



HAL
open science

Tacrolimus Exposure Prediction Using Machine Learning

Jean-baptiste Woillard, Marc Labriffe, Jean Debord, Pierre Marquet

► **To cite this version:**

Jean-baptiste Woillard, Marc Labriffe, Jean Debord, Pierre Marquet. Tacrolimus Exposure Prediction Using Machine Learning. *Clinical Pharmacology and Therapeutics*, 2021, 110 (2), pp.361-369. 10.1002/cpt.2123 . hal-04413866

HAL Id: hal-04413866

<https://unilim.hal.science/hal-04413866v1>

Submitted on 24 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.



Tacrolimus Exposure Prediction Using Machine Learning

Jean-Baptiste Woillard^{1,2,3,*}, Marc Labriffe^{1,2,3}, Jean Debord^{1,2,3} and Pierre Marquet^{1,2,3}

The aim of this work is to estimate the area-under the blood concentration curve of tacrolimus (TAC) following b.i.d. or q.d. dosing in organ transplant patients, using Xgboost machine learning (ML) models. A total of 4,997 and 1,452 TAC interdose area under the curves (AUCs) from patients on b.i.d. and q.d. TAC, sent to our Immunosuppressant Bayesian Dose Adjustment expert system (www.pharmaco.chu-limoges.fr/) for AUC estimation and dose recommendation based on TAC concentrations measured at least at 3 sampling times (predose, ~ 1 and 3 hours after dosing) were used to develop 4 ML models based on 2 or 3 concentrations. For each model, data splitting was performed to obtain a training set (75%) and a test set (25%). The Xgboost models in the training set with the lowest root mean square error (RMSE) in a 10-fold cross-validation experiment were evaluated in the test set and in 6 independent full-pharmacokinetic (PK) datasets from renal, liver, and heart transplant patients. ML models based on two or three concentrations, differences between these concentrations, relative deviations from theoretical times of sampling, and four covariates (dose, type of transplantation, age, and time between transplantation and sampling) yielded excellent AUC estimation performance in the test datasets (relative bias < 5% and relative RMSE < 10%) and better performance than maximum *a posteriori* Bayesian estimation in the six independent full-PK datasets. The Xgboost ML models described allow accurate estimation of TAC interdose AUC and can be used for routine TAC exposure estimation and dose adjustment. They will soon be implemented in a dedicated web interface.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ Tacrolimus interdose area under the curve (AUC) for individual dose adjustment is difficult to estimate in routine patient care.

WHAT QUESTION DID THIS STUDY ADDRESS?

☑ We investigated whether machine learning (ML) models could estimate tacrolimus (TAC) AUC using a limited number of blood concentrations, as well as or even better than deterministic pharmacokinetic (PK) models with Bayesian estimation.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

☑ We developed Xgboost ML models allowing accurate estimation of TAC AUC based on two or three blood concentrations in kidney, liver, or heart transplant patients with excellent performance, better than that of the PK approach previously used.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

☑ These models will be soon implemented in an expert system made available to the transplant community through a dedicated website.

Tacrolimus is a calcineurin inhibitor largely used for the prevention of rejection in solid organ transplantation.¹ It is characterized by a narrow therapeutic range and a large inter-individual variability rendering its therapeutic drug monitoring mandatory. Two main markers are currently available to adjust the individual dose: the trough whole blood level (C₀) widely used for practical and economic reasons, although it has been inconsistently associated with outcomes² and the area under the curve (AUC), which is theoretically a better marker of exposure, although no formal evidence is available.¹ AUC is more difficult to measure than C₀, especially without the use

of population pharmacokinetic (PopPK) models.³ This led us to launch, in 2005, the Immunosuppressant Bayesian Dose Adjustment (ISBA) expert system and website⁴ to share tools able to estimate the interdose AUC of immunosuppressants using maximum *a posteriori* Bayesian estimation (MAP-BE) on the basis of three blood samples and some patient characteristics (type of graft, age, post-transplantation period, and drug measurement assay).^{5,6} With ISBA, each request posted is validated in < 48 hours by a trained pharmacologist, representing a huge workload due to the large number of requests received from transplant centers worldwide (> 120,000 since 2005).

¹University of Limoges, IPPRIT, Limoges, France; ²INSERM, IPPRIT, U1248, Limoges, France; ³Department of Pharmacology and Toxicology, CHU Limoges, Limoges, France. *Correspondence: Jean-Baptiste Woillard (jean-baptiste.woillard@unilim.fr)

Received September 21, 2020; accepted November 13, 2020. doi:10.1002/cpt.2123

Additionally, we recently proposed a new metrics to dose adjust tacrolimus (TAC) in kidney transplant recipients: the AUC/C0 ratio, which has been shown, in contrast to Tac AUC or C0, to be stable in individuals along time, even in the early period after transplantation.^{6,7}

Machine learning (ML) has seldom been used in pharmacology so far.⁸ Extreme gradient boosting is a ML approach where simple regression trees are iteratively built by finding split values among all input variables that minimize prediction error. The iterative process constructs an additional regression tree of the same structure that minimizes the residual errors of the previous regression tree.⁹

In the context of PKs, we recently showed that these methods can successfully be applied to estimate iohexol clearance using a few blood concentrations and a number of patient features (i.e., demographic characteristics, laboratory test results, disease history, or associated medications).¹⁰

The objective of this study was to develop Xgboost models to estimate TAC interdose AUC, based on a limited number of blood concentrations (2 or 3) and predictors, and to compare their performance to that of MAP-BE in external validation datasets.

METHODS

Patients and data

The TAC AUC estimation and dose recommendation requests received on our ISBA website since 2007, whatever the type of transplantation, were extracted and cleaned using the tidyverse framework. We sequentially refined the dataset by selecting the requests concerning the formulations Prograf (TAC b.i.d.) or Advagraf (TAC q.d.), where TAC blood levels were measured using high-performance liquid chromatography, including at least three times of sampling at trough (C0), and ~ 60 minutes (30–100 minutes, C1) and ~ 180 minutes (140–220 minutes, C3) after drug intake. Actually, the 0, 60, and 180 minutes is the optimal sampling schedule for TAC b.i.d.¹¹ and TAC q.d.¹² The optimal sampling times requested for Envarsus are 0, 8, and 12 hours¹³ and, for this reason, this formulation was not included in the analysis. The resulting database was split into two datasets according to the TAC interdose (12 hours or 24 hours) and two independent models were developed, one for TAC b.i.d. and one for the TAC q.d. The other predictors available were the morning dose of TAC, the time elapsed between transplantation and TAC blood sampling, the type of transplant, and patient age. The code used for data cleaning can be provided upon request.

Plan of the study

The present study used supervised learning to predict the interdose AUC, whose reference value had been obtained by our ISBA expert system using MAP-BE and at least three concentrations. We developed four ML models, two for each formulation (to predict AUC_{0–12h} and AUC_{0–24h}, respectively), one based on two concentrations and the other on three. A training set was used to build the model, tune the hyperparameters, and evaluate model performance by cross-validation. Once the best model had been built, it was evaluated on an independent test set by measuring the root mean square error (RMSE; expressed in $\mu\text{g}\cdot\text{h}/\text{L}$) between the estimated and reference AUCs. As the reference (BE) AUC can be considered as an imprecise and potentially biased estimate of the “true” AUC, we investigated in a second time the performance of our Xgboost models including two or three concentrations on “true” AUCs obtained from full PK profiles using

the unbiased trapezoidal method. For that, two datasets of TAC b.i.d. in renal transplant recipients,^{7,11} one in liver transplant recipients,¹⁴ and one in heart transplant recipients,^{15,16} as well as a dataset of TAC q.d. in renal transplant recipients⁷ and another one in liver transplant recipients¹⁴ were used. The performance of the Xgboost models in these confirmation datasets was also compared with that of the BE based on the three sample limited sampling strategies available on ISBA.

Feature engineering

The TAC concentrations were binned into three theoretical time classes (concentrations at trough (C0 sampled at $t = 0$ minutes), at 1 hour (C1 sampled between 30 and 100 minutes), and at 3 hours (C3 sampled between 140 and 220 minutes), leading to three columns per patient. New variables were drawn for times 1 hour and 3 hours corresponding to the relative deviation with respect to the theoretical times. For instance, if the actual sampling time was 1.06 hours for the theoretical time 1 hour, the relative time difference was $(1.06-1)/1 = 0.06$. Other predictors corresponding to the differences between the concentrations C1–C0, C1–C3, and C3–C0 were created to add information about potentially delayed absorption peaks. The type of transplant was split into five categories: kidney, lung, heart, and liver, plus an “other” category for all the other indications (including transplantation of 2 solid organs, bone marrow transplantation, or auto-immune diseases). Finally, the set of features used to predict interdose AUCs were: the type of transplant, patient age, time elapsed between transplantation and TAC blood sampling, TAC morning dose, TAC concentrations at times 0, 1 hour, and 3 hours, relative deviation from the theoretical times, and differences between concentrations. For the models based on two concentrations (C0 and C3), the relative time difference for time = 1 hour and concentration differences including C1 were removed.

Exploratory data analyses

A correlation matrix and scatterplots were drawn to explore the correlation between AUC and predictors using the GGally R package.¹⁷

Preprocessing of the data

For all the ML analyses, the tidymodels framework was used.¹⁸ No preprocessing was applied to the data because Xgboost methods do not require normalization prior to analysis. There were no missing data in the predictors. Data splitting was performed by random selection of patients in a training set (75%) and a test set (25%).

Development of Xgboost models

The four models were tuned by searching the parameter combination associated with the lowest RMSE and highest r^2 between estimated and reference AUC values, using 10-fold cross-validation (for which the training dataset was randomly split into 10 parts). In brief, the best combination of parameters was investigated in 90% of the training dataset (analysis subset) and evaluated in the remaining 10% (assessment subset) and this process was repeated 10 times by circular permutation. The parameters tuned among a grid of 30 random combinations were: the number of predictors randomly sampled at each split (mtry, between 1 and 11), the minimum number of data points in a node required for the node to be split further (min_n between 1 and 40), the maximum depth of the tree (tree_depth, between 1 and 15), and the rate at which the boosting algorithm adapts from iteration-to-iteration (learn_rate, between 0 and 0.08). Once the best combination of hyperparameters was selected, the relative importance of the predictors was evaluated by random permutations and a variable importance plot was drawn. In a second time, the best parameter combinations were evaluated using additional 10-fold cross-validations to assess the mean RMSE and r^2 and their SDs in the training set and scatter plots of estimated vs. reference

AUC were drawn. Finally, AUC prediction was performed in the test set. The estimation performance was evaluated by the RMSE and r^2 and by calculation of the relative mean prediction error (MPE), relative RMSE, number, and proportion of estimations with an MPE value of the $\pm 20\%$ interval. Scatter plots of predicted vs. reference AUCs and residual as function of reference values were drawn.

External evaluation vs. full PK profiles and comparison with PopPK

Concentrations at 0, 1, and 3 hours, as well as dose, sampling times, and time elapsed between transplantation and TAC blood sampling were extracted from the PK databases to predict the AUC using the ML models and the MAP-BE based on three concentrations, as used in ISBA. The full concentration profiles were used to calculate the “true” trapezoidal AUC using the PKNCA R package.¹⁹

Performances of the ML models using 2 and 3 concentrations and of MAP-BE based on 3 concentrations were compared to the trapezoidal AUC in terms of the relative MPE and relative RMSE, and the proportion of bias of the ± 10 or the $\pm 20\%$ interval. Additionally, scatter plot of predicted vs. reference AUCs and residuals as a function of the reference AUCs were drawn on the same graph for the different approaches for visual comparison.

The PCCP study¹¹ included 137 PK profiles of 11 samples for TAC b.i.d. (0, 0.33, 0.66, 1, 1.5, 2, 3, 4, 6, 8, and 12 hours postdosing) collected at 7 and 14 days, and 1, 3, and 6 months after renal transplantation. The AADAPT study⁷ included 34 PK profiles of 10 samples for TAC b.i.d. (0, 0.33, 0.66, 1, 2, 3, 4, 6, 8, and 12 hours postdosing) and 41 PK profiles of 13 samples for TAC q.d. (0, 0.33, 0.66, 1, 2, 3, 4, 6, 8, 12, 13, 15, and 24 hours postdosing) collected at 7 days and 3 months postrenal transplantation. In liver transplant recipients,¹⁴ our database included 68 PK profiles

of 9 samples for TAC b.i.d. (0, 0.5, 1, 2, 3, 4, 6, 8, and 12 hours postdosing) and 91 PK profiles of 17 samples for TAC q.d. (0, 0.5, 1, 2, 3, 4, 6, 8, 12, 12.5, 13, 14, 15, 16, 18, 20, and 24 hours postdosing), collected at 7 days and 3 months post-transplantation. The Pigrec study^{15,16} included 47 PK profiles of 11 samples for TAC b.i.d. (0, 0.33, 0.66, 1, 1.5, 2, 3, 4, 6, 8, and 12 hours postdosing), collected at 7 days, 1, 3, and 12 months after heart transplantation.

RESULTS

Patients and data

A total of 4,997 TAC AUC_{0–12h} and 1,452 AUC_{0–24h} from 2,060 and 775 patients, respectively, were available in the cleaned datasets extracted from ISBA. Characteristics of the training and test sets are reported in **Table 1**. The interdose AUCs ranged between 22 and 380 $\mu\text{g}\cdot\text{h}/\text{L}$ and between 44 and 698 $\mu\text{g}\cdot\text{h}/\text{L}$ for TAC b.i.d. and q.d., respectively.

Exploratory data analyses

Correlation matrices between AUC and predictors are presented in **Figure S1a** for TAC b.i.d. and **Figure S1b** for TAC q.d. showing a strong correlation (> 0.8) between interdose AUC and C₀ or C_{3h} for TAC b.i.d.; C₀ for TAC q.d.

Xgboost model

The best-tuned parameter values for each model are presented in **Supplementary Table S1**. Results in the training set obtained after 10-fold cross-validation and in the testing set are presented

Table 1 Characteristics of the ISBA requests used for the development and validation of the models

	TAC b.i.d.		TAC q.d.	
	Train (n = 3,748)	Test (n = 1,249)	Train (n = 1,087)	Test (n = 362)
Time between transplantation and tacrolimus blood concentrations, years	1.48 [0.27–5.32]	1.94 [0.27–5.59]	2.00 [0.35–5.19]	2.18 [0.40–6.00]
Interdose AUC, $\mu\text{g}\cdot\text{h}/\text{L}$ ^a	129 [103–164]	129 [102–163]	234 [187–287]	226 [182–286]
Patient age, year	50 [31–61]	50 [30–61]	53 [42–62]	53 [41–62]
Morning dose, mg	2.5 [1.5–4.0]	2.5 [1.5–4.0]	6.0 [4.0–10.0]	6.0 [4.0–8.0]
Trough level (C ₀), $\mu\text{g}/\text{L}$	7.3 [5.7–9.0]	7.3 [5.7–9.0]	6.7 [5.3–8.2]	6.5 [5.1–8.1]
Concentration at 1 hour; C ₁ , $\mu\text{g}/\text{L}$	15.1 [10.5–21.9]	15.0 [10.2–22.2]	12.6 [9.0–18.4]	12.9 [9.1–18.4]
Concentration at 3 hours; C ₃ , $\mu\text{g}/\text{L}$	13.3 [10.3–17.4]	13.2 [10.3–17.3]	14.8 [11.1–19.2]	14.4 [11.1–18.8]
Deviation from the 1-hour theoretical time	0.00 [0.00–0.05]	0.00 [0.00–0.07]	0.00 [0.00–0.05]	0.00 [0.00–0.07]
Deviation from the 3-hour theoretical time	0.00 [0.00–0.02]	0.00 [0.00–0.02]	0.00 [0.00–0.01]	0.00 [0.00–0.02]
Concentration difference between C ₁ and C ₀	7.6 [3.3–13.4]	7.2 [3.0–13.9]	5.8 [2.4–10.7]	6.2 [2.8–11.1]
Concentration difference between C ₁ and C ₃	1.8 [–1.5 to 6.3]	1.5 [–1.5–6.6]	–1.1 [–4.1 to 1.8]	–1.1 [–3.6 to 2.0]
Concentration difference between C ₃ and C ₀	5.9 [3.8–8.8]	5.9 [3.8–8.7]	7.8 [5.0–11.5]	7.6 [4.8–11.3]
<i>Transplant type</i>				
Kidney	3577 (95.4)	1192 (95.4)	1087 (100)	362 (100)
Lung	108 (2.9)	38 (3.0)	0 (0)	0 (0)
Heart	29 (0.8)	14 (1.1)	0 (0)	0 (0)
Other	19 (0.5)	3 (0.2)	0 (0)	0 (0)
Liver	15 (0.4)	2 (0.2)	0 (0)	0 (0)

Median [IQR] are presented for continuous data and n(%) for categorical data. TAC b.i.d. is the twice daily formulation of tacrolimus, TAC q.d. the once daily formulation.

AUC, area under the curve; ISBA, Immunosuppressant Bayesian Dose Adjustment.

^aInterdose is 12 hours for the b.i.d. models and 24 hours for the q.d. models

Table 2 Performances of the models in the training and testing datasets to estimate AUCs obtained from three samples MAP-BE from ISBA

	TAC b.i.d. 2 samples	TAC b.i.d. 3 samples	TAC q.d. 2 samples	TAC q.d. 3 samples
<i>Training</i>				
RMSE \pm SD, $\mu\text{g}^*\text{h}/\text{L}^{\text{a}}$	12.4 \pm 0.38	7.66 \pm 0.54	21.4 \pm 1.36	20.3 \pm 1.26
$R^2 \pm \text{SD}^{\text{a}}$	0.938 \pm 0.003	0.976 \pm 0.002	0.930 \pm 0.005	0.937 \pm 0.006
Relative MPE, %	0.7	0.3	0.8	0.7
Relative RMSE, %	8.7	5.1	9.2	8.8
<i>Testing</i>				
RMSE, $\mu\text{g}^*\text{h}/\text{L}$	11.5	6.6	24.2	23.0
R^2	0.941	0.980	0.907	0.916
Relative MPE, %	1.1	0.5	-0.1	0.03
Relative RMSE, %	8.6	5.0	8.1	7.8
Number of MPE of the \pm 20% interval <i>n</i>	34 (2.7%)	8 (0.6%)	14 (3.9%)	12 (3.3%)

AUC, area under the curve; ISBA, Immunosuppressant Bayesian Dose Adjustment; MAP-BE, maximum *a posteriori* Bayesian estimation; MPE, mean prediction error; RMSE, root mean square error.

^aValues and SDs obtained after 10-fold cross-validation. TAC b.i.d. is the twice daily formulation of tacrolimus, TAC q.d. the once daily formulation. Interdose is 12 hours for the b.i.d. models and 24 hours for the q.d. models.

in **Table 2** and show excellent results with no difference between them (no overfitting). The relative MPEs were close to 0 and the relative RMSEs were $< 10\%$ in both the training and testing sets. The scatter plot and residual plots obtained in the test sets are presented in **Figure 1**. The variable importance plot of each model is presented in **Figure S2** and shows that concentrations at time 3 and 0 are the most important variables.

External evaluation vs. the trapezoidal AUC and comparison with PopPK estimates

The results of the external evaluation are presented in **Table 3** and show that the Xgboost model with 2 concentrations led to acceptable results in comparison to the 3-point Xgboost model or the 3-point MAP-BE.

Figures S3 and S4 present the scatter plots and residual plots of estimated vs. reference AUC for TAC b.i.d. and q.d. in the validation studies. Figures 2 and 3 present for each individual the AUC estimated using ML with 2 or 3 samples, MAP-BE with 3 samples, and the reference AUC, for both TAC b.i.d. and q.d.

DISCUSSION

In this work, we developed Xgboost ML models to estimate the interdose AUC of TAC in renal transplant recipients. We then compared the results obtained with extensive-sampling-trapezoidal AUCs (reference AUCs) and to MAP-BE estimates based on a three-point limited sampling strategy, as used by the ISBA expert system.⁴ The performances of the ML were excellent as compared with the reference AUCs, even with only two concentrations, and were better than the MAP-BE AUCs.

Previously, we successfully applied this innovative approach to iohexol in patients in the intensive care unit in a proof-of-concept study that led us to investigate their potential for the dose adjustment of immunosuppressants, using the large database of the ISBA

website. The predictions obtained in the current work are excellent in comparison to the PopPK results usually obtained. Indeed, in a review article, Brooks *et al.* observed that 71% of the MAP-BE developed for TAC and using two or more TAC concentrations showed a bias $\leq 10\%$, but that only 39% had imprecision $\leq 10\%$.²⁰ Using ML models in external datasets, we found RMSE $\leq 10\%$ and bias $< 5\%$ and a very low number of profiles out of the acceptable $\pm 20\%$ relative bias range. We did not even expect such good results in light of the performances obtained in our previous work on iohexol. This illustrates perfectly the fact that, with ML models, the higher the number of data, the better the predictions. However, beyond a certain number, the addition of redundant information will not improve the results, as mentioned by Max Kuhn in his “Applied Predictive Modeling” reference book²¹: “there are cases where adding more samples may materially improve the quality of predictions, but one should remember that big data may not mean better data.”

The expected benefits of this study are very important. Indeed, TAC AUC is not largely used as an exposure marker for routine care of transplant recipients, because the collection of three samples is difficult, expensive, and rather uncomfortable for the patients. Decreasing the number of samples without decreasing the estimation performance should facilitate the use of TAC AUC for therapeutic drug monitoring. To increase the usability of this marker, we also have to test, in future works, the performance of our ML models for AUC estimation using dried blood spot TAC concentrations.

It can be questioned whether ML will replace PopPK analysis, at least for clinical applications. Actually, the goals and possibilities of PopPK and ML are very different. For example, no simulation is possible using ML, which limits the use for such models to calculate the probability of target attainment and select the best dosing scenario, for instance. Contrary to PK-based approaches whose goal is to describe the physiological phenomena involved

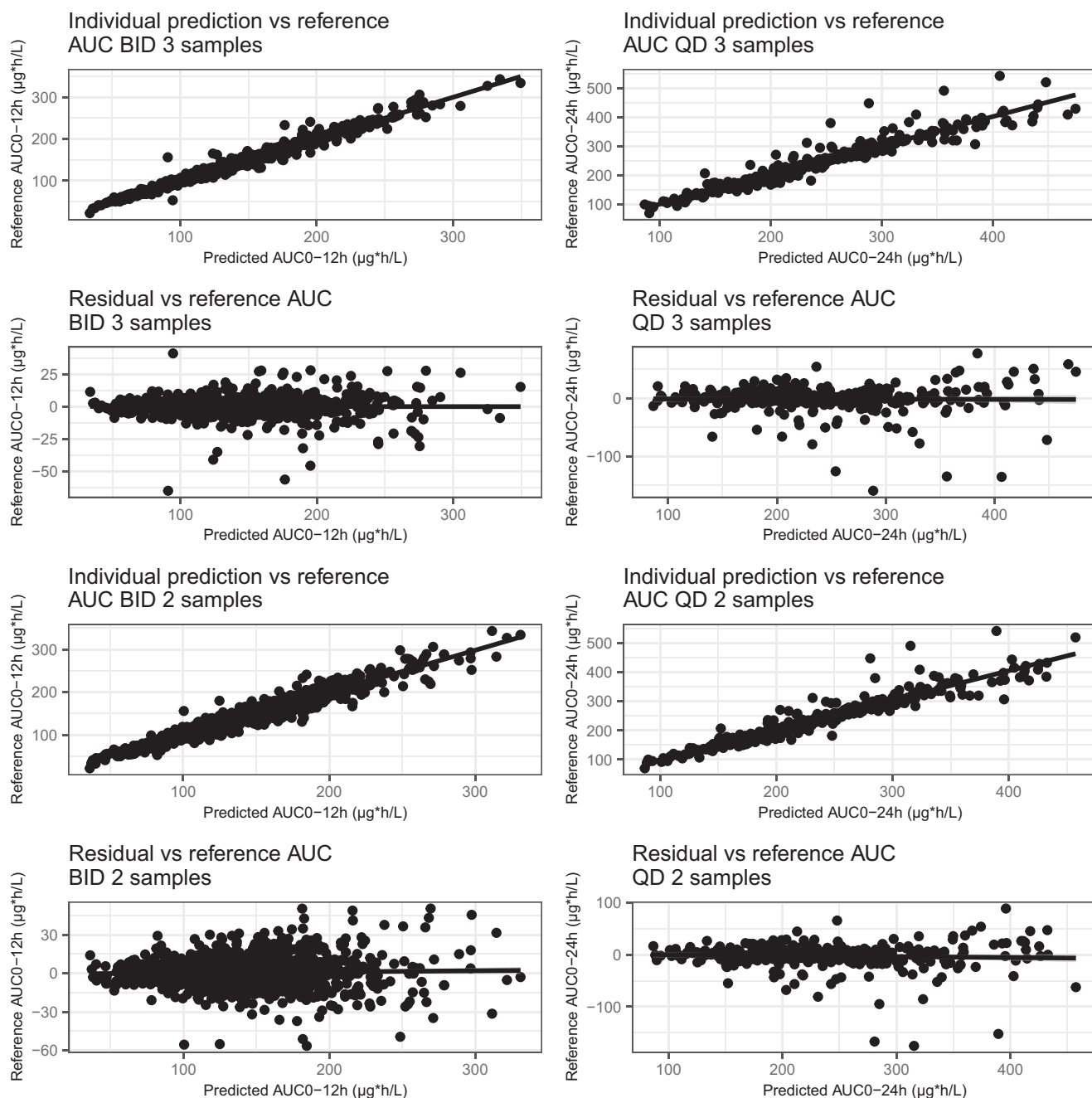


Figure 1 Scatter plots and residual plots of machine learning predicted vs. reference interdose AUCs in the test set, for TAC b.i.d. and q.d. using two or three tacrolimus blood concentrations. AUC, area under the curve; AUC_{0-12h} , area under the curve from 0 to 12 hours; AUC_{0-24h} , area under the curve from 0 to 24 hours; TAC, tacrolimus.

in drug handling by the body and in its variability using a modeling approach, the goal of ML models is accuracy-centered whatever the number and types of covariates. Thus, ML models can be successfully used as estimators (and probably give better results than PopPK combined to MAP-BE) but with the trade-off of no interpretability. Indeed, no visual representation is provided with the estimated AUC. Finally, the mathematics behind each method are very different with no PK parameter estimation for ML.

We acknowledge that developing our ML model, using as reference TAC AUCs estimated by our ISBA expert system with MAP-BE and three samples, was suboptimal because the “reference” used is itself an estimation of the true AUC. However, it allowed us to access several thousand cases as compared with only a few hundred if we had chosen the full PK profiles with trapezoidal AUCs as the references. Actually, we chose to keep the full profiles as independent datasets to best evaluate the performance of our models. Moreover, as in ISBA, several PopPK models and

Table 3 Performances of ML models and MAP-BE based on three samples of LSS to estimate reference AUCs obtained from full PK profiles using the trapezoidal rule

Study	Method	Relative MPE, %	Relative RMSE, %	Bias out of $\pm 20\%$ n (%)	Bias out of $\pm 10\%$ n (%)
TAC b.i.d. kidney 1 ¹¹ (n = 137)	Xgboost 2 concentrations	-0.6	9.0	3 (2.2)	42 (30.6)
	Xgboost 3 concentrations	2.9	8.1	2 (1.4)	26 (19.0)
	MAP-BE 3 concentrations	5.1	9.6	6 (4.4)	37 (27.0)
TAC b.i.d. kidney 2 ⁷ (n = 34)	Xgboost 2 concentrations	-1.0	9.1	2 (5.9)	9 (26.5)
	Xgboost 3 concentrations	-0.1	7.3	0 (0)	7 (20.6)
	MAP-BE 3 concentrations	2.8	9.1	1 (2.9)	9 (26.5)
TAC q.d. kidney ⁷ (n = 41)	Xgboost 2 concentrations	-0.6	10.8	3 (7.3)	14 (34.1)
	Xgboost 3 concentrations	-2.0	11.1	3 (7.3)	14 (34.1)
	MAP-BE 3 concentrations	2.1	13.2	6 (14.6)	16 (39.0)
TAC b.i.d. liver ¹⁴ (n = 68)	Xgboost 2 concentrations	3.3	12.9	7 (10.3)	30 (44.1)
	Xgboost 3 concentrations	4.1	9.9	3 (4.4)	14 (20.6)
	MAP-BE 3 concentrations	-3.4	11.2	6 (8.8)	18 (26.5)
TAC q.d. liver ¹⁴ (n = 91)	Xgboost 2 concentrations	-0.4	12.3	10 (11.0)	35 (38.5)
	Xgboost 3 concentrations	-0.9	11.5	10 (11.0)	31 (34.1)
	MAP-BE 3 concentrations	-3.5	15.7	17 (18.7)	49 (53.8)
TAC BID heart ^{15,16} (n = 47)	Xgboost 2 concentrations	-0.4	9.7	2 (4.2)	13 (27.7)
	Xgboost 3 concentrations	5.4	10.8	3 (6.4)	18 (38.3)
	MAP-BE 3 concentrations	-0.7	9.1	1 (2.1)	13 (27.7)

AUC, area under the curve; LSS, limited sampling strategy; MAP-BE, maximum *a posteriori* Bayesian estimation; ML, machine learning; MPE, mean prediction error; PK, pharmacokinetic; RMSE, root mean square error; TAC, tacrolimus.

MAP-BEs are used (for early or late post-transplant periods, different types of transplants, etc...), we wanted to check whether a single ML model taking into account the type of transplant as a covariate, but developed in a database containing mostly renal transplant patients (> 95%) could be applied successfully to other types of transplants. The results were above our expectations, with performances in heart or liver transplant recipients as excellent as in kidney transplant recipients, whatever the post-transplantation period.

Concerning the ML approach, we observed that the performances were very similar in the test set and in the training set (showing no overfitting) thanks to the resampling and cross-validation process that we used. Nevertheless, some PK profiles in the external dataset

were poorly predicted, but, in most cases, the MAP-BE predictions were unsatisfactory too. Visual inspection often showed unexpected, not to say unlikely profiles, for instance with C₀ higher than the other concentrations or much higher concentrations at 4 or 6 hours than in the first 3 hours postdose, or nonsteady-state PK profiles with C₂₄ hours much higher than C₀. Fortunately, the number and percentage of poor predictions (i.e., of $\pm 20\%$ of the reference) were quite low, indicating that these ML models, for which a patent application has been filed, can be used for routine AUC estimation. They will soon be made available to the transplant community through a web interface. In the meantime, a temporary shiny application (https://jbwoillard.shinyapps.io/App-6_tacro_ml/) has been built to test the models (for research purposes only).

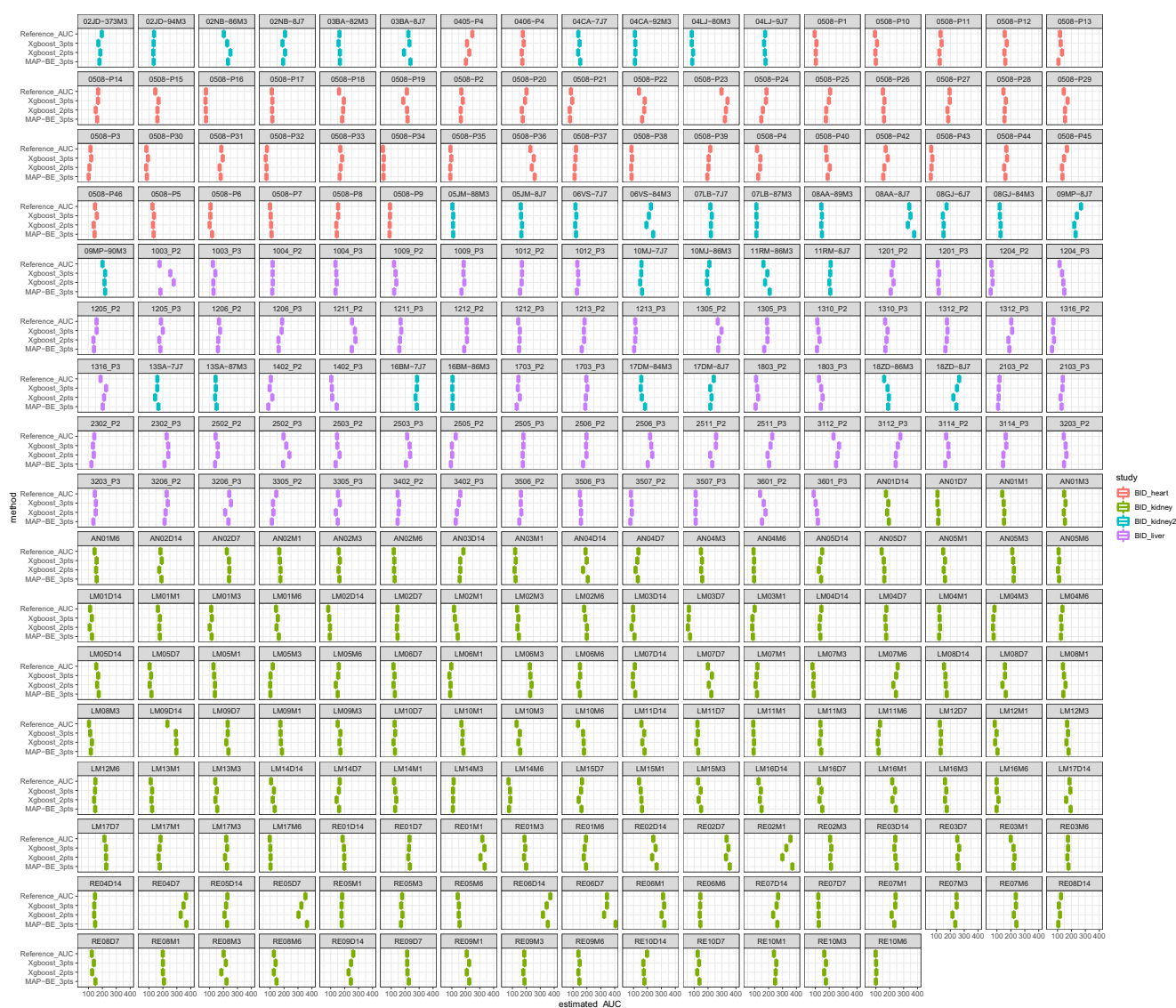


Figure 2 Comparison of the different methods of AUC estimation for TAC b.i.d. in the external validation studies. MAP-BE based on the 0, 1, and 3-hour limited sampling strategy and the parametric PopPK models available in ISBA. AUC, area under the curve; ISBA, Immunosuppressant Bayesian Dose Adjustment; MAP-BE, maximum *a posteriori* Bayesian estimation; PopPK, population pharmacokinetic; TAC, tacrolimus.

The recent results showing the stability of the AUC/C0 ratio within the post graft period^{16,7} should further increase their use, in order to define a C0 target for each patient depending on the individual AUC/C0 ratio, post-transplant period and intended use (minimization, standard, or intensification strategy).

The fair correlation ($r > 0.8$) observed between AUC and C0 in our large dataset is not different from the literature results. Indeed, in the present study, we reported the correlation coefficient (Supplementary Figure S1a,b), whereas most of the studies in the literature reported the determination coefficient (r^2). For example, the weighted average of the determination coefficients reported for TAC q.d. at three periods post-transplantation in Marquet *et al.*²² (including some of the patients used in the present paper) is

$r^2 = 0.71$, corresponding to $r = 0.84$, which is similar to the value found in the present study.

Following the publication of a recent PopPK model for everolimus AUC estimation, integrating previous exposure,²³ we checked whether taking into account anteriorities could improve AUC estimation. For this, we selected the patients with at least two requests (data not shown). However, probably due to the lower number of profiles available for the development of the ML models, the performances were much poorer than those reported above.

In conclusion, we developed Xgboost ML models allowing the accurate estimation of TAC interdose AUC and validated them in independent groups of kidney, liver, or heart transplant recipients. Similar models using the same approach are being built for

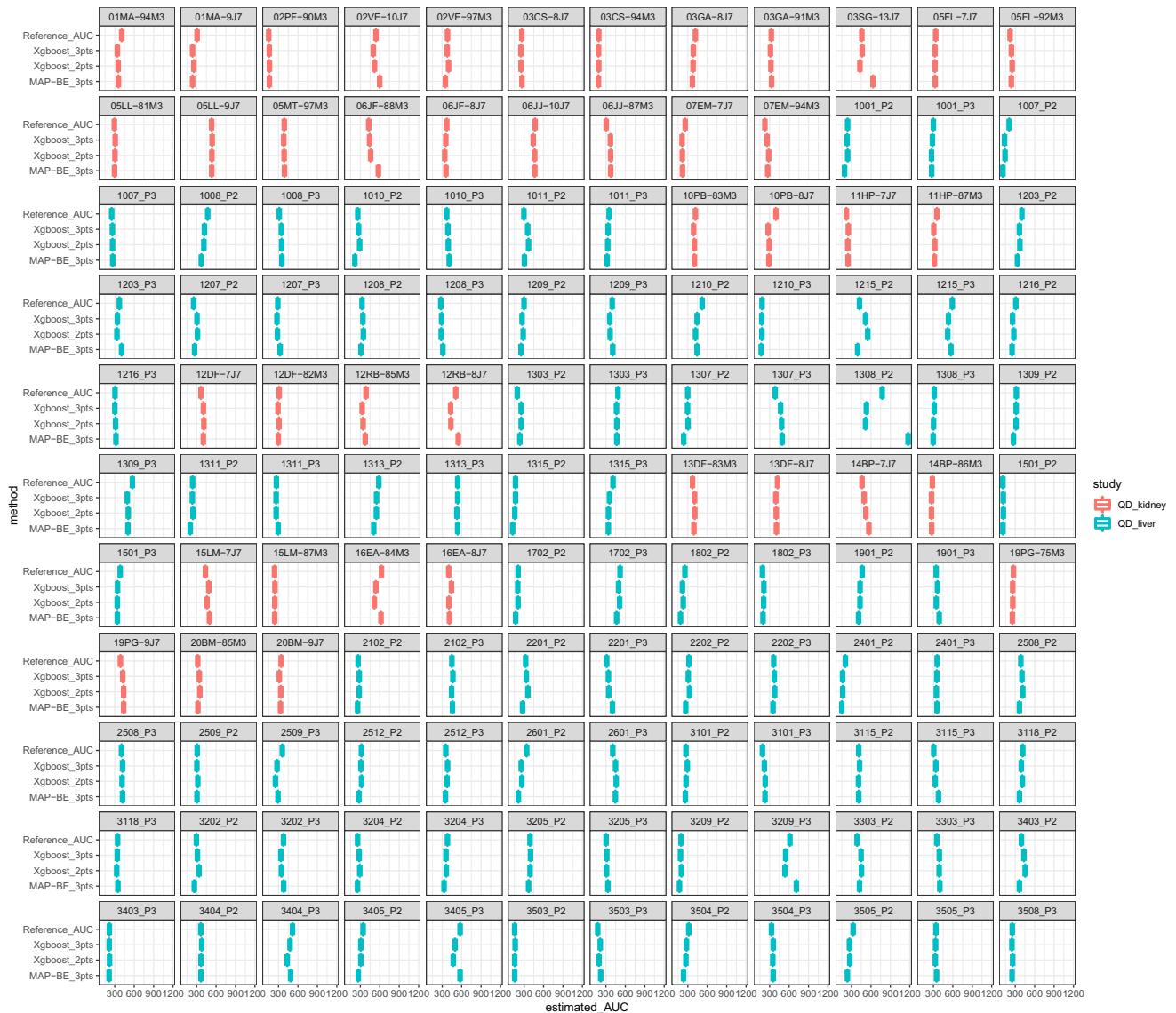


Figure 3 Comparison of the different methods of AUC estimation for TAC q.d. in the external validation studies. MAP-BE based on the 0, 1, and 3-hour limited sampling strategy and the parametric PopPK models available in ISBA. AUC, area under the curve; ISBA, Immunosuppressant Bayesian Dose Adjustment; MAP-BE, maximum a posteriori Bayesian estimation; PopPK, population pharmacokinetic; TAC, tacrolimus.

the other immunosuppressants and will soon be implemented in a dedicated web interface.

SUPPORTING INFORMATION

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

ACKNOWLEDGMENTS

The authors thank Ms. Karen Poole for manuscript editing. We thank the patients whose data were used in this study and the physicians and clinical pharmacologists for their use of the ISBA expert system.

FUNDING

No funding was received for this work.

CONFLICT OF INTEREST

P.M. has received consulting fees and travel and research grants from Sandoz and Chiesi. All other authors declared no competing interests for this work.

AUTHOR CONTRIBUTIONS

J.B.W. and P.M. wrote the manuscript. J.B.W. designed the research. J.B.W., M.L., J.D., and P.M. performed the research. J.B.W., M.L., J.D., and P.M. analyzed the data.

© 2020 The Authors *Clinical Pharmacology & Therapeutics* © 2020 American Society for Clinical Pharmacology and Therapeutics

1. Brunet, M. *et al.* Therapeutic drug monitoring of tacrolimus-personalized therapy: second consensus report. *Ther. Drug Monit.* **41**, 261–307 (2019).
2. Andrews, L.M. *et al.* Pharmacokinetic considerations related to therapeutic drug monitoring of tacrolimus in kidney transplant patients. *Expert Opin. Drug Metab. Toxicol.* **13**, 1225–1236 (2017).
3. Woillard, J.-B., Saint-Marcoux, F., Debord, J. & Åsberg, A. Pharmacokinetic models to assist the prescriber in choosing the best tacrolimus dose. *Pharmacol. Res.* **130**, 316–321 (2018).
4. ISBA website. <<https://pharmaco.chu-limoges.fr/>>. Accessed October 14, 2020.

5. Saint-Marcoux, F., Woillard, J.-B., Jurado, C. & Marquet, P. Lessons from routine dose adjustment of tacrolimus in renal transplant patients based on global exposure. *Ther. Drug Monit.* **35**, 322–327 (2013).
6. Marquet, P. *et al.* Pharmacokinetic therapeutic drug monitoring of Advagraf® in more than 500 adult renal transplant patients, using an expert system online. *Ther. Drug Monit.* **40**, 285–291 (2018).
7. Marquet, P. *et al.* 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**, 84–94 (2018).
8. Badillo, S. *et al.* An introduction to machine learning. *Clin. Pharmacol. Therapeut.* **107**, 871–885 (2020).
9. Chen, T., Guestrin, C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16, 785–794 (2016). <https://doi.org/10.1145/2939672.2939785>.
10. Woillard, J.-B. *et al.* A machine learning approach to estimate the glomerular filtration rate in intensive care unit patients based on plasma iohexol concentrations and covariates. *Clin. Pharmacokinet.* <https://doi.org/10.1007/s40262-020-00927-6>
11. Benkali, K. *et al.* Tacrolimus population pharmacokinetic-pharmacogenetic analysis and Bayesian estimation in renal transplant recipients. *Clin. Pharmacokinet.* **48**, 805–816 (2009).
12. Saint-Marcoux, F., Debord, J., Undre, N., Rousseau, A. & Marquet, P. Pharmacokinetic modeling and development of Bayesian estimators in kidney transplant patients receiving the tacrolimus once-daily formulation. *Ther. Drug Monit.* **32**, 129–135 (2010).
13. Woillard, J.-B., Debord, J., Monchaud, C., Saint-Marcoux, F. & Marquet, P. Population pharmacokinetics and Bayesian estimators for refined dose adjustment of a new tacrolimus formulation in kidney and liver transplant patients. *Clin. Pharmacokinet.* **56**, 1491–1498 (2017).
14. Riff, C., Debord, J., Monchaud, C., Marquet, P. & Woillard, J.-B. Population pharmacokinetic model and Bayesian estimator for 2 tacrolimus formulations in adult liver transplant patients. *Br. J. Clin. Pharmacol.* **85**, 1740–1750 (2019).
15. Fruit, D. *et al.* Ciclosporin population pharmacokinetics and Bayesian estimation in thoracic transplant recipients. *Clin. Pharmacokinet.* **52**, 277–288 (2013).
16. Woillard, J.-B. *et al.* Mycophenolic mofetil optimized pharmacokinetic modelling, and exposure-effect associations in adult heart transplant recipients. *Pharmacol. Res.* **99**, 308–315 (2015).
17. Schloerke, B. *et al.* GGally: Extension to 'ggplot2' <<https://CRAN.R-project.org/package=GGally>> (2020).
18. Tidymodels. <<https://www.tidymodels.org/>>. Accessed October 14, 2020.
19. Denney, B., Buckeridge, C. & Duvvuri, S. PKNCA: Perform Pharmacokinetic Non-Compartmental Analysis <<https://CRAN.R-project.org/package=PKNCA>> (2020).
20. Brooks, E., Tett, S.E., Isbel, N.M. & Staatz, C.E. Population pharmacokinetic modelling and Bayesian estimation of tacrolimus exposure: is this clinically useful for dosage prediction yet? *Clin. Pharmacokinet.* **55**, 1295–1335 (2016).
21. Applied Predictive Modeling. Applied Predictive Modeling <<http://appliedpredictivemodeling.com>>. Accessed July 16, 2020.
22. Marquet, P. *et al.* Pharmacokinetic therapeutic drug monitoring of Advagraf in more than 500 adult renal transplant patients, using an expert system online. *Ther. Drug Monit.* **40**, 285–291 (2018).
23. Zwart, T.C. *et al.* Model-informed precision dosing of Everolimus: external validation in adult renal transplant recipients. *Clin. Pharmacokinet.* <https://doi.org/10.1007/s40262-020-00925-8>