Clinical Pharmacokinetics
https://doi.org/10.1007/s40262-023-01233-7

CURRENT OPINION

External Evaluation of Population Pharmacokinetic Models for Precision Dosing: Current State and Knowledge Gaps

Mehdi El Hassani[1,2] · **Amélie Marsot**[1,2]

Accepted: 27 February 2023

Abstract
Predicting drug exposures using population pharmacokinetic models through Bayesian forecasting software can improve individual pharmacokinetic/pharmacodynamic target attainment. However, selecting the most adapted model to be used is challenging due to the lack of guidance on how to design and interpret external evaluation studies. The confusion around the choice of statistical metrics and acceptability criteria emphasises the need for further research to fill this methodological gap as there is an urgent need for the development of standards and guidelines for external evaluation studies. Herein we discuss the scientific challenges faced by pharmacometric researchers and opportunities for future research with a focus on antibiotics.

Key Points

An abundance of population models is generally available for a single drug, particularly for antibiotics.

The growing clinical awareness of model-informed precision dosing has led to several population pharmacokinetic model external evaluation studies.

Choosing the most appropriate model to be used for Bayesian forecasting is challenging given the lack of guidance on how to appropriately design and interpret an external evaluation study.

1 Introduction

Model-informed precision dosing (MIPD) is increasingly being used to support clinical decision-making for dose individualisation. It typically involves the combination of population pharmacokinetic/pharmacodynamic (PK/PD) models and knowledge about the individual patient through Bayesian forecasting software. It can either be used a priori (solely based on patient covariate data) to determine the most optimal starting dose, or a posteriori (based on covariate data in combination with one or more observed therapeutic drug monitoring [TDM] levels) to yield a Bayesian posterior parameter distribution necessary to predict the next dose that will result in optimal drug exposure [1, 2]. However, most population models are developed with clinical data (either retrospectively collected TDM data or prospective clinical trials) covering specific age groups, body compositions, genotypes, dosing regimens, indications, and comorbidities. One must therefore remain cautious when selecting a model to make predictions in an independent patient population because its predictive performance will determine the robustness of the estimated exposure and the resulting dose recommendations [3]. The development, evaluation, and clinical implementation of population PK models in clinical practice is illustrated in Fig. 1.

Model-informed precision dosing and model transferability are particularly of interest for antibiotics given that the rapid emergence of antimicrobial resistance and the shortage of new antibiotics are one of the biggest threats to health care [4]. These antibiotic-resistant infections significantly increase hospitalisation durations, mortality rates, and costs for the health care system [5, 6]. Globally, an estimated 4.95 million deaths in 2019 were associated with antimicrobial-resistant bacterial infections [7]. The importance of optimising dosing regimens of existing

✉ Mehdi El Hassani
mehdi.el.hassani@umontreal.ca

[1] Faculté de pharmacie, Université de Montréal, 2940 chemin de Polytechnique, Montréal, QC H3T 1J4, Canada

[2] Laboratoire de suivi thérapeutique pharmacologique et pharmacocinétique, Faculté de pharmacie, Université de Montréal, Montréal, Canada

Published online: 01 April 2023

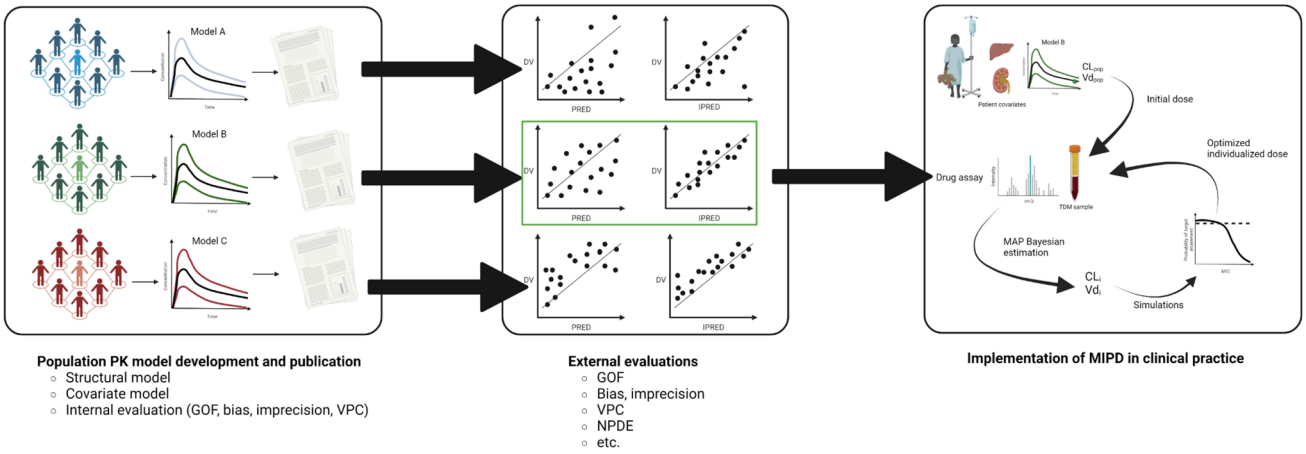

Fig. 1 Development, evaluation, and clinical implementation of population PK models. *CLi* Bayesian posterior individual clearance estimate, *CLpop* population clearance, *DV* dependant variable (observed concentration), *GOF* goodness-of-fit, *IPRED* individual-predicted concentration, *MAP* maximum a posteriori probability, *NPDE* normalised prediction distribution error, *PRED* predicted concentration, *Vdi* Bayesian posterior individual volume of distribution estimate, *Vdpop* population volume of distribution, *VPC* visual predictive check

antibiotics is therefore increasingly being recognised to achieve appropriate target concentrations at the site of infection [8].

Although several examples demonstrate the usefulness of treating individuals rather than a population [2, 9–11], MIPD has not yet become integrated into clinical practice on a large scale [12]. Some barriers towards the implementation of model-based dose individualisation in a clinical setting include high costs for some software licences, the lack of clinically oriented training in Bayesian PK and the lack of acceptance of more complex dosing strategies by prescribers [13]. There may also be some confusion regarding the choice of PK model to be used for MIPD.

Despite the importance of external evaluation studies for assessing the performance of population PK models in new patient populations, there are currently no widely accepted guidelines for conducting such studies. This lack of standardisation poses a challenge for the development and implementation of MIPD, as it makes it difficult to determine the validity of a population PK model in a new patient population. The absence of clear guidelines for conducting external evaluation studies, such as methods for selecting appropriate evaluation criteria or appropriate statistical metrics and their associated thresholds, can limit the accuracy and reliability of model predictions, which in turn hinders the uptake and integration of MIPD in clinical practice. The objective of this manuscript is to provide an overview of the current state of the field of external evaluation studies for population PK models, highlight its limitations, and identify opportunities for future research.

2 Leveraging Existing Population Models Through Evaluation

An increasing number of antibiotic population models have been developed over the past decades resulting in many models for the same drug in the same population, not necessarily adding anything new to our knowledge on the drug's PK (e.g., same reported covariates or similar parameter values [14–17]. Although Duffull and Wright have reported that repeated analyses may be desirable when authors provide new information under a specific set of study conditions, additional work on the structural and covariate model for a well characterised drug is probably unhelpful [18]. This highlights the importance of leveraging existing population models through external evaluations rather than developing new models.

There are several common methods used to evaluate population pharmacokinetic models. Prediction-based methods involve comparing the model's predictions to observed data, while simulation-based methods derive from the concept of the posterior predictive check [19], which states that if a model accurately describes a set of observations, the simulated data under that model should be similar to the observed data [20, 21]. Some common methods for prediction-based evaluation include goodness-of-fit plots or residual errors, while a common method for simulation-based evaluation includes visual predictive checks [22]. These methods are mostly used for internal evaluation of the model, which involves evaluating the model using data from the same population from which the model was derived [23]. However, as we will discuss

in the following section, external evaluation is necessary to ensure the validity and reliability of the model in a different population or clinical setting [24].

3 Lack of Guidelines in the Design of External Evaluation Studies

Considering the complex statistical methodology involved in the use of population models in a clinical setting, external evaluation is considered the most stringent method to assess their predictive performance and transferability to other clinical centres [22]. In 1997, Mentré and Ebelin defined "model validation" as the assessment of the predictability of the model and estimates for further inferences [25]. In 2001, Yano, Beal, and Sheiner proposed to use the weaker term "evaluation" rather than "validation" as the authors considered that one cannot truly validate a model [19]. It is common for the external evaluation dataset to include only a limited number of subjects under limited conditions. Therefore, an external evaluation study generally confirms a limited range of external validity [26].

No consensus has been reached on how to appropriately design and interpret an external evaluation study. The section dedicated to "model validation" in the FDA population pharmacokinetics guidance for industry, recently published in 2022, provided no additional insight on external evaluation methodology since the draft guidance published in 1999 [22]. Additionally, some guidelines on reporting population PK studies do not address the topic of external evaluations at all [27]. As a result, significant variability remains in regard to study design, statistical methodology, and reporting [24]. It has been shown that external evaluation of models for antibiotics is only performed for about 7% of published PK models, probably due to the lack of guidance. Of the 32 studies included in the review by Cheng et al. [24], 30 (93.8%) studies used prediction-based diagnostics. The accuracy of models was evaluated by using several commonly used metrics such as prediction error (PE), mean prediction error (MPE), and absolute prediction error (APE), while mean absolute prediction error (MAPE) and root mean squared error (RMSE) were commonly used to assess the precision of models. Fifteen (46.9%) studies used simulation-based diagnostics (visual predictive check [VPC] and normalised prediction distribution error [NPDE]). Eight (25%) studies used Bayesian forecasting) [28].

4 Metrics and Acceptability Thresholds

Although Sheiner and Beal were the first to propose the MPE and the RMSE as metrics to evaluate the predictive performance of a PK model in 1981 [20], prediction-based external evaluation studies were not introduced in the literature until the early 1990s with Short et al. [29] being among the first. The authors externally evaluated the bias (MPE) and imprecision (MAPE) of a propofol PK model developed by Marsh et al. [30]. Although no threshold values were prospectively defined, the model developed by Marsh et al was considered not appropriate in the Short et al study population due to high bias and imprecision (− 18.5% and 24.8%, respectively). The authors then re-estimated the PK parameters and repeated the external evaluation, after which a significant drop in bias was observed. Additional prediction-based metrics have been proposed since for a priori and posteriori dosing, including divergence, wobble, or the proportion of the PE% falling within a certain percent threshold (e.g., F_5, F_{10}, or F_{15}) [24, 31, 32], and the choice of either one for external evaluation remains subjective and arbitrary. A summary of common statistical metrics used in external evaluation studies is presented in Table 1.

Simulation-based diagnostics of PK models appeared in the mid-1990s [33] and have grown in popularity since the early 2000s. A particularly popular evaluation tool is the VPC. In 2005, Holford proposed the VPC as being more objective than standard diagnostic plots (observed vs predicted concentrations along with weighted and unweighted residuals) for model evaluation and has been widely applied to external evaluation studies since then [34–36]. However, a VPC may be uninformative if it is applied to data from clinical practice with large variability in doses and dosing intervals as well as when continuous or many combinations of different categorical covariates are used. In this regard, a variation of the VPC, the prediction-corrected VPC (pcVPC), has been proposed by Bergstrand et al in 2011 [37] to be readily applicable to these types of studies. The method involves dividing all observations in a certain time-bin by the average observed value in that bin [37]. However, a major limitation of this method is that it may not accurately reflect the model's performance if the differences in observations are due to variations in dosing or

Table 1 Common statistical metrics used for external evaluation

Abbreviation	Definition
APE	Absolute prediction error
MAPE	Mean absolute prediction error
MDAPE	Median absolute prediction error
MDPE	Median prediction error
MPE	Mean prediction error
NDPE	Normalised distribution prediction error
pcVPC	Prediction-corrected visual predictive check
PE	Prediction error
RMSE	Root mean squared error
VPC	Visual predictive check

covariate values. In 2006, Mentré and Escolano proposed another simulation-based model evaluation tool, called prediction discrepancy (pd), which takes into account the full predictive distribution of each observation [38]. The authors showed that their method exhibited improved statistical properties rather than weighted residuals, but that multiple observations per subject increased the probability of a type I error for the Kolmogorov-Smirnov normality test. Given that more than one observation is usually available per subject (i.e., observations are correlated within each subject), Brendel et al proposed a decorrelated version of the pd, the NPDE, which accounts for the correlation between multiple observations within an individual [39]. A major advantage of the NDPE is that, unlike VPCs, it provides both visual and numerical outputs.

Model acceptability criteria are not consistent across published studies for antibiotics, which hinders comparability between publications. For instance, reported threshold values for MPE range between $< \pm$ 20–30% and < 30–35% for MAPE [24]. However, authors generally reference publications from distinct pharmacological classes as justification. For example, Zhang et al. [40] externally evaluated a vancomycin population model and considered that a MAPE value of < 35% was acceptable based on the same threshold value reported by Hu et al. [41] who externally evaluated eight tacrolimus population models. Hu et al justified their MAPE threshold value by referencing a cyclosporin external evaluation study published by Mao et al. [42]. The authors stated that a model was arbitrarily considered to be clinically acceptable when the median prediction error (MDPE) $\leq \pm$ 15% and MAPE ≤30%. Metrics proposed by Varvel et al in 1992, such as the MDPE and the median absolute prediction error (MDAPE), are also typically used for external evaluations [43]. The consensus across published studies is that threshold values for model acceptability are between − 20 and 20% for bias and < 30% for imprecision. This was first reported by Miyabe-Nishiwaki et al. [44] and was justified by referencing a chapter of Miller's Anaesthesia [45]. The threshold values were arbitrarily determined by the authors because the textbook states that in many studies evaluating predictive performance of anaesthetic population PK models for target-controlled infusion pumps, MDPE values were generally around 20–30% [45–48]. These thresholds have largely been applied to other drug classes, including antibiotics, to assess model acceptability [49, 50]. As another example, Guo et al. [51] externally evaluated several vancomycin population PK models and stated that they regarded a population PK model as valid for their clinical setting when both the MPE and MDPE were less than 20%. The authors chose this cut-off value because "the risk of unrightfully adjusting the dose based on a concentration above or below the target window while the actual concentration is on target is minimal". The authors gave an example showing that a vancomycin trough being 20% under or over the median target value of 20 mg/L (for a 15–25 mg/L range, paper published before the 2020 guidelines) would still be in the therapeutic range. However, the authors failed to mention that if the actual concentrations were on either end of the therapeutic range, a 20% error could result in an inappropriate trough. Another issue limiting comparability between studies is that some authors report their external evaluation values for bias and imprecision in concentration units rather than percentages, making it difficult to assess whether they are above or below previously reported thresholds [52]. However, in some cases it may be valuable to report actual values for error as well as percent error, particularly when the error does not depend on the true value itself (i.e., additive error). Overall, the lack of a scientific basis to determine whether a model is acceptable for its intended use is problematic and could result in inappropriate dose recommendations.

We believe that the different metrics used, and their acceptability thresholds, should be tailored to the goals of the evaluation, as different models may be appropriate for different purposes, as well as consider any potential external factors that may impact the results of the evaluation.

5 Fit-for-Purpose Evaluations and Factors that Influence Them

Fit-for-purpose external evaluations are generally defined by their clinical needs [24]. Therefore, several aspects of the model should be assessed according to the objective of the evaluation. Most population models have not been developed with the intention of being used for MIPD [53]. In fact, model evaluation is usually guided by the a priori approach to predict an optimal starting dose given that the tools needed for Bayesian forecasting are not typically available [1]. However, in a MIPD setting, a priori predictive performance, depiction of PK variability components of the model using simulation-based diagnostics, and, most importantly, Bayesian forecasting (including bias and imprecision of forecasted concentrations as well as their impact on PK/PD target attainment and resulting dose recommendations) should all be evaluated [53]. However, this not always the case in published studies. For instance, some authors [51, 54] emphasise the importance of Bayesian approaches to optimise TDM and dose individualisation but fail to compute bias or imprecision for individual-predicted concentrations to assess a posteriori predictive performance. Drawing conclusions from a priori predictions when the goal of the evaluation is to use the model for dose individualisation wrongfully assumes that the population used for model development will be exactly the same as the one in which the tool will be implemented

[55]. It is expected that the independent population used for evaluation should be similar to that of the one used for model development in terms of demographic, clinical and drug administration characteristics for the validation to be successful [31]. While several external evaluation studies are available, no data exist regarding the influence of these factors on the predictive performance of a model and whether acceptability thresholds should be adjusted to these factors. More specifically, some aspects of the study design, such as sample size and number of samples per patient, are variables that could influence the results of an external evaluation. Furthermore, pharmacological class and therapeutic index are also factors that may influence the interpretation of an external evaluation study. For example, a 20% bias should not have the same clinical significance for narrow therapeutic index drugs (e.g., vancomycin) than for large therapeutic index drugs (e.g., piperacillin), suggesting that threshold values for the latter could be less conservative. There also are several factors that can influence the variability of cut-off values for the metrics used in external evaluation studies, in addition to the wide/narrow therapeutic range of the drug. One important factor is the coefficient of variation (CV%) of the assay used to measure drug concentrations, as a higher CV% results in greater variability in the measured concentrations. For instance, Graves et al performed a simulation study on the impact of assay variability on PK parameter estimation and found that bias increased with the magnitude of the assay error [56]. Clinical judgement can also play a role in determining the appropriate cut-offs, as the acceptable level of error may vary depending on the specific clinical context and the potential consequences of incorrect dosing. The objective of the evaluation (i.e., a priori or a posteriori dosing) should also be considered and clearly stated in an external evaluation study (i.e., what the authors will do with the validated model). Other variables that can influence the results of an external evaluation include number of samples per patient [57], and the magnitude of inter- and intra-individual variability [58].

Besides the fact that few investigators are fully familiar with the concept, there are some pharmaco-statistical issues that need to be considered when using maximum a posteriori probability (MAP) Bayesian control in a clinical setting. For instance, the choice of prior distribution for the model can have a significant impact on the results, particularly if the patient does not resemble the population from which the PK model was derived. Inaccurate prior assumptions can lead to biased posterior PK estimates, poor prediction accuracy, and inadequate dosing recommendations. However, without clear guidelines on how to evaluate the appropriateness of the PK model for a specific patient, the investigator may not know if the chosen model is appropriate (from either a statistical or a clinical standpoint). In this case, the investigator must be aware of the limitations and know how to evaluate the model so that it is for fit-for-purpose.

While it is still unknown how these external factors influence the results of external evaluation studies, new research is emerging, particularly in the field of machine learning, where new methods and approaches are being developed to help researchers and clinicians select the most appropriate model for their specific needs.

6 Emerging Approaches for Model Selection in MIPD

To overcome some of the challenges associated with model selection and validation, especially when many models are available for a single drug, new approaches have been developed to enable easier implementation of MIPD into clinical practice, including model selection/averaging algorithms and machine learning (boosting algorithms and neural networks).

Although there have been some attempts to use model averaging techniques to aid decision making during Phase II/III dose-finding trials [59–61], only recently has increasing interest in model averaging techniques been applied to Bayesian forecasting methods [13, 62, 63]. Multi-model selection and averaging algorithms have been proposed by Uster et al and incorporated into the TDMx software [64]. Using vancomycin as a case study, the authors derived an algorithm which automates the model selection process and finds the most adapted model for an individual patient among a set of candidate models or averages the predictions of all candidate models proportionally weighted to their retrospective model fit [13]. The authors found that predictive performance was better after model averaging than for the best single model. Similar findings were reported by the same research group in other studies [62, 63].

Machine learning (ML) is playing an increasingly important role in the optimisation of personalised dosing strategies. It has enabled models to analyse and automatically learn from a vast range of electronic health records using advanced statistical and probabilistic techniques without being explicitly programmed [65]. In fact, Brier et al were some of the first authors to apply ML algorithms to PK datasets [66]. The authors compared observed peak and trough gentamicin concentrations with predicted concentrations from a parametric non-linear mixed effects model and artificial neural networks and found that both approaches displayed equivalent accuracy. More recently, several studies have reported using ML approaches (mainly boosting algorithms) to improve TDM performance [67–69]. However, only Lee et al have applied ML to select the most appropriate vancomycin model for an individual patient to be used for Bayesian forecasting [70]. They developed a model classifier

(using learning data from 900,000 simulated patients), which assigns an individual patient to one of nine population PK models that most closely resembles the patient's PK parameters. To assess the performance of their approach, the authors performed an external evaluation using a simulated validation dataset consisting of 4000 virtual patients and predicted the vancomycin area under the concentration-time curve after a single dose and at steady-state (using the estimated clearance) using either a single model or a ML-selected or ML-weighted average model. They found that the non-weighted arithmetic mean of predictions from the nine individual models outperformed the ML approach when only troughs were available in the validation dataset, both in terms of bias and imprecision. However, as the number of samples per patient increased, the ML approach led to better accuracy and precision.

7 Future Research

It is likely that various factors will influence the results of an external evaluation study. As such, we believe that thresholds for defining model acceptability should not be the same for all drugs and study designs. For example, should threshold values be more conservative for models to be used for a posteriori dosing? We have previously proposed that bias and imprecision for individual-predicted tobramycin concentrations should be more conservative (i.e., $-5\% \leq$ MDPE $\leq 5\%$ and MDAPE $\leq 15\%$) if the goal of the evaluation is to use the validated model for Bayesian forecasting [31]. Further research could explore whether bias or imprecision is influenced by the objective of the evaluation. Also, the influence of sample size and number of samples per patient in external evaluation studies should be explored. While no evidence has been previously published to address that question, it is known that increasing sample size improves the predictive performance of a pharmacokinetic model during model development [71]. Future research could explore what the minimum sample size should be for an external evaluation study and whether an increased number of samples per patient could compensate for a reduced sample size. Furthermore, the impact of varying threshold values of prediction-based metrics and the use of "non-validated" models on dosing recommendations is still unknown and warrants investigation. Additionally, there is a need for more prospective validation studies post-implementation of PK models in Bayesian software to evaluate the impact of model-informed precision dosing on patient outcomes and to identify any potential issues or challenges that may arise in clinical practice.

In conclusion, the lack of guidelines on external evaluation studies of population PK models poses a significant challenge for the accurate and reliable use of these models in clinical practice. The confusion around the choice of statistical metrics and acceptability criteria emphasises the need for further research to fill this methodological gap as there is an urgent need for the development of standards and guidelines for external evaluation studies. Addressing these issues will ultimately lead to better informed dosing decisions and improved patient care.

Declarations

Funding Mehdi El Hassani received a scholarship from Université de Montréal, and Amélie Marsot acknowledges support from the Fonds de Recherche du Québec-Santé (FRQS) Research Scholars – Junior 1 (Young Researcher Establishment) Career Scholarship.

Conflicts of interest Mehdi El Hassani and Amélie Marsot declare that they have no potential conflicts of interest that might be relevant to the contents of this manuscript.

Author contributions ME and AM conceptualised the manuscript. ME performed the literature search and wrote the manuscript. AM revised the manuscript.

Data availability statement Data sharing is not applicable to this article as no datasets were generated or analysed.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Code availability Not applicable.

References

1. Kantasiripitak W, Van Daele R, Gijsen M, Ferrante M, Spriet I, Dreesen E. Software tools for model-informed precision dosing: how well do they satisfy the needs? Front Pharmacol. 2020;11:620.
2. Aljutayli A, Thirion DJG, Bonnefois G, Nekka F. Pharmacokinetic equations versus Bayesian guided vancomycin monitoring: Pharmacokinetic model and model-informed precision dosing trial simulations. Clin Transl Sci. 2022;15(4):942–53.
3. de Velde F, Mouton JW, de Winter BCM, van Gelder T, Koch BCP. Clinical applications of population pharmacokinetic models of antibiotics: challenges and perspectives. Pharmacol Res. 2018;134:280–8.
4. Frenette C, Sperlea D, German GJ, Afra K, Boswell J, Chang S, et al. The 2017 global point prevalence survey of antimicrobial consumption and resistance in Canadian hospitals. Antimicrob Resist Infect Control. 2020;9(1):104.
5. Mauldin PD, Salgado CD, Hansen IS, Durup DT, Bosso JA. Attributable hospital cost and length of stay associated with health care-associated infections caused by antibiotic-resistant gram-negative bacteria. Antimicrob Agents Chemother. 2010;54(1):109–15.
6. Peters L, Olson L, Khu DTK, Linnros S, Le NK, Hanberger H, et al. Multiple antibiotic resistance as a risk factor for mortality and prolonged hospital stay: A cohort study among neonatal intensive care patients with hospital-acquired infections caused

by gram-negative bacteria in Vietnam. PLoS ONE. 2019;14(5): e0215666.
7. Antimicrobial Resistance Collaborators. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet. 2022;399(10325):629–55. https://doi.org/10.1016/S0140-6736(21)02724-0. Erratum in: Lancet. 2022;400(10358):1102.
8. Leekha S, Terrell CL, Edson RS. General principles of antimicrobial therapy. Mayo Clin Proc. 2011;86(2):156–67.
9. Hughes DM, Goswami S, Keizer RJ, Hughes MA, Faldasz JD. Bayesian clinical decision support-guided versus clinician-guided vancomycin dosing in attainment of targeted pharmacokinetic parameters in a paediatric population. J Antimicrob Chemother. 2020;75(2):434–7.
10. Frymoyer A, Schwenk HT, Zorn Y, Bio L, Moss JD, Chasmawala B, et al. Model-informed precision dosing of vancomycin in hospitalized children: implementation and adoption at an Academic Children's Hospital. Front Pharmacol. 2020;11:551.
11. Brooks JT, Keizer RJ, Long-Boyle JR, Kharbanda S, Dvorak CC, Friend BD. Population pharmacokinetic model development of tacrolimus in pediatric and young adult patients undergoing hematopoietic cell transplantation. Front Pharmacol. 2021;12: 750672.
12. Darwich AS, Ogungbenro K, Vinks AA, Powell JR, Reny JL, Marsousi N, 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. 2017;101(5):646–56.
13. Heus A, Uster DW, Grootaert V, Vermeulen N, Somers A, In't Veld DH, Wicha SG, De Cock PA. Model-informed precision dosing of vancomycin via continuous infusion: a clinical fit-for-purpose evaluation of published PK models. Int J Antimicrob Agents. 2022;59(5):106579. https://doi.org/10.1016/j.ijantimicag.2022.106579.
14. Marsot A, Boulamery A, Bruguerolle B, Simon N. Vancomycin: a review of population pharmacokinetic analyses. Clin Pharmacokinet. 2012;51(1):1–13.
15. Aljutayli A, Marsot A, Nekka F. An update on population pharmacokinetic analyses of vancomycin, part I. In adults. Clin Pharmacokinet. 2020;59(6):671–98.
16. Aljutayli A, El-Haffaf I, Marsot A, Nekka F. An update on population pharmacokinetic analyses of vancomycin, part II: in pediatric patients. Clin Pharmacokinet. 2022;61(1):47–70.
17. Chung E, Sen J, Patel P, Seto W. Population pharmacokinetic models of vancomycin in paediatric patients: a systematic review. Clin Pharmacokinet. 2021;60(8):985–1001.
18. Duffull SB, Wright DF. What do we learn from repeated population analyses? Br J Clin Pharmacol. 2015;79(1):40–7. https://doi.org/10.1111/bcp.12233.
19. Yano Y, Beal SL, Sheiner LB. Evaluating pharmacokinetic/pharmacodynamic models using the posterior predictive check. J Pharmacokinet Pharmacodyn. 2001;28(2):171–92.
20. Sheiner LB, Beal SL. Some suggestions for measuring predictive performance. J Pharmacokinet Biopharm. 1981;9(4):503–12.
21. Nguyen THT, Mouksassi MS, Holford N, Al-Huniti N, Freedman I, Hooker AC, et al. Model evaluation of continuous data pharmacometric models: metrics and graphics. CPT Pharmacomet Syst Pharmacol. 2017;6(2):87–109.
22. US FDA. Population pharmacokinetics: guidance for industry. 2022. https://www.fda.gov/media/128793/download. Accessed 03 Feb 2023.
23. Brendel K, Dartois C, Comets E, Lemenuel-Diot A, Laveille C, Tranchand B, et al. Are population pharmacokinetic and/or pharmacodynamic models adequately evaluated? A survey of the literature from 2002 to 2004. Clin Pharmacokinet. 2007;46(3):221–34.
24. Cheng Y, Wang C-y, Li Z-r, Pan Y, Liu M-b, Jiao Z. Can population pharmacokinetics of antibiotics be extrapolated? Implications of external evaluations. Clin Pharmacokinet. 2021;60(1):53–68.
25. Mentre F, Ebelen ME. Validation of population pharmacokinetic–pharmacodynamic analyses: review of proposed approaches. In: Balant LP, Aarons L, editors. The population approach: measuring and managing variability in response concentration and dose (COST B1). Brussels: Office for Official Publications of the European Communities; 1997. p. 146–60.
26. Short TG, Willemsen G. Total intravenous anesthesia and target controlled infusions: a comprehensive global anthology. Anesth Analg. 2018;126(2):718.
27. Kanji S, Hayes M, Ling A, Shamseer L, Chant C, Edwards DJ, et al. Reporting guidelines for clinical pharmacokinetic studies: the ClinPK statement. Clin Pharmacokinet. 2015;54(7):783–95.
28. Brendel K, Dartois C, Comets E, Lemenuel-Diot A, Laveille C, Tranchand B, et al. Are population pharmacokinetic and/ or pharmacodynamic models adequately evaluated? Clin Pharmacokinet. 2007;46(3):221–34.
29. Short TG, Aun CS, Tan P, Wong J, Tam YH, Oh TE. A prospective evaluation of pharmacokinetic model controlled infusion of propofol in paediatric patients. Br J Anaesth. 1994;72(3):302–6.
30. Marsh B, White M, Morton N, Kenny GN. Pharmacokinetic model driven infusion of propofol in children. Br J Anaesth. 1991;67(1):41–8.
31. El Hassani M, Simard C, Pilote S, Cloutier I, Soufsaf S, Marsot A. Consideration of height-based tobramycin dosing regimens for the treatment of adult cystic fibrosis pulmonary exacerbations. Br J Clin Pharmacol. 2021;88(5):2246–55.
32. Zhao CY, Jiao Z, Mao JJ, Qiu XY. External evaluation of published population pharmacokinetic models of tacrolimus in adult renal transplant recipients. Br J Clin Pharmacol. 2016;81(5):891–907.
33. Sheiner LB, Beal SL, Dunne A. Analysis of nonrandomly censored ordered categorical longitudinal data from analgesic trials. J Am Stat Assoc. 1997;92(440):1235–44.
34. Zhao W, Kaguelidou F, Biran V, Zhang D, Allegaert K, Capparelli EV, et al. External Evaluation of Population Pharmacokinetic Models of Vancomycin in Neonates: the transferability of published models to different clinical settings. Br J Clin Pharmacol. 2013;75(4):1068–80.
35. Lv C, Lu J, Jing L, Liu TT, Chen M, Zhang R, et al. Systematic external evaluation of reported population pharmacokinetic models of vancomycin in Chinese children and adolescents. J Clin Pharm Ther. 2021;46(3):820–31.
36. Marsot A, Michel F, Chasseloup E, Paut O, Guilhaumou R, Blin O. Phenobarbital in intensive care unit pediatric population: predictive performances of population pharmacokinetic model. Fundam Clin Pharmacol. 2017;31(5):558–66.
37. Bergstrand M, Hooker AC, Wallin JE, Karlsson MO. Prediction-corrected visual predictive checks for diagnosing nonlinear mixed-effects models. AAPS J. 2011;13(2):143–51.
38. Mentré F, Escolano S. Prediction discrepancies for the evaluation of nonlinear mixed-effects models. J Pharmacokinet Pharmacodyn. 2006;33(3):345–67.
39. Brendel K, Comets E, Laffont C, Laveille C, Mentré F. Metrics for external model evaluation with an application to the population pharmacokinetics of gliclazide. Pharm Res. 2006;23(9):2036–49.
40. Zhang R, Chen M, Liu T-t, Lu J-J, Lv C-l. Comparison of the predictive performance between cystatin c and serum creatinine by vancomycin via a population pharmacokinetic models: a prospective study in a Chinese population. Eur J Drug Metab Pharmacokinet. 2020;45(1):135–49.
41. Hu C, Yin W-j, Li D-y, Ding J-j, Zhou L-y, Wang J-l, et al. Evaluating tacrolimus pharmacokinetic models in adult renal transplant

recipients with different CYP3A5 genotypes. Eur J Clin Pharmacol. 2018;74(11):1437–47.
42. Mao J-J, Jiao Z, Yun H-Y, Zhao C-Y, Chen H-C, Qiu X-Y, et al. External evaluation of population pharmacokinetic models for ciclosporin in adult renal transplant recipients. Br J Clin Pharmacol. 2018;84(1):153–71.
43. Varvel JR, Donoho DL, Shafer SL. Measuring the predictive performance of computer-controlled infusion pumps. J Pharmacokinet Biopharm. 1992;20(1):63–94.
44. Miyabe-Nishiwaki T, Masui K, Kaneko A, Nishiwaki K, Nishio T, Kanazawa H. Evaluation of the predictive performance of a pharmacokinetic model for propofol in Japanese macaques (*Macaca fuscata fuscata*). J Vet Pharmacol Ther. 2013;36(2):169–73.
45. Glass PS, Shafer S, Reves JG. Intravenous drug delivery systems. In: Miller RD, editor. Miller's Anesthesia. Churchill Livingstone: Elsevier; 2004. p. 439–80.
46. Sepúlveda P, Cortínez LI, Sáez C, Penna A, Solari S, Guerra I, et al. Performance evaluation of paediatric propofol pharmacokinetic models in healthy young children. Br J Anaesth. 2011;107(4):593–600.
47. Masui K, Upton RN, Doufas AG, Coetzee JF, Kazama T, Mortier EP, et al. The performance of compartmental and physiologically based recirculatory pharmacokinetic models for propofol: a comparison using bolus, continuous, and target-controlled infusion data. Anesth Analg. 2010;111(2):368–79.
48. Coppens M, Van Limmen JG, Schnider T, Wyler B, Bonte S, Dewaele F, et al. Study of the time course of the clinical effect of propofol compared with the time course of the predicted effect-site concentration: Performance of three pharmacokinetic-dynamic models. Br J Anaesth. 2010;104(4):452–8.
49. Yang N, Wang J, Xie Y, Ding J, Wu C, Liu J, et al. External evaluation of population pharmacokinetic models to inform precision dosing of meropenem in critically ill patients. Front Pharmacol. 2022;13: 838205.
50. Thibault C, Zuppa AF. Dexmedetomidine in children on extracorporeal membrane oxygenation: pharmacokinetic data exploration using previously published models. Front Pediatr. 2022;10:924829. https://doi.org/10.3389/fped.2022.924829.
51. Guo T, van Hest RM, Roggeveen LF, Fleuren LM, Thoral PJ, Bosman RJ, van der Voort PHJ, Girbes ARJ, Mathot RAA, Elbers PWG. External evaluation of population pharmacokinetic models of vancomycin in large cohorts of intensive care unit patients. Antimicrob Agents Chemother. 2019;63(5):e02543–18. https://doi.org/10.1128/AAC.02543-18
52. Alghanem S, Paterson I, Touw DJ, Thomson AH. Influence of multiple courses of therapy on aminoglycoside clearance in adult patients with cystic fibrosis. J Antimicrob Chemother. 2013;68(6):1338–47.
53. 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(4):928–41.
54. Zhou Y, Long E, Shi T, Wang Z, Zhao J, Liu H, et al. External validation of vancomycin population pharmacokinetic models in ten cohorts of infected Chinese patients. J Glob Antimicrob Resist. 2022;30:163–72.
55. 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(12):785–7. https://doi.org/10.1002/psp4.12353.
56. Graves DA, Locke CS Jr, Muir KT, Miller RP. The influence of assay variability on pharmacokinetic parameter estimation. J Pharmacokinet Biopharm. 1989;17(5):571–92. https://doi.org/10.1007/BF01071350.
57. El Hassani M, Marsot A. Impact of sampling times on the predictive performance of tobramycin population pharmacokinetic models [abstract no 24]. Population Approach Group in Europe 2021.
58. Wang YL, Guilhaumou R, Blin O, Velly L, Marsot A. External evaluation of population pharmacokinetic models for continuous administration of meropenem in critically ill adult patients. Eur J Clin Pharmacol. 2020;76(9):1281–9.
59. Verrier D, Sivapregassam S, Solente AC. Dose-finding studies, MCP-Mod, model selection, and model averaging: Two applications in the real world. Clin Trials. 2014;11(4):476–84.
60. Aoki Y, Hooker AC. Model averaging and selection methods for model structure and parameter uncertainty quantification. In: PAGE. Abstracts of the Annual Meeting of the Population Approach Group in Europe. ISSN 1871-6032; 2016.
61. Aoki Y, Röshammar D, Hamrén B, Hooker AC. Model selection and averaging of nonlinear mixed-effect models for robust phase III dose selection. J Pharmacokinet Pharmacodyn. 2017;44(6):581–97. https://doi.org/10.1007/s10928-017-9550-0.
62. 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(8):1045–59.
63. Heus A, Uster DW, Grootaert V, Vermeulen N, Somers A, Huis in't Veld D, et al. Model-informed precision dosing of vancomycin via continuous infusion: a clinical fit-for-purpose evaluation of published PK models. Int J Antimicrobi Agents. 2022;59(5):106579. https://doi.org/10.1016/j.ijantimicag.2022.106579.
64. Uster DW, 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 Ther. 2021;109(1):175–83.
65. Sarker IH. AI-based modeling: techniques, applications and research issues towards automation, intelligent and smart systems. SN Comput Sci. 2022;3(2):158.
66. Brier ME, Zurada JM, Aronoff GR. Neural network predicted peak and trough gentamicin concentrations. Pharm Res. 1995;12(3):406–12.
67. Huang X, Yu Z, Bu S, Lin Z, Hao X, He W, et al. An ensemble model for prediction of vancomycin trough concentrations in pediatric patients. Drug Des Devel Ther. 2021;15:1549–59.
68. Hughes JH, Keizer RJ. A hybrid machine learning/pharmacokinetic approach outperforms maximum a posteriori Bayesian estimation by selectively flattening model priors. CPT Pharmacometrics Syst Pharmacol. 2021;10(10):1150–60.
69. Woillard J-B, Labriffe M, Debord J, Marquet P. Tacrolimus exposure prediction using machine learning. Clin Pharmacol Ther. 2021;110(2):361–9.
70. Lee S, Song M, Han J, Lee D, Kim BH. Application of machine learning classification to improve the performance of vancomycin therapeutic drug monitoring. Pharmaceutics. 2022;14(5):1023. https://doi.org/10.3390/pharmaceutics14051023
71. Kim YK, Lee JH, Jang HJ, Zang DY, Lee DH. Predicting antibiotic effect of vancomycin using pharmacokinetic/pharmacodynamic modeling and simulation: dense sampling versus sparse sampling. antibiotics (Basel). 2022;11(6):743. https://doi.org/10.3390/antibiotics11060743.