

**ARTICLE**

# A hybrid machine learning/pharmacokinetic approach outperforms maximum a posteriori Bayesian estimation by selectively flattening model priors

Jasmine H. Hughes | Ron J. Keizer

InsightRX, San Francisco, California, USA

**Correspondence**Jasmine H. Hughes, 548 Market St.  
#88083, San Francisco, CA 94104, USA.  
Email: jasmine@insight-rx.com**Funding information**

No funding was received for this work.

**Abstract**

Model-informed precision dosing (MIPD) approaches typically apply maximum a posteriori (MAP) Bayesian estimation to determine individual pharmacokinetic (PK) parameters with the goal of optimizing future dosing regimens. This process combines knowledge about the individual, in the form of drug levels or pharmacodynamic biomarkers, with prior knowledge of the drug PK in the general population. Use of “flattened priors” (FPs), in which the weight of the model priors is reduced relative to observations about the patient, has been previously proposed to estimate individual PK parameters in instances where the patient is poorly described by the PK model. However, little is known about the predictive performance of FPs and when to apply FPs in MIPD. Here, FP is evaluated in a data set of 4679 adult patients treated with vancomycin. Depending on the PK model, prediction error could be reduced by applying FPs in 42–55% of PK parameter estimations. Machine learning (ML) models could identify instances where FPs would outperform MAPs with a specificity of 81–86%, reducing overall root mean squared error (RMSE) of PK model predictions by 12–22% (0.5–1.2 mg/L) relative to MAP alone. The factors most indicative of the use of FPs were past prediction residuals and bias in past PK predictions. A more clinically practical minimal model was developed using only these two features, reducing RMSE by 5–18% (0.20–0.93 mg/L) relative to MAP. This hybrid ML/PK approach advances the precision dosing toolkit by leveraging the power of ML while maintaining the mechanistic insight and interpretability of PK models.

**Study Highlights****WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?**

Model-informed precision dosing (MIPD) can improve attainment of pharmacokinetic (PK) targets and patient outcomes. However, it is unclear how to apply MIPD to patients that are not well-described by PK models.

**WHAT QUESTION DID THIS STUDY ADDRESS?**

This study evaluated the use of flattened priors during Bayesian estimation of individual PK parameters.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 Insight Rx, Inc. *CPT: Pharmacometrics & Systems Pharmacology* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics

### WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

Flattened priors (FPs) improve PK prediction accuracy for some patients. Machine learning (ML) models can identify clinical decision points benefitting from FPs.

### HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?

This work demonstrates how hybrid ML and Bayesian PK approaches can augment clinical decision making at the point of care while maintaining clinician autonomy and PK explainability.

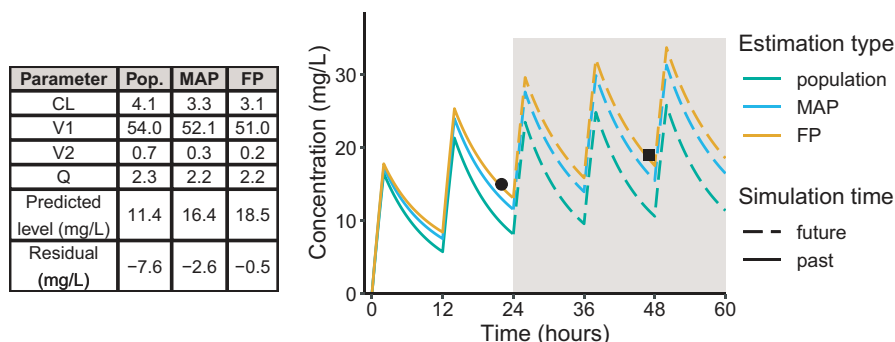
## INTRODUCTION

Model-informed precision dosing (MIPD) has shown increased adoption at the point of care over recent years, aiding clinicians in tailoring prescriptions to their patients for a variety of drugs, including antibiotics,<sup>1-4</sup> bone marrow transplant conditioning regimens,<sup>5,6</sup> monoclonal antibodies,<sup>7,8</sup> and chemotherapeutics.<sup>9</sup> Most MIPD systems in use apply some form of Bayesian estimation of the individual patient's pharmacokinetic (PK) parameters based on a population PK (PopPK) model.<sup>10</sup> In Bayesian estimation, knowledge about the distribution of parameters for the general population (i.e., the model priors), is combined with observations about the patient, typically one or more drug serum levels and/or one or more pharmacodynamic measurements. The model prior acts as an anchor, preventing PK parameter estimates from deviating too far from expected values. Often, this smoothing effect is desirable, making predictions more robust in the face of noisy clinical measurements until sufficient evidence is available to justify more unusual PK parameter estimates. In some instances, however, a patient may not be well-described by a model prior, for example, due to fluid overload or unusual levels of muscle mass that are not captured in the structural model. In these instances, it may be desirable to reduce

the weight of the model priors during PK parameter estimation, or in other words, “flatten” the distribution of model priors for the particular patient.

A hypothetical example of this approach is shown in Figure 1. Here, a level of 15 mg/L is collected at 22 h after the first dose. Using maximum a posteriori (MAP) Bayesian model prior weighting, the concentration-time curve (blue line) is simulated to lie somewhere between the measured value (solid circle) and the population prediction (green line). By using a flattened priors (FPs) approach, the PK parameter estimates are allowed to drift further from the population estimates, and the concentration-time curve (orange line) passes closer to the measured value. In this example scenario, a second level collected at 47 h is better predicted by the FP approach compared to MAP.

Although this approach may improve the predictive ability of models used in MIPD for some subset of patients that deviate substantially from the population average,<sup>11</sup> to our knowledge, there has not yet been a large-scale evaluation of the performance of FPs, nor an assessment of when this approach could prove beneficial. Here, we assess the predictive performance of FPs on a large data set of adult patients treated with vancomycin for three different PK models. Second, we trained machine learning (ML) models to identify in which



**FIGURE 1** Schematic illustrating a patient in which using flattened priors (FPs) would have improved prediction precision of the second drug level relative to maximum a posteriori (MAP) Bayesian estimation. In this hypothetical scenario, the first level collected (the solid circle at 22 h) was 15 mg/L, and was used to inform MAP Bayesian estimation of the individual's pharmacokinetic (PK) parameters (shown in the overlaid table) using either FP or MAP. These PK parameters were then used to simulate concentration-time curves, shown in the graph, and to predict the drug level at 47 h (solid square; 19 mg/L)

scenarios the FP approach would lead to more accurate descriptions of patients' future drug concentrations. Finally, we benchmarked the hybrid ML/PK approach against a previously described continuous learning approach, in which PopPK models are retrained on collected PK data.

## METHODS

### Data source

De-identified data collected during routine clinical care of adult patients treated with vancomycin and entered into the MIPD clinical decision support software InsightRX Nova between January 1, 2018, and October 30, 2020, were used as a data source and analyzed retrospectively. Records were included in the analysis if they described patients greater than or equal to 18 years old from whom at least two vancomycin serum levels were collected within a single vancomycin treatment course (defined here as vancomycin administered intermittently with no more than a 14-day gap in between doses). Records were excluded if data recording issues seemed likely (for example, biologically implausible heights or weights, or doses of vancomycin greater than 6 g, or vancomycin serum trough levels greater than 50 mg/L), or if the data could not be unambiguously interpreted (for example, multiple simultaneous doses).

### Pharmacokinetic modeling

PK models from literature were implemented in the InsightRX software. For each model and for each vancomycin treatment course, vancomycin serum levels were iteratively predicted. Using the first  $n$  levels, individual PK parameters were estimated using MAP Bayesian estimation, minimizing the log likelihood of the posterior distribution given the distribution of the random effect parameters (i.e., essentially optimizing  $\eta$  to minimize):

$$-2 \cdot \left( \sum_{i=1}^n \frac{(y_i - \hat{y}_i)^2}{\sigma_i^2} + \sum_{j=1}^m \frac{\eta_j}{\omega_j^2} \right)$$

where  $\eta$  is the vector of  $m$  random effects estimated for that individual,  $y_i$  is a measured vancomycin serum concentration,  $\hat{y}_i$  is the corresponding predicted concentration given the estimated values of  $\eta$  and the PK model,  $\sigma^2$  is the variance of the residual error for the observation, and  $\omega_j^2$  is the variance of the interindividual variability for the  $j$ th estimated PK parameter. Concentration-time curves were simulated using these parameter estimates to predict the  $(n + 1)$ th concentration. Time-varying patient covariates (weight and serum creatinine)

were censored to include only data available at the time of the  $n$ th level to avoid data leakage. PK parameter estimations were performed by multiplying the  $\sum_{j=1}^m \frac{\eta_j}{\omega_j^2}$  term in the equation above by the following values: 1.0, 0.6, 0.3, 0.125, or 0.02. The value of 1.0 produces conventional MAP estimates, the value of 0.02 produces an extremely low level of model prior weighting, and the middle three values correspond approximately to levels of FP available in the InsightRX Nova software. Maximum likelihood estimation was performed using the R package `bbmle`<sup>12</sup> and concentration-time curves were simulated using `PKPDsim`.<sup>13</sup>

The predictive performances of the two estimation methods were assessed by the root mean squared error (RMSE) and mean percent error (MPE) of these iterative model predictions relative to observed values, defined as:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2}$$

$$\text{MPE} = \frac{1}{N} \sum_{i=1}^N \frac{\hat{y}_i - y_i}{y_i}$$

where  $y_i$  is a measured vancomycin serum concentration,  $\hat{y}_i$  is the corresponding predicted vancomycin serum concentration for that individual, and  $N$  is the total number of measured drug levels across all patients. Variability in RMSE and MPE was assessed across 1000 bootstraps.

### Machine learning model development

The ML task was defined as predicting at the time of collection of each drug level whether FP or MAP would produce a more accurate prediction of the subsequent drug level. We term these time points "decision points" because they approximately align with dose adjustment decision timing at the MIPD point of care. We defined FP as producing a more accurate prediction if the absolute value of the FP prediction residual (calculated as  $(y_i - \hat{y}_i)$ ) was at least 2.5 mg/L OR 15% lower than the absolute value of the MAP prediction residual, as this was considered a clinically relevant difference.

The data set was split into three folds: a test data set (23% of patients) used only for final evaluation of model performance, a training data set (58% of patients) used for model development, and a cross-validation data set (19% of patients) used for evaluation of model performance during model development. Twenty-nine derived features hypothesized to be predictive of when to use FPs were calculated for each decision point, using only information available at the time of collection of the drug level. These features, detailed

in Table S1, included information related to: patient characteristics, such as age and body mass index; the collected level, such as the time elapsed between levels and the value of the most recent level; historical predictive performance of the PK models, such as the RMSE and MPE of past predictions; and how informative drug levels were toward MAP Bayesian estimation, such as the Mahalanobis distance and the reduction in uncertainty in PK parameters given the most recently collected level. Data were handled to avoid “data leakage” of future observations (such as future serum creatinine laboratory results) into the data available at the time of a decision point. Correlation between the features was confirmed to be under 90% for all features. Features were scaled to be approximately normally distributed and range from  $-1$  to  $1$  (continuous variables) or have a value of  $0$  or  $1$  (categorical variables) to allow for comparison of feature weighting in models.

Three types of ML models were developed in R<sup>14</sup>: logistic regression (LR; using R’s `stats::glm`), penalized logistic regression (PLR; `glmnet` package<sup>15</sup>), and XGBoost (XGB; `xgboost` package<sup>16</sup>). The LR and PLR models were fit using first-order interaction terms, and the PLR hyperparameter lambda was determined using 15-fold cross-validation. The XGB model was trained using fivefold repeated cross validation with five repeats and grid search for hyperparameter tuning. Cross validation during model training was performed using the `glmnet` package and `caret` (Kuhn, 2020). A probability threshold of 50% was selected for recommendation of FP.

Additionally, several “minimal models” were developed using only two features: (1) cumulative bias in MAP residuals and (2) the value of the last MAP residual (i.e., predicted value – observed value). A PLR model was then fit using these parameters as described previously.

The performance of these ML models was assessed using the metrics accuracy, recall (sensitivity) and precision (specificity). These metrics were calculated as follows:

$$\text{Accuracy} = \frac{\text{FP}_{\text{TRUE}} + \text{MAP}_{\text{TRUE}}}{\text{FP}_{\text{TRUE}} + \text{MAP}_{\text{TRUE}} + \text{FP}_{\text{FALSE}} + \text{MAP}_{\text{FALSE}}}$$

$$\text{Recall} = \frac{\text{FP}_{\text{TRUE}}}{\text{FP}_{\text{TRUE}} + \text{MAP}_{\text{FALSE}}}$$

$$\text{Precision} = \frac{\text{FP}_{\text{TRUE}}}{\text{FP}_{\text{TRUE}} + \text{FP}_{\text{FALSE}}}$$

where  $\text{FP}_{\text{TRUE}}$  is the number of decision points correctly classified as benefiting from FP using the threshold described above,  $\text{FP}_{\text{FALSE}}$  is the number of decision points incorrectly classified as benefiting from FP when MAP would have performed better,  $\text{MAP}_{\text{TRUE}}$  is the number of decision points correctly classified as MAP and  $\text{MAP}_{\text{FALSE}}$  is the number of decision points

identified as MAP when FP would have been more predictive. Because these terms are defined requiring an improvement of at least 2.5 mg/L or 15% for FP to be selected, we noted that many of the  $\text{FP}_{\text{FALSE}}$  instances were examples where FP indeed outperformed MAP but by a threshold of less than 2.5 mg/L and by less than 15%. We therefore also define an additional metric termed effective precision, calculated as:

$$\text{Effective Precision} = \frac{\text{FP}_{\text{Best}}}{\text{FP}_{\text{Best}} + \text{FP}_{\text{Worse}}}$$

where  $\text{FP}_{\text{Best}}$  refers to decision points where FP was recommended and produced a smaller residual than MAP, and  $\text{FP}_{\text{Worse}}$  where FP was recommended but produced a larger residual than MAP. During model development, we prioritized effective precision over recall, or, in other words, we prioritized the proportion of recommendations to use FPs that were correct at the expense of missing some scenarios where FPs would have been more appropriate than MAP. We viewed this as the more conservative mode of failure, because we considered the standard of care for MIPD in clinical practice to be MAP.

The performances of the ML models were also assessed by the RMSE and MPE of PK model predictions if the recommendations of the ML models were followed at each decision point. RMSE and MPE were calculated as described above.

## Benchmark to continuous learning

Continuous learning (CL), in which the model priors are iteratively updated to refine an existing PopPK model repeatedly, has been previously proposed by us as a method to improve PK predictive performance for precision dosing applications.<sup>11,17</sup> To benchmark the performance of the hybrid ML/PK strategy described here to CL, we re-estimated the model parameters of the three PK models on the training dataset. The PK models were implemented in NONMEM. Model coefficients, interindividual variability, and residual error were re-estimated using NONMEM version 7.4.3 (ICON plc, Ellicott City, MD) for each model using the training data set, without altering covariate structure or model structure. Predictive performance of these re-estimated (RE) models was assessed by the RMSE and MPE of iteratively predicted drug levels using MAP.

## RESULTS

### Patients and data collection

There were 4679 patients in the de-identified data set that met the inclusion criteria. Patient records were randomly

split into a training set, a cross-validation set, and a test set for final evaluation of the models. Patient characteristics for each of these data sets are summarized in Table 1. Of note, only 43% (2007/4679) of patients had more than two vancomycin serum levels collected over the course of treatment. As a result, half (4610/9052) of the predictions are based on PK parameters estimated from a single drug level.

## Predictive performance of flattened priors

Three previously published PK models were selected for evaluation. The Thomson model<sup>18</sup> was selected as an example of a PopPK model that performs well in a general adult patient population.<sup>19,20</sup> The Goti model<sup>21</sup> was selected as an example of a PopPK model that generally performs well<sup>19</sup> but produces known biases in its predictions.<sup>22</sup> The Buelga model<sup>23</sup> was selected as an example of a model with less predictive performance relative to other more recently published models.<sup>19,20</sup> Furthermore, this model has a known misspecification; it is a one-compartment model, whereas the general consensus in the literature is that vancomycin follows two-compartment kinetics.

Prediction error (RMSE) was lowest for each PK model when using a model prior weight of 0.3–0.6 (Figure 2a). However, in individual instances, lower or higher amounts of prior weighting often improved prediction precision (Figure 2b). Extremely low model prior weights (e.g., 1/50th of that of MAP) produced markedly higher RMSE relative to MAP: whereas in 26–35% of individual decision points, these predictions were more precise, occasional highly erroneous predictions inflated the imprecision in aggregate. For subsequent analyses, we set the model prior weight for FP to one-eighth, balancing a substantially different and often improved prediction relative to the MAP estimate while not perilously ignoring the model prior.

Prediction error (RMSE; Figure 2c) and bias (MPE; Figure 2d) were higher for the Buelga model relative to the

Goti model and the Thomson model, consistent with findings reported elsewhere.<sup>19,20</sup> For the Buelga model, prediction accuracy and bias could be reduced by always using FP instead of MAP: in the majority of decisions (55%), FP would produce a more accurate prediction (Figure 2e). For the Goti model and the Thomson model, prediction accuracy was similar overall for both MAP and FP. In 46% and 42% of decision points for the Goti model and the Thomson model, respectively, prediction error could have been reduced by using FPs. If the “best” estimation method had been selected at each decision point, RMSE would have been reduced by 20% for the Goti model (from 4.5 to 3.6 mg/L) and 16% for the Thomson model (from 4.3 to 3.6 mg/L). Prediction bias was lowest when naively applying FP. This finding is expected given that bias related to the anchoring effect of the model priors would be greatly reduced when using this method.

## Applying machine learning to select model prior weighting

Given the potential improvement in predictive capacity by accurately identifying when to reduce the weight of the model priors, we wondered if traditional ML methods could aid in decision making. The ML task was structured as a classification problem: for a given drug level, should FP or MAP be used to most accurately predict the subsequent drug level?

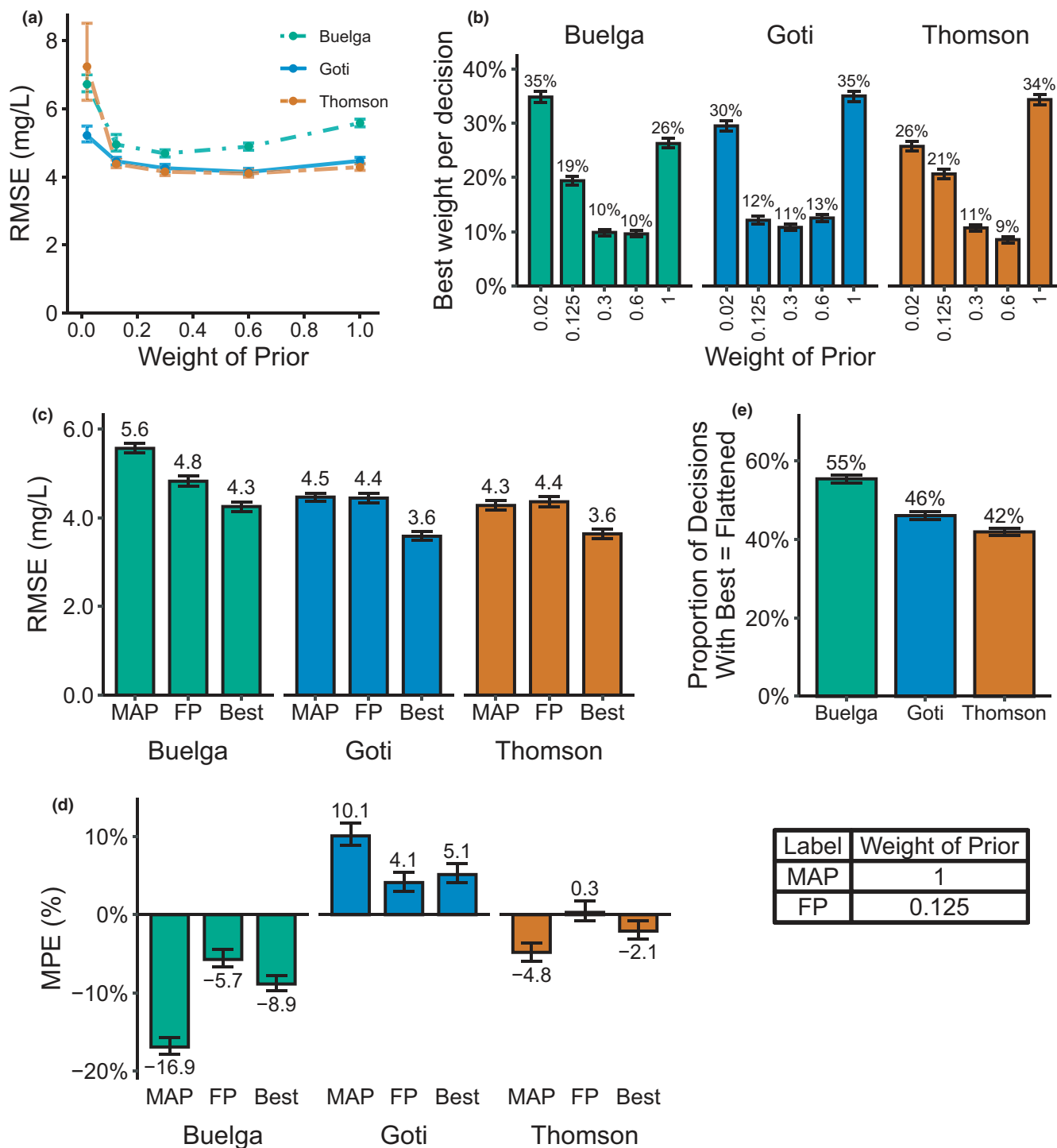
Of the three types of ML models assessed, the XGBoost models were the most predictive, with accuracies ranging from 75 to 77% (Table 2). ML model effective precision was considered to be of particular importance, because this metric reflects the proportion of recommendations that recommend a change from standard practice that were correct. For the XGBoost model trained for the Buelga model, effective precision was 86%, indicating that six out of every seven decisions where FP was recommended were indeed

**TABLE 1** Patient characteristics in training, cross-validation and testing data sets

	Unit	Training	Cross-validation	Testing
Patients	N	2700	888	1091
Drug levels	N	5153	1732	2167
Levels per patient	N	2 (2–6)	2 (2–6)	2 (2–6)
Percent female	%	40%	42%	40%
Age (years)	years	64.2 (31.2–87.5)	64.5 (30.8–88.6)	63.4 (33.1–87.6)
Weight (kg)	kg	84.2 (52–153.4)	85.6 (52.5–155.5)	86.4 (52.4–158.2)
BMI	kg/m <sup>2</sup>	28.1 (19–47)	28.7 (18.8–46.8)	28.6 (19–49.2)
Serum creatinine	mg/dl	0.9 (0.5–2.6)	0.9 (0.5–2.3)	0.9 (0.5–2.3)

Abbreviation: BMI, body mass index.

Values indicate median (5th–95th percentile) where appropriate.



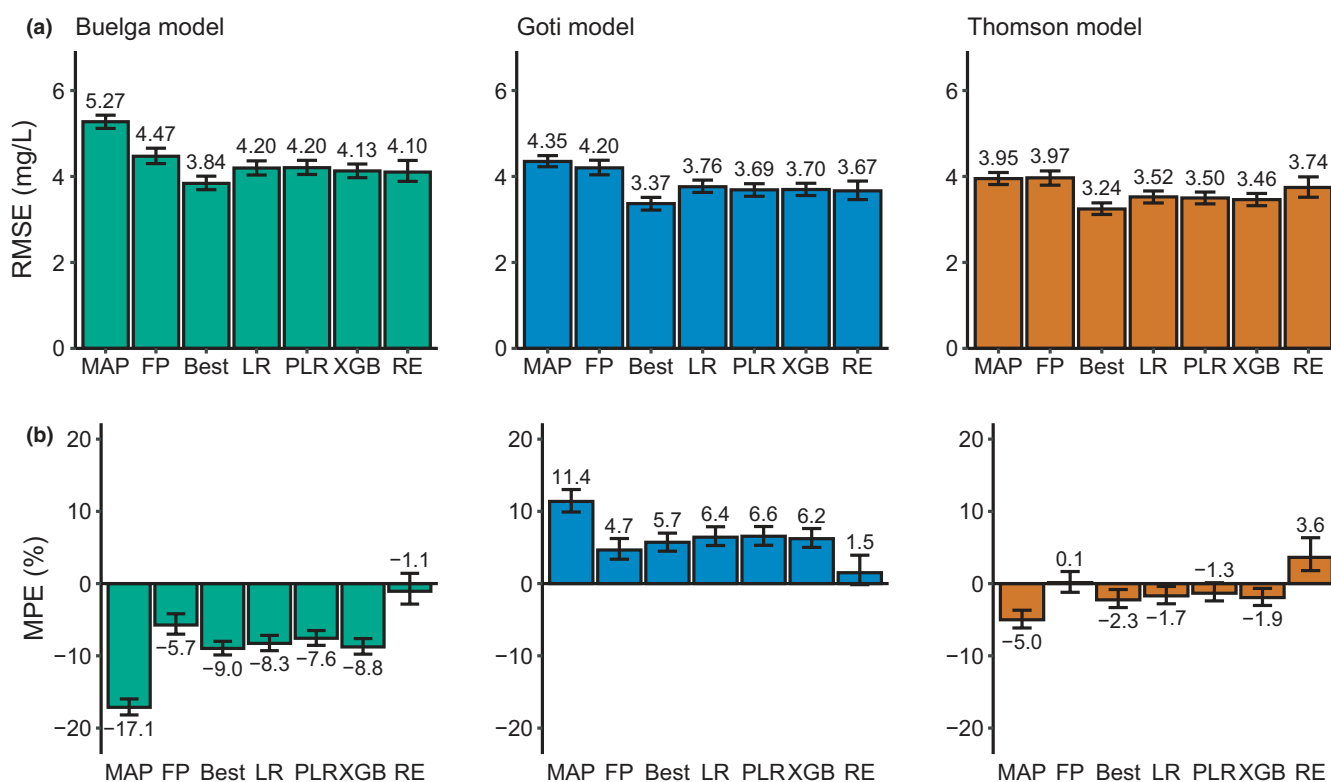
**FIGURE 2** Effect of model prior weight on prediction precision. (a) Root mean squared error (RMSE) for each of the three pharmacokinetic (PK) models, using different model prior weights. (b) Proportion of clinician decisions for which the most precise prediction using a given model was made using that particular model prior weight. (c) RMSE using maximum a posteriori (MAP) estimation, flattened priors with a prior weight of one-eighth (FP) or the best of MAP and FP for each of the three models. (d) Mean percent error (MPE) using MAP, FP, or the best of MAP and FP for each of the three models. (e) Proportion of decision points in which FPs would have resulted in a more accurate precise prediction than MAP. Bar height indicates median RMSE, MPE, or proportion for 1000 bootstraps, error bars indicate the 2.5th–97.5th percentiles of these quantities

best served by using FP. ML model performance was similar for the Goti model and the Buelga model. The performance of the ML models trained for the Thomson model were lower, perhaps reflecting this PK model's generally good predictive

performance. If the recommendations of these ML models had been followed, the RMSE of PK predictions would have been reduced by 11–22% relative to MAP (Figure 3a) and the MPE would have been reduced by 42–74% relative to MAP

Model	Accuracy	Recall	Precision	Effective precision
<b>Buelga</b>				
Logistic regression	70.6	75.2	73.6	80.1
Penalized logistic regression	71.3	75.7	74.2	81.0
XGBoost	77.0	79.8	79.7	86.3
<b>Goti</b>				
Logistic regression	70.8	62.6	72.6	77.7
Penalized logistic regression	72.1	61.5	75.7	81.7
XGBoost	75.3	70.0	76.6	82.6
<b>Thomson</b>				
Logistic regression	70.1	57.8	66.8	74.3
Penalized logistic regression	69.3	53.3	67.1	74.9
XGBoost	74.8	65.9	72.0	81.2

**TABLE 2** Performance of the machine learning models for each of the three pharmacokinetic models, evaluated on the test data set



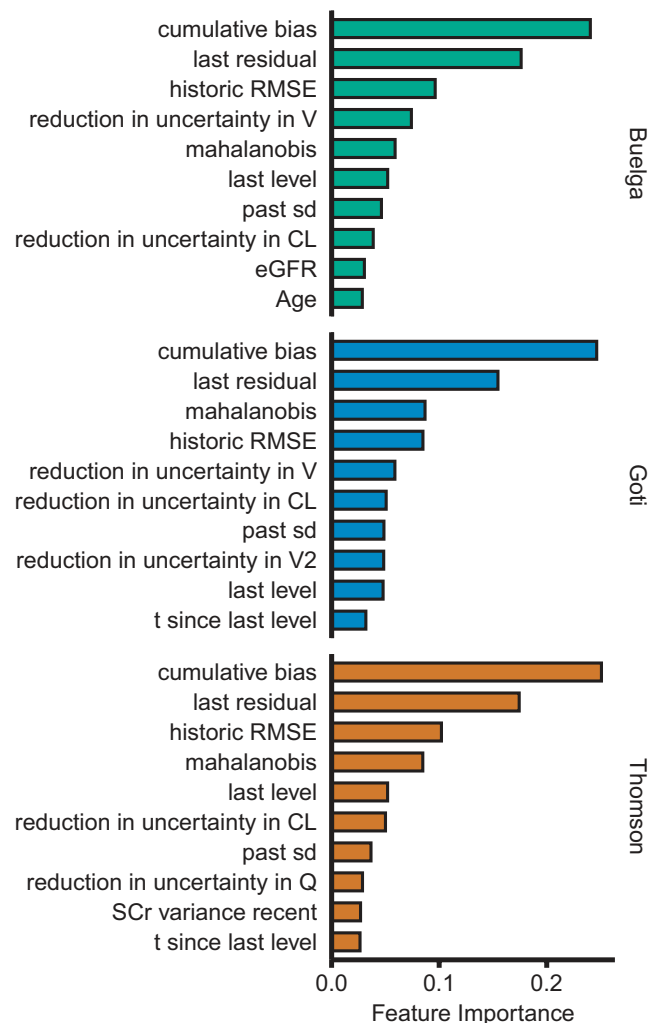
**FIGURE 3** Reduction in (a) root mean squared error (RMSE) and (b) mean percent error (MPE) for the three machine learning (ML) models (logistic regression [LR]; penalized logistic regression [PLR]; and extreme gradient boosting [XGB]), and the re-estimated (RE) PK models compared to maximum a posteriori (MAP) Bayesian estimation, to naively using flattened priors (FP), or to using the best of MAP and FP. Bar height indicates median RMSE or MPE for 1000 bootstraps, error bars indicate the 2.5th–97.5th percentiles of these quantities

(Figure 3b). The presence of highly erroneous predictions, as assessed by outliers on a measured value–predicted value plot, was similar between MAP and the ML/PK hybrid approach (Figure S1).

The most predictive features showed good agreement between the PK models. For all three XGBoost models, the two most predictive features were cumulative bias in residuals and the value of the last residual (Figure 4). These features

were also heavily weighted in the logistic regression and penalized logistic regression models (Figure S2).

These reductions in prediction error were benchmarked against RE versions of the PK models, in which the model priors were updated to reflect the data in the training data set. All three models minimized successfully, and the precision of the parameter estimates was good (Table S2). For all three PK models, the RE models were more predictive



**FIGURE 4** The 10 most important features for the XGBoost models trained to predict the need for flattened prior for the Buelga, Goti, and Thomson pharmacokinetic models. Feature importance is the fractional contribution of that feature to the model. CL, clearance learning; eGFR, estimated glomerular filtration rate; RMSE, root mean squared error; SCr, serum creatinine

than the published models and approximately matched the performance of the ML/PK hybrid approaches (Figure 3). For the Buelga and Goti models, re-estimation reduced precision bias considerably, consistent with these models having known prediction bias or misspecification.

### A minimal model for guiding clinical decision making

Because two features were consistently identified as important for ML model performance, we wondered if a minimal model, using just these two features, could achieve similar performances. The minimal model was trained on data from all three PK models using penalized logistic regression. Expressed as a formula for calculating the

probability that FP should be used ( $P(\text{flattened})$ ) this model translates to:

$$y = 0.0289X_1 - 0.812X_2 + 0.210X_1X_2 - 0.539$$

$$P(\text{flattened}) = \frac{\exp(y)}{1 + \exp(y)}$$

where  $X_1$  is the value of the most recent residual (i.e., the value predicted less the value measured), capped to a maximum of 25 mg/L and a minimum of  $-25$  mg/L, and  $X_2$  is calculated from the MPE of all prior residuals as follows:

$$X_2 = \begin{cases} -\sqrt{|MPE|} & \text{if } x < 0 \\ \sqrt{MPE} & \text{if } x \geq 0 \end{cases}$$

and capped to a minimum of  $-1$  and a maximum of 3. These data transformations were selected empirically to approximate a normal distribution with few extreme outliers in the training data set.

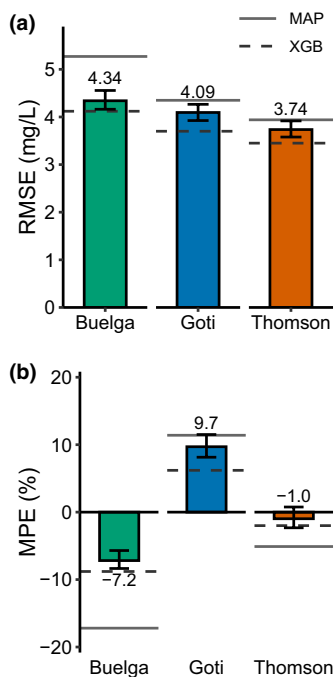
Overall, the predictive performance of this minimal model was surprisingly good despite its simplicity, with an accuracy of 65% and an effective precision of 76% (Table S3). Prediction error, as measured by RMSE, was higher than for the XGBoost models, but lower than that of MAP (Figure 5).

To evaluate the generalizability of these results across PK models, we next retrained the minimal model on data consisting of pairwise combinations of the PK models, evaluating these models on the third, unseen PK model. The accuracy (59–66%) and effective precision (71–81%) of these minimal ML models was similar to the minimal model trained on data consisting of all three PK models (Table S3), although recall was considerably lower (27–52%). RMSE was similar to that observed with the general minimal model (3.7–4.5 mg/L; Supplementary Figure S3).

## DISCUSSION

PopPK models are a powerful tool for informing precision dosing, however, not all patients are well-described by model priors. Here, we show that reducing the weight of the model prior during Bayesian estimation of individual PK parameters improves PK predictions for some patients. We combine this FP approach with traditional ML techniques to recommend when to downweigh model priors, identifying past prediction errors and consistent bias in prediction errors as PK model-independent factors to consider when deciding to use FP. This hybrid ML/PK approach is, to our knowledge, the first application of ML to improve PK modeling within an MIPD context.





**FIGURE 5** Performance of the minimal model for recommending the use of flattened priors, as measured by (a) root mean squared error (RMSE) and (b) mean percent error (MPE) relative to the median performance of maximum a posteriori (MAP) Bayesian estimation, and the full XGBoost predictive model (XGB) for each PK model. Bar height indicates median RMSE or MPE for 1000 bootstraps, error bars indicate the 2.5th–97.5th percentiles of these quantities

Applications of ML to MIPD to date have found that ML models are often able to accurately estimate past drug exposure,<sup>24,25</sup> predict future drug exposure,<sup>26–28</sup> or select doses.<sup>29–32</sup> However, the improvement in accuracy from these earlier approaches comes at the expense of pharmacological interpretability and the ability to simulate patient response to alternative dosing regimens.<sup>24,33,34</sup> An advantage of the combination of ML and PK models as described here is that clinical decision making is augmented by ML while maintaining the ability to forecast patient PKs and extract mechanistic insight from PK parameter estimates. Improving predictive performance for patients with unusual PKs by allowing more flexibility in individual PK parameters has been proposed before.<sup>35,36</sup> Our approach differs in that it allows reduction of the prior weight at each decision point rather than introducing flexibility over time.

Application of FP ameliorates one of the primary drawbacks of Bayesian dosing approaches: the requirement that the prior model should match the patient well to allow accurate predictions, especially in cases with limited sampling. Because FP downweights the prior, it relies less on the model development population, and more on the individual patient's data. However, the risk of this approach when applied in clinical practice is that the model may overfit the data. Responsible implementation of MIPD requires clinical

judgment to assess the possibility of errors, such as data entry issues, measurement error, medical considerations not captured by the PK model, or other possible explanations when a model returns an extreme or unexpected prediction. This clinical judgment is particularly important when applying FP, because the anchoring effect of the model prior is minimized. This study provides guidelines to consider when choosing whether to use MAP or FP.

This study was conducted on vancomycin PKs in adults, and it will be interesting to see how well it generalizes to other populations and drugs. PK model covariates, such as creatinine clearance and weight, were not heavily weighted by the ML models for recommending FP, likely because these covariates already inform PK model predictions. Instead, the ML models relied predominantly on metrics related to past predictive performance, and a minimal model trained only on two features related to past PK predictive performance nearly matched the performance of more complex ML models in terms of reducing PK prediction RMSE. The minimal models evaluated on unseen PK models also performed quite well. Together, these data suggest that past prediction residuals and bias in past predictions are important indicators of when to use FP, and that these results likely generalize to other PK models.

As an example of the application of the minimal model in practice, consider again the hypothetical patient shown in Figure 1. The first level was predicted to be 8.15 mg/L (using population estimates, because no drug levels were available yet), however, the first level was measured to be 15 mg/L. In the minimal model presented above, with  $X_1 = 8.15 - 15 = -6.85$  and  $X_2 = -\sqrt{6.85/15} = -0.676$ , the probability that FP should be used for estimating PK parameters is 69%. Because this probability is greater than 50%, the minimal model would recommend FP here. Upon verifying that the patient's dosing and laboratory result history was recorded accurately, the clinician could then use FP rather than MAP when adjusting the patient's dose. This minimal model could easily be incorporated within existing MIPD clinical decision support software, allowing for automated detection of the best estimation method. It will, however, be important to validate this approach prospectively before adoption of this heuristic into clinical practice.

All FP predictions made here used a factor of one-eighth to downweigh model priors, allowing the ML task to be posed as a simple binary classification problem. Future more sophisticated applications of the hybrid ML/PK approach could recommend a range of model prior weights, including more conservative or more aggressive flattening, or estimate the optimal model prior weighting.

Predictive performance was evaluated iteratively, using the first  $n$  levels to predict the  $(n + 1)$ th level. The improved predictive performance of the hybrid ML/PK approach likely reflects a more accurate estimation of individual PK parameters. As a result, past exposure estimations made

using these individual PK parameters would also likely be more accurate.

Finally, we benchmarked the performance of RE model priors against the performance of the ML models for predicting FP. In a previous study,<sup>17</sup> we found that this continuous learning approach improved the predictive performance of PK models, but that this improvement was particularly noticeable for population (or a priori) estimates, before drug levels have been collected. In contrast, in the present study, we evaluated only the performance of a posteriori predictions made using at least one drug level to inform parameter estimates, because FP is not applicable to a priori estimation. For the Goti and Buelga models, RE outperformed or matched the ML models in reducing prediction error, however, the RE Thomson was outperformed by the ML models. This finding is consistent with the Thomson model's already good performance in a general adult population and shows there is likely an additive benefit of joint application. A data-driven continuous learning MIPD system could therefore incorporate both approaches: RE to update a model to best reflect a particular population and ML to identify patients that are still poorly reflected by these model priors.

## ACKNOWLEDGEMENTS

The authors would like to thank Dominic M.H. Tong and Sirj Goswami for helpful discussions.

## CONFLICT OF INTEREST

J.H.H. and R.J.K. are employees and stockholders of InsightRX, a model informed precision dosing clinical decision support tool.

## AUTHOR CONTRIBUTIONS

J.H.H. and R.J.K wrote the manuscript. J.H.H. and R.J.K designed the research. J.H.H. performed the research. J.H.H. and R.J.K analyzed the data.

## ORCID

Jasmine H. Hughes  <https://orcid.org/0000-0003-3053-1251>

## REFERENCES

1. Stocker SL, Carland JE, Reuter SE, et al. Evaluation of a pilot vancomycin precision dosing advisory service on target exposure attainment using an interrupted time series analysis. *Clin Pharmacol Ther.* 2021;109(1):212.
2. Frymoyer A, Schwenk HT, Zorn Y, et al. Model-informed precision dosing of vancomycin in hospitalized children: implementation and adoption at an academic Children's Hospital. *Front Pharmacol.* 2020;11:551.
3. Abdulla A, Edwina AE, Flint RB, et al. Model-informed precision dosing of antibiotics in pediatric patients: a narrative review. *Front Pediatr.* 2021;9:624639.
4. Heine R, Keizer RJ, Steeg K, et al. Prospective validation of a model-informed precision dosing tool for vancomycin in intensive care patients. *Br J Clin Pharmacol.* 2020;86(12):2497-2506.
5. Shukla P, Goswami S, Keizer RJ, et al. Assessment of a model-informed precision dosing platform use in routine clinical care for personalized busulfan therapy in the pediatric hematopoietic cell transplantation (HCT) population. *Front Pharmacol.* 2020;11:888.
6. Sanghavi K, Wiseman A, Kirstein MN, et al. Personalized fludarabine dosing to reduce nonrelapse mortality in hematopoietic stem-cell transplant recipients receiving reduced intensity conditioning. *Transl Res.* 2016;175:103-115.e4.
7. Strik AS, Löwenberg M, Mould DR, et al. Efficacy of dashboard driven dosing of infliximab in inflammatory bowel disease patients; a randomized controlled trial. *Scand J Gastroenterol.* 2021;56:145-154.
8. Assa A, Matar M, Turner D, et al. Proactive monitoring of adalimumab trough concentration associated with increased clinical remission in children with Crohn's disease compared with reactive monitoring. *Gastroenterology.* 2019;157:985-996.e2.
9. Taylor ZL, Mizuno T, Punt NC, et al. MTXPK.org: a clinical decision support tool evaluating high-dose methotrexate pharmacokinetics to inform post-infusion care and use of glucarpidase. *Clin Pharmacol Ther.* 2020;108(3):635-643.
10. Darwich AS, Polasek TM, Aronson JK, et al. Annual review of pharmacology and toxicology: Model-informed precision dosing: Background, requirements, validation, implementation, and forward trajectory of individualizing drug therapy. *Annu Rev Pharmacol Toxicol.* 2021;61(1):225-245.
11. Keizer RJ, ter Heine R, Frymoyer A, et al. Model-informed precision dosing at the bedside: scientific challenges and opportunities. *CPT Pharmacometrics Syst Pharmacol.* 2018;7(12):785-787.
12. Bolker B; R Development Core Team. *bbmle: Tools for General Maximum Likelihood Estimation.* 2020.
13. Keizer RJ. PKPDSim: Simulate ODE models; 2021 <github.com/InsightRX/pkpsim>.
14. R Core Team. *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing; 2020.
15. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* 2010;33(1):1-22.
16. Chen T, et al. *xgboost: Extreme Gradient Boosting;* 2021.
17. Hughes JH, Tong DMH, Lucas SS, et al. Continuous learning in model-informed precision dosing: a case study in pediatric dosing of vancomycin. *Clin Pharmacol Ther.* 2021;109(1):233-242.
18. Thomson AH, Staatz CE, Tobin CM, Gall M, Lovering AM. Development and evaluation of vancomycin dosage guidelines designed to achieve new target concentrations. *J Antimicrob Chemother.* 2009;63:1050-1057.
19. Broecker A, Nardecchia M, Klinker KP, et al. Towards precision dosing of vancomycin: a systematic evaluation of pharmacometric models for Bayesian forecasting. *Clin Microbiol Infect.* 2019;25(10):1286.e1-1286.e7.
20. Shingde RV, Reuter SE, Graham GG, et al. Assessing the accuracy of two Bayesian forecasting programs in estimating vancomycin drug exposure. *J Antimicrob Chemother.* 2020;75(11):3293-3302.
21. Goti V, Chaturvedula A, Fossler MJ, Mok S, Jacob JT. Hospitalized patients with and without hemodialysis have markedly different vancomycin pharmacokinetics: a population pharmacokinetic model-based analysis. *Ther Drug Monit.* 2018;40(2):212-221.
22. Tong DMH, Hughes JH, Keizer RJ. Use of age-adjusted serum creatinine in a vancomycin pharmacokinetic model decreases predictive performance in elderly patients. *Ther Drug Monit.* 2021;43(1):139-140.

23. Buelga DS, del Mar Fernandez de Gatta M, Herrera EV, Dominguez-Gil A, Garcia MJ. Population pharmacokinetic analysis of vancomycin in patients with hematological malignancies. *Antimicrobial Agents Chemotherapy*. 2005;49(12):4934-4941.
24. Woillard JB, Labriffe M, Debord J, Marquet P. Tacrolimus exposure prediction using machine learning [published online ahead of print]. *Clin Pharmacol Ther*. <https://doi.org/https://doi.org/10.1002/cpt.2123>.
25. Woillard J-B, Labriffe M, Debord J, Marquet P. Mycophenolic acid exposure prediction using machine learning [published online ahead of print]. *Clin Pharmacol Ther*. <https://doi.org/https://doi.org/10.1002/cpt.2216>.
26. Brier ME, Zurada JM, Aronoff GR. Neural network predicted peak and trough gentamicin concentrations. *Pharm Res An Off J Am Assoc Pharm Sci*. 1995;12:406-412.
27. Yamamura S, Kawada K, Takehira R, et al. Artificial neural network modeling to predict the plasma concentration of aminoglycosides in burn patients. *Biomed Pharmacother*. 2004;58(4):239-244.
28. Van Looy S, Verplancke T, Benoit D, et al. A novel approach for prediction of tacrolimus blood concentration in liver transplantation patients in the intensive care unit through support vector regression. *Crit Care*. 2007;11(4):R83.
29. Tang J, Liu R, Zhang Y-L, et al. Application of machine-learning models to predict tacrolimus stable dose in renal transplant recipients. *Sci Rep*. 2017;7(1):42192.
30. Barbieri C, Molina M, Ponce P, et al. An international observational study suggests that artificial intelligence for clinical decision support optimizes anemia management in hemodialysis patients. *Kidney Int*. 2016;90(2):422-429.
31. Imai S, Takekuma Y, Miyai T, Sugawara M. A new algorithm optimized for initial dose settings of vancomycin using machine learning. *Biol Pharm Bull*. 2020;43:188-193.
32. Maier C, Hartung N, Kloft C, Huisinga W, Wiljes J. Reinforcement learning and Bayesian data assimilation for model-informed precision dosing in oncology. *CPT Pharmacometrics Syst Pharmacol*. 2021;10:241-254.
33. McComb M, Bies R, Ramanathan M. Machine learning in pharmacometrics: opportunities and challenges. *Br J Clin Pharmacol*. <https://doi.org/https://doi.org/10.1111/bcp.14801>.
34. Hutchinson L, Steiert B, Soubret A, et al. Models and machines: how deep learning will take clinical pharmacology to the next level. *CPT Pharmacometrics Syst Pharmacol*. 2019;8(3):131-134.
35. Bayard DS, Jelliffe RW. A Bayesian approach to tracking patients having changing pharmacokinetic parameters. *J Pharmacokinetic Pharmacodyn*. 2004;31:75-107.
36. Macdonald I, Staatz CE, Jelliffe RW, Thomson AH. Evaluation and comparison of simple multiple model, richer data multiple model, and sequential Interacting Multiple Model (IMM) Bayesian analyses of gentamicin and vancomycin data collected from patients undergoing cardiothoracic surgery. *Ther Drug Monit*. 2008;30:67-74.

## SUPPORTING INFORMATION

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

**How to cite this article:** Hughes JH, Keizer RJ. A hybrid machine learning/pharmacokinetic approach outperforms maximum a posteriori Bayesian estimation by selectively flattening model priors. *CPT Pharmacometrics Syst Pharmacol*. 2021;10:1150-1160. <https://doi.org/10.1002/psp4.12684>