ARTICLE

Nonparametric goodness-of-fit testing for parametric covariate models in pharmacometric analyses

Niklas Hartung[1] | **Martin Wahl**[2] | **Abhishake Rastogi**[1] | **Wilhelm Huisinga**[1]

[1]Institute of Mathematics, Universität Potsdam, Potsdam, Germany

[2]Institute of Mathematics, Humboldt-Universität zu Berlin, Berlin, Germany

Correspondence
Wilhelm Huisinga, Institute of Mathematics, Universität Potsdam, Karl-Liebknecht-Str. 24-25, 14476 Potsdam/Golm, Germany.
Email: huisinga@uni-potsdam.de

Abstract

The characterization of covariate effects on model parameters is a crucial step during pharmacokinetic/pharmacodynamic analyses. Although covariate selection criteria have been studied extensively, the choice of the functional relationship between covariates and parameters, however, has received much less attention. Often, a simple particular class of covariate-to-parameter relationships (linear, exponential, etc.) is chosen *ad hoc* or based on domain knowledge, and a statistical evaluation is limited to the comparison of a small number of such classes. Goodness-of-fit testing against a nonparametric alternative provides a more rigorous approach to covariate model evaluation, but no such test has been proposed so far. In this manuscript, we derive and evaluate nonparametric goodness-of-fit tests for parametric covariate models, the null hypothesis, against a kernelized Tikhonov regularized alternative, transferring concepts from statistical learning to the pharmacological setting. The approach is evaluated in a simulation study on the estimation of the age-dependent maturation effect on the clearance of a monoclonal antibody. Scenarios of varying data sparsity and residual error are considered. The goodness-of-fit test correctly identified misspecified parametric models with high power for relevant scenarios. The case study provides proof-of-concept of the feasibility of the proposed approach, which is envisioned to be beneficial for applications that lack well-founded covariate models.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?
Identifying covariates and establishing the relationship between covariates and model parameters is a key step in pharmacometric analyses. Although many criteria for covariate selection have been developed, there is no goodness-of-fit test for critically challenging often assumed parametric covariate-to-parameter relationships against a nonparametric alternative.

WHAT QUESTION DID THIS STUDY ADDRESS?
Development of a nonparametric statistical framework for goodness-of-fit testing of parametric covariate models, both conceptually and implementation wise, by transferring concepts from statistical learning to pharmacometrics.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?
A systematic nonparametric approach for the evaluation of parametric covariate models, which is important whenever a consensus on a covariate model is lacking, such as for the description of the maturation of metabolizing enzymes, which we investigated.
HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?
The proposed framework of goodness-of-fit testing provides a more solid statistical ground for covariate model selection.

INTRODUCTION

Pharmacokinetic/pharmacodynamic (PK/PD) models are used to describe drug concentrations/effects over time in a group of patients under treatment. Covariate models allow to describe the effect of patient characteristics (covariates) on model parameters, and play a crucial role in PK/PD model building. Common covariates are body weight, age, concentrations of important biomarkers (e.g., plasma creatinine), or genetic disposition (e.g., CYP450 polymorphisms). Additional random variability not explainable by covariates is accounted for through random effects.[1] Typically, there is detailed knowledge on the model structure (often specified in terms of compartmental models), whereas much less is known on how covariates impact drug kinetics/effects via the model parameters.

Covariate selection criteria in pharmacometric analyses have been studied extensively.[2–4] In contrast, the choice of the functional relationship between covariates and parameters has received less attention, and often, a particular parametric class of covariate-to-parameter relationships (linear, exponential, etc.) is chosen *ad hoc*.[5] To evaluate the appropriateness of a parametric covariate model class, a choice among a few candidate classes is sometimes made using likelihood ratio tests[6] or information-theoretic criteria.[7,8] However, such a comparison depends on the classes considered and does not reveal whether *any* of these classes is compatible with the PK/PD data.

Several methods have been proposed to overcome these limitations of a prespecified parametric covariate-to-parameter relationship. In the so-called two-stage approach, a covariate model is obtained by first estimating individual parameters from individual patient data, then, either choosing a suitable parametric class based on graphical analysis or solving a nonparametric regression problem.[6,9] While this approach is feasible if parameters can be identified from individual data, it is not so in the more realistic scenario of phase III clinical trials (many individuals, few datapoints per individual), where the true covariate-to-parameter relationship may be masked or wrongly attributed to other parameters.[10] Covariate models using regression splines or neural networks have also been proposed; using a small number of quadratic splines at fixed quantiles of the covariate distribution (or

similarly, a small neural network), more flexible covariate-to-parameter relationships were derived than in the commonly used classes of parametric models.[11] A nonparametric maximum likelihood algorithm was also developed and applied to clinical data; by estimating the joint distribution of parameters and covariates, it allowed to derive a nonparametric covariate-to-parameter relationship.[12]

Whereas parametric as well as nonparametric approaches have been used for the estimation problem, a nonparametric goodness-of-fit test for parametric covariate models is still lacking. For model evaluation and selection, however, a statistically sound comparison against a nonparametric alternative would be highly desirable because it allows to challenge the functional form of the covariate-to-parameter relationship more critically.

In the case of a direct covariate-to-*observable* relationship (also called nonparametric regression, or a direct problem), goodness-of-fit testing has been extensively studied in the statistical literature,[13,14] with applications in other fields, such as econometrics.[15] A standard construction is based on a mean squared distance between a parametric and a nonparametric estimator. For instance, a Nadaraya-Watson estimator can be used,[13,14] but many other nonparametric estimators work as well. The direct problem, however, is not relevant to our context, because the structural model (i.e., the parameter-to-observable relationship), is based on a class of models with established trust in the pharmacometric community; in the case of PK models with further support from reducing more detailed physiologically-based PK models.[16] Consequently, in our setting, covariate models link to (unobservable) parameters rather than to direct observations. For the estimation of a covariate-to-*parameter* model in a nonlinear parameter-to-observable relationship (also called nonlinear statistical inverse problem), a popular estimation method is Tikhonov regularization. Regularization methods have been studied in many different contexts, including inverse and ill-posed problems[17,18] and, more recently, unsupervised learning.[19–22] In the latter context, the framework of reproducing kernel Hilbert spaces (RKHS) plays a central role to address both computational and theoretical questions.[23–28]

In this paper, we propose and evaluate nonparametric goodness-of-fit tests for parametric covariate models,

transferring concepts developed in statistical learning to the pharmacological setting. We generalize known goodness-of-fit tests for the direct problem (e.g., based on kernel density estimation[13]) to nonlinear statistical inverse problems and also present a tailored numerical approach for the required computation of kernelized Tikhonov regularizers. We then demonstrate proof-of-concept in a relevant pharmacological application, the estimation of an age effect on maturation of drug-metabolizing enzymes. As a first step, here, we focus on the case with observation noise, but without random effects on the individual level.

METHODS

Nonparametric goodness-of-fit testing

We consider the statistical model

$$y_i = G\left(\theta_i, x_i\right) + \varepsilon_i, \ \theta_i = f\left(x_i\right), \ i \in \{1, \ldots, n\}, \quad (1)$$

with i.i.d. observations $(x_i, y_i)_{i=1}^n$ consisting of *covariates* $x_i \in \mathcal{X} \subset \mathbb{R}^{n_x}$ and (noisy) *observations* $y_i \in \mathbb{R}^q$, unobserved *parameters* $\theta_i \in \Theta \subset \mathbb{R}^p$, a nonlinear function $G \colon \Theta \times \mathcal{X} \to \mathbb{R}^q$ called the *mechanistic model*, independent centered noise ε_i, and a *covariate-to-parameter* mapping $f \colon \mathcal{X} \to \Theta$.

We assume the mechanistic model G to be known, which might represent a component of the solution of a system of ordinary differential equations (ODEs), observed at different time points $t_1, \ldots, t_q$, or a transformation thereof. The direct dependency of G on x_i allows to model individualized doses, and to model only particular aspects of a covariate model; this will also be used in our simulation study. Even for the simplest model, G depends nonlinearly on the parameters θ. The covariate model f is assumed to be unknown. We assume that a particular parametric class of functions $\{f_\tau, \tau \in \mathcal{T} \subset \mathbb{R}^{n_\tau}\}$ is given, chosen *ad hoc* or from domain knowledge, and aim to evaluate the hypothesis that the covariate model f belongs to this parametric class. Thus, we consider the test problem

$$H_0 \colon f \in \{f_\tau, \tau \in \mathcal{T}\} \ \text{ vs. } \ H_1 \colon f \notin \{f_\tau, \tau \in \mathcal{T}\}. \quad (2)$$

The precise formulation of the considered alternative (i.e., the space of functions for H_1) and the test statistics are still to be specified.

Testing against a nonparametric alternative

First, we consider a *nonparametric* alternative of the form $f \in \mathcal{H} \setminus \{f_\tau, \tau \in \mathcal{T}\}$, with a suitably chosen vector-valued RKHS $\mathcal{H}$ of functions $h \colon \mathcal{X} \to \mathbb{R}^p$. Briefly, an RKHS is specified via a kernel function k that is used to define a basis for $\mathcal{H}$ (see also Equation 8). Background and references on vector-valued RKHS are provided in Section S1.

We start by defining two natural estimators under the null and the alternative, namely the least squares estimator

$$f_{\widehat{\tau}} \text{ with } \quad \widehat{\tau} := \underset{\tau \in \mathcal{T}}{\operatorname{argmin}} \left[\sum_{i=1}^n \left\| y_i - G\left(f_\tau\left(x_i\right), x_i\right) \right\|^2 \right] \quad (3)$$

and the Tikhonov regularized estimator[19,22]

$$\widehat{f}^{(\lambda)} := \underset{h \in \mathcal{H}}{\operatorname{argmin}} \left[\sum_{i=1}^n \left\| y_i - G\left(h\left(x_i\right), x_i\right) \right\|^2 + \lambda \left\| h \right\|_{\mathcal{H}}^2 \right] \quad (4)$$

with regularization parameter $\lambda > 0$, respectively.[a] Based on these estimators, we consider the test statistic $T_1 = T_1^{(\lambda)}$, defined as

$$T_1 := \sum_{i=1}^n \left\| G\left(f_{\widehat{\tau}}\left(x_i\right), x_i\right) - G\left(\widehat{f}^{(\lambda)}\left(x_i\right), x_i\right) \right\|^2. \quad (5)$$

The statistic evaluates the estimated covariate-to-parameter relationships in the *space of observations* (i.e., after mapping of G); alternatively, a test statistic could be directly based on the difference of the estimated covariate-to-parameter relationships in the *space of parameters*. Smoothed versions of the parametric estimate $f_{\widehat{\tau}}$ have also been considered.[13] Both variations are introduced in Section S3.5 and all considered statistics are compared in the Discussion.

Testing against a combined parametric-nonparametric alternative

In addition to testing against a nonparametric alternative, we consider a *combined parametric-nonparametric alternative* of the form $f = f_\tau + h$ with $(\tau, h) \in \mathcal{T} \times \mathcal{H}$, but $f \notin \{f_\tau \colon \tau \in \mathcal{T}\}$. If $f_\tau \in \mathcal{H}$ for each $\tau \in \mathcal{T}$, then the two classes of functions considered under the nonparametric and combined parametric/nonparametric alternatives are identical. The reason we introduce this reformulation is to consider a different test statistic, which uses the Tikhonov-type regularization scheme

$$\widetilde{f}^{(\lambda)} := \underset{\substack{f = f_\tau + h, \\ (\tau, h) \in \mathcal{T} \times \mathcal{H}}}{\operatorname{argmin}} \left[\sum_{i=1}^n \left\| y_i - G\left(f_\tau\left(x_i\right) + h\left(x_i\right), x_i\right) \right\|^2 + \lambda \left\| h \right\|_{\mathcal{H}}^2 \right],$$

$$(6)$$

with regularization parameter $\lambda > 0$. Based on this, we define the second test statistic

$$T_2 := \sum_{i=1}^{n} \left\| G\left(f_{\hat{\tau}}\left(x_i\right), x_i\right) - G\left(\tilde{f}^{(\lambda)}\left(x_i\right), x_i\right) \right\|^2, \quad (7)$$

which is analogously defined as T_1 in the nonparametric case above.

From a modeling point of view, it is appealing that the combined model $f = f_\tau + h$ penalizes deviations from the parametric covariate model rather than the parametric covariate model itself. For example, instead of shrinking to 0 for large values of λ as the purely nonparametric covariate model does, where the mechanistic model G might even be undefined, the combined model $\tilde{f}^{(\lambda)}$ shrinks toward the parametric part, thereby avoiding this undesired behavior of the purely nonparametric model.

Critical values of test statistics

Critical values c_α for the level $0 < \alpha < 1$ and both test statistics (generically denoted T here) are approximated by a Monte Carlo procedure. Based on the data $(x_i, y_i)_{i=1}^{n}$, the least squares estimator $\hat{\tau}$ for the parametric null model is determined. Then, M synthetic datasets $(x_i, y_i^{(1)})_{i=1}^{n}, \ldots, (x_i, y_i^{(M)})_{i=1}^{n}$ are simulated under the approximate null model (with $\hat{\tau}$ instead of the unknown true value τ^*), i.e., for $m = 1, \ldots, M$,

$$y_i^{(m)} = G\left(f_{\hat{\tau}}\left(x_i\right), x_i\right) + \varepsilon_i^{(m)}, \quad i = 1, \ldots, n,$$

where $\varepsilon_i^{(m)}$ are i.i.d. realizations of the noise (we assume for simplicity that the distribution of residual errors is known). Based on the synthetic data, the statistic $T^{(m)}$ is computed as described previously. Then, letting $\widehat{F}_T$ denote the empirical distribution of T, we choose $c_\alpha = \widehat{F}_T^{-1}(1 - \alpha)$ as a critical value for the test and reject H_0 whenever $T > c_\alpha$ (because large values of T favor the alternative for all considered statistics). The workflow is depicted in Figure 1.

Efficient algorithms for estimation in a nonparametric model

Calculating the test statistics in Equations 5 and 7 requires different optimization problems to be solved, namely the least squares problem in Equation 3 to determine $f_{\hat{\tau}}$, the Tikhonov regularization problem in Equation 4 to determine $\hat{f}^{(\lambda)}$, and the combined least squares/Tikhonov regularization problem in Equation 6 to determine $\tilde{f}^{(\lambda)}$. Whereas the parametric problem is usually low-dimensional, the others are high-dimensional. Because the mechanistic model G is assumed to depend nonlinearly on the parameters θ, none of these problems can be solved in closed form. As will be shown in Section "Efficient nonparametric estimation of the maturation function", general-purpose optimizers perform poorly on the high-dimensional nonlinear problems, motivating the need for more tailored numerical approaches.
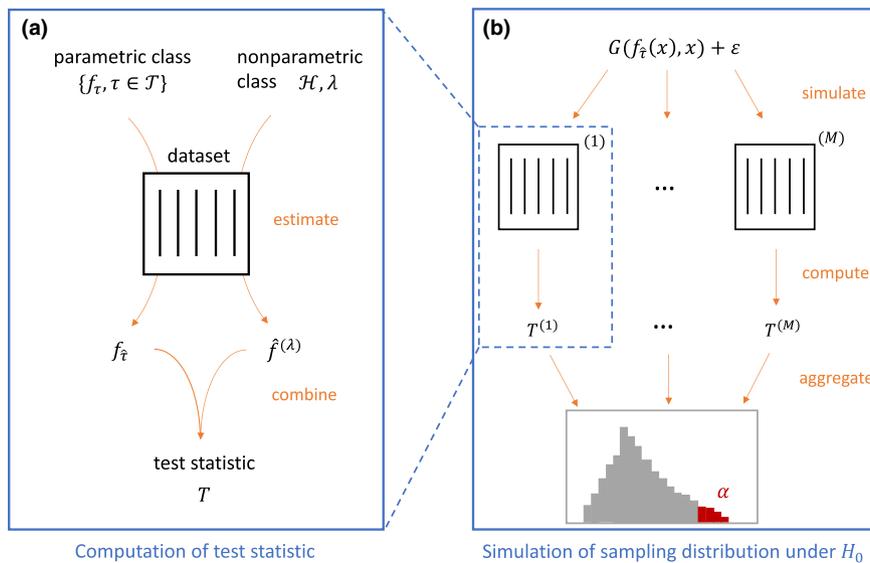

FIGURE 1 Workflow for the nonparametric goodness-of-fit test. (a) After choosing parametric and nonparametric classes, parametric and nonparametric estimators $f_{\hat{\tau}}$ and $\hat{f}^{(\lambda)}$ are computed for the given dataset. Based on these, different test statistics can be computed. (b) The sampling distribution of the test statistic under H_0 is approximated using Monte Carlo simulations under the parametric estimate $f_{\hat{\tau}}$. To this end, M different synthetic datasets are created and for each, test statistics are computed as described in (a). Subsequently, the empirical $(1 - \alpha)$-quantile of their distribution is taken as a critical value for the test decision

Parametrization of RKHS problems

The representer theorem in the theory of RKHS guarantees the existence of $\hat{f}^{(\lambda)}$ within the finite-dimensional space

$$\left\{ h_\alpha \in \mathcal{H} \,|\, h_\alpha := \sum_{i=1}^{n} k\left(\cdot, x_i\right) \alpha_i \text{ with } \alpha_1, \ldots, \alpha_n \in \mathbb{R}^p \right\}, \quad (8)$$

where k is the kernel associated with the RKHS $\mathcal{H}$.[26,27] A finite-dimensional formulation is a prerequisite for solving Equation 4 numerically. Besides the so-called *dual formulation* in Equation 8 of the finite-dimensional optimization problem, two other finite-dimensional formulations can be obtained for special types of kernels (admitting finite-dimensional feature map representations). In this case, the dimension np of the dual formulation can be reduced further through a reparameterization (leading to the so-called *primal* and *mixed* formulations). This technique is described in detail in Section S1 and exploited in the considered simulation study. For ease of readability, in the main text, we refer to the function $\hat{f}^{(\lambda)}$ solving the optimization problem in Equation 4 without specifying its underlying parametrization.

Estimation algorithms

To solve the least squares problem and obtain the parametric estimate $f_{\hat{\tau}}$, we use the Levenberg-Marquardt (LM) algorithm, a robust gradient-based method for solving nonlinear least squares problems.[29,30]

To solve the Tikhonov regularization problem in Equation 4, we propose a three-step algorithm that first solves easier approximate problems, and uses the solution of each step to obtain improved initial guesses for the subsequent step (see pseudocode in Algorithm 1):

1. *ParDir*: determine a <u>par</u>ametric estimate $f_{\hat{\tau}}$ via LM algorithm, then solve the <u>dir</u>ect nonparametric (RKHS) problem analytically (see Section S2) by considering $f_{\hat{\tau}}\left(x_i\right)$ as a surrogate for the unobservable parameters;
2. *AlyLin*: <u>analy</u>tically solve a sequence of <u>lin</u>earized nonparametric problems (see Section S2 for derivation);
3. *Nonlin*: finally, solve the original <u>nonlin</u>ear nonparametric problem in Equation 4 using a quasi-Newton method.

In this way, local minima of Equation 4 can be avoided more successfully. An analysis and benchmark of this algorithm against several alternative approaches is shown for the simulation study, where it clearly outperforms general-purpose optimizers in terms of robustness and runtime.

Algorithm 1: *ParDir-AlyLin-Nonlin for problem (4)*

```
// Step 1: "ParDir"
res := [τ ↦ (G(f_τ(x₁), x₁) − y₁, …, G(f_τ(xₙ), xₙ) − yₙ)]
τ̂ ⟵
    minimize ‖res(τ)‖² using Levenberg-Marquart with initial guess τ₀
Q_dir := [h ↦ ∑_{i=1}^{n} ‖f_τ̂(xᵢ) − h(xᵢ)‖² + λ‖h‖²_ℋ]
h_dir ⟵ minimize Q_dir(h) over ℋ analytically

// Step 2: "AlyLin"
h_lin^(0) ⟵ h_dir
for s = 1 : niter do
    G_lin^(s) ⟵ linearize model G at function h_lin^(s−1)
    Q_lin^(s) := [h ↦ ∑_{i=1}^{n} ‖yᵢ − G_lin^(s)(h(xᵢ), xᵢ)‖² + λ‖h‖²_ℋ]
    h_lin^(s) ⟵ minimize Q_lin^(s)(h) over ℋ analytically
end

// Step 3: "Nonlin"
Q := [h ↦ ∑_{i=1}^{n} ‖yᵢ − G(h(xᵢ), xᵢ)‖² + λ‖h‖²_ℋ]
h_nonlin ⟵ minimize Q(h) using quasi-Newton with initial guess h_lin^(niter)

// Output
f̂^(λ) ⟵ h_nonlin
return f̂^(λ)
```

Finally, a variant of Algorithm 1 allows to compute the combined parametric/nonparametric estimate $\tilde{f}^{(\lambda)}$ solving Equation 6. The pseudocode for this Algorithm 2 is provided in Section S3.3.

Implementation

The proposed estimation algorithms and goodness-of-fit tests were implemented in R software version 3.5.1.[31] For matrix algebra, R package "Matrix" version 1.2–14 was used.[32] The Levenberg-Marquardt algorithm was taken from R package "minpack.lm" version 1.2–1, an interface to the Fortran library MINPACK.[33] The general-purpose optimizers implemented in base R function "optim" were used for the quasi-Newton method (BFGS algorithm) and simulated annealing. The code used during the analysis is publicly available at https://zenodo.org/record/4273796.

Simulation study: Effect of enzyme maturation on drug clearance

Context of the simulation study setup

A functional relationship between body weight and drug disposition parameters, called allometric scaling, is well-established in the pharmacometric literature.[34] In young children (in particular neonates and infants), however, this weight-effect is not sufficient to describe PK data and an additional weight-independent impact of (young) age on drug clearance is accounted for by a *maturation function*.[35,36]

Many different parametric maturation functions have been proposed in the literature.[37] Therefore, goodness-of-fit testing for parametric covariate models is of particular importance in this context.

The setup of our simulation study "Effect of enzyme maturation on drug clearance" was motivated by a meta-analysis which estimated the maturation effect of the monoclonal antibody palivizumab against respiratory syncytial virus infections in young children.[35] We translated their setting to our statistical framework as follows.

Covariates

In the meta-analysis, covariates (post-gestational) age, weight, gender, ethnicity, and presence/absence of chronic lung disease were considered. For ease of presentation, we focused on the covariates (post-natal) age a in years and body weight w in kg (i.e., $x = (a, w)$). We assumed a uniform age distribution between 0 and 20 years and an age-dependent body weight distribution according to an empirical model.[38]

Mechanistic model G

As in the original study, we assumed a two-compartment PK model

$$V_1 \frac{dC_1}{dt} = Q(C_2 - C_1) - CL \cdot C_1, \tag{9}$$

$$V_2 \frac{dC_2}{dt} = Q(C_1 - C_2), \tag{10}$$

with drug concentrations C_1, C_2 in mg/L in the central and peripheral compartments with volumes V_1, V_2 in L, respectively, intercompartmental flow Q in L/day and clearance (CL) in L/day. The initial conditions were $C_1(0) = D_{rel} \cdot w / V_1$ and $C_2(0) = 0$ with an i.v. bolus administration of $D_{rel} = 15$ mg/kg body weight. We also considered a multiple dosing scenario with 30-day dosing intervals. The model can be solved analytically, see Section S3.1.

The model is parametrized in terms of the parameters $\theta = (CL, V_1, Q, V_2)$. In ref. 35 the covariate-to-parameter relationship comprised a maturation part (depending on age a) and an allometric part (depending on weight w):

$$CL = CL_0 \cdot mat(a) \left(\frac{w}{w_{ref}} \right)^{\frac{3}{4}}; \qquad V_1 = V_1^* \left(\frac{w}{w_{ref}} \right);$$

$$Q = Q^* \left(\frac{w}{w_{ref}} \right)^{\frac{3}{4}}; \qquad V_2 = V_2^* \left(\frac{w}{w_{ref}} \right);$$

TABLE 1 Parameters of the covariate model used for simulation[35]

Parameter	Value	[Unit]
α	0.589[a]	–
β	0.133	[1/year]
CL_{max}^*	198	[mL/day]
V_1^*	4090	[mL]
Q^*	879	[mL/day]
V_2^*	2230	[mL]

[a]In the original publication, $\alpha = 0.411$ was reported, inconsistent with the remaining results shown in the article. In contrast, $0.589 = 1 - 0.411$ was consistent with all other results, hence we used this value for the simulation study.

with reference body weight $w_{ref} = 70$ kg. Because allometric scaling in the stated form is widely accepted, we considered it as part of the ODEs defining the mechanistic model G,

$$V_1^* \left(\frac{w}{w_{ref}} \right) \frac{dC_1}{dt} = Q^* \left(\frac{w}{w_{ref}} \right)^{\frac{3}{4}} (C_2 - C_1) - CL^* \left(\frac{w}{w_{ref}} \right)^{\frac{3}{4}} C_1, \tag{11}$$

$$V_2^* \left(\frac{w}{w_{ref}} \right) \frac{dC_2}{dt} = Q^* \left(\frac{w}{w_{ref}} \right)^{\frac{3}{4}} (C_1 - C_2). \tag{12}$$

and considered the weight-normalized parameters $\theta = (CL^*, V_1^*, Q^*, V_2^*)$ as unknown. The observed quantity was

$$G(\theta, x) = (\ln C_1(t_1), \ldots, \ln C_1(t_q))$$

at fixed time points $t_1, \ldots, t_q$. Finally, we assumed normally distributed additive noise (i.e., $(\varepsilon_i) \sim_{iid} \mathcal{N}(0, \sigma^2)$ with $\sigma > 0$ known).

Covariate-to-parameter relationship

To generate the virtual clinical data, a saturable exponential maturation function of age a

$$CL^*(a) = (1 - \alpha e^{-\beta \cdot a}) CL_{max}^*$$

and the resulting covariate model

$$f_\tau(a) = (CL^*(a), V_1^*, Q^*, V_2^*)$$

were used, with parameter values listed in Table 1.

Simulation scenarios

We considered four simulation scenarios, varying in number of individuals, noise level, and sampling times (see Table 2).

A typical prediction with the mechanistic model is shown in Figure 2, along with the sampling times of the four considered scenarios. Three scenarios (rich, sparse, and noisy) contained sampling points from a single dosing interval, while scenario "multi" had sampling times in four dosing intervals. Because clearance is mainly informed by the terminal phase of a dosing cycle, the multiple dosing scenario is expected to allow for more precise estimates compared to the corresponding single dose scenario (noisy). With a reported noise level of $\sigma \approx 0.24$, scenarios "rich" and "sparse" were less noisy, whereas scenarios "noisy" and "multi" were more noisy than the model in ref. 35. Exemplary simulated data for each of the four scenarios are shown in Section S3.2.

Parametric covariate model classes for goodness-of-fit testing

Three different classes of covariate-to-parameter relationships were considered that differed in the parametrization of the CL^* function. All three classes have been reported in the literature.[37]

- Class based on *saturable exponential* CL* functions

$$f_\tau(a) = \left(\left(1 - \alpha e^{-\beta \cdot a} \right) CL^*_{max}, V^*_1, Q^*, V^*_2 \right),$$

with $\tau = \left(\alpha, \beta, CL^*_{max}, V^*_1, Q^*, V^*_2 \right)$. This parametric class allowed us to evaluate the type I error of the goodness-of-fit tests, because it contains the model used to generate the virtual data.

- Class based on *affine linear* CL* functions

$$f_\tau(a) = \left(\alpha + \beta\, a, V^*_1, Q^*, V^*_2 \right),$$

with $\tau = \left(\alpha, \beta, V^*_1, Q^*, V^*_2 \right)$.

- Class based on *Michaelis-Menten* type CL* functions

$$f_\tau(a) = \left(\frac{CL^*_{max} a}{K_M + a}, V^*_1, Q^*, V^*_2 \right),$$

with $\tau = \left(CL^*_{max}, K_M, V^*_1, Q^*, V^*_2 \right)$.

Choice of kernels and regularization parameters

As stated in Equation 8, the nonparametric estimate of the covariate-to-parameter function is of the form

$$f = \sum_{i=1}^n k\left(\cdot, x_i \right) \alpha_i; \quad \alpha_1, \ldots, \alpha_n \in \mathbb{R}^p. \tag{13}$$

TABLE 2 Four scenarios considered in the simulation study "Effect of enzyme maturation on drug clearance", differing in number of individuals n, standard deviation σ of the residual error, and observation time points

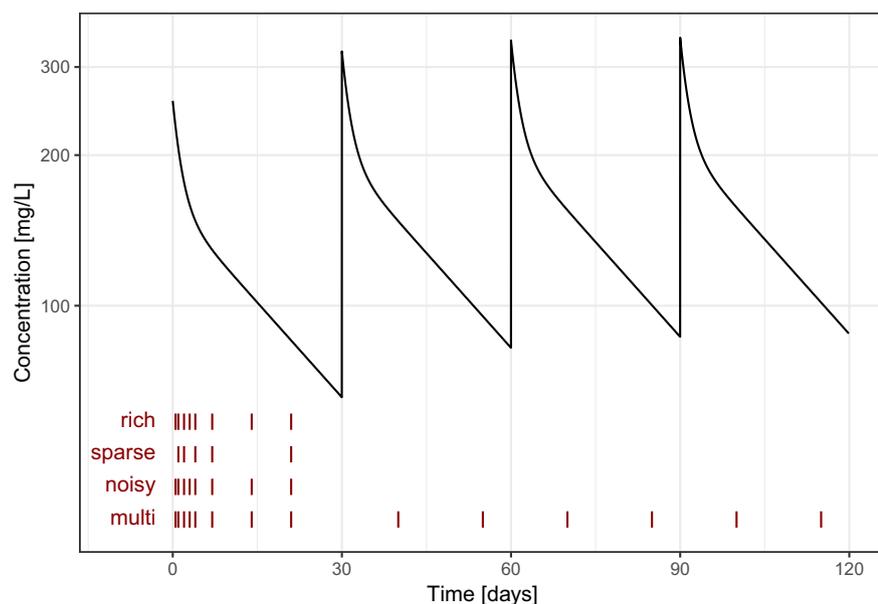| Scenario	n	σ	Observation timepoints (days)
Rich	100	0.1	0.5, 1, 2, 3, 4, 7, 14, 21
Sparse	20	0.1	1, 2, 4, 7, 21
Noisy	100	0.3	0.5, 1, 2, 3, 4, 7, 14, 21
Multi	100	0.3	0.5, 1, 2, 3, 4, 7, 14, 21, 40, 55, 70, 85, 100, 115

FIGURE 2 Typical plasma concentration-time profile for a reference adult (70 kg body weight) in the simulation study "Effect of enzyme maturation on drug clearance", based on pharmacokinetic parameters for Palivizumab from ref. 35. Four 30-day dosing cycles with a dose of 15 mg/kg body weight are simulated. The sampling times of the four considered scenarios (rich, sparse, noisy, and multi) are indicated in red

Because $f(x) = \theta$, the l-th entry of the parameter vector $\theta = \left(CL^*, V_1^*, Q^*, V_2^*\right)$ corresponds to the l-th row of the kernel $k(\cdot, \cdot)$. For the RKHS $\mathcal{H}$ in the nonparametric alternative, an age-dependent diagonal kernel of the form

$$k\left(a, a'\right) = \begin{pmatrix} \exp\left(-\dfrac{(a-a')^2}{2b^2}\right) & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (14)$$

was assumed (allometric scaling, and hence the dependency on weight w, was modeled as part of the mechanistic model G). The first component (for parameter CL*) is a Gaussian kernel with bandwidth parameter $b > 0$, the 1's on the diagonal correspond to constant kernels. Using this kernel structure, the dependency of CL* on age was modeled as a weighted sum of Gaussians (see Equation 13 above), whereas the other three parameters were constant (age-independent), because allometric scaling was part of the ODE model; see Equations 11 and 12.

For the combined parametric/RKHS model, a slightly different kernel was chosen because the age-independent components were already contained in the parametric part, leading to the kernel

$$k\left(a, a'\right) = \begin{pmatrix} \exp\left(-\dfrac{(a-a')^2}{2b^2}\right) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \quad (15)$$

We followed the practice to adapt the regularization parameter λ to a fixed bandwidth b with reasonable interaction range,[39] first choosing $b = 100$ weeks (≈ 2 years, which represents 10% of the simulated age range) as bandwidth and then $\lambda > 0$ in the Tikhonov regularization problems in Equation 4 and 6 by 5-fold cross-validation (see ref. 40, chapter 7.10.1 for details), striking a balance between goodness-of-fit (small λ) and generalizability to new data (larger λ).

RESULTS

Estimated maturation functions based on cross-validated regularization parameters

As a first step toward a nonparametric estimation of the maturation function, the regularization parameter λ was estimated by 5-fold cross-validation. We simulated one dataset per scenario (rich, sparse, noisy, and multi) and determined the cross-validated estimate $\widehat{\lambda}$ for the nonparametric estimator in Equation 4 and the combined parametric/nonparametric estimator in Equation 6 for each of the three parametric classes. The estimates $\widehat{\lambda}$ were insensitive to differences in the data scenarios. In contrast, they strongly depended on the chosen model, with $\widehat{\lambda} = 2 \cdot 10^{-5}$ in the nonparametric class and increasing from Michaelis-Menten ($\widehat{\lambda} = 10^{-5}$), to affine linear ($\widehat{\lambda} = 10^{-3}$), to the saturable exponential parametric class ($\widehat{\lambda} = 10^{-2}$) of the combined parametric/nonparametric model.

FIGURE 3 Relationship between age a and clearance predicted by parametric (black dashed line), nonparametric (green dashed line) and combined parametric/nonparametric (blue dashed line) estimates for one simulated dataset in scenario "rich". Top panel: weight-normalized clearance $CL^*(a)$; bottom panel: clearance $CL\left(a, w_{typ}(a)\right)$ for the typical weight $w_{typ}(a)$ at a certain age a (median weight predicted with the model by ref. 38). Each column corresponds to a different parametric class, left: saturable exponential (containing the true model, red solid line), middle: affine linear, and right: Michaelis-Menten. Grey crosses in top panel: age values in the simulated dataset
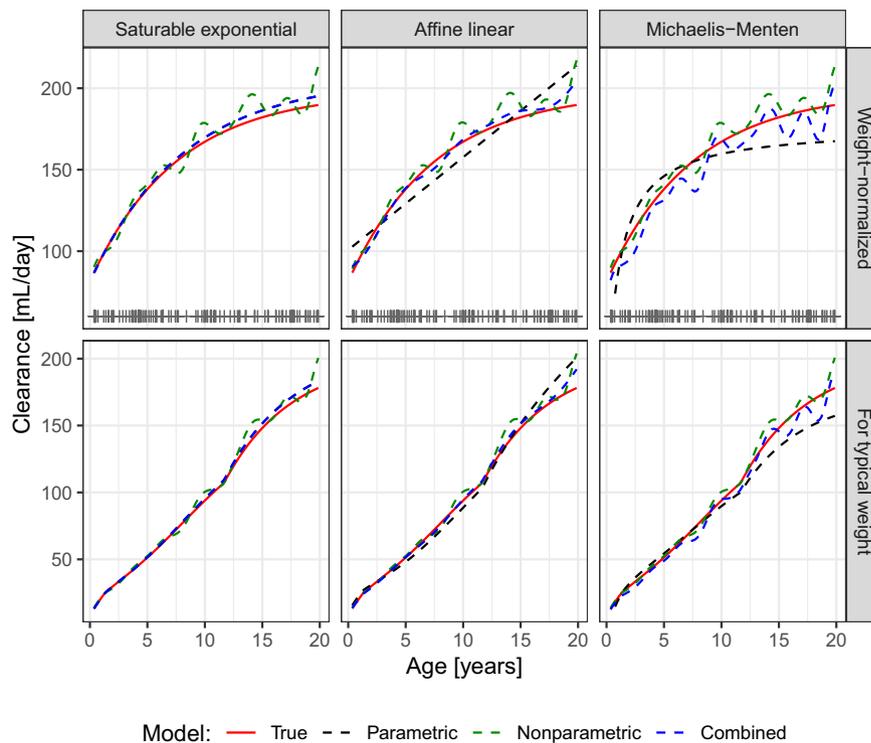
See Figure 3 for an illustration of the estimated covariate-to-parameter relationships in the data-rich scenario. Due to the inverse problem character of the estimation problem (which depends on the sensitivity of G to the parameters θ), oscillations appear in the nonparametric and combined parametric/nonparametric estimates (see also Discussion). The larger penalization parameter in the saturable exponential class, in particular compared to the Michaelis-Menten class, resulted in a much smoother combined estimate $\tilde{f}^{(\lambda)}$.

Efficient nonparametric estimation of the maturation function

From a computational point of view, the proposed goodness-of-fit tests depend crucially on the performance of the numerical algorithms used to determine the observed test statistics and, through the Monte Carlo procedure, the critical values of the test statistics. We therefore simulated 25 independent datasets for each considered scenario (rich, sparse, noisy, and multi) to evaluate convergence and runtime. The mixed RKHS formulation (see Section S1) allowed to substantially reduce the dimensionality of the corresponding estimation problems, from $4n$ to $n + 3$ parameters for the kernel structure in Equation 14 used in Equation 4 and from $4n$ to n for the kernel structure in Equation 15 used in Equation 6. For ease of presentation, here, we concentrate on Algorithm 1 in

the nonparametric problem in Equation 4; Algorithm 2 in the combined parametric/nonparametric problem in Equation 6 is evaluated in Section S3.4.

We benchmarked our proposed Algorithm 1 *ParDir-AlyLin-Nonlin* against two commonly used general-purpose optimizers, namely *quasi-Newton* (gradient-based) and *simulated annealing* (gradient-free). Additionally, we included two variants of Algorithm 1 to further elucidate the impact of the different steps in Algorithm 1: *ParDir-Nonlin* (only steps 1 + 3) and *ParDir-AlyLin* (only steps 1 + 2). The general-purpose optimizers were initialized with lognormally distributed initial conditions around 1, yielding a reasonable initial guess in the absence of more detailed knowledge. For Algorithm 1 and its variants, the parametric class of affine linear models was chosen in step 1 (see Figure 3, middle panel), and the initial guess for its coefficients was lognormally distributed within a plausible order of magnitude. All lognormal distributions had a large coefficient of variation of $\approx 130\%$ (a log-variance of 1).

The benchmark results based on our simulation study are displayed in Figure 4. Estimation with the general-purpose optimizers almost always failed to converge toward parameter values with mean squared errors close to the (expected) variances underlying the simulations, indicating inefficient exploration of the parameter space (*simulated annealing*) or convergence to local minima (*quasi-Newton*). Because *quasi-Newton* corresponds to step 3 of Algorithm 1, the
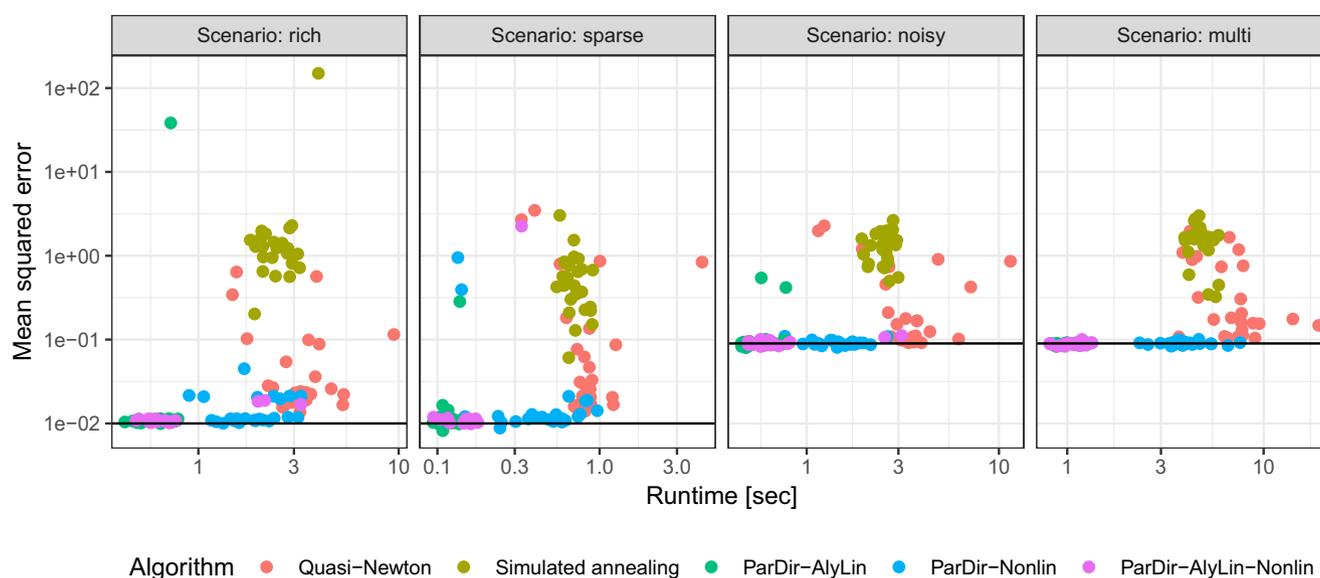

FIGURE 4 Benchmark of estimation algorithms for solving the Tikhonov regularization problem in Equation 4 in the simulation study "Effect of enzyme maturation on drug clearance". We simulated 25 independent datasets for each of the four considered data scenarios (rich, sparse, noisy, and multi) and benchmarked the following estimation algorithms: a quasi-Newton method, simulated annealing, the proposed Algorithm 1 (steps 1–3); and two variants of it: ParDir-AlyLin (only steps 1 and 2) and ParDir-Nonlin (only steps 1 and 3). For the general-purpose optimizers, random positive initial conditions were chosen. For optimizers starting with a parametric step, the parametric class of affine linear models was chosen, and the initial guess for its coefficients was lognormally distributed around values τ_0 having the correct order of magnitude, with a coefficient of variation of $\approx 130\%$. Each dot represents runtime and mean squared error for one of the 25 simulated datasets. The black horizontal line indicates the error variance σ^2 used to generate the data

performance of *ParDir-Nonlin* (steps 1 and 3) illustrates that informing the initial guess of RKHS coefficients through a parametric model largely improved frequency of successful estimations, even though the parametric class qualitatively differed from the class used to generate the data (affine linear vs. saturable exponential). Refining the initial guess (step 1) for the quasi-Newton method further by iteratively solving the linearized RKHS model (step 2) allowed to considerably reduce runtime due to a closed-form solution of the linear inverse problem. In some cases, the two steps in *ParDir-AlyLin* did not suffice to converge to nominal error levels, which showed that the final quasi-Newton step in *ParDir-AlyLin-Nonlin* was necessary. In all four scenarios (rich, sparse, noisy, and multi), the benchmark strongly supported the use of Algorithm 1 within the goodness-of-fit tests.

Goodness-of-fit testing for parametric maturation effect models

For the goodness-of-fit testing problem, datasets were simulated for each of the four scenarios (rich, sparse, noisy, and multi) using the saturable exponential model from ref. 35 as the true covariate-to-parameter relationship. Each of the three parametric classes (saturable exponential, affine linear, and Michaelis-Menten) was considered as a null model, with the saturable exponential class representing a correctly specified model, and the other two a model misspecification. Subsequently, the proposed test statistics T_1 and T_2 were computed using Algorithms 1 and 2, and their distribution under the parametric null hypothesis was approximated using $M = 500$ Monte Carlo samples. To determine the rejection frequency of the null hypothesis in each test, the entire procedure was repeated and averaged over 500 independent datasets.

The results of the goodness-of-fit tests are displayed in Table 3 (see also Section S3.5 for additional test statistics). Both tests approximately maintained the nominal type I error rate $\alpha = 0.05$ in all four data scenarios. From theory, this behavior of the Monte Carlo approximation of the sampling distributions of the test statistics is expected whenever $\hat{\tau}$ is close to the unknown true parameter $\tau^* \in \mathcal{T}$ and under an exact calculation of the test statistic. The results thus highlighted the good performance of the numerical estimation algorithm, with only a slight type I error deflation in the most challenging sparse data scenario.

As expected, the power to detect a misspecified parametric class strongly depended on the considered data scenario. For a rich data scenario, both test statistics had low type II error rates (0.4%–2.4%). In the sparse and noisy data scenarios, type II error rates increased considerably (54.2%–82.3%). In contrast, with a sampling scheme extending over several dosing intervals (scenario "multi"), the goodness-of-fit test was

TABLE 3 Type I and type II errors in goodness-of-fit testing for the simulation study "Effect of enzyme maturation on drug clearance", for test statistics T_1 and T_2 (see Section S3.5 for the complete set of test statistics, including the variants)

	Type I error			
Model class of H_0	**Rich**	**Sparse**	**Noisy**	**Multi**
Saturable exponential				
T_1	5.2%	3.8%	5.2%	4.2%
T_2	5.2%	3.0%	5.6%	4.0%
	Type II error			
Model class of H_0	**Rich**	**Sparse**	**Noisy**	**Multi**
Affine linear				
T_1	2.4%	82.3%	83.8%	7.8%
T_2	0.6%	66.2%	70.2%	2.4%
Michaelis-Menten				
T_1	0.4%	68.4%	54.2%	1.5%
T_2	1.0%	74.0%	68.8%	7.4%

Notes: For each of the simulation scenarios (rich, sparse, noisy, and multi), 500 independent datasets were simulated with an underlying saturable exponential model with parameters from Table 1. For each parametric hypothesis, the test statistics T_1 and T_2 were computed and their distribution under the null approximated with $M = 500$ Monte Carlo samples. A level $\alpha = 0.05$ was taken as a decision threshold (i.e., the empirical 0.05-fractile of the Monte Carlo samples).

again able to detect a model misspecification in presence of noisy data with type II error rates of 1.5%–7.8%. Power also depended on the test statistic and the parametric null model, although to a lesser degree than on the data scenario. Type II error rates for statistic T_2 were lower than for T_1 under the affine linear class, but larger under the Michaelis-Menten class.

DISCUSSION

We demonstrated the usefulness and practical applicability of the proposed goodness-of-fit tests for parametric covariate models in a relevant proof-of-concept study ("Effect of enzyme maturation on drug clearance"). Due to the importance of covariate modeling in pharmacometric analyses, nonparametric goodness-of-fit tests have potential for a wide applicability.

Our simulation study was based on a meta-analysis of 22 separate studies, which featured a very complex study design.[35] Although we made an effort to represent the essence of this meta-analysis, some specific aspects differed. In ref. 35, additionally to the covariates age and weight, the categorical covariates gender, ethnicity, and disease status were considered. The age distribution was nonuniform, with mainly young children and adults, because the disease is in young children (and adults were studied prior to children, as

required by regulations). Moreover, the sampling schedules for young children were sparser than for adults. All of these effects could be integrated into a simulation study, but because they would make the presentation considerably more complex, we opted not to consider them here. Exemplarily, we investigated a modified scenario "rich," with only young children (0–5 years) and young adults (15–20 years), and our goodness-of-fit test retained its properties (Section S3.7).

Moreover, the model in ref. 35 included random effects, which were not considered in our simulation study. Random effect models can be regarded as state-of-the-art in pharmacometric analyses,[1] and hence, this extension of the RKHS framework is of considerable importance. The consideration of random effects, however, requires further extensions both on a theoretical and a numerical level, which were beyond the scope of this paper.

In the main text, we compared the test statistics T_1 and T_2 defined on the observation space. In the supplement, we additionally considered a smoothed version T_1^*, as well as the corresponding statistics S_1, S_1^*, and S_2 on the parameter space (Section S3.5). The test statistics on the observation space consistently resulted in higher power than the respective statistic on the parameter space, for all considered data scenarios and both misspecified parametric classes. The additional smoothing step differentiating T_1^* from T_1 (and S_1^* from S_1) was motivated by a theoretical analysis in the direct problem, where (kernel density) smoothing increased robustness.[13] In our simulation study, such an improvement was seen on the parameter space, S_1^* showing superior power to S_1, but not on the observation space, where T_1 and T_1^* resulted in almost identical test decisions. Computationally, T_1 is preferable to T_1^*, because the smoothing step in the latter contains another Tikhonov regularization problem. A purely parametric fit with the affine linear class resulted in lower mean squared errors compared to the Michaelis-Menten class, hence the affine linear model could be considered as a less severe model misspecification. A comparison of the power of T_1 versus T_2 under these two models indicates that the combined parametric/nonparametric formulation might be beneficial for slight misspecifications, whereas the purely nonparametric formulation could have advantages for more severe misspecifications. Additionally, the more regularized formulation (i.e., with larger cross-validated $\widehat{\lambda}$) led to lower type II error rates in both cases. Further investigations are warranted to elucidate the individual characteristics of these test statistics.

The considered data scenarios covered a range of different situations, "rich" being best-case, "sparse" and "noisy" challenging the algorithms in terms of data informativeness, and "multi" being a more realistic setting. To evaluate the intrinsic difficulty of the sparse and noisy data scenarios, where type II errors were largest, we performed two additional analyses (see Section S3.6). First, the misspecified parametric models were tested against a correctly specified parametric

alternative in a likelihood ratio test, still resulting in large type II errors, even though more information was available than in the nonparametric test, which is agnostic to the true model. Second, a Fisher information-based metric showed larger uncertainties in scenarios sparse/noisy compared to rich/multi. Both analyses indicate that the sparse and noisy data scenarios were particularly challenging, and that type II error rates were not due to the structure of our nonparametric tests.

The numerical algorithms presented have been carefully evaluated based on the simulation study. Their robustness is achieved by approaching the nonlinear high-dimensional optimization problem stepwise through problems of increasing complexity. Other "coarse-graining" heuristics could be envisaged, for example, a first data pooling step for individuals with similar covariates. The advantage of Algorithm 1, however, is that no such preprocessing steps are required.

Compared to purely parametric model selection criteria, our proposed nonparametric goodness-of-fit test requires a larger computational effort, in particular to approximate the sampling distribution of test statistics under H_0. However, Monte Carlo sampling can be parallelized well, and for the investigated scenarios, test decisions could be made within the order of a minute on a desktop computer. For more complex models, runtime could be reduced further, for example, by using a compiled programming language or through another approximation of the sampling distribution.

Unlike commonly used parametric models, our proposed kernel-based estimators allowed for very flexible covariate-to-parameter relationships, including non-monotonic functions. In goodness-of-fit testing, the flexibility of nonparametric estimators is desirable because it allows to capture unexpected effects. To highlight this feature, we investigated an additional data scenario generated from a non-monotonic maturation function, representing an unknown maturation process occurring during puberty. This effect was correctly identified by the nonparametric goodness-of-fit test (see Section S3.8). If a less variable nonparametric estimate of the covariate-to-parameter relationship were desired, regularization parameters could be adapted. Indeed, it is known that optimal rates of testing and estimation may differ, with less regularization required for testing than for estimation.[41]

For our proposed nonparametric goodness-of-fit test, we used concepts from statistical learning, in particular kernel-based regularization techniques in the context of nonlinear statistical inverse problems with random design.[21] The underlying RKHS framework is general and powerful, offering many possibilities for extension. First, different classes of kernel functions allow to describe functions with a different degree of regularity, and hyperparameters like the bandwidth of the Gaussian kernel could be adapted to each data scenario (e.g., via cross-validation).[42] In addition, a nondiagonal kernel structure could be exploited to simultaneously model

parameters with a similar interpretation. Next, regularization could be dealt with differently, for example, through Landweber iterations rather than Tikhonov regularization.[43] The current approach uses a single regularization parameter λ. By using $\max_{\lambda \in \Lambda} T^{(\lambda)}$ over a suitably chosen grid Λ of regularization parameters, the tests could be made adaptive (i.e., avoiding to choose λ)[14]; yet, a successful implementation of such an approach would require a normalization of the different test statistics $T^{(\lambda)}$. Finally, in the vector-valued RKHS setting, it is natural to generalize the regularization schemes by using different regularization parameters for different components of the RKHS function.

In summary, the flexibility of the RKHS framework renders the proposed goodness-of-fit tests very versatile; our approach is envisioned to be beneficial for pharmacological applications that lack well-founded covariate models.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST
The authors declared no competing interests for this work.

AUTHOR CONTRIBUTIONS
N.H., M.W., A.R., and W.H. wrote the manuscript. N.H., M.W., and W.H. designed the research. N.H. performed the research. N.H., M.W., A.R., and W.H. analyzed the data.

ENDNOTE
[a] $\|\cdot\|$ and $\|\cdot\|_{\mathcal{H}}$ denote the Euclidean and the RKHS norm, respectively, for the latter see also Section S1.

REFERENCES
1. Standing JF. Understanding and applying pharmacometric modelling and simulation in clinical practice and research. *Br J Clin Pharmacol*. 2017;83:247-254.
2. Jonsson EN, Karlsson MO. Automated covariate model building within NONMEM. *Pharm Res*. 1998;15:1463-1468.
3. Ribbing J, Nyberg J, Caster O, Jonsson EN. The lasso–a novel method for predictive covariate model building in nonlinear mixed effects models. *J Pharmacokinet Pharmacodyn*. 2007;34:485-517.
4. Hutmacher MM, Kowalski KG. Covariate selection in pharmacometric analyses: a review of methods. *Br J Clin Pharmacol*. 2015;79:132-147.
5. Joerger M. Covariate pharmacokinetic model building in oncology and its potential clinical relevance. *AAPS J*. 2012;14:119-132.
6. Dartois C, Brendel K, Comets E, et al. Overview of model-building strategies in population PK/PD analyses: 2002–2004 literature survey. *Br J Clin Pharmacol*. 2007;64:603-612.
7. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr*. 1974;19:716-723.
8. Schwartz G. Estimating the dimension of a model. *Ann Stat*. 1978;6:461-464.
9. Mandema JW, Verotta D, Sheiner LB. Building population pharmacokinetic–pharmacodynamic models. I. Models for covariate effects. *J Pharmacokinet Biopharm*. 1992;20:511-528.
10. Savic R, Karlsson M. Importance of shrinkage in empirical bayes estimates for diagnostics: problems and solutions. *AAPS J*. 2009;11:558-569.
11. Lai T, Shih M, Wong S. A new approach to modeling covariate effects and individualization in population pharmacokinetics-pharmacodynamics. *J Pharmacokinet Pharmacodyn*. 2006;33:49-74.
12. Mesnil F, Mentre F, Dubruc C, Thenot J, Mallet A. Population pharmacokinetic analysis of mizolastine and validation from sparse data on patients using the nonparametric maximum likelihood method. *J Pharmacokinet Biopharm*. 1998;26:133-161.
13. Härdle W, Mammen E. Comparing nonparametric versus parametric regression fits. *Annal Stat*. 1993;21:1926-1947.
14. Horowitz J, Spokoiny V. An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*. 2001;69:599-631.
15. Hausman JA. Specification tests in econometrics. *Econometrica*. 1978;46:1251-1271.
16. Pilari S, Huisinga W. Lumping of physiologically-based pharmacokinetic models and a mechanistic derivation of classical compartmental models. *J Pharmacokinet Pharmacodyn*. 2010;37:365-405.
17. O'Sullivan F. Convergence characteristics of methods of regularization estimators for nonlinear operator equations. *SIAM J Numer Anal*. 1990;27:1635-1649.
18. Engl H, Hanke M, Neubauer A. *Regularization of inverse problems, vol. 375 of Mathematics and its Applications*, vol 375. Dordrecht: Kluwer Academic Publishers Group; 1996.
19. Caponnetto A, De Vito E. Optimal rates for the regularized least-squares algorithm. *Found Comput Math*. 2007;7:331-368.
20. Smale S, Zhou DX. Learning theory estimates via integral operators and their approximations. *Constr Approx*. 2007;26:153-172.
21. Lu S, Pereverzev SV. *Regularization theory for ill-posed problems, vol. 58 of Inverse and Ill-posed Problems Series*. Berlin: De Gruyter, 2013.
22. Rastogi A, Blanchard G, Mathé P. Convergence analysis of Tikhonov regularization for non-linear statistical inverse problems. *Electron J Stat*. 2020;14:2798-2841.
23. Aronszajn N. Theory of reproducing kernels. *Trans Amer Math Soc*. 1950;68:337-404.
24. Cucker F, Smale S. On the mathematical foundations of learning. *Bull Amer Math Soc*. 2002;39:1-49.
25. Schölkopf B, Smola AJ. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. Adaptive Computation and Machine Learning*. Cambridge, MA, USA: MIT Press; 2002.
26. Micchelli CA, Pontil M. On learning vector-valued functions. *Neural Comput*. 2005;17:177-204.
27. Alvarez M, Rosasco L, Lawrence ND. Kernels for vector-valued functions: a review. *Foundations Trends Mach Learn*. 2012;4:195-266.
28. Evgeniou T, Micchelli C, Poggio M. Learning multiple tasks with kernel methods. *J Mach Learn Res*. 2005;6:615-637.
29. Levenberg K. A method for the solution of certain non-linear problems in least squares. *Quart. Appl. Math*. 1944;2:164-168.

30. Marquardt D. An algorithm for least-squares estimation of nonlinear parameters. *J Soc Indust Appl Math*. 1963;11:431-441.

31. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2018.

32. Bates D, Maechler M. Matrix: Sparse and Dense Matrix Classes and Methods. R package version 1.2-14; 2018.

33. Elzhov T, Mullen K, Spiess AN & Bolker B. minpack.lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds. R package version 1.2-1; 2016.

34. Anderson BJ, Holford NH. Mechanistic basis of using body size and maturation to predict clearance in humans. *Drug Metab Pharmacokinet*. 2009;24:25-36.

35. Robbie GJ, Zhao L, Mondick J, Losonsky G, Roskos LK. Population pharmacokinetics of palivizumab, a humanized anti-respiratory syncytial virus monoclonal antibody, in adults and children. *Antimicrob Agents Chemother*. 2012;56: 4927-4936.

36. Holford N, Heo YA, Anderson B. A pharmacokinetic standard for babies and adults. *J Pharm Sci*. 2013;102:2941-2952.

37. Germovsek E, Barker CI, Sharland M, Standing JF. Scaling clearance in paediatric pharmacokinetics: all models are wrong, which are useful? *Br J Clin Pharmacol*. 2017;83:777-790.

38. Sumpter AL, Holford NH. Predicting weight using postmenstrual age–neonates to adults. *Paediatr Anaesth*. 2011;21: 309-315.

39. Hainmueller J, Hazlett C. Kernel regularized least squares: reducing misspecification bias with a flexible and interpretable machine learning approach. *Political Analysis*. 2014;22:143-168.

40. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. New York: Springer; 2009.

41. Ingster YI, Suslina IA. *Nonparametric goodness-of-fit testing under Gaussian models, vol. 169 of Lecture Notes in Statistics*. New York: Springer-Verlag; 2003.

42. Hamm T, Steinwart I. Adaptive Learning Rates for Support Vector Machines Working on Data with Low Intrinsic Dimension; 2020.

43. Hanke M, Neubauer A, Scherzer O. A convergence analysis of the Landweber iteration for nonlinear ill-posed problems. *Numer Math*. 1995;72:21-37.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.