medRxiv preprint doi: https://doi.org/10.1101/2024.05.06.24306555; this version posted May 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Scientific machine learning for predicting plasma concentrations in anti-cancer therapy

Authors

Diego Valderrama[1*], Olga Teplytska[2*], Luca Marie Koltermann[2], Elena Trunz[3], Eduard Schmulenson[2], Achim Fritsch[2], Ulrich Jaehde[2#], Holger Fröhlich[1,4#]

[1] Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Sankt Augustin, Germany

[2] Department of Clinical Pharmacy, Institute of Pharmacy, University of Bonn, Bonn, Germany

[3] Institute of Computer Science II, Visual Computing, University of Bonn, Bonn, Germany

[4] Bonn-Aachen International Center for Information Technology (B-IT), University of Bonn, Bonn, Germany

*shared first authorship, [#]shared senior authorship

Correspondence:

Holger Fröhlich, Ph.D.

Schloss Birlinghoven, 1, 53757 Sankt Augustin (Germany)

Tel. +49 151 70597940, Mail: holger.froehlich@scai.fraunhofer.de

Ulrich Jaehde, Ph.D.

An der Immenburg 4, D-53121 Bonn (Germany)

Tel. +49 228 735252, Mail: u.jaehde@uni-bonn.de

Funding

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.06.24306555; this version posted May 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

This work was partially funded by Federal Ministry of Education and Research within the projects „BNTrAinee“ (funding code 16DHBK1022).

Conflict of Interest

H.F. received grants from UCB and AbbVie. The other authors declare no competing interest for this work.

Keywords

Scientific machine learning, Machine Learning, Oncology, Drug Therapy, Pharmacokinetics, Concentration Prediction

Ethics declaration

The 5FU study (protocol-No: CESAR C-II-005, EudraCT-No; 2008-001515-37) was approved on 26.05.2008 by the ethics committee of the Albert-Ludwigs-Universität Freiburg, Freiburg, Germany. The C-IV-001 Sunitinib study (EudraCT-No: 2012-001415-23) was approved on 17.10.2012 by the ethics committee of the department of medicine of the Johann Wolfgang Goethe-Universität Frankfurt am Main, Frankfurt, Germany.

Abstract

A variety of classical machine learning approaches have been developed over the past ten years with the aim to individualize drug dosages based on measured plasma concentrations. However, the interpretability of these models is challenging as they do not incorporate information on pharmacokinetic

medRxiv preprint doi: https://doi.org/10.1101/2024.05.06.24306555; this version posted May 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

(PK) drug disposition. In this work we compare well-known population PK modelling with classical and a newly proposed scientific machine learning (SciML) framework, which combines knowledge on drug disposition with data-driven modelling. Our approach lets us estimate population PK parameters and their inter-individual variability (IIV) using multimodal covariate data of each patient. A dataset of 549 fluorouracil (5FU) plasma concentrations as example for an intravenously administered drug and a dataset of 308 sunitinib concentrations as example for an orally administered drug were used for analysis. Whereas classical machine learning models were not able to describe the data sufficiently, the proposed model allowed us to obtain highly accurate predictions even for new patients. Additionally, we demonstrated that our model could outperform traditional population PK models in terms of accuracy and greater flexibility when learning population parameters if given enough training data.

1.Introduction

The vision of precision medicine is to provide the right drug to the right patients at the right time and dosage. The choice of the optimal drug dosage is based on different criteria in which pharmacokinetics (PK) in an individual patient is an essential factor. For this purpose, population PK models are traditionally formulated as compartmental differential equations. In these models, individual differences relative to the population average are described as random statistical effects. A limitation of this established approach is that in case of incomplete or missing knowledge about specific PK processes (e.g. absorption) the formulation of model equations can become challenging. Moreover, these models do not always provide accurate predictions for new patients who have not been part of the training dataset and, therefore, might come from a different distribution.

Thus, during the last decade various machine learning techniques have been studied as a complementary strategy for the estimation of drug plasma concentrations, aiming to individualize doses. The most frequently used algorithms for dose optimization are decision trees and their ensembles, support

medRxiv preprint doi: https://doi.org/10.1101/2024.05.06.24306555; this version posted May 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

vector machines and neural networks. Additionally, reinforcement learning plays an increasing a role. The identified algorithms primarily aided in the dose individualization of anticoagulants, immunosuppressants and antibiotics [1–5]. In oncology, most of the studies we identified used reinforcement learning, including classical Q-Learning [6–9], deep Q-Learning [10, 11], deep double Q-Learning [12], fuzzy reinforcement learning [13, 14], conservative Q-Learning [15] and other approaches [16, 17]. Recently, several models have been proposed using neural networks for the prediction of drug concentrations [4, 18–20]. Lu et al. proposed to learn the initial conditions of a neural ordinary differential equation (ODE) system trained to predict PK profiles [4]. A hybrid of an ODE and a neural network was employed by Qian [18]. Similarly, Janssen et al. used a neural network to learn covariate effects which were employed in a compartment model for describing drug concentrations [19]. Valderrama et al. [20] introduced PK-SciML, a Scientific Machine Learning (SciML) [21, 22] approach for learning an unknown absorption mechanism while simultaneously estimating PK parameters. Although their model showed promising results, it was only tested on synthetic data and did not consider the inter-individual variability (IIV) of the clinical data, hence only generating populational level predictions for different dose groups. Considering that a SciML framework has the advantage of not needing to establish the relationships between covariates and parameters a priori while allowing the integration of domain expertise, we here introduce a multimodal pharmacokinetic SciML (MMPK-SciML) approach, an extension of PK-SciML, which aims to learn the IIV based on multimodal covariate data of individual patients.

As a case study, we use two real datasets for two different oncology treatments as examples for an iv and an oral treatment route and compare our model with different classical machine learning and population PK models. We demonstrate that our model produces reliable predictions, while being able to simulate new patients.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.06.24306555; this version posted May 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

2. Methods

2.1. Data collection and preprocessing

2.1.1. Fluorouracil (5FU)

In this work, plasma concentrations of patients who received fluorouracil (5FU)-based infusional chemotherapy at the Oncological Outpatient Clinic UnterEms in Leer, Germany, were retrospectively analyzed. This study was approved by the local medical ethics committee, but trial registration was not conducted due to the retrospective nature. Patients with documented therapeutic drug monitoring (TDM) of 5FU were included in the analysis. Plasma 5FU concentrations were obtained at steady state during continuous infusion and quantified using the My5-FU™ immunoassay (Saladax Biomedical Inc., Bethlehem, PA, USA) [23]. The dataset included 549 TDM samples from 157 patients and further information on demographics, blood counts and adverse events. Adverse events (AE) were graded at each patient visit according to the Common Terminology Criteria for Adverse Events (CTCAE) version 5.0 [24]. Outliers were defined as individuals with a concentration below the lower limit of quantification (< 52 ng/mL) or a clearance above 1478 L/h [25] and were excluded from the analysis. Missing data was imputed using the last observation carried forward approach within the same treatment cycle.

2.1.2. Sunitinib

Sunitinib PK data were pooled from two PK/PD studies focusing on sunitinib treatment in patients with metastatic renal cell carcinoma (mRCC) and patients with metastatic colorectal cancer (mCRC) [26, 27]. The C-IV-001 study (EudraCT-No: 2012-001415-23, date of authorisation: 17.10.2012) was a phase IV PK/PD substudy of the non-interventional EuroTARGET project, which recruited patients with mRCC at nine medical centres in Germany and the Netherlands [26]. Sunitinib doses ranged from 37,5-50 mg daily, administered orally on a 4-week on/2-week off schedule. The C-II-005 study (EudraCT-No: 2008-00151537, date of authorisation: 11.06.2008) was conducted to investigate the beneficial effect of sunitinib added to biweekly folinate, fluorouracil and irinotecan in patients with mCRC and liver metastases Patients were prescribed a daily dose of 37,5 mg sunitinib on a 4-week

medRxiv preprint doi: https://doi.org/10.1101/2024.05.06.24306555; this version posted May 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

on/2-week off schedule taken orally [27]. Both studies were performed in accordance with the Declaration of Helsinki.

A total of 308 sunitinib TDM samples were obtained from 26 mRCC and 21 mCRC patients. 345 samples were analysed for the pharmacodynamic biomarker sVEGFR2 and 337 for sVEGFR3, respectively [28]. Sunitinib measurements below the lower limit of quantitation (<0,06 ng/mL) and patients without plasma concentration data were excluded from the analysis.

2.1.3. Data preprocessing

The total dataset was split using a 10-fold cross-validation setting with a training-test split of 80/20, keeping data from one patient strictly in the same set to avoid a splitting bias. For the classical machine learning algorithms, categorical features were one-hot encoded and continuous features were scaled between zero and one. Additionally, missing data was imputed using MissForest within the cross-validation process for each split if the 'last observation carried forward' approach was not applicable.

2.2. Population pharmacokinetic modeling

The population pharmacokinetic (PopPK) model for 5FU comprised of a one-compartment model with linear elimination to describe 5FU disposition [29]. An IIV term was implemented on 5FU clearance and the volume of distribution together with its IIV were fixed to previously estimated values [29]. The residual variability was modelled as proportional and the body surface area, centered on the population median, was included as a linear covariate on clearance. Differently from the original model [29], the skeletal muscle index was not included as a covariate, because it was not available for all included patients. While Schmulenson et al. used the FOCE-I method to estimate the parameters [29], we compared this method to stochastic approximation expectation maximisation (SAEM) using NONMEM®. First, we estimated the pharmacokinetic parameters for the training patients of each split using the different methods and initial estimates as in [29]. In the next step, these values were used to simulate the expected concentrations for the test data. Mean concentration values were calculated from 1000 simulations.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.06.24306555; this version posted May 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

The structure of the population-pharmacokinetic model for sunitinib is depicted in Figure S1. A two-compartment model for sunitinib disposition and a biphasic distribution for its active metabolite SU12662 were used [28, 30]. Presystemic formation of SU12662 was modelled via a hypothetical enzyme compartment incorporated into the central compartment of sunitinib. An intercompartmental clearance connected the central compartment and the enzyme compartment and was fixed to the liver blood flow. Interindividual variability (IIV) was included for the central volumes of distribution for sunitinib and SU12662, the clearance of sunitinib and the fraction metabolized in a block matrix. Proportional errors for the parent drug and metabolite were used to describe the residual unexplained variability. Originally, the FOCE-I method was used for parameter estimation. Again, we compared this method to SAEM. Moreover, we simulated the expected concentrations for the test data and applied the same evaluation methods as for described for 5FU.

2.3. Classical Machine Learning Algorithms

Various classical machine learning methods were used for concentration prediction prediction. We compared Random Forests, Gradient Boosting, Extreme Gradient Boosting (XGBoost), Light Gradient Boosting (LGBoost), Support Vector Machines (SVM) with a radial basis function kernel and simple neural networks with a possible range of 2-10 units per layer in terms of performance. The input variables consisted of dose, weight, lean body mass (LBM), fat mass (FM), body surface area (BSA), age, sex, height and time since last dose for 5FU and sex, age, weight, height, BSA, time since last dose and the sVEGFR-2, sVEGFR-3 plasma concentrations for sunitinib. These potential covariates, despite most of them having been excluded in stepwise covariate modeling (SCM), were included to enable the machine learning algorithms to make use of potential previously missed relationships within the data as they have shown to outperform SCM in some cases [31]. All algorithms were applied with and without feature selection using XGBoost. Hyperparameter tuning was performed using the Bayesian hyperparameter optimisation software framework Optuna [32] and models were selected applying 5-fold cross validation with the mean squared error as the objective function. The neural networks were regularised applying common techniques such as drop-out, L1 regularization and gradient clipping to avoid overfitting.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.06.24306555; this version posted May 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

To investigate whether the model performance of the classical machine learning methods could be further improved by synthetic data, the training datasets were augmented for each split according to Table S1. To simulate patient data according to the correct distributions for each cross-validation fold, a PK model was fitted by NONMEM® on each fold and the obtained estimates were used for simulation. 1000 synthetic patients were created for each fold respectively and added to the clinical training data.

2.4. Multimodal Pharmacokinetic SciML model (MMPK-SciML)

The main motivation of our MMPK-SciML model was to overcome the existing limitations of PK-SciML [20] predictions, i.e., we wanted to build a model learning the IIV using neural networks and multimodal patient information. Following the classical pharmacokinetic framework, individual parameters using IIV are defined as follows:

$$\vartheta_i = \exp(log(TV_\vartheta) + \log \eta_i) \quad (1)$$

Where TV_ϑ is the typical - populational value of the parameter ϑ and η_i is the IIV also referred in this paper as eta value.

Our proposed architecture is composed of 2 main blocks: i) a neural network encoder which aims to predict the η values using patient covariates and ii) a structural well-defined ODE system to describe the PK dynamics. Therefore, given a set of k predicted patient parameters $\{\vartheta_i\}_{j=1:k}$, a dose regimen, and a time horizon, the individual concentration profiles were predicted by solving the initial value problem of the ODE system.

Following PK-SciML [20] and Lu et al. [4] the dosage was added to the first compartment of the ODE system. Additionally, we fixed the initial conditions to zero to guarantee a plausible ODE system. Model implementation is available on GitHub at https://github.com/SCAI-BIO/MMPK-SciML.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.06.24306555; this version posted May 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

2.4.1. Variational Inference

Let $y_{it}, t \in \mathcal{T}$ denote the concentration profile measured at time points $\mathcal{T}$ for patient i. Furthermore, x_i are patient-specific covariates. From a Bayesian inference perspective, we are interested in the posterior

$$p(\eta \mid \{x_i, y_{it}\}) = \frac{p(\{x_i, y_{it}\} \mid \eta)p(\eta)}{p(\{x_i, y_{it}\})} = \frac{p(\{x_i, y_{it}\} \mid \eta)p(\eta)}{\int p(\{x_i, y_{it}\} \mid \eta)p(\eta)d\eta} \quad (2)$$

Unfortunately, solving the integral is analytically intractable. Approximations via Markov Chain Monte Carlo techniques are possible but very time-consuming. To overcome this problem, Kingma et al. [33] introduced a stochastic variational inference framework for neural networks which allows quantification of epistemic uncertainty, i.e. uncertainty due to missing data. The key idea is to approximate $p(\eta \mid \{x_i, y_i(t)\})$ by a distribution $q(\eta \mid \{x_i, y_i(t)\})$, which is typically supposed to be Gaussian. The mean and (log) variance of this distribution are learned from the observed data via an encoder neural network ϕ_θ:

$$\{\mu_{\eta_i}; \log(\sigma_{\eta_i}^2)\}_{i=1:K} = \phi_\theta(x_i, \{y_{it}\}) \quad (3)$$

The initial value problem can then be solved by sampling from the distribution $N(\mu_{\eta_i}, \sigma_{\eta_i}^2)$ while taking advantage of the re-parametrization trick [34]. With that it is possible to formulate a loss function for training ϕ_θ by maximizing the so-called *Evidence Lower Bound* (ELBO) on the true posterior on log-scale:

$$\log p(\eta \mid \{x_i, y_i(t)\}) \geq \mathcal{L}_{ELBO}(\{x_i, y_{it}\}) = \underbrace{\mathbb{E}[\log p(\{x_i, y_{it}\} \mid \eta)]}_{data\ likelihood} - \underbrace{D_{KL}(q_\Phi(\eta_i|x_i, y_{it})\|p(\eta_i))}_{regularization}$$

where $D_{KL}(q_\Phi(\eta_i|x_i, y_{it})\|p(\eta_i))$ denotes the Kullback-Leibler divergence (a statistical distance measure) between the approximate posterior distribution $q_\Phi(\eta_i|x_i, y_{it})$ and a prior distribution $p(\eta_i)$. Supposing $p(\eta_i) = N(0, \lambda^2)$ and Gaussian noise, the negative ELBO can be re-written as a loss function $\ell(\{\tilde{y}_{it}\}, \{y_{it}\})$:

$$-\mathcal{L}_{ELBO}(\{x_i, y_{it}\}) \propto \ell(\{\tilde{y}_{it}\}, \{y_{it}\}) := \frac{1}{n}\sum_{i=1}^{n}\sum_{j\in\mathcal{T}} \frac{(y_{ij} - \tilde{y}_{ij})^2}{\epsilon_{ij}^2} - \frac{1}{2}\left(\frac{\sigma_{\eta_i}^2}{\lambda^2} + \frac{\mu_{\eta_i}^2}{\lambda^2} + \log\frac{\lambda^2}{\sigma_{\eta_i}^2}\right) \quad (4)$$

medRxiv preprint doi: https://doi.org/10.1101/2024.05.06.24306555; this version posted May 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

where ϵ_{ij}^2 is the variance of the measurement noise and $\tilde{y}_{ij}$ the ODE predictions. For the following experiments we assumed a proportional error $\epsilon_{ij}^2 \propto y_{ij}$. Note that a smaller $\lambda > 0$ results in a stronger regularization of $\sigma_{\eta_i}^2$ and $\mu_{\eta_i}^2$ towards their values in the prior distribution.

2.4.2. Model details

2.4.2.1. 5FU

Because all the measurements were taken at steady state there is only one measurement per individual patient. As structural ODE System we used an intravenous model which is defined as follows:

$$\frac{dC1}{dt} = \frac{D}{IT} - \frac{CL}{V}C1 \quad (5)$$

$$C1(t=0) = 0$$

Where D is the dose and IT is the infusion time.

To learn the η values, we defined ϕ_θ as an encoder network using the concatenation of the measured concentration, dose, weight, LBM, FM, BSA, age, sex and height was used as input for the first layer. Specifically, only IIV on CL was learnt for 5FU. Figure 1 (top) shows an overview of our model architecture for 5FU. Model hyperparameters and more details can be found in Appendix A

2.4.2.1. Sunitinib

Unlike the 5FU dataset, for sunitinib we took advantage of having measurements at different time points during the PK profile. We used as structural ODE system the model proposed by Diekstra et al. [28]:

$$CLIV = \frac{K_A C_1 + \frac{Q_H}{V2_S}C2}{Q_H + CL_S} \quad (6)$$

$$\frac{dC_1}{dt} = -K_A C_1$$

$$\frac{dC_2}{dt} = Q_H CLIV - \frac{Q_H}{V2_S}C_2 - \frac{Q_S}{V2_S}C_2 + \frac{Q_S}{V3_S}C_3$$

$$\frac{dC_3}{dt} = \frac{Q_S}{V2_S}C_2 - \frac{Q_S}{V3_S}C_3$$

medRxiv preprint doi: https://doi.org/10.1101/2024.05.06.24306555; this version posted May 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

$$\frac{dC_4}{dt} = F_M CL_S CLIV - \frac{CL_M}{V2_M} C_4 - \frac{Q_M}{V2_M} C_4 + \frac{Q_M}{V3_M} C_5$$

$$\frac{dC_5}{dt} = \frac{Q_M}{V2_M} C_4 - \frac{Q_M}{V3_M} C_5$$

$$C1(t=0) = C2(t=0) = C3(t=0) = C4(t=0) = C5(t=0) = 0$$

As in the original work the ODE parameters were first calculated following equation 1 and then scaled based on the weight of the patient. Specifically, $CL_S, Q_S, CL_M, Q_M, Q_H$ were scaled by a factor of $\left(\frac{Weight}{70}\right)^2$ and $V2_S, V3_S,\ V2_M, V3_M$ by a factor of $\frac{Weight}{70}$

ϕ_θ was defined as a multimodal encoder containing three blocks. The first block was an encoder for static covariates $\{x_i\}$. The second block encoded the longitudinal covariates $\{y_{it}\}$, and for this purpose we used the Time-LSTM [35]. The output of both encoders was concatenated and used by a third block, the projection encoder, with 2 subnetworks each producing K outputs which define $\left\{\mu_{\eta_i}; \log\left(\sigma^2_{\eta_i}\right)\right\}_{j=1:K}$. We defined K=4 corresponding to the IIV for $CL_S, V2_S, F_M,\ V2_M$. Figure 1 (bottom) shows an overview of our model architecture for sunitinib. Model hyperparameters and more details can be found in Appendix A.

2.5. Model Comparison

To assess predictive performance, the mean absolute error (MAE), and the root mean squared error (RMSE) were calculated and compared for the different approaches used in this project. Goodness-of-fit (GOF) plots were used to support the quantitative results. In the case of the population PK models, mean concentration values were calculated from 1000 simulations using parameters estimated on the training data as initials and then plotted. For the classical ML methods, the final predictions on the test set were used for the calculation of the metrics. For the MMPK-SciML approach, the results were obtained by using the means predicted by the model for each of the eta values because these represent the expected value.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.06.24306555; this version posted May 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Furthermore, to evaluate how well the models perform in simulating new patients, prediction corrected visual predictive checks (pcVPCs) were generated. These graphs could only be obtained for the population PK and the MMPK-SciML models because the classic machine learning approaches are not generative.

3. Results

3.1. Dataset characteristics

A dataset of 549 fluorouracil (5FU) plasma concentrations from 157 patients as example for an IV administration and another dataset of 308 sunitinib concentrations from 47 patients as example for a po administration were used for analysis. Baseline characteristics of all patients included in our analyses can be seen in Table 1.

3.2. Population pharmacokinetic modeling results

In the Pop-PK analyses, all PK parameters and their IIVs as defined in the original publications [28, 29] could be estimated for all data splits. The mean estimated parameter values were in a similar range to the original estimated values for the whole datasets as depicted in Table 2 and the η values appeared to be normally distributed for all tested methods. There were no relevant differences between the estimated parameters and the simulated concentrations for the test data of the FOCE-I and SAEM methods.

For 5FU, using both FOCE-I and SAEM, we observed balanced η distributions. However, the GOF was still relatively poor, and showed wide confidence intervals (Figures 2, 4)

In the case of sunitinib, initial convergence problems for some splits with the SAEM algorithm had to be resolved by setting the initial estimates closer to the final values obtained by Diekstra et al. [28] Once this issue was resolved, we obtained similar results to FOCE-I and comparatively good fits for both methods (Figures 3, 4).

medRxiv preprint doi: https://doi.org/10.1101/2024.05.06.24306555; this version posted May 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

3.3. Classical machine learning methods

The proposed classical ML methods were not able to learn relationships within the data and accurately predict plasma concentrations of both drugs as can be seen in the GOF plots in Figure 2, and the cross-validated accuracy metrics in Table 3. Results did not differ relevantly between algorithms with and without feature selection. Therefore, we only depict results of the algorithms where feature selection was applied. The only comparatively well performing classical algorithm on the sunitinib data was the Light Gradient Boosting algorithm. However, the performance differed vastly across splits and GOF plots suggest a rather poor performance (Figure 2).

To investigate whether the poor model performance could be improved with additional synthetic data, we augmented the training data with data that was simulated by the population PK model and repeated the model runs. As a limitation, the sunitinib biomarker data had to be simulated using the final parameter estimates from Diekstra et al. [28] due to matrix singularity in parameter estimation. However, as can be seen in Table 3, the prediction accuracy was not substantially improved for both datasets. Optimized hyperparameters and selected features are reported in Supplementary Tables S2 and S3.

3.4. MMPK-SciML

Our proposed MMPK-SciML model generates accurate predictions for the different oncology treatments routes used in this paper. Figure 2, bottom row, illustrates the GOF plots for 5FU. On the right side we depict a second version of our model using a fixed volume population parameter $TV_V = 40.0$, which will be referred to as MMPK-SciML* in the tables and figures. Opposed to classical machine learning methods a close correlation between the predictions and the real data was found. At the same time, cross-validated RMSE and MAE metrics were even lower than those of the population PK model (Tables 3, S4). Importantly, results were highly accurate for both versions of our model, indicating that the model in both cases was able to learn TV_{CL}. Especially for 5FU, we observed an at least three times better MAE than all other methods, and at least 30% improvement in RMSE for all the folds. At the same time there was a considerable difference in the predicted across model versions

medRxiv preprint doi: https://doi.org/10.1101/2024.05.06.24306555; this version posted May 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

(Table 2). Although the TV_V on average was close to the fixed volume, the variance of our standard MMPK-SciML model was larger than in the second version. Hence, during the cross-validation TV_{CL}, TV_V estimates differed.

Similarly, Figure 3 (bottom row right) illustrates the GOF plots of our MMPK-SciML model for Sunitinib. Although the GOF plots were not as good as those for 5FU, our model still showed comparable performance to the FOCE-I method, and much lower variance in prediction accuracy across data splits than the best performing classical machine learning method LightGBM (Tables 3, S4).

As can be seen in Figure 4, the MMPK-SciML models performed well in simulating new patients, as the associated statistics of the real data are within the 90% confidence intervals (shaded region) of the predictions. Additionally, our models approximated the posterior η distributions in a reliable manner (Figure S2).

4. Discussion

Using two examples of oncological treatments with different administration routes (IV and oral), our results demonstrate that generally a compartmental model structure is required to make accurate predictions of drug plasma concentrations. Overall, only the MMPK-SciML as well as the Pop-PK methods were able to adequately describe the underlying drug disposition. In contrast to MMPK-SciML, other ML-based PK models are entirely data-driven and cannot accurately learn the PK profile when working with a small number of measurements and dose schedules [4, 20, 36]. Moreover, these models have the limitation that they need to be initialized with test data to generate simulations, which in real application scenarios is usually not possible. Although our previously proposed PK-SciML [20] does not have this limitation, it only generates population level predictions and does not process multimodal data. Altogether, this demonstrates a clear advantage of our MMPK-SciML architecture proposed in this paper.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.06.24306555; this version posted May 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Remarkably, LightGBM performed well for the sunitinib analysis. Nevertheless, its performance was rather inconsistent across splits and not good enough to be used in practice, as can be seen in the GOF plots. The comparatively good performance of LightGBM may be explained by the fact that gradient boosting machines construct an ensemble of decision trees. Decision trees implicitly split the data into different subspaces, resulting in an implicit discretization. This behavior could be advantageous in situations in which the data is of multimodal nature, as typical in clinical studies.

Overall, the classic machine learning models were not able to appropriately learn key aspects of the data generating process to produce accurate PK predictions. Additionally, data augmentation could not further improve the performance of classical machine learning algorithms, suggesting that the inherent complexity of the temporal dynamics, the variance and the presence of outliers are difficult to learn by methods that have been designed for comparably simple tabular data only and thus use no information about the PK related processes. However, the results possibly could have been improved if we had more clinical training data.

4.1. 5FU

Our proposed MMPK-SciML model was able to predict the population volume for 5FU, which was impossible to obtain with FOCE-I due to convergence problems [29]. We observed that the predicted volume was often overpredicted, which could be due to outliers in the training data. We hypothesize that FOCE-I is less robust to outliers than the MMPK-SciML method and thus required the population volume to be fixed. This behavior illustrates one of the main advantages for using a SciML model for PK modeling, namely greater flexibility compared to traditional modeling approaches. However, a major limitation of our analyses was that the genotypes and the activity of the main metabolizing enzyme of 5FU, dihydropyrimidine dehydrogenase, which are important predictors for 5FU pharmacokinetics, were not available for our patient cohort. This information probably could have improved the performance of all tested models and should be reported in future studies.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.06.24306555; this version posted May 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

4.2. Sunitinib

For the sunitinib dataset we observed relatively wide confidence intervals of the MMPK-SciML estimates, while interestingly the predicted population parameters differed from those reported by Diekstra et al. [28]. We identified that although the absorption rate was predicted higher compared to Diekstra et al. (0,13/h vs. 0,31/h) and central volume of distribution was predicted lower (1820 L vs. 1352 L), the elimination and redistribution rates were similar across models in most of the cases. Especially large differences (> 40%) were observed in the estimates for the population parameters defining the concentration of the metabolite. To address that issue, we trained a model adding a proportional error between the metabolite concentration and its predictions in Eq. 4. However, this modification did not improve the performance of the model. The sunitinib analysis was more challenging than 5FU due to a small dataset composed of two different study populations increasing the variability. Moreover, more parameters had to be predicted due to the absorption process and the inclusion of metabolite concentrations, increasing the task complexity. However, considering that our model performed well despite these limitations, we expect that our MMPK-SciML method would produce more confident parameter estimates and better predictions if we had more training data.

5. Conclusion

This work shows the need to use a structural model to effectively capture the time course of plasma concentrations in patients. In this regard we proposed a novel hybrid machine learning framework, which combines the flexibility of modern neural network architectures with a compartmental model structure describing pharmacokinetic drug disposition. A limitation of our approach is the need for larger datasets compared to standard population PK modeling approaches. On the other hand, we offer the modeler the advantage that our approach does not require precisely specifying in which way PK parameters are influenced by covariates. This results in a simplification of the modeling process. A possible direction of future research is to incorporate our model architecture into more complex frameworks for dosage adjustment, e.g. via reinforcement learning.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.06.24306555; this version posted May 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

6. Study Highlights

- **What is the current knowledge on the topic?**

Machine Learning and Scientific Machine Learning (SciML) frameworks have shown promising results for pharmacokinetic modeling. However, methods for learning the inter-individual variability have not been widely investigated.

- **What question did this study address?**

How well do population pharmacokinetic (Pop-PK) and classical machine learning (ML) approaches perform in comparison to a SciML approach for PK modeling? Can a neural network be employed in a SciML framework to learn inter-individual variability while making accurate PK predictions?

- **What does this study add to our knowledge?**

The proposed MMPK-SciML model learns population PK parameters and their inter-individual variability and outperforms classical ML and Pop-PK approaches. Our approach also addresses common drug development challenges such as missing values and different time sampling.

- **How might this change clinical pharmacology or translational science?**

Our methodology focuses on comparing different machine learning methods in the PK field. Our final framework allows for the development of novel comprehensible and trustworthy strategies for individual dose adjustment.

7. Acknowledgements

We thank Ana Victoria Ponce Bobadilla for her input and explanations on how to create meaningful diagnostic plots for ML-enabled population PK models and Sven Stodtmann for the idea of encoding

medRxiv preprint doi: https://doi.org/10.1101/2024.05.06.24306555; this version posted May 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

the concentration-time trajectory into parameters. We also thank Ana Socorro Rodríguez Báez for proof-reading the manuscript and aiding with its structure.

8. Author contributions

U.J. and H.F. designed the research. D.V., O.T. and L.M.K. performed the research and analyzed the data., E.S. and A.F. developed the original population pharmacokinetic models, E.T. guided O.T. in coding the classical machine learning algorithms, and D.V., O.T., L.M.K, U.J. and H.F. wrote the manuscript.

Use of AI

The tool DeepL Write was used by OT to refine the text. The tool ChatGPT was used by OT alongside Stack Overflow to assist in the coding process.

Figure legends

Fig 1. MMPK-SciML overview. The mean and log variance of the patient eta´s distribution is predicted with a neural network. At the same time, the populational parameters are being learned and are used with an eta sample to define the patient-specific parameters, which are used with the patient dose regimen to predict the PK profile.

Fig 2. Goodness-of-fit (GOF) plots for the 5FU dataset showing predicted versus observed concentrations for all trained models.

Fig 3. Goodness-of-fit (GOF) plots for the sunitinib dataset showing predicted versus observed concentrations for all trained models.

Fig 4. Prediction-corrected Visual Predictive Checks (pcVPC) plots for the 5FU (top) and the sunitinib (bottom) dataset.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.06.24306555; this version posted May 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Supplemental Information titles

TableS1.docx **Table S1.** Description of data augmentation for 5FU and sunitinib

TableS2.docx **Table S2.** Feature selection for 5FU and sunitinib

TableS3.docx **Table S3.** Model hyperparameters for 5FU and sunitinib

TableS4.docx **Table S4.** Full cross validation metrics for 5FU and sunitinib datasets (Classic ML methods with Feature Selection)

TableS5.docx **Table S5.** Full cross validation populational parameters for 5FU and sunitinib datasets

FigS1.pdf **Fig S1.** PKPD Modell for sunitinib dataset

FigS2.pdf **Fig S2.** Posterior Eta distributions for 5FU and sunitinib datasets

AppendixA.docx Technical details

References

1. Imai S, Takekuma Y, Miyai T, Sugawara M. A New Algorithm Optimized for Initial Dose Settings of Vancomycin Using Machine Learning. *Biol Pharm Bull* 2020; 43(1): 188–93.

2. Hu PJ-H, Wei C-P, Cheng T-H, Chen J-X. Predicting adequacy of vancomycin regimens: A learning-based classification approach to improving clinical decision making. *Decision Support Systems* 2007; 43(4): 1226–41.

3. Tang J, et al. Application of Machine-Learning Models to Predict Tacrolimus Stable Dose in Renal Transplant Recipients. *Sci Rep* 2017; 7: 42192.

4. Lu J, Deng K, Zhang X, Liu G, Guan Y. Neural-ODE for pharmacokinetics modeling and its advantage to alternative machine learning models in predicting new dosing regimens. *iScience* 2021; 24(7): 102804.

5. You Dubout W. An Algorithmic Approach to Personalized Drug Concentration Predictions. Lausanne, EPFL; 2014.

6. Padmanabhan R, Meskin N, Haddad WM. Reinforcement learning-based control of drug dosing for cancer chemotherapy treatment. *Math Biosci* 2017; 293: 11–20.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.06.24306555; this version posted May 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

7. Yauney G, Shah P. Reinforcement Learning with Action-Derived Rewards for Chemotherapy and Clinical Trial Dosing Regimen Selection. In: Proceedings of Machine Learning Research; 161–226.

8. Yazdjerdi P, Meskin N, Al-Naemi M, Al Moustafa A-E, Kovács L. Reinforcement learning-based control of tumor growth under anti-angiogenic therapy. *Comput Methods Programs Biomed* 2019; 173: 15–26.

9. Ebrahimi Zade A, Shahabi Haghighi S, Soltani M. Reinforcement learning for optimal scheduling of Glioblastoma treatment with Temozolomide. *Comput Methods Programs Biomed* 2020; 193: 105443.

10. Mashayekhi H, Nazari M, Jafarinejad F, Meskin N. Deep reinforcement learning-based control of chemo-drug dose in cancer treatment. *Comput Methods Programs Biomed* 2024; 243: 107884.

11. Huo L, Tang Y. Multi-Objective Deep Reinforcement Learning for Personalized Dose Optimization Based on Multi-Indicator Experience Replay. *Applied Sciences* 2023; 13(1): 325.

12. Eastman B, Przedborski M, Kohandel M. Reinforcement learning derived chemotherapeutic schedules for robust patient-specific therapy. *Sci Rep* 2021; 11(1): 17882.

13. Treesatayapun C, Muñoz-Vázquez AJ, Suyaroj N. Reinforcement learning optimal control with semi-continuous reward function and fuzzy-rules networks for drug administration of cancer treatment. *Soft Comput* 2023; 27(22): 17347–56.

14. Alsaadi FE, Yasami A, Volos C, Bekiros S, Jahanshahi H. A New Fuzzy Reinforcement Learning Method for Effective Chemotherapy. *Mathematics* 2023; 11(2): 477.

15. Shiranthika C, Chen K-W, Wang C-Y, Yang C-Y, Sudantha BH, Li W-F. Supervised Optimal Chemotherapy Regimen Based on Offline Reinforcement Learning. *IEEE journal of biomedical and health informatics* 2022; 26(9): 4763–72.

16. Houy N, Le Grand F. Personalized oncology with artificial intelligence: The case of temozolomide. *Artif Intell Med* 2019; 99: 101693.

17. Maier C, Hartung N, Kloft C, Huisinga W, Wiljes J de. Reinforcement learning and Bayesian data assimilation for model-informed precision dosing in oncology. *CPT Pharmacometrics Syst Pharmacol* 2021; 10(3): 241–54.

18. Qian Z, Zame W, Fleuren L, Elbers P, van der Schaar M. Integrating Expert ODEs into Neural ODEs: Pharmacology and Disease Progression. In: Advances in Neural Information Processing Systems. Integrating Expert ODEs into Neural ODEs: Pharmacology and Disease Progression; 2021. Curran Associates, Inc; 11364–83.

19. Janssen A, Leebeek FWG, Cnossen MH, Mathôt RAA, the OPTI-CLOT study group, SYMPHONY consortium. Deep compartment models: A deep learning approach for the reliable prediction of time-series data in pharmacokinetic modeling. *CPT Pharmacometrics Syst Pharmacol* 2022; 11(7): 934–45.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.06.24306555; this version posted May 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

20. Valderrama D, Ponce-Bobadilla AV, Mensing S, Fröhlich H, Stodtmann S. Integrating machine learning with pharmacokinetic models: Benefits of scientific machine learning in adding neural networks components to existing PK models. *CPT Pharmacometrics Syst Pharmacol* 2024; 13(1): 41–53.

21. Baker N, et al. Workshop Report on Basic Research Needs for Scientific Machine Learning: Core Technologies for Artificial Intelligence; 2019.

22. Rackauckas C, Ma Y, Martensen J, *et al.* Universal Differential Equations for Scientific Machine Learning; 2020, arXiv preprint. http://arxiv.org/pdf/2001.04385.pdf2001.04385v4.

23. Beumer JH, et al. Multicenter evaluation of a novel nanoparticle immunoassay for 5-fluorouracil on the Olympus AU400 analyzer. *Ther Drug Monit* 2009; 31(6): 688–94.

24. Common Terminology Criteria for Adverse Events (CTCAE) v5.0. https://ctep.cancer.gov/protocolDevelopment/electronic_applications/ctc.htm#ctc_60 Accessed February 15, 2024.

25. Milano G, et al. Influence of sex and age on fluorouracil clearance. *J Clin Oncol* 1992; 10(7): 1171–5.

26. van der Zanden LFM, et al. Description of the EuroTARGET cohort: A European collaborative project on TArgeted therapy in renal cell cancer-GEnetic- and tumor-related biomarkers for response and toxicity. *Urol Oncol* 2017; 35(8): 529.e9-529.e16.

27. Mross K, et al. FOLFIRI and sunitinib as first-line treatment in metastatic colorectal cancer patients with liver metastases--a CESAR phase II study including pharmacokinetic, biomarker, and imaging data. *Int J Clin Pharmacol Ther* 2014; 52(8): 642–52.

28. Diekstra MH, et al. Population Modeling Integrating Pharmacokinetics, Pharmacodynamics, Pharmacogenetics, and Clinical Outcome in Patients With Sunitinib-Treated Cancer. *CPT Pharmacometrics Syst Pharmacol* 2017; 6(9): 604–13.

29. Schmulenson E, Zimmermann N, Müller L, Kapsa S, Sihinevich I, Jaehde U. Influence of the skeletal muscle index on pharmacokinetics and toxicity of fluorouracil. *Cancer Med* 2023; 12(3): 2580–9.

30. Yu H, et al. Integrated semi-physiological pharmacokinetic model for both sunitinib and its active metabolite SU12662. *Br J Clin Pharmacol* 2015; 79(5): 809–19.

31. Sibieude E, Khandelwal A, Girard P, Hesthaven JS, Terranova N. Population pharmacokinetic model selection assisted by machine learning. *J Pharmacokinet Pharmacodyn* 2022; 49(2): 257–70.

32. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Optuna; 2019. New York,NY,United States: Association for Computing Machinery; 2623–31.

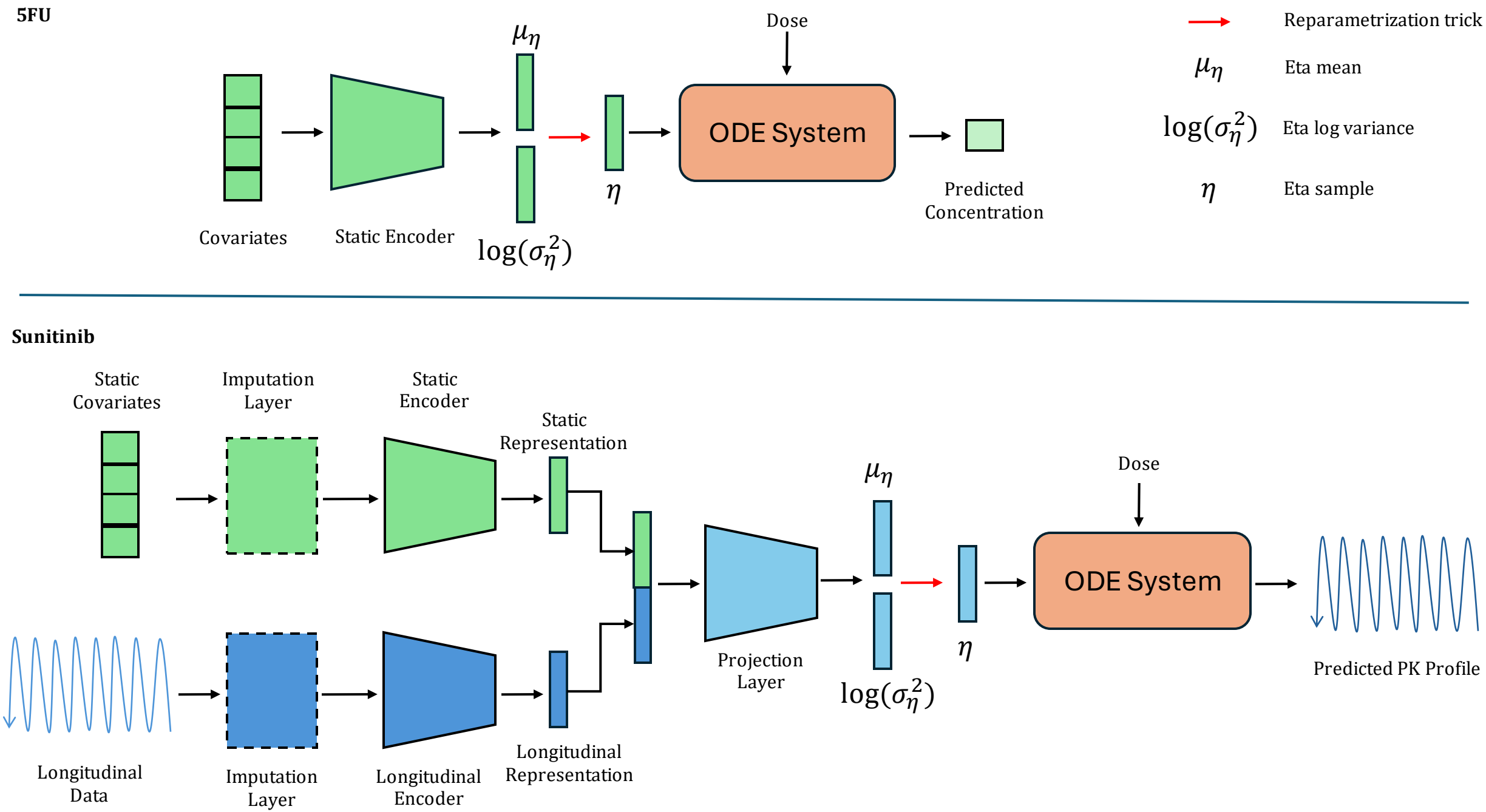
medRxiv preprint doi: https://doi.org/10.1101/2024.05.06.24306555; this version posted May 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

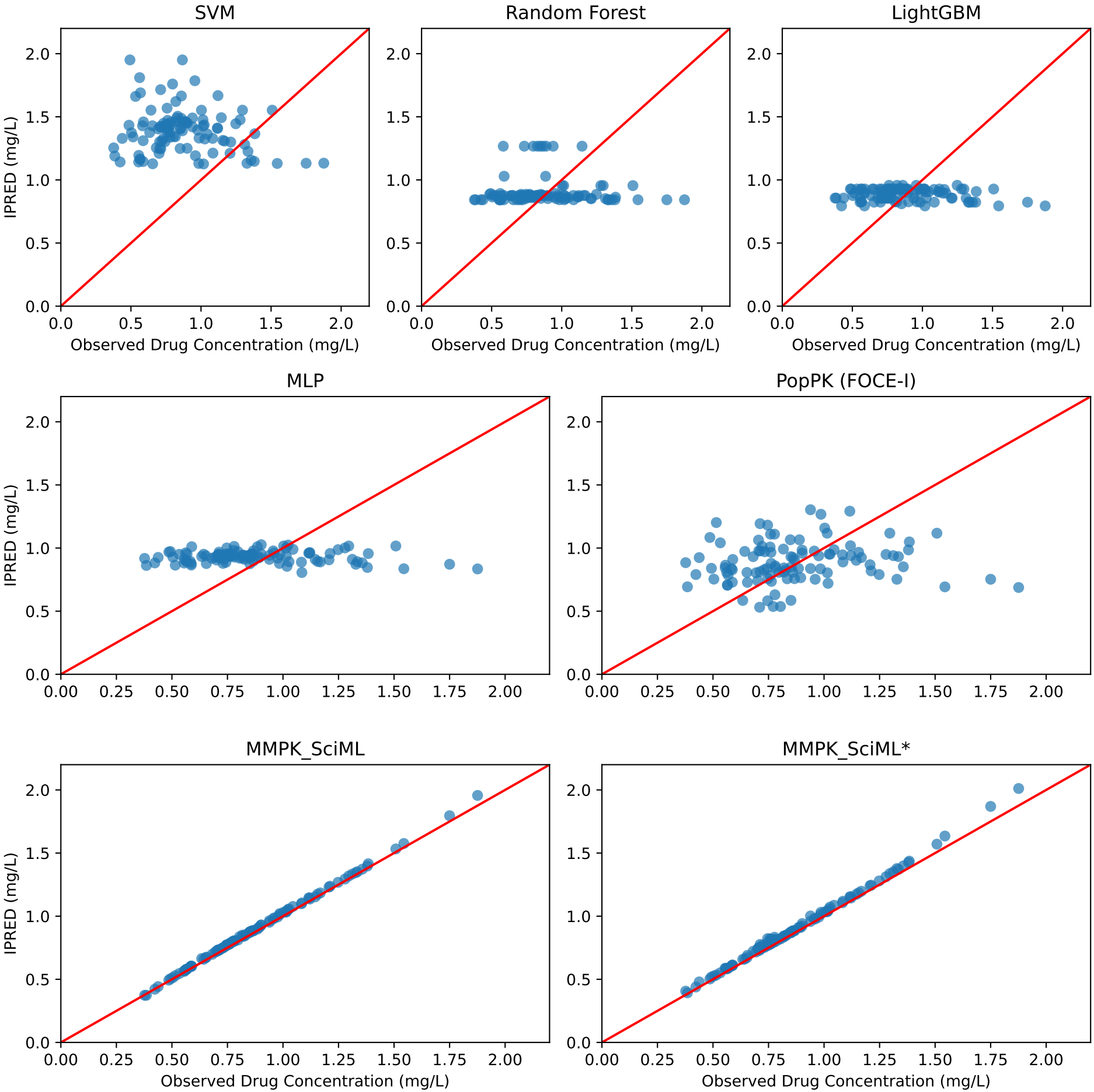
33. Kingma DP, Ba JL. Adam: A Method for Stochastic Optimization; 2014, arXiv preprint. http://arxiv.org/pdf/1412.6980.pdf1412.6980v9.

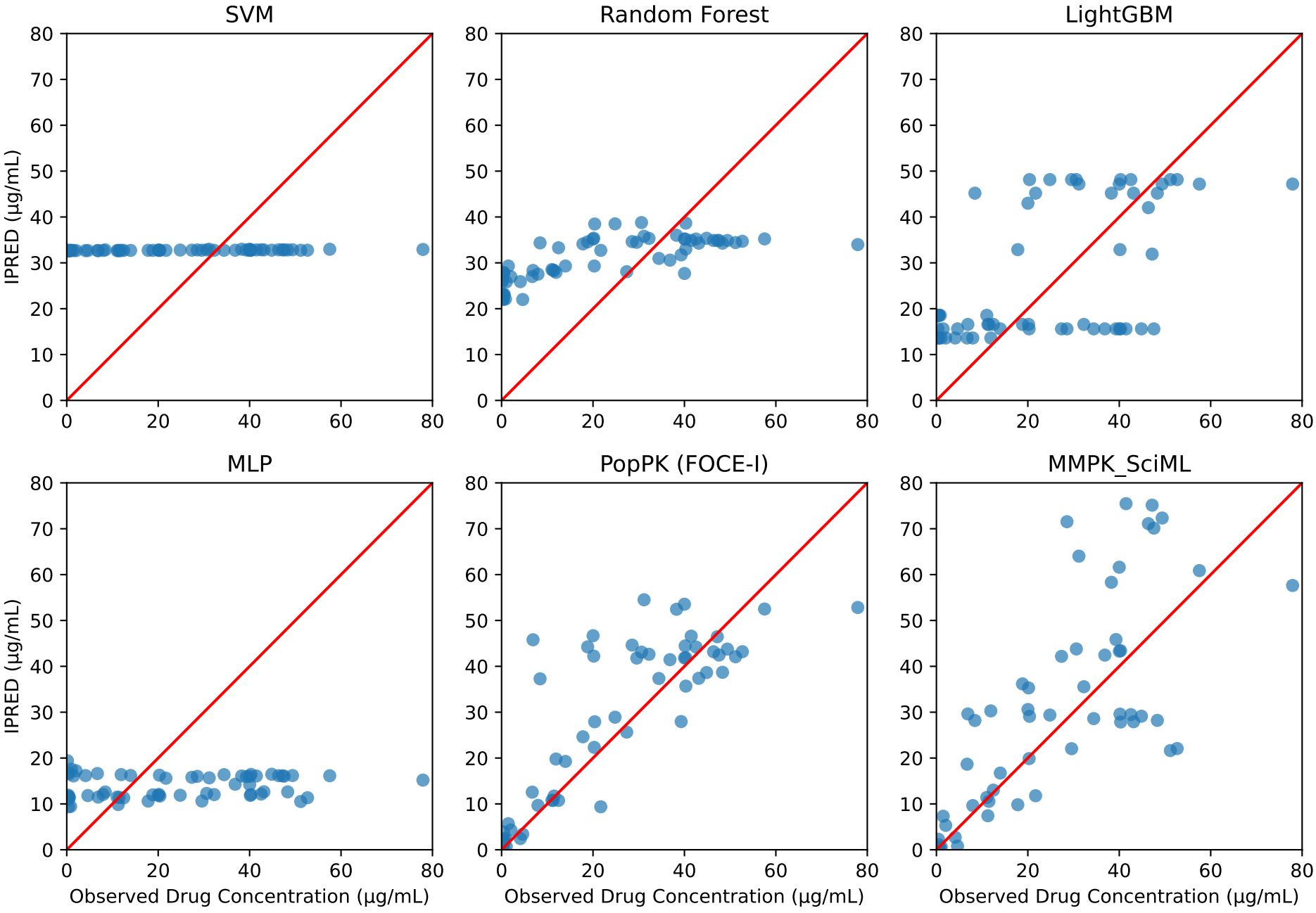
34. Kingma DP, Welling M. Auto-Encoding Variational Bayes; 2013, arXiv preprint. http://arxiv.org/pdf/1312.61141312.6114v11.

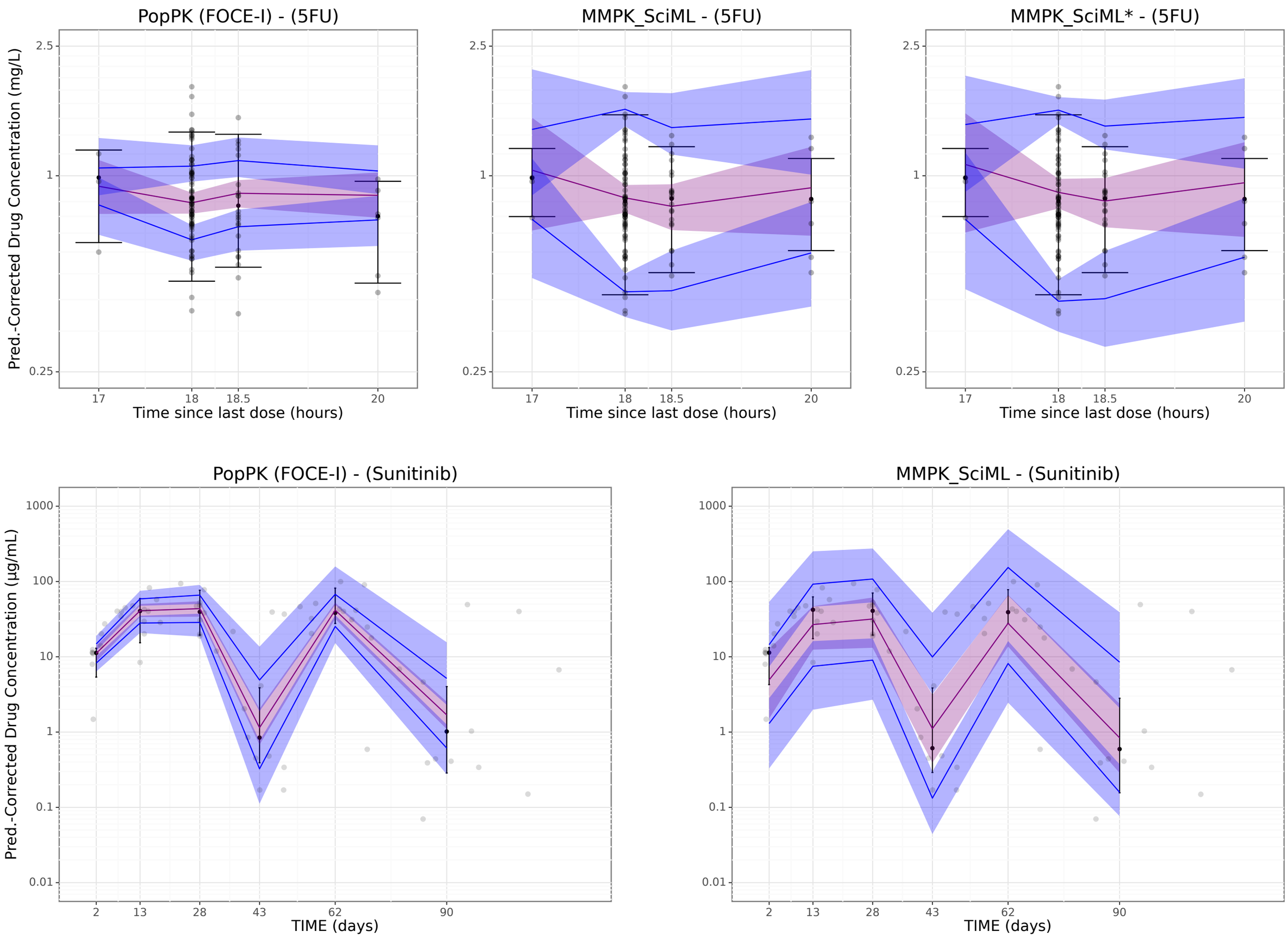
35. Baytas IM, Xiao C, Zhang X, Wang F, Jain AK, Zhou J. Patient Subtyping via Time-Aware LSTM Networks. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Patient Subtyping via Time-Aware LSTM Networks; 2017. New York, NY: ACM; 65–74.

36. Bräm DS, Parrott N, Hutchinson L, Steiert B. Introduction of an artificial neural network-based method for concentration-time predictions. *CPT Pharmacometrics Syst Pharmacol* 2022; 11(6): 745–54.


5FU
Covariates
Static Encoder
μ_η
$\log(\sigma_\eta^2)$
η
Dose
ODE System
Predicted Concentration
Reparametrization trick
μ_η Eta mean
$\log(\sigma_\eta^2)$ Eta log variance
η Eta sample
Sunitinib
Static Covariates
Imputation Layer
Static Encoder
Static Representation
Longitudinal Data
Imputation Layer
Longitudinal Encoder
Longitudinal Representation
Projection Layer
μ_η
$\log(\sigma_\eta^2)$
η
Dose
ODE System
Predicted PK Profile


SVM
Random Forest
LightGBM
MLP
PopPK (FOCE-I)
MMPK_SciML
MMPK_SciML*
IPRED (mg/L)
Observed Drug Concentration (mg/L)
0.0
0.5
1.0
1.5
2.0
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00


SVM
Random Forest
LightGBM
MLP
PopPK (FOCE-I)
MMPK_SciML
IPRED (μg/mL)
Observed Drug Concentration (μg/mL)
0
10
20
30
40
50
60
70
80


PopPK (FOCE-I) - (5FU)
MMPK_SciML - (5FU)
MMPK_SciML* - (5FU)
Pred.-Corrected Drug Concentration (mg/L)
2.5
1
0.25
17
18
18.5
20
Time since last dose (hours)
PopPK (FOCE-I) - (Sunitinib)
MMPK_SciML - (Sunitinib)
Pred.-Corrected Drug Concentration (µg/mL)
1000
100
10
1
0.1
0.01
2
13
28
43
62
90
TIME (days)

medRxiv preprint doi: https://doi.org/10.1101/2024.05.06.24306555; this version posted May 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table 1. Baseline patient characteristics (median and range), sunitinib data characteristics taken from [28]

5 FU	
Demographics	
Sex, M/F	97/60
Age (years)	64 (35-83)
Body surface area (m^2)	1,92 (1,35-2,85)
Therapy related details	
5FU dose (mg)	4000 (2700-5720)
5FU AUC (mg × h/L) [a]	18,7 (1,9-92,3)
Therapy regimen	
AIO [b]	48
FUFOX [c] (including monoclonal antibodies)	41
Paclitaxel/cisplatin/5FU/folinate	40
Other	28
Tumor entity	
Colorectal cancer	79
Gastroesophageal cancer	49
Pancreatic cancer	16
Other	13

medRxiv preprint doi: https://doi.org/10.1101/2024.05.06.24306555; this version posted May 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

	Sunitinib	
	Patients with mRCC (*n*=26) [d]	**Patients with mCRC (*n*=21)** [e]
Demographics		
Age	64 (43–75)	61 (33–85)
Sex, M/F	25/1	12/9
Weight (kg)	83 (65–106)	73 (57–106)
Height (cm)	180 (155–186)	172 (149–184)
BMI (kg/m^2) [f]	25,7 (22,5–34,5)	26,0 (13,3–39,3)

Notes:

a Calculated by multiplying the infusion time with the measured steady-state concentration.

b Weekly 5FU infusion (2600 mg/m2) over 24 h in combination with folinate (500 mg/m2).

c Weekly 5FU infusion (2000 mg/m2) over 24 h in combination with folinate (500 mg/m2) and oxaliplatin (50 mg/m2).

d,e,f (abbreviations): mRCC, metastasized renal cell carcinoma; mCRC, metastasized colorectal cancer; BMI, body mass index.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.06.24306555; this version posted May 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table 2. Cross validation average populational parameters for 5FU and Sunitinib datasets.

5FU					
Parameter	**Unit**	**FOCE-I (NM)**	**SAEM (NM)**	**MMPK-SciML***	**MMPK-SciML**
CL	L/h	216,1 ± 2,1	214,4 ± 2,80	211,38 ± 3,76	202,68 ± 28,16
V	L	46,1	46,1	46,1	38,78 ± 24,3
Sunitinib					
Parameter	**Unit**	**Diekstra et al [28]**	**FOCE-I (NM)**	**SAEM (NM)**	**MMPK-SciML**
KA	1/h	0,13	0,154 ± 0,04	0,148 ± 0,03	0,31 ± 0,02
QH^F	L/h	80	80	80	80
CL	L/h	33,9	33,42 ± 0,96	34,47 ± 1,24	35,52 ± 0,1
Q	L/h	0,37	0,36 ± 0,04	0,35 ± 0,04	0,41 ± 0,03
V2 (V2)	L	1820	1863 ± 64,64	1883 ± 85,33	1351,68 ± 6,94
$V3^F$ (V5)	L	588	588	588	588
CL_M	L/h	16,5	16,44 ± 0,43	17,13 ± 0,77	10,72 ± 0,21
Q_M	L/h	2,75	3,18 ± 0,28	2,77 ± 0,24	12,59 ± 0,24
F_M^F	-	0,21	0,21	0,21	0,21
$V2_M$ (V3)	L	730	678,1 ± 38,85	725,8 ± 80,63	393,95 ± 7,95
$V3_M$ (V4)	L	592	618 ± 35,05	624,8 ± 33,03	240,88 ± 4,85

Notes: ^{F} Fixed parameter. MMPK-SciML* refers to the model with a fixed volume for 5FU.

medRxiv preprint doi: https://doi.org/10.1101/2024.05.06.24306555; this version posted May 8, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table 3. Cross validation average metrics for 5FU and Sunitinib datasets

	5FU		Sunitinib	
Model	**MAE**	**RMSE**	**MAE**	**RMSE**
Suport Vector Machine	0,49 ± 0,06 (0,43 ± 0,06)	0,61 ± 0,1 (0,55 ± 0,03)	20,01 ± 2,66 (21,51 ± 3,51)	23,84 ± 3,41 (27,37 ± 5,27)
Random Forest	0,23 ± 0,03 (0,23 ± 0,03)	0,33 ± 0,12 (0,33 ± 0,12)	17,81 ± 2,59 (17,13 ± 3,08)	21,34 ± 3,26 (21,30 ± 4,82)
LightGBM	0,23 ± 0,03 (0.23 ± 0,03)	0,32 ± 0,12 (0,33 ± 0,12)	**13,99 ± 3,0 (19,38 ± 3,70)**	**16,86 ± 3,40 (24,43 ± 5,58)**
Multi-Layer Perceptron Two Hidden Layers	0,24 ± 0,04 (0,22 ± 0,03)	0,33 ± 0,12 (0,32 ± 0,12)	21,15 ± 4,85 (14,22 ± 3,56)	25,22 ± 5,96 (17,71 ± 4,92)
PopPK (FOCE-I)	**0,23 ± 0,03**	**0,32 ± 0,12**	**9,57 ± 2,63**	**13,90 ± 3,72**
MMPK-SciML	**0, 034 ± 0,02**	**0,105± 0,159**	**12,31 ± 3,51**	**17,99 ± 5,15**
MMPK-SciML *	**0,038 ± 0,02**	**0,09 ± 0,11**	-	-

Notes: For the classical machine learning methods, we depict the results of the versions with feature selection. In green, blue and purple we show the best, second-best and third-best model for each dataset. In brackets, we show the results with data augmentation. MMPK-SciML* refers to the model with a fixed volume for 5FU. Metrics are reported when using the sampling from the patient specific distribution for test subjects.