Pharm Res (2021) 38:625–635
https://doi.org/10.1007/s11095-021-03000-4

RESEARCH PAPER

# A Signal Processing Approach to Pharmacokinetic Data Analysis

C. O. S. Sorzano[1,2] • M. A. Pérez-de-la-Cruz Moreno[3] • F. R. Martín[2] • C. Montejo[2] • A. Aguilar-Ros[2]

Received: 25 July 2020 / Accepted: 4 November 2020 / Published online: 22 March 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

**ABSTRACT** The connection between pharmacokinetic models and system theory has been established for a long time. In this approach, the drug concentration is seen as the output of a system whose input is the drug administered at different times. In this article we further explore this connection. We show that system theory can be used to easily accommodate any therapeutic regime, no matter its complexity, allowing the identification of the pharmacokinetic parameters by means of a non-linear regression analysis. We illustrate how to exploit the properties of linear systems to identify non-linearities in the pharmacokinetic data. We also explore the use of bootstrapping as a way to compare populations of pharmacokinetic parameters and how to handle the common situation of using multiple hypothesis tests as a way to distinguish two different populations. Finally, we demonstrate how the bootstrap values can be used to estimate the distribution of derived parameters, as can be the allometric scale factors.

At the time of the conception of the paper Kinestat Pharma

Keypoints
- We provide a discrete system approach to the analysis of pharmacokinetic data that considers any therapeutic regimen.
- The main advantage of this technique is that it does not require the analytical solution of a differential equation. In this way, it can be used in situations in which this explicit solution does not exist.
- This approach allows identifying non-linearities in pharmacokinetic data.
- This approach can easily be integrated with bootstrapping so that population-wide comparisons can be easily done. In this article we illustrate the use of the method to estimate the statistical distribution of derived parameters (like the parameters of an allometric scaling, which are based on the estimation of the primary pharmacokinetic parameters).

✉ C. O. S. Sorzano
coss@cnb.csic.es

[1] National Center of Biotechnology, CSIC, Madrid, Spain

[2] University CEU San Pablo, Madrid, Spain

[3] Actual: Chemo Group, Madrid, Spain

**KEY WORDS** bootstrap · linear systems · pharmacokinetics · signal processing

## INTRODUCTION

The time evolution of a drug concentration in a two-compartment model after a single bolus injection responds to the system of differential equations

$$\begin{aligned} V_c\frac{dC(t)}{dt} &= -ClC(t)+Q\left(C_p(t)-C(t)\right) \\ V_p\frac{dC_p(t)}{dt} &= Q\left(C(t)-C_p(t)\right) \end{aligned} \tag{1}$$

with initial values $C(0)=\frac{D}{V_c}$ and $C_p(0)=0$. The time response of the concentration in the central compartment is known to be (1)

$$C(t)=D\left(Ae^{-\alpha t}+Be^{-\beta t}\right) \tag{2}$$

where $D$ is the dose

$$\begin{aligned} \kappa &= Q\left(\frac{1}{V_c}+\frac{1}{V_p}\right)+\frac{Cl}{V_c} \\ \beta &= \frac{1}{2}\left(\kappa-\sqrt{\kappa^2-4\frac{Q}{V_p}\frac{Cl}{V_c}}\right) \\ \alpha &= \frac{1}{\beta}\frac{Q}{V_p}\frac{Cl}{V_c} \\ A &= \frac{1}{V_c}\frac{\alpha-Q/V_p}{\alpha-\beta} \\ B &= \frac{1}{V_c}\frac{\beta-Q/V_p}{\beta-\alpha} \end{aligned} \tag{3}$$

Note that this response is only valid for a single, intravenous bolus given at $t=0$.



Content courtesy of Springer Nature, terms of use apply. Rights reserved.

The relationship between pharmacokinetic models and system theory was established (2–6) more than 40 years ago. System theory states that the concentration in the different compartments is the response of time-invariant systems whose inputs are the drug doses and the output is the concentration:

$$\frac{d\mathbf{C}(t)}{dt} = \mathbf{F}(\mathbf{C}(t)) + \mathbf{G}(t, \mathbf{D}(t)) \quad (4)$$

where $\mathbf{C}(t)$ is a column vector containing the concentrations at the locations of interest (the output of the system), $\mathbf{D}(t)$ is a vector of doses (the input), and **G** and **F**are functions that relate the input and output. The advantage of the system theory formulation is that it allows determining the concentrations for any therapeutic regimen. Bolus are represented by Dirac δ's, $\delta(t)$; while constant infusions are represented by Heaviside functions, $u(t)$. A bolus at time $t_0$ with a dose $D_0$ is represented by $D_0\delta(t - t_0)$, a constant infusion at rate $R_0$ between $t_1$ and $t_2$ is represented by $R_0(u(t - t_1) - u(t - t_2))$. Any therapeutic regimen can be expressed as the sum of these two types of terms. Forty years ago, the concentration used to be calculated by solving the differential equation system by analytical means such as the Laplace transform. With the advent of modern computers, the differential equation can be discretized into a difference equation and a numerical algorithm is used to calculate the time response of the system (7, 8).

In this article we present some signal processing tools that exploits the properties of systems in order to solve different problems in pharmacokinetic modelling like the use of arbitrary models with arbitrary therapeutic regimens, detection of non-linearities, comparison between population parameters, ... In order to better present the theory behind this data analysis, in some cases we immediately illustrate the ideas presented in the article with some experimental data. The data corresponds to the Justesa Imagen SAU's dimeric, nonionic, low-osmolal and low viscosity X-ray contrast agent ICJ 3393 (see Fig. 1), and currently in phase II clinical trials as a contrast agent for radiography.

## MATERIALS AND METHODS

### Animals

Beagle dogs (Biocentre, Barcelona, Spain) weighing 9.4–10.7 kg and Sprague-Dawley rats (CRIFFA, Barcelona, Spain) weighing 170–245 g were used. Animals were kept in a controlled environment (17°C to 21°C, dark/light 12/12 h). All animal procedures were conducted according to the Directive 2010/63/EU of the European Parliament and of the Council on the protection of animals used for scientific purposes, 22 September 2010. Efforts were made to minimize the number of animals used and their suffering.

### Contrast Media

The X-ray contrast agent ICJ 3393 (N,N′-Bis[3-carbamoyl-5-(2,3-dihydroxypropylcarbamoyl)-2,4,6-triiodophenyl]-N,N′-bis(2,3-dihydroxypropyl)-malonamide; INN: Iosimenol; CAS 181872–90-2) was labelled by interchange with carrier free $^{125}$INa (Amersham, Buckinghamshire, England) in presence of $Cu^{++}$ ions. The reaction mixture was filtered through an anion exchange column to remove free iodine and a cation exchange column to wash out the $Cu^{++}$ ions. The chemical and radiochemical purity of the $^{125}$I-labelled ICJ 3393 obtained were tested by HPLC and results were greater than 96%. [$^{125}$I]-ICJ 3393 was formulated with sufficient carrier to make solutions of ICJ 3393 at 30 to 320 mg I/ml.

### Pharmacokinetic Studies

#### *Studies in Rats*

The pharmacokinetic profile of ICJ 3393 after intravenous administration in rats was studied at three dose levels: 60 mg I/kg (116.5 mg ICJ 3393/kg; 0.94 MBq/ml), 300 mg I/kg (582.5 mg ICJ 3393/kg; 2.78 MBq/ml) and 1 g I/kg (1.94 g/kg ICJ 3393; 0.75 MBq/kg). For each dose level studied, the following procedures were performed. The day prior to drug administration, a catheter (Venocath-18; ABBOTT, Ireland) was implanted into the left jugular vein of a group of five rats (9) for blood sampling. After [$^{125}$I]-ICJ 3393 administration into the right tail lateral vein, blood samples of 300 µl were collected in plastic tubes. Following each blood sample, the implanted catheter was flushed with 0.2 ml of heparinised physiological saline (100 IU/ml). Plasma samples were obtained by blood centrifugation at 2500×g for 10 min. Radioactivity levels of plasma samples were measured by solid scintillation counting using an Auto-Gamma 500 automatic solid scintillation analyzer (PACKARD, Meriden, CT, USA), and the corresponding counts per minute (cpm) values were recorded. Plasma cpm values were transformed into equivalent concentrations of ICJ 3393 using the specific activity of the administered solution.

#### *Studies in Dogs*

Ten minutes prior to drug administration, a catheter (Abbocath T 20G; ABBOTT, Ireland) was implanted into the left saphenous vein of the dogs for blood sampling. After [$^{125}$I]-ICJ 3393 administration (0.6 g I/kg; 1.16 g ICJ 3393/kg; 0.74 MBq/kg) into the left antecubital vein, blood samples of 3 ml were collected in heparinised plastic tubes. Plasma samples (1 ml) were obtained after blood centrifugation at 2500×g for 10 min. Radioactivity levels in plasma samples were determined as previously stated.



Content courtesy of Springer Nature, terms of use apply. Rights reserved.

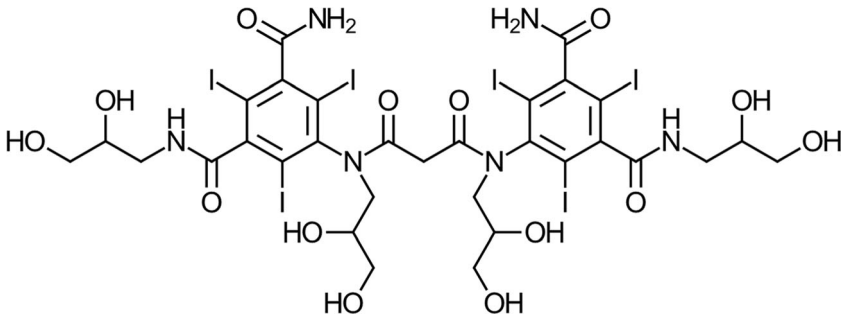

**Fig. 1** Molecular structure of ICJ 3393.

## Simulation of Pharmacokinetic Profiles for Arbitrary Systems and Arbitrary Therapeutic Regimes

Numerical analysis provides procedural tools to simulate equation systems like the one above (see Eq. (4)). The procedure normally goes through the discretization of the differential equation system. This requires transforming the continuous signals used in the differential equation into discrete signals evaluated at regular time intervals. For instance.

$$\mathbf{C}[n] = \mathbf{C}(nT_s) \tag{5}$$

$n$ is an index to refer to the sample number, $T_s$ is the sampling rate (note that this sampling is computational, and does not require sampling the drug concentration from the patient or animal). We refer to discrete signals with square brackets and to continuous signals with parentheses. Given a generic differential equation system of the form

$$\frac{d\mathbf{C}(t)}{dt} = \mathbf{H}(t, \mathbf{C}(t)) \tag{6}$$

then, the first-order, causal Euler method would iterate as

$$\mathbf{C}[n] = \mathbf{C}[n-1] + T_s\mathbf{H}((n-1)T_s, \mathbf{C}[n-1]) \tag{7}$$

In the family of systems we are interested in, **H** can be written as

$$\mathbf{H}(t, \mathbf{C}(t)) = \mathbf{F}(\mathbf{C}(t)) + \mathrm{G}(\mathrm{D}(t)) \tag{8}$$

Consequently, the first-order, causal Euler iteration would be

$$\mathbf{C}[n] = \mathbf{C}[n-1] + T_s\mathbf{F}(\mathbf{C}[n-1]) + T_s\mathbf{G}(\mathbf{D}[n-1]) \tag{9}$$

The two-compartment model belongs to a class of systems called linear systems with constant coefficients. These systems are characterized by

$$\begin{array}{rcl} \mathbf{F}(\mathbf{C}(t)) & = & A\mathbf{C}(t) \\ \mathbf{G}(\mathbf{D}(t)) & = & B\mathbf{D}(t) \end{array} \tag{10}$$

where $A$ and $B$ are square matrices. For instance, we may write the two-compartment model as

$$\begin{pmatrix} \frac{dC(t)}{dt} \\ \frac{dC_p(t)}{dt} \end{pmatrix} = \begin{pmatrix} -\frac{Cl+Q}{V_c} & \frac{Q}{V_c} \\ \frac{Q}{V_p} & -\frac{Q}{V_p} \end{pmatrix} \begin{pmatrix} C(t) \\ C_p(t) \end{pmatrix} + \begin{pmatrix} \frac{1}{V_c} & 0 \\ 0 & \frac{1}{V_p} \end{pmatrix} \begin{pmatrix} D\delta(t) \\ 0 \end{pmatrix} \tag{11}$$

It can be easily seen that $A = \begin{pmatrix} -\frac{Cl+Q}{V_c} & \frac{Q}{V_c}\frac{Q}{V_p} - \frac{Q}{V_p} \end{pmatrix}$ and $B = \begin{pmatrix} \frac{1}{V_c} & 0\ 0\frac{1}{V_p} \end{pmatrix}$.

The first-order, causal Euler iteration for linear systems would be

$$\mathbf{C}[n] = (I + AT_s)\mathbf{C}[n-1] + T_sB\mathbf{D}[n-1] \tag{12}$$

being $I$ the identity matrix and having a rest initial condition, $\mathbf{C}[n] = 0$ if $n < 0$. The first term indicates how the concentration evolves over time in the absence of external doses, the second term is an attempt to estimate the amount of drug injected in the system between one sample and the next. Actually, we can be more accurate and substitute the second term by an exact account

$$\mathbf{C}[n] = (I + AT_s)\mathbf{C}[n-1] + T_sB\int_{(n-1)T_s}^{nT_s} \mathbf{D}(\tau)d\tau \tag{13}$$

This discretization is known in numerical analysis as an explicit scheme. Although very simple, it may result in unstable systems depending on $T_s$ and $A$. Alternatively, we may use an implicit scheme, which is always stable, and for linear systems they are not too complicated. The implicit Euler scheme of first-order is of the form

$$\mathbf{C}[n] = \mathbf{C}[n-1] + T_s\mathbf{H}(nT_s, C[n]) \tag{14}$$

For a linear system, this becomes

$$\mathbf{C}[n] = \mathbf{C}[n-1] + T_sA\mathbf{C}[n] + T_sB\mathbf{D}[n] \tag{15}$$



Content courtesy of Springer Nature, terms of use apply. Rights reserved.

Solving for $\mathbf{C}[n]$

$$\mathbf{C}[n] = (I - AT_s)^{-1}\left(\mathbf{C}[n-1] + T_s B \int_{(n-1)T_s}^{nT_s} \mathbf{D}(\tau)d\tau\right) \quad (16)$$

where, as previously done, we have substituted the dose dependent term by its more accurate expression.

These approximations are of first order because the approximation error depends linearly with the sampling rate, more accurate results are obtained with finer sampling rates. We may increase the accuracy of the numerical procedure by adopting more complex numerical calculations. For instance a fourth order Runge-Kutta procedure proceeds as follows

$$\mathbf{C}[n] = \mathbf{C}[n-1] + \frac{1}{6}(h_1 + h_2 + h_3 + h_4) \quad (17)$$

with

$$\begin{aligned} h_1 &= T_s\mathbf{H}((n-1)T_s, \mathbf{C}[n-1]) \\ h_2 &= T_s\mathbf{H}\left((n-1)T_s + \frac{1}{2}T_s, \mathbf{C}[n-1] + \frac{1}{2}h_1\right) \\ h_3 &= T_s\mathbf{H}\left((n-1)T_s + \frac{1}{2}T_s, \mathbf{C}[n-1] + \frac{1}{2}h_2\right) \\ h_4 &= T_s\mathbf{H}((n-1)T_s + T_s, \mathbf{C}[n-1] + h_3) \end{aligned} \quad (18)$$

For the particular case of linear systems, we may find a simpler expression

$$\mathbf{C}[n] = \left(I + AT_s + \frac{A^2T_s^2}{2!} + \frac{A^3T_s^3}{3!} + \frac{A^4T_s^4}{4!}\right)\mathbf{C}[n-1] + T_s B \int_{(n-1)T_s}^{nT_s} \mathbf{D}(\tau)d\tau \quad (19)$$

Remarkably, the numerical methods developed in Eqs. (12), (16), and (19) can be generally applied to any linear system (even non-linear in Eq. (9)) and any dosing regimen. The two-compartment model is a linear system and, consequently, the methodologies described above can be applied. Actually, any multicompartment model with constant coefficients is also a linear system. The simplicity of reducing many different kinds of systems and therapeutic regimens (like Eq. (19)) to a single difference equation contrasts with the plethora of pharmacokinetic responses for particular cases (exemplified in Eqs. (2) and (3) and seen in books like (10)). Additionally, the numerical scheme developed above lends itself to easy computational implementation. In this way, a single formula covers a wide range of different models simply by changing the functions **F** and **G**. The validity of this approach has been demonstrated in Sec. 3.1.

## System Identification

The numerical procedure described so far has to be complemented with a way to identify the pharmacokinetic system once a therapeutic plan has been administered and the resulting drug concentrations have been measured from the patient. Suppose we have $N$ concentration measurements, $C_i$ ($i = 1,2,\ldots,N$), taken at times $t_i$. In the traditional approach we would determine the model parameters, $\Theta$ (in the two-compartments model, $\theta = \{V_c, V_p, Q, Cl\}$), that minimize a data fidelity term like

$$\min_{\theta} (C_i - C(t_i))^2 \quad (20)$$

or

$$\min_{\theta} (\log(C_i) - \log(C(t_i)))^2 \quad (21)$$

In the discrete system theoretical approach, the system identification is performed in the same way only that the predicted concentrations $C(t_i)$ are interpolated from the surrounding discrete samples.

From a theoretical point of view there is an important distinction between the two objective functions (Eqs. (20) and (21)). Both functions are assuming independence between the noise observed at each of the samples [ (7), Sec. 3.3]. However, the first objective function presumes that additive Gaussian noise while the second assumes multiplicative log-normal noise (strictly speaking, one could perform least squares, in natural or logarithmic units, without assuming normality of the residuals; however, least squares optimization corresponds to the maximum likelihood estimate if the residuals are Gaussianly distributed, we have referred to this assumption of normality in this sense). If noise were additive, it would be the same at any concentration. However, this is not the case, for instance (11) allows for bionanalysis with a maximum determination error of 15%. This implies that noise must be smaller at lower concentrations and, consequently, noise must be multiplicative.

### *Detection of Non-linearities*

Figure 2 shows the typical profile of drug concentration in plasma when the dose is fixed per weight (for instance, $D_0$ mg/kg; in our case note that the dose is expressed in mg. of iodine per kg.). Since each animal has a slightly different weight, it is given a different dose. This introduces an extra source of variation in the analysis of the data. However, for the family of linear pharmacokinetic models, and only for this family, we may computationally homogenize the different concentration profiles. This is a result of a property of linear



Content courtesy of Springer Nature, terms of use apply. Rights reserved.

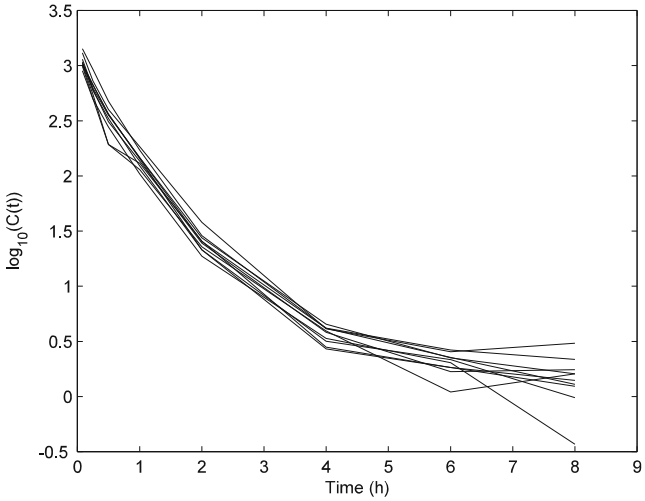


**Fig. 2** Drug concentration profiles in plasma from a set of 10 animals with a fixed dose per weight.

systems: the response of the linear system changes linearly with the input. For instance, if the response to the input dose $\mathbf{D}(t)$ is $\mathbf{C}(t)$, then the response to $a\mathbf{D}(t)$ is $a\mathbf{C}(t)$. In this way, we can safely predict what would have been the response of each one of the animals if its weight would have been different. Formally, let us assume that there are $N$ animals, each one with a weight $w_i$ ($i = 1,2,...,N$) and drug concentration response $C_i(t)$. The actual dose of each animal is

$$\mathbf{D}_i(t) = w_i D_0 \begin{pmatrix} \delta(t) \\ 0 \end{pmatrix}$$

Let $\overline{w} = \frac{1}{N}\sum_{i=1}^{N} w_i$ be the average weight of all the animals. The corresponding dose for an animal of this weight, would have been

$$\overline{\mathbf{D}}(t) = \overline{w} D_0 \begin{pmatrix} \delta(t) \\ 0 \end{pmatrix}$$

Note that

$$\overline{\mathbf{D}}(t) = \frac{\overline{w}}{w_i}\mathbf{D}_i(t) \tag{22}$$

If the response of the $i$-th animal to $\mathbf{D}_i(t)$ is $C_i(t)$, since the pharmacokinetic model is linear, then the response of the same system to $\overline{\mathbf{D}}(t)$ would be

$$\overline{\mathbf{C}}_i(t) = \frac{\overline{w}}{w_i}\mathbf{C}_i(t) \tag{23}$$

Finally, we may combine all these individual responses into a single average response of the set of animals

$$\overline{\mathbf{C}}(t) = \frac{1}{N}\sum_{i=1}^{N}\overline{\mathbf{C}}_i(t) = \frac{1}{N}\sum_{i=1}^{N}\frac{\overline{w}}{w_i}\mathbf{C}_i(t) \tag{24}$$

We may also calculate the sample standard deviation

$$\begin{aligned} s(t) &= \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(\overline{\mathbf{C}}_i(t) - \overline{\mathbf{C}}(t)\right)^2} \\ &= \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(\frac{\overline{w}}{w_i}\mathbf{C}_i(t) - \overline{\mathbf{C}}(t)\right)^2} \end{aligned} \tag{25}$$

Note that this average and standard deviation per sample is different from the raw average and standard deviation because the different weights of the animals are taken into account. To illustrate the process, Fig. 3 shows the average and standard deviation as proposed above.

This average profile may be used to fit the pharmacokinetic parameters and then use this estimate as the initialization for the fitting of the individual animals. More interestingly, we may use the ability of linear systems to scale with the dose to detect dose-dependent parameters. Figure 4 shows 3 groups of animals. Each group was given a different dose $D_0$, $D_1$, and $D_2$. We may use the scale property of linear systems to calculate which the concentration profile of each animal to a different dose would have been. To do so, let $C_i^0(t)$ be the response of an animal to dose $D_0$. If the dose were $D_1$, the animal response would be

$$\mathbf{C}_i^{0\to 1}(t) = \frac{D_1}{D_0}\mathbf{C}_i^0(t) \tag{26}$$

As we did before we can average all responses to an average weight and plot the average and standard deviation (Fig. 5). This figure shows the concentration profiles normalized by dose and weight. If the system was truly linear, then the three curves would overlap. However, we can see significant differences between the dose groups 1 or 2 h after injection. This reveals an important non-linearity in the behavior of the system. The fact that the system is not linear does not preclude a linear analysis. This is called system linearization. The trick is to assume that the system is linear in a neighborhood of the dose being analyzed (technically this is related to a Taylor first order expansion of the non-linear function connected the input and output of the system). However, we must be conscious of the limitations of the analysis and understand that the system may behave "locally" as a two-compartment pharmacokinetic model. As a consequence, the system parameters are valid only for doses similar to the one being analyzed. Away from this dose, a new system analysis must be performed.



Content courtesy of Springer Nature, terms of use apply. Rights reserved.

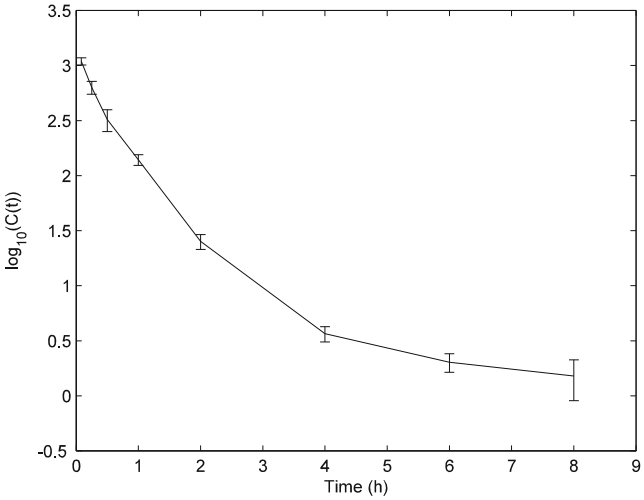


**Fig. 3** Average and standard deviation of the experimental population when responses are normalized to the mean weight.

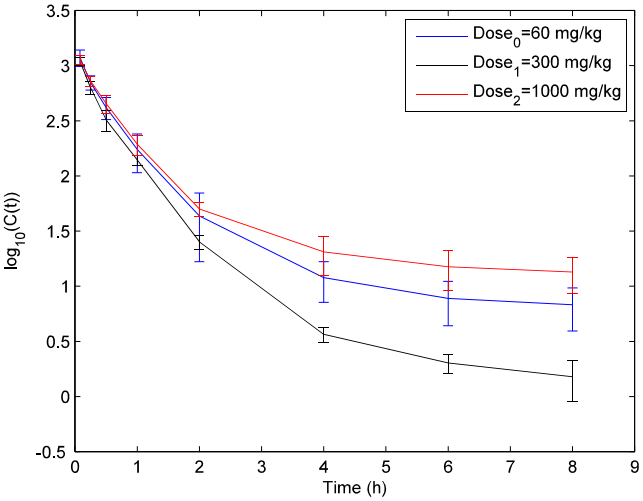


**Fig. 5** Average and standard deviation of the 3 groups after normalizing for the weight and dose. The dose of 300 mg l/kg was used as reference and the other two doses groups were normalized to be equivalent to a dose of 300 mg l/kg.

### *Comparing Two Pharmacokinetic Parameters Populations*

In many occasions we need to ascertain whether two populations of pharmacokinetic parameters are significantly different or not. For instance, we may want to determine if the population of pharmacokinetic parameters estimated for dose $D_0$ are significantly different from the pharmacokinetic parameters estimated for dose $D_1$, or if the population of parameters estimated by Method 1 is significantly different from the pharmacokinetic parameters estimated by Method 2. In general, we need to decide whether we have a single population of parameters, or two populations of parameters. We may adopt different approaches to the problem. In a parametric approach we would assume that the population of parameters has some known multivariate distribution, like the multivariate Gaussian, and use parametric tests to determine if the multivariate mean and covariance matrices of both populations are significantly different or not. Instead, we may see the number of populations as the explanatory variable and use a criterion like Bayesian Information Criterion or Akaike Information Criterion to decide if we have 1 or 2 populations. The drawback of this approach is that there is no clear threshold to determine which of the two criteria values represents reality. That is, these criteria help to choose amongst a set of models, but they cannot state whether the selected model is more or less likely.

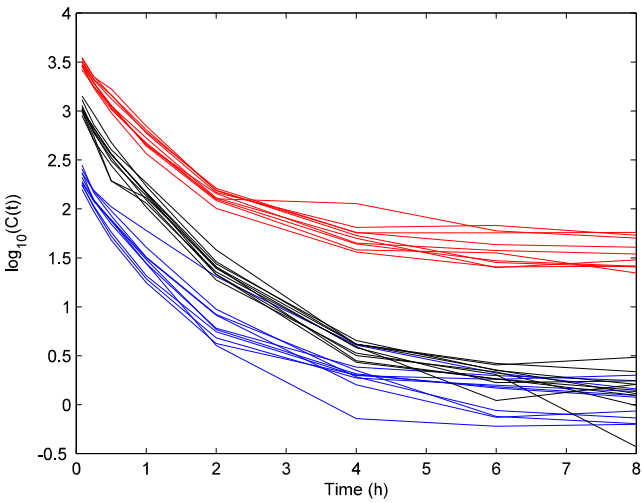


**Fig. 4** Drug concentration profiles for 3 groups of 10 animals. From bottom-up, the groups were given 60 mg l/kg, 300 mg l/kg, 1000 mg l/kg.

Alternatively, we propose in this article to perform a hypothesis test on the confusion matrix of a non-linear classifier. The advantage of this approach is that the criterion is clearly defined: the hypothesis that we have two populations is rejected if the classifier cannot be proved to perform better than random classification between the two populations (that is if the proportion of individuals from each original class in the automatically detected classes is significantly larger than 0.5). We may use a plethora of powerful non-linear classifiers coming from Machine Learning like Support Vector Machines, Neural Networks, Random Forests, etc. to discriminate between the two populations. These algorithms have proved to be useful in rather challenging conditions. We may use several kinds of classifiers to make sure that if there is a difference between the two populations, then we are capable of identifying it. For each classifier we would perform a hypothesis test, and the test would reject the hypothesis that the two populations can be discriminated or not by the classifier. The null hypothesis is that the two populations cannot be discriminated, while the alternative hypothesis is that the two populations are different.



Content courtesy of Springer Nature, terms of use apply. Rights reserved.

We may also adopt a univariate approximation to the problem of distinguishing between two populations of pharmacokinetic parameters. We can employ non-parametric hypothesis tests like Kolmogorov-Smirnov (12) to detect whether the distribution of a given parameter (like clearance) is different in one group or the other. If two multivariate populations are the same, then all the univariate tests should not detect any difference between the distribution (or any of its moments) of the individual parameters. The converse is not true: there might not be differences in the univariate distributions but there exist differences in the multivariate ones.

If we perform a total of $T$ tests, let us denote the type I and II errors of each test as $\alpha_t$ and $\beta_t$ ($t = 1,2,...,T$). Without loss of generality, let us assume that there are $R$ null hypothesis rejections (the first $R$ hypothesis tests), and $T$-$R$ non-rejections ($R$ may be 0). Let us also assume that the a priori probability that the two populations are the same is $\pi_0$ and that the probability that the two populations are not the same is $\pi_1 = 1 - \pi_0$. The probability that there are $R$ rejections if the two populations are the same is

$$\pi_0 \left( \prod_{t=1}^{R} \alpha_t \right) \left( \prod_{t=R+1}^{T} (1-\alpha_t) \right) \tag{27}$$

Similarly, the probability that there are $R$ rejections if the two populations are not the same is

$$\pi_1 \left( \prod_{t=1}^{R} (1-\beta_t) \right) \left( \prod_{t=R+1}^{T} \beta_t \right) \tag{28}$$

The Type I and Type II errors can be set by design by calculating the number of samples (13). For the sake of completeness, we give here the sample size formula (test on proportions with a single sample when the normal approximation is valid):

$$N \geq \frac{1}{4} \left( \frac{z_{1-\frac{\alpha}{2}} + z_{1-\beta}}{\Delta} \right)^2 \tag{29}$$

where $\Delta$ is the difference in proportions that we want to detect with power 1-$\beta$. The sample size formula has to be adapted to each kind of test, and the interested reader is referred to (13) for a comprehensive overview.

Assuming that a set of $R$ rejections has been observed, we may calculate the likelihood ratio between the probability of having two identical populations and the probability of having two different populations

$$L = \frac{\pi_0 \left( \prod_{t=1}^{R} \alpha_t \right) \left( \prod_{t=R+1}^{T} (1-\alpha_t) \right)}{\pi_1 \left( \prod_{t=1}^{R} (1-\beta_t) \right) \left( \prod_{t=R+1}^{T} \beta_t \right)} \tag{30}$$

If we assume that all the tests have the same confidence and statistical power, then the equation above simplifies to

$$L = \frac{\pi_0}{1-\pi_0} \left( \frac{\alpha}{1-\beta} \right)^R \left( \frac{1-\alpha}{\beta} \right)^{T-R} \tag{31}$$

If $L > 1$, then, with the number of rejections observed, it is more likely that the two populations are identical. Otherwise, it is more likely that the two populations are different. See Secs. 3.1 and 3.2 for examples of the application of this methodology.

### *Using Population Parameters to Determine Derived Parameters*

We may also use the bootstrap samples to estimate the empirical distribution of derived parameters like allometric scale parameters. Given a parameter of interest $p$ measured in two species (e.g. clearance), $A$ and $B$, we may estimate the allometric scale parameter as

$$p_B = p_A \left( \frac{BW_B}{BW_A} \right)^b \Rightarrow b = \exp \left( \frac{\log \frac{p_B}{p_A}}{\log \frac{BW_B}{BW_A}} \right) \tag{32}$$

To estimate the empirical distribution we would randomly select two animals (one of species $A$ and another of species $B$), their weights are experimentally determined as $BW_A$ and $BW_B$. The parameter of interest is randomly chosen amongst the bootstrap values estimated for the $A$ and $B$ animals. The equation above gives an estimate of the allometric scale for this specific choice of random parameters. Repeating the procedure above $N$ times, we may empirically estimate the distribution of the allometric scale factor. Note that in this way, we produce a range of values for the allometric scale factor (along with its probability density function) and not only a single value as is customary in a more traditional approach. See Sec. 3.3 for the application of this methodology.

## RESULTS

In this section we illustrate the use of the different techniques introduced above. We start by validating the use of a discrete system approach to the determination of the drug concentration profile. Then, we show how the comparison between populations can be used to ascertain differences between different therapeutic or physiological conditions. Finally, we show the use of bootstrapping for the determination of allometric scaling parameters.



Content courtesy of Springer Nature, terms of use apply. Rights reserved.

## Validation of the Discrete System Approach

We estimated the pharmacokinetic parameters for 10 rats at a dose of 300 mg I/kg (see the measured drug concentrations in Fig. 2). We estimated the pharmacokinetic parameters using the standard closed-form formulas shown in Eqs. (1), (2) and (3) (we refer to this approach as method I), and using the discretization of the differential equations introduced in Sec. 2.4 (method II). For each animal, we calculated 100 bootstrap samples, making a total of 1000 parameter estimates. The following table shows the mean and standard deviation of the four parameters under the two calculation methods:

| | *Cl [mL/min]* | $V_c$ *[mL]* | *Q [mL/min]* | $V_p$ *[mL]* |
|---|---|---|---|---|
| Method I | 1.9 ± 0.22 | 70.4 ± 11.1 | 0.29 ± 0.04 | 599.5 ± 106.0 |
| Method II | 1.8 ± 0.23 | 71.0 ± 11.3 | 0.29 ± 0.04 | 601.9 ± 105.6 |

Figure 6 shows the fitting for one of the animals using the two methods. As can be seen both result into identical estimates of the pharmacokinetic response of the animal.

We performed a Kolmogorov-Smirnov test to determine whether the distribution of any of the 4 pharmacokinetic parameters was different between the two methods. None of the 4 hypotheses was rejected (with a confidence level of 95%). Note that this test is rather general because it tests at the same time all the distribution moments (mean, variance, skewness, kurtosis, …) and percentiles.

In order to visually verify that the two populations of parameters are really the same, we would like to plot the joint probability density functions of both populations. Unfortunately, the pharmacokinetic parameters belong to a 4-dimensional space (since we have 4 parameters: $Cl$, $V_c$, $Q$ and $V_p$). For this reason we cannot visualize the overlap between the two sets of parameters. Instead, we can reduce the dimensionality of the data. With this purpose, we performed a Principal Component Analysis of the two sets at the same time (so that if there is a difference between the two, this difference can be shown on the same low-dimensional space). Figure 7 shows the first two principal components of the two sets of parameters (they explain 99.99% of the variance). We can see that the two populations perfectly overlap.

Finally, we performed 10 random subsets of 500 samples from the two populations, trained 10 Support Vector Machines and classified with each one of them the remaining 500 samples. None of the classifiers was capable of performing a classification significantly different from a random classification, implying the non-separability of the two sets of parameters.

In total we have performed 14 hypothesis tests. The statistical power with $N = 1000$ samples for testing if the proportion of misclassified individuals is 0.5 is 0.9355. None of the null hypotheses were rejected, consequently the ratio between the probability of coming from a single population and the probability of coming from different populations is approximately (see Eq. (33))

$$L = \left(\frac{0.05}{0.9355}\right)^{0}\left(\frac{0.95}{0.0645}\right)^{14} = 2.2\cdot 10^{16} \tag{33}$$

which is much larger than 1. We can, therefore, conclude that the parameters estimated by fitting a non-linear function (the classical approach) and the concentration profile generated by a discrete system are statistically indistinguishable.

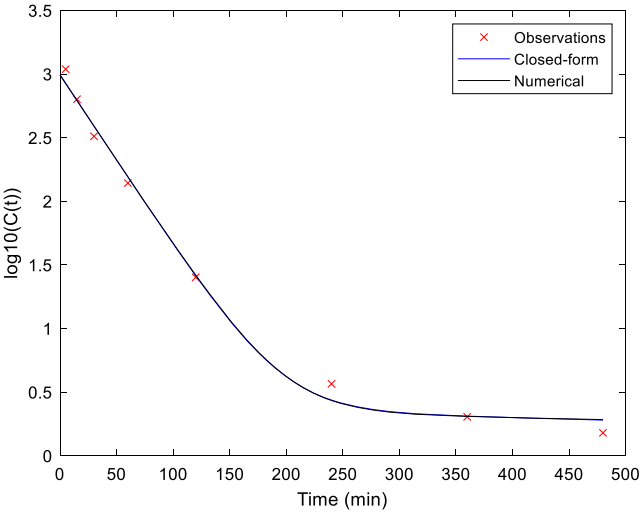


**Fig. 6** Example of the fitting of a two-compartment model using the closed form formula and the numerical approximation of the differential equation.

## Comparison between Populations

We now show how the methodology developed in this paper can actually discriminate between two different populations.

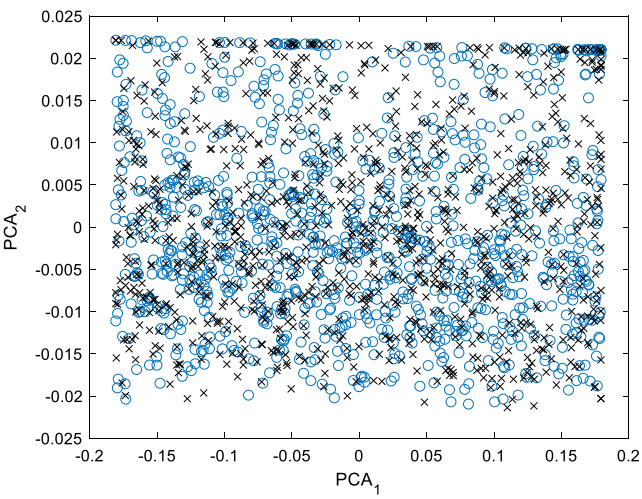


**Fig. 7** First two principal components of the pharmacokinetic parameters estimated with the classical approach (blue circles) and the discrete system approach (black crosses).



Content courtesy of Springer Nature, terms of use apply. Rights reserved.

We do so by comparing the bootstrap estimates of the pharmacokinetic parameters for a dose of 300 mg I/kg and a dose of 60 mg I/kg (see Fig. 4) and whose parameters are summarized in the following table.

| | *Cl [mL/min]* | $V_c$ *[mL]* | *Q [mL/min]* | $V_p$ *[mL]* |
|---|---|---|---|---|
| 300 mg I/kg | 1.8 ± 0.23 | 71.0 ± 11.3 | 0.29 ± 0.04 | 601.9 ± 105.6 |
| 60 mg I/kg | 1.4 ± 0.22 | 50.3 ± 11.1 | 0.25 ± 0.04 | 74.0 ± 106.0 |

We performed 14 hypothesis tests as we did in our previous examples. Figure 8 shows the projection of the parameters onto the PCA space.

In this case, all the null hypotheses were rejected. The ratio of probability of coming from a single population or from two populations is

$$L = \left(\frac{0.05}{0.9355}\right)^{14}\left(\frac{0.95}{0.0645}\right)^{0} = 1.6\cdot 10^{-18} \tag{34}$$

Consequently, we can safely conclude that there are two populations of pharmacokinetic parameters.

We performed the same kind of analysis distinguishing between female and male animals within the same dose (300 mg I/kg). The following table summarizes the parameters.

| | *Cl [mL/min]* | $V_c$ *[mL]* | *Q [mL/min]* | $V_p$ *[mL]* |
|---|---|---|---|---|
| Male | 1.9 ± 0.22 | 73.0 ± 10.2 | 0.30 ± 0.04 | 602.9 ± 108.2 |
| Female | 1.8 ± 0.23 | 69.2 ± 11.4 | 0.29 ± 0.05 | 615.0 ± 109.8 |

Three out of the four univariate tests were rejected with $p$ values well below $10^{-3}$ (the tests comparing the distributions of $Cl$, $V_c$, and $Q$), while the test corresponding to $V_p$ was not rejected. For illustration purposes we show in Fig. 9 the histogram of the $Q$ values for the female (top) and male (bottom) animals. Although the set of parameters occupy the same range of values, the distribution is not the same. The null hypothesis that these two distributions are the same was rejected by the Kolmogorov-Smirnov test with a p value of $4.8\cdot 10^{-4}$ showing the power of the methodology illustrated in this article.

None of the classifiers was capable of separating the two populations. The likelihood ratio between the probability of coming from a single population and coming from two populations is

$$L = \left(\frac{0.05}{0.9355}\right)^{3}\left(\frac{0.95}{0.0645}\right)^{11} = 1.1\cdot 10^{9} \tag{35}$$

So that it is much more likely that female and male animals do not differ in their pharmacokinetic parameters than they do. For illustration purposes, Fig. 10 shows the PCA mapping of the two populations of parameters. We can see that they are rather intertwined and probably the univariate rejections are more due to the lack of data (we only had 5 animals of each sex) than to a real difference between populations.

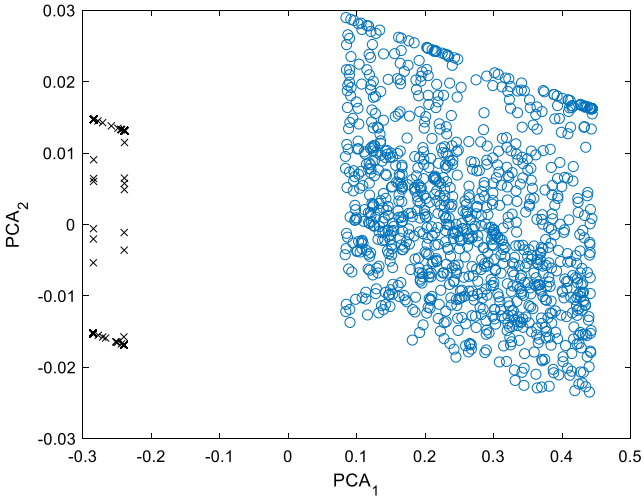


**Fig. 8** First two principal components of the pharmacokinetic parameters estimated for doses of 300 mg I/kg (blue circles) and 60 mg I/kg (black crosses).

## Distribution of Derived Parameters

As shown above, we may use the bootstrap estimates of different species to calculate the statistical distribution of derived parameters. This is a clear advantage over traditional methods where a single point estimate of the parameter is produced (Bayesian (14) or stochastic (15) methods are also capable of estimating a distribution of the model parameters). As an illustrative example the distribution of an allometric scaling factor (Eq. (32)) is presented. For the compound studied in this article, we had measurements for 10 rats at a dose of 300 mg I/kg and 3 dogs at a dose of 600 mg I/kg. We could use the methodology described in this article based on bootstrapping to calculate the parameter $b$ between any set of parameters. We did so for the scaling of the clearance $Cl$. Figure 11 shows the estimated distribution and the following table the parameters involved.

| | *Weight [g]* | *Cl [mL/min]* |
|---|---|---|
| Rats | 228 ± 16 | 1.8 ± 0.2 |
| Dogs | 9700 ± 540 | 30.1 ± 5.2 |

Note that the so calculated allometric scaling is dependent on the specific doses employed and cannot be generalized to any other dose. More experiments with different doses could have been performed, and instead of having a single scaling factor we could have calculated a surface of factors dependent on two variables (the two doses employed in the different animals).



Content courtesy of Springer Nature, terms of use apply. Rights reserved.

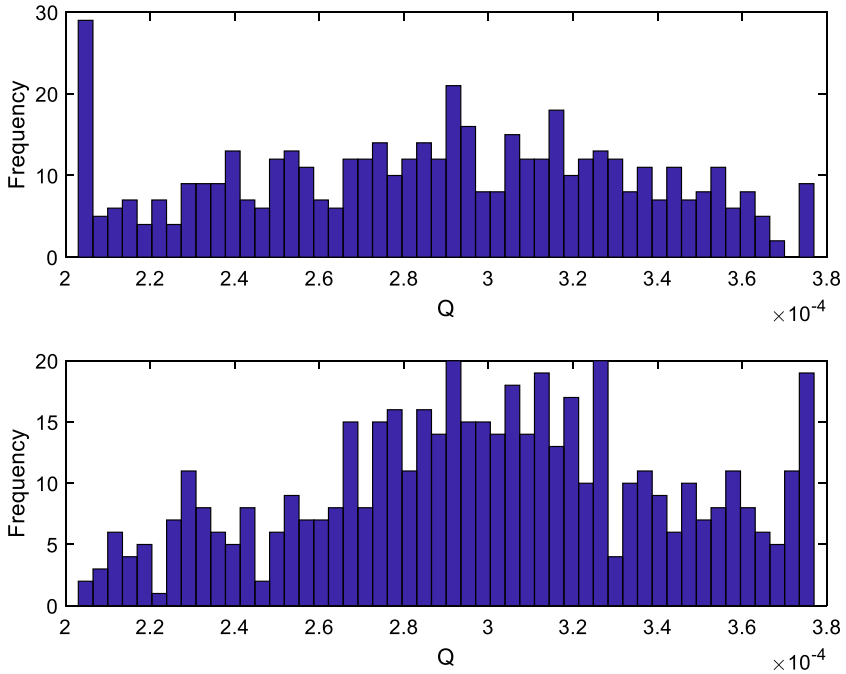


**Fig. 9** Histogram of the $Q$ values estimated by bootstrapping for the female (top) and male (bottom) animals at a dose of 300 mg l/kg.

## CONCLUSIONS

In this article we have presented a new methodology to analyze population pharmacokinetic data when there is an important lack of experimental data. The conclusions drawn from the methodology are limited by the quality and representativeness of the collected data. However, in many experimental situations it can be used as a way to propose future lines of research with the data at hand (the number of animals used in these experiments are typical in many laboratory experiments). We have also analyzed the connection between a discrete system approach and the classical pharmacokinetic approach to the identification of system parameters. We have reasoned that the discrete system approach is able of analyzing data with an arbitrary therapeutic regimen, and shown that, as expected, in the particular case of a single bolus, it produces the same results as the classical approach. For more complicated regimens, there is no closed formula in the classical

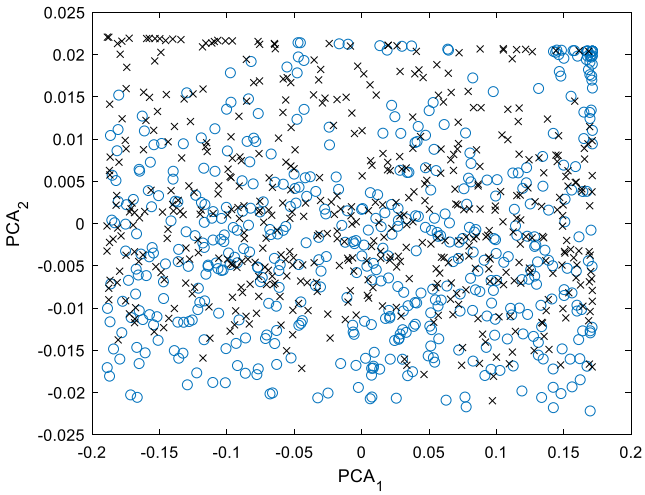


**Fig. 10** First two principal components of the pharmacokinetic parameters estimated for female (blue circles) and male (black crosses) animals at a dose of 300 mg l/kg.

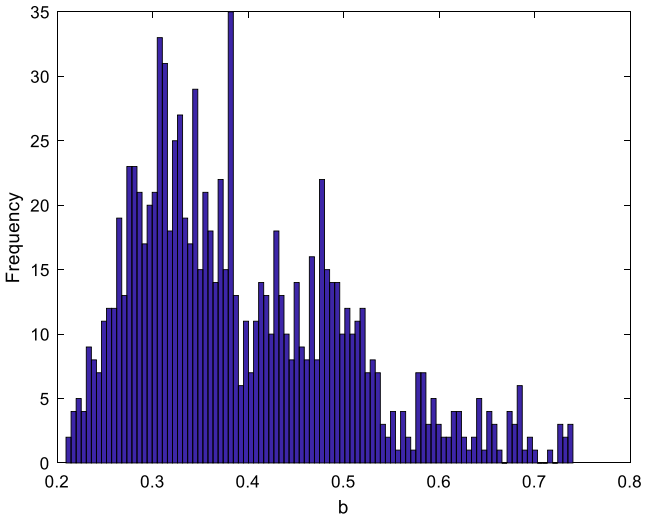


**Fig. 11** Distribution of the allometric scaling parameter for the clearance between rats at a dose of 300 mg l/kg and dogs at 600 mg l/kg.



Content courtesy of Springer Nature, terms of use apply. Rights reserved.

approach, although they are easily handled by the discrete system approach. This represents a clear advantage of the system approach over the classical approach to the calculation of the concentration profile.

## ACKNOWLEDGMENTS

The authors wish to thank Justesa Imagen SAU to grant us the use of ICJ3393 experimental data. C.O.S. Sorzano is recipient of a Ramón y Cajal fellowship from the Spanish Ministry of Economy and Competitiveness.

## REFERENCES

1. Bertrand J, Mentré F. Mathematical expressions of the pharmacokinetic and pharmacodynamic models implemented in the Monolix software. Tech Rep, INSERM U738, Paris Diderot University, 2008.
2. Cutler DJ. Linear systems analysis in pharmacokinetics. J Pharmacokinetics, Biopharmaceut. 1978;6:265–82.
3. Pedersen PV. Model-independent method of analyzing input in linear pharmacokinetic systems having polyexponential impulse response i: theoretical analysis. J Pharm Sci. 1980;69:298–305.
4. Pohjanpalo H. System identifiability based on the power series expansion of the solution. Math Biosci. 1978;41:21–33.
5. Weiss M, Föerster W. Pharmacokinetic model based on circulatory transport. Eur J Clin Pharmacol. 1979;16:287–93.
6. Schwilden H. A general method for calculating the dosage scheme in linear pharmacokinetics. Eur J Clin Pharmacol. 1981;20(5):379–86.
7. Sorzano COS, Pérez-De-La-Cruz Moreno MA, Burguet-Castell J, Montejo C, Aguilar Ros A. Cost-constrained optimal sampling for system identification in pharmacokinetics applications with population priors and nuisance parameters. J Pharm Sci. 2015;104(6):2103–9.
8. Sorzano C.O.S. A signal processing approach to pharmacokinetics, pharmacodynamics and biopharmaceutics data analysis, Ph.D. thesis, Univ. San Pablo CEU, 2015.
9. Bakar SK, Niazi S. Simple reliable method for chronic cannulation of the jugular vein for pharmacokinetic studies. J Pharm Sci. 1983;72:1027–9.
10. Gabrielsson J, Weiner D. Pharmacokinetic and pharmacodynamics data analysis: concepts and applications. 4th ed. Swedish pharmaceutical press, 2006.
11. Committee for Medicinal Products for Human Use (CHMP). Guideline on bioanalytical method validation, European Medicines Agency, 2011.
12. Sheskin DJ. Handbook of parametric and nonparametric statistical procedures, Chapman & Hall/CRC, 2004.
13. Mathews P. Sample size calculations, Mathews Malnar and Bailey, Inc. 2010.
14. Wakefield J, Aarons L, Racine-Poon A. The Bayesian approach to population pharmacokinetic/pharmacodynamic modeling. Case studies in Bayesian Statistics, pp205–265, 1999.
15. Matis JH. An introduction to stochastic compartmental models in pharmacokinetics. Pharmacokinetics. NATO ASI Series. 1988;145:113–28.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Content courtesy of Springer Nature, terms of use apply. Rights reserved.

# Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).
Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.
These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.
We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.
While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.
These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.
To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.
Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.
If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com