International Journal of Antimicrobial Agents 64 (2024) 107305

Contents lists available at ScienceDirect

International Journal of Antimicrobial Agents

journal homepage: www.elsevier.com/locate/ijantimicag

Predictive performance of multi-model approaches for model-informed precision dosing of piperacillin in critically ill patients

Lea Marie Schatz[a,*], Sebastian Greppmair[a], Alexandra K. Kunzelmann[a], Johannes Starp[a], Alexander Brinkmann[b], Anka Roehr[c], Otto Frey[c], Stefan Hagel[d], Christoph Dorn[e], Michael Zoller[a], Christina Scharf[a], Sebastian G. Wicha[f], Uwe Liebchen[a,*]

[a] Department of Anaesthesiology, LMU University Hospital, LMU Munich, Munich, Germany
[b] Department of Anesthesiology and Intensive Care Medicine, General Hospital of Heidenheim, Heidenheim, Germany
[c] Department of Pharmacy, General Hospital of Heidenheim, Heidenheim, Germany
[d] Institute for Infectious Diseases and Infection Control, University Hospital, Friedrich-Schiller-University Jena, Jena, Germany
[e] Institute of Pharmacy, University of Regensburg, Regensburg, Germany
[f] Department of Clinical Pharmacy, Institute of Pharmacy, University of Hamburg, Hamburg, Germany

ARTICLE INFO

Article history:
Received 7 May 2024
Accepted 6 August 2024

Editor: Dr C. Landersdorfer

Keywords:
Therapeutic drug monitoring
Model-informed precision dosing
Population pharmacokinetics
Intensive care unit
Sepsis
Critically ill

ABSTRACT

Objectives: Piperacillin (PIP)/tazobactam is a frequently prescribed antibiotic; however, over- or underdosing may contribute to toxicity, therapeutic failure, and development of antimicrobial resistance. An external evaluation of 24 published PIP-models demonstrated that model-informed precision dosing (MIPD) can enhance target attainment. Employing various candidate models, this study aimed to assess the predictive performance of different MIPD-approaches comparing (i) a single-model approach, (ii) a model selection algorithm (MSA) and (iii) a model averaging algorithm (MAA).

Methods: Precision, accuracy and expected target attainment, considering either initial (B1) or initial and secondary (B2) therapeutic drug monitoring (TDM)-samples per patient, were assessed in a multicentre dataset (561 patients, 11 German centres, 3654 TDM-samples).

Results: The results demonstrated a slight superiority in predictive performance using MAA in B1, regardless of the candidate models, compared to MSA and the best single models (MAA, MSA, best single models: inaccuracy ±3%, ±10%, ±8%; imprecision: <25%, <31%, <28%; expected target attainment >77%, >71%, >73%). The inclusion of a second TDM-sample notably improved precision and target attainment for all MIPD-approaches, particularly within the context of MSA and most of the single models. The expected target attainment is maximized (up to >90%) when the TDM-sample is integrated within 24 h.

Conclusions: In conclusion, MAA streamlines MIPD by reducing the risk of selecting an inappropriate model for specific patients. Therefore, MIPD of PIP using MAA implicates further optimisation of antibiotic exposure in critically ill patients, by improving predictive performance with only one sample available for Bayesian forecasting, safety, and usability in clinical practice.

1. Introduction

The broad-spectrum antibiotic piperacillin (PIP) is prescribed in 99% of all intensive care units (ICU) in Germany [1]. However, standard doses resulted in sub- and supratherapeutic concentrations in critically ill patients [2]. Underdosing may lead to therapeutic failure and development of resistant bacteria, while overdosing is associated with potential toxicity [3–5]. Hence, the international Surviving Sepsis Campaign guidelines recommend the use of therapeutic drug monitoring (TDM) and/or software tools to optimise therapy [6].

Most software tools employed population pharmacokinetics (PopPK)-models to individualise the dose for each patient, which is called model-informed precision dosing (MIPD). Before applying MIPD, an external evaluation based on independent clinical data is essential to assess which PopPK-models might be useful for dosing decisions in critically ill patients [7–9]. Usually, PopPK-models were developed to characterise the pharmacokinetics of specific populations (e.g. liver transplant patients), which implies that the applicability to a broader group of patients may not be given.

* Corresponding authors.
E-mail addresses: lea.schatz@med.uni-muenchen.de (L.M. Schatz), uwe.liebchen@med.uni-muenchen.de (U. Liebchen).

https://doi.org/10.1016/j.ijantimicag.2024.107305

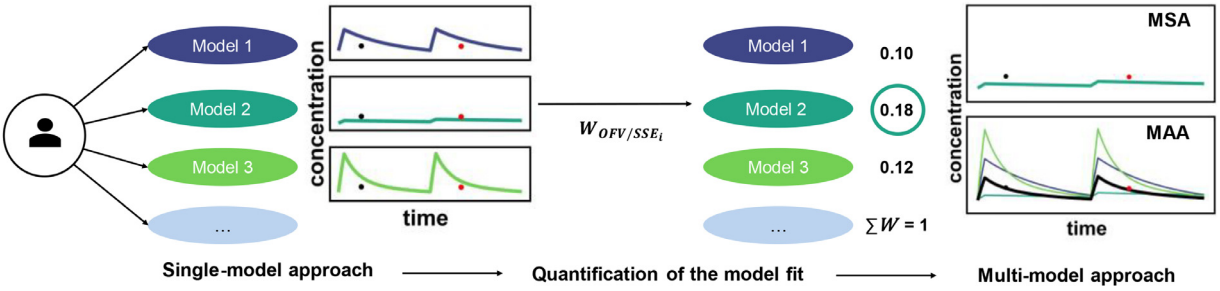

Fig. 1. Workflow from single-model approach to a multi-model approach. Single-model approach: Estimation of PopPK-model profiles for each patient using Bayesian forecasting from first TDM-sample (= black point) to future TDM-samples (= red point), each model is evaluated separately. Quantification of the model fit: For each patient the fractional individual weight of the model fit based on the first sample is calculated using a predefined weighting scheme (e.g., $W_{OFV\ or\ SSE}$). Multi-model approach: The algorithm (i) either selects the best fitting model for each patient (= maximum $W_{OFV/SSE}$) → model selection algorithm (MSA) or (ii) calculates a new prediction based on the individual weights of the models = $\sum_1^n (W_{OFV,\ SSE} \cdot c_{pred})$ → model averaging algorithm (MAA). PopPK: Population pharmacokinetic; W: calculated fractional weight; SSE: sum of squares error; OFV: objective function value; c: concentration; pred: predicted; n: set of models; i: individual patient; TDM: therapeutic drug monitoring.

We recently published a multicentre evaluation of available PIP-models and identified suitable candidates for successful MIPD. The best-performing models based on various evaluation criteria (e.g. goodness-of-fit-plots, statistical analysis) were the Kim-model [10] for the entire dataset, the Klastrup-model [11] for continuous infusion, and the Udy-model [12] for intermittent infusion, all derived from ICU-patient data [13]. Furthermore, subgroup analyses revealed that different models are beneficial for distinct situations and patient groups. As a consequence, the user is expected to be aware of the diversity of models and their preferred application, and to manually select the most appropriate model in each case, which hinders clinical applicability. This challenge could be addressed by employing a multi-model approach, as different models can be used simultaneously for predictions. The theory behind two existing multi-model approaches originally published by Uster et al. [14] is (i) to automatically select the best available model for each patient (model selection algorithm, MSA) or (ii) to calculate individual predictions based on a combination of PopPK-models (model averaging algorithm, MAA).

The objective of this analysis was to compare the predictive performance of a single-model approach with MSA and MAA by examining two clinical Bayesian scenarios: First, including only the initial TDM-sample per patient into the prediction (B1), and second, considering the first two measured TDM-samples per patient (B2).

2. Material and methods

2.1. Clinical data and candidate models

The external dataset included 561 patients from 11 study centres with 3654 TDM-samples. A detailed overview of the patients´characteristics as well as an overview of the 24 previously identified PopPK-models was published by Greppmair et al. [13] and is available in the Supplements S1.

2.2. The principle of multi-model approaches

The multi-model approach can be utilised if at least one TDM-sample is available. Predictions are calculated using multiple models integrated into the software, eliminating the need for the user to make a unique model selection.

The initial step in both multi-model approaches involves quantifying the model's prediction based on the included TDM-sample(s), i.e. a weighting scheme is applied to assess the quality of fit between the predicted concentration(s) (c_{pred}) and the measured TDM-concentration(s) (c_{obs}). Uster et al. [14] proposed two different weighting schemes, the objective function value (OFV), which is a goodness-of-fit criterion for evaluating PopPK-models (Eq. 1), and the sum of squares error (SSE), which is a mathematical criterion for quantifying the dispersion of observed values around predicted values (Eq. 2) [14].

$$W_{SSE_{model}} = \frac{e^{(-0.5 \cdot SSE_{model})}}{\sum_1^n e^{(-0.5 \cdot SSE_{model})}} = \frac{e^{-0.5 \cdot \sum (c_{obs,j} - c_{pred,j})^2}}{\sum_1^n e^{-0.5 \cdot \sum (c_{obs,j} - c_{pred,j})^2}} \quad (1)$$

$$W_{OFV_{model}} = \frac{LL_{model}}{\sum_1^n LL_{model}} = \frac{e^{(-0.5 \cdot OFV_{model})}}{\sum_1^n e^{(-0.5 \cdot OFV_{model})}} \quad (2)$$

W : calculated fractional weight; SSE : sum of squares error; OFV : objective function value; c : concentration; obs : observed/measured; pred : predicted; LL : Likelihood; $_n$: set of models; $_j$: observation

In a second step, depending on the multi-model approach, the algorithm either selects the most appropriate model for the individual patient based on this weighting scheme (select the model with the highest value in the weighting scheme = MSA) or computes new predictions by weighting the model predictions according to one of the aforementioned weighting schemes (= MAA) (Figure 1).

2.3. Weighting schemes

Based on the first TDM-sample, the model fit according to SSE and OFV was calculated for all selected models. Then, the fractional individual weight of each model and its impact on model selection/averaging were investigated. This was achieved by quantifying how often each model was selected by the algorithm (MSA) for all patients. The individual fractional weights for an individual patient (MAA) were calculated and the impact of SSE and OFV on statistical criteria in predictive performance were assessed. Influential models were defined as models with a fractional individual weight >0.001. In the present analysis, three preselection-scenarios were compared to evaluate the robustness of the multi-model approaches across different candidate models:

(1) All models (excluding the Asín-Prieto-model [15] were considered for MSA/MAA algorithms (Supplements S2)).
(2) Only the best models (models of Kim [10], Roberts-DM [16], Udy [12], Tamme [17], Klastrup [11]) identified in a previous study [13], were included.
(3) Five different models were selected, consisting of two with strong bias, two with moderate bias (one instance of positive bias and one instance of negative bias), and one model without bias (model of Chen [18], Bue [19], Sukarnjanaset [20], Hahn [21], Kim [10]).

2.4. Predictive performance

The predictive performance of the three preselection-scenarios was assessed and compared by evaluating two steps: (i) only the first measured TDM-sample per patient was included to predict all future concentrations (Bayesian 1=B1, 3093 predicted concentrations), and (ii) the first two TDM-samples per patient were utilised for Bayesian predictions (Bayesian 2=B2, 2532 predicted concentrations).

2.5. Statistical criteria

The median prediction error (MPE) for accuracy and median absolute prediction error (MAPE) for precision were calculated for all single models and the multi-model approaches. Additionally, the spread of the data was determined using the interquartile range (IQR) of the prediction errors (PE). The 95% confidence interval (CI) of PE was calculated to investigate significant differences in model predictions.

Based on the assumption that the true PIP-concentration (either observed or predicted concentration) is uncertain, a modified version of the Sheiner and Beal formula was employed [13,22–24]:

$$PE_{ij} = \frac{c_{pred,i,j} - c_{obs,i,j}}{(c_{pred,i,j} + c_{obs,i,j})/2} \quad (3)$$

$$MPE\ [\%] = median\left(\left\{PE_{1,1},\ \ldots,\ PE_{i,j}\right\}\right) \cdot 100 \quad (4)$$

$$MAPE\ [\%] = median\left(\left\{\ |PE_{1,1}|,\ |\ldots|,\ |PE_{i,j}|\right\}\right) \cdot 100 \quad (5)$$

PE : prediction error; MPE : median prediction error;
MAPE : median absolute prediction error c : concentration;
obs : observed; pred : predicted; $_i$: individual patient;
$_j$: observation

2.6. Clinical criteria

Since there is no clearly defined target range for PIP, percentage of TDM-samples within ±20%, ±30%, ±40%, ±50%-PE ranges (p20, p30, p40, p50, respectively) were employed for evaluation. The expected target attainment corresponds to p50 and was calculated based on an exemplary target concentration of 64 mg/L (4x the epidemiological cutoff (ECOFF) for *Pseudomonas aeruginosa* according to the European Committee on Antimicrobial Susceptibility Testing (EUCAST)) and an absolute target range of 32 mg/L to 96 mg/L (2-6x ECOFF) [6,25,26].

2.7. Time after TDM

TDM-samples were grouped and analysed depending on the time after the integration of TDM-data (TaTDM) (24h, 48h, 72h, 96h, >120h (binned: ±12h, respectively)). This analysis was conducted using data from study 1 and 2 (continuous infusion) only, comprising a total of 489 patients with 1306/817 predicted TDM-samples for B1/B2 respectively. In these studies, TDM-samples were collected over observation periods that exceeded 7 d, which was not the case in the other datasets.

2.8. Software

The models were previously encoded and the predictions were executed within NONMEM®7.4 utilising Perl Speaks NONMEM® and Pirana Version 2.9.9. All numerical and graphical analyses were performed using R®4.2.3/R Studio® 2023.03.0.

3. Results

3.1. Weighting schemes

Several differences were observed when comparing the two different weighting schemes SSE and OFV with overall advantages for utilising OFV. When evaluating model fit using SSE, few models received considerably higher weighting compared to the use of OFV. This implies that in case of MAA, these particular models exert a dominant influence on predictions owing to the weighting within the averaging process (e.g. $MSA\_SSE_{all\ models}$: 2 influential models vs. $MSA\_OFV_{all\ models}$: 17 influential models). When employing SSE for model selection, only one model emerged as the preferred. In contrast, the use of OFV led to a more balanced selection of models, taking into account a broader spectrum of model performances (e.g. $MSA\_SSE_{all\ models}$, the Udy-model was selected in 530/561 patients vs. $MSA\_OFV_{all\ models}$ the Udy-model was selected in 149/561 patients). In terms of predictive performance, there was no significant difference in precision observed between the utilisation of SSE or OFV (95% CI were overlapping). In the $MAA_{all\ models}$, OFV exhibited a slight advantage ($MPE_{OFV/SSE}$ 0.39/4.24%; $MAPE_{OFV/SSE}$ 23.3/23.7%), whereas in the MSA, SSE performed slightly better ($MPE_{OFV/SSE}$ 10.0/-2.16%; $MAPE_{OFV/SSE}$ 30.6/27.9%). However, the predictive performance of MSA using SSE was comparable to that of the Udy-model, which was selected most frequently. Subsequently, for all further analyses, OFV was employed as weighting scheme (for further illustration see Supplements S2).

3.2. Predictive performance

3.2.1. Statistical criteria

The predictive performance of single models varied more widely compared to MSA and MAA (single models/ MSA/ MAA, B1: MPE -77.3-72.3%/ -4.54-10.0%/ -1.60-3.19%, MAPE 26.2-77.7%/ 28.4-30.6%/ 23.3-24.5%, B2: MPE -55.8-65.3%/ 3.69-8.0%/ 3.70-5.61%, MAPE 23.7-67.2%/ 22.3-25.0%/ 22.6-23.4%). Overall, MAA demonstrated significantly greater precision compared to MSA and most single models for B1, as evidenced by the non-overlapping 95% confidence intervals. Only the models by Udy et al. [12] and Fillâtre et al. [27] exhibited similar precision, with a MAPE of less than 27.0%. Although the 95% confidence intervals for MPE overlapped substantially, $MAA_{all\ models}$ was the only approach where the 95% confidence interval of MPE included 0. Furthermore, $MAA_{all\ models}$ exhibited significantly greater precision than all other approaches (B1_MAA MPE: 0.39%; MAPE: 23.3% (95% CI: 22.1-24.3%)). Integrating the second sample improved precision in most single models and MSA, resulting in comparable precision between MAA and MSA, as well as for half of the 24 single models. Only the best single model by Kim et al. showed no bias for B2 (95% CI for MPE included 0) (Table 1, Fig. 2, Supplements S3).

3.2.2. Clinical criteria

When using the $MAA_{all\ models}$, the integration of a single sample ($MAA_{all\ models}$ B1) led to as many samples within the error margin (p20-p50) as the best single model when including two samples (e.g. p30 B1: $MAA_{all\ models}$ vs. B2: Kim 59.5% vs. 59.6%, respectively). This corresponded to an absolute improvement of up to 6.3% more predicted concentrations falling within the calculated error margin for B1 compared to Kim-model [10] (B1: $MAA_{all\ models}$ vs. Kim p20: +6.3%, p30: +4.9%; p40: +2.7% p50: +1.9% samples within the error margin Supplements S4).

3.2.3. Time after TDM

Overall, the precision and accuracy of all approaches were better when the sampling of the integrated sample(s) occurred no

Table 1
Predictive performance of the best single models vs. the different scenarios of the multi-model approaches.

Model/Scenario	MPE (95% CI) [%]	MAPE (95% CI) [%]	p30 [%]	p50 [%]
Bayesian 1				
Kim	-5.85 (-7.72 – -3.88)	27.3 (26.1 – 28.3)	54.6	**77.6**
Klastrup	-7.71 (-9.49 – 5.99)	28.0 (26.7 – 29.0)	53.2	73.5
Udy	**-3.61 (-5.01 – -1.81)**	**26.2 (25.0 – 27.5)**	**55.0**	75.2
$MSA_{all\ models}$	10.0 (8.13 – 11.8)	30.6 (28.9 – 32.2)	49.4	71.0
$MSA_{best\ models}$	-4.54 (-6.11 – -2.98)	28.4 (27.2 – 29.5)	**52.6**	**75.1**
$MSA_{diff.\ models}$	**2.30 (0.804 – 4.42)**	**28.4 (27.3 – 29.9)**	52.1	74.6
$MAA_{all\ models}$	**0.39 (-1.05 – 1.68)**	**23.3 (22.1 – 24.3)**	**59.5**	**79.5**
$MAA_{best\ models}$	-1.60 (-3.23 – -0.316)	24.5 (23.3 – 26.0)	57.9	77.4
$MAA_{diff.\ models}$	3.19 (1.25 – 4.78)	24.5 (23.0 – 25.8)	58.0	78.6
Bayesian 2				
Kim	**-0.92 (-2.78 – 0.987)**	**23.7 (22.5 – 24.9)**	**59.6**	**80.6**
Klastrup	6.65 (4.63 – 8.70)	28.4 (27.1 – 30.0)	52.0	73.9
Udy	7.50 (6.03 – 9.09)	23.8 (22.5 – 25.1)	58.5	76.3
$MSA_{all\ models}$	8.0 (6.17 – 9.32)	24.9 (23.8 – 26.1)	57.1	77.2
$MSA_{best\ models}$	**3.69 (1.84 – 5.1)**	**22.3 (21.3 – 23.7)**	**61.4**	**80.3**
$MSA_{diff.\ models}$	4.32 (2.3 – 6.07)	25.0 (24.0 – 26.1)	57.9	78.8
$MAA_{all\ models}$	5.61 (4.14 – 7.42)	23.4 (22.3 – 24.5)	60.3	**79.9**
$MAA_{best\ models}$	**3.70 (2.12 – 5.47)**	**22.6 (21.7 – 23.8)**	**61.6**	79.4
$MAA_{diff.\ models}$	3.85 (2.36 – 5.44)	23.1 (22.2 – 24.2)	60.4	**79.9**

Abbreviations: MPE: median prediction error; MAPE: median absolute prediction error; CI: confidence interval; p30/p50: percentage of TDM-samples within ±30%, ±50%-prediction error-ranges for the complete dataset; MSA: model selection algorithm; MAA: model averaging algorithm; all: all models; best: best models; diff.: different models.

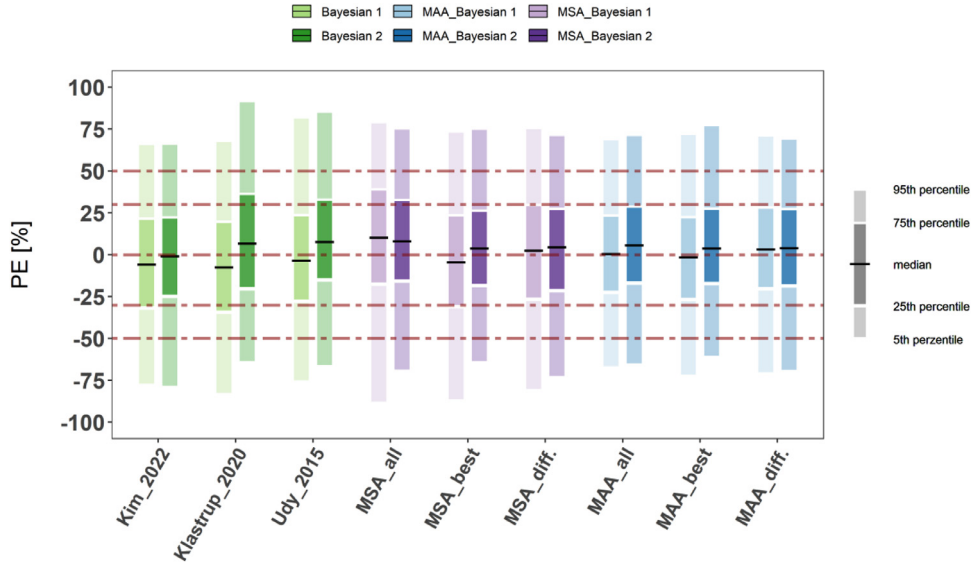

Fig. 2. Prediction error (PE) of the predicted versus the observed piperacillin concentrations of the best models in the single-model approach and the three preselection-scenarios of the model selection algorithm (MSA) or model averaging algorithm (MAA). Bayesian 1 (B1): Prediction of concentrations integrating the first TDM-sample per patient; Bayesian 2 (B2): Prediction of concentrations integrating the first and second TDM-samples per patient; selected models for MSA/MAA: all: all models; best: best models; diff.: different models. TDM: therapeutic drug monitoring. Red dashed lines represent a PE of -50%, -30%, 0%, 30%, 50%.

longer than 24 h ago. The most pronounced difference in predictive performance, with multi-model approach superiority, was observed when integrating one TDM-sample within 24h TaTDM (median MPE/MAPE: single-model approach vs. multi-model approach -6.79/ 22.0% vs. -1.82/ 17.7%, respectively) (Supplements S5 and S6). A time dependence in expected target attainment (p50) was also evident in all approaches with an expected target attainment of over 80% up to >90% for TDM-samples within 24h TaTDM using a multi-model approach or one of the best single models (Table 2, Figure 3). The expected target attainment across all scenarios, in best single-models and multi-model approaches, declined from 24h to 48h TaTDM ranging between -6% to -10% for B1. Despite this decrease, the expected target attainment remained consistently above 70% over the course of >120h TaTDM (Supplements S7).

4. Discussion

The standard dosage of PIP leads to adequate drug concentrations in less than one-fifth of patients with sepsis [2]. The previously published external evaluation of 24 PopPK models by Greppmair et al. in 2023 demonstrated that MIPD is a promising approach for individualising PIP dosing in critically ill patients and may result in an increased target attainment [13]. This study advocates MIPD using MAA to facilitate clinical implementation, while

Table 2
Predictive performance of the best single models vs. the different scenarios of the multi-model approaches within 24 h after the last integrated TDM sample.

Model/Scenario	MPE (95% CI) [%]	MAPE (95% CI) [%]	p30 [%]	p50 [%]
Bayesian 1				
Kim	-6.57 (-10.6–2.58)	21.4 (18.0-24.2)	64.0	87.3
Klastrup	-5.07 (-8.18–2.29)	**16.5 (14.1-18.4)**	**73.6**	88.0
Udy	**3.08 (-1.17-6.46)**	17.4 (15.0-20.0)	71.2	86.3
$MSA_{all\ models}$	**-1.21 (-4.91-1.91)**	18.3 (15.7-22.0)	70.2	**88.7**
$MSA_{best\ models}$	-1.24 (-3.77-2.98)	**17.1 (15.0-21.4)**	**71.2**	87.7
$MSA_{diff.\ models}$	-3.04 (-8.35-1.89)	20.7 (17.8-23.4)	69.5	88.0
$MAA_{all\ models}$	-2.70 (-5.53-0.087)	**15.5 (13.3-18.5)**	72.9	89.0
$MAA_{best\ models}$	**-0.17 (-2.94-2.58)**	15.7 (14.0-19.1)	**73.3**	89.0
$MAA_{diff.\ models}$	-2.40 (-5.57–0.106)	19.0 (15.4-21.4)	70.9	**89.7**
Bayesian 2				
Kim	**1.49 (-4.72-7.57)**	19.7 (17.0-24.6)	**68.2**	**91.1**
Klastrup	1.72 (-2.78-8.44)	**19.2 (15.3-22.9)**	67.8	90.2
Udy	12.6 (7.86-17.8)	22.3 (18.2-26.5)	61.2	83.6
$MSA_{all\ models}$	8.83 (4.45-12.8)	20.7 (17.0-24.0)	65.0	87.4
$MSA_{best\ models}$	7.85 (1.80-12.4)	21.1 (19.0-23.4)	66.4	88.8
$MSA_{diff.\ models}$	**6.94 (-0.15-10.8)**	**20.0 (17.0-23.7)**	**67.8**	**89.7**
$MAA_{all\ models}$	6.95 (0.29-11.4)	20.2 (17.5-23.4)	68.2	87.9
$MAA_{best\ models}$	7.29 (3.07-11.6)	20.1 (17.2-23.4)	66.8	87.9
$MAA_{diff.\ models}$	**3.24 (-0.41-9.57)**	**19.7 (17.0-22.5)**	**66.8**	**88.3**

Abbreviations: TaTDM: Time after therapeutic drug monitoring; MPE: median prediction error; MAPE: median absolute prediction error; CI: confidence interval; p30/p50: percentage of TDM-samples within ±30%, ±50%-prediction error-ranges within 24h TaTDM; MSA: model selection algorithm; MAA: model averaging algorithm; all: all models; best: best models; diff. different models.

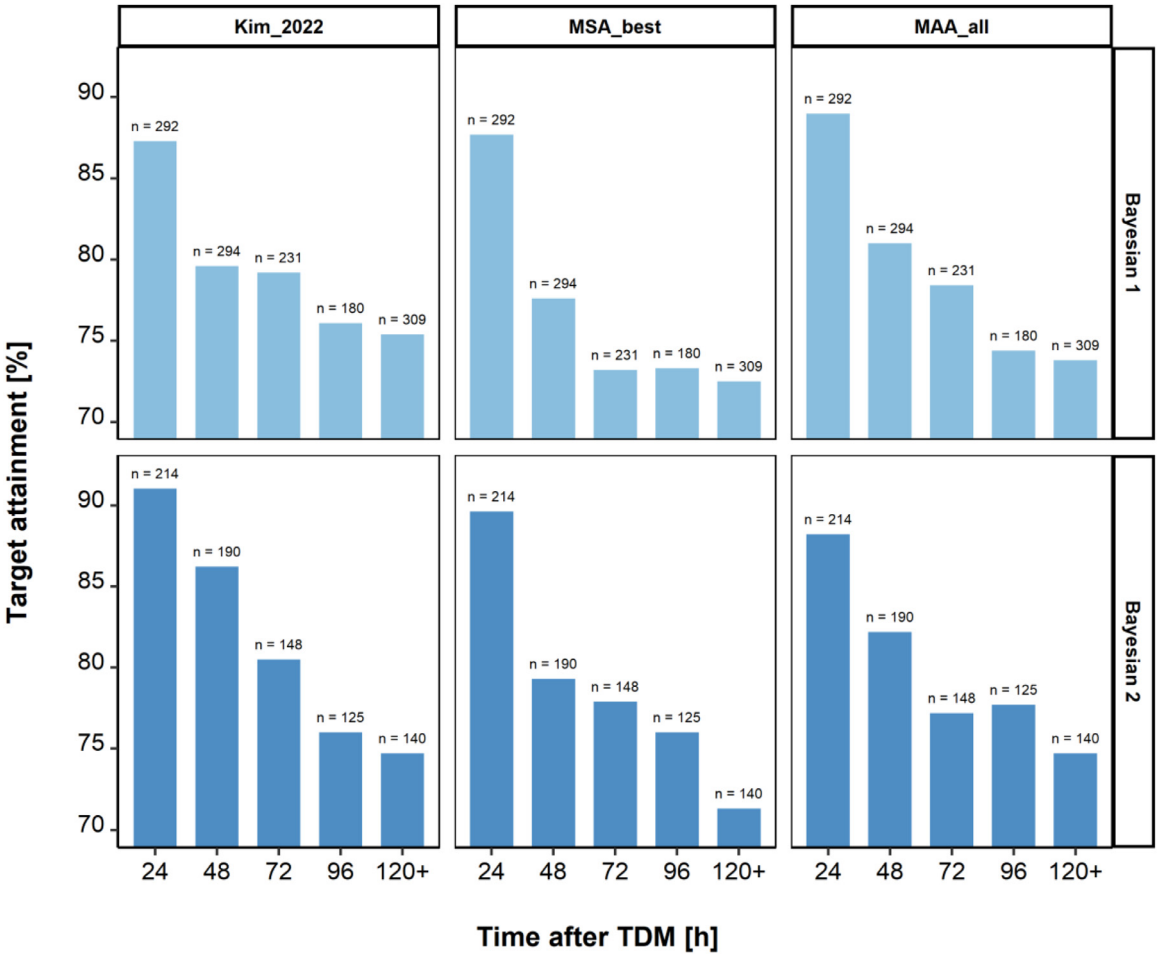

Fig. 3. Expected target attainment (p50) [%] versus time interval [h] since the last integrated therapeutic drug monitoring (TDM) concentration for Bayesian forecasting of the best single model and the best scenario of the model selection algorithm (MSA) or model averaging algorithm (MAA). Bayesian 1 (B1): Prediction of concentrations integrating the first TDM sample per patient; Bayesian 2 (B2): Prediction of concentrations integrating the first and second TDM samples per patient; selected models for MSA/MAA: all: all models; best: best models. TDM: therapeutic drug monitoring.

maintaining predictive performance and expected target attainment.

Our evaluation of the multi-model approaches MAA and MSA included the determination of an optimal weighting scheme, the identification of appropriate candidate models, and will imminently be integrated into the open-access software platform TDMx® for practical application [28].

Overall, the OFV yielded a more appropriate weighting scheme compared to SSE, even though the results regarding predictive performance were comparable. The model fit assessment in SSE relies solely on the absolute discrepancy between predicted and observed concentrations. Therefore, the selection of the most probable individual parameters with a small residual error is primarily driven by the concentration and hardly influenced by the model. Consequently, models with an extremely small residual errors, such as Udy et al. [12], were favoured in the MSA without a particular emphasis on improving future predictions compared to other models due to overfitting. Utilising the MSA when a single model predominantly prevails is not substantively justifiable and beneficial in multi-model approaches, yielding comparable predictive performance compared to Udy et al. [12].

In MAA the fractional individual weight becomes very small due to the squaring operation and exponential function. Consequently, certain models had minimal influence on the analysis when using MAA. Hence, when the SSE was employed as the evaluation criterion, neither MAA nor MSA confers any discernible advantage. Conversely, when the OFV was employed, it incorporates not only absolute discrepancy between predicted and observed concentrations but also prior information into the evaluation of model fit. This resulted in the selection of diverse models within MSA, and the quantification of the impact of various models on predictions within MAA became more measurable. The use of OFV as a weighting scheme was also recommended as the most suitable by Uster et al. [14] investigating MIPD in vancomycin and by Kantasiripitak et al. [29] investigating MIPD during infliximab therapy in adults. However the SSE proved superior in the MIPD analysis of infliximab in children, which can be attributed either to the model structure or the lower target values [30].

Besides the weighting scheme, another challenge posed by a multi-model approach is the selection of candidate models. Particularly in the case of MAA, no significant differences in predictive performance were observed between the scenarios tested. This implies that this approach is highly robust concerning the selection of candidate models. In neither of the previously published multi-model publications by Uster et al. [14], Kantasiripitak et al. [29] Kantasiripitak et al. [30] three different preselection-scenarios were tested and compared to the single-model approach. In the study by Uster et al. [14], six candidate models were selected with different model populations, and in both studies by Kantasiripitak et al. [29,30], all models were selected as candidate models. Kantasiripitak et al. [29] also evaluated the predictive performance with 4 best model candidates. Furthermore, Uster et al. [14] and Kantasiripitak et al. [29] studies confirmed that reducing the number of candidate models had only a minor impact on predictive performance. In summary, all of these studies have concluded that the multi-model approach, especially MAA, provided a clear advantage. Our study, comprising 561 patients and over 3600 samples, is, to our knowledge, the most comprehensive investigation to evaluate the multi-model approach and confirmed these results.

In our evaluation, the most notable disparity between the approaches was apparent in B1. MAA yielded predictive performance slightly superior to the best-performing single model, which also aligns with the three above-mentioned multi-model approach-publications [14,29,30]. Even when including more TDM-samples, the predictive performance of the multi-model approach, MAA in particular, was generally more consistent and comparable or even superior to the single-model approach. It is noteworthy that the comparison to the best-performing is based on the best-case scenario in the single-model approach. The disparities in predictive performance among single models were considerably more pronounced than those observed among the preselection-scenarios within the multi-model approaches. Therefore, the selection of candidate models and a comprehensive external evaluation with subgroup analyses is of higher importance in the single-model approach compared to MAA or MSA.

Including a second TDM-sample enhanced predictive performance for B2 in most models using the single-model approach since it allows a more precise quantification of the variability in an individual patient. However, the MAA performed well even with a single sample, surpassing both MSA and single- model approaches. This can be explained by the approach itself. MSA operates on an "all-or-nothing" principle. After a model is selected, its prediction is used exclusively, with no further influence from other models on predictive performance. In contrast, MAA does not adhere to this principle, as predictions are calculated proportionally through a weighting scheme. In MAA a high weighting carries significant influence on the prediction, yet it does not lead to a distinct model choice. By calculating model predictions based on the weighting scheme, the aim is to achieve an optimal forecast. This assumption was supported by the results of MAA for B1.

However, with the inclusion of a second sample, this advantage in predictions was no longer evident. The second sample increased the probability of selecting the most appropriate model in the context of MSA. In contrast, during the MAA, the influence of predictions from inappropriate models may prevail.

Since different patient subgroups benefit from manual model selection to achieve the most reliable and efficient dosing recommendations, different candidate models were implemented into the software tool TDMx® [13]. The decision-making process in clinical routines is made by health-care professionals and becomes especially challenging when a patient aligns with multiple model populations, such as continuous administration of PIP and continous kidney replacement therapy. Inappropriate model selection may lead to inaccurate dosage calculations. Conversely, MSA and MAA significantly mitigates this probability in clinical use. Additionally, it cannot be ruled out that the multi-model approach may confer a particular advantage, especially for highly specific patient profiles, as the model selection or model averaging provides more flexibility for adaptation.

Unfortunately, there are currently no uniform validation criteria for model evaluation in MIPD and a wider set of diagnostic tools for model qualification for MIPD may still be needed [31]. We were able to demonstrate that regardless of the approach, adhering to a TaTDM of 24 h led to a significantly higher expected target attainment rate, reaching about 90%. This was likely attributed to the pharmacokinetic variability over time observed in critically ill patients, where clinical conditions can change rapidly. Nevertheless, both statistical and clinical criteria were able to demonstrate that the risk for an incorrect choice in candidate models remains lower in the context of MAA and MSA compared to a single-model approach and the use of a regularly updated TDM-sample is advantageous.

MIPD/TDM is a topic that has been poorly investigated in clinical trials and the results of the studies conducted so far are highly controversial. A recently published systematic review and meta-analysis by Codina et al. [32] reported a decrease in mortality, albeit without reaching statistical significance. This observation was associated with a higher level of target attainment. The review included two large-scale randomised controlled trials by Hagel et al. [2] and Ewoldt et al. [33] with 254 and 388 patients, respectively. In both studies, the superiority of individualised antibiotic therapy (TDM- Hagel et al. [2] or MIPD- Ewoldt et al. [33]) was not ob-

served, however, the reported target attainment was under 60%. Our analysis suggests that the application of MIPD using a multi-model approach may further enhance target attainment, potentially leading to outcome benefits (e.g. the Anderson-model used in Ewoldt et al. [33], resulted in an expected target attainment of 66.8% in our evaluation, compared to 79.5% achieved by the $MAA_{\text{all models}}$). In particular, studies that employ MIPD with evaluated models and approaches in clinical settings, testing and linking them to clinical outcomes, are warranted.

Some limitations of our study should be mentioned: First, the expected target attainment served as an illustration for clinical practice-based assumptions. Besides, the evaluation over time was only assessed on patients with continuous infusion (since only these datasets including longer monitoring periods and multiple samples). But it is reasonable to assume that these findings also apply to intermittent infusion. Second, since MAA and MSA rely on the principle of Bayes, our results are only applicable when at least one TDM-sample is available. Initial model selection still poses a challenge demanding research in the future [34]. Furthermore, there is potential for additional research by assessing the efficacy of combining Bayesian model averaging with the "synthetic model combination" [34]. In this process, a machine learning approach is employed to determine the best model ensemble, even without TDM-samples, upon which the initial dosage is calculated. Third, even though we had a very large dataset to evaluate the multi-model approaches there remained a residual risk for a bias in the predictive performance. Nonetheless, our results were consistent with previous studies [14,29,30] and making for example, "overtuning" unlikely. Fourth, candidate models for the preselection scenarios were chosen randomly, thereby reducing the likelihood of "overtuning". Fifth, model averaging or selection became necessary due to the absence of a perfect model, emphasising the value of improved models. Additionally, developing new models based on diverse datasets, covering various subgroups, and enhancing the understanding of PIP pharmacokinetics in critically ill patients (e.g., using Physiologically-based pharmacokinetic models) could be a promising approach.

In conclusion, the use of a MAA resulted in predictive performance comparable or slightly better than the best single models. In addition, the absence of the need for model selection when employing a TDM-sample represents a crucial advancement in the implementation of MIPD. Since the predictive performance was less affected by the candidate models, the need for a comprehensive external evaluation is reduced. Therefore, our study is an important step towards the clinical implementation of individualised dosing of PIP in critically ill patients.

Funding

This work was supported by the German Research Foundation (grant number: 504995322).

Ethical approval

Ethical approval for this study was waived by the local institutional review board (No.: 21–1162 KB).

Sequence information

Not applicable.

Declaration of competing interests

None declared.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ijantimicag.2024.107305.

References

[1] Liebchen U, Paal M, Scharf C, Schroeder I, Grabein B, Zander J, et al. The ONTAI study – a survey on antimicrobial dosing and the practice of therapeutic drug monitoring in German intensive care units. J Crit Care 2020;60:260–6. doi:10.1016/j.jcrc.2020.08.027.
[2] Hagel S, Bach F, Brenner T, Bracht H, Brinkmann A, Annecke T, et al. Effect of therapeutic drug monitoring-based dose optimisation of piperacillin/tazobactam on sepsis-related organ dysfunction in patients with sepsis: a randomized controlled trial. Intensive Care Med 2022;48:311–21. doi:10.1007/s00134-021-06609-6.
[3] Imani S, Buscher H, Marriott D, Gentili S, Sandaradura I. Too much of a good thing: a retrospective study of β-lactam concentration-toxicity relationships. J Antimicrob Chemother 2017;72:2891–7. doi:10.1093/jac/dkx209.
[4] Scharf C, Liebchen U, Paal M, Taubert M, Vogeser M, Irlbeck M, et al. The higher the better? Defining the optimal beta-lactam target for critically ill patients to reach infection resolution and improve outcome. J Intensive Care 2020;8:1–10. doi:10.1186/s40560-020-00504-w.
[5] Cusumano JA, Klinker KP, Huttner A, Luther MK, Roberts JA, LaPlante KL. Towards precision medicine: Therapeutic drug monitoring–guided dosing of vancomycin and β-lactam antibiotics to maximize effectiveness and minimize toxicity. Am J Health-Syst Pharmacy 2020;77:1104–12. doi:10.1093/ajhp/zxaa128.
[6] Guilhaumou R, Benaboud S, Bennis Y, Dahyot-Fizelier C, Dailly E, Gandia P, et al. Optimization of the treatment with beta-lactam antibiotics in critically ill patients-guidelines from the French Society of Pharmacology and Therapeutics (Société Française de Pharmacologie et Thérapeutique-SFPT) and the French Society of Anaesthesia and Intensive Care Medicine (Société Française d'Anesthésie et Réanimation-SFAR). Crit Care 2019;23:104. doi:10.1186/s13054-019-2378-9.
[7] Wicha SG, Märtson A-G, Nielsen EI, Koch BCP, Friberg LE, Alffenaar J-W, et al. From therapeutic drug monitoring to model-informed precision dosing for antibiotics. Clin Pharmacol Ther 2021;109:928–41. doi:10.1002/cpt.2202.
[8] Cotta MO, Lipman J, De Waele J. Advancing precision-based antimicrobial dosing in critically ill patients. Intensive Care Med 2023;49:324–6. doi:10.1007/s00134-022-06969-7.
[9] Bulman ZP, Wicha SG, Nielsen EI, Lenhard JR, Nation RL, Theuretzbacher U, et al. Research priorities towards precision antibiotic therapy to improve patient care. The Lancet Microbe 2022;3:e795–802. doi:10.1016/S2666-5247(22)00121-5.
[10] Kim YK, Kim HS, Park S, Kim H, Lee SH, Lee D-H. Population pharmacokinetics of piperacillin/tazobactam in critically ill Korean patients and the effects of extracorporeal membrane oxygenation. J Antimicrob Chemother 2022;77:1353–64. doi:10.1093/jac/dkac059.
[11] Klastrup V, Thorsted A, Storgaard M, Christensen S, Friberg LE, Öbrink-Hansen K. Population pharmacokinetics of piperacillin following continuous infusion in critically ill patients and impact of renal function on target attainment. Antimicrob Agents Chemother 2020;64:e02556-19. doi:10.1128/aac.02556-19.
[12] Udy AA, Lipman J, Jarrett P, Klein K, Wallis SC, Patel K, et al. Are standard doses of piperacillin sufficient for critically ill patients with augmented creatinine clearance? Crit Care 2015;19:28. doi:10.1186/s13054-015-0750-y.
[13] Greppmair S, Brinkmann A, Roehr A, Frey O, Hagel S, Dorn C, et al. Towards model-informed precision dosing of piperacillin: multicenter systematic external evaluation of pharmacokinetic models in critically ill adults with a focus on Bayesian forecasting. Intens Care Med 2023;49:966–76. doi:10.1007/s00134-023-07154-0.
[14] Uster D, Stocker SL, Carland JE, Brett J, Marriott DJE, Day RO, et al. A model averaging/selection approach improves the predictive performance of model-informed precision dosing: vancomycin as a case study. Clin Pharmacol Therap 2021;109:175–83. doi:10.1002/cpt.2065.
[15] Asín-Prieto E, Rodríguez-Gascón A, Trocóniz IF, Soraluce A, Maynar J, Sánchez-Izquierdo JÁ, et al. Population pharmacokinetics of piperacillin and tazobactam in critically ill patients undergoing continuous renal replacement therapy: application to pharmacokinetic/pharmacodynamic analysis. J Antimicrob Chemother 2014;69:180–9. doi:10.1093/jac/dkt304.
[16] Roberts DM, Liu X, Roberts JA, Nair P, Cole L, Roberts MS, et al. A multicenter study on the effect of continuous hemodiafiltration intensity on antibiotic pharmacokinetics. Crit Care 2015;19:84. doi:10.1186/s13054-015-0818-8.
[17] Tamme K, Oselin K, Kipper K, Tasa T, Metsvaht T, Karjagin J, et al. Pharmacokinetics and pharmacodynamics of piperacillin/tazobactam during high volume haemodiafiltration in patients with septic shock. Acta Anaesthesiologica Scandinavica 2016;60:230–40. doi:10.1111/aas.12629.
[18] Chen R, Qian Q, Sun M, Qian C, Zou S, Wang M, et al. Population Pharmacokinetics and Pharmacodynamics of Piperacillin/Tazobactam in Patients with Nosocomial Infections. Eur J Drug Metab Pharmacokinet 2016;41:363–72. doi:10.1007/s13318-015-0276-3.
[19] Bue M, Sou T, Okkels ASL, Hanberg P, Thorsted A, Friberg LE, et al. Population pharmacokinetics of piperacillin in plasma and subcutaneous tissue in patients

on continuous renal replacement therapy. Int J Infect Dis 2020;92:133–40. doi:10.1016/j.ijid.2020.01.010.
[20] Sukarnjanaset W, Jaruratanasirikul S, Wattanavijitkul T. Population pharmacokinetics and pharmacodynamics of piperacillin in critically ill patients during the early phase of sepsis. J Pharmacokinet Pharmacodyn 2019;46:251–61. doi:10.1007/s10928-019-09633-8.
[21] Hahn J, Min KL, Kang S, Yang S, Park MS, Wi J, et al. Population pharmacokinetics and dosing optimization of piperacillin-tazobactam in critically ill patients on extracorporeal membrane oxygenation and the influence of concomitant renal replacement therapy. Microbiol Spectrum 2021;9:e0063321. doi:10.1128/Spectrum.00633-21.
[22] Broeker A, Nardecchia M, Klinker KP, Derendorf H, Day RO, Marriott DJ, et al. Towards precision dosing of vancomycin: a systematic evaluation of pharmacometric models for Bayesian forecasting. Clin Microbiol Infect 2019;25:12861–7. doi:10.1016/j.cmi.2019.02.029.
[23] Sheiner LB, Beal SL. Some suggestions for measuring predictive performance. J Pharmacokinetics Biopharm 1981;9:503–12. doi:10.1007/BF01060893.
[24] Schatz LM, Brinkmann A, Röhr A, Frey O, Greppmair S, Weinelt F, et al. Systematic evaluation of pharmacokinetic models for model-informed precision dosing of meropenem in critically ill patients undergoing continuous renal replacement therapy. Antimicrob Agents Chemother 2023;67:0010423. doi:10.1128/aac.00104-23.
[25] Roberts JA, Paul SK, Akova M, Bassetti M, Waele JJ, Dimopoulos G, et al. DALI: defining antibiotic levels in intensive care unit patients: are current β-lactam antibiotic doses sufficient for critically ill patients? Clin Infect Dis 2014;58:1072–83. doi:10.1093/cid/ciu027.
[26] The European Committee on Antimicrobial Susceptibility Testing - EUCAST 2022 n.d. https://mic.eucast.org/search/?search[method]=mic&search[antibiotic]=-1&search[species]=411&search[disk_content]=-1&search[limit]=50. Accessed on 10 September, 2024.
[27] Fillâtre P, Lemaitre F, Nesseler N, Schmidt M, Besset S, Launey Y, et al. Impact of extracorporeal membrane oxygenation (ECMO) support on piperacillin exposure in septic patients: a case–control study. J Antimicrob Chemother 2021;76:1242–9. doi:10.1093/jac/dkab031.
[28] Wicha SG, Kees MG, Solms A, Minichmayr IK, Kratzer A, Kloft C. TDMx: a novel web-based open-access support tool for optimising antimicrobial dosing regimens in clinical routine. Int J Antimicrob Agents 2015;45:442–4. doi:10.1016/j.ijantimicag.2014.12.010.
[29] Kantasiripitak W, Outtier A, Wicha SG, Kensert A, Wang Z, Sabino J, et al. Multi-model averaging improves the performance of model-guided infliximab dosing in patients with inflammatory bowel diseases. CPT Pharmacometrics Syst Pharmacol 2022;11:1045–59. doi:10.1002/psp4.12813.
[30] Kantasiripitak W, Wicha SG, Thomas D, Hoffman I, Ferrante M, Vermeire S, et al. A model-based tool for guiding infliximab induction dosing to maximize long-term deep remission in children with inflammatory bowel diseases. J Crohns Colitis 2023;17:896–908. doi:10.1093/ecco-jcc/jjad009.
[31] Keizer RJ, ter Heine R, Frymoyer A, Lesko LJ, Mangat R, Goswami S. Model-informed precision dosing at the bedside: scientific challenges and opportunities. CPT Pharmacometrics Syst Pharmacol 2018;7:785–7. doi:10.1002/psp4.12353.
[32] Codina MS, Bozkir HÖ, Jorda A, Zeitlinger M. Individualised antimicrobial dose optimisation: a systematic review and meta-analysis of randomised controlled trials. Clin Microbiol Infect 2023;29:845–57. doi:10.1016/j.cmi.2023.03.018.
[33] Ewoldt TMJ, Abdulla A, Rietdijk WJR, Muller AE, de Winter BCM, Hunfeld NGM, et al. Model-informed precision dosing of beta-lactam antibiotics and ciprofloxacin in critically ill patients: a multicentre randomised clinical trial. Intensive Care Med 2022;48:1760–71. doi:10.1007/s00134-022-06921-9.
[34] Chan A, Peck R, Gibbs M, van der Schaar M. Synthetic Model combination: a new machine-learning method for pharmacometric model ensembling. CPT: Pharmacometrics Syst Pharmacol 2023;12:953–62. doi:10.1002/psp4.12965.