Bayesian Population Modeling of Phase I Dose Escalation Studies: Gaussian Process Versus Parametric Approaches

Alberto Russu*, Italo Poggesi, Roberto Gomeni, and Giuseppe De Nicolao, *Senior Member, IEEE*

***Abstract*—The early stages of the drug development process are often characterized by a limited number of subjects participating the study and a limited number of measurements per individual that can be collected, mainly due to technical, ethical, and cost reasons. The so-called *dose escalation* studies, performed during phase I, usually involve about 40 subjects or less, and feature observations at no more than three (rarely four or five) dose levels-per-subject. Depending on the complexity of the underlying pharmacokinetics, simple linear models or nonlinear ones (e.g., power, $E_{\max}$ models) may be appropriate to describe the relationship between the metrics of systemic exposure to the drug ($C_{\max}$, AUC) and the administered dose. However, in such data-poor scenarios, formulating models based on parametric descriptions is generally hard, and may easily result in model misspecification. Hence, nonparametric or "model-free" solutions, borrowed from the machine learning field, are deemed appealing. We resort to Gaussian process theory to work out Bayesian posterior expectations of a population (a.k.a mixed-effects) regression problem, namely Population Smoothing Splines (PSS). We show that in seven experimental dose escalation studies, Population Smoothing Splines improve on three widely used parametric population methods. Superiority of the model-free technique is confirmed by a simulated benchmark: Population Smoothing Splines compare very favorably even with the true parametric model structure underlying the simulated data.**

***Index Terms*—Bayesian population model, dose escalation, Gaussian process, mixed effects model, phase I trials.**

I. Introduction

ONE of the most interesting identification problems arising in biomedical data analysis is the characterization of a population of subjects. Classical examples are found in pharmacokinetics (PK) and pharmacodynamics (PD), where multiple subjects are sampled in order to obtain both the average and individual response to the administered drug. If a sufficiently large number of samples are collected in each individual, it is possible to identify a distinct model for each subject. The typical response of the population could then be obtained from the distribution of the individual models. However, the specific nature of biomedical experiments often poses technological cost or ethical constraints that permit to collect only few data in each subject. When the separate identification of individual models is not viable, an effective solution is provided by the so-called *population* modeling approach [1]–[3].

Population methods analyze all the data jointly, yielding an average model and individual ones, as well as an estimate of the inter-individual variability [1]–[9].

In the drug development process, the use of population approaches has been recommended by the Food and Drug Administration, in order to obtain a reliable assessment of intra- and inter-individual variabilities [10]. However, the use of such models is not restricted to pharmacology but is being extended to data analysis problems arising in several contexts ranging from medical imaging [11] and diagnosis of metabolic disorders [12] to genomics [13].

At the early stages of drug development or when the mechanistic understanding of the biological and physiological process involved in the drug action are not available, reliable parametric models [14]–[18] may be difficult to formulate. Hence the need for flexible nonparametric population approaches that reduce the structural assumptions to a minimum [19]. Along this direction, an example is provided by the so-called semiparametric methods that model the response curves as regression splines [20], [21]. A potential difficulty underlying the use of these techniques is the optimization of the number and location of the knots of regression splines, which could suffer from the presence of local minima. More recently, in order to develop a fully nonparametric approach, within a Bayesian paradigm it has been proposed to model the individual curves as realizations of discrete- or continuous-time stochastic processes, e.g., random walks or integrated Wiener processes [22], [23]. In these works, each individual curve is seen as the sum of an average curve (common to all subjects) and an individual shift (varying from subject to subject). In particular, both the average curve and the individual shifts are assumed to be Gaussian processes whose statistics are specified by few hyperparameters. For instance, if the curve is an integrated Wiener process, the hyperparameter is the corresponding intensity. Hyperparameter tuning can be carried out via likelihood maximization. For a given choice of the hyperparameters, the posterior expectations of the processes

Manuscript received December 13, 2010; revised May 6, 2011 and May 27, 2011; accepted July 23, 2011. Date of publication August 15, 2011; date of current version October 19, 2011. *Asterisk indicates corresponding author.*

*A. Russu is with the Department of Computer Engineering and Systems Science, University of Pavia, Pavia 27100, Italy (e-mail alberto.russu@unipv.it).

I. Poggesi was with the Clinical Pharmacology, Modeling & Simulation, GlaxoSmithKline, Verona, Italy at the time of manuscript preparation. He is now with Advanced Modeling & Simulation, Janssen Pharmaceutical Companies of Johnson & Johnson, Milan, Italy (e-mail: ipoggesi@its.jnj.com).

R. Gomeni was with Pharmacometrics, GlaxoSmithKline, Upper Merion, PA, USA. He is now with Alleantis LLC, Raleigh-Durham, NC, USA (e-mail: rgomeni@gmail.com).

G. De Nicolao is with the Department of Computer Engineering and Systems Science, University of Pavia, Pavia, 27100 Italy (e-mail: giuseppe.denicolao@unipv.it).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TBME.2011.2164614

given the data provide point estimates of the average and individual curves. In particular, when the prior is formulated in terms of integrated Wiener processes, the estimated curves are cubic splines [23]. Recently, a Bayesian MCMC approach able to return the full posterior of hyperparameters and unknown functions has been also worked out [24].

Dose-escalation procedures are often conducted in the phase I of clinical development of experimental drugs, especially when the compound is given to human subjects for the first time. In this context, human healthy subjects are typically given the experimental drug in different occasions, at increasing dose levels, to assess safety, tolerability, and pharmacokinetics. In addition to this, markers of pharmacological activity can be evaluated for helping the identification of the clinically relevant doses and/or systemic exposure [25], [26].

In case of first-time-in-human studies, the escalation starts at such a dose that no pharmacological response is expected. Doses are then gradually increased, carefully monitoring safety and tolerability, up to a stopping limit. For experimental drugs, metrics of systemic exposure, such as peak plasma concentration of the compound ($C_{\max}$) or area under the plasma concentration-time profile AUC), are often used as surrogate of safety stopping limits. During the assessment, pharmacokinetic data are collected in order to propose the next dose level that can be administered without exceeding the predefined pharmacokinetic limits in the same subjects or in a different cohort of subjects [25], [27].

In recent years, there has been a growing interest in Bayesian methods for the sequential estimation of either the toxicity probability [25], [28]–[33], or the relationship between dose and metrics of drug exposure, both in the parametric [34] and nonparametric case [35], [36] (the latter two not resorting to a population approach). In the field of cancer dose finding trials, a notable example is provided by the work of O'Quigley and coworkers [28], which resort to parametric models (e.g., logistic) of the toxicity probability as a function of the dose. Although a particular parametric model of the dose-toxicity relationship may be misspecified, their approach allows efficient estimation of so-called maximum tolerated dose, which is a major goal in dose finding studies.

In real dose escalation scenarios, limited number of subjects ($\sim$ 10 to 40) and samples-per-subject ($\sim$ 1 to 5) is available. Selection of the most appropriate dose-exposure model is therefore a crucial issue. The purpose of this paper is to provide a novel methodology that overcomes possible model misspecification issues during a dose-escalating process. To achieve this objective, we resort to a flexible, model-free approach based on Gaussian process theory. Our nonparametric technique relies on a mixed-effects (a.k.a. population) approach, with the aim to characterize the dose-exposure relationship of the subjects enrolled in the trial as well as a generic subject who was never observed. Three parametric Bayesian population models were used as a benchmark for the model-free method. Model comparison based on complexity criteria and cross validation was applied to identify the most appropriate model. We compare performances on both simulated and experimental datasets obtained for GlaxoSmithKline investigational compounds entering the clinical phase.

TABLE 1
DESIGN CHARACTERISTICS OF THE SEVEN DOSE ESCALATION STUDIES

Study	# Subjects	# Obs.	# DLs/subject*	Dose range
A	25	100	5	0.05–1.6
B	40	69	3	0.25–3.5
C	12	33	3	2–100
D	22	44	3	5–75
E	21	50	3	2–120
F	24	51	3	3–180
G	21	50	3	5–80

*Dose levels per subject.

The paper is organized as follows. A description of materials is provided in Section II. Section III describes the Population Smoothing Splines (PSS) method. Results obtained in experimental and simulated scenarios are illustrated in Section IV. Section V discusses our model-free approach to modeling dose escalation data. Estimation algorithm, details of the simulation study, and references to the source code in R language [37] are reported in the Appendices.

II. MATERIALS

Methods described in the present paper were tested in real scenarios. Experimental datasets from seven different dose escalation studies were used. Data refer to compounds in early clinical development proposed for the treatment of psychiatric disorders. Each subject received placebo and up to three or five ascending doses of the experimental drug (see Table I). The initial doses were chosen as an appropriate submultiple of the expected pharmacologically active dose. Subjects were recruited until predefined safety or pharmacokinetic stopping limits were reached. These limits are typically defined using data collected in preclinical toxicological experiments. Safety stopping limits are defined as a given incidence and severity of unwanted effects (as expected based on preclinical experiments) in the subjects enrolled in the dose escalation study. The pharmacokinetic stopping limits are defined as the metrics of systemic exposure to the drug (peak plasma concentration ($C_{\max}$), area under the plasma concentration-time curves (AUC)) calculated in animals given the highest dose level without safety/tolerability concerns (typically, the maximal dose level at which no adverse effects were observed in animals). The seven studies differed in the number of subjects (12 to 40), number of observations per subject (1 to 5), total number of observations (33 to 100), number of dose levels per subject (3 to 5), and dose range (see also Table I).

Additionally, 300 simulated studies generated with three parametric models described in Section III (100 datasets each) were analyzed in a comparative benchmark.

III. METHODS

Consider the problem of estimating continuous functions $z^j, j = 1, \ldots, N$ from sparse measurements Y_k^j taken at discrete doses D_k^j on N subjects. In the following, the mathematical relationship between doses and measurements (e.g., metrics of systemic exposure such as peak concentration, $C_{\max}$, or area under the plasma concentration-time curve, AUC) will be modeled after a logarithmic transformation: $y_k^j := \log(Y_k^j)$ and

$d_k^j := \log(D_k^j)$. For the *j*th subject, the following measurements are available:

$$y_k^j = z^j(d_k^j) + v_k^j, \qquad k = 1, \ldots, n_j$$

where v_k^j are mutually independent, normally distributed measurement errors with zero mean and variance σ^2.

Individual curves $z^j(d)$ may be decomposed as:

$$\begin{aligned} \bar{z}^*(d) &= \phi^T(d)\zeta + \bar{z}(d) \\ z^j(d) &= \bar{z}^*(d) + \tilde{z}^j(d) \end{aligned} \tag{1}$$

where $\phi^T(d) = [1 \;\; d]$ and $\zeta \sim N(0, \infty\mathbf{I})$. Note that the term $\bar{z}^*(d)$, which does not depend on j, represents the so-called *typical curve*, that is, the typical behavior of the population.

In vector notation, (1) is expressed by

$$\begin{aligned} \bar{\mathbf{z}} &:= \left[\bar{z}(d_1^1)\ldots\bar{z}(d_{n_1}^1) \quad \ldots \quad \bar{z}(d_1^N)\ldots\bar{z}(d_{n_N}^N)\right]^T \\ \tilde{\mathbf{z}} &:= \left[\tilde{z}^1(d_1^1)\ldots\tilde{z}^1(d_{n_1}^1) \quad \ldots \quad \tilde{z}^N(d_1^N)\ldots\tilde{z}^N(d_{n_N}^N)\right]^T \\ \mathbf{v} &:= \left[v_1^1\ldots v_{n_1}^1 \quad \ldots \quad v_1^N\ldots v_{n_N}^N\right]^T \\ \mathbf{\Phi} &:= \left[\phi(d_1^1)\ldots\phi(d_{n_1}^1) \quad \ldots \quad \phi(d_1^N)\ldots\phi(d_{n_N}^N)\right]^T \\ \mathbf{y} &= \mathbf{\Phi}\zeta + \bar{\mathbf{z}} + \tilde{\mathbf{z}} + \mathbf{v}. \end{aligned}$$

Unknown functions $\bar{z}(d)$ and $\tilde{z}^j(d)$ are modeled as Gaussian stochastic processes with zero mean and autocovariance functions $\bar{R}(d,\delta)$ and $\tilde{R}(d,\delta)$, respectively (defined in Appendix A). More precisely, $\bar{z}(d)$ and $z^j(d)$ are assumed to be integrated Wiener processes:

$$\begin{aligned} \dot{\bar{x}}(d) &= \mathbf{A}\bar{x}(d) + \mathbf{B}\bar{w}(d) \qquad \bar{x}(0) = 0 \\ \bar{z}(d) &= \mathbf{C}\bar{x}(d) \\ \dot{\tilde{x}}^j(d) &= \mathbf{A}\tilde{x}^j(d) + \mathbf{B}\tilde{w}^j(d) \qquad \tilde{x}^j(0) \sim N\left(0, \begin{bmatrix} \sigma_0^2 & 0 \\ 0 & 0 \end{bmatrix}\right) \\ \tilde{z}^j(d) &= \mathbf{C}\tilde{x}^j(d) \\ \mathbf{A} &= \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \mathbf{C} = [1 \quad 0]. \end{aligned}$$

Random processes $\bar{w}(d)$ and $\tilde{w}^j(d)$ are zero-mean white Gaussian noises with variance $\bar{\lambda}^2$ and $\tilde{\lambda}^2$, respectively. The variance parameters $\bar{\lambda}^2$, $\tilde{\lambda}^2$, σ_0^2, and σ^2 are also called hyperparameters of the model. Moreover, $\bar{w}(d)$, $\tilde{w}^j(d)$, $\bar{x}(0)$, $\tilde{x}^j(0)$, ζ, and v_k^j are assumed independent $\forall j, k$.

Note that the term $\phi^T(d)\zeta$ in (1) allows modeling completely uncertain initial conditions of $\bar{x}(d)$, which could not be obtained by imposing infinite variance on $\bar{x}(0)$ (see [23] for a detailed discussion).

In a Bayesian setting, the above stochastic model is widely used to formalize prior knowledge about the smoothness of an unknown function. As a matter of fact, the second derivatives of $\bar{z}(d)$ and $\tilde{z}^j(d)$ have finite variance, which is equivalent to assume that realizations of $\bar{z}(d)$ and $\tilde{z}^j(d)$ are continuous up to the first derivative. According to the Bayesian paradigm, the estimate is just the posterior expectation given the observed data $\mathbf{y}$. Assuming hyperparameters as known, the posterior expectation can be obtained in closed form since the stochastic model is completely linear. Notably, the solution yields so-called PSS [23], that is, piecewise cubic polynomials. As noted elsewhere, PSS can be interpreted as a kind of regularization network, i.e., a basis function network whose weights are computed by solving a Tychonov-type regularization problem [38], [39]. The detailed estimation procedure is given in Appendix A, where also the issue of hyperparameter tuning is addressed.

Moreover, population parametric models were used to describe the dose-measurement relationship. In particular, we explored:

1) A power model:

$$Y_k^j = \alpha^j (D_k^j)^{\beta^j}(1 + V_k^j)$$

2) An $E_{\max}$ model:

$$Y_k^j = \frac{E_{\max}^j D_k^j}{E_{50}^j + D_k^j}(1 + V_k^j)$$

3) A Hill (sigmoidal) model:

$$Y_k^j = \frac{E_{\max}^j (D_k^j)^\gamma}{(E_{50}^j)^\gamma + (D_k^j)^\gamma}(1 + V_k^j).$$

The three parametric models were fitted with WinBUGS 1.4.3 [40], whereas PSS using R 2.8.0 [37]. In all cases, a Bayesian population approach was adopted for model estimation. Noninformative priors were chosen for population parameters and measurement error variance. Priors of interindividual variances in the parametric models were automatically tuned through a preliminary two-stage fitting.

IV. Results

A. Analysis of Experimental Dose Escalation Datasets

The PSS method was tested on seven experimental dose-escalation datasets described in Section II and labeled A to G in the following. All datasets were successfully fitted. As an example, results of parameter estimation from dataset B are presented in Fig. 1, which shows the population dose-exposure relationship and 90% predictive limits superimposed on the whole dataset. The corresponding fitting of individual data is shown in Fig. 2 for a subset of six subjects. The population response curve has been superimposed as a reference, so as to point out the heterogeneity of the subjects in terms of dose-exposure relationship. Note that, given the small number of observations per subject, the individual dose-exposure curve is a compromise between the population curve and the individual data.

Furthermore, Fig. 3 shows estimated population curve, predictive limits and exposure data for the other six dose escalation studies. Interestingly, uncertainty in the distribution of dose-exposure relationships is accounted for by larger predictive limits in dose regions where no data are available (see for example study C). Also note that, although in study C the distribution of data points at the higher doses appears shifted toward the lower percentile, this should not be regarded as a symptom of model misspecification: since the escalation is controlled by safety criteria, subjects characterized by a dose-exposure curve

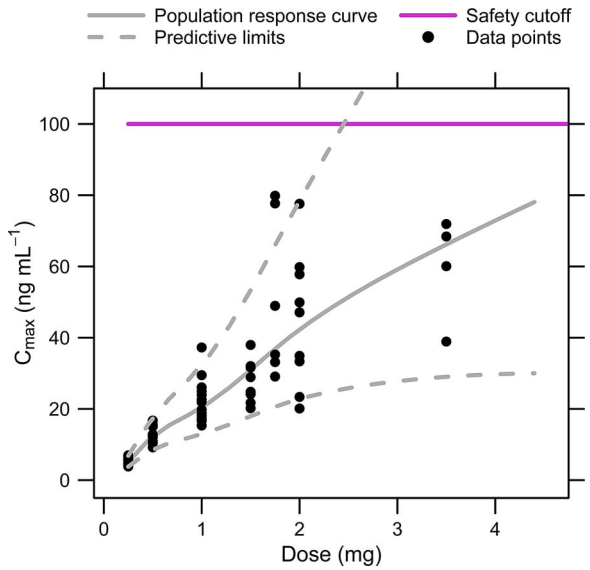


Fig. 1. Population response curve relative to experimental dataset B. The 90% predictive limits, which represent the population distribution, were obtained through MonteCarlo simulation. The safety cutoff is the pharmacokinetic stopping limit (see Section I).

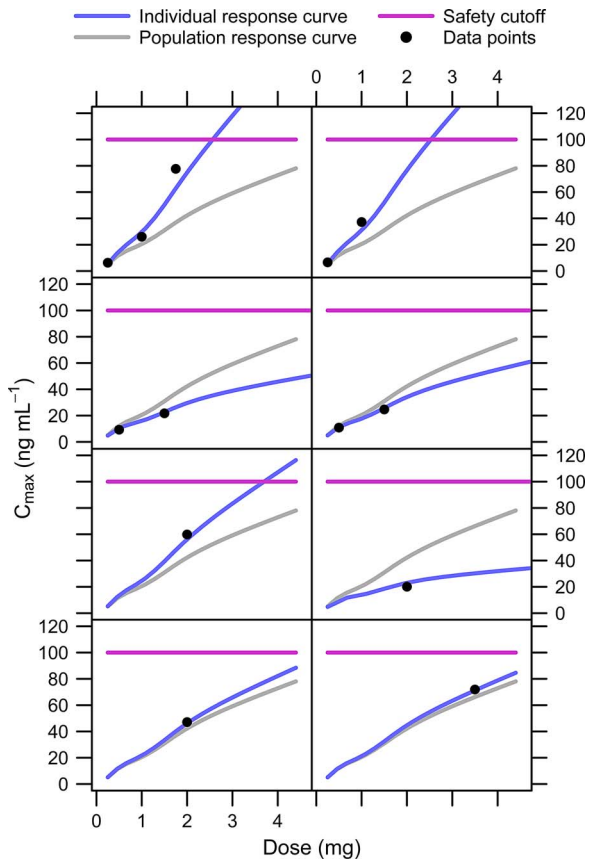


Fig. 2. Individual predictions for a subset of six subjects from study B. The heterogeneity of subjects is apparent. Given the small number of observations, the individual dose-exposure curve is a compromise between the population curve and the individual data.

lying above the population curve are, in fact, less likely to receive high doses, as opposed to those who respond less (i.e., whose curves lie below the population curve).

In addition to PSS, the three parametric models described in Section III were fitted on the seven experimental datasets. Performances of the four approaches were evaluated by means of three metrics:

1) Identification root mean square error (RMSE), which reflects goodness-of-fit (Fig. 4);
2) Bayesian information criterion (BIC), as a metric of model complexity (Fig. 5);
3) Root mean square error obtained from cross validation (cross-validatory RMSE) [41], to assess the predictive capabilities of PSS relative to the three parametric methods (Fig. 6). Each dataset was split into an identification set and a validation set, the latter including higher doses than the former. Amount of data included in the identification set ranged from 41% to 82% with respect to the total number of observations. Identification sets were made large enough to allow identifiability of all four models.

B. Simulation Study

Performances of PSSs were tested in a simulated benchmark, as detailed in Appendix B. Each parametric model (Power, $E_{\max}$, Hill) was used to generate 100 synthetic datasets. The generating model and PSS were then fitted, and compared in terms of identification RMSE, BIC, and cross-validatory RMSE, so as to emphasize the contribution of goodness-of-fit, model complexity and predictive capability to model performances. Fig. 7 shows boxplots of identification RMSE, BIC, and cross-validatory RMSE for each comparison, i.e., generating model versus PSS.

V. Discussion

The availability of effective nonparametric methods is of great interest in the population pharmacokinetic field. Especially in the early stages of drug development (e.g., phases 0 and 1), in absence of reliable parametric models, nonparametric estimation may help both establishing a dose-exposure relationship [22] and checking for misspecification of candidate parametric models.

The proposed nonparametric model for the population analysis of multiple experiments has been successfully employed to model dose escalation data in phase I clinical trials, although in principle our approach could as well be applied to the analysis of longitudinal pharmacokinetic or pharmacokinetic/pharmacodynamic (PK/PD) data. The average curve as well as the individual ones were modeled as Gaussian processes. The posterior expectation of such processes, given the available data (Bayes estimate), takes the form of a regularization network [23], [38], [39], i.e., the linear combination of autocovariance functions centered at the sampling knots. The network weights were computed by solving a system of linear equations. Modeling the average and individual curves as integrated Wiener processes yields estimate cubic splines as estimates.

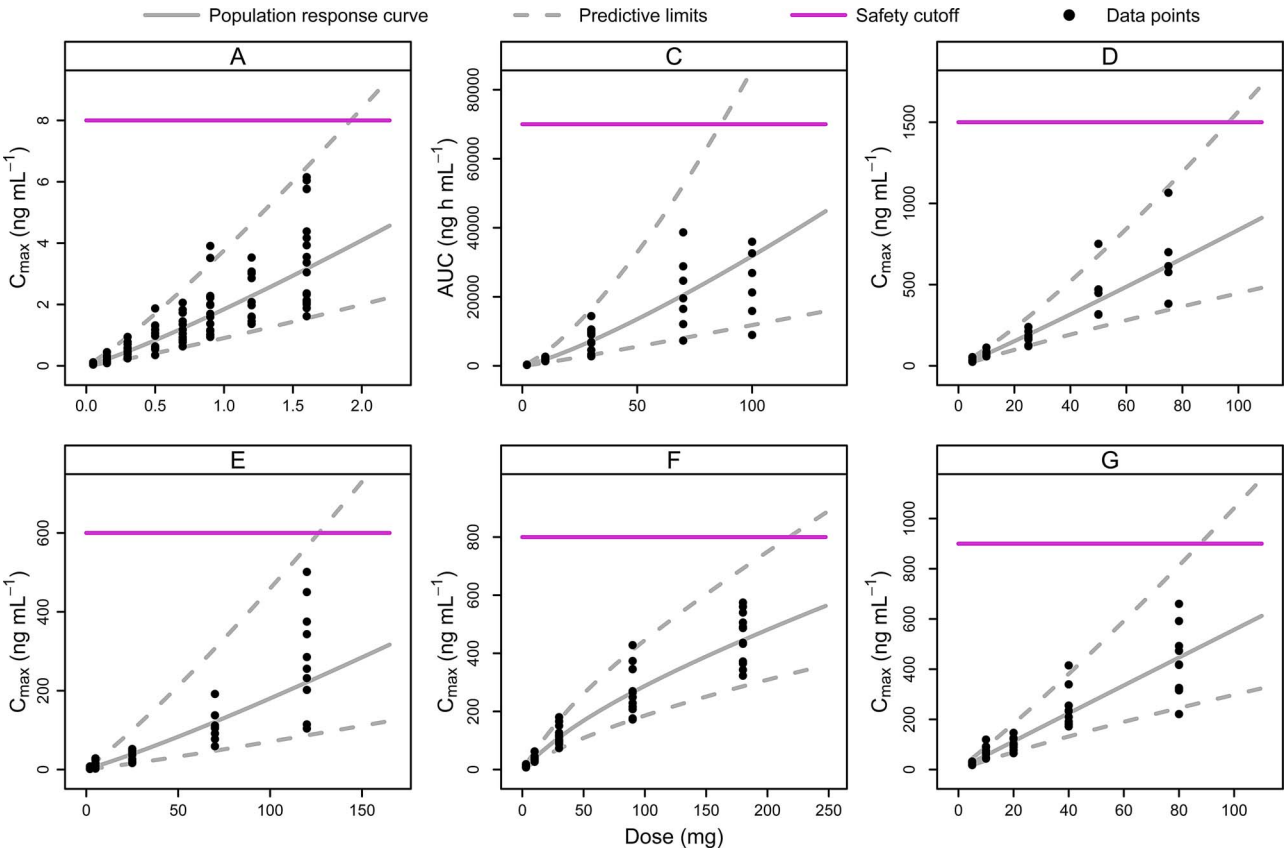


Fig. 3. Population response curve and 90% predictive limits for the six experimental datasets A and C–G.

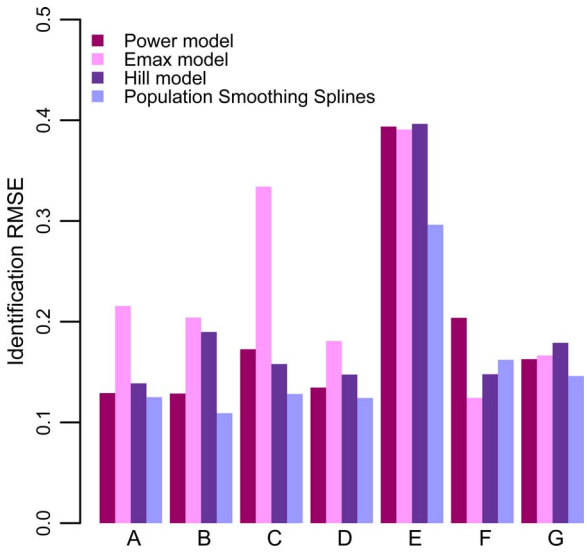


Fig. 4. Identification RMSEs obtained from the seven experimental datasets.

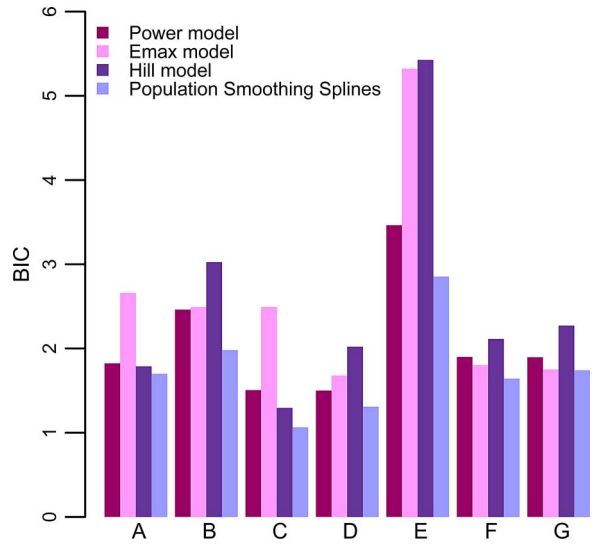


Fig. 5. BIC scores obtained from the seven experimental datasets.

A thorough model comparison procedure was applied, based on residuals, model complexity, and cross validation. Evaluating several performance metrics allows a useful cross check when one is faced with the problem of finding the most adequate model for a given study.

In experimental studies, PSS robustly handled a variety of scenarios, thus overcoming possible misspecification problems. The proposed approach yielded excellent goodness-of-fit, BIC, and cross-validatory RMSE, performing comparably to, or better than parametric methods. In particular, PSS achieved better identification RMSE than parametric methods in six studies out of seven, with PSS still better than the Power model in study F (Fig. 4). With respect to model complexity, analysis of BIC scores (Fig. 5) highlighted PSS as the preferred method in all

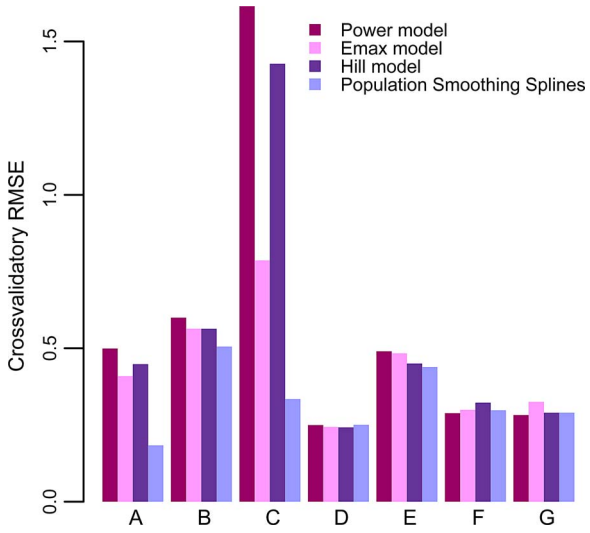


Fig. 6. Cross-validatory RMSEs obtained from the seven experimental datasets.

studies, although in study G the $E_{\max}$ model achieves comparable performance. Such outcome of the model complexity analysis is not trivial at all, given the generally large number of degrees-of-freedom of spline models [42]. Last, PSS performed well also in cross-validation (Fig. 6). PSS won the comparison in four cases (studies A–C, and E). Moreover, cross-validatory RMSEs obtained with PSS were substantially equivalent to the winning methods in the remaining three studies.

In the simulated benchmark, our nonparametric method outperformed parametric techniques on their own home ground. In the six comparisons relative to PSS versus Power and $E_{\max}$ models (Fig. 7), PSS yielded better results even if the datasets were generated by the parametric methods themselves. Although it may look surprising that the Gaussian process approach improves on using the true model structure, there are other recent results (see Fig. 4 in [43]) showing the effectiveness of nonparametric estimation complemented with automatic hyperparameter tuning (see the likelihood maximization A1 in Appendix A) in comparison with traditional parametric modeling. In our benchmark, the percentage of PSS "winning matches" (as determined by lower identification RMSE, BIC, or cross-validatory RMSE) ranged from 70% to 91%, with respect to the number of simulated datasets (100 for each parametric model).

In the three cases concerning PSS versus the Hill model, percentages amount to 28%, 40%, and 41% relative to identification RMSE, BIC, or cross-validatory RMSE, respectively. Note, however, that such results, in a real-life perspective, still look comforting: If we were to analyze an experimental dataset, without knowing that the underlying structure is that of a Hill model, we would still be able to apply PSS successfully (in terms of model complexity and cross-validatory performance) with about 40% chance.

In view of the above results, an added value of PSS is evident when few data are available to suggest a parametric model, and one wants to perform reliable predictions: evaluation of cross-validatory RMSE yielded percentages of PSS wins equal to 85%, 70%, and 41% (PSS versus Power, $E_{\max}$ and Hill models, respectively).

Finally, it is worth adding some comments on the issue of extrapolation, which is obviously fundamental in dose-escalation studies. In this respect, one may feel more comfortable with parametric models rather than nonparametric ones, as the latter seem more empirical. While in some cases this might be true, when there is not enough information to support a specific parametric model (as typically happens in dose-escalation studies), the application of a nonparametric model may be less biased. To make an example, if the parametric model is linear in log-log scale (i.e., a power model is assumed for the original data), extrapolations at high doses are based on intercept and slope parameters that optimize the fit over the whole dose range. In case of a model mismatch, such an extrapolation is all but safe because predictions for high doses will be affected by the need to fit data collected at low doses. Conversely, the proposed nonparametric approach yields linear predictions in log-log scale (i.e., power model extrapolations in the natural scale) that extrapolate the behavior observed at nearby doses, without being overly affected by the need to accommodate the data collected at lower doses. This extrapolation strategy is conceptually more robust than postulating the existence of a *true* parametric model. Indeed, our results from simulated and experimental studies confirm the effectiveness of Gaussian process nonparametric extrapolation.

In conclusion, the approach proposed in this paper for relating exposure to doses is a very valuable alternative to parametric models, especially when the amount of information collected does not allow to resort to parametric models, or there is a danger of model misspecification.

APPENDIX A
PSS ESTIMATION

Since the statistics of the processes ($\bar{\lambda}^2$, $\tilde{\lambda}^2$, and σ_0^2), as well as the magnitude of the measurement error (σ^2), are generally not known, an empirical Bayes scheme [44] is here adopted: first, hyperparameters $\left\{\bar{\lambda}^2, \tilde{\lambda}^2, \sigma_0^2, \sigma^2\right\}$ are estimated via maximum likelihood (ML), then their ML estimates are plugged into the Bayes estimator. Therefore, since all the involved processes (conditional on the hyperparameters) are jointly Gaussian, the posterior distributions are Gaussian as well. The following formulas provide the posterior means for the typical curve and the individual ones. The detailed derivation of the formulas described later, as well as the expression of the posterior variance (not employed in the present work), are reported in [23].

Posterior means are obtained as the linear combination of autocovariance functions $\bar{R}(t, t_k^j)$ and $\tilde{R}(t, t_k^j)$, evaluated at the sampling knots t_k^j, weighted with coefficients c_k^j:

$$\hat{\bar{z}}^*(d) := E\left[\bar{z}^*(d) \mid \mathbf{y}\right] = \sum_{j=1}^{N} \sum_{k=1}^{n_j} c_k^j \bar{R}(d, d_k^j) + \phi^T(d)\mathbf{b}$$

$$\hat{z}^j(d) := E\left[z^j(d) \mid \mathbf{y}\right] = \hat{\bar{z}}^*(d) + \sum_{k=1}^{n_j} c_k^j \tilde{R}(d, d_k^j).$$

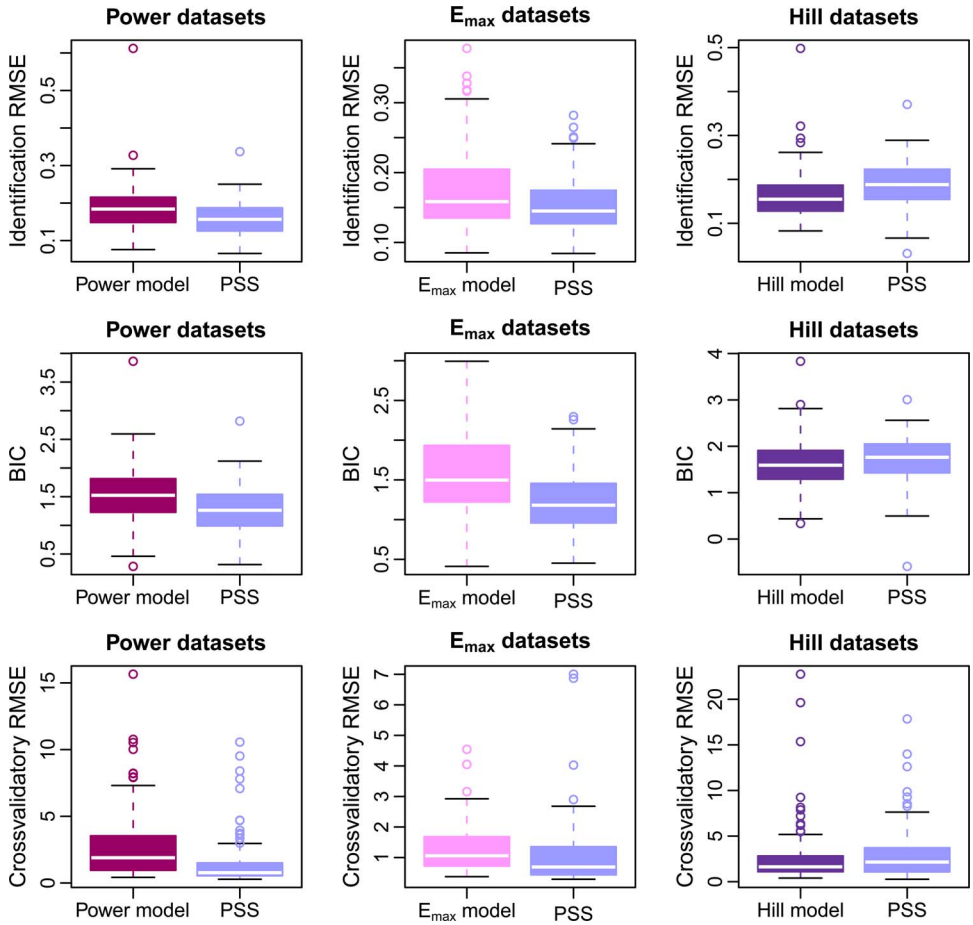


Fig. 7. Identification RMSE, BIC, and cross-validatory RMSEs obtained from 100 simulated datasets.

Autocovariances $\bar{R}(t, t_k^j)$ and $\tilde{R}(t, t_k^j)$ are given by

$$\bar{R}(d,\delta) = \bar{\lambda}^2 \begin{cases} \frac{d^2}{2}\left(\delta - \frac{d}{3}\right) & d \le \delta \\ \frac{\delta^2}{2}\left(d - \frac{\delta}{3}\right) & d > \delta \end{cases}$$

$$\tilde{R}(d,\delta) = \sigma_0^2 + \tilde{\lambda}^2 \begin{cases} \frac{d^2}{2}\left(\delta - \frac{d}{3}\right) & d \le \delta \\ \frac{\delta^2}{2}\left(d - \frac{\delta}{3}\right) & d > \delta \end{cases}.$$

Moreover, the variance of vectors $\bar{\mathbf{z}}$ and $\tilde{\mathbf{z}}$ is expressed in matrix form by

$$\bar{\mathbf{R}} = Var\left[\bar{\mathbf{z}}\right] := \begin{bmatrix} \bar{R}(d_1^1, d_1^1) & \dots & \bar{R}(d_1^1, d_{n_N}^N) \\ \vdots & \vdots & \vdots \\ \bar{R}(d_{n_N}^N, d_1^1) & \dots & \bar{R}(d_{n_N}^N, d_{n_N}^N) \end{bmatrix}$$

$$\tilde{\mathbf{R}} = Var\left[\tilde{\mathbf{z}}\right] := blockdiag\left\{\tilde{\mathbf{R}}^1, \dots, \tilde{\mathbf{R}}^N\right\}$$

$$\tilde{\mathbf{R}}^j := \begin{bmatrix} \tilde{R}(d_1^j, d_1^j) & \dots & \tilde{R}(d_1^j, d_{n_j}^j) \\ \vdots & \vdots & \vdots \\ \tilde{R}(d_{n_j}^j, d_1^j) & \dots & \tilde{R}(d_{n_j}^j, d_{n_j}^j) \end{bmatrix}.$$

Weights c_k^j can be interpreted as the weights of a *regularization network* [38], [39], and are computed as

$$\mathbf{M} = \bar{\mathbf{R}} + \tilde{\mathbf{R}} + \boldsymbol{\Sigma}_v \qquad \mathbf{N} = \left(\boldsymbol{\Phi}^T \mathbf{M}^{-1} \boldsymbol{\Phi}\right)^{-1}$$

$$\boldsymbol{\gamma}^T \boldsymbol{\gamma} = \mathbf{y}^T \mathbf{M}^{-1} \mathbf{y} - \mathbf{y}^T \mathbf{M}^{-1} \boldsymbol{\Phi} \mathbf{N} \boldsymbol{\Phi}^T \mathbf{M}^{-1} \mathbf{y}$$

$$\mathbf{b} = \mathbf{N} \boldsymbol{\Phi}^T \mathbf{M}^{-1} \mathbf{y} \qquad \mathbf{c} = \mathbf{M}^{-1}\left(\mathbf{y} - \boldsymbol{\Phi}\mathbf{b}\right).$$

In view of the above definitions, maximum likelihood (ML) estimates of hyperparameters are computed as

$$\boldsymbol{\Theta}^{ML} := \arg\min_{\Theta} \left\{\log(|\mathbf{M}|) - \log(|\mathbf{N}|) + \boldsymbol{\gamma}^T \boldsymbol{\gamma}\right\}$$
$$\left\{\bar{\lambda}^2, \tilde{\lambda}^2, \sigma_0^2, \sigma^2\right\} = \boldsymbol{\Theta}^{ML}. \tag{A1}$$

Appendix B
Simulation Study

Performances of PSS were compared to Power, $E_{\max}$, and Hill models using a MonteCarlo procedure. Each parametric model (Power, $E_{\max}$, Hill) was used to generate 100 synthetic datasets containing 12 subjects and 3 observations/subject. A first cohort of four subjects received doses {0.25, 0.5, 1 mg}, second one received {1, 1.5, 1.75 mg}, third one received {1.75, 2, 3.5 mg}, so as to mimic a realistic dosing schedule (such a dose grid is also featured in experimental dataset B). Inter- and intra-individual variances were chosen so as to yield a realistic degree of heterogeneity in the simulated trials. Parameter values used to simulate datasets are reported in Table II. A constant coefficient of variation of 20% was used to generate the measurement error.

TABLE II
SIMULATION PARAMETERS

Model	Parameter	Population value	IIV
Power	α	10	43%*
	β	0.5	20%
E_{max}	E_{max}	80	14%*
	E_{50}	0.5	46%*
Hill	E_{max}	80	14%*
	E_{50}	1.5	78%*
	γ	3	–

Population values and their percentage inter-individual variability (IIV) are reported.
*Relative to the logarithm of the population value.

In order to apply the cross-validation, each simulated dataset was split into an identification set, featuring cohort 1 (all doses) and 2 (1 and 1.5 mg), and a validation set, featuring cohort 2 (1.75 mg) and 3 (all doses). Posterior means of the individual response curves were used as point predictions. The RMSE between such point predictions and the validation data were then considered as the performance metrics.

APPENDIX C
R SOURCE CODE

The PSS method was implemented and tested using R version 2.8.0 [37] under Windows XP Professional. The PSS source code in R language as well as example datasets are available from the corresponding author, or can be downloaded at `http://aimed11.unipv.it/PSS`.

REFERENCES

[1] S. L. Beal and L. B. Sheiner, "Estimating population kinetics," *Crit. Rev. Biomed. Eng.*, vol. 8, pp. 195–222, 1982.

[2] M. Davidian and D. M. Giltinan, *Nonlinear Models for Repeated Measurement Data*. New York: Chapman and Hall, 1995.

[3] L. B. Sheiner, "The population approach to pharmacokinetic data analysis: Rationale and standard data analysis methods," *Drug Metab. Rev.*, vol. 15, pp. 153–171, 1984.

[4] L. Aarons, "Software for population pharmacokinetics and pharmacodynamics," *Clin. Pharmacokinet.*, vol. 36, pp. 255–264, 1999.

[5] V. Guardabasso, P. J. Munson, and D. Rodbard, "A versatile method for simultaneous analysis of families of curves," *FASEB J.*, vol. 2, pp. 209–215, 1988.

[6] L. B. Sheiner, B. Rosenberg, and V. V. Marathe, "Estimation of population characteristics of pharmacokinetic parameters from routine clinical data," *J. Pharmacokin. Biopharm.*, vol. 5, pp. 445–479, 1977.

[7] L. B. Sheiner and J. L. Steimer, "Pharmacokinetic/pharmacodynamic modeling in drug development," *Annu. Rev. Pharmacol. Toxicol.*, vol. 40, pp. 67–95, 2000.

[8] S. Vozeh, J. L. Steimer, M. Rowland, P. Morselli, F. Mentré, L. P. Balant, and L. Aarons, "The use of population pharmacokinetics in drug development," *Clin. Pharmacokinet.*, vol. 30, pp. 81–93, 1996.

[9] L. Yuh, S. Beal, M. Davidian, F. Harrison, A. Hester, K. Kowalski, E. Vonesh, and R. Wolfinger, "Population pharmacokinetic/pharmacodynamic methodology and applications: A bibliography," *Biometrics*, vol. 50, pp. 566–575, 1994.

[10] Center for Drug Evaluation and Research, *Guidance for industry: Population pharmacokinetics*. United States Department of Health and Human Services, Food and Drug Administration, 1999.

[11] A. Bertoldo, G. Sparacino, and C. Cobelli, "Population approach improves parameter estimation of kinetic models for dynamic PET data," *IEEE Trans. Med. Imag.*, vol. 23, pp. 297–306, 2004.

[12] P. Vicini and C. Cobelli, "The iterative two-stage population approach to IVGTT minimal modeling: Improved precision with reduced sampling," *Am. J. Physiol. Endocrinol. Metab.*, vol. 280, pp. 179–186, 2001.

[13] F. Ferrazzi, P. Magni, and R. Bellazzi, "Bayesian clustering of gene-expression time series," in *Proc. 3rd Int. Workshop Bioinform. Manag. Anal. Interpretation Microarray Data (NETTAB 2003)*, pp. 53–55.

[14] S. L. Beal and L. B. Sheiner, *NONME Users Guide.*. San Francisco: Univ. California, 1998.

[15] R. Jelliffe, A. Schumitzky, M. V. Guilder, X. Wang, and R. Leary, "Population pharmacokinetic and dynamic models: Parametric (P) and nonparametric (NP) approaches," in *Proc. 14th IEEE Symp. Computer-Based Med. Syst.*, 2001, pp. 407–412.

[16] R. Leary, R. Jelliffe, A. Schumitzky, and M. V. Guilder, "An adaptive grid non-parametric approach to pharmacokinetic and dynamic (PK/PD) population models," in *Proc. 14th IEEE Symp. Computer-Based Med. Syst.*, 2001, pp. 389–394.

[17] J. Wakefield and J. Bennett, "The Bayesian modelling of covariates for population pharmacokinetic models," *JASA*, vol. 91, pp. 917–927, 1996.

[18] J. Wakefield, A. F. M. Smith, A. Racine-Poon, and A. Gelfand, "Bayesian analysis of linear and nonlinear population models using the Gibbs sampler," *Appl. Statist.*, vol. 43, pp. 201–221, 1994.

[19] I. A. Ibragimov and R. Z. Khasminskii, *Statistical estimation: Asymptotic theory*. New York: Springer-Verlag, 1981.

[20] K. E. Fattinger and D. Verotta, "A nonparametric subject-specific population method for deconvolution: I. description, internal validation, and real data examples," *J. Pharmacokin. Biopharm.*, vol. 23, pp. 581–610, 1995.

[21] K. Park, D. Verotta, T. F. Blaschke, and L. B. Sheiner, "A semiparametric method for describing noisy population pharmacokinetic data," *J. Pharmacokin. Biopharm.*, vol. 25, pp. 615–642, 1997.

[22] P. Magni, R. Bellazzi, G. D. Nicolao, I. Poggesi, and M. Rocchetti, "Nonparametric AUC estimation in population studies with incomplete sampling: A Bayesian approach," *J. Pharmacokin. Pharmacodyn.*, vol. 29, pp. 445–471, 2002.

[23] M. Neve, G. D. Nicolao, and L. Marchesi, "Nonparametric identification of population models via Gaussian processes," *Automatica*, vol. 43, pp. 1134–1144, 2007.

[24] "Nonparametric identification of population models: an MCMC approach," *IEEE Trans. Biomed. Eng.*, vol. 55, pp. 41–50, 2008.

[25] S. Piantadosi and G. Liu, "Improved designs for dose escalation studies using pharmacokinetic measurements," *Stat. Med.*, vol. 15, pp. 1605–1618, 1996.

[26] S. Piantadosi, "Dose-finding designs," in *Clinical trials: a methodologic perspective*, S. Piantadosi, Ed. Hoboken, NJ: Wiley, 2005, pp. 223–250.

[27] F. Rombout, L. Aarons, M. O. Karlsson, A. Man, F. Mentré, P. Nygren, A. Racine, H. Schaefer, J. L. Steimer, and I. Troconiz, "Modelling and simulation in the development and use of anti-cancer agents: An underused tool?," *J. Pharmacokin. Pharmacodyn.*, vol. 31, pp. 419–440, 2004.

[28] J. O'Quigley, M. Pepe, and L. Fisher, "Continual Reassessment Method: A practical design for Phase I clinical trials in cancer," *Biometrics*, vol. 46, pp. 33–48, 1990.

[29] P. F. Thall and K. E. Russell, "A strategy for dose-finding and safety monitoring based on efficacy and adverse outcomes in Phase I/II clinical trials," *Biometrics*, vol. 54, pp. 251–264, 1998.

[30] J. M. Heyd and B. P. Carlin, "Adaptive design methods in the Continual Reassessment Method for Phase I studies," *Stat. Med.*, vol. 18, pp. 1307–1321, 1999.

[31] M. Gasparini and J. Eisele, "A curve-free method for Phase I clinical trials," *Biometrics*, vol. 56, pp. 609–615, 2000.

[32] S. Zohar and S. Chevret, "The continual reassessment method: comparison of Bayesian stopping rules for dose-ranging studies," *Stat. Med.*, vol. 20, pp. 2827–2843, 2001.

[33] Y. Zhou, J. Whitehead, E. Bonvini, and J. W. Stevens, "Bayesian decision procedures for binary and continuous bivariate dose-escalation studies," *Pharm. Stat.*, vol. 5, pp. 125–133, 2006.

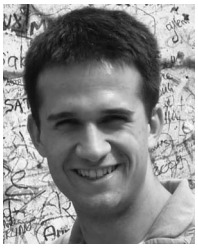
[34] J. Whitehead, Y. Zhou, A. Mander, S. Ritchie, A. Sabin, and A. Wright, "An evaluation of Bayesian designs for dose-escalation studies in healthy volunteers," *Stat. Med.*, vol. 25, pp. 433–445, 2006.

[35] D. A. Berry, P. Müller, A. P. Grieve, M. Smith, T. Parke, R. Blazek, N. Mitchard, and M. Krams, "Adaptive Bayesian designs for dose-ranging drug trials," in *Case Studies in Bayesian Statistics V*, C. Gatsonis, R. E. Kass, B. Carlin, A. Carriquiry, A. Gelman, I. Verdinelli, and M. West, Eds. New York: Springer-Verlag, 2001, pp. 99–181.

[36] M. K. Smith, I. Jones, M. F. Morris, A. P. Grieve, and K. Tan, "Implementation of a Bayesian adaptive design in a proof of concept study," *Pharm. Stat.*, vol. 5, pp. 39–50, 2006.

[37] R Development Core Team. (2008). *R: A language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, [Online]. Available: http://www.R-project.org.

[38] T. Poggio and F. Girosi, "Networks for approximation and learning," *Proc. IEEE*, vol. 78, pp. 1481–1497, 1990.
[39] F. Girosi, M. Jones, and T. Poggio, "Regularization theory and neural networks architectures," *Neural Comput.*, vol. 7, pp. 219–269, 1995.
[40] D. J. Lunn, A. Thomas, N. Best, and D. J. Spiegelhalter, "WinBUGS – A Bayesian modelling framework: concepts, structure, and extensibility," *Stat. Comput.*, vol. 10, pp. 325–337, 2000.
[41] A. Vehtari and J. Lampinen, "Bayesian model assessment and comparison using cross-validation predictive densities," *Neural Comput.*, vol. 14, pp. 2439–2468, 2002.
[42] G. Wahba, *Spline Models for Observational Data*. Philadelphia, PA: SIAM, 1990.
[43] G. Pillonetto and G. D. Nicolao, "A new kernel-based approach to linear system identification," *Automatica*, vol. 46, pp. 81–93, 2010.
[44] H. Robbins, "An Empirical Bayes approach to statistics," in *Proc. Third Berkeley Symp. Math. Statist. Prob.*. Berkeley: Univ. California Press, 1956, pp. 157–163.



Alberto Russu received the Master's degree *cum laude* in computer engineering in 2006 and the Ph.D. degree in bioengineering and bioinformatics with the thesis "Bayesian population approaches to pharmacometrics and clinical trials", in 2010, both from the University of Pavia, Italy.

In 2006, he visited the European Space Research and Technology Center at the European Space Agency (ESA-ESTEC), Noordwijk, The Netherlands. In 2009, he visited the Pharmacology Division at the Leiden/Amsterdam Center for Drug Research, Leiden, The Netherlands, and the Department of Clinical Pharmacology and Discovery Medicine, GlaxoSmithKline, Greenford, UK. Since 2011, he is a Postdoctoral Researcher with the Department of Computer Engineering and Systems Science, University of Pavia, Italy. His research interests include Bayesian methods, population modeling in pharmacometrics, and clinical trial analysis and design.

Italo Poggesi received the Laurea degree (*cum laude*) in chemistry from the University of Pavia,Italy, in 1986.

Since 1988, he held different positions in the pharmaceutical industry, in the areas of pharmacokinetics, metabolism, and drug development. He was Head of Pharmacokinetic Data Analysis and Prediction and Modeling groups within Pharmacia, Milan, Italy, and, subsequently Director and Psychiatry Leader in the Clinical Pharmacokinetics/Modeling and Simulation Unit, Psychiatry CEDD, GlaxoSmithKline, Verona, Italy. He is currently Director and Advanced Modeling and Simulation Leader within the Clinical Pharmacology Group of Janssen R&D, Beerse, Belgium. His current research interests include the development and evaluation of modeling techniques with the aim of accelerating drug development and reducing the attrition of drug candidates. Since 2005, he is part of the editorial board of the journal *Expert Opinion on Drug Metabolism & Toxicology*. He is author or coauthor of more than 100 publications, more than 50 of which have been published in peer-reviewed journals. He is also coinventor of two patented methodologies.

Roberto Gomeni received a degree in mathematics from the University of Milan, Italy, and the Ph.D. degree in pharmacokinetics from the University of Montpellier, France.

He is currently President of Alleantis, a clinical research organization based in Research Triangle Park, NC, USA. He was the Head of Pharmacometrics at GlaxoSmithKline R&D, King of Prussia, PA, USA. He is responsible for strategic and leveraged utilization of model-based approach for improving efficiency in drug development projects (from lead optimisation to marketing). His current research interests are focused on the identification and implementation of strategies based on modeling approaches to enhance drug development process, drive decision making, and risk management using drug and disease progression models, clinical trial simulation, Bayesian modeling, and knowledge-based computer-assisted drug development processes. He is an author of more than 145 original research papers published in international scientific journals on individual and population PK/PD analysis and mathematical modeling in drug discovery, preclinical and clinical pharmacology.

Giuseppe De Nicolao (SM'97) received the *cum laude* degree in electronic engineering from the Polytechnic of Milan, Italy, in 1986.

From 1987 to 1988, he was with the Biomathematics and Biostatistics Unit of the Institute of Pharmacological Researches "Mario Negri," Milan. In 1988, he joined the Italian National Research Council (CNR) as a Researcher Scientist of the Center of System Theory in Milan. In 1991, he held a visiting fellowship at the Department of Systems Engineering of the Australian National University, Canberra. In 1998, he was a keynote speaker at the IFAC workshop on "Nonlinear model predictive control: Assessment and future directions for research." From 1992 to 2000, he was an Associate Professor, and, since 2000, he has been a Full Professor of model identification in the Department of Computer Science and Systems Engineering of the University of Pavia, Italy. From 1999 to 2001, he was an Associate Editor of the IEEE TRANSACTIONS ON AUTOMATIC CONTROL and, from 2007 to 2010, he was an Associate Editor of *Automatica*. His research interests include Bayesian learning, neural networks, model predictive control, optimal and robust filtering and control, deconvolution techniques, modeling, identification and control of biomedical systems, advanced process control and fault diagnosis for semiconductor manufacturing. On these subjects, he has authored or coauthored more than 100 journal papers and is coinventor of three patents.