a misspecification in the model increases with the magnitude of this misspecification. This method can therefore be used efficiently, possibly in combination with other diagnostic tools, to drive model building in population PKPD analyses.

Our numerical experiments confirmed that EBEs for assessing the distribution of the individual parameters and/or the correlation structure of the random effects may introduce strong biases when  $\eta$ -shrinkage is important. However, interestingly, in our example, we show that the effect of a continuous covariate on a PK parameter is correctly detected using either the EBEs or the sampled parameters. The numerical tests also revealed that the sampling of conditional distribution can also suffer and results in lack of power in presence of  $\eta$ -shrinkage (see Figure 5).

Thus, even if using EBEs can be helpful for the search of misspecifications, it appears not to be a reliable methods for validation of the final model and sampled parameters should always be used for this aim.

Herein, we only addressed the problem of diagnosing the model for the individual parameters, but the same approach could be developed for other diagnostic plots and for testing other components of the model including residual error model, structural model, and handling of BLQ data.

## References

- [1] P. L. Bonate, *Pharmacokinetic-pharmacodynamic modeling and simulation*. Springer, 2011.
- [2] M. Lavielle, *Mixed Effects Models for the Population Approach: Models, Tasks, Methods and Tools*. Chapman and Hall/CRC, 2014.
- [3] E. Comets, K. Brendel, and F. Mentré, “Computing normalised prediction distribution errors to evaluate nonlinear mixed-effect models: the npde add-on package for R,” *Computer Methods and Programs in Biomedicine*, vol. 90, no. 2, pp. 154–166, 2008.
- [4] E. Comets and K. Brendel, “Model evaluation in nonlinear mixed effect models, with applications to pharmacokinetics,” *Journal de la Société Française de Statistique*, vol. 151, no. 1, pp. 106–128, 2010.
- [5] M. Karlsson and R. Savic, “Diagnosing model diagnostics,” *Clinical Pharmacology & Therapeutics*, vol. 82, no. 1, pp. 17–20, 2007.
- [6] M. Lavielle and K. Bleakley, “Automatic data binning for improved visual diagnosis of pharmacometric models,” *Journal of Pharmacokinetics and Pharmacodynamics*, vol. 38, no. 6, pp. 861–871, 2011.
- [7] Y. Yano, S. L. Beal, and L. B. Sheiner, “Evaluating pharmacokinetic/pharmacodynamic models using the posterior predictive check,” *Journal of pharmacokinetics and pharmacodynamics*, vol. 28, no. 2, pp. 171–192, 2001.
- [8] F. Combes, S. Retout, N. Frey, and F. Mentré, “Powers of the likelihood ratio test and the correlation test using empirical bayes estimates for various shrinkages in population pharmacokinetics,” *CPT: pharmacometrics & systems pharmacology*, vol. 3, no. 4, pp. 1–9, 2014.

- [9] R. Savic and M. Karlsson, “Importance of shrinkage in empirical Bayes estimates for diagnostics: problems and solutions,” *The AAPS Journal*, vol. 11, no. 3, pp. 558–569, 2009.
- [10] R. Drikvandi, G. Verbeke, A. Khodadadi, and V. P. Nia, “Testing multiple variance components in linear mixed-effects models,” *Biostatistics*, vol. 14, no. 1, pp. 144–159, 2013.
- [11] Z. Li and L. Zhu, “A new test for random effects in linear mixed models with longitudinal data,” *Journal of Statistical Planning and Inference*, vol. 143, no. 1, pp. 82–95, 2013.
- [12] J. Mun and M. J. Lindstrom, “Diagnostics for repeated measurements in linear mixed effects models,” *Statistics in Medicine*, vol. 32, no. 8, pp. 1361–1375, 2013.
- [13] A. Alonso, S. Litière, and G. Molenberghs, “A family of tests to detect misspecifications in the random-effects structure of generalized linear mixed models,” *Computational Statistics & Data Analysis*, vol. 52, no. 9, pp. 4474–4486, 2008.
- [14] X. Huang, “Diagnosis of random-effect model misspecification in generalized linear mixed models for binary response,” *Biometrics*, vol. 65, no. 2, pp. 361–368, 2009.
- [15] E. F. Vonesh, V. M. Chinchilli, and K. Pu, “Goodness-of-fit in generalized nonlinear mixed-effects models,” *Biometrics*, vol. 52, no. 2, pp. 572–587, 1996.
- [16] G. Claeskens and J. Hart, “Goodness-of-fit tests in mixed models,” *Test*, vol. 18, no. 2, pp. 213–239, 2009.
- [17] C. Ritz, “Goodness-of-fit tests for mixed models,” *Scandinavian journal of statistics*, vol. 31, no. 3, pp. 443–458, 2004.
- [18] S. G. Meintanis and S. Portnoy, “Specification tests in mixed effects models,” *Journal of Statistical Planning and Inference*, vol. 141, no. 8, pp. 2545–2555, 2011.

- [19] J. Parke, “A procedure for generating bootstrap samples for the validation of nonlinear mixed-effects population models,” *Computer Methods and Programs in Biomedicine*, vol. 59, no. 1, pp. 19–29, 1999.
- [20] C. M. Laffont and D. Concordet, “A new exact test for the evaluation of population pharmacokinetic and/or pharmacodynamic models using random projections,” *Pharmaceutical research*, vol. 28, no. 8, pp. 1948–1962, 2011.
- [21] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. Springer Texts in Statistics, 2004.
- [22] N. Holford, “Clinical pharmacokinetics and pharmacodynamics of warfarin. understanding the dose-effect relationship,” *Clinical Pharmacokinetics*, vol. 11, no. 6, pp. 483–504, 1986.