Leveraging QSP Models for MIPD: A Case Study for Warfarin/INR

Undine Falkenhagen[1,2], Larisa H. Cavallari[3], Julio D. Duarte[3], Charlotte Kloft[4], Stephan Schmidt[5] and Wilhelm Huisinga[2,*]

Warfarin dosing remains challenging due to substantial inter-individual variability, which can lead to unsafe or ineffective therapy with standard dosing. Model-informed precision dosing (MIPD) can help individualize warfarin dosing, requiring the selection of a suitable model. For models developed from clinical data, the dependence on the study design and population raises questions about generalizability. Quantitative system pharmacology (QSP) models promise better extrapolation abilities; however, their complexity and lack of validation on clinical data raise questions about applicability in MIPD. We have previously derived a mechanistic warfarin/international normalized ratio (INR) model from a blood coagulation QSP model. In this article, we evaluated the predictive performance of the warfarin/INR model in the context of MIPD using an external dataset with INR data from patients starting warfarin treatment. We assessed the accuracy and precision of model predictions, benchmarked against an empirically based reference model. Additionally, we evaluated covariate contributions and assessed the predictive performance separately in the more challenging outpatient data. The warfarin/INR model performed comparably to the reference model across various measures despite not being calibrated with warfarin initiation data. Including *CYP2C9* and/or *VKORC1* genotypes as covariates improved the prediction quality of the warfarin/INR model, even after assimilating 4 days of INR data. The outpatient INR exhibited higher unexplained variability, and predictions slightly exceeded observed values, suggesting that model adjustments might be necessary when transitioning from an inpatient to an outpatient setting. Overall, this research underscores the potential of QSP-derived models for MIPD, offering a complementary approach to empirical model development.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ Warfarin dose requirements differ substantially between individuals. Model-informed precision dosing (MIPD) can improve individualized dosing. Typically, MIPD is performed using empirically based models.

WHAT QUESTION DID THIS STUDY ADDRESS?

☑ Can we leverage the mechanistic knowledge in quantitative systems pharmacology (QSP) models for MIPD for warfarin?

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

☑ A mechanistic warfarin/INR model derived from a blood coagulation QSP model, without calibration to clinical data, performed comparably to an empirical reference model for early INR prediction in an external evaluation. Including CYP2C9 and VKORC1 genotypes as covariates substantially enhanced the prediction. Model adjustments might be necessary when transitioning from inpatient to outpatient data.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

☑ QSP-derived modeling offers a complementary approach to empirical modeling to develop PK/PD models for MIPD more systematically, more mechanistically, and more transparently and with less dependence on specific clinical data.

Warfarin, despite its widespread and long-term use as an anticoagulant, poses challenges to clinicians due to significant variability in individual dose requirements. The warfarin effect is often quantified by the prothrombin time, typically expressed as the international normalized ratio (INR), with a target range of 2–3 for many indications. This target range is narrow in comparison with the interindividual variability (IIV), leading to more than 10-fold differences in required maintenance dose.[1] In recent trials

[1]PharMetrX Graduate Research Training Program, Berlin/Potsdam, Germany; [2]Institute of Mathematics, Mathematical Modelling and Systems Biology, University of Potsdam, Potsdam, Germany; [3]College of Pharmacy, Department of Pharmacotherapy and Translational Research, University of Florida, Gainesville, Florida, USA; [4]Institute of Pharmacy, Department of Clinical Pharmacy and Biochemistry, Freie Universität Berlin, Berlin, Germany; [5]College of Pharmacy, Department of Pharmaceutics, Center for Pharmacometrics and Systems Pharmacology, University of Florida, Orlando, Florida, USA. *Correspondence: Wilhelm Huisinga (huisinga@uni-potsdam.de)

Received January 29, 2024; accepted April 5, 2024. doi:10.1002/cpt.3274

with genotype-guided algorithms, the mean time within target INR range was only 54.7% and 67.4%.[2,3] A high probability of INRs outside the target range can result in an increased risk of bleeding or ineffective therapy.

To guide individual dose selection, especially during the dose-finding phase, model-informed precision dosing (MIPD) algorithms have been employed. One such example is regression algorithms, which predict the dose based on covariates such as weight, age, cytochrome P450 isoenzyme 2C9 (*CYP2C9*) genotype or vitamin K epoxide reductase complex 1 (*VKORC1*) genotype.[4,5] Although these models succeed at explaining parts of the observed variability, additional unexplained IIV remains, thus preventing a priori determination of the optimal individual dose based on covariates alone. In addition to stratifying by covariates, pharmacokinetics/pharmacodynamics (PK/PD)-based algorithms can help to update the optimal dose estimate based on early INR measurements.[6–8] When correctly accounting for the delayed effect, early measurements from the non-steady-state phase can inform the PK/PD model predictions of the steady-state INR.

There is a large variety in PK/PD models that model the warfarin/INR relationship.[6–9] Empirically based PK/PD models depend heavily on the underlying population and study design. Without access to the underlying data, it is difficult to compare the models and assess whether the results can be extrapolated beyond the population used for model development. In a previous article, we proposed to use model reduction in QSP models to derive mechanistic PK/PD models as an alternative to the standard empirical model development based on data.[10] This model development approach makes underlying assumptions explicit and provides a basis for data and hypothesis evaluation and discussion. Using this approach, we have previously derived a mechanistic warfarin/INR model from a QSP model of blood coagulation.[10]

To assess the warfarin/INR model for applicability in MIPD, this article addresses three clinically relevant questions. We first assessed the accuracy and precision of early prediction of INR data from a diverse patient cohort in an external evaluation. We then considered which covariates are most informative for the INR prediction, distinguishing between a priori and early prediction after 4 days. We assessed the dose optimization by comparing predicted optimal doses and doses at discharge. Finally, we assessed how well the prediction extrapolates to outpatient data. For the INR prediction within the MIPD algorithm, we employed a full Bayesian method implemented using a particle filter.[11] The prior knowledge encoded in the mechanistic warfarin/INR model and the covariate information is combined with early INR data to yield a posterior INR distribution. Assessing the uncertainty associated with the INR prediction allows us to evaluate clinically relevant properties such as the probability of achieving a steady-state INR within the target range.

METHODS

A robust and mechanistic understanding is essential to advance MIPD for warfarin. While we previously derived a reduced warfarin/INR model from a QSP framework,[10] its clinical application requires thorough validation. To this end, we evaluated the predictive performance of the warfarin/INR model using an external dataset from an ethnically diverse cohort of patients initiating warfarin,[12] which was not used in the model development. Our assessment leveraged a full Bayesian method, enabling a comprehensive evaluation of the model's precision and accuracy. The analysis was implemented in MATLAB 2021a, and the model code is accessible from https://doi.org/10.5281/zenodo.10844967.

Diverse study population

The data were previously collected in a prospective study to assess pharmacogenetic dosing in a real-world setting and include 258 adults initiating warfarin treatment.[12] The dataset comprises demographic, pharmacogenetic, and clinical data. No PK data were available. The data consisted of 0–10 measurements from an inpatient setting and 0–10 measurements from an outpatient setting per patient. For patients with both inpatient and outpatient data, the outpatient data were collected after the patient was discharged from the hospital. For inpatient data, the daily warfarin dose was reported; for outpatient data, the average daily dose in the week before the respective INR measurement was available. This value was used in the simulation as described in **Supplementary Materials Section S1**. Patients were genotyped for *VKORC1* c.1639G>A; *CYP2C9**2, *3, *5, *6, *8 and *11; *CYP4F2* p.V433M and a novel polymorphism in the CYP2C cluster (rs12777823).[13] For details on data cleaning, see **Supplementary Materials Section S2**.

To evaluate the predictive performance in the inpatient data, we divided the dataset into training data used to compute individual posterior predictions (Days 0 to 4) and evaluation data (Days 5 to 9). This separation was chosen to balance between accumulating sufficient data for informed predictions and retaining sufficient data for subsequent evaluations. We thus excluded patients from this analysis if they had no INR data on Day 5 or later. Out of 640 INR values from 81 patients fulfilling the conditions, 392 were early INR values used for individual predictions, and 248 were late INR values used for evaluation. To assess prediction quality in outpatient data, all patients with outpatient data were considered, which left 149 patients with 554 inpatient and 1,070 outpatient INR values (50 of these patients have no inpatient data). **Table 1** shows baseline characteristics of the two partly overlapping patient populations of (i) patients with inpatient INRs after at least 5 days (for the inpatient analysis) and (ii) patients with any outpatient data (for the outpatient analysis).

Previously developed mechanistic warfarin/INR model

We previously derived a small-scale warfarin/INR model[10] (**Figure 1a**) by reducing a blood coagulation QSP model[14] with assumed variability on all parameters, which we summarize in this section. The model reduction ensures the reduced model inherits biological interpretability from the QSP model while approximating the warfarin/INR relationship of the QSP model for realistic populations. Therefore, a virtual population was created by assuming a distribution across different *CYP2C9* genotypes (different clearance values) and different *VKORC1* genotypes (different IC_{50} values),[9] as well as random variations in parameter values, realized by a multiplicative log-normal distribution with 40% CV. The model reduction was based on model order reduction using sensitivity-based input-response indices[15] and model simplification, including simplified rate reactions and analytical solutions of sub-models.[10]

The QSP model in Ref. [14] describes the time courses of important coagulation factors *in vivo* as well as the *in vitro* blood coagulation test to determine the prothrombin time (PT), from which the INR is obtained by normalization. It models the extrinsic and intrinsic pathways, explicitly including activation, complex formation, reduction, and oxidation of coagulation factors, in addition to stimulation, production,

Table 1 Baseline characteristics of patients with (i) late inpatient INRs (on or later than day 5), which were used in the inpatient analysis and (ii) patients with any outpatient data, which were used in the outpatient analysis. The two datasets partly overlap, with 43 patients in both datasets

Characteristics	Patients with inpatient data at Day 5 or later ($n=81$)	Patients with outpatient data ($n=149$)
Age (years), mean ± SD	52 ±14	52 ±17
Weight (kg), mean ± SD	97 ±37	92 ±29
Female sex, %	40.7	46.7
Baseline INR, mean ± SD	1.19 ±0.19	1.18 ±0.19
Self-reported race/ethnicity, *N* (%)		
Black or African American	41 (50.6)	86 (57.7)
White	14 (17.3)	13 (8.7)
Hispanic	11 (13.6)	36 (24.2)
Other	15 (18.5)	14 (9.4)
Patients with variant alleles, *N* (%)		
CYP2C9 *2	9 (11.1)	15 (10.0)
CYP2C9 *3	2 (2.5)	5 (3.4)
CYP2C9 *5	0 (0)	2 (1.3)
CYP2C9 *6	2 (2.5)	3 (2.0)
CYP2C9 *8	5 (6.2)	6 (4.0)
CYP2C9 *11	1 (1.2)	2 (1.3)
VKORC1 c.1639G>A	29 (35.8)	61 (40.9)
CYP4F2	27 (33.3)	37 (24.8)
rs12777823	27 (33.3)	46 (30.9)

degradation, and inhibition of degradation reactions. The QSP model also models details of the vitamin K cycle for simulating the response to warfarin therapy. The PT is calculated using a threshold on the cumulative fibrin concentration after activation by the tissue factor.[14] The reduced model still contains the key components of the blood coagulation cascade to predict the effect of warfarin on the INR. These are the coagulation factors II, VII, and X in addition to the vitamin K hydroquinone (VKH2). *In vivo*, warfarin inhibits VKH2, a stimulator of synthesis of coagulation factors II, VII, and X. *In vitro*, the product of these three coagulation factor concentrations determines the PT and the INR. The coagulation factors are indirectly inhibited by warfarin, that is, warfarin acts on VKH2, which in turn acts on the synthesis rates of the coagulation factors. This induces a delay between PK and PD, comparable to a turnover model in classical PK/PD modeling. See **Supplementary Materials Section S11** for additional insights into the modeled delay and its dependence on genotypes. See **Supplementary Materials Section S3** for a complete model definition.

To evaluate the predictive quality of our model, we compared its performance to an established empirical reference model developed by Hamberg et al.[9] (Figure 1b). In the reference model, the warfarin effect on the INR is modeled using two transit compartment chains that account for the delayed effect and an INR equation that translates the terminal compartment concentrations into the INR.

Prior uncertainty in the mechanistic warfarin/INR model

To use the mechanism-based warfarin/INR model in a Bayesian setting, a prior uncertainty must be defined for all parameters. As we want to predict the INR for individual patients, the individual prior predictive distribution depends on the covariate information of the patient. Given the covariate information, the prior parameter uncertainty for any single patient before the start of the treatment equals the IIV from the population setting. The prior INR uncertainty for a single patient additionally depends on the residual unexplained variability (RUV). For mathematical details, see **Supplementary Materials Section S4**.

We considered age, weight, *CYP2C9* genotype, and rs12777823 (only in patients who self-identified as Black or African American) as covariates affecting the PK.[9,13,16] INR baseline, *VKORC1*, and *CYP4F2* (affecting Vitamin K metabolism) genotypes were considered covariates affecting the PD.[9,17] Covariate relationships and parameters were taken from the literature; for details, see **Supplementary Materials Section S5**. Covariates were included based on prediction quality improvements and difficulty of obtaining the information. For IIV, we used a lognormal distribution with 40% CV on all parameters in the warfarin/INR model as in the model reduction.[10] During the model reduction in the previous article, a deliberately high IIV was chosen to ensure a good approximation in a diverse population; however, a smaller IIV might suffice to model a specific patient population. We thus alternatively considered a more realistic IIV (with CV between 10 and 40%) on the different parameters if literature data were available; see **Supplementary Materials Section S6**. For RUV, we used a lognormal distribution with 20% CV on the INR.[9] Compared with empirical models, a mechanistic model with IIV on all parameters might explain more variability and thus leave a smaller RUV. Thus, we alternatively considered an RUV with 18% CV.

Individual posterior prediction and dose optimization

We used a Bayesian statistical approach to infer individual parameters from patient data and the prior, and used posterior predictions to forecast individual therapy. INR data until Day 4 (the data horizon) were assimilated for each patient to obtain their posterior parameter distribution (see **Supplementary Materials Section S4**) using the particle filter approach presented by Maier et al.[11] The particle filter is an efficient simulation-based method for Bayesian updating based on sequential data.[11] For each patient, 1,000 parameter realizations (the "particles") with the same covariates were drawn from the prior parameter distribution (the IIV). They were used to simulate the prior INR distribution, which additionally includes the RUV. Then, the particles were weighted based on the likelihood of the sequentially considered data points given this parameter realization. The result is a virtual population with parameters distributed according to the posterior parameter distribution.[11] Then, the predictions made with the posterior were compared with the actual data on Days 5 to 9 for evaluation. For the outpatient analysis, each INR prediction was performed by assimilating all INRs up to the previous visit.

An optimal daily dose maximizes the probability of the INR being within the target window at steady state. Depending on time and dosing history, the probability of being within the target window can be calculated from the posterior predicted INR distribution. Using the posterior parameter distribution, based on data up to Day 4, and the RUV, we simulated the posterior INR distribution, assigning a dose from a range between 0.5 mg and 20 mg after the reported inpatient doses (starting latest at Day 10). We chose the dose with the highest predicted probability of the INR being within the target window on Day 30, at which point the INR was expected to be in steady state. The same particle filter method with data up to Day 4 and the same dose optimization was used with the reference model.

Measures for prediction quality

To assess the model quality, we evaluated the quality of the INR predictions and compared the predicted optimal dose with the actually prescribed doses. Both accuracy and precision of INR predictions are essential for dose optimization.

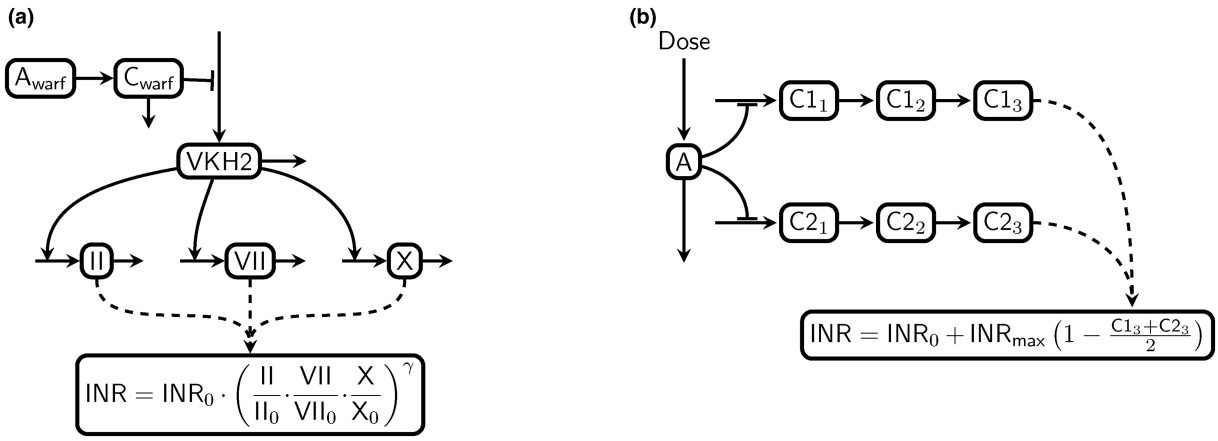

Figure 1 (**a**) Mechanistic warfarin/INR model. Ordinary differential equations model the warfarin effect on coagulation factors II, VII, and X. The INR is calculated in terms of these coagulation factors. (**b**) Empirically based reference model.[9] Warfarin affects two transit chains whose terminal compartments are translated into the INR. A_{warf}, amount of warfarin in absorption compartment; C_{warf}, warfarin concentration in central compartment; VKH2, vitamin K hydroquinone; II_0, pre-treatment concentration of factor II; VII_0, pre-treatment concentration of factor VII; X_0, pre-treatment concentration of factor X.

We employed a prediction-corrected visual predictive check (pcVPC)[18] to visually interpret the prediction quality. A pcVPC normalizes all predictions and data around the median prediction in order to make data resulting from differing covariates or dosing histories comparable. To quantitatively measure the INR prediction accuracy, we compared the INR measurements obtained between Days 5 and 9 with the corresponding INR predictions, which incorporated the individual's INR data from Days 1 to 4; see **Figure 2a** for an illustration. For the outpatient analysis, we compared outpatient INR measurements with the corresponding INR predictions that incorporated all INR data up to the previous visit. For a comprehensive evaluation of the predicted distribution, instead of relying solely on point predictions, we calculated the proportion of the INR measurements falling within their respective 90% prediction intervals. The 90% prediction intervals are, for a specific patient and a specific time point, defined as the range between the 5th and 95th percentile of the posterior distribution for the INR including RUV. We examined INR prediction precision by assessing the probability (averaged over the population) of the INR falling within the target window at the steady state, given the calculated optimal dose; see **Figure 2b** for an illustration. The accuracy and precision are evaluated for the same posterior distribution. Notably, the precision of predicted INR_{ss} is meaningful only when accuracy is good since it is solely derived from the prediction. As we expected the accuracy and precision to depend on the considered population, we performed case resampling bootstrapping to assess the uncertainty. We drew 1,000 populations ($n = 81$) from the actual population via drawing with replacement, obtaining 1,000 accuracy and precision values. The 90% confidence

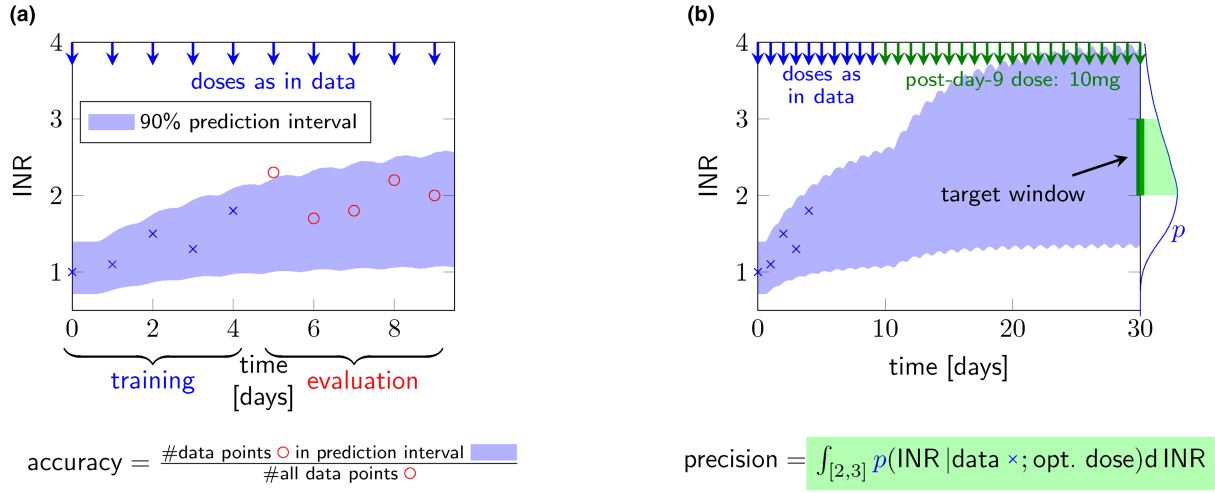

Figure 2 Accuracy and precision measures illustrated for a single patient. (**a**) As a measure for prediction accuracy, we calculate the coverage probability of the 90% prediction interval. Given INR data (blue cross) until Day 4, the posterior prediction (blue area) is simulated. The coverage probability is the share of evaluation data points (red circles) lying within their respective prediction interval. (**b**) To measure the prediction precision, we simulate the predicted probability of an INR within the target window at Day 30 given optimal dose. Doses until Day 9 were taken from the data, INR data until Day 4 was assimilated. The prediction for a dose of 10 mg is shown for illustration. The posterior predictive distribution $p(\cdot \mid \text{data}; \text{post-day-9 dose})$ of the INR prediction at Day 30 (blue line) is shown, and the probability of an INR within the target window is colored (green area). The optimal dose was defined as the post-day-9 dose that maximized the precision and was determined by simulating the posterior INR distribution for a range of doses between 0.5 mg and 20 mg given daily after Day 9. In the actual data, the doses until Day 9 typically vary, but in this illustration, a constant dose is assumed for the first 9 days for simplicity.

intervals constructed with the percentile method are shown in **Figure 4**, in addition to the results for the actual population.

Finally, we wanted to evaluate the dose optimization. While the actual optimal dose is unknown, the dose at discharge represents the dose considered optimal or acceptable by the prescribing physician at the time of patient discharge. Thus, we compared the predicted optimal doses (based on INR data up to Day 4) with the discharge doses.

RESULTS

We employed a full Bayesian approach to optimize doses in an MIPD setting using a mechanistic warfarin/INR model. INR_0, age, weight, *CYP2C9*, and *VKORC1* were chosen as covariates in the final model; see **Section Covariate contributions to prediction quality** for details. Individual INR measurements were assimilated to generate posterior predictive distributions.

Model fit and prediction quality

Figure 3 shows the prediction-corrected visual predictive check (pcVPC)[18] showing the 10th, 50th, and 90th percentile of the data together with the simulation's 95% confidence intervals of these percentiles. As inferred from the *Model Performance* pcVPCs (top), the posterior INR simulations based on all individual data describe the data well for both the mechanistic warfarin/INR model and the reference model. To assess the model's ability to predict future INR data, posterior predictions were generated using individual INR data until Day 4. Subsequently, these predictions were contrasted with actual data from Day 5 onwards, constituting an external validation. **Figure 3** (bottom) presents the corresponding pcVPCs, providing insights into predictive performance on unseen data. The pcVPCs of our model and the reference model look similarly good. The lower data percentiles align well with the lower prediction confidence intervals for both models and data horizons. The data median and upper data percentiles are initially below the prediction confidence intervals but eventually align within them. Regardless of the data horizon, both models show a slight overprediction initially; in other words, the INR increase is modeled to be faster than what is observed in the data.

The accuracy and precision measures detailed in **Section Measures for prediction quality** assess the quality of future data predictions. **Figure 4a** presents these measures, along with uncertainty obtained from bootstrapping, for various variability hyperparameters for IIV and RUV (**Section Prior uncertainty in the mechanistic warfarin/INR model**) in comparison with the reference model. For an accurate prediction, close to 90% of the data points should be within the 90% prediction interval. A high predicted probability of being within the target window at steady state indicates high precision. In general, we can observe a trade-off between accuracy and precision for different hyperparameters. We considered accuracy more important, as precision is solely derived from the model predictions and, therefore, only meaningful if the predictions are accurate. A large uncertainty in accuracy is observed, which is due to the strong dependence of the accuracy measure on single data points. The precision depends on the predictions and not directly on the data and thus is less sensitive to the bootstrapped population. We also assessed whether the prediction quality varies for specific subpopulations, particularly patients who self-identified as Black or African American, who are well represented in the dataset. When considering the *5, *6, *8, and *11 alleles in addition to *2 and *3, we would expect a similar prediction performance for the African American subpopulation as when considering *2 and *3 only in European ancestry populations, while the failure to test for CYP2C9 variants *5, *6, *8, and *11 is related to poor dose prediction in African Americans.[19,20] In this study, these variants are tested for and the prediction quality for patients who self-identified as Black or African American does not significantly differ from that of the entire population. Overall, the mechanistic warfarin/INR model and the reference model have comparable accuracy and precision.

Covariate contributions to prediction quality

Covariates can partly explain the large variability in warfarin response and thus enhance dose selection, especially at the treatment start. We investigated how integrating covariates mechanistically into the model impacted prediction quality. Data were available on baseline INR, age, weight, *CYP2C9* genotype, *VKORC1* genotype, *CYP4F2* genotype, and rs12777823 genotype. *CYP4F2* was not informative beyond baseline INR changes (see **Supplementary Materials Section S9**) and thus is not further considered.

Figure 4b shows accuracy and precision for the mechanistic warfarin/INR model with different subsets of the covariates. Prior predictions are shown in green and posterior predictions with INR data up to Day 4 in violet. In general, with more information, the prediction quality improves in both accuracy and precision. At the start of treatment, the dose selection depends solely on the covariates. In this case, including *VKORC1* or *CYP2C9* improves accuracy and precision over using only clinical covariates (INR_0, age, and weights). Incorporating rs12777823 for patients who self-identified as Black or African American, however, did not notably improve the prediction, possibly due to the small number of patients affected (only 8 patients who self-identified as Black or African American had a variant allele). In summary, as expected, incorporating genotype information enhances dosing recommendations when individual INR data are absent. When assimilating INR data up to Day 4, the inclusion of genotype information still notably enhances precision; however, it does not improve the already high accuracy of predictions. The precision enhancement could be explained by the covariates bringing prior predictions within the correct range, allowing the INR data to refine them further. The inclusion of *CYP2C9* appears to have a more pronounced positive impact than *VKORC1*, possibly due to the *VKORC1* effect being observable earlier in the INR than the *CYP2C9* effect. It has been observed before that the INR takes longer to respond in patients with specific *CYP2C9* variant alleles.[6,21] The addition of rs12777823 genotype for patients who self-identified as Black or African American does not further improve the prediction after INR data until day 4. In summary, incorporating genotype information enhances confidence in attaining the target window with selected doses, even after assimilating INR data up to Day 4. Based on these results, we used the following covariates in the remainder of this article. INR_0, age, and weight were included, given

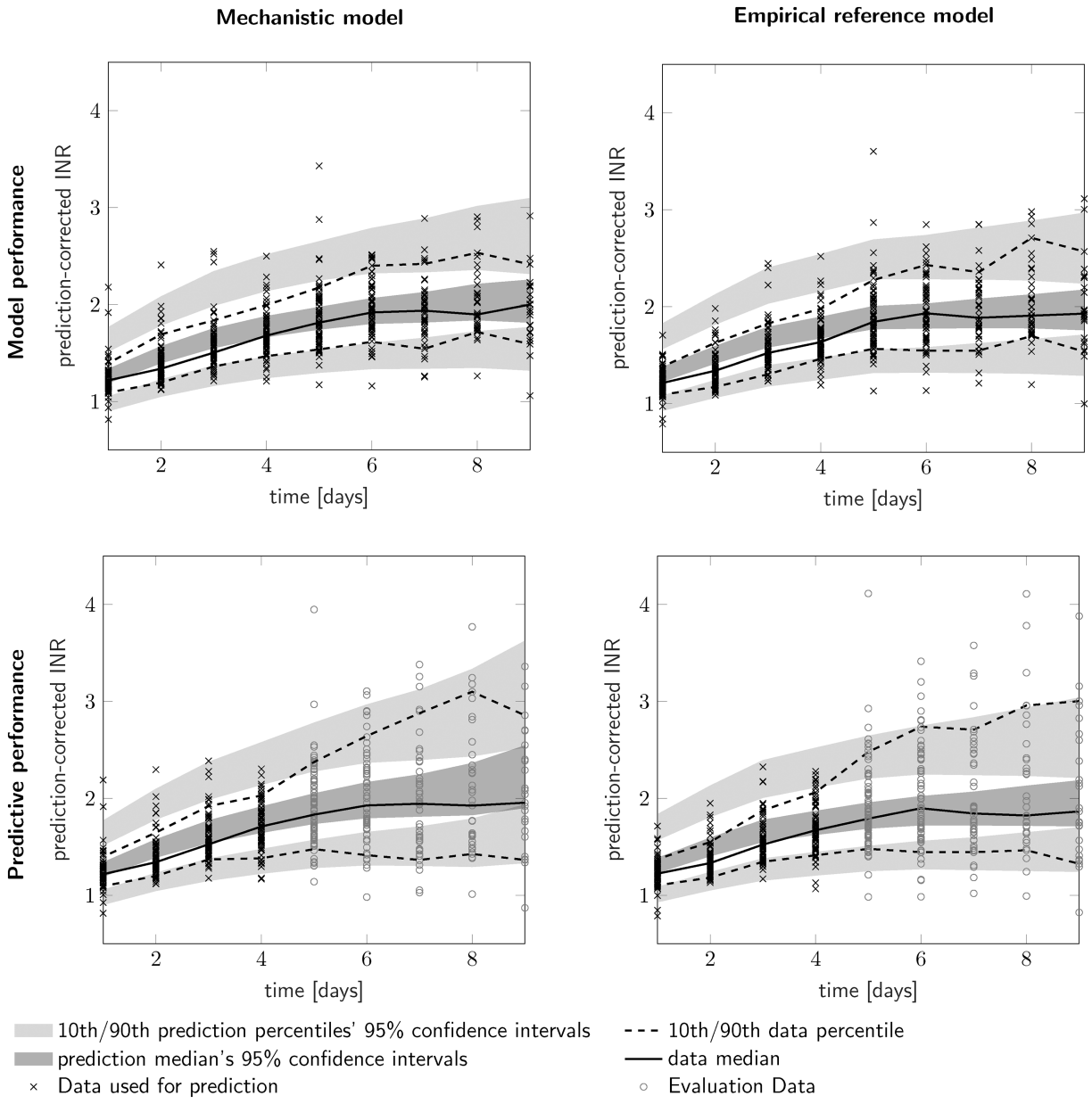

Figure 3 Prediction-corrected visual predictive checks (pcVPCs) for model performance and predictive performance. Left: mechanistic warfarin/INR model; Right: empirical reference model. Top: Posterior simulation of INR data based on individual INR data up to Day 9. Bottom: Posterior prediction of INR data for Days 5–9 based on individual INR data up to Day 4. The same data are shown in all panels.

their accessibility and solid mechanistic foundation. Additionally, *CYP2C9* and *VKORC1* genotypes were included, as they substantially enhanced prediction quality both pre-treatment and after 4 days. However, we excluded the rs12777823 genotype as a covariate in the model, due to its lack of contribution to prediction improvement within this data set and its relatively unclear mechanism.

Dose prediction

The purpose of the individual INR prediction is the selection of optimal individual maintenance doses. Ideally, one would compare between predicted and actual optimal doses. However, most patients did not achieve stable maintenance doses during the reported time frame. Of the 81 patients in the inpatient analysis, only 15 reached a stable dose based on the available data (including their outpatient data). These patients likely represent the subset most responsive to treatment, rendering a meaningful comparison between maintenance and predicted doses unfeasible in the present data. As a substitute, we benchmarked model-predicted optimal maintenance doses (based on INR data until Day 4) against the actual dose at discharge, which was available for 68 of the 81 patients. Those doses were determined by the prescribing physician guided by the dose recommended by a clinical pharmacist based on INR values up to 7 days after warfarin initiation using either the Gage et al. algorithm or the IWPC algorithm.[4,5,22]

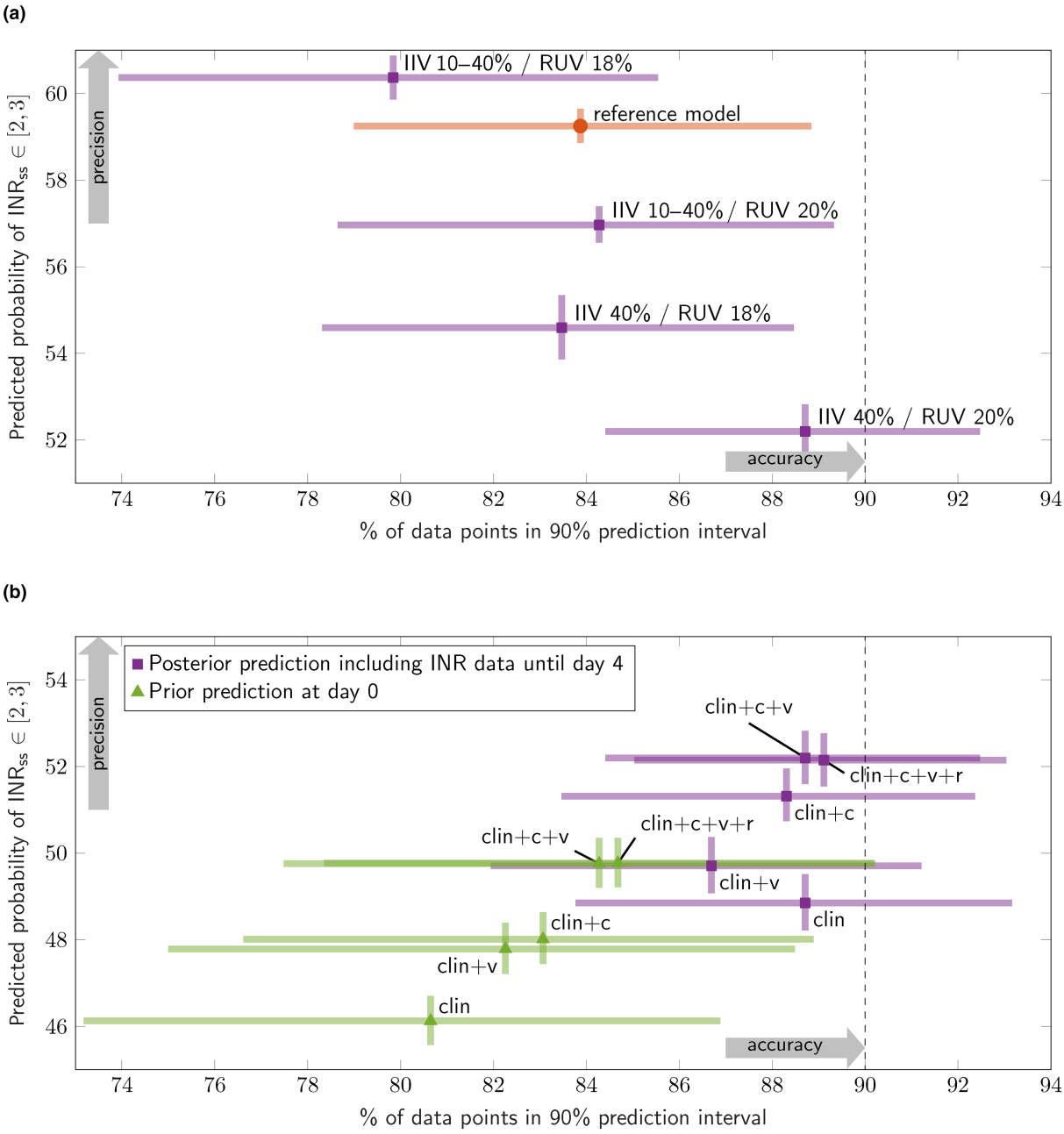

Figure 4 Accuracy and precision of the mechanistic warfarin/INR model with various settings and the reference model for comparison. (**a**) Different hyperparameters IIV and RUV and comparison with the reference model. For IIV, either all parameters were distributed with 40% CV or parameters had CV according to literature (between 10 and 40%), as described in **Section Prior uncertainty in the mechanistic warfarin/INR model**. (**b**) Influence of different covariate sets on prediction quality before start of treatment (green) and after 4 days of INR data (violet). The point labeled "IIV 40%/RUV 20%" in Panel a is the same as the point labeled "clin+c+v" in Panel b. Accuracy is defined as the percentage of data points within 90% prediction interval (evaluated on INR measurements from days 5–9, which were not used for prediction). Precision is defined as the predicted probability of INR within target window at steady state (posterior prediction given optimal individual dose). Uncertainty (90% confidence interval) obtained from bootstrapping. pred: predicted, TW: target window, SS: steady state, IIV: interindividual variability, RUV: residual unexplained variability, clin: clinical covariates (age, weight, baseline INR), c: *CYP2C9*, v: *VKORC1*, r: rs12777823 (only for patients who self-identified as Black or African American).

Figure 5 shows the actual doses at discharge versus predicted optimal doses (based on INR data until day four). Based on the mechanistic warfarin/INR model, 42.7% of the predicted doses were within ± 20 % of the dose at discharge, compared to 30.9% of predicted doses based on the reference model. Hence, the proposed mechanistic warfarin/INR model slightly outperforms the reference model. Of note, while our model predicts lower doses than actually prescribed, the reference model rather predicts higher doses.

The patients with variant rs12777823 alleles self-identifying as Black or African American ($n = 8$) had a smaller mean predicted optimal dose than the remaining population (4.1 mg vs. 4.8 mg),

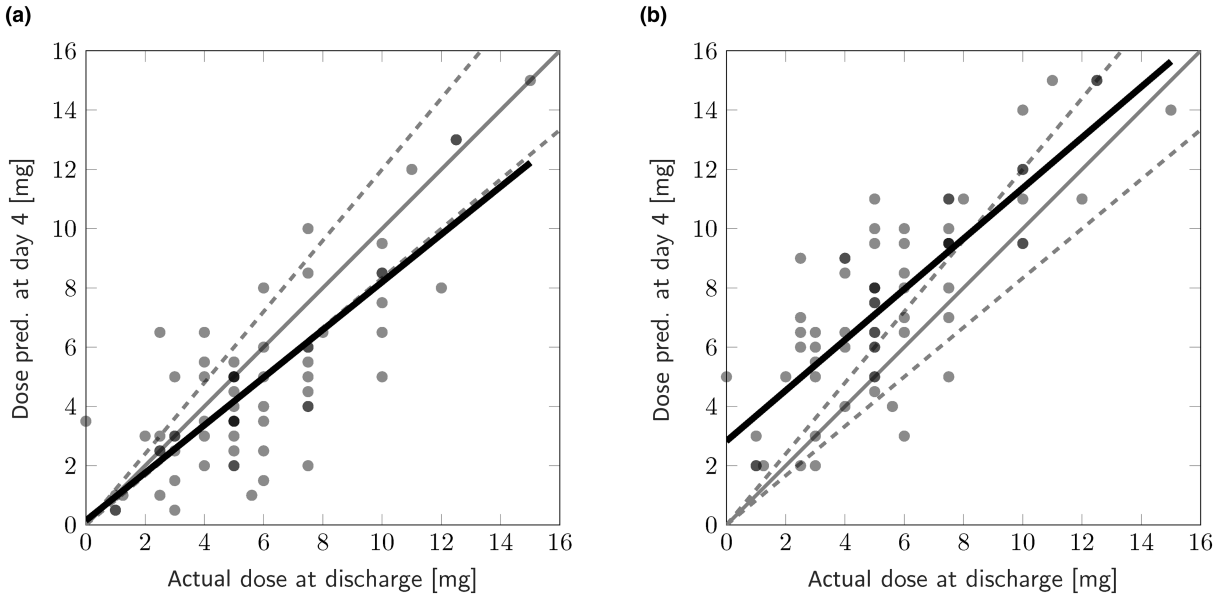

Figure 5 Comparison of predicted optimal dose and actually prescribed dose at discharge. (**a**) Mechanistic warfarin/INR model; (**b**) reference model. Optimal dose predictions based on INR data up to day 4. A linear regression line (black) is plotted together with a perfect correlation line (gray) and 20% deviations (dashed gray).

even though the rs12777823 genotype was not included as a covariate in the final model. This is in line with previous findings,[23] although not very significant due to the small number of affected patients. The patients with any variant CYP2C9 allele (*2, *3, *5, *6, *8 or *11, $n = 17$) had smaller mean predicted doses than the remaining population (2.5 mg vs. 5.3 mg); however, this was already accounted for in the prior model by the inclusion of CYP2C9 as a genotype.

Outpatient INR prediction

Accurately predicting INRs is expected to be more challenging for outpatients than in highly controlled inpatient environments. We evaluated whether the data observed up to the previous visit was sufficient to predict the INR at the subsequent visit in outpatients. We also analyzed if the information at discharge suffices to predict the future outpatient data, see **Supplementary Materials Section S7**.

Figure 6 shows the pcVPC for (i) *Model performance* of the posterior simulation based on all INR data of the respective individual, and (ii) *Predictive performance* of the posterior prediction based on individual INR data up to the previous visit. These data are well-captured in the *Model performance* pcVPC. In the *Predictive performance* setting, a slight, consistent overprediction by approximately 10% can be observed. The uncertainty in predictions is relatively well captured, although the range between upper and lower data percentiles is slightly wider than the range between the prediction percentiles. Concerning accuracy, only 74.4% of the data points are within their respective 90% prediction interval; ideally, this value would be closer to 90%.

Figure 7 shows the pcVPC for outpatient data and predictions with the empirical reference model for comparison. The pcVPCs show a *Model performance* and *Predictive performance* similar to the mechanistic warfarin/INR model, and the accuracy is also very similar, with 74.2% of the data points being within their respective 90% prediction interval. In contrast to the mechanistic model, the empirical reference model rather underpredicts the data in the *Predictive performance* setting.

DISCUSSION

In this study, we evaluated the predictive quality of a mechanistic warfarin/INR model on clinical data when used in MIPD. The warfarin/INR model was previously derived from a QSP model of blood coagulation by model reduction. We conducted an external validation using clinical INR data from a diverse patient cohort and compared it against the external validation of an empirical reference model.[9] Due to observing accuracy and precision comparable to the reference model and the added advantage of a mechanistic foundation, we gained confidence in applying the mechanistic warfarin/INR model in MIPD in similar datasets.

In the presented mechanistic approach, the starting point is the prior mechanistic knowledge encoded in the underlying QSP model. By design, the reduced model depends on the underlying QSP model; this enables a critical discussion of the reduced model since its basis is explicitly given. In contrast, empirical PK/PD models depend on the underlying clinical study used to develop them, most importantly on the study design (patient cohort, dosing regimen, etc). As a result, empirical models can vary substantially (as is also the case for warfarin, see Ref. [7–9]), leaving the user to choose a model without typically knowing explicitly the reason for the differences in model structure. QSP and physiologically based PK models have previously been used to predict individual outcomes based on patient covariates.[24–26] By gathering relevant patient information (e.g., clinical covariates and genotype information) and incorporating it into the mechanistic model, a virtual patient or "digital twin" can be simulated to assess individual outcomes.

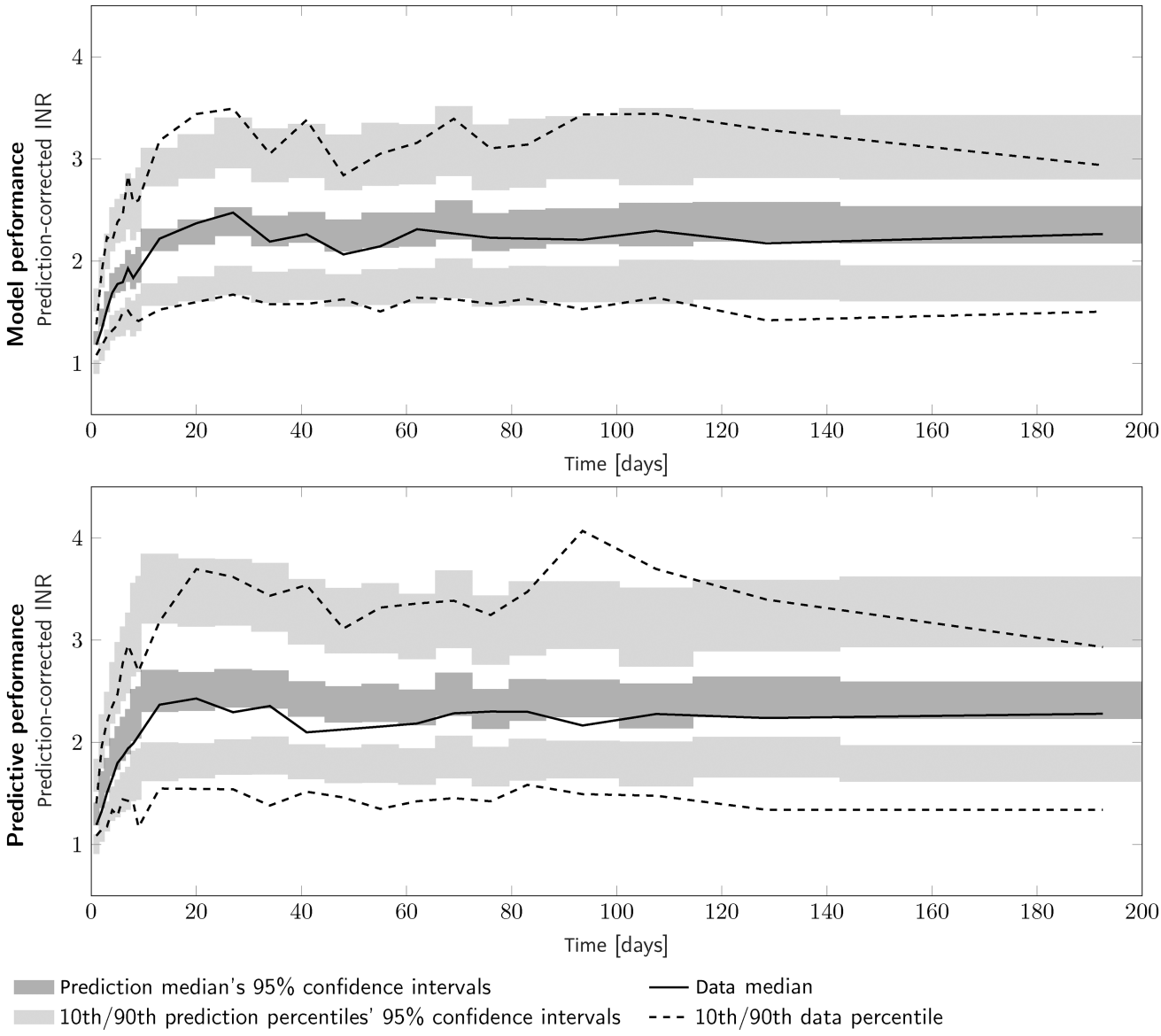

Figure 6 Mechanistic model's prediction-corrected visual predictive check (pcVPC) for patients with outpatient data. Top: Posterior simulation of INR based on all individual (inpatient and outpatient) INR data. Bottom: Posterior prediction of INR based on the individual (inpatient and outpatient) INR data only up to the previous visit. Times were binned to have at least approx. 20 data points per bin; the boxes illustrate the bins. The last bin includes times up to day 242 (box not fully shown).

Importantly, our approach integrates QSP and pharmacometric approaches by additionally assimilating individual outcome data in a Bayesian context (enabled through model reduction). In comparison with a complex QSP model, a reduced model also facilitates communication across disciplines, in addition to faster parameter estimation, allowing for Bayesian updating in real time. Therefore, when made available to clinicians through a web interface, the reduced mechanistic model could aid dose adaptation within a bedside-ready decision support tool.

Often, model prediction quality is assessed by the root mean squared error (RMSE) or (pc)VPCs. However, this focus on population-level metrics, although assessing the population variability, may mask individual prediction inaccuracies. Also, a quantification of uncertainties in the individual predictions is missing. To overcome these limitations in model assessment, we used a full Bayesian approach[11] and used two measures for the accuracy of the individual predictive distribution and the individual predicted precision. These quantitative accuracy and precision measurements provide a comprehensive evaluation of the predictive performance, including its uncertainty. This approach enabled us to evaluate clinically relevant model quality markers, notably the probability of achieving a steady state INR within the target range.

We observed initial overpredictions of the INR values by both the mechanistic and the reference model (**Figure 3**). The models predicted a faster INR increase than what is observed in the data. A possible reason could be a sample bias, as we considered only patients hospitalized for at least 5 days. Indeed, those patients had

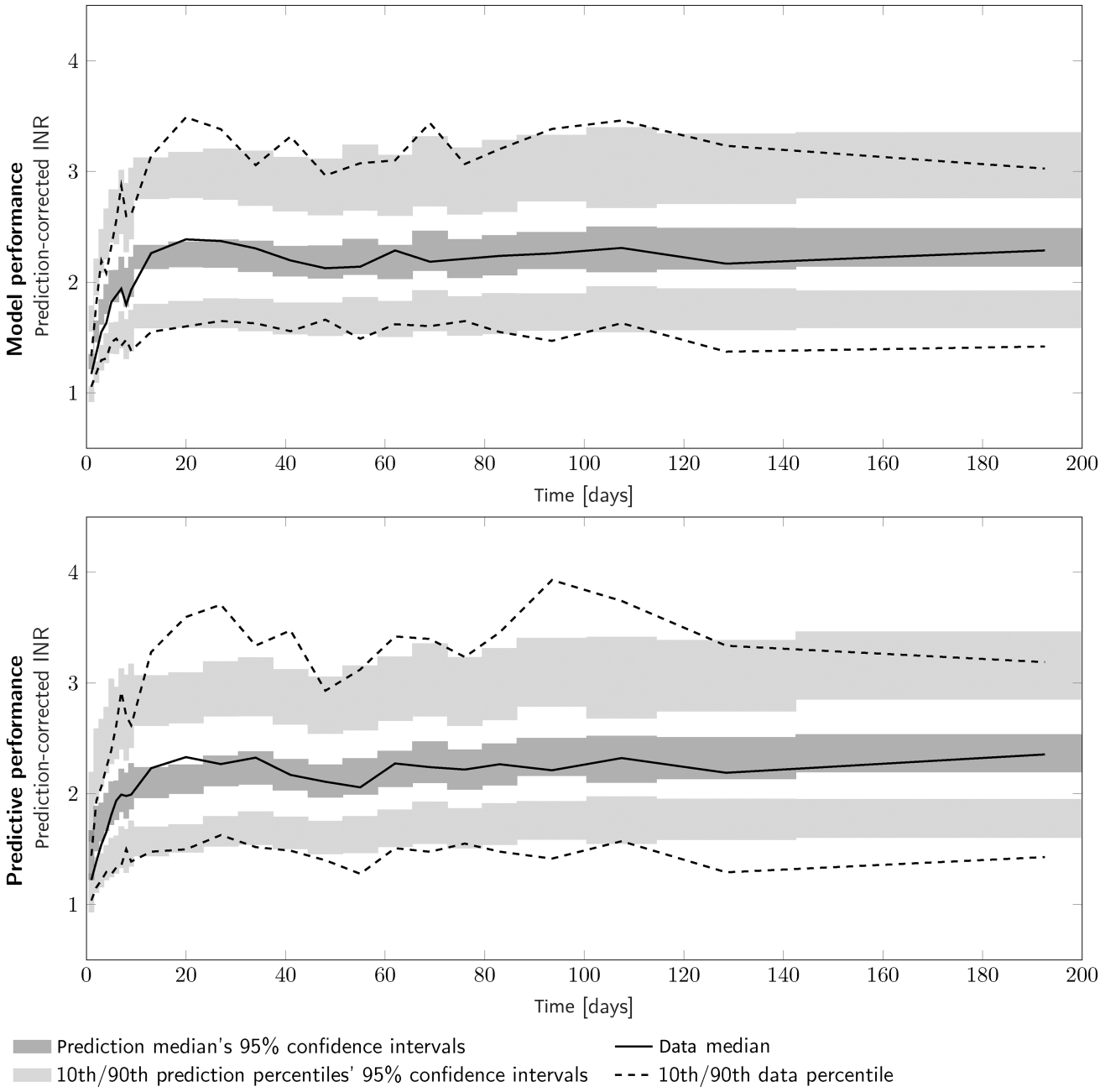

Figure 7 Empirical reference model's prediction-corrected visual predictive check (pcVPC) for patients with outpatient data. Top: Posterior simulation of INR based on all individual (inpatient and outpatient) INR data. Bottom: Posterior prediction of INR based on the individual (inpatient and outpatient) INR data only up to the previous visit. Times were binned to have at least approx. 20 data points per bin; the boxes illustrate the bins. The last bin includes times up to day 242 (box not fully shown).

a lower mean INR on Day 4 than others. Irrespective of the cause, there are two possible solutions to account for the initial overprediction. First, weighing the data more heavily by using a flattened prior[27] would lead to a more substantial adaptation of the prediction to the data. However, this approach resulted in a better agreement between early data and prediction but a worse agreement for later time points when using data until day four. The second option involves adapting the entire model via a hierarchical Bayesian framework. In the present article, only individual patient data was assimilated to calculate the individual posterior prediction, but the model prior was not calibrated using patient data. As a next step, the data from the whole patient cohort could be used to refine the prior in a continued learning approach.[28] Merging the data with the mechanistic prior knowledge is expected to enhance performance for future predictions.

The balance between accuracy and precision in the model prediction is largely modulated by the two hyperparameters, RUV and IIV. Reducing the variability hyperparameters enhances precision at the expense of accuracy. However, a prediction that is precise yet inaccurate is not meaningful. Therefore, we prioritized

accuracy over precision. When selecting hyperparameters, we suggest establishing a cutoff value for the accuracy measure, for example, requiring a minimum of 80% of values to fall within the 90% prediction interval and then optimizing for precision. Of the considered hyperparameters, the RUV with 20% CV and IIV according to literature (10–40%) are the best given these criteria. The optimal hyperparameter values could also be estimated in a hierarchical Bayesian framework.

Interestingly, the magnitude of the RUV considerably limits the probability of achieving INR values within the target range; this is illustrated in **Supplementary Materials Section S10**. For exponential RUV with 20% CV,[9] the precision cannot exceed 68.9%. The corresponding expected time within target range defined by the Rosendaal method[29] was 78.8% in a numerical simulation, see **Supplementary Materials Section S10** for details. Attaining this bound would require complete identification of the interindividual variability from INR data.

Our analysis identified several influential covariates: baseline INR, weight, age, *CYP2C9*, and *VKORC1* genotype. The genotype information retains its predictive significance even after assimilating INR data up to Day 4, as demonstrated in **Figure 4b** and consistent with previous findings.[30] The prediction enhancements observed in **Figure 4b** suggest that considering the *CYP2C9* genotype may have significant predictive value even in the absence of *VKORC1* data (compare points labeled "clin" and "clin+c"). This diverges from the Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for Pharmacogenetics-Guided Warfarin Dosing,[1] which recommends a genotype-guided approach only if both *CYP2C9* and *VKORC1* genotypes are available. This may be clinically relevant because *CYP2C9* may already be on record for some patients, given its significance for several other drugs. We also evaluated rs12777823 for patients who self-identified as Black or African American and *CYP4F2* as potential covariates. Regarding *CYP4F2*, our model suggests that this enzyme primarily affects the baseline INR without influencing the subsequent warfarin therapy trajectory. The rs12777823 genotype was previously found to be associated with altered warfarin clearance in African Americans[13,31] and to influence the required warfarin dose.[23] In the present analysis, only 8 patients who self-identified as African American had a variant rs12777823 allele, thus the analysis remains inconclusive. We excluded the rs12777823 genotype as a covariate in the model due to its lack of contribution to prediction improvement within this patient population (see **Figure 4b**, compare points labeled "clin+c+v" and "clin+c+v+r").

Dose optimization is a key objective in model-informed precision dosing. Ideally, if stable maintenance doses were available for the majority of the population, they would be the preferred outcome metric. However, relying solely on stable maintenance doses would exclude patients who do not achieve stable dosing, potentially masking underperformance for those patients. The presented analysis uses the dose at discharge as a substitute for the stable maintenance dose as most patients did not achieve a stable maintenance dose within the available data frame. The mechanistic warfarin/INR model slightly outperforms the reference model by predicting (based on INR data until Day 4) optimal doses that are closer to the actual doses at discharge. While the discharge doses might differ from stable maintenance doses, they were deemed optimal by a physician at a later time point than the prediction was made. In practice, the dose optimization approach would involve determining the individual optimal dose based on the prior and all available data up to the point in time at which a dose adaptation decision is to be made. The individually optimal starting dose would be determined based only on covariates. Later doses would be based on covariates and additionally on all available INR data up to that time. Consequently, the dose would be adjusted daily during the patient's clinic stay and at each subsequent clinic visit thereafter. The optimized dose based on INR data until Day 4, which we used in this article for calculation of the precision and comparison with actual doses, represents a snapshot of this approach at Day 4. Of note, the particle filter method facilitates the assimilation of new data to update the optimal dose.[11]

We finally explored the difference in prediction quality in inpatient and outpatient settings. In the prediction of outpatient data, we observed a slight bias, with INR predictions from the mechanistic model being larger than observed values (**Figure 6** bottom) and predictions from the reference model being smaller (**Figure 7** bottom). This discrepancy might be attributed to issues with adherence, an increased Vitamin K diet or a potential habituation effect. Consistent with our expectations, the variability in outpatient data appears slightly higher than predicted, likely due to inconsistent adherence, less controlled dosing times, varying diet and other unforeseen factors. It is plausible that any model developed on data from the highly controlled hospital setting would require adjustments, particularly in the residual error component, when adapted to an outpatient setting. An increase in the RUV to 30% CV would suffice for acceptable accuracy of the outpatient prediction but comes with a reduction in precision. With a model update in the hierarchical Bayes framework, including an estimation of the RUV, the accuracy might be improved without consequences on the precision.

In conclusion, this article endorses QSP-derived modeling as a complementary approach to empirical modeling to produce PK/PD models suitable for MIPD. In a case study, the mechanistic warfarin/INR model, even without prior calibration with clinical data, performed well compared to an empirically based model. When integrated with empirical insights, the mechanistic warfarin/INR model shows potential for MIPD with the added advantage of physiologic interpretability. In the future, the mechanistic model development concept based on QSP model reduction and empirical concepts could be combined to yield better PK/PD models for MIPD.

SUPPORTING INFORMATION

Supplementary information accompanies this paper on the *Clinical Pharmacology & Therapeutics* website (www.cpt-journal.com).

ACKNOWLEDGMENTS

UF acknowledges the assistance provided by ChatGPT version 4, an AI language model developed by OpenAI. This tool supported refining the manuscript's language, structure, and conciseness, contributing to the revision of the abstract, introduction, and discussion sections. Its role was limited to providing language editing suggestions and did not extend to any contribution to the conceptual framework, data analysis, or interpretation of results. Open Access funding enabled and organized by Projekt DEAL.

FUNDING
Funding provided by the graduate research training program PharMetrX: Pharmacometrics & Computational Disease Modeling, Berlin/Potsdam, Germany. Funded by NIH Grant R01HG011800.

CONFLICT OF INTEREST
C.K. and W.H. report research grants from an industry consortium (AbbVie Deutschland GmbH & Co. K.G., AstraZeneca, Boehringer Ingelheim Pharma GmbH & Co. K.G., Grünenthal GmbH, F. Hoffmann-La Roche Ltd., Merck KGaA, Novo Nordisk A/S, and SANOFI) for the PharMetrX program. In addition, C.K. reports research grants from the Innovative Medicines Initiative-Joint Undertaking ("DDMoRe") and Diurnal Ltd. C.K. reports grants from the Federal Ministry of Education and Research within the Joint Programming Initiative on Antimicrobial Resistance Initiative. All other authors declared no competing interests for this work.

AUTHOR CONTRIBUTIONS
U.F., L.H.C., J.D.D., C.K., S.S., and W.H. designed the research and wrote the manuscript; U.F. performed the research; U.F., L.H.C., J.D.D., S.S., and W.H. analyzed the data.

1. Johnson, J. *et al*. Clinical Pharmacogenetics implementation consortium (CPIC) guideline for Pharmacogenetics-guided warfarin dosing: 2017 update. *Clin. Pharmacol. Ther.* **102**, 397–404 (2017).
2. Gage, B.F. *et al*. Effect of genotype-guided warfarin dosing on clinical events and anticoagulation control among patients undergoing hip or knee arthroplasty. *JAMA* **318**, 1115–1124 (2017).
3. Pirmohamed, M. *et al*. A randomized trial of genotype-guided dosing of warfarin. *N. Engl. J. Med.* **369**, 2294–2303 (2013).
4. Gage, B. *et al*. Use of Pharmacogenetic and clinical factors to predict the therapeutic dose of warfarin. *Clin. Pharmacol. Ther.* **84**, 326–331 (2008).
5. The International Warfarin Pharmacogenetics Consortium. Estimation of the warfarin dose with clinical and Pharmacogenetic data. *N. Engl. J. Med.* **360**, 753–764 (2009).
6. Hamberg, A.K. *et al*. A PK-PD model for predicting the impact of age, CYP2C9, and VKORC1 genotype on individualization of warfarin therapy. *Clin. Pharmacol. Ther.* **81**, 529–538 (2007).
7. Ohara, M. *et al*. Determinants of the over-anticoagulation response during warfarin initiation therapy in Asian patients based on population pharmacokinetic-Pharmacodynamic analyses. *PLoS One* **9**, e105891 (2014).
8. Xue, L. *et al*. Theory-based pharmacokinetics and pharmacodynamics of S- and R-warfarin and effects on international normalized ratio: influence of body size, composition and genotype in cardiac surgery patients. *Br. J. Clin. Pharmacol.* **83**, 823–835 (2016).
9. Hamberg, A.K. *et al*. A Pharmacometric model describing the relationship between warfarin dose and INR response with respect to variations in CYP2C9, VKORC1, and age. *Clin. Pharmacol. Ther.* **87**, 727–734 (2010).
10. Falkenhagen, U., Knöchel, J., Kloft, C. & Huisinga, W. Deriving mechanism-based pharmacodynamic models by reducing quantitative systems pharmacology models: an application to warfarin. *CPT Pharmacometrics Syst. Pharmacol.* **12**, 432–443 (2023).
11. Maier, C., Hartung, N., Wiljes, J., Kloft, C. & Huisinga, W. Bayesian data assimilation to support informed decision making in individualized chemotherapy. *CPT Pharmacometrics Syst. Pharmacol.* **9**, 153–164 (2020).
12. Arwood, M. *et al*. Anticoagulation endpoints with clinical implementation of warfarin Pharmacogenetic dosing in a real-world setting: a proposal for a new Pharmacogenetic dosing approach. *Clin. Pharmacol. Ther.* **101**, 675–683 (2016).
13. Perera, M.A. *et al*. Genetic variants associated with warfarin dose in African-American individuals: a genome-wide association study. *Lancet* **382**, 790–796 (2013).
14. Wajima, T., Isbister, G.K. & Duffull, S.B. A comprehensive model for the humoral coagulation network in humans. *Clin. Pharmacol. Ther.* **86**, 290–298 (2009).
15. Knöchel, J., Kloft, C. & Huisinga, W. Understanding and reducing complex systems pharmacology models based on a novel input–response index. *J. Pharmacokinet. Pharmacodyn.* **45**, 139–157 (2018).
16. Hamberg, A.K. *et al*. Warfarin dose prediction in children using pharmacometric bridging—comparison with published pharmacogenetic dosing algorithms. *Eur. J. Clin. Pharmacol.* **69**, 1275–1283 (2013).
17. McDonald, M.G., Rieder, M.J., Nakano, M., Hsia, C.K. & Rettie, A.E. CYP4F2 is a vitamin K_1 oxidase: an explanation for altered warfarin dose in carriers of the V433M variant. *Mol. Pharmacol.* **75**, 1337–1346 (2009).
18. Bergstrand, M., Hooker, A.C., Wallin, J.E. & Karlsson, M.O. Prediction-corrected visual predictive checks for diagnosing nonlinear mixed-effects models. *AAPS J.* **13**, 143–151 (2011).
19. Drozda, K. *et al*. Poor warfarin dose prediction with pharmacogenetic algorithms that exclude genotypes important for African Americans. *Pharmacogenet. Genomics* **25**, 73–81 (2015).
20. Kimmel, S.E. *et al*. A pharmacogenetic versus a clinical algorithm for warfarin dosing. *N. Engl. J. Med.* **369**, 2283–2293 (2013).
21. Deng, J., Vozmediano, V., Rodriguez, M., Cavallari, L.H. & Schmidt, S. Genotype-guided dosing of warfarin through modeling and simulation. *Eur. J. Pharm. Sci.* **109**, S9–S14 (2017).
22. Nutescu, E.A. *et al*. Feasibility of implementing a comprehensive warfarin Pharmacogenetics service. *Pharmacotherapy: J Human Pharmacol drug Therapy* **33**, 1156–1164 (2013).
23. Limdi, N.A. *et al*. Race influences warfarin dose changes associated with genetic factors. *Blood* **126**, 539–545 (2015).
24. Hartmann, S., Biliouris, K., Lesko, L.J., Nowak-Göttl, U. & Trame, M.N. Quantitative systems pharmacology model-based predictions of clinical endpoints to optimize warfarin and rivaroxaban anti-thrombosis therapy. *Front. Pharmacol.* **11**, 1041 (2020).
25. Polasek, T.M., Shakib, S. & Rostami-Hodjegan, A. Precision dosing in clinical medicine: present and future. *Expert. Rev. Clin. Pharmacol.* **11**, 743–746 (2018).
26. Joslyn, L.R., Huang, W., Miles, D., Hosseini, I. & Ramanujan, S. Digital twins elucidate critical role of Tscm in clinical persistence of TCR-engineered cell therapy. *NPJ Syst. Biol. Appl.* **10**, 11 (2024).
27. Hughes, J.H. & Keizer, R.J. A hybrid machine learning/pharmacokinetic approach outperforms maximum a posteriori Bayesian estimation by selectively flattening model priors. *CPT Pharmacometrics Syst. Pharmacol.* **10**, 1150–1160 (2021).
28. Maier, C., de Wiljes, J., Hartung, N., Kloft, C. & Huisinga, W. A continued learning approach for model-informed precision dosing: updating models in clinical practice. *CPT Pharmacometrics Syst. Pharmacol.* **11**, 185–198 (2021).
29. Rosendaal, F.R., Cannegieter, S.C., van der Meer, F.J.M. & Briët, E. A method to determine the optimal intensity of oral anticoagulant therapy. *Thromb. Haemost.* **69**, 236–239 (1993).
30. Lenzini, P. *et al*. Pharmacogenetic warfarin dose refinements remain significantly influenced by genetic factors after one week of therapy. *Thromb. Haemost.* **107**, 232–240 (2012).
31. Ar, S. *et al*. Association of cytochrome P450 2C19 genotype with the antiplatelet effect and clinical efficacy of Clopidogrel therapy. *JAMA* **302**, 849 (2009).