A Model Averaging/Selection Approach Improves the Predictive Performance of Model-Informed Precision Dosing: Vancomycin as a Case Study

David W. Uster[1], Sophie L. Stocker[2,3], Jane E. Carland[2,3], Jonathan Brett[2,3], Deborah J.E. Marriott[3,4], Richard O. Day[2,3] and Sebastian G. Wicha[1,*]

Many important drugs exhibit substantial variability in pharmacokinetics and pharmacodynamics leading to a loss of the desired clinical outcomes or significant adverse effects. Forecasting drug exposures using pharmacometric models can improve individual target attainment when compared with conventional therapeutic drug monitoring (TDM). However, selecting the "correct" model for this model-informed precision dosing (MIPD) is challenging. We derived and evaluated a model selection algorithm (MSA) and a model averaging algorithm (MAA), which automates model selection and finds the best model or combination of models for each patient using vancomycin as a case study, and implemented both algorithms in the MIPD software "TDMx." The predictive performance (based on accuracy and precision) of the two algorithms was assessed in (i) a simulation study of six distinct populations and (ii) a clinical dataset of 180 patients undergoing TDM during vancomycin treatment and compared with the performance obtained using a single model. Throughout the six virtual populations the MSA and MAA (imprecision: 9.9–24.2%, inaccuracy: less than ± 8.2%) displayed more accurate predictions than the single models (imprecision: 8.9–51.1%; inaccuracy: up to 28.9%). In the clinical dataset, the predictive performance of the single models applying at least one plasma concentration varied substantially (imprecision: 28–62%, inaccuracy: −16 to 25%), whereas the MSA or MAA utilizing these models simultaneously resulted in unbiased and precise predictions (imprecision: 29% and 30%, inaccuracy: −5% and 0%, respectively). MSA and MAA approaches implemented in TDMx might thereby lower the burden of fit-for-purpose validation of individual models and streamline MIPD.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ Bayesian forecasting using population pharmacokinetic models is increasingly recognized as a useful tool to inform optimal dose selection, particularly for drugs used to treat infectious diseases.

WHAT QUESTION DID THIS STUDY ADDRESS?

☑ The selection of the most accurate model to inform optimal dosing remains challenging, particularly for use in heterogenous and complex patient populations and novel, pragmatic approaches to guide model selection are required.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

☑ The use of an automated model averaging/selection approach allows for model structure uncertainty, while retaining at least the performance level of the most appropriate model in the algorithm for the individual patient. Implemented in the open-access software TDMx makes the approach easily accessible.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

☑ The developed algorithms may increase the accuracy of precision dose calculations associated with a better response as well as lowering the burden of fit-for-purpose validation of pharmacometric models for model-informed precision dosing.

[1]Department of Clinical Pharmacy, Institute of Pharmacy, University of Hamburg, Hamburg, Germany; [2]Department of Clinical Pharmacology and Toxicology, St. Vincent's Hospital, Sydney, New South Wales, Australia; [3]St. Vincent's Clinical School, University of New South Wales, Sydney, New South Wales, Australia; [4]Department of Clinical Microbiology and Infectious Diseases, St. Vincent's Hospital, Sydney, New South Wales, Australia.
*Correspondence: Sebastian G. Wicha (sebastian.wicha@uni-hamburg.de)

Received July 17, 2020; accepted September 12, 2020. doi:10.1002/cpt.2065

Pharmacometric models, if implemented in model-informed precision dosing (MIPD) software, can support dose individualization through forecasting future drug responses. The process, also referred to as Bayesian forecasting, usually includes the computational combination of patient information, individual plasma concentrations, and prior information in the form of population pharmacokinetic (PopPK) models to generate individual estimates. The estimated responses, in turn, can be used to assess whether (future) pharmacokinetic (PK)/pharmacodynamic targets can be attained with or without dose adaptations.

High-impact examples of applied MIPD[1–4] demonstrate the usefulness of treating individuals instead of populations.[5] Nonetheless, MIPD as a crucial pillar of precision medicine has not yet become integrated into clinical practice on a large scale.[6] A number of barriers toward the implementation of MIPD have been identified, including the lack of clinically oriented training in MIPD, the lack of acceptance of more complex dosing strategies (by prescribers),[7] unclear reimbursement, or the lack of pharmaceutical industry support.[8] Another challenge associated with MIPD is the selection of an appropriate pharmacometric model and the related fit-for-purpose validation required.[9]

The model selection process is challenging as PopPK models are usually developed and validated to provide quantitative insight into the PKs of a specific population, but their forecasting performance in diverse real-world populations undergoing therapeutic drug monitoring (TDM) is rarely evaluated. Selecting the "incorrect" model can potentially result in inappropriate dose recommendations and therefore lead to patient harm and/or suboptimal treatment outcomes and repeatedly validating models in different populations is arduous and costly.[9–12]

The objective of this study was to mitigate the challenge of the clinical application of MIPD by developing and evaluating a model selection algorithm (MSA) and a model averaging algorithm (MAA), which automates the model selection process for use in Bayesian forecasting. Both algorithms were compared with individual pharmacometric models in comprehensive simulation studies as well as in a heterogenous real-world clinical TDM dataset of patients administered vancomycin and subsequently implemented in the web-based MIPD software TDMx.[13] Vancomycin was exemplarily chosen because the latest international consensus guideline for TDM of vancomycin recommends using area under the curve (AUC) guided dosing through MIPD[14] and therefore suitable PopPK models are urgently required.

METHODS

Single-model approach

Contemporary Bayesian forecasting software commonly utilizes the single-model approach to inform dose selection.[15] Six published vancomycin PopPK models, developed in distinct patient populations, including extremely obese, critically ill, hospitalized, and those with sepsis, trauma, and post-heart surgery[16–21] were encoded in NONMEM (version 7.4.3; ICON plc, Dublin, Ireland). A detailed overview of the model properties can be found in **Table S1**. As an external reference, the two-compartment model of Goti *et al.*,[22] which was recently determined as the most accurate for vancomycin Bayesian forecasting was used.[10] Furthermore, to compare the predictive performance of the MSA and MAA to a best case of a single model Bayesian forecast, we re-estimated the parameters of the Goti model[22] using our clinical dataset (**Table S1**).

Multimodel approach

Two automated multimodel approaches were developed with the same six candidate PopPK models to be used simultaneously for forecasting individual PK profiles of either simulated or real-world patients (see below). Both algorithms comprised three essential steps (**Figure 1**): (1) the PK parameter estimation, (2) the automated comparison of the model fits, and (3) the adjustment of the forecasted concentration-time profiles.

Model selection algorithm. First, the individual PK parameters were estimated with each of the six PopPK models on observed data. Second, the forecasted (i.e., predicted) concentration-time data was automatically weighted using the weighting schemes described below. Third, the MSA selected the best fitting model via the obtained weightings and the competing models were discarded.

Model averaging algorithm. In the MAA, the available concentration-time data was used to average the model predictions under consideration balancing the predictions of each model using different weighting schemes that reflect individual goodness of fit. Similar to the MSA, the PK parameters of the individual patients were estimated with each PopPK model and individual weightings assigned. Then, in contrast to the MSA, the MAA averaged the model-predicted vancomycin concentrations at each forecasted time point using the set of PopPK models jointly with the data-derived weighting scheme. Furthermore, to investigate the vancomycin-specific target, the MAA also averaged the forecasted AUC.

Weighting schemes. The second step of the MSA/MAA requires a criterion to quantify the individual model fits with respect to the candidate models. Therefore, three different weighting schemes that summarize different model fit metrics were investigated: the objective function value (OFV), an adjusted Akaike information criterion (AIC) or the squared prediction errors (SSEs). Subsequently, the most suitable weighting scheme was implemented in the multimodel approaches.

The first weighting scheme compared the maximum likelihood (LL) obtained through the OFV of the ith model relative to the set of n models as follows:

$$W_{\mathrm{OFV}_i} = \frac{\mathrm{LL}_i}{\sum_1^n \mathrm{LL}_n} = \frac{e^{\left(-0.5\times \mathrm{OFV}_i\right)}}{\sum_1^n e^{\left(-0.5\times \mathrm{OFV}_n\right)}} \quad (1)$$

The second weighting scheme (W_{AIC}) consisted of an adjusted AIC with two main parts, adopting the approach proposed by Aoki *et al.*[23]: the LL and a penalizing term. In contrast to conventional AIC,[24] solely the number of random-effect parameters that quantify the magnitude of explained variability (i.e., interindividual variability (IIV)), were included in the penalizing term k (Eq. 2).

$$W_{\mathrm{AIC}_i} = \frac{e^{\left(\ln(\mathrm{LL})_i - k\right)}}{\sum_1^n e^{\left(\ln\ (\mathrm{LL})_n - k\right)}} \quad (2)$$

The third weighting scheme utilized the SSEs, which, in contrast to Eqs. 1 and 2, excludes the influence of the model structure on the predictions (Eq. 3). Here, *true* represents the measured and *pred* means the predicted value of the jth observation, respectively.

$$W_{\mathrm{SSE}_i} = \frac{e^{\left(-0.5\times \mathrm{SSE}_i\right)}}{\sum_1^n e^{\left(-0.5\times \mathrm{SSE}_n\right)}} = \frac{e^{\left(-0.5\times \sum\left(\mathrm{true}_j - \mathrm{pred}_j\right)\right)}}{\sum_1^n e^{\left(-0.5\times \sum\left(\mathrm{true}_j - \mathrm{pred}_j\right)\right)}} \quad (3)$$

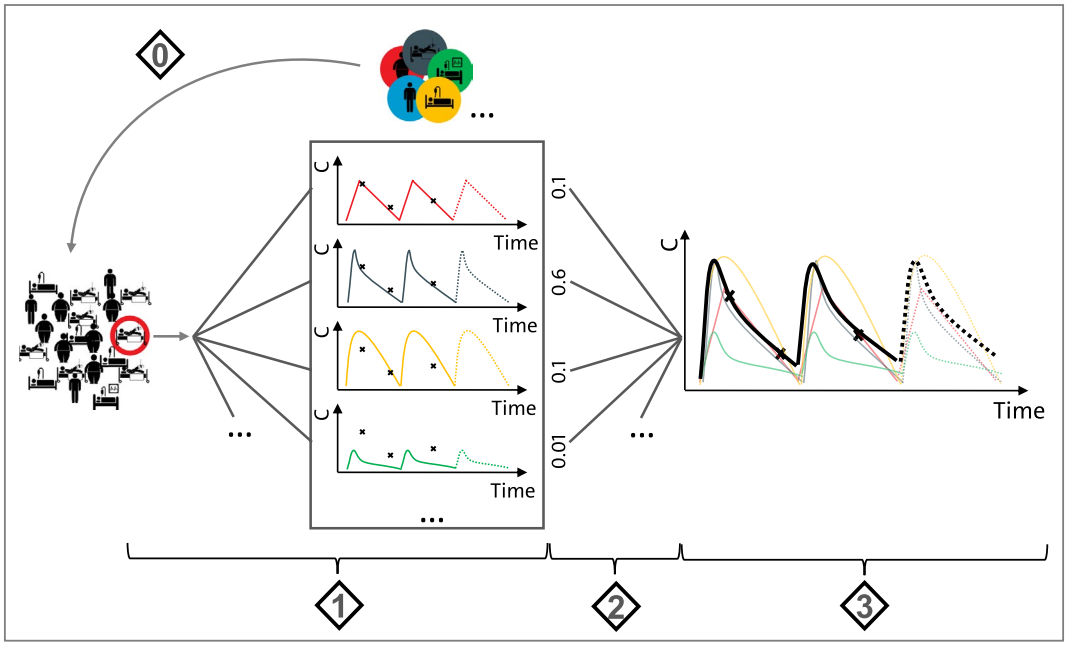

Figure 1 Model averaging scheme applied to a patient from a heterogenous or unknown population (red circle) with information on dosing, plasma concentrations (black cross), and the relevant covariates to forecast a future pharmacokinetic (PK) profile (dotted line). The algorithm comprises three steps: (1) Estimation of the PK profiles with a certain number of selected models. (2) Automated comparison of the individual model fit via a predefined criterion (e.g., the objective function value) and calculation of individual weights (e.g., W_{OFV}). The better the fit of the model, the higher the weighting. (3) Adjustments of the predictions by the respective weighting and building a weighted average (black line) with the best fitting model given the highest influence (model averaging algorithm) or being selected, while others are discarded (model selection algorithm). (0) is not part of the algorithm, but explains the simulation study graphically. Colored icons – models developed in distinct populations; colored lines – estimated PK profile of the chosen patient using the models, respectively. [Colour figure can be viewed at wileyonlinelibrary.com]

Robustness of the algorithm. In order to evaluate how many candidate models per multimodel approach were required for accurate predictions, we sequentially excluded the model(s) with the best performance metrics and estimated the vancomycin PK profiles of patients in the clinical dataset (see below) using the MSA and MAA with the weighting scheme W_{OFV}. The predictive performance of the algorithms with the smaller set of models was compared via the forecasting performance metrics (Eqs. 4 and 5).

Simulation study

A virtual population of 1,000 patients with randomly acquired covariates receiving the same vancomycin dose every 12 hours was created. The covariates were either sampled from a normal (age, body mass index, and body height), log normal (serum creatinine), or a binomial distribution (sex) to mimic a real adult population. Parameter details and correlations can be found in **Table S2**. With each of the six models, PK data were simulated, including a peak and a trough plasma concentration from three dosing intervals and the AUC between 24 and 36 hours (i.e., true AUC). Subsequently, the PK data obtained from the 6,000 simulated patients (1,000 per model) were evaluated using the single-model approach as well as the two multimodel approaches.

Clinical data

The predictive performance of the algorithms was evaluated in previously published heterogeneous clinical datasets.[10,25] Data from hospitalized and critically ill patients (n = 374) with 1,967 routine vancomycin TDM samples between January 2010 and August 2019 were retrospectively collected. This dataset includes 180 patients for whom plasma vancomycin concentrations were available for 3 dosing occasions (samples = 741). Patient demographics and clinical data are summarized in **Table S3**. The studies were approved by the St. Vincent's Hospital Human Research Ethics Committee in Sydney (2019/ETH09850, 2019/ETH02942 and 2019/ETH03054). Vancomycin concentrations were determined by standard immunoassay (EMIT 2000, Siemens Healthineers). Eleven datapoints (0.6% of all plasma concentrations) were below the limit of quantification of 2 mg/dL and were excluded. The age, bodyweight, and height were available in every patient. If specific continuous covariates were not available, either the median of the dataset or the median of the model population was imputed. If categorical covariates were missing, the data were assumed to be in the negative category (e.g., not receiving concomitant furosemide).

Forecasting performance metrics

The predictive performance of the MSA and MAA were compared with the single-model approach using the six PopPK models as well as to the "reference" PopPK model.[22] Predictive performance was assessed via the differences between the predicted and observed plasma vancomycin concentrations (clinical data) or the simulated and predicted AUC. Three different scenarios (outlined below) were used to predict plasma concentrations and the AUC in the third observed dosing occasion (ODO), where observed plasma vancomycin concentrations were "hidden" from the models or algorithms. The third ODO was chosen, as in the clinical dataset information from at least two previous dosing intervals were available in the majority (52%) of the patients. The ODOs were defined as not necessarily consecutive dosing intervals for which one or more observed vancomycin concentration(s) were available.

The first and second scenarios mimicked the bedside process of achieving a clinical target. First, predictions were made using solely the patient covariates (*a priori*). Second, the vancomycin concentrations from one or two previous ODOs were used in addition to the covariate information to forecast the PK profiles in the third (hidden) ODO. The third scenario

was used to retrospectively determine the general fit of the models/algorithms to the data by including the third ODO. Relative root mean square error (rRMSE, Eq. 4) and relative bias (rBias, Eq. 5) were used to determine (im-)precision and (in-)accuracy of the forecasted parameters, respectively.[26] The metrics were calculated relative to the observed plasma concentrations in the clinical dataset and the predicted AUC relative to the true AUC obtained in the simulations.

$$rRMSE = \sqrt{\frac{1}{n} \times \sum_{1}^{i} \left(\frac{(\text{predicted}_i - \text{true}_i)^2}{\text{true}_i^2} \right)} \times 100 \quad (4)$$

$$rBias = \frac{1}{n} \times \sum_{1}^{i} \left(\frac{(\text{predicted}_i - \text{true}_i)}{\text{true}_i} \right) \times 100 \quad (5)$$

The performance of the models was considered clinically acceptable if the rBias was between −20% and 20%, with the 95% confidence intervals (CI) including zero.[25] Additionally, the precision metric (rRMSE) should be as small as possible.

The simulations and all data fitting processes were done in NONMEM in conjunction with PsN (version 4.9.0).[27] The "tidyverse" package (version 1.3.0)[28] in R (version 3.6.1)[29] was used to develop the algorithms as well as to assess the results graphically and mathematically.

MIPD software for vancomycin

For translation of the herein presented results into clinical practice, we developed a vancomycin dosing module in the open-access MIPD software TDMx.[13] The software module contains the single models as well as the MSA and MAA algorithms and allows for AUC-based dose calculations of vancomycin. The software can be accessed under http://www.tdmx.eu/.

RESULTS

Evaluation of the weighting schemes

Three weighting schemes (i.e., W_{OFV}, W_{AIC}, and W_{SSE}) were evaluated. Whereas the weighting scheme W_{OFV} consisted solely of the likelihood, obtained through the NONMEM calculated OFV, the W_{AIC} further accounted for model complexity through penalizing IIV. Due to the low number of observations per patient in the Bayesian estimation of the PK parameters (e.g., maximum four observations in the simulation study), the penalizing term dominated the resulting W_{AIC}, and therefore shifted the influence of the more complex models toward the less penalized ones (**Figure 2** vs. **Figure S1**). In comparison to W_{OFV}, the model, including the highest number of IIV terms (Thomson[21]), was less often selected in the forecasting of hospitalized patients (simulated by the Thomson model[21]) using W_{AIC} (one occasion: 63.9% vs. 0.3% and two occasions 71.3% vs. 8.9%). Simultaneously, the Adane[16] and Revilla[19] models were increasingly selected (< 13% to > 28%). Despite this shift in the distribution of the weights, the performance metrics of the MSA and MAA using the W_{OFV} were just slightly superior over the W_{AIC} (e.g., MSA one occasion rBias/rRMSE 0.6%/18.6% vs. 1.8%/19.4%; **Figure S2**). The W_{OFV} was slightly preferable to W_{SSE} with an rRMSE being 0.9% and an rBias 0% smaller on average. Therefore, the W_{OFV} was used in the following analyses.

Simulation study

Across all six simulated populations, the proportion of patients, for which the selected model corresponded with the model used to simulate the data, ranged from 21.1% in the Medellín-Garibay model[18] to 63.9% in the Thomson model[21] when using data from one ODO (**Figure 2**, one occasion). Although the PK parameters of the other patients were estimated with the algorithm assigning the highest weight to the other five models (i.e., a "wrong" model), the MSA led to similar metrics as the best single-model approach in the respective (sub-)population (**Figures S3 and S4**).

Overall, the metrics of the MAA and MSA were in good alignment throughout the 6,000 simulated patients with slightly higher imprecision and inaccuracy in the MAA approach (mean

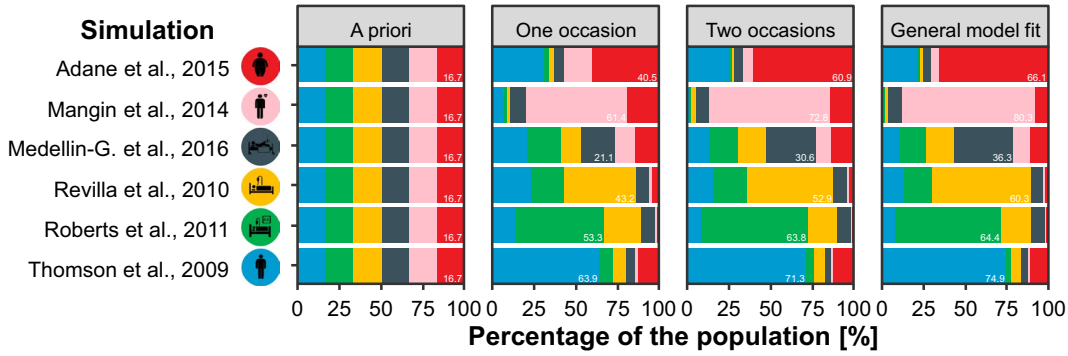

Figure 2 Influence of the models in the multimodel approaches using the weighting scheme W_{OFV} and stratified by the simulated populations (*y-axis*). Each patient was stained in the color of the particular model which obtained the highest weight in the algorithm. *(A priori)* prediction using the patient covariates only; Bayesian forecasting using plasma vancomycin concentrations from (One occasion) the second (i.e., most recent) observed dosing occasion and (Two occasions) the first and second observed dosing occasion; (General model fit) Bayesian estimation using plasma vancomycin concentrations from all three dosing intervals. White numbers – numerical value of the biggest portion in the subpopulation and scenario, respectively. In each simulation, the pharmacokinetic parameters of majority of the individuals were predicted solely by (for model selection algorithm) or mostly influenced by (for model averaging algorithm) the same model used in the underlying simulation. This pattern increased when more data was supplied and indicates that the algorithm detected the underlying true simulation model (i.e., the optimal model for each individual patient). [Colour figure can be viewed at wileyonlinelibrary.com]

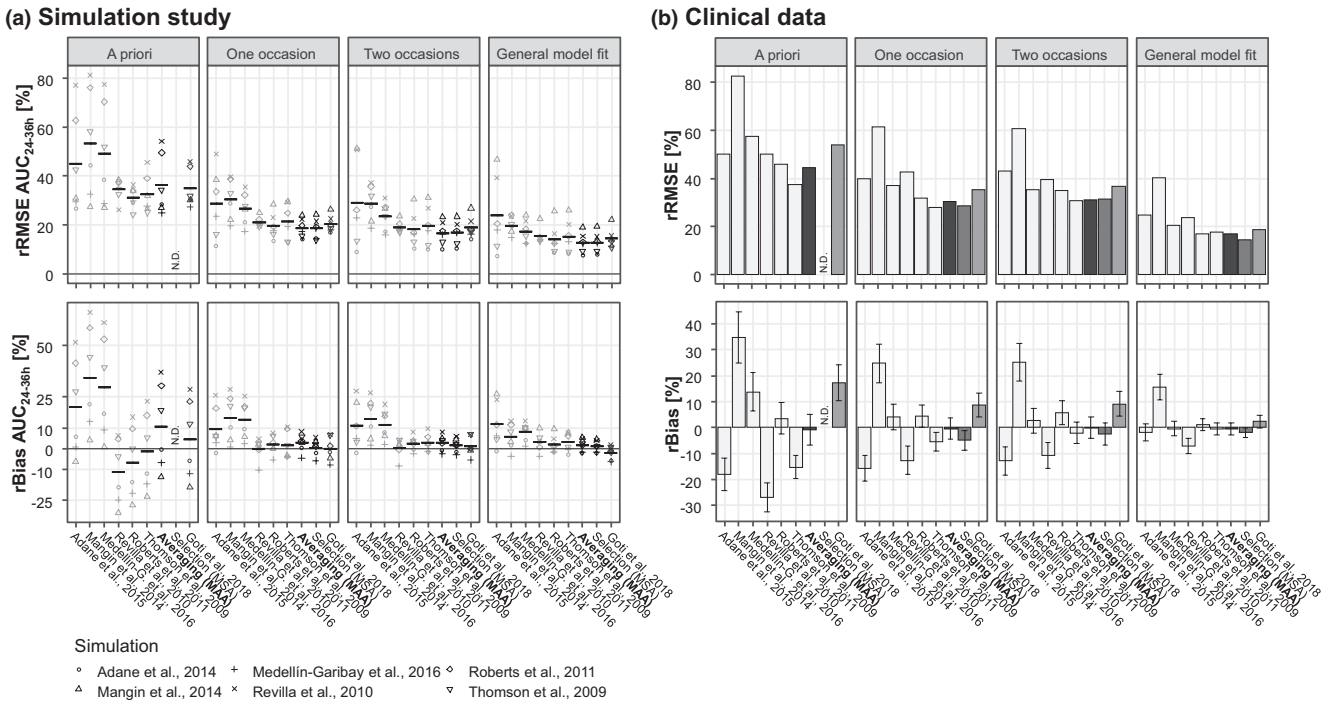

Figure 3 The relative root mean square error (rRMSE) and relative bias (rBias) when predicting either the area under the curve (AUC) or concentration-time data in the third dosing occasion, which is blinded to the models/algorithms from in various settings: *A priori* prediction using the patient covariates only; Bayesian forecasting using plasma vancomycin concentrations from (One occasion) the second (i.e., most recent) observed dosing occasion, and (Two occasions) the first and second observed dosing occasions; (General model fit) Bayesian estimation using plasma vancomycin concentrations from all three dosing occasions. (**a**) Simulation study; the predicted vs. the simulated AUC between 24 hours and 36 hours calculated in the whole 6,000 simulated patients (horizontal line) and the (sub-)populations (shapes), respectively, and (**b**) clinical data; the predicted vs. the vancomycin concentrations in the third observed dosing occasion obtained in the clinical studies. The ordinate is displaying the six single model approaches (light grey), the model averaging algorithm (black), the model selection algorithm (dark grey), and the external model (grey) per scenario. Whiskers cover the 95 % confidence interval of the relative bias calculated via the standard error. N.D., not defined.

absolute difference rRMSE 0.1% and rBias: 1.7% in the forecasting; **Figure 3a**). In comparison with the single-model performance throughout the 6,000 heterogenous virtual patients, the MSA displayed the most precise predictions ranging from an rRMSE of 18.6% (18.8% MAA) in the forecasting using one occasion to 12.9% (12.8% MAA) in the general model fit vs. the single model approaches with an imprecision of 19.3–30.3% in the forecasting or 14.1–23.8% in the general model fit. The rBias of the MAA and MSA was always less than ± 9% and < 15% using the single models, except in the *a priori* setting, which is by default an average of the model predictions with equal weightings (MAA) or not defined (MSA). In comparison to the *a priori* forecast, the inclusion of concentrations from one ODO led to an improvement of the precision and accuracy of the MAA by a factor of two and four, respectively. Additional concentrations from a second ODO improved predictions only marginally.

In each simulated population, the predictions of the MSA and MAA outperformed most of the single-model approaches with an rBias between −5.9% and 8.2% and an rRMSE always < 24.2% in the forecasting and general model fit, whereas the single models varied between −10.3% and 28.9% (rBias) and an rRMSE of up to 51.1% (**Figure 3a**–shapes). The single-model approach was only slightly

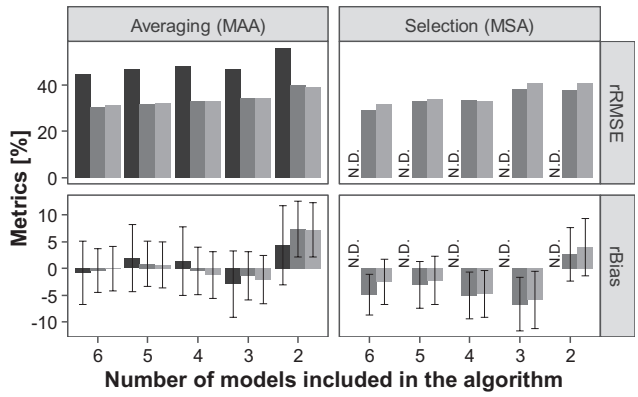

Figure 4 Robustness of the model averaging algorithm (MAA; left) and the model selection algorithm (MSA; right) displayed via the relative root mean square error (rRMSE) and relative bias (rBias) of the clinical data. The successively excluded models are: Thomson, Roberts, Medellín-G., Adane (from left to right), remaining: Mangin, Revilla. (black) *a priori* prediction using the patient covariates only; Bayesian forecasting using plasma vancomycin concentrations from the second (i.e., most recent) dosing occasion (dark grey), and the first and second dosing occasions (light grey). Whiskers cover the 95% confidence interval of the rBias calculated via the standard error. N.D., not defined.

more precise, when it was used to forecast the PK within the respective population the model was developed for (**Figures S3 and S4**).

In comparison to the "reference" model of Goti,[22] the MSA and MAA were more precise (rRMSE) while being similarly accurate (rBias < ± 10%). The rRMSE using the Goti model[22] ranged from 20.3% to 19.0% (one and two occasions), whereas the precision of the multimodel approaches ranged from 18.8% to 16.8%.

Clinical data

Forecasting performance. The MSA and MAA were applied to the clinical dataset and confirmed the simulation study findings. Although the forecasting performance of the single models varied substantially (rBias: −16 to 25%, rRMSE: 28–62%), the MAA using these models simultaneously resulted in unbiased and precise predictions using data from one (rBias: 0%, rRMSE: 30%) and two previous ODOs (rBias 0%, rRMSE: 31%; **Figure 3b**), matching the simulation study results that additional concentrations only marginally improved the predictions. The rRMSE of the MAA was always in the range of the best single model (absolute deviation to Thomson model[21]: *a priori*: 7.2%, forecasting: 2.7%/0.5%, and general model fit: −0.7%).

Although both MSA and MAA displayed an rRMSE in the forecasting between 29% and 32%, the MAA was slightly more accurate (rBias one occasion/two occasions −0.4%/−0.1% MAA vs. −5.0%/−2.5% MSA). Although the 95% CI of the two algorithms were overlapping in large parts, only the 95% CI of the rBias in the MAA included 0 in all settings. The MAA, unlike all single-models, met the clinical acceptance criteria in every scenario. The approaches performed better than the recently evaluated best model of Goti[10] (rBias one occasion/two occasions: 8.6%/9.1%, rRMSE: 35% and 37%). Even if

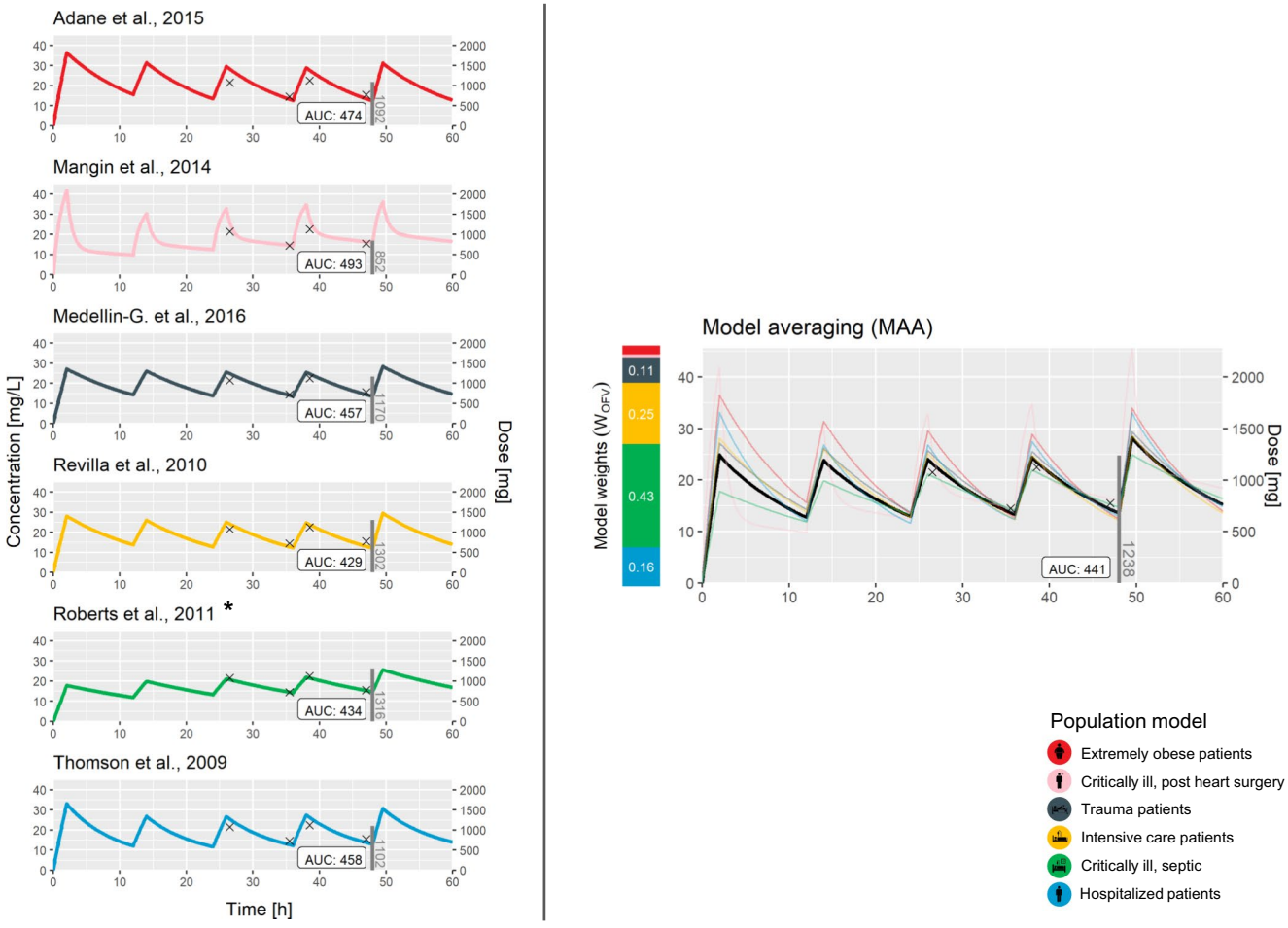

Figure 5 Calculation of an optimal vancomycin dosing on day 3 in a critically ill example patient (male, 70 kg, 1.75 m, 56 years, serum creatinine: 80 µmol/L) to attain the target area under the curve $(AUC)_{24h}$/minimal inhibitory concentration (MIC) ratio of 500 with the model-informed precision dosing software TDMx. The patient received 1,000 mg vancomycin twice daily, following a loading dose of 2,000 mg and 2 peak and 2 trough vancomycin plasma concentrations were measured between 23 and 14 mg/L. With the single-models (left), model fit was heterogeneous and the determined AUC_{24h}/MIC ratio varied from 429 to 493. The single model-derived dose recommendations varied between 852 mg and 1,316 mg (grey). The model selection algorithm (MSA) selected the Roberts model as indicated by the highest weight (*). The model averaging algorithm (MAA; black line, right panel) predicted an AUC_{24h}/MIC ratio of 441 prior dose adjustment, and the subsequent dose recommendations was 1,238 mg. The MAA was mainly influenced by the Roberts and Revilla model, both derived in the critically ill population, as indicated by the model weights (barplot). Both the MSA and the MAA calculated a plausible dose adjustment being in line with the most accurate single model. [Colour figure can be viewed at wileyonlinelibrary.com]

the parameters of the best reference model[22] were re-estimated based on our clinical dataset and the predictive performance of the adjusted model improved, the two algorithms were similar to, or performed better than, the single model (**Figure S5**).

Robustness of the algorithm. In order to evaluate how many models were needed in the MSA and MAA, we successively excluded the model with the best weighting and assessed the performance of the algorithms consisting of six to two models. The predictive performance of the MAA was "stable" even when only three instead of six models were used, where the rRMSE increased from 30.4% to 34.2% in the forecasting (44.4–46.6% *a priori*) and the rBias varied between −2.1% and 0.7% (−2.9 to 1.9% *a priori*; **Figure 4**). The 95% CI of the rBias included 0 in all scenarios and never exceeded ± 7.4%. When using two models in the MAA, the rRMSE increased by 5.7% in the forecasting (9.2% *a priori*) with a rBias of 7.3% (4.3% *a priori*). The exclusion of up to four models in the MSA resulted in greater imprecision (rRMSE 28.7% to 40.4%) and a rBias between −6.7% to 3.9% (**Figure 4**). In comparison to the MAA, the forecasting of the MSA was less accurate (mean rBias MSA −2.9%; MAA 1.0%) and less precise with a larger rRMSE in 8 of 10 settings.

MIPD software for vancomycin

The vancomycin module in TDMx was cross-validated against NONMEM (version 7.4.3) indicating virtually identical results of the model predictions as well as objective function values (**Figures S6 and S7**).

A patient case example using the single-model approaches as well as MSA and MAA is presented in **Figure 5**. Single models did not only provide a very heterogenous fit to the example patient, but also derived dose recommendations to attain an AUC_{24h} over minimal inhibitory concentration ratio of 500 were highly variable ranging from 852–1,316 mg, whereas the MSA and MAA predicted an optimal dose of 1,316 and 1,238 mg, respectively.

DISCUSSION

For vancomycin—one of the most commonly used antibiotics in clinical practice[30]—more than 30 PopPK models have been developed in diverse patient populations.[10] However, to choose and validate a model for an individual patient might not always be within the skill-set of the decision maker. We therefore provide two new multimodel approaches using automated model selection/averaging in Bayesian forecasting, with a better forecasting performance than a single PopPK model. When implemented in MIPD software, the clinical decision maker does not need to rely on one predetermined model but can automatically allow the algorithms to find the most suitable predictions for an individual patient. Thereby the precision dosing process will be streamlined and the burden of local model validation lowered. Moreover, by implementing the MSA and MAA into the web-based MIPD software TDMx,[13] the developed algorithms are readily available to the scientific and clinical community.

The MSA/MAA required predictions to be adequately weighted. Therefore, three different weighting schemes were compared and the most suitable identified. The weightings derived from the W_{OFV}, which represent a balance between those from the W_{AIC} and W_{SSE}, provided superior weightings, although there was little difference in the predictive performance among the three schemes.

Although single pharmacometric models are usually evaluated in a specific population prior to their publication, extrapolating from a single model might not guarantee suitable predictions of concentration-time profiles in another, potentially very different, patient population. Even if the underlying population were known, the patient could still display atypical PK parameters. This implies that more flexibility is required when predicting PK parameters in clinical settings with patients from heterogenous populations. We demonstrated that the multimodel approaches provided this essential flexibility to forecast "future" vancomycin PK profiles of a heterogeneous population more accurately than using a single model.

Several factors could contribute to the superior performance of MSA and MAA over the single-model approach. First, predicting PK parameters with a single model ignores uncertainty in the structural model that could affect the predictive performance.[23,31] Specifically, a single model might not reflect the most suitable compartmental structure or parameterization. Furthermore, individuals differ in their physiology and therefore in their drug disposition. Extremely obese patients, for example, display a typical volume of distribution of vancomycin of 0.5 L/kg, one third of the value in patients with sepsis, whereas the clearance is comparable.[16,20] In contrast, the clearance of vancomycin determined in trauma patients is significantly lower than in extremely obese and more similar to the clearance in critically ill patients undergoing heart surgery.[17,18] These alterations in key PK parameters are especially relevant in critically ill patients[32] and highlight the importance of dosing decisions to be informed through careful selection of a pharmacometric model in MIPD. To ensure generalizability of the study findings the PopPK models selected to evaluate MSA and MAA were developed in diverse populations (i.e., extreme obese, critically ill, and hospitalized patients, among others). Second, a single model might include misinterpreted covariates (e.g., for burn status)[33] or lack covariates, which are critical in particular patients (e.g., correction for age, body weight, and kidney function).[34] However, the concurrent use of various models will include a larger range of covariates. Third, single models developed in small, homogenous cohorts of patients while displaying good internal predictivity, might not display external predictivity due to selection bias in the covariate submodel[35] and underestimated parameter uncertainty in studies with small patient numbers.[36] Finally, single models may have been developed based upon routine clinical data, which often exhibits uncertainty in documented dosing and sampling time. Uncertain sampling inflates the residual error of the models and thus can lead to a worse individual fit using the model in a sparsely sampled TDM setting due to a lower "trust" in the observed values and a higher impact of the Bayesian prior during Bayesian forecasting.[37]

When implementing MSA or MAA it might be challenging to select the candidate models to be included in the algorithms. To

this end, we evaluated the robustness of the MSA and MAA in this case study by successively excluding the best models (**Figure 4**). This stepwise exclusion of the candidate models resulted only in a minor decrease in the predictive performance of the MSA and MAA until only four and three models remained, respectively. Even with the two models that performed worst in the single model approach, a considerably improved forecasting performance could be observed when used in MSA, and even better in MAA (**Figure 4**). Whether this behavior is generalizable and whether a similar approach could be used to identify the best set of models would need to be evaluated in further MIPD settings beyond vancomycin, but this indicates that the algorithms in the evaluated case study are not dependent on a distinct single, well-performing model. This robustness makes the MAA and MSA an attractive approach also in settings where only a few PopPK models are available. Hence, MSA and MAA can use all relevant collective knowledge simultaneously to inform precision dosing calculations. Further, this process is automatable and enables the re-allocation of resources which would otherwise be required for the time-consuming task of identifying and validating the most suitable model for MIPD. Although Bayesian averaging in patient care[38] and drug development[23,39] has been discussed before, we, for the first time, systematically developed and evaluated such an approach in the context of MIPD.

Other popular approaches in MIPD are nonparametric in nature. The developed MSA and MAA combine the flexibility of nonparametric approaches[40] with the rigor of a parametric framework, without needing to know the nonparametric distribution (i.e., support points) of the population, which are rarely published. A comparison between MSA/MAA and nonparametric approaches with regard to handling outliers might be worth pursuing in future studies. Moreover, all PopPK models were coded from publications whereas nonparametric models require the derived sets of support points/raw data, which are rarely publicly available.

Some limitations of this study are acknowledged. The MSA and MAA require TDM data to weigh the predictions and find the best model during the course of therapy. For the first dose calculation (i.e., the *a priori* setting), a manual selection of the model is still required. Furthermore, none of the investigated vancomycin population models included interoccasion variability (IOV), a term quantifying PK variability between dosing occasions within the same individual. Although the inclusion of IOV during model development has been shown to be beneficial,[41] this variability can be challenging to handle in MIPD using the single-model approach.[42] Hence, investigation of MSA/MAA approaches with IOV should be evaluated in future studies.

Although the time required to complete the MSA and MAA per individual (mean 7.0 seconds with SD < 0.6 seconds using two samples and a set of six models) in the TDMx-based Bayesian forecasting was acceptable, calculating the uncertainty in the PK-profile predictions is only feasible with the MSA post-autoselection. Simulating the uncertainty with each model of the MAA set (e.g., using the Monte Carlo method) and averaging the obtained CI based on the weighting is highly time-consuming, and more research is required to provide assurance that the calculated CI will be statistically correct.

Moreover, in order to fully benefit from MIPD tools in critical therapies, the transferability of the present case study with vancomycin to other drugs in or outside the field of infectious diseases should be investigated.

In conclusion, the present study comprehensively evaluated model averaging and selection algorithms in MIPD using vancomycin as a case study. The algorithms overcome one of the major difficulties associated with the implementation of MIPD into clinical practice—selection of the most appropriate PopPK model for an individual patient. The developed algorithms can provide a more reliable Bayesian forecast when compared with using a single model.

SUPPORTING INFORMATION

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

ACKNOWLEDGMENTS

The authors want to thank the St Vincent's Curran Foundation Endowment Fund, Australia, for support to generate the clinical dataset from St Vincent's hospital. Furthermore, the authors acknowledge that parts of this analysis were submitted to the European Congress of Clinical Microbiology and Infectious Diseases 2020 and can be found in the Abstract book (Abstract 4886). Open access funding enabled and organized by Projekt DEAL.

FUNDING

No funding was received for this work.

CONFLICT OF INTEREST

The authors declared no competing interests for this work.

AUTHOR CONTRIBUTIONS

D.W.U. and S.G.W. wrote the manuscript. D.W.U. and S.G.W. designed the research. D.W.U., S.G.W., S.L.S., J.E.C., J.B., D.J.E.M., and R.O.D. performed the research. D.W.U. and S.G.W. analyzed the data.

1. Liang, T.J. & Ghany, M.G. Therapy of hepatitis C — back to the future. *N. Engl. J. Med.* **370**, 2043–2047 (2014).
2. Marsousi, N. *et al*. Coadministration of ticagrelor and ritonavir: toward prospective dose adjustment to maintain an optimal platelet inhibition using the PBPK approach. *Clin. Pharmacol. Ther.* **100**, 295–304 (2016).
3. Mould, D., D'Haens, G. & Upton, R. Clinical decision support tools: the evolution of a revolution. *Clin. Pharmacol. Ther.* **99**, 405–418 (2016).
4. Vinks, A., Emoto, C. & Fukuda, T. Modeling and simulation in pediatric drug therapy: application of pharmacometrics to define the right dose for children. *Clin. Pharmacol. Ther.* **98**, 298–308 (2015).
5. Deitchman, A.N. The risk of treating populations instead of patients. *CPT Pharmacometrics Syst. Pharmacol.* **8**, 256–258 (2019).
6. Darwich, A.S. *et al*. 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 (2017).

7. Peck, R.W. The right dose for every patient: a key step for precision medicine. *Nat. Rev. Drug Discov.* **15**, 145–146 (2016).
8. Polasek, T.M. *et al*. What does it take to make model-informed precision dosing common practice? Report from the 1st Asian Symposium on Precision Dosing. *AAPS J.* **21**, 17 (2019).
9. Heine, R. *et al*. Prospective validation of a model-informed precision dosing tool for vancomycin in intensive care patients. *Br. J. Clin. Pharmacol*. 2020; https://doi.org/10.1111/bcp.14360.
10. Broeker, A. *et al*. Towards precision dosing of vancomycin: a systematic evaluation of pharmacometric models for Bayesian forecasting. *Clin. Microbiol. Infect.* **25**, 1286.e1–1286.e7 (2019).
11. Guo, T. *et al*. External evaluation of population pharmacokinetic models of vancomycin in large cohorts of intensive care unit patients. *Antimicrob. Agents Chemother.* **63**, 1–9 (2019).
12. Turner, R.B. *et al*. Review and validation of Bayesian dose-optimizing software and equations for calculation of the vancomycin area under the curve in critically ill patients. *Pharmacother. J. Hum. Pharmacol. Drug Ther.* **38**, 1174–1183 (2018).
13. Wicha, S.G. *et al*. TDMx: a novel web-based open-access support tool for optimising antimicrobial dosing regimens in clinical routine. *Int. J. Antimicrob. Agents* **45**, 442–444 (2015).
14. Rybak, M.J. *et al*. Therapeutic monitoring of vancomycin for serious methicillin-resistant Staphylococcus aureus infections: a revised consensus guideline and review by the American Society of Health-System Pharmacists, the Infectious Diseases Society of America, the Pediatric Infectious Diseases Society, and the Society of Infectious Diseases Pharmacists. *Am. J. Health Pharm.* **77**, 835–864 (2020).
15. Sheiner, L.B., Beal, S., Rosenberg, B. & Marathe, V.V. Forecasting individual pharmacokinetics. *Clin. Pharmacol. Ther.* **26**, 294–305 (1979).
16. Adane, E.D., Herald, M. & Koura, F. Pharmacokinetics of vancomycin in extremely obese patients with suspected or confirmed staphylococcus aureus infections. *Pharmacother. J. Hum. Pharmacol. Drug Ther.* **35**, 127–139 (2015).
17. Mangin, O., Urien, S., Mainardi, J.-L., Fagon, J.-Y. & Faisy, C. Vancomycin pharmacokinetic and pharmacodynamic models for critically ill patients with post-sternotomy mediastinitis. *Clin. Pharmacokinet.* **53**, 849–861 (2014).
18. Medellín-Garibay, S.E. *et al*. Pharmacokinetics of vancomycin and dosing recommendations for trauma patients. *J. Antimicrob. Chemother.* **71**, 471–479 (2016).
19. Revilla, N., Martín-Suárez, A., Pérez, M.P., González, F.M. & Fernández de Gatta, M.D.M. Vancomycin dosing assessment in intensive care unit patients based on a population pharmacokinetic/pharmacodynamic simulation. *Br. J. Clin. Pharmacol.* **70**, 201–212 (2010).
20. Roberts, J.A. *et al*. Vancomycin dosing in critically ill patients: robust methods for improved continuous-infusion regimens. *Antimicrob. Agents Chemother.* **55**, 2704–2709 (2011).
21. Thomson, A.H., Staatz, C.E., Tobin, C.M., Gall, M. & Lovering, A.M. Development and evaluation of vancomycin dosage guidelines designed to achieve new target concentrations. *J. Antimicrob. Chemother.* **63**, 1050–1057 (2009).
22. Goti, V., Chaturvedula, A., Fossler, M.J., Mok, S. & Jacob, J.T. Hospitalized patients with and without hemodialysis have markedly different vancomycin pharmacokinetics. *Ther. Drug Monit.* **40**, 212–221 (2018).
23. Aoki, Y., Röshammar, D., Hamrén, B. & Hooker, A.C. Model selection and averaging of nonlinear mixed-effect models for robust phase III dose selection. *J. Pharmacokinet. Pharmacodyn.* **44**, 581–597 (2017).
24. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19**, 716–723 (1974).
25. Cunio, C.B. *et al*. Towards precision dosing of vancomycin in critically ill patients: an evaluation of the predictive performance of pharmacometric models in ICU patients. *Clin. Microbiol. Infect.* 2020; https://doi.org/10.1016/j.cmi.2020.07.005.
26. Sheiner, L.B. & Beal, S.L. Some suggestions for measuring predictive performance. *J. Pharmacokinet. Biopharm.* **9**, 503–512 (1981).
27. Keizer, R.J., Karlsson, M.O. & Hooker, A. Modeling and simulation workbench for NONMEM: tutorial on Pirana, PsN, and Xpose. *CPT Pharmacometrics Syst. Pharmacol.* **2**, 1–9 (2013).
28. Wickham, H. *et al*. Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
29. R Core Team. *R: A Language and Environment for Statistical Computing* (Vienna, Austria: R Foundation for Statistical Computing, 2019).
30. Magill, S.S. *et al*. Prevalence of antimicrobial use in US Acute Care Hospitals, May-September 2011. *JAMA* **312**, 1438 (2014).
31. Buckland, S.T., Burnham, K.P. & Augustin, N.H. Model selection: an integral part of inference. *Biometrics* **53**, 603 (1997).
32. Udy, A.A., Roberts, J.A. & Lipman, J. Clinical implications of antibiotic pharmacokinetic principles in the critically ill. *Intensive Care Med.* **39**, 2070–2082 (2013).
33. Colin, P.J. *et al*. Vancomycin pharmacokinetics throughout life: results from a pooled population analysis and evaluation of current dosing recommendations. *Clin. Pharmacokinet.* **58**, 767–780 (2019).
34. Marsot, A., Boulamery, A., Bruguerolle, B. & Simon, N. Vancomycin. *Clin. Pharmacokinet.* **51**, 1–13 (2012).
35. Ribbing, J. & Niclas Jonsson, E. Power selection bias and predictive performance of the population pharmacokinetic covariate model. *J. Pharmacokinet. Pharmacodyn.* **31**, 109–134 (2004).
36. Broeker, A. & Wicha, S.G. Assessing parameter uncertainty in small-n pharmacometric analyses: value of the log-likelihood profiling-based sampling importance resampling (LLP-SIR) technique. *J. Pharmacokinet. Pharmacodyn.* **4**, 219–228 (2020).
37. Alihodzic, D. *et al*. Impact of inaccurate documentation of sampling and infusion time in model-informed precision dosing. *Front. Pharmacol.* **11**, 1–12 (2020).
38. Mould, D.R., Upton, R.N. & Wojciechowski, J. Dashboard systems: implementing pharmacometrics from bench to bedside. *AAPS J.* **16**, 925–937 (2014).
39. Buatois, S., Ueckert, S., Frey, N., Retout, S. & Mentré, F. Comparison of model averaging and model selection in dose finding trials analyzed by nonlinear mixed effect models. *AAPS J.* **20**, 56 (2018).
40. Jelliffe, R.W. *et al*. Model-based, goal-oriented, individualised drug therapy. *Clin. Pharmacokinet.* **34**, 57–77 (1998).
41. Karlsson, M.O. & Sheiner, L.B. The importance of modeling interoccasion variability in population pharmacokinetic analyses. *J. Pharmacokinet. Biopharm.* **21**, 735–750 (1993).
42. Abrantes, J.A., Jönsson, S., Karlsson, M.O. & Nielsen, E.I. Handling interoccasion variability in model-based dose individualization using therapeutic drug monitoring data. *Br. J. Clin. Pharmacol.* **85**, 1326–1336 (2019).