



ARTICLE

Machine learning algorithms to estimate everolimus exposure trained on simulated and patient pharmacokinetic profiles

Marc Labriffe^{1,2} | Jean-Baptiste Woillard^{1,2} | Jean Debord^{1,2} | Pierre Marquet^{1,2}

¹Pharmacology & Transplantation, INSERM U1248, Université de Limoges, Limoges, France

²Department of Pharmacology, Toxicology and Pharmacovigilance, CHU de Limoges, Limoges, France

Correspondence

Pierre Marquet, Department of Pharmacology, Toxicology and Pharmacovigilance, University Hospital of Limoges, CBRS, 2 rue Bernard Descottes, 87000 Limoges, France.
Email: pierre.marquet@unilim.fr

Funding information

No funding was received for this work.

Abstract

Everolimus is an immunosuppressant with a small therapeutic index and large between-patient variability. The area under the concentration versus time curve (AUC) is the best marker of exposure but measuring it requires collecting many blood samples. The objective of this study was to train machine learning (ML) algorithms using pharmacokinetic (PK) profiles from kidney transplant recipients, simulated profiles, or both types, and compare their performance for everolimus AUC_{0-12h} estimation using a limited number of predictors, as compared to an independent set of full PK profiles from patients, as well as to the corresponding maximum a posteriori Bayesian estimates (MAP-BE). XGBoost was first trained on 508 patient interdose AUCs estimated using MAP-BE, and then on 500–10,000 rich interdose PK profiles simulated using previously published population PK parameters. The predictors used were: predose, ~1 h, and ~2 h whole blood concentrations, differences between these concentrations, relative deviations from theoretical sampling times, morning dose, patient age, and time elapsed since transplantation. The best results were obtained with XGBoost trained on 5016 simulated profiles. AUC estimation achieved in an external dataset of 114 full-PK profiles was excellent (root mean squared error [RMSE] = 10.8 µg*h/L) and slightly better than MAP-BE (RMSE = 11.9 µg*h/L). Using more profiles ($n = 10,035$) did not improve the ML algorithm performance. The contribution of mixing patient and simulated profiles was significant only when they were in balanced numbers, with ~500 for each (RMSE = 12.5 µg*h/L), compared with patient data alone (RMSE = 18.0 µg*h/L).

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

Assessing everolimus area under the concentration-time curve (AUC) requires either collecting many blood samples or using a pharmacokinetic (PK) model and Bayesian estimator with a few blood samples. Machine learning (ML) algorithms represent an alternative, provided they can be trained on large enough databases.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *CPT: Pharmacometrics & Systems Pharmacology* published by Wiley Periodicals LLC on behalf of American Society for Clinical Pharmacology and Therapeutics.

WHAT QUESTION DID THIS STUDY ADDRESS?

It evaluated the contribution and limits of simulated data to train ML models to estimate everolimus AUC_{0-12h} .

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

An optimal amount of simulated data ($n = 5016$ PK profiles) optimized XGBoost AUC_{0-12h} prediction, even rendering patient data useless, and yielded better performance than Bayesian estimation.

HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?

When limited data is available to train ML algorithms, simulations can be used. However, too many simulated data expose to overfitting, highlighting the need for independent patient datasets for external validation.

INTRODUCTION

Everolimus is an inhibitor of the mammalian target of rapamycin (mTOR) activity, in particular in lymphocytes.¹ It is a non-nephrotoxic drug that shows a synergistic immunosuppressive effect with calcineurin inhibitors (CNIs).^{2,3} It is characterized by a narrow therapeutic range and a large interindividual variability requiring concentration-based dose adjustments, similar to CNIs. Therapeutic drug monitoring (TDM) is therefore recommended—or, in certain countries, compulsory—for everolimus and, due to its high distribution in erythrocytes, dose individualization is generally based on trough whole blood concentrations.⁴

Two main markers are currently available to individually adjust everolimus dose: the trough blood level (C₀), which is widely used for practical and economic reasons, although it has inconsistently been associated with clinical outcomes, and the interdose area under the curve (AUC_{0-12h}),⁵ which reflects overall exposure and is theoretically a better predictor of the drug pharmacodynamics. Kovarik et al.⁶ actually found exposure-response relationships between everolimus AUC and the incidence of thrombocytopenia, hypertriglyceridemia, and hypercholesterolemia. In the same study, he measured the interindividual (coefficient of variation [CV] = 85.4%) and the intra-individual (interoccasion) (CV = 40.8%) variability of the AUC, suggesting that TDM is needed and feasible, respectively. Another study showed that the interindividual variability is larger for everolimus C₀ than AUC_{0-12h} (CV% = 55% and 31%, respectively), as is the intra-individual variability (45% and 27%, respectively).⁷ However, the interdose AUC is more difficult to measure than C₀ because it requires collecting and analyzing many samples. In practice, everolimus AUC has been estimated using population pharmacokinetic (PopPK) models and maximum a posteriori Bayesian estimation (MAP-BE)

based on limited sampling strategies.⁸ In 2005, the Immunosuppressant Bayesian Dose Adjustment (ISBA) expert system and website (<https://abis.chu-limoges.fr/login>) were launched to share tools able to estimate the interdose AUC of immunosuppressants using MAP-BE on the basis of three blood samples and some patient characteristics (type of graft, age, post-transplantation period, and drug measurement assay).⁹ In 2018, a new everolimus model was made available on ISBA, where each request posted is validated in less than 48 h by a trained pharmacologist.

Over the last 2 decades, machine learning (ML) has been successfully used in many applications in pharmacology, thanks to the huge and ever-increasing amount of data and computational power as well as to the improvement of learning algorithms.^{9,10} Extreme gradient boosting (XGBoost) is an ML algorithm where simple regression trees are iteratively built by finding split values among all input variables to minimize prediction error. The iterative process constructs an additional regression tree of the same structure to minimize the residual errors of the previous regression tree.¹¹ We found that XGBoost was particularly suited to estimate the AUC of other immunosuppressive drugs using limited sampling strategies and covariates.^{12,13} For tacrolimus in particular, we even trained in parallel such algorithms on massive simulated data rather than actual patient data, showing again better performance than usual MAP-BE.¹⁴ This was an important finding because, for many drugs such as everolimus, there is not enough patient data available to train ML algorithms. However, the full potential of simulated data combined with patient data has not been explored yet. Is there an optimal number of simulations? Is patient data, even in rather low volume, still useful if a potentially infinite number of PK profiles can be simulated? If so, is there an optimal balance between patient and simulated data?

The objective of this study was to compare different combinations of patient and simulated PK profiles for the training of an XGBoost algorithm able to estimate everolimus AUC_{0-12h} using a limited number of predictors. The true performance was evaluated in external validation datasets of full concentrations profiles from kidney transplant recipients, and then compared to that of MAP-BE in the same datasets.

METHODS

Patients and actual data

The everolimus AUC estimation and dose recommendation requests received on our ISBA website since 2018 for recipients of a renal transplant were extracted and cleaned using the Tidyverse framework. Data collection was approved by the regional ethics committee, and all patients gave their informed consent to participate in the study (EudraCT number 2006–0068 32–23 and 2009–0135 41–28). Blood was collected at three sampling times at least: predose (C0), ~60 min (30–100 min, C1), and ~120 min (115–220 min, C2) after drug intake. Everolimus blood levels were measured using high-performance liquid chromatography coupled to tandem mass spectrometry. The other predictors available were the morning dose of everolimus, the time elapsed between transplantation and everolimus blood sampling, and patient age. The code used for data cleaning and data that support the findings of this study are available upon request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Simulated data

We used the parameters of a previously published pharmacokinetic (PK) model developed for everolimus in a population of adult kidney transplant recipients.¹⁶ PK profiles were simulated at steady-state over a 12-h interval, uniformly for different drug doses (0.5, 1, 1.25, 1.5, 1.75, 2, 2.25, 2.5, 2.75, 3, 3.25, 3.5, 4, and 4.5 mg), using the *mrsgolve* R package.¹⁷ The proportional error was significantly diminished to 0.01% (as compared to 13.9% in the original paper) to obtain less noisy, smoother simulated PK profiles, deliberately neglecting measurement errors to optimize algorithm training on unaltered, “true” AUC values. However, at the prediction step, gaussian noise with mean = 1.2% and SD = 3.9% (minimum = 0.0% and maximum = 130.9%) was randomly added to the simulated C1 and C2 sampling times, using the *sdcMicro* R package.¹⁸ The aim was to introduce uncertainty on

input data so as to observe the algorithm prediction performance in more realistic conditions. In addition, we kept the interindividual variability of the PK parameters described in the initial study (eta values), as well as that brought by the most important covariate, the ideal body weight. Indeed, in the original model, apparent volume of distribution of the central compartment after oral administration (V1/F) was a function of the ideal body weight. We simulated ideal body weight values using a truncated random normal distribution with mean \pm SD (minimum–maximum) = 68 \pm 7.5 (52–83) kg (in accordance with the original article).

Datasets and analysis strategy

The present study used supervised learning from different training datasets to predict the interdose AUC, whose reference values had been obtained either through our ISBA expert system using MAP-BE and three everolimus blood concentrations for kidney transplant patients; or using the trapezoidal rule with the PKNCA R package for the simulated profiles.¹⁹ In order to maximize the diversity of the training sets when patient and simulated data were mixed, the simulated profiles were not obtained using our in-house PK model but using a model from the literature.¹⁶ We trained many ML algorithms, based on patient, simulated, or mixed data. Each training dataset was used in turn to build an XGBoost algorithm, tune the hyperparameters, and evaluate its performance by a single 10-fold cross-validation (random partition of the training set into 10 parts). The algorithms were then evaluated on independent subsets of the training sets by calculating the root mean square error (RMSE; expressed in $\mu\text{g}\cdot\text{h}/\text{L}$) between the estimated and reference AUCs. Finally, the different algorithms were comparatively evaluated using two independent datasets of everolimus full PK profiles in kidney transplant recipients.

Feature engineering

Everolimus blood concentrations (whether actually measured or simulated) were divided into three theoretical time classes: concentrations at trough (C0 sampled at $t = 0$ min), 1 h (C1 sampled between 30 and 100 min), and 2 or 3 h (C2 sampled between 115 and 220 min). New variables were drawn for times 1 and 2 h corresponding to the relative deviation with respect to the theoretical times. For instance, if the sampling time was 1.06 h, the relative time difference with the theoretical time 1 h was $(1.06 - 1)/1 = 0.06$. Other predictors corresponding to the differences between the concentrations C1–C0, C1–C2, and

C2–C0 were created to add information about potentially delayed absorption peaks. Finally, the features tested as predictors of the interdose AUCs in the training set from actual patients were: patient age, time elapsed between transplantation and everolimus blood sampling, everolimus morning dose, everolimus concentrations at times 0, 1 h, and 2 h, relative deviation from the theoretical times, and differences between concentrations. In the training set of simulated profiles, as well as in the mixed training sets, the potential predictors were limited to: everolimus concentrations at times 0, 1 h, and 2 h, relative deviation from the theoretical times, differences between concentrations, and everolimus morning dose.

Exploratory data analyses

A correlation matrix and scatterplots were drawn to explore the correlations between AUC and predictors in the actual patient dataset, using the GGally R package.²⁰

Preprocessing of the data

For all the ML analyses, the tidymodels framework was used. 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 between training datasets (75%) and test datasets (25%) was performed by random selection of patients (or simulated cases).

Training of XGBoost algorithms

The algorithms were tuned by searching the hyperparameter combination associated with the lowest RMSE and highest R^2 between estimated and reference AUC values, using 10-fold cross-validation. In brief, the best combination of hyperparameters was investigated in 90% of each training dataset in turn (analysis subset) and evaluated in the remaining 10% (assessment subset) and this process was repeated 10 times by circular permutation. The hyperparameters 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 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 adapted from iteration-to-iteration (learn_rate, between 0 and 0.08). In a second time, the best hyperparameter combinations were evaluated by means of another set of 10-fold cross-validation to

assess the mean RMSE and R^2 and their SDs in the corresponding training dataset and draw the scatter plots of estimated versus reference AUC. Finally, AUC estimation was evaluated in the respective test datasets by calculating RMSE, R^2 normalized RMSE (RMSE divided by the mean of reference AUCs), relative mean prediction error (MPE), as well as through the number and proportion of estimates with absolute MPE greater than 20%. The code used for the simulation of PK profiles and XGBoost training is provided as [Supplementary Text S1](#).

External evaluation of machine learning AUC estimates by comparison with full PK profiles, and comparison with maximum a posteriori Bayesian estimates

The independent validation dataset comprised full PK profiles from the PIGREC trial (NCT00812786; 0, 0.33, 0.66, 1, 1.5, 2, 3, 4, 6, 8, 9, and 12 h postdose) and from the Everold trial (NCT01028092; 0, 0.33, 0.75, 1, 1.66, 2, 4, 6, 8, 10, and 12 h postdose). Concentrations at 0, 1, and 2 h, everolimus dose, blood sampling times, and time elapsed between transplantation and everolimus blood sampling were extracted from the independent PK databases to predict the AUC using the ML algorithms as compared with the MAP-BE used in ISBA. The full concentration profiles were used to calculate the trapezoidal AUC (chosen as the reference) using the DescTools package.²¹ The performance of the ML algorithms and of MAP-BE was evaluated by comparing the estimated AUCs to the trapezoidal AUCs in terms of RMSE and relative MPE, and the proportion of bias out of the $\pm 20\%$ interval. Additionally, the scatter plots of predicted versus reference AUCs and residuals versus predicted AUCs were drawn on the same graph for visual comparison of the different approaches.

RESULTS

Patients and data

The cleaned dataset extracted from ISBA used as patient data training set consisted of 508 everolimus AUC_{0-12h} from 177 patients. The characteristics of the training and test sets of patient data are reported in [Table 1](#). The median AUC_{0-12h} was 101 (interquartile range [IQR] 73, 142) $\mu\text{g}\cdot\text{h}/\text{L}$. The independent validation patient dataset comprised 114 PK profiles of 10–12 samples and in this group the coefficient of determination R^2 between C0 and the reference trapezoidal AUC_{0-12h} ($n = 114$) was only 0.776.

TABLE 1 Characteristics of the features used for the training and validation of the first XGBoost algorithm based on 508 patient pharmacokinetic profiles

	Train set (<i>n</i> = 381)	Test set (<i>n</i> = 127)	External validation set (<i>n</i> = 114)
Time between transplantation and tacrolimus blood concentrations, months	3.95 [1.97, 11.84]	3.95 [1.97, 11.84]	14.76 [4.97, 105.90]
AUC _{0-12h} , µg*h/L	102 [74, 142]	101 [73, 145]	96 [69, 125]
Patient age, year	47 [35, 57]	47 [39, 57]	50 [40, 59]
Morning dose, mg	1.50 [0.75, 1.50]	1.50 [0.75, 1.50]	1.00 [0.56, 2.00]
Trough level (C ₀), µg/L	5.4 [3.7, 7.9]	5.5 [3.5, 8.6]	5.4 [3.9, 7.4]
Concentration at 1 h; C ₁ , µg/L	14.2 [9.2, 20.5]	13.3 [8.7, 19.9]	15.2 [10.2, 21.0]
Concentration at 2 h; C ₂ , µg/L	13.2 [9.0, 18.4]	12.9 [9.2, 18.3]	11.5 [8.4, 15.6]
Deviation from the 1-h theoretical time, %	0 [0, 0]	0 [0, 0]	0 [0, 0]
Deviation from the 2-h theoretical time, %	0 [0, 4]	0 [0, 2]	0 [0, 0]
Concentration difference between C ₁ and C ₀	8.5 [4.2, 13.1]	7.2 [3.3, 11.9]	9.5 [5.6, 14.1]
Concentration difference between C ₁ and C ₂	1.5 [-1.3, 4.6]	0.3 [-2.1, 3.6]	3.6 [0.6, 6.4]
Concentration difference between C ₂ and C ₀	7.1 [4.7, 10.8]	7.3 [4.3, 10.0]	5.8 [3.9, 9.1]
Reference AUCs: number of samples	3	3	10–12
Reference AUCs: method used	Same MAP-BE		Trapezoidal rule

Note: Medians [interquartile ranges] are presented here.

Abbreviations: AUC, area under the curve; ISBA, Immunosuppressant Bayesian Dose Adjustment; MAP-BE, maximum a posteriori Bayesian estimation; XGBoost, extreme gradient boosting, an optimized gradient boosting machine learning method.

Exploratory data analyses

The correlation matrix between everolimus AUC_{0-12h} and predictors from patient PK profiles is presented in Figure S1, showing that the strongest correlations (>0.8) were between AUC_{0-12h} and C₀ or C_{2h}.

XGBoost algorithms, training, and test sets

The best-tuned hyperparameter values for each algorithm are presented in Table S1. The results in the training sets obtained after 10-fold cross-validation and in the respective test sets are shown in Table 2. Among the test sets, the lowest RMSE (6.7 µg*h/L) was obtained using 10,035 simulated profiles.

External evaluation versus the trapezoidal AUC in an independent dataset

The best results (RMSE = 10.8 µg*h/L) were obtained using 5016 simulated profiles without patient data (Table 2). Focusing on the simulated profiles, Figure 1 presents the performances of XGBoost in the training and the external validation datasets according to the number of simulations used (including additional models trained on *n* = 250, 500 or 15,051 simulated profiles). It shows that

the higher the number, the better the performance in the training set, whereas in the independent dataset RMSE followed a U shape curve with a minimal value for 5016 simulations.

Figure 2 presents the scatter plots and residual plots of estimated versus reference AUCs in the external validation dataset for four models: the algorithm based on the patient data only (*n* = 508); the best model mixing patient and simulated data (*n* = 508 and 500, respectively); the best model using only simulated data (*n* = 5016), and MAP-BE for comparison. There was no systematic bias. For our best model (5016 simulations), we also explored the possibility of adding more variability in the sampling times (Table S3), and it negatively affected the results in the test set, and a little bit in the validation set.

DISCUSSION

In this work, based on our previous experience with ML tools to estimate overall exposure to other immunosuppressive drugs,¹³⁻¹⁵ we used XGBoost ML algorithms to estimate the interdose AUC of everolimus in renal transplant recipients. Because we had a more limited training dataset with actual patient data than with other immunosuppressants, we trained ML algorithms on patient, simulated, and mixed data and compared the estimates obtained in an independent database from kidney

TABLE 2 Performance of the XGBoost algorithms at estimating everolimus AUC_{0-12h} in the different sorts of training, testing, and external validation datasets

		Train set (n = 75%)	Test set (n = 25%)	External validation set (n = 114 full PK profiles)	
		XGBoost	XGBoost	XGBoost (n = 114)	MAP-BE (n = 94 ^a)
508 patient PK profiles	RMSE, µg*h/L	15.2	15.4	18.0	11.9
	Normalized RMSE (%)	13.5	13.8	17.2	11.2
	R ²	0.921	0.922	0.873	0.952
	Relative MPE (%)	1.9	4.5	4.5	3.0
	Number of MPE out of the ±20% interval n	41 (10.8%)	20 (15.7%)	17 (14.9%)	7 (7.4%)
500 simulated + 508 patient PK profiles	RMSE, µg*h/L	32.1	23.3	12.5	11.9
	Normalized RMSE (%)	24.9	17.2	11.9	11.2
	R ²	0.880	0.942	0.939	0.952
	Relative MPE (%)	-0.4	0.8	0.0	3.0
	Number of MPE out of the ±20% interval n	50 (6.6%)	26 (10.3%)	5 (4.4%)	7 (7.4%)
1003 simulated PK profiles	RMSE, µg*h/L	18.5	19.0	18.6	11.9
	Normalized RMSE (%)	12.6	12.8	17.8	11.2
	R ²	0.970	0.970	0.919	0.952
	Relative MPE (%)	1.7	1.6	9.4	3.0
	Number of MPE out of the ±20% interval n	39 (5.2%)	13 (5.2%)	22 (19.3%)	7 (7.4%)
1003 simulated + 508 patient PK profiles	RMSE, µg*h/L	19.2	10.7	14.1	11.9
	Normalized RMSE (%)	13.6	7.8	13.4	11.2
	R ²	0.967	0.986	0.924	0.952
	Relative MPE (%)	0.5	0.0	1.2	3.0
	Number of MPE out of the ±20% interval n	50 (4.4%)	17 (4.5%)	8 (7.0%)	7 (7.4%)
2508 simulated PK profiles	RMSE, µg*h/L	13.4	15.1	11.4	11.9
	Normalized RMSE (%)	8.6	10.2	10.9	11.2
	R ²	0.987	0.982	0.951	0.952
	Relative MPE (%)	0.1	0.1	1.4	3.0
	Number of MPE out of the ±20% interval n	8 (0.4%)	4 (0.6%)	8 (7.0%)	7 (7.4%)
2508 simulated + 508 patient PK profiles	RMSE, µg*h/L	12.5	14.7	12.2	11.9
	Normalized RMSE (%)	8.5	10.2	11.7	11.2
	R ²	0.987	0.981	0.942	0.952
	Relative MPE (%)	0.3	0.2	2.2	3.0
	Number of MPE out of the ±20% interval n	39 (1.7%)	17 (2.3%)	7 (6.1%)	7 (7.4%)
5016 simulated PK profiles	RMSE, µg*h/L	14.1	11.2	10.8	11.9
	Normalized RMSE (%)	9.3	7.3	10.3	11.2
	R ²	0.985	0.990	0.956	0.952
	Relative MPE (%)	0.1	0.1	1.6	3.0
	Number of MPE out of the ±20% interval n	9 (0.2%)	2 (0.2%)	7 (6.1%)	7 (7.4%)

(Continues)

TABLE 2 (Continued)

		Train set (<i>n</i> = 75%)	Test set (<i>n</i> = 25%)	External validation set (<i>n</i> = 114 full PK profiles)	
		XGBoost	XGBoost	XGBoost (<i>n</i> = 114)	MAP-BE (<i>n</i> = 94 ^a)
5016 simulated + 508 patient PK profiles	RMSE, $\mu\text{g}\cdot\text{h}/\text{L}$	11.1	9.2	12.7	11.9
	Normalized RMSE (%)	7.4	6.4	12.1	11.2
	R^2	0.990	0.992	0.939	0.952
	Relative MPE (%)	0.2	0.3	2.7	3.0
	Number of MPE out of the $\pm 20\%$ interval <i>n</i>	44 (1.1%)	16 (1.2%)	6 (5.3%)	7 (7.4%)
10,035 simulated PK profiles	RMSE, $\mu\text{g}\cdot\text{h}/\text{L}$	7.6	6.7	12.6	11.9
	Normalized RMSE (%)	5.0	4.3	12.1	11.2
	R^2	0.996	0.997	0.942	0.952
	Relative MPE (%)	0.0	0.0	-1.2	3.0
	Number of MPE out of the $\pm 20\%$ interval <i>n</i>	3 (0.0%)	0 (0.0%)	7 (6.1%)	7 (7.4%)
10,035 simulated + 508 patient PK profiles	RMSE, $\mu\text{g}\cdot\text{h}/\text{L}$	7.6	7.5	13.7	11.9
	Normalized RMSE (%)	5.0	4.9	13.1	11.2
	R^2	0.996	0.996	0.929	0.952
	Relative MPE (%)	0.1	0.1	2.6	3.0
	Number of MPE out of the $\pm 20\%$ interval <i>n</i>	41 (0.5%)	13 (0.5%)	9 (7.9%)	7 (7.4%)

Note: The performance of the MAP-BE actually used in the online ISBA expert system is displayed here, in the last column of the table, for comparison purposes.

Abbreviations: AUC, area under the curve; ISBA, Immunosuppressant Bayesian Dose Adjustment; MAP-BE, maximum a posteriori Bayesian estimation; MPE, mean prediction error; Normalized RMSE, root mean square error divided by the mean of reference AUCs; PK, pharmacokinetic; RMSE, root mean square error; XGBoost, extreme gradient boosting, an optimized gradient boosting machine learning method.

^aFor 20 profiles, MAP-BE could not be used because the morning dose was missing.

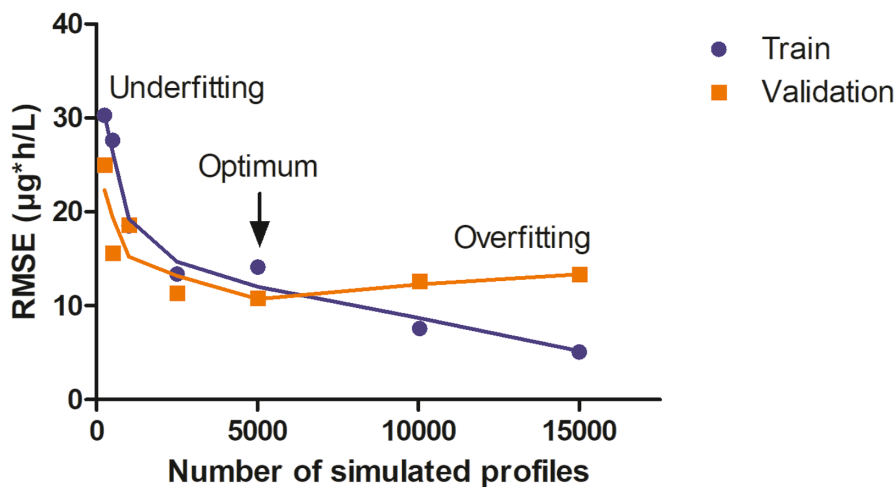
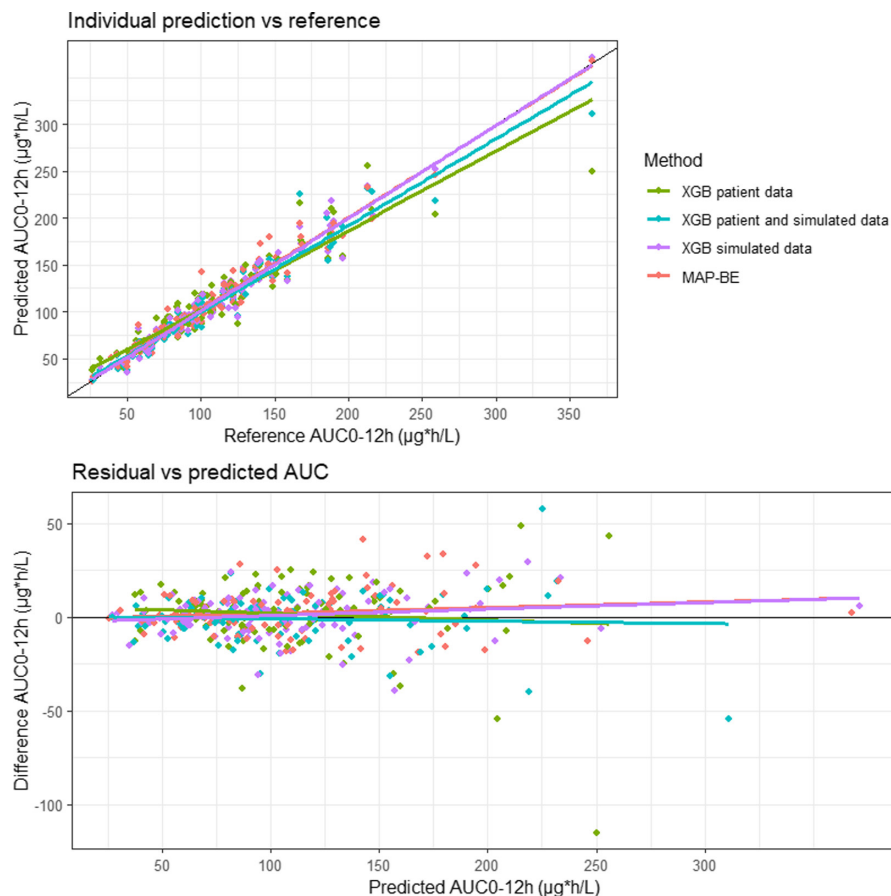


FIGURE 1 Plot of everolimus $\text{AUC}_{0-12\text{h}}$ prediction RMSE in the training (blue) and the external validation (orange) datasets, according to the number of simulations used to train the XGBoost algorithm. Points represent the performance of each XGBoost model, lines are a smoothed representation of trends. $\text{AUC}_{0-12\text{h}}$, 0–12-h area under the concentration-time curve; RMSE, root mean square error; XGBoost, extreme gradient boosting, an optimized gradient boosting machine learning method

FIGURE 2 Scatter plots and residual plots of machine learning predicted versus reference everolimus AUC_{0-12h} in the external validation dataset. The thin black line represents $y = x$. The colored lines were obtained by linear regression for each version of the XGBoost algorithm: in green, the model trained on patient data only ($n = 508$); in blue, the model trained on a balanced mix of patient ($n = 508$) and simulated ($n = 500$) data; in purple the best model, based on simulated data only ($n = 5016$); in red, the MAP-BE currently available through our online expert system ISBA. AUC, area under the curve; ISBA, Immunosuppressant Bayesian Dose Adjustment; MAP-BE, maximum a posteriori Bayesian estimation; XGBoost, extreme gradient boosting, an optimized gradient boosting machine learning method



transplant recipients with extensive-sampling with trapezoidal AUCs (reference AUCs) and MAP-BE AUC estimates based on a three-point limited sampling strategy, as used by our ISBA expert system. Different sizes of simulated training datasets were therefore compared based on several indicators, but primarily imprecision (i.e., RMSE). The performances of the ML algorithms trained on 5016 simulations without patient data yielded the best results. However, RMSE represents imprecision in the dataset and cannot be interpreted as the absolute error in a given patient. In our study, Figure 2 shows that absolute errors were lower for smaller reference AUCs.

The results in the training sets obtained after 10-fold cross-validation and in the respective test sets (Table 2) gradually improved with the number of simulated profiles (from 1003 up to the 10,035), yielding RMSE from 19.0 down to $6.7 \mu g \cdot h/L$ in the test sets. This apparently unlimited decrease in RMSE was a sign of overfitting. When we evaluated the mixing of simulated and patient data, we noted that adding 500 simulated to the 508 patient profiles seemed to penalize the model as compared to patient data alone (RMSE = $23.3 \mu g \cdot h/L$ and 15.4 in the test set, respectively). This is probably due to the wide diversity of profiles to be handled in this fairly small training set. In contrast, adding the 508 patients' profiles to 10,035

simulations increased the RMSE from 6.7 to $7.5 \mu g \cdot h/L$ in the test set. This was a second sign of model overfitting due to the huge number of simulated profiles in the training set.

In a second step, all the models were externally evaluated using as references the trapezoidal AUCs of an independent patient dataset. Adding the 508 patient AUCs to the 2508, 5016, or 10,035 simulated profiles for training did not improve the performances of the algorithm at this validation step. With 5000 simulations or less, prediction RMSE was roughly equivalent in the training and the validation datasets. With 10,000 simulations or more, RMSE was still decreasing in the training datasets, whereas it was slowly increasing in the independent patient dataset, showing overfitting to the parametric model used for simulations.

In addition, as shown in Figure 2, the two algorithms trained on patient data (even partially) did worse for the highest AUC values than those trained only on simulated data ($n = 5016$), or than MAP-BE. The lowest RMSE ($10.8 \mu g \cdot h/L$) in the external validation dataset (optimum) was obtained with 5016 simulated profiles. However, this precise optimal number of simulated profiles may not be generalized to all types of datasets, depending on the type of PK profiles, interindividual variability, data quality, etc.

It is worth noting that C0 values were not so well-correlated with the reference AUCs ($R^2 = 0.776$) calculated using the trapezoidal rule with greater than or equal to 10 samples from an unprecedented number of full PK profiles in kidney transplant recipients (114 profiles at median 14.8 [IQR = 5.0, 105.9] months post-transplantation in our independent validation database). Estimating AUC from C0 could therefore lead to great uncertainty (Table S2). Chan et al. compared C0 to incomplete AUC (AUC_{0.5}) in 92 patients at 1, 3, and 6 months post-transplantation and found R^2 values of 0.59, 0.81, and 0.83, respectively.⁵ Our results suggests that C0 is not as good a surrogate of the AUC, contrary to what was repeatedly claimed (e.g., Shipkova et al. TDM 2016⁷), and, in this respect, no better than for tacrolimus (Brunet et al. TDM 2019²²). Consequently, adjusting everolimus dose on the AUC rather than on C0, at least in certain patients or situations, might improve patient outcome.

One strength of the present study is to have trained and validated ML algorithms of everolimus exposure prediction on two independent datasets, both larger than those generally used in PK studies. In 2012, Moes et al.¹⁶ trained a PK model on 52 PK profiles and developed an MAP-BE with a 2-point LSS (C0 and C2) yielding $R^2 = 0.90$ in the training step, but it was not validated in an independent validation dataset. Similarly, Robertsen et al. proposed a PK model to describe everolimus PK in whole blood and in peripheral blood mononuclear cells,⁸ with a slightly better performance than our model trained on patient data (RMSE = 9.9% and 10.6%, respectively), but their Bayesian estimator was trained on a very small dataset ($n = 20$) and validated on an even smaller one ($n = 4$). In the study by Zwart et al.,²³ a model was built using only one observed concentration (C0) to estimate the AUC_{0-12h}, based on 322 PK profiles. The normalized RMSE was 17.4% and 16.3% at less than or equal to 6 and greater than or equal to 6 months post-transplant, respectively. However, because their data were collected from routine clinical care, their reference AUCs were calculated using less samples for each profile ($n = 4$, up to 7 in the best case) than in our study ($n = 10-12$).

Despite training on a rather large set of data (as compared with previous literature reports), the performance of our XGBoost algorithm was just slightly better than that of our previous MAP-BE. We even observed some absolute errors >20% in the validation dataset. These atypical cases were either overestimated because of unusual flat profiles or underestimated because of very high peak concentrations at 1 h (see Figure S2). These situations were probably better covered by the simulation approach, anticipating a large number of possibilities in an artificial

way (variability in doses, clearances, ideal body weights...). Figure 2 clearly shows improved predictions when AUC greater than 200 $\mu\text{g}\cdot\text{h}/\text{L}$. Because we receive many less requests for everolimus AUC_{0-12h} estimation on our ISBA expert system than for tacrolimus or mycophenolic, it would take many more years to reach the same amount of actual patient data to improve the ML algorithms and reach the same level of accuracy and precision as we did for these two drugs.^{13,14} For this reason, we chose to enhance ML by extending the training dataset through mixing patient and simulated profiles, or by relying only on a large number of simulations, as a test that could be extrapolated to other drugs. The last strategy was found to be the most powerful because it yielded very good performances across the whole range of AUC values in the external validation dataset.

The predictions obtained in the current work with ML are of similar quality to those obtained with MAP-BE. The latter only used the morning dose, three concentrations, and their respective times to provide an excellent estimation of the full trapezoidal AUC_{0-12h}, whereas the XGBoost algorithm trained on actual patient data used two more predictors, the time elapsed between transplantation and everolimus blood sampling, and patient age. However, these additional predictors presumably did not add much information to estimate AUC_{0-12h}, because algorithms built exclusively from simulated data without such covariates performed as well or better. In the context of MAP-BE, this has been initially described years ago by Sheiner et al.²⁴ for digoxin: the addition of covariates does not carry as much information as one additional plasma concentration. This is confirmed by the excellent prediction performance in the independent patient dataset, where age and post-transplantation time were not available. Moreover, long before this work, the MAP-BE used here as a comparator was built using ca. 30–40 profiles. Consequently, this study also illustrates a fundamental difference between XGBoost and MAP-BE: data-driven algorithms cannot be any better than the data available (e.g., here, the training set included very few patient PK profiles with AUC values <50 or >200 $\mu\text{g}\cdot\text{h}/\text{L}$) whereas compartmental PK models, if well designed, are expected to be valid even beyond values used to develop them.

We trained XGBoost ML algorithms on a large number of everolimus PK profiles simulated using a PopPK model from the literature and obtained better results than algorithms trained on a 10-fold smaller database of patient data, or on mixed databases of patient and simulated PK data. XGboost estimation based on three concentration-time points only (no other covariate) provided accurate and precise estimation of everolimus interdose AUC in a

large independent dataset of everolimus full PK profiles from kidney graft recipients. These algorithms can be used as alternatives to our previously developed Bayesian estimator available through our ISBA expert system (<https://abis.chu-limoges.fr/login>) and will soon be implemented in a dedicated web interface (for research purposes only), together with the recently published ML algorithms for tacrolimus and mycophenolate mofetil.

AUTHOR CONTRIBUTIONS

M.L. and P.M. wrote the manuscript. J.-B.W., P.M., and M.L. designed the research. J.D. designed one of the modeling programs. M.L. and J.-B.W. trained the algorithms. M.L., P.M., and J.-B.W. analyzed the data.

CONFLICTS OF INTEREST

The authors declared no competing interests for this work.

ORCID

Marc Labriffe  <https://orcid.org/0000-0001-5840-8904>

Jean-Baptiste Woillard  <https://orcid.org/0000-0003-1695-0695>

Jean Debord  <https://orcid.org/0000-0002-3309-1100>

Pierre Marquet  <https://orcid.org/0000-0001-7698-0760>

REFERENCES

- Schuler W, Sedrani R, Cottens S, et al. SDZ RAD, a new rapamycin derivative: pharmacological properties in vitro and in vivo. *Transplantation*. 1997;64(1):36-42. doi:10.1097/00007890-199707150-00008
- Nashan B, Curtis J, Ponticelli C, et al. Everolimus and reduced-exposure cyclosporine in de novo renal-transplant recipients: a three-year phase II, randomized, multicenter, open-label study. *Transplantation*. 2004;78(9):1332-1340. doi:10.1097/01.tp.0000140486.97461.49
- Vitko S, Tedesco H, Eris J, et al. Everolimus with optimized cyclosporine dosing in renal transplant recipients: 6-month safety and efficacy results of two randomized studies. *Am J Transplant*. 2004;4(4):626-635. doi:10.1111/j.1600-6143.2004.00389.x
- Kirchner GI, Meier-Wiedenbach I, Manns MP. Clinical pharmacokinetics of everolimus. *Clin Pharmacokinet*. 2004;43(2):83-95. doi:10.2165/00003088-200443020-00002
- Chan L, Hartmann E, Cibrik D, Cooper M, Shaw LM. Optimal everolimus concentration is associated with risk reduction for acute rejection in de novo renal transplant recipients. *Transplantation*. 2010;90(1):31-37. doi:10.1097/TP.0b013e3181de1d67
- Kovarik JM, Kahan BD, Kaplan B, et al. Longitudinal assessment of everolimus in de novo renal transplant recipients over the first post-transplant year: pharmacokinetics, exposure-response relationships, and influence on cyclosporine. *Clin Pharmacol Ther*. 2001;69(1):48-56. doi:10.1067/mcp.2001.112969
- Shipkova M, Hesselink DA, Holt DW, et al. Therapeutic drug monitoring of Everolimus: a consensus report. *Ther Drug Monit*. 2016;38(2):143-169. doi:10.1097/FTD.0000000000000260
- Robertsen I, Debord J, Åsberg A, Marquet P, Woillard JB. A limited sampling strategy to estimate exposure of Everolimus in whole blood and peripheral blood mononuclear cells in renal transplant recipients using population pharmacokinetic modeling and Bayesian estimators. *Clin Pharmacokinet*. 2018;57(11):1459-1469. doi:10.1007/s40262-018-0646-5
- Saint-Marcoux F, Woillard JB, Jurado C, Marquet P. Lessons from routine dose adjustment of tacrolimus in renal transplant patients based on global exposure. *Ther Drug Monit*. 2013;35(3):322-327. doi:10.1097/FTD.0b013e318285e779
- Badillo S, Banfai B, Birzele F, et al. An introduction to machine learning. *Clin Pharmacol Ther*. 2020;107(4):871-885. doi:10.1002/cpt.1796
- Woillard JB, Salmon Gandonnière C, Destere A, 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*. 2021;60(2):223-233. doi:10.1007/s40262-020-00927-6
- 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*. Association for Computing Machinery; 2016:785-794. doi:10.1145/2939672.2939785
- Woillard JB, Labriffe M, Debord J, Marquet P. Tacrolimus exposure prediction using machine learning. *Clin Pharmacol Ther*. 2020;30:361-369. doi:10.1002/cpt.2123
- Woillard JB, Labriffe M, Debord J, Marquet P. Mycophenolic acid exposure prediction using machine learning. *Clin Pharmacol Ther*. 2021;24:370-379. doi:10.1002/cpt.2216
- Woillard JB, Labriffe M, Prémaud A, Marquet P. Estimation of drug exposure by machine learning based on simulations from published pharmacokinetic models: the example of tacrolimus. *Pharmacol Res*. 2021;167:105578. doi:10.1016/j.phrs.2021.105578
- Moes DJAR, Press RR, den Hartigh J, van der Straaten T, de Fijter JW, Guchelaar HJ. Population pharmacokinetics and pharmacogenetics of everolimus in renal transplant patients. *Clin Pharmacokinet*. 2012;51(7):467-480. doi:10.2165/11599710-000000000-00000
- Elmokadem A, Riggs MM, Baron KT. Quantitative systems pharmacology and physiologically-based pharmacokinetic modeling with mrgsolve: a hands-on tutorial. *CPT Pharmacometrics Syst Pharmacol*. 2019;8(12):883-893. doi:10.1002/psp4.12467
- Templ M, Kowarik A, Meindl B. Statistical disclosure control for micro-data using the R package sdcMicro. *J Stat Softw*. 2015;67(1):1-36. doi:10.18637/jss.v067.i04
- Denney W, Duvvuri S, Buckeridge C. Simple, automatic noncompartmental analysis: the PKNCA R package. *J Pharmacokinet Pharmacodyn*. 2015;42:S65-S107. doi:10.1007/s10928-015-9432-2
- Schloerke B, Cook D, Larmarange J, et al. *GGally: Extension to "Ggplot2."*; 2021. Accessed April 8, 2021. <https://CRAN.R-project.org/package=GGally>
- Signorell A, Aho K, Alfons A, et al. *DescTools: Tools for Descriptive Statistics*; 2021. Accessed April 8, 2021. <https://CRAN.R-project.org/package=DescTools>
- Brunet M, van Gelder T, Åsberg A, et al. Therapeutic drug monitoring of tacrolimus-personalized therapy: second consensus report. *Ther Drug Monit*. 2019;41(3):261-307. doi:10.1097/FTD.0000000000000640

23. Zwart TC, Moes DJAR, van der Boog PJM, et al. Model-informed precision dosing of Everolimus: external validation in adult renal transplant recipients. *Clin Pharmacokinet.* 2021;60(2):191-203. doi:[10.1007/s40262-020-00925-8](https://doi.org/10.1007/s40262-020-00925-8)
24. Sheiner LB, Beal S, Rosenberg B, Marathe VV. Forecasting individual pharmacokinetics. *Clin Pharmacol Ther.* 1979;26(3):294-305. doi:[10.1002/cpt1979263294](https://doi.org/10.1002/cpt1979263294)

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Labriffe M, Woillard J-B, Debord J, Marquet P. Machine learning algorithms to estimate everolimus exposure trained on simulated and patient pharmacokinetic profiles. *CPT Pharmacometrics Syst Pharmacol.* 2022;11:1018-1028. doi:[10.1002/psp4.12810](https://doi.org/10.1002/psp4.12810)