



HAL
open science

Estimation of drug exposure by machine learning based on simulations from published pharmacokinetic models: The example of tacrolimus

Jean-Baptiste Woillard, Marc Labriffe, Aurélie Prémaud, Pierre Marquet

► To cite this version:

Jean-Baptiste Woillard, Marc Labriffe, Aurélie Prémaud, Pierre Marquet. Estimation of drug exposure by machine learning based on simulations from published pharmacokinetic models: The example of tacrolimus. *Pharmacological Research*, 2021, 167, pp.105578. 10.1016/j.phrs.2021.105578 . hal-04414355

HAL Id: hal-04414355

<https://unilim.hal.science/hal-04414355>

Submitted on 29 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Estimation of drug exposure by machine learning based on simulations from published pharmacokinetic models: the example of tacrolimus.

Jean-Baptiste Woillard^{a, b, c}, Marc Labriffe^{a, b, c}, Aurélie Prémaud^{a, b}, Pierre Marquet^{a, b, c}

- a. Univ. Limoges, IPPRITT, F-87000 Limoges, France.
- b. INSERM, IPPRITT, U1248, F-87000 Limoges, France.
- c. Department of Pharmacology and Toxicology, CHU Limoges, F-87000 Limoges, France.

ORCID: 0000-0003-1695-0695

Corresponding author:

Jean-Baptiste Woillard

Title: PharmD, PhD

Adress: CBRS, Service de Pharmacologie, Toxicologie et Pharmacovigilance, CHU Limoges, 2 rue du Pr Descottes, F-87000 Limoges, France.

Phone: +33 5 55 05 61 40

Fax: +33 5 55 05 61 62

Email: jean-baptiste.woillard@unilim.fr

Abstract

We previously demonstrated that Machine learning (ML) algorithms can accurately estimate drug area under the curve (AUC) of tacrolimus or mycophenolate mofetil (MMF) based on limited information, as well as or even better than maximum a posteriori Bayesian estimation (MAP-BE). However, the major limitation in the development of such ML algorithms is the limited availability of large databases of concentration vs. time profiles for such drugs. The objectives of this study were: (i) to develop a Xgboost model to estimate tacrolimus inter-dose AUC based on concentration-time profiles obtained from a literature population pharmacokinetic (POPPK) model using Monte Carlo simulation; and (ii) to compare its performance with that of MAP-BE in external datasets of rich concentration-time profiles.

The population parameters of a previously published PK model were used in the mrgsolve R package to simulate 9000 rich interdose tacrolimus profiles (one concentration simulated every 30 minutes) at steady-state. Data splitting was performed to obtain a training set (75%) and a test set (25%). Xgboost algorithms able to estimate tacrolimus AUC based on 2 or 3 concentrations were developed in the training set and the model with the lowest RMSE in a ten-fold cross-validation experiment was evaluated in the test set, as well as in 4 independent, rich PK datasets from transplant patients.

ML algorithms based on 2 or 3 concentrations and a few covariates yielded excellent AUC estimation in the external validation datasets (relative bias <5% and relative RMSE <10%), comparable to those obtained with MAP-BE.

In conclusion, Xgboost machine learning models trained on concentration-time profiles simulated using literature POPPK models allow accurate tacrolimus AUC estimation based on sparse concentration data. This study paves the way to the development of artificial intelligence at the service of precision therapeutic drug monitoring in different therapeutic areas.

Keywords: machine learning; Monte Carlo simulations; tacrolimus; Xgboost; therapeutic drug monitoring; population pharmacokinetics

Abbreviations

artificial intelligence (AI)

area-under the curve (AUC)

maximum a posteriori Bayesian estimates (MAP-BE)

machine learning (ML)

mean prediction error (MPE)

population pharmacokinetic (POPPK)

Root mean square error (RMSE)

therapeutic drug monitoring (TDM)

1. Introduction

We recently developed Xgboost machine learning (ML) models for tacrolimus area-under the curve (AUC) estimation based on 2 or 3 concentrations and 4 covariates (dose, type of transplantation, age and time between transplantation and sampling) [1]. These models were developed using a large database of maximum a posteriori Bayesian estimates (MAP-BE) of tacrolimus interdose AUC in transplant patients on twice-daily tacrolimus, derived from blood concentrations measured pre-dose and 1 and 3h post-dose. The performance of our Xgboost models with respect to gold-standard AUCs obtained with the trapezoidal rule applied to full PK profiles in 4 external datasets from kidney, liver or heart transplant patients, was excellent and better than that of MAP-BE. This approach looks promising but requires large datasets to build accurate artificial intelligence (AI) algorithms that can cope with the complexity of the relationships between features and the output value(s). Actually, the availability of such large datasets is the major limitation to the dissemination of this approach for precision therapeutic drug monitoring based on overall drug exposure.

A theoretical solution would be to simulate datasets and develop ML models based on these simulations, as is sometimes done for population pharmacokinetics models [2]. Indeed, simulations are largely used in population pharmacokinetics modeling for evaluation or inference, but it must be interpreted with the clear understanding of the limitations and assumptions inherent to the model (e.g., extrapolations requiring assumption of linearity) [3].

In a previous study, we compared different machine learning approaches (Multivariate Adaptive Regression Spline [4] and Extreme Gradient Boosting; Xgboost[5]) to estimate iohexol clearance and we found that Xgboost based on the iterative construction of regression tree minimizing the residual errors of the previous regression tree, was associated with the best performances (smallest Mean Prediction Error and Root mean square error; RMSE) [6]. We further employed Xgboost to estimate tacrolimus and mycophenolic acid inter-dose AUC using 2 or 3 concentration-time points, with better efficacy than MAP-BE [1,7].

The objectives of the present study were: (i) to develop a Xgboost algorithm to estimate tacrolimus inter-dose AUC based on concentration-time profiles obtained using Monte Carlo simulation from a literature population pharmacokinetic (POPPK) model; and (ii) to compare its performance with that of MAP-BE in external datasets of rich concentration-time data.

2. Material and methods

2.1. Simulation process

The population parameters of a previously published pharmacokinetics model of tacrolimus in adult kidney transplant recipients [8] were used in the mrgsolve R package[9] to simulate 1000 rich tacrolimus PK profiles (one sample every 30 minutes) at steady-state over a 12h interval, for each of 9 drug doses (0.5, 1, 1.5, 2, 3, 3.5, 4, 6.5, 8 mg), representing a total of 9000 PK profiles. The mrgsolve model used (presented in the ESM1) consisted in a 2-compartment structural model with linear elimination from the central compartment and delayed absorption described by an erlang model with 3-transit compartments. The parameters of the final model reported in the original publication [8] were used with a few modifications: inter-occasion variability was not implemented and the combined residual error was drastically diminished (additive error = 0.1 $\mu\text{g/L}$ and proportional error = 0.01%) in order to obtain noisier, more realistic simulated PK profiles. For the same reason, filters were applied to remove profiles characterized by C_0 and/or C_{12} higher than 20 $\mu\text{g/L}$ and/or by an $\text{AUC}_{0-12\text{h}} < 40$ or $> 350 \mu\text{g}\cdot\text{h/L}$ so as to obtain a range of values compatible with observed data. Another filter was applied to remove profiles characterized by $C_0 > C_{0.5\text{h}}$ and $C_0 > C_{1\text{h}}$.

The interindividual variability in profiles was increased by simulating covariates: hematocrit values were simulated using a truncated random uniform distribution with mean \pm SD [min-max]= 32 ± 7 [21;46] % (in accordance to the original article) and cytochrome P450 3A5*1/*3 status using a binomial distribution with 90% of expressers for dose>6mg bid and 10% of expressers otherwise. Additionally, a random noise of 1% was independently added on simulated blood sampling times

using the `sdcMicro` R package in order to introduce uncertainty on this parameter [10]. Finally, the 'true' AUC was calculated for each simulated profile using the trapezoidal rule in the `PKNCA` R package[11].

2.2. Preparation of the data and feature engineering

Concentrations at approximative times 0, 1 and 3h were extracted from the simulated concentration vs. time profiles, together with the exact time of sampling (artificially noised as stated before), the simulated dosing information, CYP status, and hematocrit. Tacrolimus concentrations were binned into 3 theoretical time classes (concentrations at trough (C0 sampled at $t=0\text{min}$), at 1h (C1 sampled between 0.85 and 1.15h) and at 3h (C3 sampled between 2.8 and 3.2h), leading to 3 columns per patient. Features engineering (corresponding to the creation of artificial predictors to add information for the algorithm) was applied as previously described [1], leading to the generation of additional predictors: the relative deviation with respect to the theoretical time at t1h and t3h and the differences between concentrations (C3-C1, C3-C0 and C1 - C0). AUC prediction was finally performed using the tacrolimus morning dose, tacrolimus concentrations at approximative times 0, 1h and 3h, relative deviations from the theoretical times, differences between concentrations, hematocrit and CYP status. For the models based on 2 concentrations only (C0 and C3), the relative time difference for time = 1h and concentration differences including C1 were removed.

2.3. Machine learning analysis

A classical approach was used that includes splitting of the simulation dataset into a training (75%) and a test (25%) subsets. Ten-fold cross-validation was applied to the training set to tune the hyperparameters and evaluate the model performance by cross-validation, in order to prevent overfitting. Finally, the models with the best performance in terms of root mean square error (RMSE, expressed in $\mu\text{g}\cdot\text{h}/\text{L}$) and r^2 with respect to the 'true' AUC in the training subset were evaluated on the test subset by calculating the relative mean prediction error (MPE), relative RMSE and number of profiles out of the $\pm 20\%$ MPE interval between the estimated and 'true' AUCs.

2.4. External validation

The two final ML models (based on 2 and 3 concentrations, respectively) were externally validated on actual full-PK profiles measured in transplant patients, the AUC of which was estimated using the trapezoidal rule (PKNCA R package). The PCCP study [12] included 137 PK profiles of 11 samples for TAC BID (0, 0.33, 0.66, 1, 1.5, 2, 3, 4, 6, 8, 12h post dosing) collected at 7 and 14 days, 1, 3 and 6 months after kidney transplantation. The AADAPT study [13] included 34 PK profiles of 10 samples for TAC BID (0, 0.33, 0.66, 1, 2, 3, 4, 6, 8, 12h post dosing) collected at 7 days and 3 months post kidney transplantation. For liver transplantation [14] our database included 68 PK profiles of 9 samples for TAC BID (0, 0.5, 1, 2, 3, 4, 6, 8, 12h post dosing) collected at 7 days and 3 months post-transplantation. The PIGREC study [15,16] included 47 PK profiles of 11 samples for TAC BID (0, 0.33, 0.66, 1, 1.5, 2, 3, 4, 6, 8, 12h post dosing), collected at 7 days, 1, 3 and 12 months after heart transplantation. The performance of the Xgboost models in these independent validation datasets was compared to that of MAP-BE based on the 3-point limited sampling strategies available on our online expert system ISBA (<https://abis.chu-limoges.fr/login>) by calculating the relative MPE and relative RMSE, and the proportion of relative differences out of the ± 10 or the $\pm 20\%$ intervals. Additionally, scatter plots of predicted vs. reference AUCs and residuals vs. reference AUCs for the different approaches were presented on the same graph for visual comparison.

3. Results

3.1. Simulation and data

The simulated profiles are presented in the ESM 2. After applying the filters described above, 4392 simulated profiles were exploitable (19 profiles were filtered out because they had a C0 higher than C0.5 or C1, 4192 profiles had both C0 and C12 > 20 $\mu\text{g/L}$, 23 had either C0 or C12 > 20 $\mu\text{g/L}$ and 351 profiles had AUC out of the 40-350 $\mu\text{g}\cdot\text{h/L}$ interval). Histograms of the corresponding simulated concentrations and AUCs are presented in Figure 1. The correlation plots between AUCs and features is presented in the ESM 3. The correlation coefficients between AUC and C0, C1 or C3 were 0.97, 0.87

and 0.94, respectively. Typical values of the simulated concentrations, covariates and AUCs, as split into the training and test subsets, are presented in Table 1.

3.2. Xgboost model

The best-tuned parameter values of each model are presented in ESM4. Ten-fold cross-validation in the training subset and external validation in the test subset (Table 2) showed very accurate AUC estimation, with no difference in accuracy and precision between the two validation steps (no overfitting): the relative MPE was close to 0% as the relative RMSE lower than 5% in both. The scatter plots and residual plots obtained in the test subset are presented in Figure 2. The variable-importance plot of each model is presented in ESM5 and shows that concentrations at time 3 and 0 are the most important for AUC estimation.

3.3. External evaluation vs. the trapezoidal AUC and comparison with MAP-BE

The results of external evaluations (Table 3) show that the Xgboost model with 2 concentrations led to almost as accurate and precise results as the 3-point Xgboost model, and better than the 3-point MAP-BE in all patient datasets except in liver and heart transplant recipients for whom a larger imprecision ~12% was observed (Figure 3). The AUCs estimated using ML with 2 or 3 samples, MAP-BE with 3 samples and the reference AUC, for each of the 286 individuals enrolled in these independent studies used for external validation are compared in Figure 4 and show overall comparable performances.

4. Discussion

In this work, we used a previously published POPPK model of tacrolimus administered twice daily to kidney transplant patients to simulate rich concentration vs. time profiles and we developed an Xgboost algorithm to estimate tacrolimus AUC_{0-12h} using a limited number of “features”, be it tacrolimus whole blood concentrations, sampling times, patient CYP3A5*1/*3 genotype or hematocrit. The ML algorithms developed based on 2 and 3 samples compared favorably with

previous MAP-BE models for AUC estimation in external datasets of actual and rich concentration – time profiles, including from liver or heart transplant recipients and exhibited similar performances in comparison to those recently developed from actual patient data[1]. The present proof-of-concept study is the first to demonstrate that pharmacokinetic modelling and machine learning can be combined to derive accurate estimators of drug exposure, through simulation of rich concentration vs. time profiles. This innovative approach represents a breakthrough with respect to the previous situation where large databases had to be available to train a ML algorithm, with hardly any independent datasets left to validate it.

The present ML algorithm exhibits slightly lower performances in liver or heart than in kidney transplant recipients. This is not totally unexpected, as simulations of PK profiles were performed using a POPK model developed in renal transplant patients. It probably points to a limitation of the proposed strategy, i.e. that extrapolation to patient groups not covered by the initial model may result in higher imprecision.

Since modeling is always simplifying, simulations based on models lead to a loss of information in comparison to real data (especially the relationships with covariates). Additionally, the correlation between C₀ and AUC_{0-12h} ($r=0.94$, corresponding to a $r^2= 0.88$) was clearly overestimated as compared with classically reported r^2 values of 0.60 to 0.80 [17]. Moreover, the median simulated C₀ was slightly higher than the average values observed in routine practice (~10 vs ca. 7 µg/L, respectively [1]). The simulated C₀ in the original publication [8] were also visually (VPC plot) largely higher than those classically observed in transplant patients. Simulation generally leads to a larger range of values than those observed in real-life, but the application of drastic filters allowed us to limit the simulated profiles (after exclusion of ~50% of them) to a realistic range and to obtain C₀ and AUC distributions closer to those observed in routine practice. All these elements could have led to a biased model but the use of a boosting approach very different from the nonlinear mixed effect model use to generate the data prevented us from this bias.

Concerning the choice of filtering C0.5 or C1 higher than C0, we acknowledge that in some case, such a scenario can be observed, but this is not so frequent and in such cases, there is doubt about the correct timing of drug intake (large decreases are generally the consequence of an accident such as for example drug intake before the theoretical time). With such filtering, we wanted to get rid of really unrealistic PK profiles and only 0.2% of them were excluded with this filter.

External validation in independent clinical datasets demonstrated the validity of our approach and showed that even if the simulated data were not exactly similar to real data, this did not affect the estimation performance. Moreover, we found the same variable importance ranking as for the ML algorithms previously developed based on actual patient data[1]. This is probably due to the principle of Xgboost that splits continuous variables according to percentile distribution and choose only one splitting threshold at each node, meaning that it is little influenced by extreme or unrealistic values contrary to regression methods for example.

Interestingly, we initially simulated concentration-time profiles using the residual error values observed in the PK study [8] (additive error=0.71 $\mu\text{g/L}$ and proportional error= 11.3%) but that led to sometimes very unrealistic profiles with very strong gaps between 2 adjacent concentrations in a simulated profile (e.g. concentrations increasing from 5 to 30 $\mu\text{g/L}$ between samples virtually collected at 9 and 9.5h post dosing). This is because the residual error is applied independently to all the simulated concentrations of a profile (and not to the PK parameters), and this led us to fix the residual error close to 0 and to generate most of the variability between simulated profiles using the dose and covariate distributions. Of note, the AUC was not normally distributed because the simulated doses did not follow a normal distribution.

The hyperparameters obtained by tuning in the final 3-sample Xgboost algorithm were identical to those of our previous algorithm[1], but they were different between the 2-sample algorithms (tree depth = 1 in the present study while 10 in the previous one. As, only 3 variables were considered

important for the 2 sample ML algorithm as compared to 7 for the 3-sample algorithm, that translates into a lower depth of the tree for the former.

Implementation of literature models has largely been used to perform therapeutic drug monitoring (TDM) based on MAP-BE in different therapeutic areas (for example [18–23]). The models chosen have to be carefully selected based on convincing data fitting and validation. As mentioned above, attention has to be paid to patient characteristics in the population where the model was developed in order to limit extrapolations [3]. Very importantly, the models developed have to be validated in independent, actual datasets before routine use can be considered. The POPPK model used for simulations in the present study was characterized by a rather complex structure and a large number of parameters ($n=9$), eliciting close fitting of measured concentration-time profiles [3] and accurate Bayesian estimation of tacrolimus AUC in a validation set obtained by data splitting. It is important to note that not all published models may be convenient, and that sophisticated models may be better suited when it comes to simulations, to prevent oversimplification and unrealistic profiles. We hypothesize that the better the initial PK model, the better the ML algorithm, and we will investigate what makes a good POPPK model for this purpose in future works.

As mentioned above, the innovative approach tackled here may facilitate the development of ML algorithms dedicated to precision TDM. Additionally, ML algorithms derived from PK simulations could be further refined through the addition of actual data. This strategy may be particularly appropriate to estimate exposure based on limited sampling strategies for drugs for which POPPK models but no MAP-BE are available [24]. Actually, the PK data usually obtained as part of phase I and phase II clinical studies may be sufficient for this. Reciprocally, these models may be useful in the development of candidate drugs, in addition to POPPK, to help estimate the AUC in case of drug-drug interactions, or in special populations (children, pregnant women) in whom it is difficult to draw many blood samples.

This approach can also be theoretically extended to more complicated PK models, describing for instance free drug concentrations, or concentrations in tissues, or to PK/PD models.

5. Conclusion

In conclusion, we trained Xgboost machine learning algorithms on concentration-time curves of tacrolimus simulated using a literature POPPK model developed in kidney transplant recipients and demonstrated that they yielded very accurate tacrolimus AUC_{0-12h} estimation in independent and rich PK datasets from patients with a kidney, liver or heart transplant. This study paves the way to the development of ML algorithms for precision TDM based on exposure estimation in different therapeutic areas.

Declaration

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflict of interest: None in relation to this work

Data availability: The datasets generated and analysed during the current study are available from the corresponding author on request.

Code availability: available from the corresponding author on request

Author Contributions: JBW and PM designed the research and wrote the manuscript, JBW performed the research and analyzed the data with help from AP and ML. All coauthors reviewed and approved the final version of the manuscript.

Acknowledgments

We thank Ms. Karen Poole for manuscript editing.

References

- [1] J.-B. Woillard, M. Labriffe, J. Debord, P. Marquet, Tacrolimus exposure prediction using machine learning, *Clinical Pharmacology & Therapeutics*. n/a (2020).
<https://doi.org/10.1002/cpt.2123>.
- [2] T.M. Nanga, T.T.P. Doan, P. Marquet, F.T. Musuamba, Toward a robust tool for pharmacokinetic-based personalization of treatment with tacrolimus in solid organ transplantation: A model-based meta-analysis approach, *Br J Clin Pharmacol*. 85 (2019) 2793–2823. <https://doi.org/10.1111/bcp.14110>.
- [3] D.R. Mould, R.N. Upton, Basic concepts in population modeling, simulation, and model-based drug development, *CPT Pharmacometrics Syst Pharmacol*. 1 (2012) e6.
<https://doi.org/10.1038/psp.2012.4>.
- [4] J.H. Friedman, Multivariate Adaptive Regression Splines, *The Annals of Statistics*. 19 (1991) 1–67.
- [5] T. Chen, C. Guestrin, XGBoost: A Scalable Tree Boosting System, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*. (2016) 785–794. <https://doi.org/10.1145/2939672.2939785>.
- [6] J.-B. Woillard, C. Salmon Gandonnière, A. Destere, S. Ehrmann, H. Merdji, A. Mathonnet, P. Marquet, C. Barin-Le Guellec, A Machine Learning Approach to Estimate the Glomerular Filtration Rate in Intensive Care Unit Patients Based on Plasma Iohexol Concentrations and Covariates, *Clin Pharmacokinet*. (2020). <https://doi.org/10.1007/s40262-020-00927-6>.
- [7] J.-B. Woillard, M. Labriffe, J. Debord, P. Marquet, Mycophenolic acid exposure prediction using machine learning, *Clin Pharmacol Ther*. (2021). <https://doi.org/10.1002/cpt.2216>.
- [8] J.-B. Woillard, B.C.M. de Winter, N. Kamar, P. Marquet, L. Rostaing, A. Rousseau, Population pharmacokinetic model and Bayesian estimator for two tacrolimus formulations--twice daily Prograf and once daily Advagraf, *Br J Clin Pharmacol*. 71 (2011) 391–402.
<https://doi.org/10.1111/j.1365-2125.2010.03837.x>.
- [9] mrgsolve.github.io, (n.d.). <https://mrgsolve.github.io/> (accessed December 17, 2020).
- [10] M. Templ, A. Kowarik, B. Meindl, Statistical Disclosure Control for Micro-Data Using the R Package *sdcMicro*, *Journal of Statistical Software*. 67 (2015) 1–36.
<https://doi.org/10.18637/jss.v067.i04>.

- [11] W.S. Denney, S. Duvvuri, C. Buckeridge, Simple, Automatic Noncompartmental Analysis: The PKNCA R Package, *Journal of Pharmacokinetics and Pharmacodynamics*. 42 (2015) 11-107,S65. <https://doi.org/10.1007/s10928-015-9432-2>.
- [12] K. Benkali, A. Prémaud, N. Picard, J.-P. Rérolle, O. Toupance, G. Hoizey, A. Turcant, F. Villemain, Y. Le Meur, P. Marquet, A. Rousseau, Tacrolimus population pharmacokinetic-pharmacogenetic analysis and Bayesian estimation in renal transplant recipients, *Clin Pharmacokinet*. 48 (2009) 805–816. <https://doi.org/10.2165/11318080-000000000-00000>.
- [13] P. Marquet, L. Albano, J.-B. Woillard, L. Rostaing, N. Kamar, C. Sakarovitch, P. Gatault, M. Buchler, B. Charpentier, E. Thervet, E. Cassuto, Comparative clinical trial of the variability factors of the exposure indices used for the drug monitoring of two tacrolimus formulations in kidney transplant recipients, *Pharmacol. Res*. 129 (2018) 84–94. <https://doi.org/10.1016/j.phrs.2017.12.005>.
- [14] C. Riff, J. Debord, C. Monchaud, P. Marquet, J.-B. Woillard, Population pharmacokinetic model and Bayesian estimator for 2 tacrolimus formulations in adult liver transplant patients, *Br J Clin Pharmacol*. 85 (2019) 1740–1750. <https://doi.org/10.1111/bcp.13960>.
- [15] D. Fruit, A. Rousseau, C. Amrein, F. Rollé, N. Kamar, L. Sebbag, M. Redonnet, E. Epailly, P. Marquet, A. Prémaud, Ciclosporin population pharmacokinetics and Bayesian estimation in thoracic transplant recipients, *Clin Pharmacokinet*. 52 (2013) 277–288. <https://doi.org/10.1007/s40262-013-0037-x>.
- [16] J.-B. Woillard, F. Saint-Marcoux, C. Monchaud, R. Youdarène, L. Pouche, P. Marquet, Mycophenolic mofetil optimized pharmacokinetic modelling, and exposure-effect associations in adult heart transplant recipients, *Pharmacol. Res*. 99 (2015) 308–315. <https://doi.org/10.1016/j.phrs.2015.07.012>.
- [17] F. Saint-Marcoux, J.-B. Woillard, C. Jurado, P. Marquet, Lessons from routine dose adjustment of tacrolimus in renal transplant patients based on global exposure, *Ther Drug Monit*. 35 (2013) 322–327. <https://doi.org/10.1097/FTD.0b013e318285e779>.
- [18] C. Ogami, Y. Tsuji, Y. Nishi, H. Kawasuji, H. To, Y. Yamamoto, External evaluation of population pharmacokinetics and pharmacodynamics in linezolid-induced thrombocytopenia: the transferability of published models to different hospitalized patients, *Ther Drug Monit*. (2020). <https://doi.org/10.1097/FTD.0000000000000816>.
- [19] J. Methaneethorn, N. Leelakanok, Pharmacokinetic variability of phenobarbital: a systematic review of population pharmacokinetic analysis, *Eur J Clin Pharmacol*. (2020). <https://doi.org/10.1007/s00228-020-03011-x>.
- [20] T. Guo, R.M. van Hest, L.F. Roggeveen, L.M. Fleuren, P.J. Thoral, R.J. Bosman, P.H.J. van der Voort, A.R.J. Girbes, R.A.A. Mathot, P.W.G. Elbers, External Evaluation of Population

- Pharmacokinetic Models of Vancomycin in Large Cohorts of Intensive Care Unit Patients, *Antimicrob Agents Chemother.* 63 (2019). <https://doi.org/10.1128/AAC.02543-18>.
- [21] M. Tazuin, J.-M. Tréluyer, R. Nabbout, T. Billette de Villemeur, I. Desguerre, R. Aboura, I. Gana, Y. Zheng, S. Benaboud, N. Bouazza, C. Chenevier-Gobeaux, C. Freihuber, D. Hirt, Simulations of Valproate Doses Based on an External Evaluation of Pediatric Population Pharmacokinetic Models, *J Clin Pharmacol.* 59 (2019) 406–417. <https://doi.org/10.1002/jcph.1333>.
- [22] E. Santacana, L. Rodríguez-Alonso, A. Padullés, J. Guardiola, F. Rodríguez-Moranta, K. Serra, J. Bas, F. Morandeira Biology, H. Colom, N. Padullés, External Evaluation of Population Pharmacokinetic Models of Infliximab in Patients With Inflammatory Bowel Disease, *Ther Drug Monit.* 40 (2018) 120–129. <https://doi.org/10.1097/FTD.0000000000000476>.
- [23] M.F. Hwang, R.J. Beechinor, K.C. Wade, D.K. Benjamin, P.B. Smith, C.P. Hornik, E.V. Capparelli, S. Duara, K.A. Kennedy, M. Cohen-Wolkowicz, D. Gonzalez, External Evaluation of Two Fluconazole Infant Population Pharmacokinetic Models, *Antimicrob Agents Chemother.* 61 (2017). <https://doi.org/10.1128/AAC.01352-17>.
- [24] F. Gallais, L. Ysebaert, F. Despas, S. De Barros, L. Dupré, A. Quillet-Mary, C. Protin, F. Thomas, L. Obéric, B. Allal, E. Chatelut, M. White-Koning, Population Pharmacokinetics of Ibrutinib and Its Dihydrodiol Metabolite in Patients with Lymphoid Malignancies, *Clin Pharmacokinet.* 59 (2020) 1171–1183. <https://doi.org/10.1007/s40262-020-00884-0>.

Table 1: Characteristics of the simulated profiles in the training and test subsets

	Training subset	Test subset
n	3294	1098
Trough level (C0) ($\mu\text{g/L}$); median [range]	9.2 [1.1, 20.0]	9.4 [1.8, 20.0]
Concentration at 1h; C1 ($\mu\text{g/L}$); median [range]	15.7 [2.6, 81.2]	15.6 [3.1, 70.7]
Concentration at 3h; C3 ($\mu\text{g/L}$); median [range]	13.7 [2.9, 43.9]	14.0 [3.5, 42.2]
Deviation from the 1h theoretical time; median [range]	0.00 [-0.13, 0.13]	-0.00 [-0.12, 0.12]
Deviation from the 3h theoretical time; median [range]	0.00 [-0.05, 0.04]	-0.00 [-0.03, 0.04]
Morning dose (mg); median [range]	1.5 [0.5, 8.0]	1.5 [0.5, 8.0]
Hematocrit (%); median [range]	32 [21, 46]	32 [21, 46]
CYP450 3A5 expressors; number (%)	383 (11.6)	139 (12.7)
AUC0-12h ($\text{mg}\cdot\text{h/L}$); median [range]	142 [40, 349]	145 [40, 350]
Concentration difference between C1 and C0 ($\mu\text{g/L}$); median [range]	5.1 [-0.0, 66.1]	5.0 [0.0, 58.6]
Concentration difference between C1 and C3 ($\mu\text{g/L}$); median [range]	1.6 [-20.8, 52.2]	1.5 [-14.6, 45.4]
Concentration difference between C3 and C0 ($\mu\text{g/L}$); median [range]	3.6 [0.1, 29.0]	3.7 [0.2, 30.5]

Table 2: Performance of the models in the training and test datasets to estimate tacrolimus simulated 'true' AUC_{0-12h}.

		Xgboost model with 2 samples	Xgboost model with 3 samples
Training subset	*RMSE ± SD (µg*h/L)	7.32 ± 0.147	4.34 ± 0.138
	*R² ± SD	0.990 ± 0.0004	0.997 ± 0.0002
	Relative MPE (%)	0.33	0.15
	Relative RMSE (%)	4.66	2.66
Test subset	RMSE (µg*h/L)	7.53	4.48
	R²	0.990	0.997
	Relative MPE (%)	0.54	0.13
	Relative RMSE (%)	4.60	2.61
	Number of MPE out of the ±20% interval / n	1 / 1098	0 / 1098

*Mean and standard deviation obtained after 10-fold cross-validation.

Table 3: Comparative performance of ML algorithms trained of simulated data and of MAP-BE to estimate reference AUCs obtained from actual full PK profiles using the trapezoidal rule.

Study	Method	Relative MPE (%)	Relative RMSE (%)	Relative errors out of $\pm 20\%$ n (%)	Relative errors out of $\pm 10\%$ n (%)
Kidney transplantation 1 [12] (n = 137)	Xgboost 2 concentrations	1.08	8.55	3 (2.2)	30 (21.9)
	Xgboost 3 concentrations	3.28	7.98	1 (0.7)	25 (18.2)
	MAP-BE 3 concentrations	5.1	9.6	6 (4.4)	37 (27.0)
Kidney transplantation 2 [13] (n = 34)	Xgboost 2 concentrations	0.69	9.0	2 (5.9)	7 (20.6)
	Xgboost 3 concentrations	0.27	8.1	0 (0)	8 (23.5)
	MAP-BE 3 concentrations	2.8	9.1	1 (2.9)	9 (26.4)
Liver transplantation [14] (n = 68)	Xgboost 2 concentrations	5.9	12.9	6 (8.8)	23 (33.8)
	Xgboost 3 concentrations	3.7	8.8	2 (2.9)	14 (20.6)
	MAP-BE 3 concentrations	-3.4	11.2	6 (8.8)	18 (26.5)
Heart transplantation [15,16] (n = 47)	Xgboost 2 concentrations	3.2	11.4	4 (8.5)	13 (27.7)
	Xgboost 3 concentrations	3.4	9.1	2 (4.2)	10 (21.3)
	MAP-BE 3 concentrations	-0.7	9.1	1 (2.1)	13 (27.7)

Figure legends

Figure 1: Distributions of the simulated tacrolimus concentrations and AUC_{0-12h}

Figure 2: Scatter plots and residual plots of machine learning predicted vs. reference AUC_{0-12h} in the simulated test subset using 2 or 3 tacrolimus blood concentrations.

Figure 3: Scatter plots and residual plots of MAP-BE and machine learning predicted vs reference AUC_{0-12h} in the external validation set (actual profiles) using 2 or 3 tacrolimus blood concentrations. (MAP-BE is maximum a posteriori Bayesian estimation based on the 0, 1 and 3h limited sampling strategy and the parametric POPPK models available in ISBA, <https://abis.chu-limoges.fr/login>).

Figure 4: Comparison of the different methods of tacrolimus AUC_{0-12h} estimation in the external validation datasets. (MAP-BE is maximum a posteriori Bayesian estimation based on the 0, 1 and 3h limited sampling strategy and the parametric POPPK models available in ISBA).