frontiers | Frontiers in Pharmacology

TYPE Original Research
PUBLISHED 19 October 2023
DOI 10.3389/fphar.2023.1270443

OPEN ACCESS

EDITED BY
Fawzy Elbarbry,
Pacific University, United States

REVIEWED BY
Thomas Polasek,
Monash University, Australia
Wilhelm Huisinga,
University of Potsdam, Germany

*CORRESPONDENCE
David Augustin,
david.augustin@cs.ox.ac.uk

RECEIVED 02 August 2023
ACCEPTED 05 October 2023
PUBLISHED 19 October 2023

CITATION
Augustin D, Lambert B, Robinson M, Wang K and Gavaghan D (2023), Simulating clinical trials for model-informed precision dosing: using warfarin treatment as a use case.
Front. Pharmacol. 14:1270443.
doi: 10.3389/fphar.2023.1270443

Simulating clinical trials for model-informed precision dosing: using warfarin treatment as a use case

David Augustin[1]*, Ben Lambert[2], Martin Robinson[1], Ken Wang[3] and David Gavaghan[1]

[1]Department of Computer Science, University of Oxford, Oxford, United Kingdom, [2]College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, United Kingdom, [3]Research and Early Development, F. Hoffmann-La Roche AG, Basel, Switzerland

Treatment response variability across patients is a common phenomenon in clinical practice. For many drugs this inter-individual variability does not require much (if any) individualisation of dosing strategies. However, for some drugs, including chemotherapies and some monoclonal antibody treatments, individualisation of dosages are needed to avoid harmful adverse events. Model-informed precision dosing (MIPD) is an emerging approach to guide the individualisation of dosing regimens of otherwise difficult-to-administer drugs. Several MIPD approaches have been suggested to predict dosing strategies, including regression, reinforcement learning (RL) and pharmacokinetic and pharmacodynamic (PKPD) modelling. A unified framework to study the strengths and limitations of these approaches is missing. We develop a framework to simulate clinical MIPD trials, providing a cost and time efficient way to test different MIPD approaches. Central for our framework is a clinical trial model that emulates the complexities in clinical practice that challenge successful treatment individualisation. We demonstrate this framework using warfarin treatment as a use case and investigate three popular MIPD methods: 1. Neural network regression; 2. Deep RL; and 3. PKPD modelling. We find that the PKPD model individualises warfarin dosing regimens with the highest success rate and the highest efficiency: 75.1% of the individuals display INRs inside the therapeutic range at the end of the simulated trial; and the median time in the therapeutic range (TTR) is 74%. In comparison, the regression model and the deep RL model have success rates of 47.0% and 65.8%, and median TTRs of 45% and 68%. We also find that the MIPD models can attain different degrees of individualisation: the Regression model individualises dosing regimens up to variability explained by covariates; the Deep RL model and the PKPD model individualise dosing regimens accounting also for additional variation using monitoring data. However, the Deep RL model focusses on control of the treatment response, while the PKPD model uses the data also to further the individualisation of predictions.

KEYWORDS

MIPD, precision dosing, clinical trial simulation, inter-individual variability, PKPD modelling, reinforcement learning, deep learning, warfarin

1 Introduction

Model-informed precision dosing (MIPD) is an emerging technique used to individualise dosing regimens of otherwise difficult-to-administer drugs (Sheiner, 1969; Keizer et al., 2018; Darwich et al., 2021). Typical examples that would benefit from MIPD are drugs with narrow therapeutic windows and large treatment response variability, such as warfarin, docetaxel and infliximab (Wadelius and Pirmohamed, 2007; Gill et al., 2016; Ma et al., 2021). Other examples include antibiotics, like vancomycin, where MIPD has been suggested and partially implemented to guide dosing strategies for critically ill patients, balancing the treatment of severe infections and the risks for harmful adverse events (Broeker et al., 2019; Wicha et al., 2021; Matsumoto et al., 2022). MIPD may also be used to efficiently adapt dosing regimens to continuously changing conditions, for example to stabilise the blood glucose levels of diabetes patients with insulin treatment (Wang et al., 2019; Zhu et al., 2020).

The most prominent MIPD methods are regression, reinforcement learning (RL) and pharmacokinetic and pharmacodynamic (PKPD) modelling (Johnson et al., 2011; Johnson et al., 2017; Keizer et al., 2018; Ribba et al., 2020). These methods may be further categorised into different subvariants. RL variants include, for example, model-free algorithms such as Q learning (Zadeh et al., 2023) and model-based algorithms such as Monte Carlo tree search (Maier et al., 2021). MIPD variants of PKPD modelling include maximum a posteriori-guided dosing and Bayesian data assimilation-guided dosing (Maier et al., 2020). Other variants include the modelling of virtual twins using physiology-based PK (PBPK) and quantitative systems pharmacology (QSP) models (Polasek and Rostami-Hodjegan, 2020). Although differences across MIPD methods exist, all have in common that they need to process data in two steps in order to make individualised predictions. First models are fitted to population data. This step establishes a relationship between patient characteristics and dosing strategies. In a second step, data specific to the to-be-treated patient is used to predict individualised dosing regimens. For example, regression models have been fitted using records of genetic information and dosages across patients (Gage et al., 2008; Klein et al., 2009; Gong et al., 2011), enabling the prediction of individualised dosages based on genetic information.

The data used for model fitting and dosing regimen individualisation differ substantially across MIPD methods and include clinical factors, genetic factors and monitoring data. The type and volume of data are key to both the accuracy of predictions and the ease of implementation in clinical practice (Darwich et al., 2017; Ribba et al., 2022). The more data are collected, the better dosing regimens can be individualised. However, practical constraints limit how much and what type of data may be available for MIPD. As a result, the trade-off between accuracy and practicality needs to be considered when applying MIPD approaches to medicines. The systematic study of this trade-off for a specific application is, however, complicated by the astronomical costs of clinical trials, rendering repeated clinical trials for different MIPD methods infeasible. In this study, we propose a framework for the simulation of clinical MIPD trials, facilitating a resource-efficient way to test and develop MIPD approaches.

Using simulations to understand MIPD approaches is not a new concept and several MIPD simulation studies exist in the literature. Moore et al. (2004) and Ribba et al. (2022) use simulated treatment responses to investigate RL approaches as a strategy to individualise dosages of anaesthetics in intensive care units. For the simulations, they used a semi-mechanistic PKPD model to emulate the time course of treatment responses and a nonlinear mixed effects (NLME) model structure to capture inter-individual variability (IIV). A similar approach is adopted by Maier et al. (2020, 2021) to study the individualisation of paclitaxel-based chemotherapy using different MIPD approaches, including PKPD modelling, RL and a hybrid PKPD-RL approach. An NLME model simulation approach is also used by Zadeh et al. (2023) to study deep RL as a candidate for MIPD of warfarin. Abrantes et al. (2019) extend this NLME simulation approach, adding inter-occasion variability (IOV) of treatment responses as an extra dimension to the MIPD simulation. They model IOV following Karlsson and Sheiner (1993) and randomly vary the PKPD model parameters of virtual patients over time. An analogous approach is used by Keutzer and Simonsson (2020) to understand MIPD-based treatment individualisation of rifampicin.

We propose an extended framework for the simulation of clinical MIPD trials. Our framework complements the emulation of IIV and IOV by other previously established elements of clinical trial simulation (Holford et al., 2000; Holford et al., 2010). In particular, our framework includes treatment response emergence from complex interactions of pharmacological and physiological processes as a central feature of the trial simulation. In addition, deviations of dose administrations from nominal dosing regimens, as well as delayed monitoring measurements are incorporated in the simulation to more faithfully represent practical limitations of monitoring-based MIPD approaches.

The article is divided into three sections: methods; results and discussion; and conclusion. In the methods, we first introduce the general framework for MIPD trial simulation and subsequently demonstrate its implementation in terms of a clinical trial model for the warfarin use case. We then introduce the three MIPD models, which we will investigate with the help of the clinical trial model. The considered models are: 1. a neural network regression model; 2. a deep reinforcement learning model; and 3. a PKPD model. In the results and discussion, we use the clinical trial model to simulate MIPD trials for each of the three models and analyse the strengths and limitations of the MIPD models. In our analysis, we pay careful attention to attributing generic strengths and limitations to the MIPD methodology and specific strengths and limitations to our implementations of the models. In the conclusion, we summarise our findings and propose future directions for MIPD research.

2 Methods

We first introduce the proposed framework for clinical MIPD trial simulation and discuss the specific implementation for warfarin. We then outline the investigated MIPD methods. The data, models and scripts used in this article are hosted on GitHub (https://github.com/DavAug/mipd-warfarin). A user-friendly API

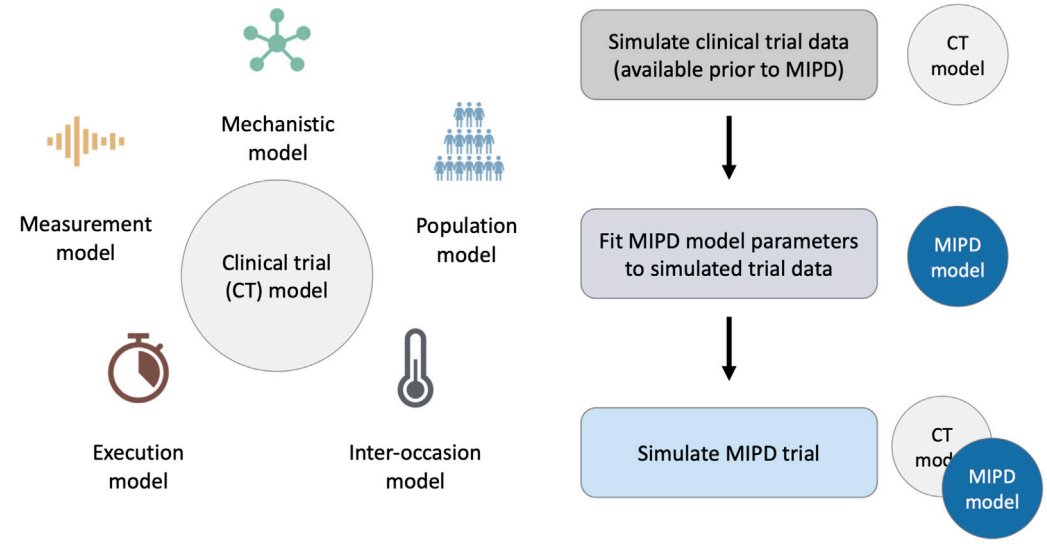

FIGURE 1
Framework for MIPD trial simulation. The left panel shows the components of the clinical trial (CT) model: 1. A mechanistic model; 2. A population model; 3. An inter-occasion model; 4. An execution model; and 5. A measurement model. The right panel shows the workflow leading up to the MIPD trial simulation. First, the CT model is used to simulate typical clinical trial data, available prior to the MIPD trial. Second, the simulated trial data is used to fit the parameters of the MIPD model. This emulates the starting point of MIPD trials in practice. Finally, the CT model and the fitted MIPD model are used to simulate the MIPD trial.

for MIPD approaches using PKPD modelling has been implemented in the open source Python package chi (Augustin, 2021). MIPD approaches using neural networks were implemented in PyTorch (Paszke et al., 2019).

2.1 General framework for MIPD trial simulation

The objective of MIPD is to achieve desired treatment outcomes across individuals despite large treatment response variability by optimising individual dosing regimens. Challenges for this individualisation are nonlinear and delayed treatment responses (Mager, 2006; Dirks and Meibohm, 2010; Véronneau-Veilleux et al., 2020), making it difficult to adjust and extrapolate dosages based on feedback from monitoring. Variability of the treatment response as a result of time-varying pharmacokinetics, concomitant food-intake, comedication and disease progression further complicate the dosing regimen adjustment (Keutzer and Simonsson, 2020). A realistic assessment of MIPD approaches in a simulated trial therefore needs to account not only for IIV, but also for treatment response delays, nonlinearities and IOV. However, PKPD-related aspects are not the only factors influencing the success of MIPD methods. There are also practical limitations for MIPD, including imperfect measurements and deviations from nominal dosing and monitoring schedules (Holford et al., 2000; Holford et al., 2010). While measurement noise is commonly included in simulations, the variability in the execution of the trial often remains neglected. Deviations from the nominal schedule can impact the success of MIPD methods, especially when they are not fed back into the model.

To faithfully emulate the performance of MIPD in clinical practice, our simulation framework accounts for these PKPD-related and practical challenges using a clinical trial (CT) model composed of five components: 1. a mechanistic model; 2. a population model; 3. an inter-occasion model; 4. an execution model and 5. a measurement model (see left panel in Figure 1). We describe each of these components in Section 2.1.2. The right panel of the figure illustrates the workflow of using the CT model for simulating MIPD trials.

2.1.1 Workflow of MIPD trial simulation

Our method of MIPD trial simulation involves three steps (see right panel in Figure 1): 1. simulation of pre-MIPD clinical trial data; 2. fitting of the MIPD model to this simulated data; 3. simulation of the MIPD trial. The first two steps emulate the MIPD model development that takes place in practice prior to MIPD trials. In our method, we first simulate typical clinical data from e.g. phase I trials, phase II trials and/or phase III trials using the CT model. We then fit the MIPD model parameters to the simulated data. The details of the fitting process are specific to the MIPD model and are presented in Section 2.3. It is important that the simulated data are used for the fitting, even if real clinical data are available, in order to facilitate a clear attribution of limitations observed in a simulated MIPD trial to the MIPD model. If instead, the MIPD model was fitted to real clinical data, the approximation error of the CT model with respect to the data-generating process of the real clinical trial may also contribute to limitations observed in the simulated trials, making it harder to draw conclusions about the MIPD model. The real data should, instead, be used to calibrate the CT model prior to the data simulation in order to minimise the approximation error as much as possible. The final step of the workflow is the simulation of the MIPD trial using the fitted MIPD model and the CT model.

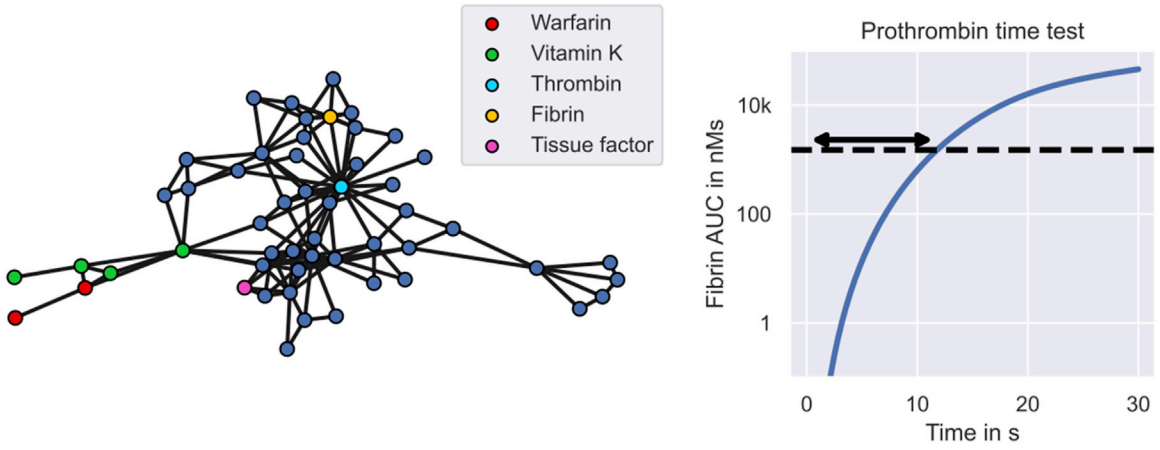

FIGURE 2
Warfarin clinical trial simulation - the mechanistic model. The figure shows Wajima et al.'s model of the warfarin treatment response mechanism. The left panel shows the model. Nodes represent states of the model, including warfarin (red), vitamin K (green) and different coagulation factors (blue). Central coagulation factors include thrombin (light blue) and fibrin (yellow). Multiple nodes for warfarin refer to the drug amount in different compartments, while multiple nodes for vitamin K also refer to different forms of vitamin K, such as vitamin K epoxide and vitamin K hydroquinone. Interactions between states, such as transitions, inhibitions or activations, are represented by edges. The right panel shows the model simulation of the prothrombin time test. The arrow indicates the prothrombin time which marks the time elapsed between exposure to tissue factor and reaching a fibrin AUC of 1500 nMs.

2.1.2 Components of the CT model

We now describe the components of the CT model. By design, the components of the simulation are non-overlapping and are as modular as possible. This simplifies the model structure and enables an iterative development of CT models, making it possible to replace or further develop individual components without requiring changes in other model components.

1. The mechanistic model: This component models the dynamics of treatment responses as a function of time, t, and the dosing regimen, r,

$$\bar{y}(t, r, \psi). \tag{1}$$

$\bar{y}$ denotes quantities of interest that may be monitored in clinical practice, and ψ denotes the parameters of the model. The main purpose of the mechanistic model is to faithfully reflect nonlinearities and delays of the treatment response, emergent from complex cascades of pharmacological and physiological processes. Emulating this complexity provides a tool to test the ability of different MIPD approaches to approximate the treatment response and predict individualised dosing regimens. Popular choices to simulate treatment response dynamics include PKPD models and quantitative systems pharmacology (QSP) models (Ribba et al., 2020; Azer et al., 2021; Maier et al., 2021).

An example mechanistic model of warfarin treatment developed by Wajima et al. (2009) is illustrated in Figure 2. Warfarin is an oral anticoagulant widely used for the prevention and treatment of venous thrombosis, pulmonary embolism and thromboembolic complications associated with atrial fibrillation and/or cardiac valve replacement (FDA, 2010). The left panel shows the 53 blood components described by the model, including warfarin, vitamin K and different coagulation factors, such as thrombin and fibrin. Edges between the components represent interactions, such as transitions, inhibitions or activations. The monitored quantity of the treatment response, $\bar{y}$, is the prothrombin time. The prothrombin time measures the time it takes for plasma to clot after exposure to a thromboplastin reagent and is routinely measured in clinical practice. In Wajima et al.'s model, this prothrombin time test is simulated by measuring the time elapsed between adding the reagent (300 nM tissue factor) and reaching a fibrin area-under-the-curve (AUC) of 1500 nMs (see right panel in Figure 2). The prothrombin time is commonly reported in terms of the international normalised ratio (INR), which measures the prothrombin time of a patient's blood sample in units of the prothrombin time of a reference sample. We will use this model in our warfarin clinical trial simulation (see Section 2.2 for details).

2. The population model: This component models the variability in the treatment response across individuals using a mixed effects model extension of the mechanistic model. A mixed effects model defines a population distribution of model parameters

$$p(\psi|\theta), \tag{2}$$

capturing the differences between individuals, i.e., the IIV (Lavielle, 2014; Augustin et al., 2023). θ denotes the parameters of the population distribution. Each sample, ψ, from the population distribution represents an individual with treatment response $\bar{y}(t, r, \psi)$. Thus, differences between individuals arise in this model structure from differences in the mechanistic model parameters. For some applications, these differences can be partially explained by covariates, χ, which may divide the population into subpopulations, $p(\psi|\theta, \chi)$. The full population distribution across covariates is then given by the average of the supopulations weighted by the relative frequency of the covariates, $p(\psi|\theta) = \mathbb{E}_{\chi}[p(\psi|\theta, \chi)]$.

Covariates of the variability can range from clinical factors to genetic factors. For example for warfarin treatment, the VKORC1 genotype

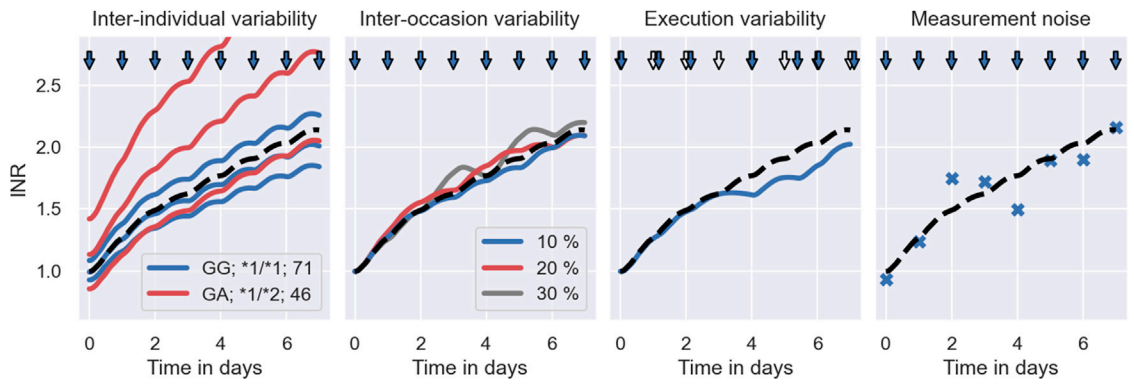

FIGURE 3
Warfarin clinical trial simulation–sources of treatment response variability. The figure shows each contribution to the treatment response variability in isolation. Panel 1: Illustrates the effect of the population model on the clinical trial simulation. The solid lines indicate the treatment responses of 6 simulated individuals to daily warfarin administrations: 3 of which have the covariates χ = (GG, * 1/* 1, 71) (see blue lines); and the remaining 3 have the covariates χ = (GA, * 1/* 2, 46) (see red lines). The typical treatment response across individuals is indicated by a dashed line. The administration times are indicated by blue arrows. Panel 2: Illustrates the effect of the inter-occasion model on the clinical trial simulation. The dashed line shows the treatment response of an individual with a constant vitamin K input rate, i.e., no IOV. The solid lines show the treatment response of the same individual with vitamin K input rates that randomly vary by 10% (blue), 20% (red) and 30% (grey) between days. Panel 3: Illustrates the effect of the execution model on the clinical trial simulation. The dashed line indicates the treatment response of an individual associated with the nominal dosing regimen (hollow arrows). The blue line indicates the treatment response of the same individual associated with the delayed dose administrations (blue arrows). Panel 4: Illustrates the effect of the measurement model on the clinical trial simulation. The dashed line shows the simulated treatment response of an individual and the scatter points show the associated monitoring measurement.

explains 27% of the observed response variability (Wadelius et al., 2009). Other covariates of warfarin treatment include mutations in the CYP2C9 gene and the age of the patient. The link between covariates and pharmacological or physiological processes makes it possible to define mixed effects models that reflect the mechanistic relationship between covariates and the treatment response variability (Hamberg et al., 2007; Hamberg et al., 2010; Hartmann et al., 2016; Hartmann et al., 2020). For example, warfarin's mode of action is the inhibition of the vitamin K epoxide reductase complex (VKORC), and mutations in VKORC's subunit 1 (VKORC1) affect the inhibitory activity, which can be implemented with a reduced EC50 parameter in Wajima et al.'s mechanistic model.

The left panel of Figure 3 illustrates the emergence of inter-individual treatment response variability in our warfarin clinical trial model. The figure shows simulated INR treatment responses of 6 individuals to daily administrations of warfarin. 3 individuals have the GG genotype (VKORC1), the *1/*1 genotype (CYP2C9) and are 71 years old (see blue lines). The remaining 3 individuals have the GA genotype (VKORC1), the *1/*2 genotype (CYP2C9) and are 46 years old (see red lines). The dashed line indicates the treatment response of an average 71 year old individual with the GG and *1/*1 genotypes. We can see that individuals with χ = (GG, *1/*1, 71) tend to respond less strongly to warfarin treatment than individuals with χ = (GA, *1/*2, 46). However, there remains substantial IIV that is not explained by covariates.

3. The inter-occasion model: This component models the variability of the treatment response over time using time-varying modifications of the model parameters

$$\psi \rightarrow \psi'(t) = \psi\,\eta(t). \tag{3}$$

η denotes the alterations of the model parameters, and ψ' denotes the new model parameters. The treatment response of an individual with parameters ψ is now given by the mechanistic model simulation using the time-varying parameters, $\bar{y}(t, r, \psi'(t))$ (Karlsson and Sheiner, 1993). The role of the inter-occasion model is to implement changes of the treatment response that are not accounted for by the mechanistic model. Such changes can be externally driven, e.g., by concomitant food intake or comedication, or of completely unknown origin (Keutzer and Simonsson, 2020). The dynamics of η, $p(\eta|t)$, are a modelling choice.

A source of inter-occasion variability for warfarin is, for example, the time-varying consumption of vitamin K (Xue et al., 2016), changing the amount of vitamin K available in the blood. Since warfarin's mode of action is to inhibit VKORC – a complex converting one form of vitamin K into another, clotting factor-activating form of vitamin K – an increased availability of vitamin K can reverse the treatment effects of warfarin. In the 2nd panel of Figure 3 we illustrate the effects of varying vitamin K consumption on the warfarin treatment response in the clinical trial simulation. The dashed line shows the treatment response simulation with a constant vitamin K input rate parameter, i.e., no variability in the vitamin K consumption. The solid lines show the treatment response simulations with vitamin K input rates that randomly vary by 10% (blue), 20% (red) and 30% (grey) from day to day.

4. The execution model: This component models unintended deviations from nominal dosing regimens and monitoring schedules during the trial. Nominal dosing regimens are defined by a sequence of doses and administration times, $r = \{(d_j, t_j)\}$, where d_j denotes the jth dose and t_j denotes the associated administration time. Nominal monitoring schedules are similarly defined by a sequence of measurement times. The actual doses and times are modelled in the execution model using random deviations from the nominal schedules

$$d_j \rightarrow d_j' = d_j + \Delta d, \qquad t_j \rightarrow t_j' = t_j + \Delta t, \tag{4}$$

where Δd and Δt denote the deviations. The treatment response corresponding to the actual dosing regimen, $r' = \{(d'_j, t'_j)\}$, is simulated using $\bar{y}(t, r', \psi'(t))$. The role of the execution model is to test the robustness of MIPD approaches in clinical practice by emulating the limited control over dose administrations and monitoring. By choosing to report nominal schedules as opposed to actual schedules, the execution model can also be used to mirror common inaccuracies of clinical data. The distributions of dose and time deviations, $p(\Delta d)$ and $p(\Delta t)$, are modelling choices and may differ between dosing and measurement schedules. While not considered in this article, the execution model may be extended to include missed administrations or missed measurements during the trial. Assuming no persistence, this can be implemented using draws of Bernoulli random variables associated with each time point, indicating whether or not a dose was administered or a measurement was taken (Holford et al., 2000; Holford et al., 2010). For infusions, deviations in the duration of the administration may also be modelled.

In the 3rd panel of Figure 3 we illustrate the effects of delayed dose administrations on the treatment response in the warfarin trial simulation. Doses are administered daily. The nominal dosing regimen is illustrated by hollow arrows. The actual dosing regimen with delayed administrations is illustrated by blue arrows. The associated treatment response simulations are indicated by the dashed line (nominal) and by the solid line (actual).

5. The measurement model: This component models the limited accuracy of treatment response measurements

$$y = \bar{y}(t, r', \psi'(t)) + \varepsilon, \tag{5}$$

where y denotes the measurement and ε denotes the measurement error. This defines a distribution of measurements around the mechanistic model output, $\bar{y}$, at each time point t, $p(y|t, r', \psi')$, where measurements may be expected. The wider the measurement distribution, the larger the measurement noise. During the trial simulation, monitoring measurements are sampled from the measurement distribution. This model can be extended to include noise also in the measurement process of covariates, for example to reflect genotyping errors of the VKORC1 or CYP2C9 genes. This is however not considered in this article.

In the 4th panel of Figure 3, we illustrate INR measurements sampled from the measurement distribution during warfarin treatment (blue scatter points). The dashed line shows the treatment response simulation of the mechanistic model without measurement noise.

2.2 Implementation of warfarin trial simulation

We develop the warfarin clinical trial model following the framework for MIPD trial simulation introduced in Section 2.1. The mechanistic model is implemented using Wajima et al.'s model of the humoral coagulation network (see Figure 2). The model provides a mechanistic description of warfarin's PKPD using a system of nonlinear differential equations

$$\frac{da_d}{dt} = -k_a a_d + r(t), \quad \frac{da_c}{dt} = k_a a_d - k_e a_c, \quad \frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}, a_c, \psi). \tag{6}$$

a_d and a_c describe the pharmacokinetics of warfarin and denote the amount of the drug in the dose compartment and the central compartment, respectively. k_a denotes the absorption rate and k_e denotes the elimination rate. $r(t)$ denotes the dose rate and implements the dosing regimen. The pharmacodynamics of warfarin are captured by $\mathbf{x}$, denoting the 51 remaining states of the model. The prothrombin time is simulated as a function of the states, $\bar{y}(t, r, \psi) = g(\mathbf{x}(t, r, \psi), \psi)$, involving the computation of the fibrin AUC after exposure to 300 nM tissue factor. For full details of the mechanistic model, we refer to (Wajima et al., 2009) and Supplementary Appendix S1. A systems biology markup language (SBML) specification is provided on GitHub (https://github.com/DavAug/mipd-warfarin) for simplified cross-platform implementation of the model.

The population model is implemented using a hybrid of two mixed effects models developed by (Hamberg et al., 2010; Hartmann et al., 2016; 2020). Hartmann et al. (2016) provide a mixed effects model extension of Wajima et al.'s model, capturing the treatment response variability emergent from varying production rates of selected coagulation factors, including prothrombin, protein S, protein C and coagulation factors V, VII, IX, X, XI, XII and XIII. In a subsequent publication, they extend their model to include variability explained by covariates, such as the genotypes of the VKORC1 gene and the CYP2C9 gene (Hartmann et al., 2020). VKORC1 is used to model the variability in warfarin's EC50, while CYP2C9 is used to model the variability in warfarin's elimination rate. In our population model, we modify Hartmann et al.'s model further using elements from Hamberg et al.'s model to incorporate age and heterozygosity in the genotypes as covariates of the IIV (Hamberg et al., 2010). This results in a population distribution whose subpopulations, $p(\psi|\theta, \chi)$, are defined by the age of a patient, one of three VKORC1 genotypes (GG; GA; AA) and one of six CYP2C9 genotypes (*1*1; *1*2; *1*3; *2*2; *2*3; *3*3). For full details of the population model, we refer to Supplementary Appendix S1. The population model parameters, θ, used to simulate the clinical trial are provided on GitHub (https://github.com/DavAug/mipd-warfarin).

The inter-occasion model is implemented using time-varying vitamin K input rates (see 2nd panel in Figure 3). The input rate alterations are assumed to be normally distributed

$$p(\eta|t) = \mathcal{N}(\eta|\mu_\eta, \sigma_\eta^2), \tag{7}$$

with constant mean and standard deviation: $\mu_\eta = 1$ and $\sigma_\eta = 0.1$. To allow a change of vitamin K consumption over time, we sample a new η for each simulation day, such that the altered input rate may be interpreted as the daily average of the vitamin K consumption.

The execution model is implemented using exponentially distributed delays of the administration and monitoring times

$$p(\Delta t|\tau) = \frac{1}{\tau} e^{-\Delta t/\tau}, \tag{8}$$

where $\tau = 30$ min denotes the average delay. For each nominal administration time and each nominal monitoring time, we independently sample delays from $p(\Delta t|\tau)$ and compute the

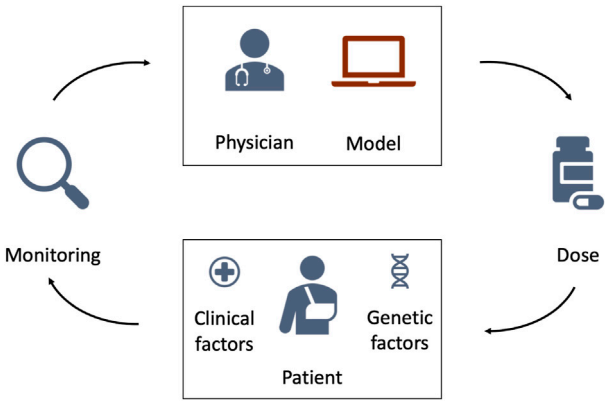

FIGURE 4
Schematic illustration of model-informed precision dosing. The top of the figure shows the treating physician and the MIPD model. The bottom of the figure shows the to-be-treated patient characterised by clinical and genetic factors. The physician uses the MIPD model and the patient characteristics to predict an individualised dose. If the treatment response is monitored over time, the physician can use the treatment response measurements and the MIPD model to iteratively refine the dose predictions.

actual times according to Eq. 4. If dose administrations and monitoring measurements are scheduled at the same nominal times, we only draw one delay random variable for both events.

The measurement model is implemented using lognormally distributed measurements around the mechanistic model output

$$p(y|t, r', \psi') = \mathrm{LN}(y|\mu, \sigma), \tag{9}$$

where $\mu = \bar{y}(t, r', \psi')$ denotes the median, and $\sigma = 0.1$ denotes the scale of the distribution. This implements a measurement error that scales proportionally with the measured quantity, producing measurement errors of approximately 10% of the mechanistic model output.

2.3 MIPD methods

We investigate three MIPD models for warfarin treatment individualisation: 1. a neural network regression model ('Regression model'); 2. a deep reinforcement learning model ('Deep RL model'); and 3. a pharmacokinetic and pharmacodynamic model ('PKPD model'). All three models use data specific to the to-be-treated patient to predict individualised dosing regimens. While the details differ substantially between the models, Figure 4 illustrates their general workflow. The bottom of the figure shows the to-be-treated patient characterised by covariates, such as clinical and genetic factors. The top of the figure shows the treating physician and the MIPD model. The physician uses the model and the patient characteristics to predict an individualised dose. For some MIPD approaches, this dose can be iteratively refined using measurements of the patient's treatment response. Some models can also predict the full future dosing regimen at each iteration rather than just the next dose. Below, we discuss the three methods investigated in this study in more detail.

1. Regression model – This MIPD approach follows Anderson et al. (2012) and Verhoef et al. (2013) and uses a static model of the daily maintenance dose to individualise treatments. The maintenance dose refers to the constant warfarin dose administered daily to maintain a desired INR level. The maintenance dose is modelled as a function of the desired response, $y^\star$, and the patient's covariates, χ,

$$d^\star = d^\star(\chi, y^\star). \tag{10}$$

$d^\star$ denotes the maintenance dose. This makes it possible to predict individualised maintenance doses based on the covariates of an individual. In contrast to the generalised workflow in Figure 4, the model does not iteratively update its predictions using monitoring measurements.

The model can be implemented using a variety of regression approaches, including linear regression, spline regression and tree regression (Klein et al., 2009). In this article, we choose a neural network approach. Neural networks are universal function approximators and can therefore learn to approximate the maintenance dose function in Eq. 10 from data, even when the relationship between the dose and $(\chi, y^\star)$ is nonlinear.

For those who are more familiar with neural network regression, in summary we compose the network of three sequential fully connected layers implemented in PyTorch (Paszke et al., 2019); the two inner layers of this network are of width 1024 and have ReLU activations. The output layer uses a sigmoid activation. The network is trained on simulated trial data (see Section 3.2) to minimise the mean squared error objective function using Adam (Kingma and Ba, 2014). For full details on the implementation and training of the model, we refer to Supplementary Appendix S2.

2. Deep RL model – This MIPD approach follows a deep reinforcement learning approach similar to (Zadeh et al., 2023) and uses a model of the next-to-administer dose to individualise treatments. The dose is modelled as a function of the covariates and the current monitoring data

$$d_j = d_j(\chi, y_j). \tag{11}$$

d_j denotes the dose at time t_j and y_j denotes the INR measurement at time t_j. This makes it possible to iteratively predict individualised dosages based on monitoring data and the covariates of an individual, as illustrated in Figure 4. The target treatment response, $y^\star$, is specified before the training of the model (Zadeh et al., 2023).

Conceptually, RL learns dosing strategies from trial and error: the model sequentially administers dosages and evaluates the 'goodness' of the dose decisions based on the feedback from the treatment response. The learned dosing strategy can be shown to optimally target the desired treatment response under certain technical assumptions and takes the form of a function for the next-to-administer dose (see Eq. 11). We discuss the limitations of these assumptions in Section 3.5. While trial and error in a clinical setting raises ethical questions, RL models can also be trained on treatment response emulators (Ribba et al., 2022). Popular treatment response emulators are PKPD models (Zadeh et al., 2023). To reduce the number of trial and error iterations needed for convergence of the dosing strategy, RL can be performed in conjunction with function approximators (Baird, 1995). In this article, we choose a

deep neural network as the function approximator – an approach commonly referred to as deep RL.

For those who are more familiar with deep RL: we use a DQN model to implement the Deep RL model (Mnih et al., 2013). We compose the network of four sequential fully connected layers implemented in PyTorch (Paszke et al., 2019): three hidden layers of widths (256, 128, 64) with ReLU activations, and the output layer of width 58. The outputs of the network are the predicted Q-values (Mnih et al., 2013). The dose with the maximum Q-value is suggested for administration. Following (Zadeh et al., 2023), the network is trained online using a PKPD model as a treatment response emulator. Prior to the training, the PKPD model is fitted to simulated trial data (see Section 3.2). We train the model to minimise the temporal difference error using Double Q-learning (Van Hasselt et al., 2016) and the Adam optimiser (Kingma and Ba, 2014). For full details on the implementation and training of the model, we refer to Supplementary Appendix S3.

3. PKPD model – This MIPD approach follows a PKPD modelling approach similar to Hamberg et al. (2015) and uses a model of the dosing regimen to individualise treatments. The dosing regimen is modelled as a function of the covariates, the desired treatment response and the monitoring data

$$r = r\left(\chi, y^{\star}, \mathcal{D}_j\right). \tag{12}$$

$\mathcal{D}_j = \{(y_1, t_1), \ldots, (y_j, t_j), r_j\}$ denotes the monitoring data available up to and including time t_j, where we use r_j to denote the administered dosing regimen for $t < t_j$. The function makes it possible to predict individualised dosing regimens based on monitoring data and the covariates of an individual (see Figure 4). The predictions can be iteratively refined as more monitoring data becomes available.

PKPD modelling is a semi-mechanistic modelling approach which makes use of approximate descriptions of the physiological and pharmacological processes to predict the treatment response dynamics. Conceptually, PKPD modelling is similar to the mechanistic model component of the CT model (see e.g., Wajima et al.'s models in Figure 2), but generally model the biological processes in lower detail. An example PKPD model is illustrated in Supplementary Figure S4.7 in Supplementary Appendix S4. By comparing the predicted and desired treatment responses, this approach is able to determine the optimal dosing regimens for each individual. Analogously to the CT model in Section 2.1.2, inter-individual variability is described by differences in the model parameters across individuals. Patient-specific parameters are derived from the patient's covariates and monitoring data.

For those who are more familiar with PKPD modelling, in summary we use Hamberg et al.'s model (Hamberg et al., 2010) implemented in chi (Augustin, 2021) to model the warfarin treatment response dynamics. We fit the model to simulated trial data (see Section 3.2) using hierarchical Bayesian inference and the No-U-Turn sampler (NUTS) (Hoffman and Gelman, 2014) implemented in pints (Clerx et al., 2019). Individualised dosing regimens are predicted in two steps: 1. The model is fit to an individual's monitoring data using the population model and the individual's covariates as prior knowledge (Maier et al., 2020); and 2. The dosing regimen is optimised to minimise the mean squared error between the model predictions and the desired treatment response. We use Bayesian inference and pints' implementation of the adaptive covariance matrix Markov chain Monte Carlo (ACMC) sampler to fit the model to an individual's data. For the dosing regimen optimisation, we use pints' implementation of the covariate matrix adaption evolution strategy (CMA-ES) optimiser (Hansen et al., 2003). For full details on the implementation and the dosing regimen prediction, we refer to Supplementary Appendix S4.

3 Results and discussion

We simulate three MIPD trials – one trial for each MIPD model – and analyse their relative strengths and limitations. To this end, we follow the workflow illustrated in Figure 1 and, first, fit the MIPD models to simulated clinical trial data to emulate a typical starting point for MIPD trials. The data imitate typical data collected during each of the three phases of clinical trials, and are simulated using the CT model. After the model fitting, we simulate the MIPD trials.

3.1 Simulating trial phases prior to MIPD

At present, most clinical trials are conducted not having MIPD in mind. As a result, the data available for the fitting of MIPD models will often not be tailored to the needs of the method. To reflect this practical limitation in our study, we simulate data from three typical phases of clinical trials for the warfarin use case, not taking the data-requirements for MIPD into account. All MIPD models in this article are fitted using only the data from these trials. The simulated trial data as well as the code to reproduce the trials are hosted on GitHub (https://github.com/DavAug/mipd-warfarin).

Trial phase I: Phase I trials are primarily used to establish the safety of drugs and often monitor a drug's absorption, distribution, metabolism and elimination (ADME) in a relatively small cohort of patients. We emulate such a trial by mirroring a clinical trial reported in (Hamberg et al., 2007). In this trial, the pharmacokinetics of $N = 60$ individuals is monitored. Each individual is administered with a single 10 mg dose of warfarin. Following the nominal administration time, the warfarin concentration in the blood is measured at 10 h, 35 h and 60 h after the administration. The data collected during the trial are the warfarin concentration measurements, the dosing regimens and the covariates for each of the 60 individuals. The simulated warfarin concentrations are illustrated in the left panel of Figure 5. The demographics of the cohort are reported in Supplementary Table S1. Pseudo-code outlining the implementation of the trial is presented in Supplementary Algorithm S1.

The panel shows the simulated warfarin concentration measurements. Measurements are illustrated using scatter points. Measurements taken from the same individual are connected using a solid line. The nominal administration time of the warfarin dose is indicated using a blue arrow.

Trial phase II: Phase II trials are primarily used to establish the efficacy of drugs and monitor a drug's pharmacodynamics. We simulate a phase II trial using a design similar to trials reported in

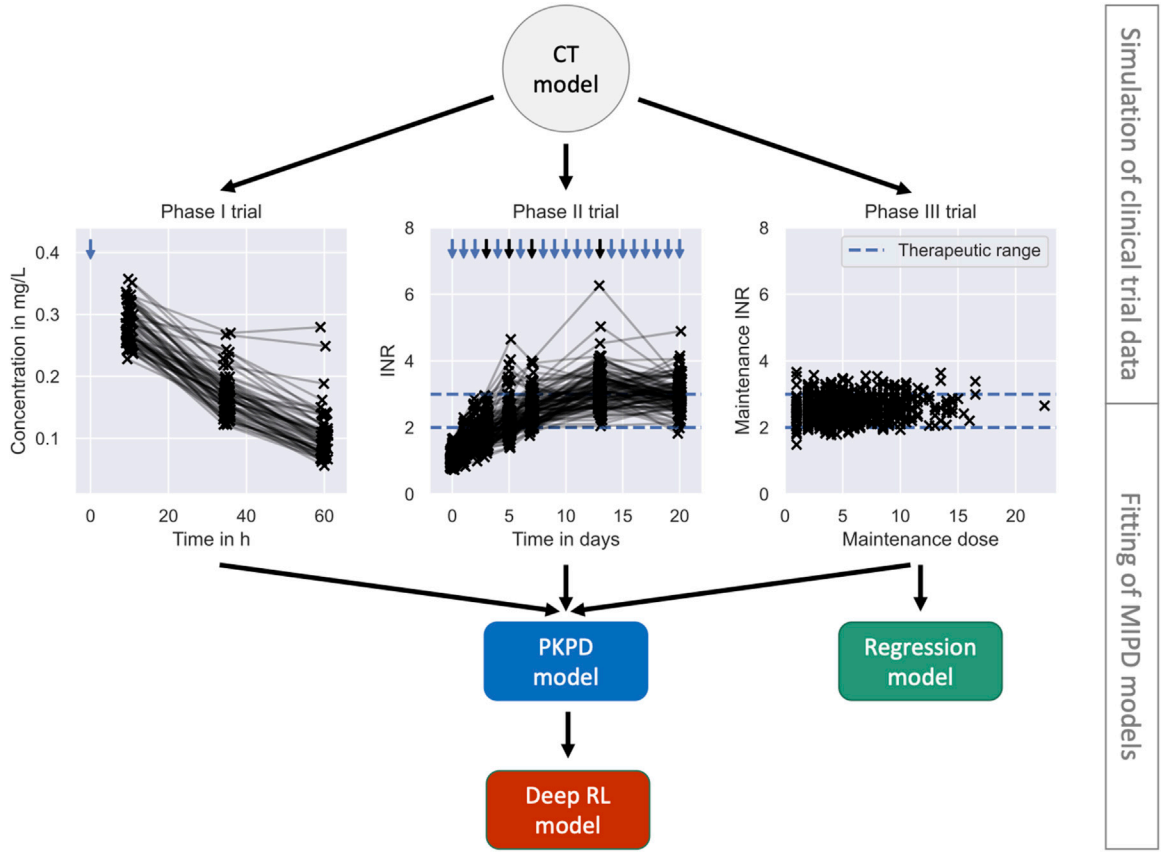

FIGURE 5
Pre-MIPD trial workflow. The figure shows the two steps performed prior to the MIPD trial simulation: 1. Simulation of clinical trial data (top); and 2. Fitting of MIPD models (bottom). The CT model is used to simulate the trial data. The simulated data is illustrated in the middle of the figure: phase I (left); phase II (middle); phase III (right). Measurements of the blood warfarin concentration or the INR are illustrated using scatter points. Measurements connected by solid lines are taken from the same individual at different time points. Nominal administration times are illustrated by blue arrows. Dose administrations with individualised, on-the-fly adjustments of the dose amount are highlighted in black. The therapeutic range is illustrated in the middle and right panel using blue dashed lines. The bottom of the figure shows the fitting workflow of the MIPD models: The PKPD model is fitted to the data from all trials; the Regression model is fitted only to the data from clinical trial phase III; and the Deep RL model is fitted to treatment responses simulated with the PKPD model.

(Hamberg et al., 2010). We monitor the INR response of 100 simulated individuals for 3 weeks. Each individual is treated with daily warfarin doses. During the first 3 days, all individuals receive the same treatment: $d_1 = 10$ mg; $d_2 = 7.5$ mg; and $d_3 = 5$ mg. The 4th dose is adjusted for each individual based on the INR treatment response by a medical professional, here emulated using a simple linear heuristic: $d_j = d_{j-1}\, y^*/y_j$, where y_j denotes the INR measurement taken just before the jth dose administration. This heuristic computes a personalised warfarin dose targeting the desired treatment response, y^*, assuming a linear relationship between the INR and the dose amount. The dose amounts are adjusted three more times for each individual on day 5, 7 and 13 of the trial using the same heuristic. The INR of the individuals is closely monitored during the induction phase of the trial (day 0, day 1, day 2, day 3) and less frequently measured as the trial progresses (day 5, day 7, day 13 and day 20). To emulate safety constrains of real clinical trials, the trial is discontinued when an individual displays three consecutive INR measurements above 5. The data collected during the trial are the INR measurements, the nominal dosing regimens and the covariates for each of the 100 individuals.

The trial was not terminated early for any of the simulated individuals. The simulated INR measurements are illustrated in the middle panel of Figure 5. The demographics of the cohort are reported in Supplementary Table S1. Pseudo-code outlining the implementation of the trial is presented in Supplementary Algorithm S2. The panel shows the simulated INR measurements. Measurements are illustrated using scatter points. Measurements taken from the same individual are connected using a solid line. The nominal administration times of the warfarin doses are indicated using arrows. Black arrows indicate personalised adjustments of the dose amount. The therapeutic range is indicated using blue dashed lines.

Trial phase III: Phase III trials can vary in scope, but tend to involve larger cohorts and have a stronger focus on treatment endpoints. We emulate a phase III trial by mirroring a clinical trial reported in (Klein et al., 2009). The focus of the trial is to understand the variability of the maintenance warfarin dose. We simulate the trial analogously to trial phase II, but with a larger cohort, $N = 1000$, and for a longer duration (8 weeks). Following the initial, identical phase of the trial, the daily doses are adjusted two

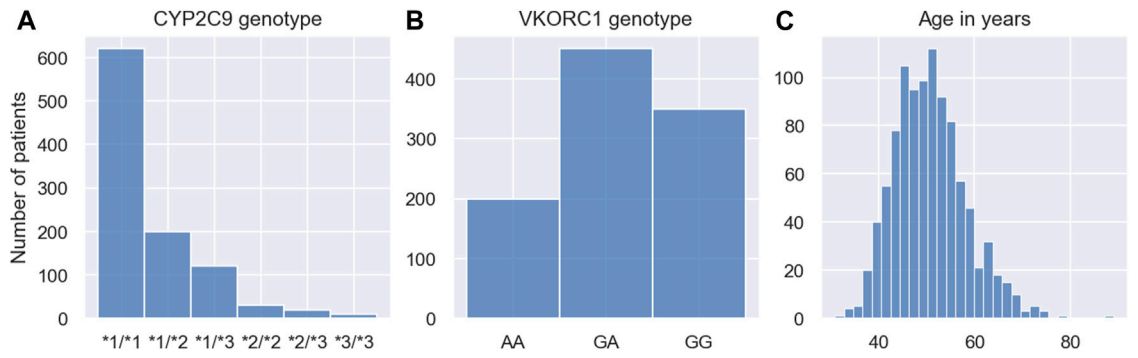

FIGURE 6
Demographics of MIPD trial cohort. The figure shows the CYP2C9 genotype distribution **(A)**, the VKORC1 genotype distribution **(B)** and the age distribution **(C)** of the MIPD trial cohort. The cohort includes 1000 simulated individuals.

more times on day 27 and 34. During the final 3 weeks of the trial, the dose amounts remain unchanged to guarantee that individuals equilibrate to the maintenance treatment response by the end of the trial. To emulate safety constrains of real clinical trials, the trial is discontinued when an individual displays three consecutive INR measurements above 5. The data collected during the trial are the INR measurements at the end of the trial (day 55), the maintenance warfarin doses and the covariates for each of the 1000 individuals.

The trial was not terminated early for any of the simulated individuals. The simulated maintenance doses and INR measurements are illustrated in the right panel of Figure 5. The demographics of the cohort are reported in Supplementary Table S1. Pseudo-code outlining the implementation of the trial is presented in Supplementary Algorithm S3. The panel shows the maintenance dose on the x-axis and the measurement of the INR on the y-axis. Measurements are illustrated using scatter points. The therapEUtic range is indicated using blue dashed lines.

3.2 Fitting the MIPD models

We fit the MIPD models to the simulated clinical trial data, as illustrated in Figure 5. The fitting is the second step of the MIPD trial simulation workflow (see Figure 1). Only one of the models – the PKPD model – can be fitted to all of the available clinical trial data. The Regression model is fitted using the data from the phase III trial. The Deep RL model is trained indirectly on the trial data through simulations from the fitted PKPD model. For details on the fitting, we refer to Section 2.3, Supplementary Appendix S2, Supplementary Appendix S3 and Supplementary Appendix S4.

3.3 Simulating the MIPD trials

We simulate one MIPD trial for each of the three methods in Section 2.3. All trials are conducted using the same cohort. The cohort includes $N = 1000$ individuals. The demographics of the cohort are modelled after a trial reported in (Hamberg et al., 2010) and are visualised in Figure 6. The figure shows the CYP2C9 genotype distribution, the VKORC1 genotype distribution and the age distribution in the cohort.

In each trial, individuals are treated with daily warfarin doses for 19 days. The dose amounts are individualised using the respective MIPD model, as described in Section 2.3, and target a treatment response of $y^* = 2.5$. The data available for the individualisation are the INR measurements taken daily before each dose administration and the covariates of the individual. The Regression model individualises the treatments by predicting the maintenance dose based on the covariates of the patients. This maintenance dose is administered every day throughout the trial. The Deep RL model and the PKPD model predict the warfarin doses iteratively based on a patient's covariates and INR measurements. In contrast to the simulated trials in the previous section, we do not emulate safety constrains for the MIPD trials in order to expose possible weaknesses of the models more clearly. Pseudo-code outlining the implementation of the trial is presented in Supplementary Algorithm S4.

3.4 Results of the MIPD trials

We use three metrics to quantify the success, the safety, and the efficiency of the dosing regimen individualisation: the maintenance INR; the peak INR; and the time in the therapeutic range (TTR). The success of the individualisation is quantified using the maintenance INR measured on the last day of the trial. INR measurements inside the therapeutic range indicate success, while measurements outside the therapeutic range indicate poor dosing regimen individualisation. The safety of the individualisation is quantified using the largest INR measurement recorded during the trial. This peak of the INR response indicates the risk for major bleeding events while transitioning into maintenance treatment. The efficiency of the individualisation is quantified using the TTR, i.e., the number of INR measurements inside the therapeutic range. For successful individualisations, the TTR indicates how quickly the desired treatment response has been achieved.

The results of the trials are visualised in Figure 7. Row 1 shows the results for the Regression model, row 2 shows the results for the Deep RL model, and row 3 shows the results for the PKPD model. The left panel shows the maintenance INR distribution in the cohort. INRs inside the therapeutic range (see blue dashed lines)

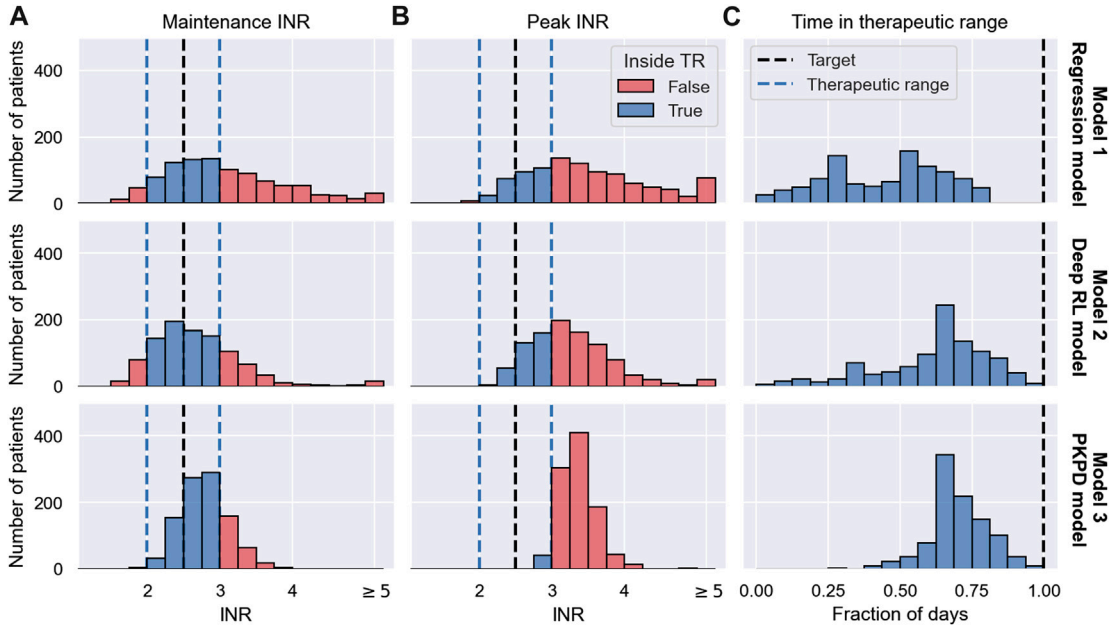

FIGURE 7
MIPD trial results. The figure shows the outcome of three MIPD trials conducted with an identical cohort of size N = 1000 with different MIPD models. The top row shows the results for the Regression model; the middle row shows the results for the Deep RL model; and the bottom row shows the results for the PKPD model. The outcome of the trials is illustrated using three metrics: the maintenance INR measured on the last day of the trial **(A)**; the largest INR value measured during the trial **(B)**; and the time in the therapeutic range **(C)**. The panels show the distributions of these metrics across individuals. The therapeutic range is illustrated using blue dashed lines. INR measurements inside the therapeutic range are highlighted in blue, and INR measurements outside the therapeutic range are highlighted in red. Target values of the dosing regimen individualisation are visualised using black dashed lines.

are highlighted in blue and INRs outside the therapeutic range are highlighted in red. The target INR is illustrated using a black dashed line. The middle panel shows the peak INR distribution in the cohort. The right panel shows the TTR distribution across individuals. The target TTR is illustrated using a dashed line.

The maintenance INR distributions (left panel) show that in a time span of 19 days all three MIPD methods are able to successfully target the therapeutic window for a large number of individuals. The Regression model successfully individualises the dosing regimen for 47.0% of the individuals, while the Deep RL model and the PKPD model have success rates of 65.8% and 75.1%, respectively (see blue histograms). For the remaining individuals, the severity of the failed dosing regimen individualisation varies substantially. For the Regression model, 149 individuals display maintenance INRs above 4 with the largest value being 7.51. For the Deep RL model, only 32 individuals exceed maintenance INRs of 4. However, the largest maintenance INR is 29.03 — a value almost 4 times larger than for the Regression model, raising serious safety concerns. For the PKPD model the largest maintenance INR is 3.90. 97.8% of the individuals display maintenance INR measurements less than 3.5, showing that the PKPD model is able to most consistently achieve maintenance INRs close to the therapeutic window.

The peak INR distributions (middle panel) show that, for the majority of the cohort, the methods are not able to individualise the dosing regimens without overshooting the therapeutic window. In the Regression model trial, 67.1% of the individuals display peak INR measurements above the therapeutic range. The Deep RL model controls the treatment response marginally better, overshooting the therapeutic range for 65.1% of the individuals. The PKPD model misses the therapeutic window for almost all individuals (95.9%) before reaching maintenance treatment. However, the panel also shows that the largest INR value measured in the PKPD model trial is substantially smaller than the largest INR values in the other two trials (Regression model: 7.99; Deep RL model: 29.03; PKPD model: 4.91), indicating that the PKPD model is the safest MIPD approach among the tested models.

The TTR distributions (right panel) show that the time spent inside the therapeutic range varies between individuals and MIPD approaches. For the Regression model, the median TTR is 45% of the trial duration, with TTRs ranging between 0% and 85%. The other two methods achieve substantially larger TTRs across individuals. For the Deep RL model, the median TTR is 68% of the trial duration, with individual values ranging between 5% and 95%. The PKPD model achieves a median TTR of 74%, with a minimum TTR of 26% and a maximum TTR of 100%. This shows that across individuals the PKPD model takes the least time to successfully reach the therapeutic window.

3.5 Degrees of dosing regimen individualisation

The simulated trials in Section 3.4 show that different MIPD approaches have different strengths and limitations. In this section, we study the dosing strategies of the models in more detail to gain a

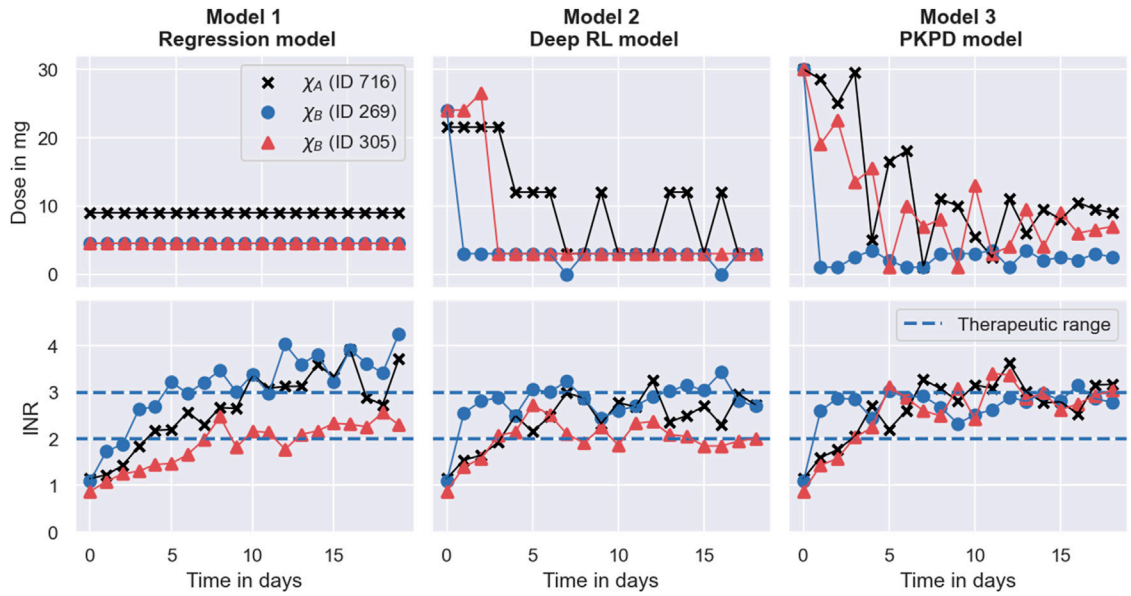

FIGURE 8
Degrees of dosing regimen individualisation. The figure shows the dosing regimen individualisations achieved by the MIPD models in the simulated MIPD trials for three representative individuals. The left panel shows the results for the Regression model, the middle panel shows the results for the Deep RL model, and the right panel shows the results for the PKPD model. The top row illustrates the administered dose amounts and the bottom panel illustrates the INR monitoring data. Dose amounts or measurements belonging to the same individual are connected using solid lines. The therapeutic range is illustrated using dashed lines. One of the individuals (ID 716) is characterised by the covariates χ_A = (*1* 1, GG, 50). The other two individuals (ID 269; ID 305) are both characterised by the covariates χ_B = (*1* 2, GA, 50).

better understanding about their practical and methodological differences. We pay particular attention to attributing generic strengths and limitations to the methodology and specific strengths and limitations to our implementation.

We investigate the dosing strategies by studying the dose decisions suggested by each of the models for three representative individuals from the simulated trials in Section 3.4. The first individual, with ID 716, is characterised by the covariates χ_A = (*1*1, GG, 50). The other two individuals (ID 269; ID 305) are both characterised by the covariates, χ_B = (*1*2, GA, 50). The different dosing strategies and treatment responses are visualised in Figure 8. The figure shows the doses administered during the trials in the top panel and the corresponding treatment response measurements in the bottom panel. Doses or measurements belonging to the same individual are connected using solid lines. The therapeutic range is illustrated using dashed lines. The left panel shows the trial results for the Regression model, the middle panel shows the trial results for the Deep RL model, and the right panel shows the trial results for the PKPD model.

3.5.1 The Regression model

The left panel of the figure shows that the Regression model predicts a maintenance dose of 9 mg for the individual with ID 716 (black scatter points), and a maintenance dose of 4.5 mg for the other two individuals (see top left panel in Figure 8). These maintenance doses are administered daily throughout the trial. For the individual with ID 305, the maintenance INR measurement at the end of the trial is inside the therapeutic range, while the maintenance INR measurements for the other two individuals overshoot the therapeutic range (see bottom left panel in Figure 8). Notably, the individual with the successful individualisation (ID 305) has the same covariates as one of the individuals with the failed individualisation (ID 269).

This illustrates both a strength and a limitation of the Regression model: a strength of the model is that it is able to predict individualised dosages using information only about the covariates of individuals, making the model easier to implement in clinical practice than monitoring-based approaches. However, the figure also shows that dosing regimens exclusively derived from covariates will, at best, successfully target the desired treatment response for an average individual characterised by the covariates. Inter-individual differences that are not explained by covariates are not accounted for. In this case, the individual with ID 305 happens to respond similarly to an average individual[1] with the covariates, χ_B, resulting in a successful dosing regimen individualisation. The individual with ID 269, on the other hand, responds more strongly to warfarin than an average individual[2], yielding a maintenance INR above the therapeutic range. Not being represented by an average individual also explains the failed individualisation for the individual with ID 716. The inability to account for unexplained IIV is a generic limitation of approaches exclusively based on covariates.

1 In this case, the close-to-average response can be explained by the individual's model parameters, ψ, being close to the means of the population distribution (Eq. 2).

2 Similarly to the above footnote, the stronger-than-average response can be explained by key model parameters being further away from the means of the population distribution.

TABLE 1 Summary of the strengths and limitations of the MIPD models. The properties of the models are grouped into data-related properties (rows 1–3), variability-related properties (rows 4–8), and strategy-related properties (rows 9–13). The parentheses around the property of the PKPD model in the top right corner indicate that the model, as defined in Section 2.3, can be used without covariates, but, in the simulated MIPD trial in Section 3.4, we did not explore this possibility. Similarly, parentheses around the property below indicate that the PKPD model can be used without the use of monitoring data.

	Regression model	Deep RL model	PKPD model
Requires covariates	✓	✓	(✓)
Requires monitoring	✗	✓	(✓)
Requires indefinite monitoring	✗	✓	✗
Accounts for explained IIV	✓	✓	✓
Accounts for unexplained IIV	✗	✓	✓
Accounts for IOV	✗	✓	✓
Accounts for EV	✗	✓	✓
Accounts for treatment response delays	✗	✗	✓
Robust to unseen treatment responses	✓	✗	✓
Robust to model misspecification	✓	✗	✗
Learns individual-specific dosing regimens	✗	✗	✓

In addition to this limited ability to account for IIV, the figure also shows that the Regression model is incapable of accounting for inter-occasional variation and differences in the execution of the treatment. This limitation is, again, a direct consequence of exclusively using covariates for the dosing regimen individualisation. Without quantifying the individual-specific IOV and EV, predicted dosing regimens can, at best, be successful when the IOV and the EV of the treated individual happen to be close to the average IOV and EV observed in the dataset used for the model fitting.

Another limitation, illustrated in Figure 8, is that the Regression model cannot account for treatment response delays. The model only predicts maintenance dosages, and therefore provides no guidance for individualisation of the induction phase of the treatment. In our implementation, we choose to overcome this limitation of the model by administering the predicted maintenance dosages from the beginning of the trial, not attempting to individualise induction dosing regimens without model guidance. This has the consequence that treatment responses take time to reach their maintenance level, limiting the efficiency attainable by the model. From Figure 8, we can, for example, see that the INR measurements of the individual with ID 305 reach the therapeutic range for the first time on day 8 of the treatment. After that, two more measurements are outside the therapeutic range, giving rise to a TTR of 10 days during the 19 days trial. Together with the top right panel in Figure 7, this indicates that even when the Regression model individualises maintenance dosages successfully, it achieves, at best, an average TTR of around 55%–65%. This limit to the efficiency is specific to the treatment response delay of warfarin and our implementation of the Regression model.

A summary of the strengths and limitations specific to the Regression model are presented in the left column of Table 1.

3.5.2 The Deep RL model

In comparison, the Deep RL model is able to predict more individualised dosing regimens than the Regression model (see middle panel in Figure 8). The panel shows that for all three individuals, the model begins the treatment with increased warfarin dosages of more than 20 mg. During the maintenance phase, it reduces the dosages. For the individual with covariates χ_A (ID 716), the model alternates between administering dosages of 12 mg and 3 mg. For the other two individuals with covariates χ_B, the model administers a constant dose of 3 mg, from which it only occasionally deviates for one of the individuals (ID 269). Overall, the figure shows that the Deep RL model is more successful in individualising the dosing regimens of the individuals, as indicated by their maintenance INR measurements at the end of the trial. The individualisation of the induction phase reduces the time needed to reach the therapeutic range, leading to all three individuals displaying INR measurements inside the therapeutic range within the first four treatment days.

The main reason for the improved performance of the Deep RL model is the use of feedback control from the monitoring data in addition to the covariate information. The model derives predictions from both monitoring data and covariates using the dose function, defined in Eq. 11, which predicts the next-to-administer dose based on the most recently measured INR value and the covariates of the to-be-treated individual. We visualise the predictions of the dose function for the three individuals in Figure 9. For illustrative purposes, the figure focuses on INR values between 0.5 and 7. The predicted dose values are illustrated using a black line for individuals with the covariates χ_A and a red-blue line for individuals with the covariates χ_B. The target treatment response is illustrated using a black dashed line, and the therapeutic range is indicated using blue dashed lines.

The figure shows that the Deep RL model has a clear strategy for INR monitoring measurements below 4, and a less clear strategy for INR measurements above 4. We will first focus on the dose decisions

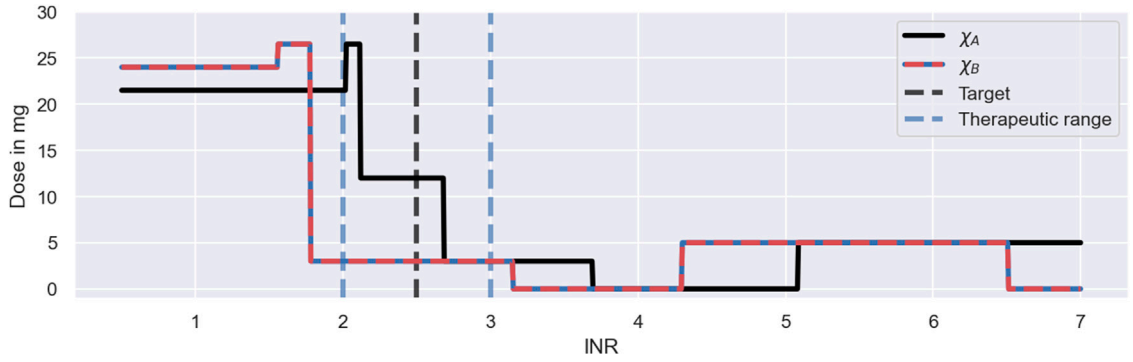

FIGURE 9
Dosing strategy of the Deep RL model. The figure illustrates the dose function of the Deep RL model, defined in Eq. 11, for a range of possible INR monitoring measurements and two sets of covariates: χ_A = (*1*1, GG, 50) (black line); and χ_B = (*1*2, GA, 50) (red-blue line). The function determines the next-to-administer dose based on the most recently measured INR value (bottom axis) and the covariates of an individual. The target treatment response is illustrated using a black dashed line. The therapeutic range is indicated using blue dashed lines.

for INRs below 4: for INR measurements below the therapeutic range, the model administers large warfarin doses; for INRs inside the therapeutic range, the model administers intermediate warfarin doses; and for INRs above the therapeutic range, the model administers low warfarin doses. The exact change points and dose amounts are specific to the covariates. For example, for individuals with the χ_A covariates, the model starts the treatment with a dose of 21.5 mg and switches to an intermediate dose of 12 mg for INR measurements between 2.1 and 2.7 (see black line in Figure 9). This dosing strategy is consistent with the dose decisions observed in Figure 8.

The dose function illustrates both a strength and a limitation of the Deep RL model (see Figure 9). A strength of the model is that it bases its dose decisions on a simple and interpretable feedback control mechanism: when INR values are too low, it increases the warfarin dose; and when INR values are too high, it decreases the warfarin dose. These dose decisions are tailored to an individual based on the individual's covariates. This strategy leads to a substantially higher success rate and efficiency of the dosing regimen individualisation relative to the Regression model (see Figure 7), as the feedback control enables the model to maintain treatment responses within a desired range, even when unexplained IIV, IOV and EV are present. This strength is generic to reinforcement learning approaches.

However, while the ability to utilise monitoring data for control is a strength, the model's inability to learn from the measurements is a limitation of the Deep RL model, making the approach indefinitely reliant on monitoring data. The reliance on monitoring data is specific to the Deep RL model and is a consequence of its learning approach: the model establishes the dose function (Eq. 11) prior to the treatment of individuals (see Section 3.2), never individualising the dose function based on the individual-specific monitoring data. Instead, the model treats individuals with dosing strategies exclusively based on covariates, where the monitoring data is only used as a control mechanism to steer treatment responses back to the desired treatment response, if needed (see e.g., the dosing strategy in Figure 8 for ID 269). In principle, reinforcement learning approaches can update the dose function using individual-specific monitoring data (Maier et al., 2021). However, for the Deep RL model, meaningful updates of the dose function are challenging, as the large number of model parameters of its neural network complicates the balance between fine-tuning and overfitting to the limited number of measurements.

The absence of fully individualised dosing strategies explains the observed tendency of the Deep RL model to overshoot the therapeutic range (see middle panel in Figure 7): to successfully target the therapeutic range across individuals with the same set of covariates, the initial warfarin dose predicted by the dose function needs to be high enough, so that even the weakest responders in a subpopulation reach the therapeutic range. This dose leads to INR responses above the therapeutic range for strong responders in the same subpopulation due to the substantial level of warfarin IIV not explained by covariates. The over-treatment of the strong responders is compensated for by reducing the dose for INRs above the therapeutic range so much that even the strongest responders are guaranteed to regress back to the therapeutic range (see large dose steps in Figure 9). This explains the substantial fraction of individuals with peak INRs above the therapeutic range in Figure 7, and points to a general lack of precision of the Deep RL model. This 'control over precision' strategy is a generic limitation of reinforcement learning approaches that do not individualise their dose functions based on the individual-specific monitoring data.

The dose function in Figure 9 also demonstrates that the Deep RL model can fail to learn meaningful dosing strategies for the full range of possible INR measurements. When INR measurements become large (≥ 5.05 for χ_A; and ≥ 4.3 for χ_B), the model suggests increasing the warfarin dose to 5 mg, despite the fact that higher warfarin dosages inevitably lead to even higher INR values. The root for these dose decisions lies in the training of the Deep RL model (see Section 3.2): large INR monitoring measurements remain under-explored during the training, leading to poorly tested dose decisions for large INR measurements. Those dose decisions can remain without consequence, when the model is able to control treatment responses well enough to never reach large INR values (see middle panel in Figure 8). But when an individual responds more strongly to warfarin than expected, for example due to unexplained IIV, IOV or EV, poorly tested dose decisions for

large INR values can cause a failure of the model's feedback control mechanism. This failure explains the severe over-treatments of a few individuals observed in the simulated MIPD trial in Figure 7.

The lack of exploration during the training, despite the use of a standard training procedure (see ϵ-greedy policy in Supplementary Appendix S3), is the consequence of a mismatch between the technical assumptions of the Deep RL model and the reality of warfarin treatment responses: the Deep RL model assumes that there is no delay between dose administrations and the feedback from INR measurements, when, in fact, warfarin treatment responses have delays of up to 10 days (see e.g., Figure 3). This reduces the ability of the standard training procedure to contribute to the exploration of large INR measurements. The mismatch between the model assumptions and the treatment response delay is generic to reinforcement learning models and difficult to overcome, as the convergence and optimality of reinforcement learning centrally relies on the assumption that transition dynamics can be modelled by a Markov decision process with i.i.d. actions and i.i.d. states (Sutton and Barto, 2018). For the Deep RL model, this implies that the model has to assume that INR measurements on 1 day depend only on the INR measurement and the administered dose on the previous day, i.e., any dose administrations and INR measurements on earlier days cannot be taken into account. Despite those technical constraints, Zadeh et al. (2023) show that their deep reinforcement learning model is able to improve its performance, when the i.i.d. assumption of the states is explicitly violated and dose decisions are conditioned on more than one recent INR measurement. As a result, the technical assumptions of reinforcement learning, needed for convergence and optimality, may limit the ability of reinforcement learning models to account for treatment response delays, in theory, but in practice, it may be possible to overcome those limitations (Gaon and Brafman, 2020).

A summary of the strengths and limitations specific to the Deep RL model are presented in the middle column of Table 1.

3.5.3 The PKPD model

The right panel in Figure 8 shows that the PKPD model achieves the highest degree of dosing regimen individualisation among the tested models. The model begins the treatment for all three individuals with a high dose of 30 mg, followed by a gradual reduction of the dose over the following days. For the individual with ID 269, this dose reduction is more rapid than for the other two individuals. Towards the end of the trial, the dosages converge to constant, individual-specific maintenance dosages. The maintenance INRs are located at the upper threshold of the therapeutic range in close proximity to each other, indicating success of the dosing regimen individualisation. All three individuals display INR measurements inside the therapeutic range within the first three treatment days.

The main reason for the good performance of the PKPD model is its use of both covariate information and monitoring data similar to the Deep RL model. The model uses the covariates to predict initial dosing regimens that target the desired treatment response for average individuals. These initial dosing regimens are further individualised based on the individual-specific monitoring data, achieving an increasing degree of individualisation as more monitoring data becomes available. This increasing degree of individualisation provides a distinct advantage over the other two MIPD models. The PKPD model derives predictions from both monitoring data and covariates using the dosing regimen function, defined in Eq. 12, which predicts individualised dosing regimens based on all available INR measurements of the to-be-treated individual, the covariates, and the already-administered dosages. We visualise the values of the dosing regimen function for the three individuals in Figure 10. For illustrative purposes, the figure focuses on 4 days of the simulated trial: day 1; day 2; day 7; and day 16. The predicted dose values are illustrated using scatter points in two opacity levels: opaque scatter points indicate already administered dosages; and faded scatter points indicate future dose administrations. The treatment day is illustrated using dashed lines.

The figure shows that the dosing regimen predictions are iteratively updated as the treatment progresses. On day 1, the model provides rough estimates of the dosing regimens, scheduling maintenance dosages of 8 mg (ID 716), 4 mg (ID 269), and 5.5 mg (ID 305) after an initial induction phase of the treatment. With the next INR measurement on day 2, both the induction dosing regimen as well as the maintenance dosages are updated (see second panel in Figure 10). On day 16, the model predicts fully individualised dosing regimens which are almost identical to the dosing regimens administered in the simulated trial (see right panel in Figure 8).

The dosing regimen function illustrates both a strength and a limitation of the PKPD model (see Figure 10). A strength of the model is that it predicts full dosing regimens from any number of monitoring measurements, making the dosing schedule transparent, foreseeable, and less reliant on frequent or regular monitoring. The prediction of full dosing regimens is enabled by the model's explicit description of the pharmacological processes in terms of a semi-mechanistic model. This permits the prediction of treatment responses, and thus optimal dosing regimens, for any future time points. It also helps the model to account for treatment response delays and nonlinearities of the dose-response relationship (see Section 2.3). However, the explicit model of the treatment response bears a risk for model misspecification (Merlé et al., 2004). In particular, neglecting or oversimplifying important treatment response mechanisms can lead to inaccurate treatment response predictions, which, in turn, can impact the quality of the dosing regimen individualisation. For example, the results from the simulated MIPD trial show that the PKPD model tends to administer too large warfarin dosages during the trial, resulting in a systematic bias towards INR measurements larger than the target INR (see bottom panel of Figure 7). This indicates that the PKPD model oversimplifies crucial elements of the treatment response mechanisms, resulting in a tendency to underestimate the treatment response of individuals. The risk for model misspecification is a generic limitation of PKPD modelling, which needs to be mitigated prior to clinical applications, for example by quantifying the structural uncertainty of PKPD models using model selection criteria or probabilistic model averaging (Uster et al., 2021; Augustin et al., 2022).

The dosing regimen function in Figure 10 also illustrates that the updates of the dosing regimens themselves can result in a bias of the treatment strategy. The comparison between the predicted dosing regimens shows that although more individual-specific monitoring measurements are available on day 7 of the treatment, the maintenance dosages predicted on day 2 are closer to the actual maintenance dosages administered towards the end of the trial. This suggests that the degree of the dosing regimen individualisation can temporarily decrease with the number of monitoring measurements

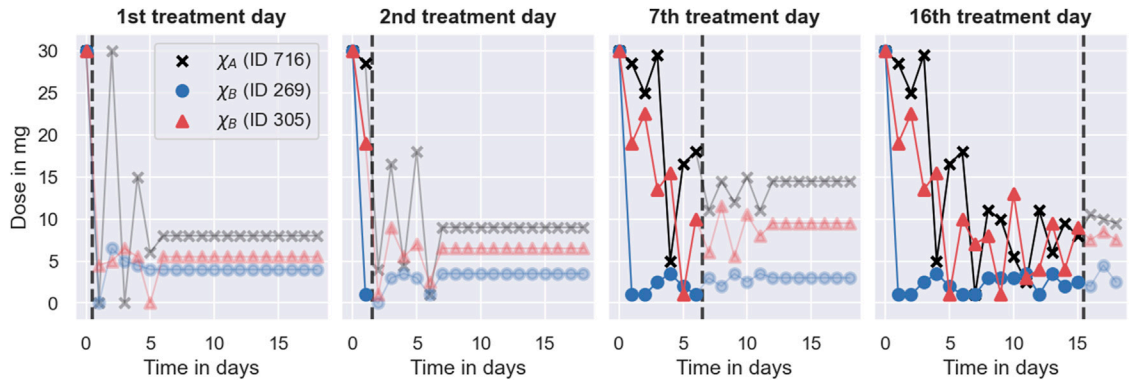

FIGURE 10
Dosing strategy of the PKPD model. The figure illustrates the dosing regimen function of the PKPD model, defined in Eq. 12, for the three individuals from Figure 8 on 4 days of the trial: on the 1st treatment day (panel 1), on the 2nd treatment day (panel 2); on the 7th treatment day (panel 3) and on the 16th treatment day (panel 4). Dosages are illustrated using scatter points in two opacity levels: opaque scatter points for already administered dosages; faded scatter points for future dose administrations. Dosages belonging to the same individual are connected using solid lines. The day of the treatment is illustrated using black dashed lines.

– a limitation specific to our implementation of the dosing regimen individualisation. The potential for worse dosing regimens despite more monitoring data is related to the estimation of the individual – specific model parameters from the monitoring data: we estimate the model parameters using Bayesian inference (see Section 2.3). The result of this estimation is a distribution of parameter values consistent with the data, also known as posterior distribution. In our implementation, we estimate the individual-specific model parameters by the modes of the distribution, also known as maximum *a posteriori* (MAP) estimates. The MAP estimates are a popular choice to reduce posterior distributions to just one set of model parameters (Sheiner et al., 1979). However, by disregarding the other model parameters that are also consistent with the data, it is possible to introduce biases in the treatment response predictions with consequences for the dosing regimen individualisation. In particular for nonlinear treatment responses, the treatment response predicted with the MAP estimates is, generally, not the treatment response that maximises the predictive probability (Maier et al., 2020). This increases the risk for inaccurate treatment response predictions.

A bias of the MAP-based treatment response predictions explains the lack of safety observed during the simulated MIPD trial (see bottom panel in Figure 7). The MAP-based predictions tend to underestimate the warfarin treatment response for small INR values which leads to individualised dosing regimens with elevated dose amounts early in the trial. As the treatment progresses, the uncertainty about the model parameters becomes smaller, reducing the error of the MAP estimation, and with it, the bias of the treatment response predictions (see maintenance INR distribution in Figure 7). This bias is also supported by the goodness-of-fit plot from the model fit to the pre-MIPD trial data in Section 3.2, where also the predictions with the maximum probability parameters of the population distribution show a tendency to provide biased treatment response predictions (see middle and bottom right panel in Supplementary Figure S4.9). This limitation of our PKPD model can be mitigated by predicting the treatment response with each parameter set that is consistent with the data, i.e., the parameters in the posterior distribution, producing a distribution of treatment responses which reflects the uncertainty in the treatment response predictions (Maier et al., 2020). The distribution of treatment responses can be optimised to obtain a dosing regimen with less risk for bias (Maier et al., 2021).

A summary of the strengths and limitations specific to the PKPD model are presented in the right column of Table 1.

4 Conclusion

Simulated clinical trials provide a resource-efficient way to test and develop fit-for-purpose models for precision dosing. We show that we can emulate clinical trials using a clinical trial model with five independent model components: 1. a mechanistic model; 2. a population model; 3. an inter-occasion model; 4. an execution model; and 5. a measurement model. Each model component captures a different complexity of clinical practice that challenges the successful individualisation of treatments, ranging from PKPD-related challenges, such as nonlinear and delayed treatment responses, to practical challenges, such as unintentional deviations from nominal dosing schedules (see Section 2.1). The modularity of the model components simplifies the development process of the clinical trial model, allowing for independent updates of its components throughout the drug development pipeline. This makes it possible to iteratively improve the trial simulations and develop a promising companion MIPD tool as more understanding and information about the drug under trial becomes available (Polasek et al., 2019).

Simulating trials for precision dosing of warfarin, we find that different MIPD models have different strengths and limitations. These strengths and limitations can be generic to the methodology or specific to the model implementation. Modelling approaches that predict dosing regimens exclusively based on covariates of the treatment response variability are generically limited when unexplained IIV, IOV and EV are present (see Section 3.5.1). However, when the majority of the treatment response variability can be explained by covariates, and those covariates are available in clinical practice, such approaches provide an excellent solution to the

individualisation of dosing regimens. Otherwise, MIPD models based on monitoring data are better at accounting for treatment response variability, and achieve a higher degree of dosing regimen individualisation in the simulated warfarin trials (see Sections 3.4, 3.5).

But there are also challenges with monitoring-based MIPD approaches. We find that the Deep RL model adopts a 'control over precision' treatment strategy, where doses are adjusted based on the feedback response from the monitoring data with limited foresight (see Section 3.5.2). This lack of precision makes the model reliant on indefinite monitoring and limits its ability to account for treatment response delays. Deep reinforcement learning may, nevertheless, provide a good solution to the individualisation of dosing regimens for applications where monitoring is not a challenge, such as for treatments in the intensive care unit (Moore et al., 2004) or for dosing devices that are physically attached to patients, like insulin pumps (Zhu et al., 2020).

The PKPD model achieves the highest degree of dosing regimen individualisation among the tested models. The model predicts fully individualised dosing regimens whose level of individualisation increases with the amount of available monitoring data. This indicates that across applications PKPD models are the most promising approach for precision dosing. However, PKPD models are more susceptible to model misspecifications than the other two approaches (see Section 3.5.3), necessitating a careful evaluation of the predictive uncertainty of the model prior to MIPD, for example by means of model selection criteria or probabilistic model averaging (Augustin et al., 2022).

Overall we find that MIPD approaches vary in their ability to account for the different sources of treatment response variability, meaning an ideal approach depends on the context. Distinguishing four sources of treatment response variability – 1. unexplained IIV; 2. explained IIV; 3. IOV; and 4. EV – our results suggest that PKPD modelling is more successful than the other two approaches in cases where the treatment response variability is dominated by unexplained IIV. PKPD models account for unexplained IIV by individualising treatment response predictions from just a small number of monitoring measurements. We expect PKPD models to also be the favourable choice when the treatment response variability is dominated by EV, as the other tested models cannot adapt their predictions to irregular administration and monitoring schedules. However, when IOV dominates the treatment response variability, deep reinforcement learning may perform best, as its 'control over precision' dosing strategy is agnostic to random fluctuations in the treatment response. Lastly, when the treatment response variability is dominated by explained IIV, all three models likely provide similar maintenance predictions, making regression the preferred choice as long as treatment response delays are not a concern.

While the sources of the variability can indicate which MIPD approach may perform best, the volume and type of the available data determine which MIPD approach is possible. PKPD modelling works with both large and limited amounts of data, provided the treatment response mechanisms are well known and the available data can be linked to the processes involved. In contrast, regression and deep reinforcement learning need more data to be feasible but can process more complex data types and do not rely on a mechanistic understanding of the treatment response.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

DA: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing–original draft, Writing–review and editing. BL: Supervision, Writing–review and editing. MR: Supervision, Writing–review and editing. KW: Supervision, Writing–review and editing. DG: Funding acquisition, Supervision, Writing–review and editing.

Funding

This work was supported by the UK Engineering and Physical Sciences Research Council (grant number EP/S024093/1); and the Biotechnology and Biological Sciences Research Council (grant number BB/P010008/1). DA acknowledges EPSRC for studentship support via the Doctoral Training Centre in Sustainable Approaches to Biomedical Science: Responsible and Reproducible Research, as well as the Clarendon Fund for studentship support. BL, MR, and DG acknowledge support from the EPSRC Centres for Doctoral Training Programme. DG acknowledge support from a Biotechnology and Biological Sciences Research Council project grant. KW are employees of F. Hoffmann La Roche Ltd. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of interest

Author KW was employed by F. Hoffmann-La Roche AG.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2023.1270443/full#supplementary-material

References

Abrantes, J. A., Jönsson, S., Karlsson, M. O., and Nielsen, E. I. (2019). Handling interoccasion variability in model-based dose individualization using therapeutic drug monitoring data. *Br. J. Clin. Pharmacol.* 85, 1326–1336. doi:10.1111/bcp.13901

Anderson, J. L., Horne, B. D., Stevens, S. M., Woller, S. C., Samuelson, K. M., Mansfield, J. W., et al. (2012). A randomized and clinical effectiveness trial comparing two pharmacogenetic algorithms and standard care for individualizing warfarin dosing (coumagen-ii). *Circulation* 125, 1997–2005. doi:10.1161/CIRCULATIONAHA.111.070920

Augustin, D. (2021). *Chi - an open source python package for treatment response modelling*.

Augustin, D., Lambert, B., Wang, K., Walz, A.-C., Robinson, M., and Gavaghan, D. (2023). Filter inference: a scalable nonlinear mixed effects inference approach for snapshot time series data. *PLOS Comput. Biol.* 19, e1011135. doi:10.1371/journal.pcbi.1011135

Augustin, D., Walz, A.-C., Wang, K., Lambert, B., Clerx, M., Robinson, M., et al. (2022). Treatment response prediction: is model selection unreliable? *bioRxiv*, 2022–2103. doi:10.1101/2022.03.19.483454

Azer, K., Kaddi, C. D., Barrett, J. S., Bai, J. P., McQuade, S. T., Merrill, N. J., et al. (2021). History and future perspectives on the discipline of quantitative systems pharmacology modeling and its applications. *Front. physiology* 12, 637999. doi:10.3389/fphys.2021.637999

Baird, L. (1995). "Residual algorithms: reinforcement learning with function approximation," in *Machine learning proceedings 1995* (Elsevier), 30–37.

Broeker, A., Nardecchia, M., Klinker, K., Derendorf, H., Day, R., Marriott, D., et al. (2019). Towards precision dosing of vancomycin: a systematic evaluation of pharmacometric models for bayesian forecasting. *Clin. Microbiol. Infect.* 25, 1286. doi:10.1016/j.cmi.2019.02.029

Clerx, M., Robinson, M., Lambert, B., Lei, C. L., Ghosh, S., Mirams, G. R., et al. (2019). Probabilistic inference on noisy time series (PINTS). *J. Open Res. Softw.* 7, 23. doi:10.5334/jors.252

Darwich, A., Ogungbenro, K., Vinks, A. A., Powell, J. R., Reny, J.-L., Marsousi, N., et al. (2017). Why has model-informed precision dosing not yet become common clinical reality? lessons from the past and a roadmap for the future. *Clin. Pharmacol. Ther.* 101, 646–656. doi:10.1002/cpt.659

Darwich, A. S., Polasek, T. M., Aronson, J. K., Ogungbenro, K., Wright, D. F., Achour, B., et al. (2021). Model-informed precision dosing: background, requirements, validation, implementation, and forward trajectory of individualizing drug therapy. *Annu. Rev. Pharmacol. Toxicol.* 61, 225–245. doi:10.1146/annurev-pharmtox-033020-113257

Dirks, N. L., and Meibohm, B. (2010). Population pharmacokinetics of therapeutic monoclonal antibodies. *Clin. Pharmacokinet.* 49, 633–659. doi:10.2165/11535960-000000000-00000

FDA (2010). *Coumadin® tablets (warfarin sodium tablets, usp) crystalline coumadin® for injection (warfarin sodium for injection, usp)*. https://www.accessdata.fda.gov/drugsatfda_docs/label/2010/009218s108lbl.pdf. Accessed: 2022-October-03

Gage, B., Eby, C., Johnson, J., Deych, E., Rieder, M., Ridker, P., et al. (2008). Use of pharmacogenetic and clinical factors to predict the therapeutic dose of warfarin. *Clin. Pharmacol. Ther.* 84, 326–331. doi:10.1038/clpt.2008.10

Gaon, M., and Brafman, R. (2020). Reinforcement learning with non-markovian rewards. *Proc. AAAI Conf. Artif. Intell.* 34, 3980–3987. doi:10.1609/aaai.v34i04.5814

Gill, K. L., Machavaram, K. K., Rose, R. H., and Chetty, M. (2016). Potential sources of inter-subject variability in monoclonal antibody pharmacokinetics. *Clin. Pharmacokinet.* 55, 789–805. doi:10.1007/s40262-015-0361-4

Gong, I. Y., Tirona, R. G., Schwarz, U. I., Crown, N., Dresser, G. K., LaRue, S., et al. (2011). Prospective evaluation of a pharmacogenetics-guided warfarin loading and maintenance dose regimen for initiation of therapy. *Blood, J. Am. Soc. Hematol.* 118, 3163–3171. doi:10.1182/blood-2011-03-345173

Hamberg, A.-K., Dahl, M.-L., Barban, M., Scordo, M. G., Wadelius, M., Pengo, V., et al. (2007). 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. doi:10.1038/sj.clpt.6100084

Hamberg, A.-K., Hellman, J., Dahlberg, J., Jonsson, E. N., and Wadelius, M. (2015). A bayesian decision support tool for efficient dose individualization of warfarin in adults and children. *BMC Med. Inf. Decis. Mak.* 15, 7–9. doi:10.1186/s12911-014-0128-0

Hamberg, A.-K., Wadelius, M., Lindh, J., Dahl, M.-L., Padrini, R., Deloukas, P., et al. (2010). 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. doi:10.1038/clpt.2010.37

Hansen, N., Müller, S. D., and Koumoutsakos, P. (2003). Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (cma-es). *Evol. Comput.* 11, 1–18. doi:10.1162/106365603321828970

Hartmann, S., Biliouris, K., Lesko, L. J., Nowak-Göttl, U., and Trame, M. N. (2020). Quantitative systems pharmacology model-based predictions of clinical endpoints to optimize warfarin and rivaroxaban anti-thrombosis therapy. *Front. Pharmacol.* 11, 1041. doi:10.3389/fphar.2020.01041

Hartmann, S., Biliouris, K., Lesko, L., Nowak-Göttl, U., and Trame, M. (2016). Quantitative systems pharmacology model to predict the effects of commonly used anticoagulants on the human coagulation network. *CPT pharmacometrics Syst. Pharmacol.* 5, 554–564. doi:10.1002/psp4.12111

Hoffman, M. D., Gelman, A., et al. (2014). The no-u-turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* 15, 1593–1623.

Holford, N. H., Kimko, H. C., Monteleone, J. P., and Peck, C. C. (2000). Simulation of clinical trials. *Annu. Rev. Pharmacol. Toxicol.* 40, 209–234. doi:10.1146/annurev.pharmtox.40.1.209

Holford, N., Ma, S., and Ploeger, B. (2010). Clinical trial simulation: a review. *Clin. Pharmacol. Ther.* 88, 166–182. doi:10.1038/clpt.2010.114

Johnson, J. A., Caudle, K. E., Gong, L., Whirl-Carrillo, M., Stein, C. M., Scott, S. A., et al. (2017). Clinical pharmacogenetics implementation consortium (cpic) guideline for pharmacogenetics-guided warfarin dosing: 2017 update. *Clin. Pharmacol. Ther.* 102, 397–404. doi:10.1002/cpt.668

Johnson, J. A., Gong, L., Whirl-Carrillo, M., Gage, B. F., Scott, S. A., Stein, C., et al. (2011). Clinical pharmacogenetics implementation consortium guidelines for cyp2c9 and vkorc1 genotypes and warfarin dosing. *Clin. Pharmacol. Ther.* 90, 625–629. doi:10.1038/clpt.2011.185

Karlsson, M., and Sheiner, L. (1993). The importance of modeling interoccasion variability in population pharmacokinetic analyses. *J. Pharmacokinet. Biopharm.* 21, 735–750. doi:10.1007/BF01113502

Keizer, R. J., Ter Heine, R., Frymoyer, A., Lesko, L. J., Mangat, R., and Goswami, S. (2018). Model-informed precision dosing at the bedside: scientific challenges and opportunities. *CPT pharmacometrics Syst. Pharmacol.* 7, 785–787. doi:10.1002/psp4.12353

Keutzer, L., and Simonsson, U. S. (2020). Individualized dosing with high inter-occasion variability is correctly handled with model-informed precision dosing—using rifampicin as an example. *Front. Pharmacol.* 11, 794. doi:10.3389/fphar.2020.00794

Kingma, D. P., and Ba, J. (2014). Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980*

Klein, T. E., Altman, R. B., Eriksson, N., Gage, B. F., Kimmel, S. E., et al. (2009). Estimation of the warfarin dose with clinical and pharmacogenetic data. *N. Engl. J. Med.* 360, 753–764. doi:10.1056/NEJMoa0809329

Lavielle, M. (2014). *Mixed effects models for the population approach: models, tasks, methods and tools*. 1 edn. Chapman and Hall/CRC).

Ma, Y., Zhao, X., Chen, X., Huang, X., Lin, Q., Lin, Y., et al. (2021). Therapeutic drug monitoring of docetaxel by pharmacokinetics and pharmacogenetics: a randomized clinical trial of auc-guided dosing in nonsmall cell lung cancer. *Clin. Transl. Med.* 11, e354. doi:10.1002/ctm2.354

Mager, D. E. (2006). Target-mediated drug disposition and dynamics. *Biochem. Pharmacol.* 72, 1–10. doi:10.1016/j.bcp.2005.12.041

Maier, C., Hartung, N., de Wiljes, J., Kloft, C., and Huisinga, W. (2020). Bayesian data assimilation to support informed decision making in individualized chemotherapy. *CPT pharmacometrics Syst. Pharmacol.* 9, 153–164. doi:10.1002/psp4.12492

Maier, C., Hartung, N., Kloft, C., Huisinga, W., and de Wiljes, J. (2021). Reinforcement learning and bayesian data assimilation for model-informed precision dosing in oncology. *CPT pharmacometrics Syst. Pharmacol.* 10, 241–254. doi:10.1002/psp4.12588

Matsumoto, K., Oda, K., Shoji, K., Hanai, Y., Takahashi, Y., Fujii, S., et al. (2022). Clinical practice guidelines for therapeutic drug monitoring of vancomycin in the framework of model-informed precision dosing: a consensus review by the Japanese society of chemotherapy and the Japanese society of therapeutic drug monitoring. *Pharmaceutics* 14, 489. doi:10.3390/pharmaceutics14030489

Merlé, Y., Aouimer, A., and Tod, M. (2004). Impact of model misspecification at design (and/or) estimation step in population pharmacokinetic studies. *J. Biopharm. Statistics* 14, 213–227. doi:10.1081/BIP-120028516

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., et al. (2013). *Playing atari with deep reinforcement learning. arXiv preprint arXiv:1312.5602*.

Moore, B. L., Sinzinger, E. D., Quasny, T. M., and Pyeatt, L. D. (2004). Intelligent control of closed-loop sedation in simulated icu patients. *Flairs Conf.* 109–114.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). Pytorch: an imperative style, high-performance deep learning library. *Adv. neural Inf. Process. Syst.* 32.

Polasek, T. M., Rayner, C. R., Peck, R. W., Rowland, A., Kimko, H., and Rostami-Hodjegan, A. (2019). Toward dynamic prescribing information: codevelopment of companion model-informed precision dosing tools in drug development. *Clin. Pharmacol. drug Dev.* 8, 418–425. doi:10.1002/cpdd.638

Polasek, T. M., and Rostami-Hodjegan, A. (2020). Virtual twins: understanding the data required for model-informed precision dosing. *Clin. Pharmacol. Ther.* 107, 742–745. doi:10.1002/cpt.1778

Ribba, B., Bräm, D. S., Baverel, P. G., and Peck, R. W. (2022). Model enhanced reinforcement learning to enable precision dosing: a theoretical case study with dosing of propofol. *CPT Pharmacometrics Syst. Pharmacol.* 11, 1497–1510. doi:10.1002/psp4.12858

Ribba, B., Dudal, S., Lavé, T., and Peck, R. W. (2020). Model-informed artificial intelligence: reinforcement learning for precision dosing. *Clin. Pharmacol. Ther.* 107, 853–857. doi:10.1002/cpt.1777

Sheiner, L. B., Beal, S., Rosenberg, B., and Marathe, V. V. (1979). Forecasting individual pharmacokinetics. *Clin. Pharmacol. Ther.* 26, 294–305. doi:10.1002/cpt1979263294

Sheiner, L. B. (1969). Computer-aided long-term anticoagulation therapy. *Comput. Biomed. Res.* 2, 507–518. doi:10.1016/0010-4809(69)90030-5

Sutton, R. S., and Barto, A. G. (2018). *Reinforcement learning: an introduction*. MIT press.

Uster, D. W., Stocker, S. L., Carland, J. E., Brett, J., Marriott, D. J., Day, R. O., et al. (2021). A model averaging/selection approach improves the predictive performance of model-informed precision dosing: vancomycin as a case study. *Clin. Pharmacol. Ther.* 109, 175–183. doi:10.1002/cpt.2065

Van Hasselt, H., Guez, A., and Silver, D. (2016). Deep reinforcement learning with double q-learning. *Proc. AAAI Conf. Artif. Intell.* 30 (1). doi:10.1609/aaai.v30i1.10295

Verhoef, T. I., Ragia, G., de Boer, A., Barallon, R., Kolovou, G., Kolovou, V., et al. (2013). A randomized trial of genotype-guided dosing of acenocoumarol and phenprocoumon. *N. Engl. J. Med.* 369, 2304–2312. doi:10.1056/NEJMoa1311388

Véronneau-Veilleux, F., Ursino, M., Robaey, P., Lévesque, D., and Nekka, F. (2020). Nonlinear pharmacodynamics of levodopa through Parkinson's disease progression. *Chaos Interdiscip. J. Nonlinear Sci.* 30, 093146. doi:10.1063/5.0014800

Wadelius, M., Chen, L. Y., Lindh, J. D., Eriksson, N., Ghori, M. J., Bumpstead, S., et al. (2009). The largest prospective warfarin-treated cohort supports genetic forecasting. *Blood, J. Am. Soc. Hematol.* 113, 784–792. doi:10.1182/blood-2008-04-149070

Wadelius, M., and Pirmohamed, M. (2007). Pharmacogenetics of warfarin: current status and future challenges. *pharmacogenomics J.* 7, 99–111. doi:10.1038/sj.tpj.6500417

Wajima, T., Isbister, G., and Duffull, S. (2009). A comprehensive model for the humoral coagulation network in humans. *Clin. Pharmacol. Ther.* 86, 290–298. doi:10.1038/clpt.2009.87

Wang, Y., Zhu, H., Madabushi, R., Liu, Q., Huang, S.-M., and Zineh, I. (2019). Model-informed drug development: current us regulatory practice and future considerations. *Clin. Pharmacol. Ther.* 105, 899–911. doi:10.1002/cpt.1363

Wicha, S. G., Märtson, A.-G., Nielsen, E. I., Koch, B. C., Friberg, L. E., Alffenaar, J.-W., et al. (2021). From therapeutic drug monitoring to model-informed precision dosing for antibiotics. *Clin. Pharmacol. Ther.* 109, 928–941. doi:10.1002/cpt.2202

Xue, L., Holford, N., and Miao, L. (2016). *Warfarin pkpd: theory, body composition and genotype*. Lisbon: Annual Meeting of the Population Approach Group in Europe.

Zadeh, S. A., Street, W. N., and Thomas, B. W. (2023). Optimizing warfarin dosing using deep reinforcement learning. *J. Biomed. Inf.* 137, 104267. doi:10.1016/j.jbi.2022.104267

Zhu, T., Li, K., Kuang, L., Herrero, P., and Georgiou, P. (2020). An insulin bolus advisor for type 1 diabetes using deep reinforcement learning. *Sensors* 20, 5058. doi:10.3390/s20185058