Siivola, Eero; Weber, Sebastian; Vehtari, Aki

Qualifying drug dosing regimens in pediatrics using Gaussian processes

Published in:
STATISTICS IN MEDICINE

DOI:
10.1002/sim.8907

Published: 10/05/2021

Document Version
Publisher's PDF, also known as Version of record

Published under the following license:
CC BY

Please cite the original version:
Siivola, E., Weber, S., & Vehtari, A. (2021). Qualifying drug dosing regimens in pediatrics using Gaussian processes. STATISTICS IN MEDICINE, 40(10), 2355-2372. https://doi.org/10.1002/sim.8907
Qualifying drug dosing regimens in pediatrics using Gaussian processes

Eero Siivola1,2 | Sebastian Weber1 | Aki Vehtari2

1Advanced Exploratory Analytics, Novartis Pharma AG, Basel, Switzerland
2Department of Computer Science, Aalto University, Espoo, Finland

Correspondence
Eero Siivola, Department of Computer Science, Aalto University, P.O.Box 11000, 00076 Espoo, Finland.
Email: eero.siivola@gmail.com

Funding information
Academy of Finland, Grant/Award Numbers: 298742, 313122; Novartis AG

Drug development commonly studies an adult population first and then the pediatric population. The knowledge from the adult population is taken advantage of for the design of the pediatric trials. Adjusted drug doses for these are often derived from adult pharmacokinetic (PK) models which are extrapolated to patients with smaller body size. This extrapolation is based on scaling physiologic model parameters with a body size measure accounting for organ size differences. The inherent assumption is that children are merely small adults. However, children can be subject to additional effects such as organ maturation. These effects are not present in the adult population and are possibly overlooked at the design stage of the pediatric trials. It is thus crucial to qualify the extrapolation assumptions once the pediatric trial data are available. In this work, we propose a model based on a non-parametric regression method called Gaussian process (GP) to detect deviations from the made extrapolation assumptions. We introduce the theoretical background of this model and compare its performance to a parametric expansion of the adult model. The comparison includes simulations and a clinical study data example. The results show that the GP approach can reliably detect maturation trends from sparse pediatric data.

KEYWORDS
Gaussian processes, organ maturation, pharmacokinetic models

1 INTRODUCTION

Pediatric drug development programs are an essential part of making any drug available for safe and effective use in kids. While in the past new drugs were only developed in the adult population, it is nowadays compulsory for pharmaceutical companies to perform pediatric clinical trials in addition to adult trials for new drugs in the EU1 and the USA.2 As the pediatric trials typically follow the adult trials, the pediatric trials can take advantage of the established knowledge of the adult population.3 Despite more knowledge being available at the start of the pediatric drug development programs, there are substantial challenges with patient recruitment due to ethical and legal constraints. In addition to this, sample collection and clinical endpoint measurement to decide the efficacy and safety of the drug might differ for the pediatric and adult populations. These factors often lead to very sparse pediatric datasets and/or the necessity to use different clinical endpoints for efficacy and safety measurement.
A critical piece of information obtained usually during a phase II adult program is the drug dosing regimen allowing for effective and safe use of the drug. Identifying and confirming the same for the pediatric population is the primary goal of the pediatric program. As the physiology of the pediatric patients differs from adults, their dosing regimen may need to be adjusted. A commonly used key assumption is that the so-called therapeutic window is similar in adult and pediatric patients. The therapeutic window is commonly defined in terms of drug concentration in the systemic blood circulation of a patient. This drug concentration must not exceed levels causing safety issues nor must they fall below levels leading to insufficient efficacy. Hence, the usual goal of a pediatric drug development program is to study drug dosing regimens that generate similar drug concentrations in blood as in adults. The relationship of drug administration and resulting blood concentration over time is described by a pharmacokinetic (PK) model (see, eg, Lin and Lu⁴) that describes the absorption, distribution, metabolism, and excretion (ADME) processes of a drug in the body. Mathematically PK models predict the distribution of the amount of the drug in different parts of the body at a given time as a function of the dosing history (see, eg, Wakefield⁵). The PK model from adults is used to design the pediatric trials. As pediatric patients have significantly smaller body sizes than adults, the existing PK model is extrapolated to very small body sizes to select the adequate doses for the pediatric trials.

One approach to adapt PK models to smaller body sizes is allometric scaling.⁶ The basic principle is to model the key PK parameters such as clearance and blood volume as a function of some body size metric. Commonly used metrics for body size are body weight, body surface area, or related measures. When the dosing regimen of a pediatric trial is decided, kids are essentially treated as small adults, only taking into account their smaller size. However, since very young pediatric patients can have immature organs, allometric scaling may not suffice to account for the implied physiological effect. Because of this, additional organ maturation may need to be accounted for. Hence, at the design stage, such an effect can be overlooked, since no pediatric data have yet been collected. In some cases, detailed knowledge on the drug may suggest the presence of certain maturation effects which can be accounted for based on knowledge from the literature. In any case, only the actual data from the pediatric trial will allow detecting the presence of any actual organ maturation. After the pediatric trial data has been collected, the used dosing regimen needs to be verified by comparing the PK model predictions to the actual measurements.⁷,⁸

The influence of immature organs is commonly taken into account by adding one or more maturation functions to the PK model. These maturation functions are parametric functions of age that multiply the existing model parameters and approach unity the closer to adulthood the child gets. They can be understood to represent the relative capacity of a child organ in relation to the full capacity of an adult organ. While the parametric functions in use often work in practice, they are not derived from first principles as opposed to the rest of the PK model. A sigmoid-shaped Hill model is the most popular maturation function,⁷,⁸ but other approaches exist as well.⁹,¹⁰ As no single parametric function is sufficiently general, a model selection procedure should be used to select the final model from a set of alternatives.

Due to ethical, legal, and logistic challenges, collecting pediatric data are hard and the resulting datasets are very sparse.¹¹ The data are usually not only sparse in the number of patients but also in the number of observations per patient due to limiting the burden and pain caused by, for example, taking a venous blood sample.¹² In addition to this, pediatric data are often considered to be more heterogeneous than adult data due to the mix of a potential maturation effect, vastly different sizes, and individual random variation.¹¹ These factors in the pediatric data lead to substantial aleatoric and epistemic uncertainty which makes it hard to verify that the assumptions about the maturation function are correct. For defining the pediatric dosing regimen, it is hence critical to account for the uncertainty sources adequately, for example, by using Bayesian approaches or rely on bootstrapping. However, model prediction and uncertainty estimates are only reliable if the model assumptions are reasonable (including any parametric functions).

In this article, we consider the situation where the sparse pediatric trial data are collected and a potential maturation effect in the PK model must be detected from an adjusted adult model used to design the pediatric trial. In a way, this can be understood as a model misspecification problem. Our contributions are:

- We propose the use of the non-parametric Gaussian processes (GPs)¹³ to detect the model misspecification and to model the maturation effect. Our approach does not introduce bias to the model and avoids the need for model selection. The key question to address is if the non-parametric GP approach is suitable for the sparse data as collected in pediatric trials. Often non-parametric approaches tend to require a significant amount of data and it is critical to assess the performance of the GP approach using commonly used pediatric trial designs.
- We propose the use of non-observed data points, the so-called virtual observations, to encode prior information about the monotonicity of the maturation effect. The goal of this is to use the existing but sparse data more efficiently and to reduce posterior uncertainty.
• We compare the proposed non-parametric method to a parametric maturation function approach. The comparison is done in a simulation study and a real data example. The simulation study consists of four simulation studies with realistic and sparse data designs. Different simulated maturation trends are explored (no maturation, relevant maturation, and bi-phasic maturation). At last, a case study from a real pediatric study is evaluated.

This article is structured as follows. Section 2 describes pharmacokinetic modeling as used for sparse pediatric data and connects it to drug responses and dosing. Section 3 describes the proposed approach for modeling maturation effects. Section 4 shows the simulated case studies that show the advantages and limitations of the proposed approach. In Section 5, we compare the proposed approach to existing parametric approaches on real pediatric data. The article concludes with a discussion in Section 6.

2 PHARMACOKINETIC MODELING IN SPARSE DATA CONTEXTS

Pharmacokinetic (PK) models are nonlinear mixed-effect models that aim to describe the ADME processes of a drug. The aim is to predict the concentration of a given drug in the blood as a function of time for an individual patient. The prediction is based on the dosing history and details of the patient, like body weight. As each patient differs from one another, most of the parameters in a PK model are specific to the patient and vary around typical parameter values. Thus, most of the parameters are modeled in a hierarchical approach, where each patient have their own set of parameter values, which are modeled as normally distributed random effects centered at the typical population estimates. The typical population estimates are modeled as a function of the patient’s individual traits such as body weight. PK models are trained with data that consist of patients, their dosing history and drug concentration measurements in blood. Kallen provides a good introduction on the theory and derivations of the most used PK models and their efficient computation.

This work concentrates on sparse data designs as they are common in pediatric trials. Pediatric trials usually have a very limited sampling design of the blood concentrations. This leads to restrictions as to what can be inferred from the data. Due practical issues, blood concentrations are usually only measured just before the next dose is given. In this case, it is only possible to learn how patients reach and stay at the steady-state conditions when subject to continuous and regular dosing administration schemes. As no detailed information on the absorption process is measured by such a design, it is possible to only consider a very simple one-compartment model. In this work, we concentrate on a 1st order absorption process corresponding to an oral route of administration. Physiologically this model assumes that the drug is administrated through mouth, from which it transfers through stomach to the central blood circulation and is then eliminated. Assuming that patient \( i \in [1, \ldots, N] \) has a dosing history \( \{D_j, t_j\}_{j=1}^{d_i} \) for doses \( D_j \) at times \( t_j \), this model can be described with the following first-order process:

\[
\begin{align*}
\frac{da_1^{(i)}(t)}{dt} &= -k_u^{(i)}a_1^{(i)}(t), \quad a_1^{(i)}(t) = \lim_{t \to t_j^{-}} a_1^{(i)}(t) + D_j \quad \forall j \in 1 \ldots d_i, \\
\frac{da_2^{(i)}(t)}{dt} &= k_u^{(i)}a_1^{(i)}(t) - \frac{CL^{(i)}}{V^{(i)}}a_2^{(i)}(t), \quad a_1^{(i)}(0) = 0, \quad \text{and} \quad a_2^{(i)}(0) = 0,
\end{align*}
\]

where \( a_1^{(i)}(t) \) is the mass (mg) of the drug in absorption compartment (stomach), \( a_2^{(i)}(t) \) is the mass (mg) of the drug in central compartment (blood circulation), \( k_u^{(i)} \) is the absorption rate constant (1/h), \( CL^{(i)} \) is the clearance (l/h), and \( V^{(i)} \) is the apparent volume of blood (l). This system of differential equations has an analytic solution for an arbitrary dosing regimen.

Some drugs violate the assumption of a first-order absorption process. One such example is Everolimus, which is used in the real clinical data example of Section 5. To avoid the use of more complex models, which lack an analytic solution, an alternative approach is to use a relative bioavailability. Given a dose \( D_j \), relative bioavailability is modeled as

\[
F(D_j) = \left( \frac{D_j}{D_{ref}} \right)^f.
\]

From modeling perspective, relative bioavailability scales the administered doses and non-zero \( f \) is used to add non-linearity to the absorption process.
The dynamics system in Equation (1) contains patient dependent clearance, apparent volume and absorption parameters. The clearance for an individual $i$ is modeled as a nonlinear mixed-effect model

$$\log(C_{l(i)}(t)) = \log(C_{l,ref}) + \log(A_{Cl}(\text{weight}(i))) + \log(M_{Cl}(\text{age}(i))) + \eta_{Cl}^{(i)}, \quad \eta_{Cl}^{(i)} \sim N(0, \omega_{Cl}^2),$$  

where the reference clearance $C_{l,ref}$ is the population average value, the allometric scaling function $A_{Cl}(\text{weight}(i))$ depends on body weight, the maturation function $M_{Cl}(\text{age}(i))$ varies with age, and random effect $\eta_{Cl}$ accounts for the heterogeneity between individuals not explained by other factors. The age used in this manuscript always refers to the post-menstrual age, which on average is 42 weeks at birth. The apparent volume of an individual is modeled with a similar function, although in this work we do not use the maturation function, since we are mainly interested in the clearance which is relevant for characterizing steady-state conditions. The absorption rate also varies between the individuals, but the between individual heterogeneity is not identifiable from a common pediatric study design. Because of this, this work models the absorption rate only using the reference value, or population average. The described 1-compartment oral PK model is illustrated in Figure 1.

In pediatric drug trials, the drugs are usually administrated at a fixed frequency and with a constant dose. This means that the doses are given at equidistant times, at interval $\tau$, such that a steady-state is attained after a number of doses. The steady-state is characterized formally as $a_2(t+\tau) = a_2(t)$ for any $t$ after the steady-state has been achieved. The average concentration at steady-state is fully determined by the clearance (and relative bioavailability) of the drug. Because of this, we wish to learn the values of the clearance as accurately as possible. As the emphasis of this article is in modeling the maturation effect, it is important to understand how the clearance differs between the pediatric and the adult populations. The potential deviation observed in the clearance parameters will be considered as a maturation effect. Should a significant maturation effect be detected from the pediatric data, the pediatric dosing regimen may need to be adjusted by using the detected maturation effect as a scaling factor for the dose.

The parameters of a PK model are learned from data that are obtained from clinical trials. These data consist of details of patients, their dosing histories and drug concentration measurements. Since the measurements $y_{ij}$ are drug concentrations $c_{ij}$ rather than drug amounts, the solution to Equation (1) must be rewritten in terms of the drug concentration $c_{ij}(t)$. This can be done by using apparent volume $V_{ij}$ as a divisor. The error of a concentrations measurement $y_{ij}$ is modeled on the log scale as,

$$\log(y_{ij}) \sim N\left(\log(c_{ij}(t_j)), \sigma_p^2 + \left(\frac{\sigma_c}{c_{ij}(t_j)}\right)^2\right),$$

where $\log(y_{ij})$ is the noisy log concentration measurement of patient $i$ at time-point $t_j$, $c_{ij}(t_j)$ is the corresponding concentration, $\sigma_p^2$ is the constant part of the noise, and $\sigma_c^2$ is the noise term proportional to the predicted concentration.

In this article, we model the problem using the Bayesian framework. Weakly informative priors are applied for each parameter when fitting the adult dataset and the resulting (adult) posterior distributions are used as informative priors for the pediatric dataset. The details of these are presented with the respective examples below.

### 3 | Modeling Maturation with Gaussian Processes

In this section, we first present how a GP can be used to model a maturation function before showing how prior assumptions about the maturation function can be incorporated into a GP. Finally, we describe how the performances of different models will be compared.
3.1 Model specification

GPs are a powerful non-parametric regression method dating back to the 1940s\cite{15} and were first used in geostatistics,\cite{16} where they are better known as Kriging. The modern Bayesian interpretation of the GP framework was proposed in the 1970s.\cite{17,18} Mathematically, GPs are a class of stochastic processes which define prior distributions over unobservable latent functions. GPs provide a flexible framework for solving regression and classification problems and they have become a popular go-to-method in many disciplines. A comprehensive description of GPs is available from Rasmussen and Williams,\cite{13} which provides a good introduction for the interested reader. Key advantages of Bayesian GPs relevant to the analysis at pediatric trials include: (a) they avoid bias due to their non-parametric formulation, (b) they allow for uncertainty in their learned functional shape, and most importantly, (c) they enable one to include domain knowledge about the target function. To our knowledge, GPs have not yet been used in the context of pediatric PK models.

We use GPs to model the maturation function and the inter-patient variation of the logarithmic maturation function values which depend on age. Mathematically, a function follows a GP distribution if the values of that function follow a multivariate Gaussian distribution across the target function. To our knowledge, GPs have not yet been used in the context of pediatric PK models.

Let us assume \( N \) estimates of the maturation function \( \{ \log(\text{CL}_{\text{age}}^{(i)}) \} \) such that \( \mathbf{g} = [ \log(\text{CL}_{\text{age}}^{(1)}), \ldots, \log(\text{CL}_{\text{age}}^{(N)})]^{\top} \) at \( \mathbf{A} = [\text{age}^{(1)}, \ldots, \text{age}^{(N)}]^{\top} \) such that \( \mathbf{K}_{ij} = \text{cov}(\log(\text{CL}_{\text{age}}^{(i)}), \log(\text{CL}_{\text{age}}^{(j)})) = k(\text{age}^{(i)}, \text{age}^{(j)}) \). To connect the estimates \( \mathbf{m}_{\text{CL}} = [\hat{m}_{\text{CL}}^{(1)}, \ldots, \hat{m}_{\text{CL}}^{(N)}]^{\top} \) to the true function values \( \mathbf{g} \), we assume that given \( \mathbf{g} \), estimates are unbiased and normally distributed with variance \( \omega_{\text{CL}}^{2} \) and are conditionally independent of other estimates,

\[
p(\mathbf{m}_{\text{CL}} \mid \mathbf{g}, \omega_{\text{CL}}^{2}) = \prod_{i=1}^{N} N(\log(\text{CL}_{\text{age}}^{(i)}), 0, \omega_{\text{CL}}^{2}).
\]

We use the Bayes’ theorem to derive the predictive distribution for \( \mathbf{N} \) logarithmic maturation function values \( \mathbf{g}_{\ast} = [\log(\text{CL}_{\ast}^{(1)}), \ldots, \log(\text{CL}_{\ast}^{(N)})]^{\top} \) for previously unseen patients with ages \( \mathbf{A}_{\ast} = [\text{age}_{\ast}^{(1)}, \ldots, \text{age}_{\ast}^{(N)}]^{\top} \),

\[
p(\mathbf{g}_{\ast} \mid \mathbf{m}_{\text{CL}}, \mathbf{A}_{\ast}) = \frac{\int \int p(\mathbf{g}, \mathbf{g}_{\ast} \mid \mathbf{A}, \mathbf{A}_{\ast}) \prod_{i=1}^{N} N(\log(\text{CL}_{\ast}^{(i)}), 0, \omega_{\text{CL}}^{2})d\mathbf{g}}{\int \int p(\mathbf{g}, \mathbf{g}_{\ast} \mid \mathbf{A}, \mathbf{A}_{\ast}) \prod_{i=1}^{N} N(\log(\text{CL}_{\ast}^{(i)}), 0, \omega_{\text{CL}}^{2})d\mathbf{g}}.
\]

where

\[
p\left(\mathbf{g}, \mathbf{g}_{\ast} \mid \mathbf{A}, \mathbf{A}_{\ast}\right) = \mathcal{N}\left(\begin{bmatrix} \mathbf{g} \\ \mathbf{g}_{\ast} \end{bmatrix} \mid \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathbf{K} & \mathbf{K}_{\ast}^{\top} \\ \mathbf{K}_{\ast} & \mathbf{K}_{\ast}\mathbf{K}_{\ast}^{\top} \end{bmatrix}\right).
\]

Here the covariance between the latent values at the training and test inputs is \( \text{cov}(\mathbf{g}, \mathbf{g}_{\ast}) = \mathbf{K}_{\ast} \) and the covariance matrix of the latent values at the test inputs is \( \text{cov}(\mathbf{g}_{\ast}, \mathbf{g}_{\ast}) = \mathbf{K}_{\ast}\mathbf{K}_{\ast}^{\top} \).

As all distributions in Equation (6) are Gaussian, the predictive distribution has an analytic form

\[
\mathbf{g}_{\ast} \mid \hat{\mathbf{m}}_{\text{CL}} \sim \mathcal{N}(\mu_{\ast}, \Sigma_{\ast}), \quad \text{where} \quad \mu_{\ast} = \mathbf{K}^{-1} \hat{\mathbf{m}}_{\text{CL}}, \quad \Sigma_{\ast} = \mathbf{K}_{\ast} - \mathbf{K}_{\ast}\mathbf{K}_{\ast}^{\top},
\]
and thus the predictive distribution of the logarithmic maturation function can be written explicitly for any age.

Each element of the covariance matrices $K_\mathbf{I}$ and $K_\mathbf{II}$ is computed with a positive definite covariance function, which is also known as kernel function. As the chosen kernel function encodes the assumed covariance between the latent maturation function values, it is important to choose this so that it is aligned with prior assumptions. Different kernels can be used to capture different prior assumptions about the function smoothness or shape. See Duvenaud$^{19}$ for the most commonly used kernels. The kernel we will use is the product of a linear and a squared exponential kernel:

$$
K_{ij} = k_{\text{lin}}(\text{age}(i), \text{age}(j)) \times k_{\text{exp}}(\text{age}(i), \text{age}(j))
$$

$$
= \left( \rho_b + \rho_v(\text{age}(i) - c)(\text{age}(j) - c) \right) \times \left( \omega^2 \exp \left( \frac{(\text{age}(i) - \text{age}(j))^2}{2\ell^2} \right) \right). \tag{8}
$$

The squared exponential kernel determines the overall smoothness, that is, how much the logarithmic maturation function is allowed to change with the age. The linear kernel acts as a multiplier: it allows the maturation function to change more quickly with age for younger patients, which may be appropriate because we assume younger patients to mature at a faster rate than older patients. Within the linear kernel $\rho_b$ determines the covariance of at input $c$, while $\rho_v$ defines the overall variability of the linear part. For the squared exponential kernel, $\omega$ defines the average distance of the function from the mean and length-scale $\ell$ defines how smooth the function is. To make the product of these two kernels identifiable, either $\rho_v/\rho_b$ or $\omega$ has to be fixed.

### 3.2 Including domain knowledge in the Gaussian process model to handle sparsity

As pediatric trials generate typically very sparse data, relevant leveraging prior information can improve model fitting. Even though the exact parametric form of the logarithmic maturation function is unknown, prior physiological knowledge can be incorporated into the GP model.

The logarithmic maturation effect, $\log \left( M_{Cl}(\text{age}) \right)$, should be 0 for mature children. This is because once pediatric patients reach a certain age, they will have fully working organs just like adults. Moreover, it is reasonable to assume that the maturation function is smooth when transitioning to the adult model. This can be guaranteed by assuming that the derivative of the maturation function is 0 for mature children. The term “mature child” is with depend on the drug under investigation, since the maturation curves of different drugs can be very different. It is also reasonable to assume that the younger the pediatric patient is, the greater the maturation effect is such that the logarithmic maturation function will take smaller values. Consequently, the logarithmic maturation function can be assumed to increase monotonically. If a function is monotonically increasing within interval, its derivative values are greater than or equal to zero at all points within the interval.

These assumptions can be incorporated into the GP model by encoding them as “hypothetical data.” This approach is also referred to as “virtual observations” in the GP literature.$^{20}$ For instance, to guarantee guaranteeing continuity between an adult regime where no maturation occurs and a regime where maturation can occur, we can add an unobserved logarithmic maturation function value of 0 for an age that corresponds to a mature patient (e.g., 18 years of age). Ensuring constraints on the functional shape of the logarithmic function, such as monotonicity, is also possible by adding unobserved derivative observations of the logarithmic maturation function to the GP. Since differentiation is a linear operator, the partial derivative of a (mean-square differentiable) GP remains a Gaussian process.$^{21}$ As covariance functions are bilinear mappings with respect to their derivative,

$$
\text{cov} \left( \frac{\partial g(i)}{\partial \text{age}(i)}, g(j) \right) = \frac{\partial}{\partial \text{age}(i)} \text{cov} \left( g(i), g(j) \right), \tag{9}
$$

$$
\text{cov} \left( \frac{\partial g(i)}{\partial \text{age}(i)}, \frac{\partial g(j)}{\partial \text{age}(j)} \right) = \frac{\partial^2}{\partial \text{age}(i) \partial \text{age}(j)} \text{cov} \left( g(i), g(j) \right), \tag{10}
$$

the covariance matrices in Equations (4) and (7) can be extended to include partial derivatives either as observations or as values to be predicted.

In addition to the covariance function, to completely define the unobserved data we need an observation model to connect the assumption to the latent maturation function. For unobserved data on the function itself, it is natural to stipulate a normal likelihood similar to the one assumed in Equation (5) for the actual observations. However, for the
monotonicity information, the connection is not as straightforward. Following Riihimäki and Vehtari,\textsuperscript{20} we denote by \( g_s \in \{-1, 1\} \) the partial derivative sign value of \( g \) at age “age.” Then, the probability of observing a positive or a negative derivative value is assumed to follow a probit likelihood with a control parameter \( \nu \)

\[
p \left( g_s \mid \frac{\partial g}{\partial \text{age}} \right) = \Phi \left( \frac{\partial g}{\partial \text{age}} \frac{g_s}{\nu} \right), \quad \text{where } \Phi(z) = \int_{-\infty}^{z} N(t \mid 0, 1)dt. \quad (11)
\]

In practice, the probit likelihood, \( \Phi(\cdot) \), is often used in classification tasks and results in high likelihoods for latent derivatives of same sign. The control parameter can be used to adjust the steepness of the likelihood function which helps to ensure robustness to erroneous monotonicity information. To encode monotonicity, virtual derivative observations have to be placed in a dense enough grid to force the function to be monotonic across some interval.\textsuperscript{20} Just how many are needed will depend on the assumed length scale of the latent logarithmic maturation function.

The unobserved data representing prior information are added as additional terms to the GP model likelihood as follows. Let us assume that we have \( N \) log maturation observations \( \mathbf{m}_{\text{cl}} \) at \( \mathbf{A} = [\text{age}^{(1)}, \ldots, \text{age}^{(N)}]^T \), \( N_s \) virtual log maturation observations \( \mathbf{m}_s \) at \( \mathbf{A}_s \), \( N_d \) virtual log derivative observations \( \mathbf{g} \) at \( \mathbf{A}_d \) and \( N_s \) virtual log derivative sign observations \( \mathbf{g}_s \) at \( \mathbf{A}_s \). Furthermore, assume normal likelihoods with variances \( \omega_{\text{cl}}^2, \omega_{s}^2, \omega_d^2 \) for estimates, hypothetical estimates, and observations on derivative values, respectively, and a probit likelihood for the derivative sign observations. Then the posterior predictive distribution \( p(\mathbf{g} \mid \mathbf{A}_s, \mathbf{A}_d, \mathbf{m}_{\text{cl}}, \mathbf{m}_s) \) of the latent values \( g \) at ages \( \mathbf{A}_s \) is given by

\[
Z^{-1} \prod_{i=1}^{N_s} \prod_{j=1}^{N} \prod_{l=1}^{N_d} \prod_{m=1}^{N_s} \Phi \left( \frac{\partial g^l}{\partial \text{age}} \frac{g_s^m}{\nu} \right) \int_{-\infty}^{\infty} N(t \mid 0, 1)dt. \quad (12)
\]

with normalizing constant \( Z, \mathbf{g}_s = [\mathbf{g}_s^0, \frac{\partial g_s^0}{\partial \text{age}}, \frac{\partial g_s^0}{\partial \text{age}}]^T \) and \( \mathbf{A}_s = [\mathbf{A}_s, \mathbf{A}_d, \mathbf{A}_d]^T \). \( p(\mathbf{g} \mid \mathbf{A}_s, \mathbf{A}_d, \mathbf{m}_{\text{cl}}, \mathbf{m}_s) \) is similar to Equation (7), but the cross covariances between the latent function and the latent function derivatives (eg, \( \text{cov} \left( g_s, \frac{\partial g}{\partial \mathbf{A}_s} \right) \)) can be computed using Equation (9) and the cross covariance between two latent function derivatives (eg, \( \text{cov} \left( \frac{\partial g_s}{\partial \mathbf{A}_s}, \frac{\partial g_s}{\partial \mathbf{A}_d} \right) \)) can be computed with Equation (10). The partial derivatives of the covariance function needed for this can be computed using the product rule. The needed partial derivatives for the squared exponential can be found from Riihimäki and Vehtari,\textsuperscript{20} but the derivatives of the linear kernel are omitted for simplicity. Since the likelihood of the derivative sign observations (Equation (11)) is not Gaussian, the posterior (Equation (12)) is not analytically tractable, but can be approximated using random draws. As the posterior distribution of the PK model is intractable, the fact that the logarithmic maturation function is intractable does not increase the computational burden of fitting the model.

In addition to these virtual observations, we also know that younger patients mature faster than older patients. This makes it natural to use the logarithm of age as an input for GPs. As logarithmic function grows faster for smaller values, we can use the logarithm to stretch the relative differences between young patients. This way the GP model can have relatively larger variability for younger patients. In practice, we replace \( \mathbf{A} \) in Equation (4) by \( \log(\mathbf{A}) = [\log(\text{age}^{(1)}), \ldots, \log(\text{age}^{(N)})]^T \). This allows for larger changes in maturation for smaller kids without making the function too non-smooth for old adolescents. Since the logarithm of age is used, ages close to birth (or premature births) would cause numerical issues. Instead we use the logarithm of the postmenstrual age (pma), which is the time from the first day of the last menstrual period the mother had before giving birth. This can be approximated by adding 0.7 years to the patients age.

### 3.3 Model comparison

To quantitatively compare the proposed non-parametric model with other popular parametric alternatives, we score models based on their ability to predict data from previously unseen patients. We use the fully Bayesian measure of leave-one-patient-out expected log predictive density (loo-elpd). This measure captures the predictive power of the model as it reflects how well the model is able to predict the drug concentration of unseen pediatric patients. This choice deviates from the oftentimes used Bayesian information criteria (BIC)\textsuperscript{22} for model selection when using frequentist inference.
The reasons is that BIC relies on point estimates and as such does not fully account for the uncertainty represented by the posterior distribution. Nonetheless, BIC targets the approximation of the Bayes factor which is also the case for loo-elpd. Hence both metrics target the same quantity of interest, but in a fully Bayesian framework the loo-elpd is expected to more accurately approximate the Bayes factor. For details on leave-one-patient-out cross validation, please refer to Vehtari et al. 23

The loo-elpd measure is defined as

$$\text{loo-elpd} = \sum_{i=1}^{N} \log \left( \int p(y_i \mid \Theta)p(\Theta \mid y_{-i})d\Theta \right), \quad (13)$$

where $p(\Theta \mid y_{-i})$ is the posterior distribution of the model parameters $\Theta$ given measurements of all other patients but $i$ and $p(y_i \mid \Theta)$ is the posterior predictive distribution of an unseen patient $i$.

Another way to compare models is to see how strongly individual clearance estimates $\eta_{\text{Cl}}^{(i)}$ are shrunk toward the population mean. This can be done by computing a pooling factor $\lambda$ which is the Bayesian equivalent of the shrinkage estimate for Frequentist models. 25,26 The key difference between a pooling factor and the shrinkage estimate is that the former is based on the entire posterior distribution while the latter uses empirical point estimates of the random effects. The pooling factor varies between 0 and 1 and is defined as

$$\lambda = 1 - \frac{\text{Var}\left( E\left( \eta_{\text{Cl}}^{(i)} \right) \right)}{E\left( \text{Var}\left( \eta_{\text{Cl}}^{(i)} \right) \right)} \quad (14)$$

Values close to 1 indicate that the individual estimates are shrunk toward the population mean, or that the population mean explains a lot of the variability of the individual values. In contrast, values close to 0 indicate that the individual estimates are further away from the population mean. In the context of maturation models, if the maturation model is wrong and the pooling factor is high, the estimated individual estimates will be biased. On the other hand, if the pooling factor is high, this can also be an indication of overlearning for the non-parametric models like GPs.

### 4 | SIMULATION STUDIES

In this section, we present the results of four simulation studies conducted to compare the performance of the proposed approach with and without additional prior information about monotonicity with that of the most commonly used parametric modeling approaches. First we give implementation details, before going through three different simulated data scenarios which were each represented by 50 random dataset. Finally, we show how the GP model behaves under different levels of data sparsity.

#### 4.1 | Implementation details

The purpose of the simulation studies is to mimic a realistic pediatric trial design when the adult PK model is accurately known. The scenario is that the pediatric trial is planned under the assumption that the adult PK model is correct also for the pediatric patients. This means that maturation effects are assumed to be absent in the pediatric population. Hence, the dosing regimen of the pediatric trial is planned with dosing adjusted only for the size of the patients (with allometric scaling), but without taking the age of the patients into account (with maturation effects). In contrast, the pediatric data are simulated from three different models which deviate from this assumption. By analyzing the simulated pediatric data, we aim to understand how well the proposed method works in realistic scenarios where the ground truth, the model used to generate the pediatric data, is known.

In all three cases, the data are generated using a one-compartment model with oral administration. Allometric scaling

$$A(\text{weight}) = \left( \frac{\text{weight}}{75\text{kg}} \right)^{\alpha} \quad (15)$$
is assumed for both clearance and volume. Three alternatives for the maturation of clearance are considered in the different examples: Logarithm of Hill function (Section 4.2), no maturation (Section 4.3), and logarithm of Double Hill (Section 4.4). Absorption is assumed to be constant for the whole population in all examples. The exact model parameters can be found in Table 1.

In each simulation study (Sections 4.2 to 4.4), we apply three different approaches to model a potential maturation effect. The first maturation model is a Hill function,

$$\log(M_{\text{Cl}}(\text{age})) = \log \left( \frac{\text{age}^h}{\text{MA50}^h + \text{age}^h} \right),$$

(16)

which is a popular choice for the maturation model (henceforth referred to as the Hill model). The second model is a GP, which is implemented as described in Section 3 without including any virtual derivative sign observations to ensure monotonicity (subsequently referred to as the naive GP model). The third model is a GP with 8 virtual derivative sign observations placed log linearly between 0 and 3 years, and one virtual observation for maturation and its derivative for a 5-year-old child (labeled as the GP model from this on). The parameter values of all three models (Hill model, naive GP model and GP model) are described in Table 2. Priors for the rest of the model parameters can be found from Table 1.

In all simulation studies, the data are generated with 20 patients, with linearly spaced ages between 0 and 5 years, and weights taken from a population curve without additional noise. This choice for ages mimics an idealized situation, where the pediatric data are evenly spread across the pediatric age range of the drug of interest. Each patient is treated with a dose regimen tailored to their body weight. The dosing regimen for a 75 kg reference patient includes first a loading dose of 1500 mg followed by 10 more doses administered daily every 24 hours after the previous dose. Measurements are always taken before a new dose is given, excluding the first dose. Each patients dose is determined using allometric scaling based on body weight with the same scaling factor as is used for the clearance parameter ($dose_i = dose_{\text{ref}} \left( \frac{\text{weight}_i}{75 \text{ kg}} \right)^a$). The simulation models and data scenario are chosen so as to be comparable to real clinical data situations. As the simulation model includes random parameters, and as the drug concentration measurements are assumed to be stochastic, we simulate 50 trials for each scenario, to better take into account the aleatoric uncertainty. Simulations are performed using Stan 2.18.2 and ggplot2 is used to visualize the results. The source code is build from native R and Stan and is not dependent on any other population PK software. The source code is openly available at https://github.com/esiivola/admegp.

### Table 1 Parameters to generate the simulated data in all simulation studies and our prior assumptions

| Parameter name | Symbol and true value | Prior |
|----------------|------------------------|-------|
| Absorption     | $k_{\text{ref}} = 0.35$ l/h | $\log k_{\text{ref}} \sim N(\log(0.35), \log(2.01) / 1.96)$ |
| Clearance for 75 kg | $Cl_{\text{ref}} = 0.58$ l/h | $\log Cl_{\text{ref}} \sim N(\log(0.58), \log(1.05) / 1.96)$ |
| Volume for 75 kg | $V_{\text{ref}} = 10$ l | $\log V_{\text{ref}} \sim N(\log(10), \log(1.05) / 1.96)$ |
| Power for allometric scaling in clearance | $\alpha_{\text{Cl}} = 3/4$ | Fixed to true value |
| Power for allometric scaling in volume | $\alpha_V = 1$ | Fixed to true value |
| Noise standard deviation of log absorption | $\omega_{k_a} = 0.1$ l/h | Fixed to true value |
| Noise standard deviation of log clearance | $\omega_{\text{Cl}} = 0.207 = \log(15) / 1.96$ l/h | $\omega_{\text{Cl}} \sim N(0.207, \log(1.1) / 1.96)$ |
| Noise standard deviation of log volume | $\omega_V = 0.207 = \log(15) / 1.96$ l | $\omega_V \sim N(0.207, \log(1.1) / 1.96)$ |
| Constant noise standard deviation of log measurement | $\sigma_p = 0.1$ mg/l | $\sigma_p \sim N(0.1, \log(1.05) / 1.96)$ |
| Varying noise standard deviation of log measurement | $\sigma_c = 0$ mg/l | Fixed to true value |
| Nonlinear relative bioavailability | $f = 0$ | Fixed to true value |
| Reference dose | $D = 1$ mg | Fixed to true value |
| Maturation | $\log M(\text{age})$ | See Table 2 |

*a* Priors are for log parameters.

*b* Powers for allometric scaling are derived from first principles.
### Table 2

| Maturation model | Parameter name | Symbol in model | Prior |
|------------------|----------------|-----------------|-------|
| Hill model       | Hill coefficient | $h$ | $h \sim N \left( 2.94, \frac{\log(2)}{1.96} \right)$ |
| Hill model       | Half maturation age | MA50 | $MA50 \sim N \left( \frac{53}{52}, \frac{\log(2)}{1.96} \right)$ |
| (naive) GP model | Noise variances | $\omega^2_{Ct}$, $\omega^2_{Dv}$ | 0.05 |
| (naive) GP model | Control parameter | $\nu$ | 10 |
| (naive) GP model | Constant variance | $\rho_b$ | Fixed to 1 |
| (naive) GP model | Age when the variance is fixed | $c$ | Fixed to $\log(0.7)$ |
| (naive) GP model | Overall variability of the linear part | $\rho_v$ | $\frac{L-\mu_L}{\log(5.71-\log L)}$, where $L \sim N(0.8, 0.1)$ |
| (naive) GP model | Variability of the squared exponential | $\omega^2$ | $\omega^2 \sim GIG(a = 5, b = 1, p = -1)$ |
| (naive) GP model | Length scale | $\ell'$ | $\ell' \sim GIG(a = 5, b = 2, p = -6)$ (for Naive GP model: $\ell' \sim GIG(a = 5, b = 1, p = -1)$) |

**Note:** Means of $h$ and $MA50$ are values reported for a drug Sirolimus. GI$G$ stands for generalized inverse Gaussian, which is recommended in Stan manual for length scale of squared exponential. Virtual observations of the GP model are placed as follows: 8 virtual derivative observations logarithmically evenly spaced between 0 and 16 years; 1 fixed derivative observation and 1 fixed direct observation at age 18.

## 4.2 Hill maturation

In the first example, data are generated from a Hill type maturation model (Equation (16)). Parameters are equal to the values reported for immunosuppressant Sirolimus ($h = 2.94$, $MA50 = \frac{53}{52}$), a drug used after transplantation in order to avoid organ rejection. In this scenario, using the Hill model for the maturation function means there is no model misspecification.

Figure 2 illustrates the exponent of the maturation function ($M_{Cl}(\cdot)$) for clearance as estimated by the Hill and GP models. Note that unlike in Equation (3), results are not plotted on the logarithmic scale so that the figure shows the factor (between 0 and 1) by which the adult clearance (and doses) must be multiplied in order to obtain the pediatric clearance (and doses). The widths of the posterior predictive intervals for the GP models better reflect the sparsity of the data and also increase for younger patients. The tendency of the GP model estimates to bend toward the model prior ($\exp(0)$) can be seen for very young patients where the data are also sparse. This trend is more obvious for the naive GP model.

Table 3 summarizes the means and standard deviations of the loo-elpd differences between the three different models and the baseline (adult model) computed from the results of 50 different dataset initializations. Since the data are sparse, the random initialization affects the results substantially and the standard deviation across all realizations is large. To be able to compare different models, the table also summarizes their ranks for the same data realization. The table shows that the Hill model (labeled as parametric model) outperforms all other models when the data are simulated according to the Hill model. Table 4 summarizes the pooling factors estimated by the different models for the Hill data averaged over 50 random dataset realizations. The results for the Hill model and both the GP models are similar in this simulation study.

## 4.3 No maturation

In the second example, the data are generated without a maturation effect, that is, $\log(M(age)) = 0$. In other words, the data are simulated from the adult model. This maturation function means the Hill model is misspecified. Figure 3 shows the maturation function detected by the Hill and GP models. All models manage to model the maturation function reasonably well. The Hill model results in small values for $MA50$ and large values for $h$ so that the nonzero maturation can
not only be detected for very young patients. The GP model learns very small values for the lengthscale $\ell$ so that the derivative is slightly negative where the virtual derivative observations are, but the detected maturation still stays reasonably close to zero. The naive GP model results in a smoother approximation to the maturation function due to larger lengthscale values, but it still overlearns.

Table 3 summarizes the means and standard deviations of the loo-elpd differences between different models and the baseline (adult model) computed from the results of 50 different dataset realizations. The table shows that the model without maturation (labeled as the adult model) outperforms the other models, but both of the GP models perform slightly better than the Hill model. This result is to be expected, as the adult model is the model used to generate the
**TABLE 3** Model comparison for three simulated datasets (with known maturation function) and for Everolimus data (with unknown maturation function)

| Experiment name     | Adult | Parametric model | Naive GP model | GP model |
|---------------------|-------|------------------|----------------|----------|
|                     | loo-elpd difference | 0 (0) | **109 (21)** | 81 (34) | 95 (23) |
| Hill maturation     | rank  | 3.9 (0.3)        | **1.0 (0.2)** | 2.1 (0.4) | 2.9 (0.5) |
|                     | probability of winning | 0% | **96%** | 2% | 2% |
| No maturation       | loo-elpd difference | 0 (0) | −10 (10) | −13 (19) | −8 (13) |
|                     | rank  | **1.4 (0.8)** | 3.1 (0.9) | 3.0 (1.0) | 2.5 (0.9) |
|                     | probability of winning | **78%** | 6% | 8% | 8% |
| Double Hill maturation | loo-elpd difference | 0 (0) | 74 (25) | 69 (36) | **86 (27)** |
|                     | rank  | 3.9 (0.3)        | 2.2 (0.7) | 2.6 (0.7) | **1.3 (0.6)** |
|                     | probability of winning | 0% | 16% | 6% | **78%** |
| Everolimus data     | loo-elpd difference | 0 | −15 | −3 | −13 |

**Note:** For each simulated experiment, the values are reported as the means (and standard deviations) of differences of loo-elpd values between each method and the adult model computed over 50 random dataset realizations. Because of this, the values for the adult model are 0. For the loo-elpd differences, larger values are better. For Everolimus data, the loo-elpd differences are reported without repetitions. For the simulated datasets, the table also shows the mean ranks (and standard deviations) of the loo-elpd values (second row). Ranks are computed so that the highest loo-elpd gets rank 1 and lowest gets rank 4. In addition to this, the probability of each model being the best is reported on the third row by computing the proportion of times the model has the highest loo-elpd value for each dataset. The best model for each experiment is on bold.

**TABLE 4** Pooling factors of the individual clearance estimates for different models in three simulated scenarios (with known maturation function) and for Everolimus data (with unknown maturation function)

| Experiment name     | Adult | Parametric | Naive GP | GP |
|---------------------|-------|------------|----------|----|
| Hill maturation     | 0.03 (0.02) | 0.24 (0.05) | 0.32 (0.12) | 0.34 (0.14) |
| No maturation       | 0.24 (0.07) | 0.25 (0.09) | 0.36 (0.13) | 0.32 (0.15) |
| Double Hill maturation | 0.02 (0.00) | 0.11 (0.03) | 0.31 (0.15) | 0.29 (0.17) |
| Everolimus data     | 0.66 | 0.73 | 0.75 | 0.72 |

**Note:** The values for the simulated scenarios are reported as mean (standard deviation) over 50 dataset realizations. The values for Everolimus data are displayed as the pooling factors of the reported dataset.

Data. Table 4 summarizes the pooling factors of the different models for 50 different datasets. Again, the results for the Hill model and GP models are similar.

### 4.4 Double Hill maturation

In the third example, the data are generated from a double Hill type maturation model:

\[
\log(M(\text{age})) = \log \left( p_1 \frac{\text{age}^{h_1}}{MA50_1^{h_1} + \text{age}^{h_1}} + (1 - p_1) \frac{\text{age}^{h_2}}{MA50_2^{h_2} + \text{age}^{h_2}} \right),
\]  

with \( p_1 = 0.5, h_1 = 0.7, MA50_1 = 6.0, h_2 = 3.5, \) and \( MA50_2 = 7.0. \) Figure 4 shows the maturation function detected by the Hill and GP models. In this case, the Hill model is unable to capture the true function and the resulting estimates are overconfident, especially for kids of ages 1 to 3 years. Both GP models are able to capture the true maturation function more accurately and unlike for the Hill model, the widths of the posterior credible intervals are not unduly narrow for ages 1 to 2.5 years.

Table 3 summarizes the means and standard deviations of the loo-elpd differences between different models and the baseline (adult model) computed from the results of 50 different dataset realizations. The table shows that the GP model
outperforms naive GP and Hill models as it is more flexible than the Hill model but more constrained (monotonic) than the naive GP. Table 4 summarizes the means and standard deviations of the pooling factors estimated by the different models for the double Hill maturation model averaged over 50 random dataset realizations. The pooling factors for the GP models are higher than for the Hill model, which indicates that the Hill model is not flexible enough to fit the data. Figure 4 shows that the Hill model does not fit the data well, which can also been seen in the small value of the pooling factor.

4.5 Effects of sparsity on the model performance

On average, the more patients and measurements there are, the better the estimates for the maturation are. But how does data sparsity affect the results? We repeat the simulations described above, still using the Hill type maturation model (see Section 4.2) but this time varying the number of patients and the number of observations per patient. We then compute the average estimation error as

$$\frac{1}{N} \sum_{i=1}^{N} \left| \log \tilde{M}(\text{age}_{i}^{(i)}) - \log M(\text{age}_{i}^{(i)}) \right|,$$

where $\tilde{M}(\text{age})$ is median of the estimated maturation and $\{\text{age}_{i}^{(i)}\}^{N}_{i=1}$ are 100 ages selected linearly between 0 and 5 years.

Figure 5 visualizes the average error for 0-to-5-year-old patients for the Hill and the GP models as a function of the number of patients and concentration measurements per patient. The results are averaged over 50 dataset realizations for each plotted data point. Other parameters and prior distributions of the model stay unchanged. The Hill model has smaller average errors than the GP models, as this is the true data generating model.

The figure shows that the average error drops as the number of patients increases. However, there is no clear trend when the number of measurements per patient increases. The figure also shows that in the case of a fixed budget of concentration measurements, within the scope of the grid, it is better to select more patients with fewer measurements per patient rather than fewer patients with more measurements. Certainly these findings are specific to the chosen scenario for which the within-patient error is relatively small as compared to the between patient heterogeneity. While this is a common real data situation, the results may differ in other circumstances. A consequence of this is that only few measurements per patient are needed to accurately estimate the individual maturation function values.
FIGURE 5  Average error in the logarithm of maturation for patients of age 0 to 5 years as a function of the number of patients and observations per patient. The values are averaged over 50 dataset realizations from simulation using the Hill maturation function. The maturation is modeled with, A, the Hill model, B, the naive GP model, and with, C, the GP model. Three black lines show different combinations of patients and observations per patients when the total number of observations is limited to 100, 200, or 300 [Colour figure can be viewed at wileyonlinelibrary.com]

5  ANALYSIS OF PEDIATRIC CLINICAL DATA OF EVEROLIMUS

Everolimus is an immunosuppressant that is used to prevent the rejection of organ transplants. As an example to demonstrate our approach, we re-analyze a pediatric trial investigating Everolimus used in liver transplantation. The pediatric trial and the corresponding adult study have been modeled with a one-compartment model with oral administration and dose-dependent bioavailability (see Section 4.1). Kovarik et al.30 presented the PK model for adults and Dumortier and Looby31 presented the study design for the pediatric trial. No allometric scaling has been used in the adult PK model, but for the pediatric trial data, the clearance has been modeled using the weight dependent function

\[ M(\text{weight}) = \left( \frac{\text{weight}}{20\text{kg}} \right)^{\theta_p}. \]  

(18)

This modeling approach for the pediatric trial data empirically accounts for size differences that were not accounted for in the adult model. Therefore, we consider this term as a deviation from the adult model and treat it as a maturation function. The difference of this maturation function to the ones considered earlier is that it is a function of body weight instead of age. The pediatric dataset consists of 21 patients with 6.2 concentration observations on average per patient. The model parameters and their informative prior distributions learned from the adult data can be found in Table 5. Prior distributions of the model parameters are listed in Table 6.

A standard approach to assess if the pediatric data are consistent with the adult model is to perform a visual predictive check (VPC). This entails using the adult model to predict conditional on the pediatric trial data (baseline covariates and dosing history) the data as collected in the pediatric trial.

If the pediatric data are aligned with the predictions from the adult model, then this is interpreted as confirmation that the adult model is applicable and correctly specified for the pediatric population as well. Figure 6 illustrates the predictions generated with the adult PK model for the patients of the real pediatric data and the actual measurements of those patients. This visual predictive check (VPC) compares at the the predicted lower 10%, median, and 90% quantile with the quantiles of the actual data. The figure shows that the median of the data appears to be adequately predicted by the model, while the outer quantiles are difficult to conclude on. The figure also illustrates the limitations of the VPC in the context of model checking for pediatric data. First, in realistic data, some patients have a longer record of measurements and the data become more sparse the more time goes on. This can be seen in the figure as the visualized bins become wider on the right even though they have an equal amount of data. This also distracts the reader from the illustration as points on the right get more visual weight. Second, realistic pediatric data are often very sparse in the number of patients and number of measurements per patient. Sparse data limit the number of bins as a small number of data points in each bin makes estimating the percentiles rather inaccurate, especially if they are far from the median (eg, see the 10% and 90% percentiles in the figure). Furthermore, different age ranges can differ vastly in their baseline covariates, such as
Table 5  Prior assumptions of the model parameters

| Parameter name (unit)                       | Symbol in model | Prior                                      |
|---------------------------------------------|-----------------|--------------------------------------------|
| Absorption (1/h)                            | $k_{ref}$       | Fixed to 6.1                               |
| Clearance for 75 kg (l/h)                   | $Cl_{ref}$      | $\log Cl_{ref} \sim N \left( \log(34.7), \frac{\log(1.05)}{1.96} \right)$ a |
| Volume for 75 kg (l)                        | $V_{ref}$       | Fixed to 4170                               |
| Power for allometric scaling in clearance (no unit) | $\alpha_{CI}$   | Fixed to 0                                 |
| Power for allometric scaling in volume (no unit)   | $\alpha_V$     | Fixed to 0                                 |
| Noise standard deviation of log absorption (1/h) | $\omega_{ka}$  | Fixed to 0                                 |
| Noise standard deviation of log clearance (l/h)   | $\omega_{CI}$  | $\omega_{CI} \sim N \left( 0.34, \frac{\log(1.004)}{1.96} \right)$ |
| Noise standard deviation of log volume (l)       | $\omega_V$     | Fixed to 0                                 |
| Constant noise standard deviation of log measurement (mg/l) | $\sigma_p$      | $\sigma_p \sim N \left( 0.33, \frac{\log(1.04)}{1.96} \right)$ |
| Varying noise standard deviation of log measurement (mg/l) | $\sigma_c$      | $\sigma_c \sim N \left( 0.81, \frac{\log(1.04)}{1.96} \right)$ |
| Non linear relative bioavailability (no unit)    | $f$             | $f \sim N \left( 0.34, \frac{\log(1.01)}{1.96} \right)$ |
| Reference dose (mg)                          | $D$             | Fixed to 5 mg                               |

aNote that prior is for log clearance!

Table 6  Prior distributions of the maturation models used in the clinical data example

| Maturation model | Parameter name | Symbol in model | Prior                                      |
|------------------|----------------|-----------------|--------------------------------------------|
| Parametric model | power of allometric scaling       | $\theta_p$      | $\theta_p \sim N \left( 1.02, \frac{\log(7.2)}{1.96} \right)$ |
| (naive) GP model | Noise standard deviation | $\omega_{CI}, \omega_v$ and $\omega_{ad}$ | Fixed to 0.05 |
| (naive) GP model | Control parameter          | $\nu$           | Fixed to 10                                |
| (naive) GP model | Constant variance           | $\rho_b$        | Fixed to 1                                 |
| (naive) GP model | Age when the variance is fixed | $c$             | Fixed to $\log(0.7)$                       |
| (naive) GP model | Overall variability of the linear part | $\rho_e$       | $\frac{L-\exp}{\log(5.7)-log(0)}, \text{ where } L \sim N \left( 0.8, 0.1 \right)$ |
| (naive) GP model | Variability of squared exponential | $\omega$       | $\omega \sim GIG(a = 1, b = 1, p = 0)$ |
| (naive) GP model | Length scale                | $\ell$          | $\ell \sim GIG(a = 1, b = 2, p = 0)$       |

Note: GIG stands for generalized inverse Gaussian, which is recommended in Stan manual for length scale of the squared exponential. Virtual observations of the GP model are placed as follows: 8 virtual derivative observations logarithmically evenly spaced between 0 and 3 years; 1 fixed derivative observation and 1 fixed direct observation at age 5.

Body size, leading to generally wider prediction bands (colored bands in the figure) when comparing to VPC plots that are generated for adult data. Splitting the pediatric dataset into smaller age ranges amplifies the sparsity problem even more. These factors make reading the VPC plot harder and concluding whether or not the model is correct becomes more prone to subjectivity. A VPC is still a valuable tool since it allows to evaluate the model on its original scale, which is often helpful. However, very sparse data can limit its usefulness. Hence it should only be used together with other analysis methods to conclude if the model is correct.

Figure 7 illustrates the maturation function for clearance detected by the GP models and the parametric maturation model (Equation (18)). To make the parametric and GP methods comparable, the weights in the parametric model are transferred to ages with a standard population curve. The figure shows that the GP model is very close to value 1 for the whole age range. This results suggests only very moderate or no maturation effect.

Table 3 summarizes the loo-elpd differences between the different models and the baseline (adult model). The table shows that there is not much proof in the data that any maturation function would be needed as the loo-elpd values are very close to each other. The adult model slightly outperforms other models, but the differences are small.

Table 4 summarizes the pooling factors of the individual clearance estimates for the different models. The table shows little difference between the models, but the pooling factor for the naive GP model is higher than for the rest, which may
FiguRe 6  Data and dashed lines visualizing the 10%, 50%, and 90% percentiles for each measurement time in the pediatric data. The data are divided so that each of the eight visualized bins have an equal number of measurements in them. The continuous lines of the same color represent the same percentiles for predictions made with the posterior fit of the model used for the adults (logarithmic maturation function set to zero) and the shaded area around each line shows 2.5% and 97.5% percentiles of the percentiles [Colour figure can be viewed at wileyonlinelibrary.com]

FiguRe 7  Maturation function median, 68.27% and 95.45% central posterior intervals (corresponds to ±1 and 2 standard deviations) detected by, A, the parametric model, B, the naive GP model, and, C, the GP model. The figure also shows the estimated mean maturation and the adjusted individual clearance values. The ages with virtual derivative observations used in the GP model are indicated with grey squares. The reference weight for the parametric maturation function is 20 kg. The x-axis is presented on a logarithmic scale to better visualize that the maturation function changes more for younger kids [Colour figure can be viewed at wileyonlinelibrary.com]

indicate overfitting of the naive GP to the sparse data. This suggests that the naive GP model also learns variation between individuals and not only the maturation function.

To summarize, the proposed GP approach does not detect any noticeable deviation from the adult model. No maturation function or any other allometric effect is apparent from the Everolimus pediatric trial data. The clinical data also show a clear shortcoming of the frequently used parametric approach chosen here, which is the implied discontinuity. Recall that the adult model holds for subjects with age beyond 18 years. At the boundary age of 18 years, the two different models would imply a discontinuous change in the apparent clearance. This discontinuity is avoided with the GP modeling approach which is anchored at the age of 18 years to have no deviation from the adult model.

6  |  SUMMARY AND DISCUSSIONS

The choice of the dosing regimen for pediatric trials is in most cases based on extrapolation of an adult PK model in the absence of actual pediatric trial data. Once the pediatric trial data becomes available it is essential to confirm the appropriateness of the dosing regimen used in the trial. In this article, we proposed the use of a constrained non-parametric GP to detect deviations from the original extrapolation model. The GP model approach is a data-driven approach that is used to estimate possible model deviations. We evaluated our approach in four simulation studies using realistically sparse datasets. The evaluation showed that our approach is capable of correctly (not) detecting an unexpected maturation
function if (not) present. The results are summarized in Table 3. In addition to this, we were able to reproduce similar results and performance on real pediatric clinical data when comparing to the results of the parametric model published earlier.

Extreme data sparsity is a problem to non-parametric approaches like GPs. By using virtual data to add prior assumptions like monotonicity to the model, sparsity becomes less of a problem. GPs were shown to behave unsteadily if the data and prior assumptions disagree. In this case, the virtual observations introduce extra bias to the model, but not as much as a non-realistic parametric model. If the prior assumptions agree with the data, virtual observations were shown to solve the problems caused by the sparsity.

A key advantage of a non-parametric GP approach is the lower risk of model misspecification when comparing to the parametric approaches. The GP learns the functional relationship of maturation and age only locally in ranges supported by the data. In contrast, a parametric approach will be informed from data in a local age range, but result in a comparatively confident global relationship across large age ranges. Thus, the use of a GP protects against over-confident conclusions that are far outside the domain of the data. For regions close to the domain of collected data or sparsely sampled data regions, the GP will infer a smooth functional relationship and preserve the continuous differentiability of the maturation function. Thus, discontinuous changes in the maturation function are avoided.

In summary, we have shown that a non-parametric GP approach can reliably detect whether dosing regimen adjustments are needed or if the chosen dosing regimen is adequate. In comparison to model checking approaches, which rely on possibly subjective evaluation of model diagnostics, the proposed GP approach is data-driven and thus less subjective. In addition to this, GPs avoid bias and overconfident conclusions when compared to parametric approaches. Overall, the presented GP approach can lead to better decisions in pediatric drug development.

ACKNOWLEDGEMENTS
The work of Aki Vehtari was supported by Academy of Finland grants number 298741 and 313122. We would like to address our gratitude to Lisa Hampson for helping us clarify the language. We would also like to thank Thomas Dumortier for his useful feedback on the manuscript, experiments and our methodology. Last, we want to thank the two anonymous Reviewers for taking the time necessary to review the manuscript. We appreciate their comments, which greatly helped us to improve the quality of the manuscript.

DATA AVAILABILITY STATEMENT
Research data are not shared.

ORCID
Eero Siivola  https://orcid.org/0000-0002-3926-9651

REFERENCES
1. Regulation (EC) No 1902/2006 of the European Parliament and of the Council of 12 December 2006 amending Regulation 1901/2006 on medicinal products for paediatric use; December 2006. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32006R1901. Accessed June 1, 2020.
2. Best Pharmaceuticals for Children Act (BPCA); March 2012. https://www.nichd.nih.gov/research/supported/bpca. Accessed June 1, 2020.
3. Dunne J, Rodriguez WJ, Murphy MD, et al. Extrapolation of adult data and other data in pediatric drug-development programs. *Pediatrics*. 2011;128(5):1242-1249.
4. Lin JH, AYHL. Role of pharmacokinetics and metabolism in drug discovery and development. *Pharmacol Rev*. 1997;49(4):403-449.
5. Wakefield J. The Bayesian analysis of population pharmacokinetic models. *J Am Stat Assoc*. 1996;433(91):62-75.
6. West GB, Brown JH, Enquist BJ. A general model for the origin of allometric scaling laws in biology. *Science*. 1997;276(5309):122-126.
7. Rhodin MM, Anderson BJ, Peters AM, et al. Human renal function maturation: a quantitative description using weight and postmenstrual age. *Pediatr Nephrol*. 2009;24(1):67-76.
8. Holford N, Heo Y-A, Anderson B. A pharmacokinetic standard for babies and adults. *J pharm Sci*. 2013;102(9):2941-2952.
9. Anderson BJ, Fons G, Autret-Leca E, Allegaert K, Boccard E. Pediatric intravenous paracetamol (propacetamol) pharmacokinetics: a population analysis 1. *Pediatr Anesth*. 2005;15(4):282-292.
10. Ince I, de Wildt SN, Wang C, et al. A novel maturation function for clearance of the cytochrome P450 3A substrate midazolam from preterm neonates to adults. *Clin Pharmacokinet*. 2013;52(7):555-565.
11. Bruusse JM, Calvier EAM, Krekels EHH, et al. Children in clinical trials: towards evidence-based pediatric pharmacotherapy using pharmacokinetic-pharmacodynamic modeling. *Expert Rev Clin Pharmacol*. 2016;9(9):1235-1244.
12. Krekels EHJ, van Hasselt JGC, van den Anker JN. Evidence-based drug treatment for special patient populations through model-based approaches. *Eur J Pharm Sci*. 2017;109:S22-S26.
13. Williams CKI, Rasmussen CE. *Gaussian Processes for Machine Learning*. Cambridge, MA: The MIT Press; 2006.
14. Kallen A. *Computational Pharmacokinetics*. Boca Raton, FL: CRC Press; 2007.
15. Kac M, Siegert AJF. An explicit representation of a stationary Gaussian process. *Ann Math Stat*. 1947;18(3):438-442.
16. Matheron G. Principles of geostatistics. *Econ Geol*. 1963;58(8):1246-1266.
17. Blight BJN, Ott L. A Bayesian approach to model inadequacy for polynomial regression. *Biometrika*. 1975;62(1):79-88.
18. O’Hagan A. Curve fitting and optimal design for prediction. *J Royal Stat Soc Ser B (Methodol)*. 1978;40(1):1-24.
19. Duvenaud D. Automatic Model Construction with Gaussian Processes [PhD thesis]. University of Cambridge; 2014.
20. Riihimäki J, Vehtari A. Gaussian processes with monotonicity information. Paper presented at: Proceedings of the 13th International Conference on Artificial Intelligence and Statistics. Sardinia, Italy; 2010:645-652.
21. Solak E, Murray SR, Leithead WE, Leith DJ, Rasmussen CE. Derivative observations in gaussian process models of dynamic systems. *Adv Neural Inf Process Syst*. 2003;13:1033-1040.
22. Schwarz G. Estimating the dimension of a model. *Ann Stat*. 1978;6(2):461-464.
23. Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput*. 2017;27(5):1413-1432.
24. Gelman A, Pardoe I. Bayesian measures of explained variance and pooling in multilevel (hierarchical) models. *Technometrics*. 2006;48(2):241-251.
25. Xu XS, Yuan M, Karlsson MO, Dunne A, Nandy P, Vermeulen A. Shrinkage in nonlinear mixed-effects population models: quantification, influencing factors and impact. *AAPS J*. 2012;14(4):927-936.
26. Savic RM, Karlsson MO. Importance of shrinkage in empirical Bayes estimates for diagnostics: problems and solutions. *AAPS J*. 2009;11(3):558-569.
27. Emoto C, Fukuda T, Mizuno T. Characterizing the developmental trajectory of sirolimus clearance in neonates and infants. *CPT Pharmacometrics Syst Pharmacol*. 2016;5(8):411-417.
28. Stan Development Team Stan modeling language users guide and reference manual. version 2.17.1; 2018. http://mc-stan.org.
29. Hadley W. *ggplot2: Elegant Graphics for Data Analysis*. 2nd ed. New York, NY: Springer-Verlag; 2016.
30. Kovarik JM, Hsu C-H, McMahon L, Berthier S, Rordorf C. Population pharmacokinetics of everolimus in de novo renal transplant patients: impact of ethnicity and comedications. *Clin Pharmacol Therapeut*. 2001;70(3):247-254.
31. Dumortier T, Looby M. Supporting a pediatric investigational plan for everolimus - defining the extrapolation plan; May 2016. https://www.ema.europa.eu/en/documents/presentation/presentation-supporting-pediatric-investigational-plan-everolimus-defining-extrapolation-plan-thomas_en.pdf. Accessed June 1 2020.

**How to cite this article:** Siivola E, Weber S, Vehtari A. Qualifying drug dosing regimens in pediatrics using Gaussian processes. *Statistics in Medicine*. 2021;40:2355–2372. https://doi.org/10.1002/sim.8907