



Application of machine learning to predict tacrolimus exposure in liver and kidney transplant patients given the MeltDose formulation

Laure Ponthier, Pierre Marquet, Dirk Jan A. R. Moes, Lionel Rostaing, Bart van Hoek, Caroline Monchaud, Marc Labriffe, Jean Baptiste Woillard

► To cite this version:

Laure Ponthier, Pierre Marquet, Dirk Jan A. R. Moes, Lionel Rostaing, Bart van Hoek, et al.. Application of machine learning to predict tacrolimus exposure in liver and kidney transplant patients given the MeltDose formulation. European Journal of Clinical Pharmacology, 2022, 79 (2), pp.311-319. 10.1007/s00228-022-03445-5 . hal-04414330

HAL Id: hal-04414330

<https://unilim.hal.science/hal-04414330>

Submitted on 26 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.

Application of machine learning to predict tacrolimus exposure in liver and kidney transplant patients given the meltdose formulation.

Laure Ponthier^{1,2}, Pierre Marquet^{1,3}, Dirk Jan A.R. Moes⁴, Lionel Rostaing⁵, Bart van Hoek⁶, Caroline Monchaud^{1,3}, Marc Labriffe^{1,3}, JB Woillard^{1,3}

¹ P&T, Univ. Limoges, INSERM, Limoges, France.

² Department of Pediatrics, CHU Limoges, Limoges, France.

³ Department of Pharmacology and Toxicology, CHU Limoges, Limoges, France.

⁴ Department of Clinical Pharmacy & Toxicology, Leiden University Medical Center, The Netherlands.

⁵ Nephrology, hemodialysis, apheresis and kidney transplantation department, Grenoble University Hospital, Grenoble, and Grenoble Alpes University, Grenoble, France

⁶ Department of Hepatology, Leiden University Medical Center, The Netherlands.

ORCID: 0000-0003-1695-0695

Corresponding author:

Jean-Baptiste Woillard

Title: PharmD, PhD

Adress: CBRS, rue du Pr Descottes, 87042 Limoges Cedex, France

Phone: +33 5 55 05 61 40

Fax: +33 5 55 05 61 62

Email: jean-baptiste.woillard@unilim.fr

Running title: machine learning to predict tacrolimus melt dose exposure

Acknowledgements: The authors gratefully thank Veloxis and Chiesi for providing the phase II pharmacokinetic data.

Abstract

Purpose: Machine Learning (ML) algorithms represent an interesting alternative to Maximum-a-Posteriori Bayesian-estimators (MAP-BE) for tacrolimus AUC estimation, but it is not known if training an ML model using a lower number of full pharmacokinetic (PK) profiles (=“true” reference AUC) provides better performances than using a larger dataset of less accurate AUC estimates. The objectives of this study were: to develop and benchmark ML algorithms trained using full PK profiles to estimate MeltDose®-tacrolimus individual AUCs using 2 or 3 blood concentrations; and to compare their performance to MAP-BE.

Methods: Data from liver (n=113) or kidney (n=97) transplant recipients involved in Meltdose-tacrolimus PK studies were used for the training and evaluation of the ML algorithms. The “true” AUC_{0-24h} was calculated for each patient using the trapezoidal rule on the full PK profile. ML algorithms were trained to estimate tacrolimus true AUC using 2 or 3 blood concentrations. Performances were evaluated in 2 external sets of 16 (renal) and 48 (liver) transplant patients.

Results: The best estimation performances were obtained with the MARS algorithm and the following limited sampling strategies (LSS): predose (0), 8h and 12h post-dose (rMPE=-1.28%, rRMSE=7.57%), or 0 and 12h (rMPE=-1.9%, rRMSE=10.06%). In the external dataset, the performances of the final ML algorithms based on 2 samples in kidney (rMPE=-3.1%, rRMSE=11.1%) or liver transplant recipients (rMPE=-3.4%, rRMSE=9.86%) were as good as or better than those of MAP-BEs based on 3 time-points.

Conclusion: The MARS ML models developed using “true” MeltDose®-tacrolimus AUCs yielded accurate individual estimations using only two blood concentrations.

Keywords: machine learning; population pharmacokinetics; tacrolimus melt dose; transplantation; exposure

What is already known about this subject:

- Machine Learning (ML) algorithms is an interesting alternative to Maximum-a-Posteriori Bayesian-estimators (MAP-BE) for tacrolimus AUC estimation
- It is not known if training an ML model using a lower number of full pharmacokinetic (PK) profiles (=“true” reference AUC) provides better performances than using a larger dataset of less accurate AUC estimates.

What this study adds:

- ML algorithms were trained from liver (n=113) or kidney (n=97) transplant recipients involved in Meltdose-tacrolimus PK studies in whom the reference AUC_{0-24h} was calculated using the trapezoidal rule on the full PK profiles
- External validation showed that the performances of the final ML algorithms (MARS) based on 2 samples were as good as or better than those of MAP-BEs based on 3 time-points.
- The performances in an external validation set obtained in the present study were similar to, and no better than, those obtained using a larger dataset of less accurate AUC estimates.

Introduction

Tacrolimus is a calcineurin inhibitor very frequently employed in the prevention and treatment of allograft rejection in solid organ transplantation. It exhibits a narrow therapeutic index and a large interindividual and long-term intraindividual variability, making therapeutic drug monitoring and individual dose adjustment essential [1,2]. Several factors influence its pharmacokinetics and explain part of this inter or intra-individual variability, such as patient age, drug-drug or drug-food interactions, genetic polymorphisms of CYP3A isoenzymes, hematocrit, serum albumin concentration, etc. [3].

Therapeutic drug monitoring (TDM) helps to prevent or correct overexposure that may increase the risk of adverse effects as well as underexposure that may increase the risk of allograft rejection [4]. In routine care and as recommended in the Summary of Product Characteristics, tacrolimus TDM and dose adjustment for all the available formulations is performed using the trough concentration (C_0) [1,2]. Indeed, C_0 is a surrogate of the area under the concentration-time curve (AUC) as they exhibit a good but variable correlation [5]. However, AUC seems to be a more precise marker of exposure and has been associated with tacrolimus efficacy/safety [6,7]. One of the limits of the use of AUC is that it requires collecting a relative large number of blood samples.

Envarsus ® is a prolonged-release, once-daily formulation of tacrolimus developed by Veloxis, using their patented technology MeltDose® (Veloxis Pharmaceuticals, Hørsholm, Denmark). Population pharmacokinetic (POPPK) models and Maximum a Posteriori Bayesian estimators (MAP-BE) were developed for Envarsus ®. They allow AUC estimation based on a 3-point limited sampling strategy (LSS): predose (0), 8h and 12h post-dose, one

model for kidney and one for liver transplantation [8], or 0, 4h and 8h in liver transplant recipients [9].

Machine learning (ML) techniques [10] can also be used to estimate the AUC of immunosuppressive drugs based on patient features and observed concentrations. ML encompasses several methods (e.g., support vector machine (SVM), partial least squares discriminant analysis (PLS-DA), random forest, boosting, multivariate adaptive regression splines (MARS), etc.) that involve complex algorithmic designs, including large numbers of free parameters and complex interactions, to minimize the errors between predicted and observed values by means of an error function.

Recently, Woillard et al. trained Xgboost (eXtreme Gradient Boosting) machine learning (ML) algorithms to estimate immediate release tacrolimus AUC using only two blood concentrations (0 and 3h) [11]. These algorithms were trained from very large numbers of requests gathered on a web platform where the reference AUC was estimated using a 3-point LSS and MAP-BE (<https://abis.chu-limoges.fr/login>). In this case, the “reference” AUC corresponded to the real AUC value + uncertainty (“noise”) due to MAP-BE estimation. It is possible that training ML algorithms with a smaller dataset of more accurate “true” AUCs (calculated from rich concentration-time profiles using the trapezoidal rule for instance) would yield as good or even improved performance.

The objectives of this study were: (i) to train ML algorithms on a dataset of MeltDose®-tacrolimus full PK profiles to estimate individual AUCs using only 2 or 3 blood concentrations; and (ii) to compare their performance to those of MAP-BE with the LSSs 0, 8h and 12h or 0, 4h and 8h in independent set of patients.

Material and methods

Patients

Clinical and pharmacokinetic data from 113 liver and 97 kidney transplant patients from two phase II, open-label, multicentre prospective US clinical trials conducted in stable adult kidney and liver transplant patients, who were converted from Prograf® capsules twice daily to Envarsus® (MeltDose®-tacrolimus) tablets once daily were used for training and testing ML algorithms. These trials complied with the Declaration of Helsinki amended in Tokyo and all the patients enrolled gave their written informed consent. These data have been previously used to develop a POPPK and derive a MAP-BE based on a LSS [8]. Briefly, all patients transplanted for at least 6 months were switched from Prograf® to Envarsus® on day 8 of the study and had two PK assessments on days 14 and 21. For each patient, 13 bloods samples were collected pre-dose and 0.5, 1, 1.5, 2, 3, 4, 6, 8, 12, 20 and 24 hours after dosing. All samples were measured using a validated liquid chromatography - tandem mass spectrometry method with a lower limit of quantitation of 0.2 ng/ml.

The “true” AUC was calculated for each patient using the trapezoidal rule on the full PK profiles in the PKNCA R package [12].

Preparation of the data and feature engineering

Concentrations and sampling times were extracted and binned into theoretical time classes: concentrations at 0 (“C0” sampled at $t = 0$ min), 1 h (“C1” sampled between 0.7 and 1.1 h), 2 h (“C2”, 1.6 - 2.5 h), 3 h (“C3”, 2.6 - 3.5 h), 4 h (“C4”, 3.6 - 4.5 h), 6 h (“C6”, 4.6 - 6.5 h), 8 h (“C8”, 6.6 - 8.5 h), 12 h (“C12”, 8.6 - 12.5h), 20 h (“C20”, 16.6 - 20.6 h), and 24 h (“C24”, 20.7 - 25.5 h). The range of the time classes were investigated and selected manually in order to have every time represented in each patients. When relevant, deviation from the theoretical sampling times was taken into account by creating a new variable corresponding to the relative deviation with respect to the theoretical time in each bin.

Finally, AUC prediction was attempted based on 2 or 3 concentrations, the relative time deviation and other predictors including demographic data (age, sex, time elapsed since

transplantation, transplanted organ (liver or kidney), and hematocrit). The individual status regarding cytochrome P450 3A5 polymorphisms were not known in the original study and could not be investigated further.

Machine learning analysis

All pre-processing and machine learning analyses were performed using the tidymodels framework in R version 4.0.5 [13]. Missing concentration data were imputed using the k nearest neighbors (with $k=5$). Data splitting between a training set (75%) and a test set (25 %) was performed by random selection of patients. The training set was secondarily split into an analysis set (80%) and an assessment set (20%) in order to benchmark different ML algorithms and select the best without wasting the test set for this purpose. Preprocessing consisted in normalization (centering and scaling) of numeric variables, and one hot encoding of categorical features.

Xgboost [14], MARS [15] and GLMNET (generalized linear model via penalized maximum likelihood) [16] algorithms with different combinations of 3 concentration-time points were first trained and compared based on the root mean square error (RMSE) and coefficient of determination (R^2). The combinations investigated were 0/8/12 hours, 0/2/6 hours, 0/2/4 hours, 0/4/6 hours, 0/4/8 hours, 0/2/12 hours and 0/6/12 hours, based on previous studies [8,9].

Combination of C0 with the 2 most important concentration-time points in variable importance plots were then investigated for the development of the ML algorithms based on 2 points and the selection was based on the performance (RMSE).

For each algorithm and combination, the hyperparameters were tuned using ten-fold cross-validation in the analysis set. Once optimized, the ML algorithms were evaluated in the assessment set to select the best in terms of prediction performance. This best algorithm was

refined using the analysis and assessment sets combined and was finally evaluated in the test set. The global procedure is summarized in supplemental Figure 1.

External evaluation and comparison with others approaches

The final ML algorithms based on 2 or 3 concentrations were evaluated in two external, independent datasets in which tacrolimus was measured using validated liquid chromatography-tandem mass spectrometry methods.

The first comprised 16 full PK profiles (13 blood samples collected at pre-dose and 0.5, 1, 2, 3, 4, 6, 8, 10, 12, 13, 15 and 24 hours after dosing) of MeltDose®-tacrolimus from stable renal transplant patients. The performances of the different ML algorithms were compared to those of 2 MAP-BE using two different 3-point LSSs: 0/8/12h, intended for renal transplant patients [17]; and 0/4/8h intended for liver transplant patients [9]. The reference AUC was calculated using the trapezoidal rule on all the available samples.

The second external set comprised data from 51 stable liver transplant patients at 2 weeks after conversion (9 blood samples collected at pre-dose and 1, 2, 3, 4, 6, 8, 12, 24h after dosing), used to develop the POPPK model and the MAP-BE based on the 0/4/8h LSS [9]. Three patients with sampling time deviation >5% or 1 hour from the theoretical LSS were excluded from the analysis. In this dataset, the reference AUC was obtained by application of the POPPK model and the MAP-BE on the 9 samples. The different ML algorithms were compared to MAP-BE using the 0/8/12h LSSs [17] and to the results reported in the original study in liver transplant patients [9].

Results

Data

Two hundred and ten PK profiles were available in the original dataset and were randomly assigned to the training (n=158) or test (n=52) sets. In the original dataset used for the development of the ML algorithm, the proportion of imputed data were 0% for C0, 2.3% for C0.5, 1.9% for C1, 1.4% for C1.5, 1.4% for C2, 0.95% for C3, 0.95% for C4, 0% for C6, 1.4% for C8, 0.5% for C12, 0.5% for C16, 0.95% for C20, 0% for C24. In the training set, 130 were randomly assigned to the analysis and 28 to the assessment set. In one of the external datasets, among the 51 PK profiles from liver transplant recipients, only 48 had concentrations at 0, 8 and 12h available. Patient characteristics in the training, assessment, test sets and in the 2 external validation sets are provided in Table 1.

Comparison of algorithms

The results obtained with each ML algorithm and the different 3-point combinations showed, as previously observed [17], that the 0/8/12h LSS was always associated with the best performances in the assessment set (Supplemental Table 1). While GLMNET displayed slightly better results in the analysis set, the MARS algorithm yielded the best performances in the assessment set (relative RMSE= 7.57%, relative MPE = -1.28%) followed by Xgboost (relative RMSE= 9.17 %, relative MPE = 0.43%) and GLMNET (relative RMSE= 10.12%, relative MPE = 0.90%) (Table 2). Figure 1 shows the scatter plot and Bland–Altman plot of the individually-predicted vs reference AUC0-24hs for the 0/8/12h LSS in the test set with the MARS algorithm.

The variable importance plot for the MARS algorithm showed that the concentrations at 8h, 12h and 0h were the most important, in this order (Figure 2) leading us to investigate different combinations of these 3 concentration-time points for the 2-point ML algorithms.

Performances of the 2-point limited sampling strategies

The two-point LSSs with concentrations measured at time 0 and 12h performed best (Supplemental Table 2). The achievements of the different ML algorithms with this 0/12h LSS in the analysis and assessment sets are compared in Table 2. The MARS algorithm again performed best and was therefore the only one evaluated in the test set, showing rMPE/rRMSE/number(%) out of the $\pm 20\%$ interval of -0.36/10.4/2(3.8%), as compared to 1.62/9.8/5(9.4%) with the 3-point LSS. The scatter plot and Bland–Altman of predicted vs reference AUCs with the MARS algorithm and 2-point LSS in the test set is presented in Figure 1.

Evaluation in external datasets

The results of the final MARS algorithm based on 2 or 3 points and of the MAP-BEs based on 3 points in both validation sets are compared in Table 3. The MARS algorithms with 2 points: (i) displays similar performances to that with 3 points the two independent datasets, of kidney and liver transplant recipients respectively; (ii) had a lower MPE than MAP-BE in kidney transplant patients; (iii) showed a RMSE almost half of that of the MAP-BE based on the 0/8/12h LSS in the same patient group while similar to that of the MAP-BE based on the 0/4/8 hours LSS developed in liver transplant patients; (iv) yielded RMSE similar to that of MAP-BE based on the 0/8/12h LSS in liver transplant patients and similar bias or number of patient out of the 20% bias range in comparison to the original study [9]. The corresponding Bland-Altman plot of predicted vs reference AUCs are presented in Figures 3 and 4.

Discussion

The MARS models developed here provide accurate AUC estimation for melt-dose tacrolimus, using only concentrations measured at 0 and 12 hours and a very limited number of other features.

The sample at 12h may be difficult to draw in routine practice, as it is unlikely that outpatients are kept for 12h at the hospital. However, as previously observed [17] and shown in the variable importance plot (figure 2), it was the one of the most important features in our algorithm. Capillary blood microsampling devices may make collection of this late point easier, even if it could introduce another source of variability. Indeed, there is an additional bias and imprecision that can vary between the types of device [18–20]. Another MAP-BE developed for liver transplant recipients required a late sample at 8h to yield accurate AUC estimation [9]. However, the ML algorithms including 8h instead of 12h were not associated with better performance in the present study. Interestingly, C8 was the most important time-point in the VIP plot but its combination with C0 led to decreased performances in comparison with the C0 & C12 combination.

We investigated 3 different algorithms that rely on different approaches. Glmnet is a penalized regression and is based on a linear relationship between predictors and AUC, the 2 others are based on nonlinear relationships (Xgboost is an ensemble method that aggregate decision trees and MARS is a non-parametric derived regression that break a given distribution into small linear pieces).

In our previous works on machine learning tacrolimus AUC estimation [11], we either used as references 3-point MAP-BE estimates of the AUC, or PK profiles simulated using a POPPK model [21]. Here, the reference AUC was calculated for each patient using the trapezoidal rule on rich PK profiles (samples collected pre-dose, 0.5h, 1h, 1.5h, 2h, 3h, 4h, 6h, 8h, 12h, 20h and 24h after dosing). We hypothesized that using complete profiles for model development would decrease the noise and improve estimation performance, even if the

training set was smaller. Actually, the performances in an external validation set obtained in the present study were similar to, and no better than, those obtained in these previous studies. However, this is indirect comparison since these studies were performed for different tacrolimus formulations.

When evaluated in an external dataset of kidney transplant recipients, the ML models developed outperformed the MAP-BEs previously developed on the same datasets. A possible explanation is that some PK profiles in the dataset were not necessarily at steady-state, which may mislead mechanistic models more than ML algorithms. Another explanation would be that these patients were quite unusual and that the ML algorithm is more flexible to model them since it is less constrained. On the contrary, for liver transplant recipients, the performances obtained in the same external dataset by a MAP-BE previously developed for this graft type were similar to those of the ML algorithms trained in a dataset of both renal or liver transplant recipients.

When evaluated in liver transplant patients at 2 weeks after the switch (i.e., against the profiles used to develop the POPPK model [9]), the performances of the 2- or 3-sample MARS algorithm and of our MAP-BE previously developed in liver transplant patients [8] were excellent and exhibits bias and number of profile with bias out of the 20% interval similar to the ones reported in the original study [9]. Of note, the performances of the MAP-BE were numerically better than those of the MARS algorithm in the validation dataset. This could be explained by the limited number of samples available for the development of the ML algorithm (data driven approach), whereas the POPPK models (mechanistic approach) require a lower number of samples to provide very accurate results. Recently, we have shown that the addition of simulations from a POPPK model to experimental data could improve the learning performances of the ML algorithms [22].

The methodology used in the present study for the development of the model can be regarded as complex since we performed 2 consecutive data splitting. However, this two-set approach allowed us to prevent overfitting and to benchmark different algorithms in the train set while keeping the test set for the final evaluation only.

The range of the time classes were investigated and selected manually in order to have the most patients as possible with all the bins. Even if it was a subjective approach, we think that it has not biased the analysis. For the external validation, we also removed subjectively profiles with sampling time deviation $>5\%$ or 1 hour from the theoretical LSS. Interestingly, the features “relative deviation with respect to the theoretical sampling time” were not selected as important by the algorithms. Nevertheless, further studies with external data or simulations studies have to be performed to clarify this point.

This study has some limitations. First, pharmacogenetic data were not available. It is well established that the CYP3A5 genotype influences tacrolimus clearance and that patients expressing CYP3A5 require a higher tacrolimus dose than non-expressors, even this influence might be less for Envarsus due to its more distal absorption [9]. Anyway, its implication has not been studied so far in machine learning estimators of tacrolimus AUC and even without this information the accuracy and precision of the models are excellent. Secondly, the size of the external validation set is small for renal transplant patients, but full PK profiles are not easily available. Finally, all the patient used to develop and validate the algorithm were at stable transplant period (at least 6 months post transplantation). The performances of the algorithms developed should be further investigated in patient in the early phase post transplantation.

In conclusion, a MARS estimator developed from “true” reference AUCs very accurately estimated melt-dose tacrolimus AUC using only 2 or 3 blood concentrations and a few numbers of other features, and performs as well as ML estimators we previously developed

for other tacrolimus formulations, where 3-point LSS and MAP-BE AUCs, or trapezoidal AUCs from simulated PK profiles, were used as references. The MARS estimator based on 2 time points may improve the use of AUC-based tacrolimus individual dose adjustment.

Declarations

Ethical Approval

These trials complied with the Declaration of Helsinki amended in Tokyo and all the patients enrolled gave their written informed consent.

Competing interests

Laure Ponthier, Dirk Jan A.R. Moes, Bart van Hoek, Caroline Monchaud, Marc Labriffe1 have no conflicts of interest that are relevant to the content of this manuscript. Pierre Marquet and L Rostaing has received research grants and speaker fees from Chiesi, JB Woillard has received speaker fees from Chiesi,

Authors' contributions

LP, JBW, DJM, PM, ML conceived of or designed study, LP, JBW, DJM, CM, LR, BvH performed research, LP, JBW, DJM, PM, ML analyzed data, LP, PM, DJM, ML, JBW wrote the paper

Funding

No funding was received to perform this study

Availability of data and materials

The datasets analysed during the current study are the property of Veloxis and Chiesi and are not publicly available.

References

1. Brunet M, van Gelder T, Åsberg A, Haufrøid V, Hesselink DA, Langman L, et al. Therapeutic Drug Monitoring of Tacrolimus-Personalized Therapy: Second Consensus Report. *Ther Drug Monit.* 2019;41:261–307.
2. Woillard JB, Monchaud C, Saint-Marcoux F, Labriffe M, Marquet P. Can the Area Under the Curve/Trough Level Ratio Be Used to Optimize Tacrolimus Individual Dose Adjustment? Transplantation. Volume-Issue-10.1097. 2022 Nov 2;
3. Staatz CE, Tett SE. Clinical pharmacokinetics and pharmacodynamics of tacrolimus in solid organ transplantation. *Clin Pharmacokinet.* 2004;43:623–53.
4. Hesselink DA, Bouamar R, Elens L, van Schaik RHN, van Gelder T. The role of pharmacogenetics in the disposition of and response to tacrolimus in solid organ transplantation. *Clin Pharmacokinet.* 2014;53:123–39.
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.* 2013;35:322–7.
6. Woillard J-B, Saint-Marcoux F, Debord J, Åsberg A. Pharmacokinetic models to assist the prescriber in choosing the best tacrolimus dose. *Pharmacol Res.* 2018;130:316–21.
7. Bouamar R, Shuker N, Hesselink DA, Weimar W, Ekberg H, Kaplan B, et al. Tacrolimus predose concentrations do not predict the risk of acute rejection after renal transplantation: a pooled analysis from three randomized-controlled clinical trials(†). *Am J Transplant.* 2013;13:1253–61.
8. 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.* 2017;56:1491–8.
9. Martial LC, Biewenga M, Ruijter BN, Keizer R, Swen JJ, van Hoek B, et al. Population pharmacokinetics and genetics of oral meltdose tacrolimus (Envarsus) in stable adult liver transplant recipients. *Br J Clin Pharmacol.* 2021;
10. Lima AN, Philot EA, Trossini GHG, Scott LPB, Maltarollo VG, Honorio KM. Use of machine learning approaches for novel drug discovery. *Expert Opin Drug Discov.* 2016;11:225–39.
11. Woillard J-B, Labriffe M, Debord J, Marquet P. Tacrolimus Exposure Prediction Using Machine Learning. *Clin Pharmacol Ther.* 2020;
12. Denney WS, Duvvuri S, Buckeridge C. Simple, automatic non compartmental analysis: the PKNCA R package. *J Pharmacokinet Pharmacodyn* [Internet]. 2015 [cited 2021 May 9];42:11–107. Available from: <http://link.springer.com/10.1007/s10928-015-9432-2>
13. Kuhn M, Wickham H. tidymodels: Easily Install and Load the ‘Tidymodels’ Packages version 0.1.0 from CRAN. [Internet]. Available from: <https://rdrr.io/cran/tidymodels/>
14. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Internet]. San Francisco California USA: ACM; 2016 [cited 2021 Jun 30]. p. 785–94. Available from: <https://dl.acm.org/doi/10.1145/2939672.2939785>
15. Friedman JH, Roosen CB. An introduction to multivariate adaptive regression splines. *Stat Methods Med Res.* 1995;4:197–217.
16. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J Stat Softw.* 2010;33:1–22.
17. 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.*

2017;56:1491–8.

18. Deprez S, Stove CP. Fully Automated Dried Blood Spot Extraction coupled to Liquid Chromatography-tandem Mass Spectrometry for Therapeutic Drug Monitoring of Immunosuppressants. *J Chromatogr A*. 2021;1653:462430.

19. Capiou S, Veenhof H, Koster RA, Bergqvist Y, Boettcher M, Halmingh O, et al. Official International Association for Therapeutic Drug Monitoring and Clinical Toxicology Guideline: Development and Validation of Dried Blood Spot-Based Methods for Therapeutic Drug Monitoring. *Ther Drug Monit*. 2019;41:409–30.

20. Delahaye L, Veenhof H, Koch BCP, Alffenaar J-WC, Linden R, Stove C. Alternative Sampling Devices to Collect Dried Blood Microsamples: State-of-the-Art. *Ther Drug Monit*. 2021;43:310–21.

21. Woillard J-B, 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.

22. 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;

Table 1: Patient characteristics in the analysis, assessment, train, test and external sets.

Variables	Analysis set N= 130	Assessment set N= 28	Training set (Analysis + Assessment Set) N=158	Test set N=52	External set Renal N=16	External Set Liver N=48
Age (years)	49 (11)	47 (11)	49 (11)	48 (12)	49 (16)	54 (12)
Daily dose (mg)	5.0 (3.0)	3.6 (1.7)	4.8 (2.9)	5.2 (3.0)	4.2 (2.4)	2.5(1.5)
Sex Male (%)	82 (63.1%)	14 (50.0%)	96 (60.8%)	36(69.2%)	9 (56.2%)	35(68.6%)
Hematocrit (%)	40.7(4.8)	40.2 (3.7)	40.7 (4.6)	39.8 (4.6)	40.4 (4.3)	40.6 (4.5)
AUC0-24 (µg*h/L)	207 (74.)	209 (61)	207 (73)	205 (53)	202 (78)	152 (66)
Liver n (%)	68 (52.31%)	19(67.86%)	87 (100.00%)	26 (50%)	0(0%)	48(100%)
Kidney n (%)	62 (47.69%)	9 (32.14%)	71 (100.00%)	26 (50%)	16(100%)	0(0%)
Concentration at 0 h	6.2 (2.3)	6.4 (2.1)	6.2 (2.3)	6.3 (2.1)	6.3 (2.2)	5.0 (2.1)

Continuous variables are presented as mean (SD) and categorical variables as n (%)

Table 2: Performance of different ML algorithms and limited sampling strategies to estimate tacrolimus AUC_{0-24h} as compared with the reference AUC_{0-24h}: in the analysis set after 10-fold cross-validation; and in the assessment set.

ML ALGORITHM dataset	2-sample LSS (0/12 hours)				3-sample LSS (0/8/12 hours)			
	Relative MPE (%)	Relative RMSE (%)	Number of estimates out of \pm the 20% interval n (%)	Number of estimates out of \pm the 10% interval n (%)	Relative MPE (%)	Relative RMSE (%)	Number of estimates out of \pm the 20% interval n (%)	Number of estimates out of \pm the 10% interval n (%)
GLMNET *Analysis set	1.50	11.0	9(6.9)	35(26.9)	0.97	9.36	7(5.3)	29(22.3)
GLMNET Assessment set	-2.9	10.4	2(7.1)	9(32.1)	0.90	10.12	1(3.6)	5(17.9)
XGBOOST *Analysis set	2.12	17.6	13(10.0)	56(43.1)	0.79	13.1	14(10.7)	40(30.8)
XGBOOST Assessment set	-0.9	11.4	3(10.7)	10(35.7)	0.43	9.17	2(7.1)	8(28.6)
MARS *Analysis set	1.23	11.1	9(6.9)	35(26.9)	0.45	9.8	7(5.3)	33(25.3)
MARS Assessment set	-1.9	10.0	0(0)	9(32.1)	-1.28	7.57	1(3.6)	6(21.4)

*Results in the analysis set were obtained after 10-fold cross validation; GLMNET is LASSO and Elastic-Net Regularized Generalized Linear Models, XGBOOST is extreme gradient boosting and MARS is Multivariate Adaptive Regression Splines

Table 3: Performance of the final MARS algorithm based on 2 concentrations (0 and 12h) and 3 concentrations (0, 8h and 12h) to estimate AUCs in two **independent validation sets** as compared with previous MAP-BE using 3-point LSSs.

		Relative MPE (%)	Relative RMSE (%)	Relative errors out of $\pm 20\%$ n (%)	Relative errors out of $\pm 10\%$ n (%)
Kidney transplant recipients (n=16)	MARS 0/12h	-3.1	11.1	2 (12.5%)	5 (31.3%)
	MARS 0/8/12h	-4.2	10.1	1 (6.3%)	5 (31.3%)
	MAP-BE 0/8/12h (developed in renal transplant patients) [8]	7.3	23.0	8 (53.3%)	12 (80%)
	MAP-BE 0/4/8h (developed in liver transplant patients) [9]	-7.4	11.2	1 (6.3%)	7 (43.8%)
Liver transplant recipients (n =48) [9]	MARS 0/12h	-3.4	9.86	3 (6.2%)	13 (27.1%)
	MARS 0/8/12h	-3.4	9.14	2 (4.1%)	12 (25%)
	MAP-BE 0/8/12h (developed in liver transplant patients) [8]	-1.2	8.84	1 (2.1%)	9 (20.8%)
	MAP-BE 0/4/8h (developed in the same liver transplant patients) [9]	1.8*	8.64*	0 (0%) (n = 53)*	4 (7.50%) (n=53)*

AUC is Area under the curve, MAP-BE is Maximum a Posteriori Bayesian estimation, MARS is Multivariate Adaptive Regression Splines; *corresponds to the results reported in the original study in which the same patients were used to developed the model and evaluate the LSS

Figure Legends

Figure 1: Scatter plot of AUC0-24s estimated using the MARS algorithm based on 2 points at 0 and 12h (A) or 3 points at 0, 8h and 12h (B) vs reference trapezoidal AUC0-24 in the test set, and corresponding Bland–Altman plots (C) and (D). Difference is the difference between the Reference and the MARS AUC0-24s, and mean is the average of both.

Figure 2: Variable importance plot for the MARS algorithm in the analysis set. “C0” is concentration sampled at $t = 0$ min, “C8” is the concentration sampled between 6.6 and 8.5 h and “C12” is the concentration sampled between 8.6 and 12.5h. Importance is relative calculated using a generalized cross-validation (GCV) statistic.

Figure 3: Bland Altman of reference vs predicted AUC0-24s using MARS algorithm based on 2 points (0/12 hours; A) or 3 points (0/8/12 hours ; B), and MAP-BE based on 3 points (0/8/12 hours; C or 0/4/8 hours ; D) in the renal transplant patients external dataset.

Figure 4: Bland Altman of reference vs predicted AUCs using MARS algorithm based on 2 points (0/12 hours; A) or 3 points (0/8/12 hours ; B), and MAP-BE based on 3 points (0/8/12 hours; C) in the liver transplant patients external dataset[9].