ORIGINAL RESEARCH ARTICLE

Multivariate Exact Discrepancy: A New Tool for PK/PD Model Evaluation

Sarah Baklouti[1,2] · Emmanuelle Comets[3,4] · Peggy Gandia[1,2] · Didier Concordet[2]

Abstract

Background Pharmacokinetic models are evaluated using three types of metrics: those based on estimating the typical pharmacokinetic parameters, those based on predicting individual pharmacokinetic parameters and those that compare data and model distributions. In the third groups of metrics, the best-known methods are Visual Predictive Check (VPC) and Normalised Prediction Distribution Error (NPDE). Despite their usefulness, these methods have some limitations, especially for the analysis of dependent concentrations, i.e., evaluated in the same patient.

Objective In this work, we propose an evaluation method that accounts for the dependency between concentrations.

Methods Thanks to the study of the distribution of simulated vectors of concentrations, the method provides one probability per individual that its observations (i.e., concentrations) come from the studied model. The higher the probability, the better the model fits the individual. By examining the distribution of these probabilities for a set of individuals, we can evaluate the model as a whole.

Results We demonstrate the effectiveness of our method through two examples. Our approach successfully detects misspecification in the structural model and identifies outlier kinetics in a set of kinetics.

Conclusion We propose a straightforward method for evaluating models during their development and selecting a model to perform therapeutic drug monitoring. Based on our preliminary results, the method is very promising but needs to be validated on a larger scale.

1 Introduction

Population pharmacokinetic (PK) models are commonly used in the field of pharmacology. Building a model involves several steps, including an evaluation step, that is crucial in ensuring the model's predictive performance. The criteria used to evaluate a model belong to three families.

The first family of criteria comprises graphs that rely on estimating the typical PK parameters, which in turn enable the estimation of typical concentrations. These graphs serve the purpose of assessing the suitability of the structural model and provide visual insights into the variability between individuals as well as within individuals. They offer graphic representations that aid the evaluation of the appropriateness of the model and provide information regarding inter-individual and intra-individual dispersion. The second family of criteria groups together graphs

that are based on predicting the individual PK parameters. These graphs allow to check the appropriateness of the distribution assumptions, such as Individuals Weighted Residuals (IWRES) versus time or observed concentration versus individual prediction. However, due to shrinkage [1, 2], the prediction of individual PK parameters can sometimes be poor, especially when there are only a small number of concentrations per individual. While there are some corrections of shrinkage that can be considered [3], the average distance (actually the rate of convergence) to the actual value of these parameters cannot be smaller than a value inversely proportional to the number of subjects.

The third family of criteria aims to evaluate whether the concentrations used to build the model reasonably belong to the concentration distribution imposed by the model, which is obtained through simulation [4]. The gold standards of these methods are the Visual Predictive Check (VPC) and Normalised Prediction Distribution Errors (NPDE) [1]. The VPC is a graph that displays both the prediction interval provided by the model and the observed

---

Extended author information available on the last page of the article

Key Points

This study presents a novel methodology for evaluating pharmacokinetic models during their construction or selecting consistent models for therapeutic drug monitoring.

The main advantage of the method is its ability to account for the dependence between concentrations measured in the same individual.

The performances of the method have been evaluated through two simulation examples: to detect a misspecification in the structural model and to identify outliers in a set of kinetics.

concentrations at each time point [5–7]. Although this tool has been corrected in several forms to improve its performance, it does not account for the dependence between concentrations assessed in the same individual.

The second method involves calculating the NPDE, which is a tool derived from Prediction Discrepancy (PD) [8]. The PD of an observation corresponds to its percentile in its own predictive distribution. However, since the concentrations observed in the ith individual are not independent, it is not possible to evaluate the distribution of PD as soon as there is more than one observation per subject in the study. As a consequence, Brendel et al., proposed decorrelating PD to make it independent, resulting in the NPDE [9, 10]. However, as noted by the authors themselves, this implies independence only when the vector of concentrations of each individual is drawn from a Gaussian distribution [11]. A basic theorem of statistics says that the unique function of independent Gaussian random variables preserving normality is the linear (more precisely affine) function. As the nonlinear mixed effects model (NLME) is a nonlinear function of Gaussian random variables, the resulting vector of concentrations cannot be Gaussian. As a consequence, as the vector of concentrations is not Gaussian, decorrelation does not imply independence.

In the end, PD and NPDE are very useful in detecting significant deviations from the model when a single concentration, or several concentrations spaced in time, are available for each individual. In all other cases, it is expected that some concentrations will be incorrectly classified as outliers, while others will be erroneously considered compatible with the studied model.

In this article, we propose a method that allows to evaluate a model accounting for the dependence between concentrations in the same individual. This method is based on a free adaptation to our context of the ideas presented in the first chapter of the book by Chacón and Duong [12].

2 Materials and Methods

This section comprises four subsections. The first subsection presents the model and notation that will be used throughout the paper. In the second subsection, we explain the theoretical basis of the proposed method. The third subsection shows the implementation of the method in the field of pharmacology, and the fourth subsection is a simulation study that (1) illustrates the method's application and (2) evaluates its performance in two aspects: detecting misspecifications in the structural model and identifying individuals with outlier kinetic profiles.

2.1 Model and Notation

Let us consider the following population PK model

$$\begin{cases} Y_t = m(t, \varphi, C) + g(t;\varphi;C;\sigma)\varepsilon_t \text{ where } \varepsilon_t \sim_{\text{iid}} \mathbb{N}(0, 1) \\ \varphi = \qquad\qquad h(\theta, C, \eta) \text{ where } \eta \sim \mathbb{N}(0, \Omega) \end{cases} \quad (1)$$

This model describes the evolution of the concentrations $(Y_t)_t$ with time t in an individual according to some of its individual characteristics C (covariates) and known population PK parameters (θ, Ω, σ).

φ represents the vector of individual PK parameters.

m represents the function describing the structural model, g represents the function describing the error model and h represent the function describing the relation between the individual PK parameters φ, the covariate C, the population parameters θ and the individual random effects η.

Assume now that n_i concentrations $(Y_{i1}, \ldots, Y_{in_i})$ have been observed at times $(t_{i1}, \ldots, t_{in_i})$ in the ith individual of a sample of N individuals. The covariates values of all individuals (C_i) are also assumed to be known.

2.2 The Theoretical Basis

Consider a random variable $X \in \mathcal{R}^p$ with a probability distribution function (pdf) denoted f. For technical reasons that can easily be relaxed, we assume that f is non-constant on all non-empty subsets of $\mathcal{R}^p$. For all $\lambda \geq 0$, we consider the level set of all x such that $f(x)$ is greater than λ:

$$A_\lambda = \{x \in \mathcal{R}^p, f(x) \geq \lambda\}$$

Notice that the sequence of sets $\left(A_\lambda\right)_{\lambda \geq 0}$ is decreasing for the inclusion. In other terms, for all $0 \leq \lambda_1 \leq \lambda_2, A_{\lambda_2} \subseteq A_{\lambda_1}$.

Let us now consider the function

$$F(\lambda) = 1 - \int f(x) 1_{A_\lambda}(x) \mathrm{d}x = \int f(x) 1_{A_\lambda^c}(x) \mathrm{d}x$$

where $A_\lambda^c = \{x \in \mathcal{R}^p, f(x) < \lambda\}$ and $1_{A_\lambda^c}(x) = 1$ when $x \in A_\lambda^c$ and $1_{A_\lambda^c}(x) = 0$ otherwise.

Using these sets, we can see that for all $\lambda \geq 0$,

$$1 - F(\lambda) = P\left(X \in A_\lambda\right) = P(f(X) \geq \lambda) \tag{2}$$

The sets A_λ are, therefore, the prediction regions of the random vector X. By their very construction, they are even the smallest (with respect to their volume) prediction regions of X that can be built. This idea for constructing small prediction regions has already been proposed for multidimensional discrete random variables [13].

Now, let us inverse the process of Eq. (2), i.e., we fix X and look for the λ verifying $X \in A_\lambda$. We thus consider the random variable

$$\lambda(X) = \sup\{\lambda \geq 0 \text{ so that } f(X) \geq \lambda\} \tag{3}$$

For any $\lambda_1 \geq 0$, $P(\lambda(X) > \lambda_1) = P(f(X) > \lambda_1) = 1 - F(\lambda_1)$. From this last equality we derive that the cumulative distribution function of the univariate random variable $\lambda(X)$ is F. This allows to write the following property:

Let X be a multidimensional random variable, $F(\lambda)$ and $\lambda(X)$ be respectively defined as in Eqs. (2) *and* (3), *then $P(X) \triangleq F(\lambda(X))$ is distributed according to a uniform distribution on]0,1[.*

A noticeable property of $P(X) = F(\lambda(X))$ is that its distribution does not depend on the dimension of X.

2.3 Application to Population Pharmacokinetic Model

This theoretical property can be applied to PK models. Returning to our notations, we consider N vectors, $y_1, \ldots, y_N$, containing, respectively, the concentrations observed in N independent individuals. An unknown model, denoted M_0, is at the origin of the observed concentrations.

We want to test whether the observed concentrations can be considered to come from the population PK model described by Eq. (1). Because this model depends on the unknown parameter $\tau = (\theta, \Omega, \sigma)$ we denote this model M_τ. Let us denote $\hat{\tau}$ the estimation of τ obtained from the observed concentrations $y_1, \ldots, y_N$. Conditionally on $\hat{\tau}$, i.e.

assuming that the actual value of τ is $\hat{\tau}$, our initial question, "is the model M_τ valid?" translates into "is $M_{\hat{\tau}}$ equal to M_0?".

To answer this question, we first have to compute the probability $P(y_i)$ of observing the concentration vector y_i for each individual.

For the ith individual, using the population model given by the Eq. (1), we simulate vectors of concentrations $(Y_{it_1}, \ldots, Y_{it_n})$ at the same time points at which the individual was sampled. We set f to be the pdf of the joint distribution of the concentration vector.

We need to compute:

1. the function $F(\lambda) = P\left(f\left(Y_{it_1}, \ldots, Y_{it_n}\right) \leq \lambda\right)$.
2. $\lambda(y_i)$, where $y_i = (y_{i1}, \ldots, y_{in})$ is the vector containing the concentrations observed in the ith individual; y_i should not be confused with the theoretical concentrations $(Y_{it_1}, \ldots, Y_{it_n})$ simulated with the model (Eq. 1). A large value of $\lambda(y_i)$ indicates that the observed kinetic profile in the individual is in the "heart" of the distribution of the kinetic profiles given by the model. On the contrary, a low value of $\lambda(y_i)$ indicates that the observed kinetic profile is far from the one expected by the model.
3. the multivariate exact discrepancy (MED) is defined as $P(y_i) = F(\lambda(y_i))$ that, when multiplied by 100, gives the percentage of kinetic profiles given by the model that are more "extreme" than the observed profile. For example, if $F(\lambda(y_i)) = 1 \times 10^{-6}$, it means that only one individual in a million has a kinetic profile from the model that is more extreme than the one observed in our individual. In other words, the model is inadequate for describing the kinetic of the individual.

Because f is the pdf of a marginal distribution, it is quite difficult to obtain directly through computation. Instead, we propose to approximate f through simulations as follows:

(a) Simulate K vectors $\left(Y_{it_1}^{(k)}, \ldots, Y_{it_n}^{(k)}\right)_{k=1,\ldots,K}$ using model Eq. (1).
(b) Compute an approximation $\tilde{f}$ of f using a kernel estimator based on the K vectors simulated at step (a). To approximate $\tilde{f}$, we chose to use the function kde available in the package ks in R.
(c) For a given λ, an approximation $\tilde{F}(\lambda)$ of $F(\lambda)$ is given by the percentage of $\tilde{f}\left(Y_{it_1}^{(k)}, \ldots, Y_{it_n}^{(k)}\right)$ smaller than λ; in other terms,

$$\tilde{F}(\lambda) = \frac{1}{K} \sum_{k=1}^{K} 1_{\left[\tilde{f}\left(Y_{it_1}^{(k)}, \ldots, Y_{it_n}^{(k)}\right) \leq \lambda\right]} \tag{4}$$
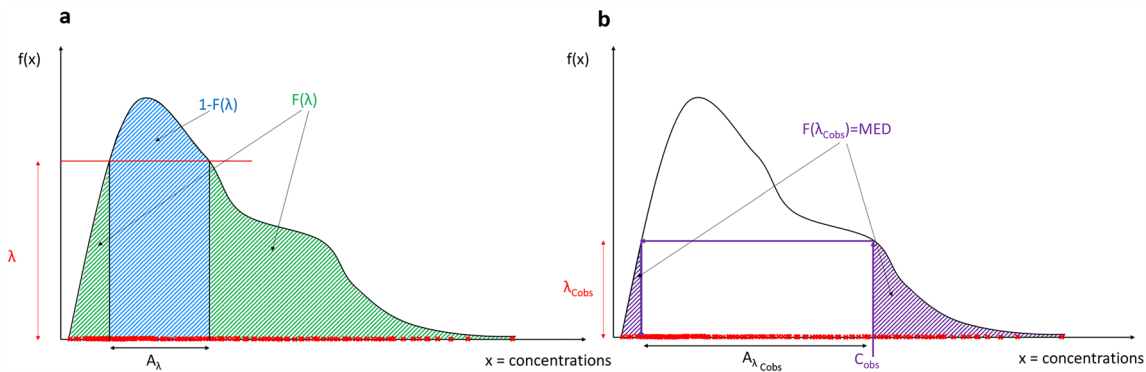
Fig. 1 Schematic representation of the multivariate exact discrepancy (MED) computation in one dimension. The **a** (on the left) represents the different notation used to compute the MED. The x-axis represents the concentrations that can be observed with the tested model at a given time with fixed values of covariates. The y-axis represents the probability density function $f(x)$ of these concentrations. The red crosses represented on the x-axis are concentrations simulated with the model. A large value of $f(x)$ indicates a large density of (simulated) concentrations. For a given value of λ, we can compute A_λ, the set of concentrations whose density is higher than λ. From A_λ, we can calculate $1 - F(\lambda) = P(X \in A_\lambda)$, that can be seen as the percentage of simulated concentrations with a density higher than λ, and $F(\lambda)$, the percentage of simulated concentrations with a density lower than lambda. Now that these sets have been defined, let's see how to use them to define MED. The figure **b** (on the right) illustrates how to compute the MED for a patient for whom we have measured one concentration (C_{obs}) represented on the x-axis. The y-axis still represents the pdf $f(x)$. We can calculate the set $A_{\lambda C_{obs}}$ for which C_{obs} is at the boundary and the corresponding λ that will be denoted $\lambda_{C_{obs}}$. The MED are defined as the probability that a concentration does not belong to $A_{\lambda C_{obs}}$ that can be seen as the percentage of simulated concentrations standing in a region where $f(x) \le \lambda_{C_{obs}}$. If the MED is large, the observed concentration stands in a region of high density and the concentration is consistent with the model. On the contrary, if the MED is low, we have a low probability of observing the concentration with the model tested

The quality of the approximation of f and $F(\lambda)$ depends solely on the number of simulated vectors K. The greater the number of concentrations per individual, the greater K. We have no guidelines for choosing K at present.

A summary of the MED calculation for one patient is shown in Fig. 1. To make it easier to understand, the Fig. 1 shows the MED calculation when only one concentration is available for one patient.

If we repeat the proposed method to each individual, we obtain one probability per individual, $P(y_1), \dots, P(y_N)$.

If we return to the question of whether or not $M_{\hat{\tau}}$ describes the data well, we can state that if $M_{\hat{\tau}} = M_0$, the $P(y_i)$ values should be distributed according to a uniform distribution on $[0, 1]$. Therefore, if the distribution of the $P(y_i)$ values is not uniform, $M_{\hat{\tau}} \ne M_0$ and the model $M_{\hat{\tau}}$ is invalidated.

It remains to be checked whether the distribution of the $P(y)$'s is uniform on $[0, 1]$. Any test that compares a sample distribution to a reference distribution (uniform on $]0; 1[$) can be used. In this context, tests such as the Kolmogorov-Smirnov test, Anderson Darling test, chi-square tests are appropriate; each having its own advantages and drawbacks. In the remainder of this article, we will use the Kolmogorov-Smirnov test without discussing this choice.

To summarise the interpretation of MED, we know that MED coming from the true model are expected to be drawn from a uniform distribution over $[0, 1]$. There is thus a small probability ($<$ alpha $= 5\%$) to observe a MED less than 5%

coming from the true model. In this respect, MED can be interpreted as a *p*-value.

The great advantage of this method is its ability to identify outliers. Indeed, an individual's probability determines whether or not the individual's kinetic belongs to the model. To do that, a probability threshold must be defined: if an individual's probability falls below this threshold, then the individual is identified as an outlier. When an individual is considered as an outlier, their entire PK profile is rejected, not just certain concentrations. Detection of outliers can be useful both in building a model and in performing therapeutic drug monitoring.

Usually in statistics, an outlier is defined as a value that has a small probability to come from the studied model. Similarly, we will classify an individual as an outlier if their kinetic profile has a small probability of being generated by the population PKs under investigation. The search of outlier kinetics is quite distinct from the search of outlier concentrations, which involves comparing each concentration to a chosen model without considering the other concentrations obtained in the same individual.

Consider N vectors, $y_1, \dots, y_N$, containing the concentrations observed in N independent individuals. These vectors do not necessarily have the same size because the number of concentrations per individual may vary. Additionally, these concentrations may not have been obtained at the same times post-administration. Our goal is to identify individuals with

a kinetic profile that can be considered as outliers with respect to the model described by Eq. (1). This model contains estimation of population parameters, $\hat{\tau}$, obtained from the observed concentrations $y_1, \ldots, y_N$. We assume that $\hat{\tau}$ is the true value of τ and is known without any imprecision. For all individuals, we can compute $P(y_1), \ldots, P(y_N)$ using the method described above. When N is large, we can expect to observe very low and very large values of the $P(y_i)$ even if none of the kinetics are outliers. The threshold value below which the $P(y_i)$'s can be considered abnormal can be derived by their minimum.

Let us consider LL_α the limit defined by

$$P\left(\min\left\{P(y_1), \ldots, P(y_N)\right\} \le LL_\alpha\right) = \alpha \qquad (5a)$$

The meaning of this limit is the following: only $100 \times \alpha\%$ of the samples of N individuals should exhibit a kinetics (y_i) coming from the model (Eq. 1) with $P(y_i) \le LL_\alpha$.

In other terms, if α is chosen small enough (for example, $\alpha = 0.05$), a kinetic profile y_i with $P(y_i) \le LL_\alpha$ should be considered as outlier.

From the property, we know that when $y_1, \ldots, y_N$ are drawn in the model described by (Eq. 1) then $P(y_i) \sim_{iid} U([0, 1])$. It follows that

$$LL_\alpha = 1 - (1 - \alpha)^{1/N} \qquad (5b)$$

2.4 Simulation Study

The objective of the simulation section is (1) to illustrate how the method works using a very simple example and (2) to evaluate the performance of the method in detecting a problem in the model through two examples.

All the population PK models used in this article have arbitrarily chosen population parameters and variance. They were chosen to ensure a good illustration of the method.

2.4.1 Illustrative Example

This simplified example demonstrates the practical application of the method within the context of TDM (Therapeutic Drug Monitoring). In this scenario, we have measured two

concentrations in a patient and our objective is to identify compatible models.

Assuming a dose $D = 100$ was administered intravenously to an individual at time 0. We observed two concentrations $(y_1, y_2) = (1.22, 0.011)$ in this individual at times $t_1 = 39$ and $t_2 = 99$, respectively.

We would like to determine if the following published model can well describe the individual's kinetics.

$$\begin{cases} Y_t &= \frac{D}{V}e^{-\frac{cl}{V}t}(1 + 0.1\varepsilon_t) \text{ where } \varepsilon_t \sim_{iid} \mathbb{N}(0, 1) \\ \begin{pmatrix} cl \\ V \end{pmatrix} &= \begin{pmatrix} 0.1\exp(\eta_{cl}) \\ 1\exp(\eta_V) \end{pmatrix}, \text{ where } \begin{pmatrix} \eta_{cl} \\ \eta_V \end{pmatrix} \sim \mathbb{N}\left(0, \begin{pmatrix} 0.2^2 & 0 \\ 0 & 0.1^2 \end{pmatrix}\right) \end{cases} \qquad (6)$$

The model described in Eq. (6) has the same structure as the one of Eq. (1) with $\theta = \begin{pmatrix} cl \\ V \end{pmatrix}$, there is no covariate so no need to write a matrix C containing their values, $m(t, \varphi) = \frac{D}{V}e^{-\frac{cl}{V}t}$, $g(t, \varphi, \sigma) = 0.1 \times \frac{D}{V}e^{-\frac{cl}{V}t}$, $h(\theta, \eta) = \begin{pmatrix} 0.1\exp(\eta_{cl}) \\ \exp(\eta_V) \end{pmatrix}, \Omega = \begin{pmatrix} 0.2^2 & 0 \\ 0 & 0.1^2 \end{pmatrix}$.

We simulated $K = 100{,}000$ vectors (Y_{39}, Y_{99}) using the model Eq. (6).

The kernel density estimator was constructed using simulations with the ks library in R.

The individual's probability was calculated using Eq. (4). For comparison, we analysed the results provided by VPC and MED graphically.

2.4.2 Evaluation of a Population Pharmacokinetic Model

This first example evaluated the method's ability to detect a misspecification in the structural model. In some published population PK studies, the PK model used to describe individual kinetics is a one compartmental model. These studies can exhibit important residual error. For instance, models with a combined or proportional residual error model that show a coefficient of variation of the residual error much greater than the coefficient of variation of the analytical method used.

The following simulations evaluate the performance of the proposed method in detecting a model that is too simplistic for analysing data. The data are assumed to come from a bi-compartmental model (described by Eq. 7 below) but are analysed using a one-compartmental model (Eq. 8).

The following 2-compartment model was used to generate $L = 200$ sets of $N = 200$ kinetics, each of these including 2 time points.

$$\begin{cases} y_{ij}^{(l)} = \left(A_i^{(l)}e^{-\alpha_i^{(l)}t_{ij}^{(l)}} + B_i^{(l)}e^{-\beta_i^{(l)}t_{ij}^{(l)}}\right) \times \left(1 + 0.2\varepsilon_{ij}^{(l)}\right) \text{ where } \varepsilon_{ij}^{(l)} \sim_{iid} \mathbb{N}(0, 1) \\ \log\left(A_i^{(l)}, \alpha_i^{(l)}, B_i^{(l)}, \beta_i^{(l)}\right) \sim_{iid} \mathbb{N}\left((\ln(100), \ln(0.1), \ln(10), \ln(\beta)), 0.1^2 I_4\right), \, l = 1, \ldots, L, \, i = 1, \ldots, n, \, j = 1, 2 \end{cases} \qquad (7)$$
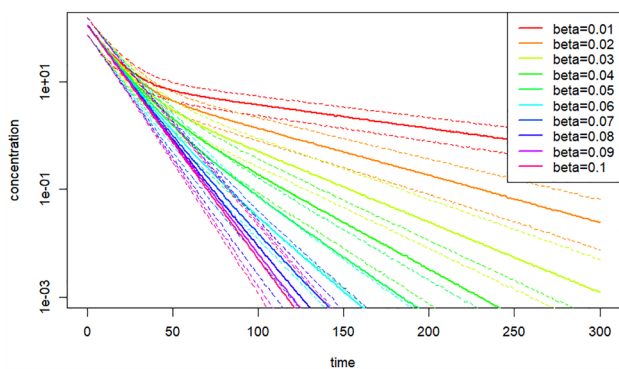
Fig. 2 This figure represents the 5%, 50% (median) and 95% percentiles of concentrations (on a logarithmic scale) simulated with a 2-compartment model (Eq. 7) for several values of β (slow or terminal decay slope) ranging from the 0.01 to 0.1. The curves obtained with $\beta = 0.1$ corresponds to a one-compartment model (straight line). The curves obtained from $\beta = 0.06$ to $\beta = 0.09$ are almost indistinguishable from the one obtained with $\beta = 0.1$

Where iid means independent and identically distributed and I4 refers to a 4×4 identity matrix

For each value of β within the interval [0.01, 0.1] with an increment of 0.01, 200 sets of 200 simulations were performed. The kinetic profiles resulting of these simulations are shown in Fig. 2.

For all simulated individuals, the first sampling time $t_{i1}^{(l)}$ was drawn from a uniform distribution on [0, 125] and the second sampling time was set to $t_{i1}^{(l)} + 125$. This way of choosing the sampling times ensures that the two time points in a kinetic profile are sufficiently spaced. Next, for each value of β, the closest one compartment model was looked for by analysing several datasets generated with Eq. (7) using the following model:.

$$\begin{cases} Y_{ij} = A_i e^{-\alpha_i u_{ij}} \times \left(1 + \sigma \varepsilon_{ij}\right) & \text{where } \varepsilon_{ij} \sim_{\text{iid}} \mathbb{N}(0,1) \\ \log\left(A_i, \alpha_i\right) \sim_{\text{iid}} \mathbb{N}\left(\left(\ln\left(\theta_A\right), \ln\left(\theta_\alpha\right)\right), \begin{pmatrix} \omega_A^2 & 0 \\ 0 & \omega_\alpha^2 \end{pmatrix}\right) \end{cases} \quad (8)$$

We built the models Eq. (8) using Monolix. To define the best one-compartmental models (i.e., the models closest to model Eq. 7), we used 1500 individuals with 10 observations per individual. The times $\left(u_{ij}\right)$ for all individuals were uniformly selected from the time interval [0, 250].

The model described by Eq. (8) was used to simulate K = 10,000 vectors at each time point $t_{ij}^{(l)}$ to compute $P_i^{(l)} = \tilde{F}_i^{(l)}\left(\left(y_{i1}^{(l)}, y_{i2}^{(l)}\right)\right)$ as described by Eq. (4). For each set (each $l = 1, \dots, L$), the distribution of the $\left(P_i^{(l)}\right)_{i=1,\dots,n}$ was compared to a uniform distribution over [0, 1] using a Kolmogorov-Smirnov test.

The 200 sets built with $\beta = 0.1$ (i.e., the true model) allows us to evaluate the type I error, as the percentage of the 200 sets where the hypothesis of uniformity distribution of the $\left(P_i^{(l)}\right)_{i=1,\dots,n}$ was rejected. Similarly, the power of detecting the wrong model (beta different from 0.1) was evaluated by the percentage of p-values given by the Kolmogorov-Smirnov test that were less than 0.05, indicating rejection of the uniformity hypothesis made on the distribution of the $\left(P_i^{(l)}\right)_{i=1,\dots,n}$.

We chose to compare the results obtained with MED and NPDE as a reference method. However, as we were only interested in comparing the power of the tests, we used the Prediction Distribution Error (PDE), a non-standardized variable, instead of NPDE. PDE were constructed in the same way as MED (10,000 concentration vectors simulated for each individual) and tested for uniformity using a Kolmogorov-Smirnov test. As for MED, we evaluated the type I error and the power of PDE.

The power tests obtained with our method and with PDE were compared graphically.

2.4.3 Detection of Individuals with Outlier Kinetics

The second example evaluates the performance of the method to detect outlier's kinetic of individuals. The aim of this example is to evaluate the ability of the method to reject a model based on the percentage of outliers.

Let us consider the following population PK model

$$\begin{cases} Y_{ij} = \frac{100 Ka_i}{V_i(Ka_i - Cl_i/V_i)}\left(e^{-\frac{Cl_i}{V_i}t_j} - e^{-Ka_i t_j}\right) \times \left(1 + 0.1 \times \varepsilon_{ij}\right) & \text{where } \varepsilon_{ij} \sim_{\text{iid}} \mathbb{N}(0,1) \\ \begin{pmatrix} Ka_i \\ V_i \\ Cl_i \end{pmatrix} = \begin{pmatrix} \exp(\eta_i^{Ka}) \\ 10 \exp(\eta_i^{V}) \\ \theta_{Cl} \exp(\eta_i^{Cl}) \end{pmatrix}, & \text{where } \begin{pmatrix} \eta_i^{Ka} \\ \eta_i^{V} \\ \eta_i^{Cl} \end{pmatrix} \sim_{\text{iid}} \mathbb{N}\left(0, \begin{pmatrix} 0.1^2 & & \\ 0 & 0.1^2 & \\ 0 & 0 & 0.1^2 \end{pmatrix}\right) \end{cases}$$
$$(9)$$

This model, with $\theta_{cl} = 2$ L/h, serves as the reference model. Therefore, kinetics simulated with this model ($\theta_{cl} = 2$ L/h) were used to compute $\tilde{f}$ as previously described.

Next, a sequence of values of $\theta_{cl} \in [1;3]$ by step of 0.1 was used to generate outlier kinetics. Three sets of arbitrarily fixed observations times were used. The first set of observation times contained 2 times: $t_1 = 3$ and $t_2 = 33$; the second set contained 3 times: $t_1 = 3, t_2 = 16$ and $t_3 = 33$ and finally, the third set contained 4 times: $t_1 = 3, t_2 = 10, t_3 = 20$, and $t_4 = 33$.

For each simulated outlier kinetic (y), we computed a corresponding $P(y)$. The kinetics (y) for which $P(y) \leq LL_{0.05}$ were identified as outliers. We computed the percentages of

detected outlier kinetics (also called power) for $N = 100$, $N = 200$ and $N = 10,000$.

As there are no other methods available to detect outliers, we did not compare our tool with any other method.

3 Results

3.1 Illustrative Example

The plot of percentiles provided by the model described by Eq. (6) are presented in Fig. 3. We have plotted the two observed concentrations in red on the plot. It is evident that this plot suggests these two concentrations to plausibly come from the model described in Eq. (6).

However, when looking at the shape of the kinetics (in a logarithmic scale), all the log-concentrations of a kinetics are grouped around a straight line. The plot of percentiles roughly show the limits of the straight lines that can be encountered. It is also clear that the "beginning" of the kinetics represented by a dot line cannot belong to the 90% prediction interval.

Figure 4 represents the value of $\ln(Y_{99})$ as a function of $\ln(Y_{39})$, with different coloured areas representing the sets A_λ for various values of λ. The darker coloured areas were obtained with the largest values of λ and they correspond to the area with the highest density of the simulated points. Similarly, the areas drawn with light colours correspond to low density of simulated points. The red point $(\ln 1.22, \ln 0.011)$ represents the observed concentrations
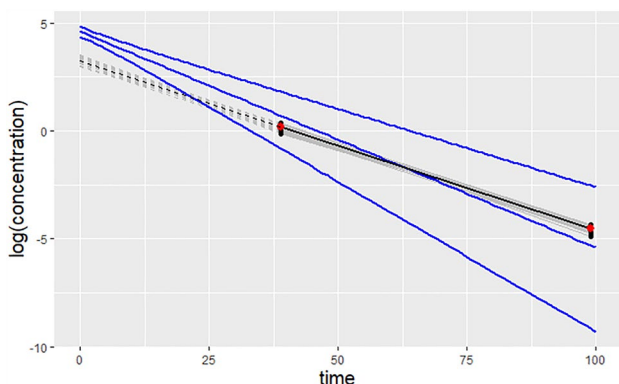

Fig. 4 The x and y axes of this figure represent, respectively, the concentrations (in log-scale) simulated with the 1-compartment model (Eq. 6) at times 39 and 99 (C_39 and C_99). The shapes represented (close to ellipses) are the sets A_λ for various values of λ. The sets shown in red were obtained with the largest values of λ and correspond to the area where most of the concentrations simulated with a 1-compartment model (Eq. 6) are found (in logarithmic scale). On the contrary, the sets represented in light yellow correspond to the area where only few concentrations from the 1-compartment model (Eq. 6) can be observed. The area in white (the rest of the picture) corresponds to the area where none of the 100,000 concentrations simulated with the 1-compartment model (Eq. 6) were observed. In this white area, lambda is close to 0. Finally, the point in red correspond to the two concentrations observed in the individual. These two concentrations fall in a white area indicating that the 1-compartment model (Eq. 6) cannot be used to describe the individual's kinetics

in the individual. The value of λ from which the level set of the pdf contains $(\ln 1.22, \ln 0.011)$ is close to zero. More precisely $\lambda(\ln 1.22, \ln 0.011) \simeq 7 \times 10^{-10}$. Finally, $\tilde{F}(\lambda(\ln 1.22, \ln 0.011)) = 0$ indicating that none of the simulated vectors given by Eq. (6) have a lower pdf value, showing that $\tilde{F}((\ln 1.22, \ln 0.011)) \leq 1/100,000$.

In conclusion, there is a probability $\leq 10^{-5}$ that these concentrations come from the model Eq. (6).

3.2 Evaluation of a Population Pharmacokinetic Model

In Table 1, we present the estimated parameter values obtained by fitting a one-compartment model to the two-compartment model simulations according to the value of β indicated in the first row. As the kinetics become more bi-compartmental (decreasing beta), the one-compartment model becomes less suitable for the data, resulting in an increase in residual variability.

Figure 5 shows the type I error rate when $\beta = 0.1$ (i.e., one-compartment model). The power of detecting the wrong model by our method and by PDE is presented for different values of β.

With our method, the kinetics simulated with a value of β less than 0.05 are always identified as being simulated with a model that is not a one-compartment model: the power is

Fig. 3 The blue lines represent the 5%, 50% (median) and 95% percentiles of concentrations (in logarithmic scale) simulated with a 1-compartment model (Eq. 6). The two red points represent the observed concentrations in a patient and the black points the variability expected around these points. As both points fall within the 90% prediction interval, the usual analysis of this figure is misleading. Indeed, both observed concentrations belong to the prediction interval, which suggests that the two concentrations observed on the patient are consistent with the model. However, taken as a whole the possible kinetic profiles are not compatible with the 1-comparmental model described in the Eq. (6)
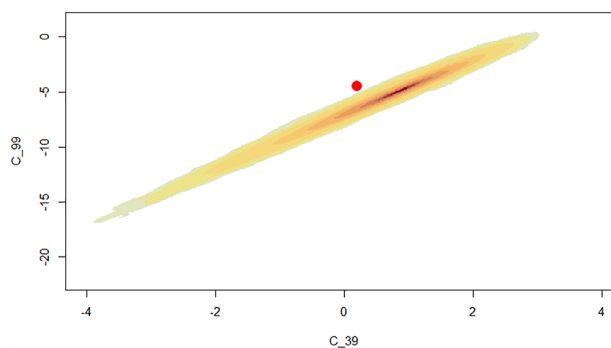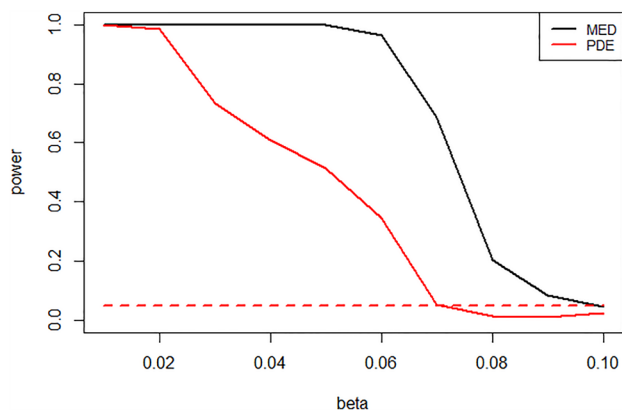
Table 1 Mono-compartmental population pharmacokinetic model based on β

β	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.1
θ_A	30.30	32.26	34.48	37.04	40.00	47.62	58.82	76.92	90.91	110
θ_α	0.017	0.027	0.037	0.048	0.058	0.069	0.078	0.087	0.094	0.1
ω_A	0.099	0.094	0.1	0.12	0.14	0.19	0.26	0.27	0.24	0.1
ω_α	0.069	0.083	0.085	0.087	0.088	0.091	0.089	0.083	0.084	0.1
σ	0.52	0.52	0.51	0.5	0.48	0.44	0.37	0.3	0.26	0.2

Fig. 5 This figure represents the probability of detecting that a 2-points kinetics simulated with the 2-compartment model (Eq. 7) does not come from the 1-compartment model (Eq. 8) for 2 different methods: the method proposed in this paper and the Prediction Distribution Error (PDE) (reference method). Several values of β (beta) were used to simulate the 2-points kinetics
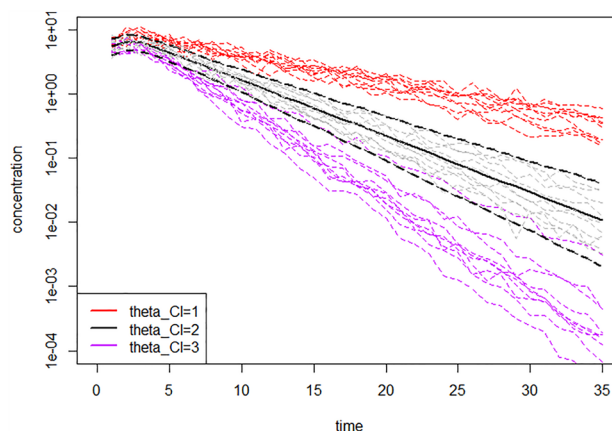

Fig. 6 This figure displays the 5th, 50th (median) and 95th percentiles of concentrations (in log-scale) simulated with a 1-compartment model and an extravascular administration (Eq. 9) for the reference clearance (2 L/h), represented in black. Examples of kinetic profiles obtained with different θ_{cl} are superimposed. Specifically, profiles with a θ_{cl} equal to 1 L/h, 2 L/h and 3 L/h are, respectively, presented in purple, grey and red

100%. This probability decreases when beta get closer to 0.1, which corresponds to a one-compartment model. The advantage of the proposed method is that it allows distinguishing a two-compartment model from a one-compartment model, even for $\beta = 0.06$ or 0.07, while visual examination of the data cannot achieve this (as shown in Fig. 2). The "power" obtained with $\beta = 0.1$ corresponds to the level of the statistical test: 0.05.

With the PDE method, only the kinetics simulated with a value of β equal to 0.01 are always identified as being simulated with a model that is not a one-compartment model. For the value of β between 0.07 and 0.09, the PDE method always fails to distinguish a two-compartment model from a one-compartment model.

3.3 Detecting |Individuals with Outlier PK

The detection of outliers is presented with various numbers of available subjects, denoted N. Increasing N allows for the evaluation of the method's power to detect outliers based on the number of available data. When $N = 10,000$, the number of detected outliers converges to its true value.

Figure 6 presents the plot of percentiles of the true model (θ_{cl} equal to 2 L/h) along with examples of outlier kinetic profiles simulated with θ_{cl} of 1 L/h (in red), 2 L/h (in grey) and 3 L/h (in purple).

The results obtained with 2 and 3 observations are presented in Fig. 7. The curves obtained with four observation times were found to be superimposable with the one obtained using only 3 observation times.

The decision rule used to identify a kinetic profile as an outlier was designed to identify the most extreme (i.e., with the smallest $P(y)$) observable kinetics among N kinetics. Therefore, this decision rule depends on the number of kinetics among which we search to identify outliers. Figure 7 shows that as the value of θ_{cl} deviates further from 2 L/h, it becomes easier for the proposed method and decision rule to identify kinetics simulated with θ_{cl} different from 2 and analysed with $\theta_{cl} = 2$. On the other hand, simulated kinetics with θ_{cl} between 1.5 and 2.5 are not identified as outliers, regardless of the number of observation times used.
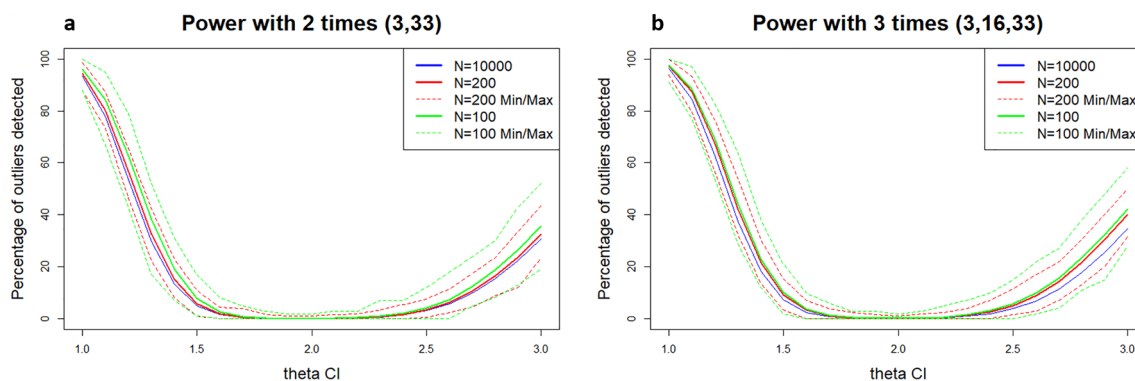
Fig. 7 The left (resp. right) figure shows the percentage of 2-point (resp. 3-point) kinetic profiles simulated with the 1-compartment model and an extravascular administration (Eq. 9) with θ_{cl} ranging from 1 to 3 L/h and identified as coming from a 1-compartment model and an extravascular administration (Eq. 9) with $\theta_{cl} \neq 2.0$ L/h. The decision rule used to declare a kinetic profile as not compatible with $\theta_{cl} = 2.0$ L/h is described by Eqs. (5a) and (5b)

4 Discussion

In this article, we propose a new tool for evaluating population PK models. Our method enables the computation of a probability for each individual that their concentration profile comes from the model. The implementation of this method is done in 3 steps: (a) simulating kinetics using the tested model; (b) calculating the pdf with the simulated kinetics; and (c) calculating the multivariate exact discrepancy for each observed kinetics ($P(y)$) with the pdf. The proposed method can be used regardless of individual sampling times and their covariate values.

We demonstrated the performance of our method through two simulation examples. The first example evaluated its ability to detect a misspecification in the structural model. Our method showed good statistical properties and outperformed PDE in detecting the incorrect structural model. One possible explanation is that our method fully exploits the non-independence between the concentrations measured in an individual.

The second example evaluates the ability of the method to detect outliers kinetic. This novel application is highly interesting since there is currently no other method to identify outliers. The rationale behind this application is that, during the model building process, if too many observed kinetics are not compatible with the model, the model should be changed. However, we must recognise that the identification of outlier kinetics based on the law of extremes (the minimum) did not give the expected results. Indeed, we were unable to identify as outliers the simulated kinetics whose mean clearance values were close to the reference clearance value. Further work on this subject, including exploring other decision rules, is necessary to enable better identification of kinetics that deviate significantly from the model.

Another direct application of this method is to use it for selecting a population PK model in the context of therapeutic drug monitoring (TDM). Currently, in hospital practice, model evaluation remains difficult due to the lack of guidelines [14]. However, selecting the correct model is a crucial step to ensure that the predictions made are accurate and that the resulting dosage regimen is appropriate. The main difficulty is due to the lack of an appropriate tool, i.e., a method to select a model at the individual level [15]. The proposed method shows great promise for this application as it provides one probability per individual that their kinetics belong to a model. This probability could therefore be used to select a model consistent with the observed concentrations at an individual level rather than a population level. It can thus be used to flag individuals with atypical kinetics, and this information can be used to indicate uncertainty in dose recommendations, as well as to request additional blood samples to better characterise its PK profile. But, as the method evaluates an entire kinetic profile, an erroneous observation (due to a wrong dosage or sampling time) could lead to an individual being wrongly identified as an outlier. This is a limitation of this version of the method.

A much more detailed simulation analysis should be carried out to confirm (or refute) these preliminary findings. We still need to test the method's ability to detect various other specification errors encountered in the model-building process (e.g., evaluation of a covariate model…) as well as its ability to perform external validation for TDM purposes.

Moreover, while the results obtained seem encouraging, it is important to note that, like all other model evaluation methods, we reason conditionally based on the population parameters. In other words, the imprecision of the population parameters is not accounted for. This is a limit of the

method that we do not know how to overcome. This limitation requires us to work with the best available estimators of the population parameters, i.e., with the maximum likelihood.

Finally, there is still important work to be done on the estimation of the pdf $\tilde{f}$.

When there are few observations per individual (less than 4), the ks library in R can quickly calculate the MED. However, when the number of observations per individual is greater than 5, ks takes a long time to estimate f. This is probably because ks tries to estimate the optimal bandwidth (smoothing parameter), which can be a time-consuming process. It is likely, but still needs to be verified, that crude (non-optimal) bandwidths would allow model evaluation and outlier search questions to be answered efficiently while also being computed quickly.

In conclusion, the method proposed in this article shows great promise for both model evaluation and selection of model to perform TDM. However, it needs to be tested on a large scale in order to fully assess its effective power, especially in the model building process.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40262-023-01296-6.

Declarations

Funding No external funding was used for this work.

Conflicts of interest Sarah Baklouti, Emmanuelle Comets, Peggy Gandia, and Didier Concordet declare that they have no conflict of interest that might be relevant to the contents of this manuscript.

Availability of data and material Not applicable.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Code availability The code for example 1 is available in supplemental data.

Author contributions DC proposed the methodology for model evaluation, SB and DC performed simulations. DC, EC, PG, and SB drafted the paper. All authors critically reviewed several drafts and approved the final manuscript.

References

1. Nguyen THT, Mouksassi M-S, Holford N, Al-Huniti N, Freedman I, Hooker AC, et al. Model evaluation of continuous data pharmacometric models: metrics and graphics. CPT Pharmacometr Syst Pharmacol. 2017;6:87–109.
2. Karlsson MO, Savic RM. Diagnosing model diagnostics. Clin Pharmacol Ther. 2007;82:17–20.
3. Baklouti S, Gandia P, Concordet D. "De-Shrinking" EBEs: the solution for bayesian therapeutic drug monitoring. Clin Pharmacokinet. 2022;61:749–57.
4. Sheiner LB. Learning versus confirming in clinical drug development. Clin Pharmacol Ther. 1997;61:275–91.
5. Post TM, Freijer JI, Ploeger BA, Danhof M. Extensions to the visual predictive check to facilitate model performance evaluation. J Pharmacokinet Pharmacodyn. 2008;35:185–202.
6. Holford N. The visual predictive check—superiority to standard diagnostic (rorschach) plots; 2005. p. 14.
7. Karlsson MO, Holford N. A tutorial on visual predictive checks. In: PAGE. Abstracts of the Annual Meeting of the Population Approach Group in Europe. 2008. https://www.page-meeting.org/?abstract=1434.
8. Mentré F, Escolano S. Prediction discrepancies for the evaluation of nonlinear mixed-effects models. J Pharmacokinet Pharmacodyn. 2006;33:345–67.
9. Brendel K, Comets E, Laffont C, Laveille C, Mentré F. Metrics for external model evaluation with an application to the population pharmacokinetics of gliclazide. Pharm Res. 2006;23:2036–49.
10. Comets E, Mentré F. Developing tools to evaluate non-linear mixed effect models: 20 years on the npde adventure. AAPS J. 2021;23:75.
11. Comets E, Brendel K, Mentré F. Model evaluation in nonlinear mixed effect models, with applications to pharmacokinetics. J Soc Franç Stat. 2010;151:106–28.
12. Chacón JE, Duong T. Multivariate kernel smoothing and its applications. 1st ed. New York: Chapman and Hall/CRC; 2018.
13. Chafai D, Concordet D. Confidence regions for the multinomial parameter with small sample size. J Am Stat Assoc. 2009;104:1071–9.
14. El Hassani M, Marsot A. External evaluation of population pharmacokinetic models for precision dosing: current state and knowledge gaps. Clin Pharmacokinet. 2023;62:533–40.
15. Keizer RJ, Ter Heine R, Frymoyer A, Lesko LJ, Mangat R, Goswami S. Model-informed precision dosing at the bedside: scientific challenges and opportunities. CPT Pharmacometr Syst Pharmacol. 2018;7:785–7.

Authors and Affiliations

Sarah Baklouti[1,2] · Emmanuelle Comets[3,4] · Peggy Gandia[1,2] · Didier Concordet[2]

✉ Didier Concordet
   d.concordet@envt.fr

[1] Laboratoire de Pharmacocinétique et Toxicologie, Institut Fédératif de Biologie, CHU de Toulouse, Toulouse, France

[2] INTHERES, Université de Toulouse, INRAE, ENVT, Toulouse, France

[3] Université Paris Cité and Université Sorbonne Paris Nord, Inserm, IAME, 75018 Paris, France

[4] Université de Rennes, Inserm, EHESP, Irset-UMR_S 1085, 35000 Rennes, France