Received: 3 May 2021 | Revised: 20 December 2021 | Accepted: 31 December 2021

DOI: 10.1002/psp4.12758

ARTICLE

Predicting model-informed precision dosing: A test-case in tacrolimus dose adaptation for kidney transplant recipients

Ruben Faelens[1] | **Nicolas Luyckx**[2] | **Dirk Kuypers**[3] | **Thomas Bouillon**[1,4] | **Pieter Annaert**[1,4]

[1]Department of Pharmaceutical and Pharmacological Sciences, KU Leuven, Leuven, Belgium

[2]SGS Exprimo NV, Mechelen, Belgium

[3]Department of Nephrology, University Hospitals Leuven, Leuven, Belgium

[4]BioNotus GCV, Niel, Belgium

Correspondence
Pieter Annaert, Drug Delivery and Disposition, Herestraat 49-box 921, 3000 Leuven, Belgium.
Email: pieter.annaert@kuleuven.be

Present address
Nicolas Luyckx, Calvagone SARL, Lyon, France

Funding information
This work was supported by FWO TBM grant T003117N

Abstract

Before investing resources into the development of a precision dosing (model-informed precision dosing [MIPD]) tool for tacrolimus, the performance of the tool was evaluated in silico. A retrospective dataset of 315 de novo kidney transplant recipients was first used to identify a one-compartment pharmacokinetic (PK) model with time-dependent clearance. MIPD performance was subsequently evaluated by calculating errors to predict future concentrations, which is directly related to dosing precision and probability of target attainment (PTA). Based on the identified model residual error, the theoretical upper limit was 45% PTA for a target of 13.5 ng/ml and an acceptable range of 12–15 ng/ml. Using empirical Bayesian estimation, this limit was reached on day 5 post-transplant and beyond. By incorporating correlated within-patient variability when predicting future individual concentrations, PTA improved beyond the theoretical upper limit. This yielded a Bayesian feedback dosing algorithm accurately predicting future trough concentrations and adapting each dose to reach a target concentration. Simulated concentration-time profiles were then used to quantify MIPD-based improvement on three end points: average PTA increased from 28% to 39%, median time to three concentrations in target decreased from 10 to 7 days, and mean log-squared distance to target decreased from 0.080 to 0.055. A study of 200 patients was predicted to have sufficient power to demonstrate these nuanced PK end points reliably. These simulations supported our decision to develop a precision dosing tool for tacrolimus and test it in a prospective trial.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THIS TOPIC?

Tacrolimus is an immunosuppressor used to reduce graft rejection risk in solid organ transplant recipients. It has a narrow therapeutic index, potentially benefiting from model-informed precision dosing (MIPD). However, this improvement has not been quantified in silico.

WHAT QUESTION DID THIS STUDY ADDRESS?

This study predicts the effect on probability of target attainment of tacrolimus in silico during the first 14 days post renal transplant. Based on data previously obtained in 315 kidney transplant recipients, trough concentrations under MIPD-recommended dosing were simulated. These were used to compare standard of care with MIPD using three criteria: time to reach target, per-patient average target attainment, and distance from target. These data were used to calculate the statistical power of a prospective clinical trial, and to optimize the design of this trial.

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

This study is the first published in silico prediction of MIPD in tacrolimus. We propose the model-predictive control/MIPD estimation method to take parameter drift into account.

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

This allows optimization of precision dosing implementation, and the design of a sufficiently powered study to demonstrate impact on clinical care.

INTRODUCTION

Optimal dosage of drugs with a narrow therapeutic index is an active area of research. These drugs display high pharmacokinetic (PK) or pharmacodynamic (PD) variability, exceeding the safe and effective variability[1] and therefore require individual dose adaptation. Covariate-based dose adaptation may be attempted first, as, for example, patient bodyweight can be easily measured. In case of large unexplained variability, however, regular follow-up of the patient is needed to adapt the dose. In its simplest form, the drug is titrated until efficacy and safety is reached for the patient. Typically, drug concentration is used as a quantitative surrogate, aiming for an exposure previously established as having sufficient efficacy and acceptable safety.

One such drug requiring blood concentration monitoring is tacrolimus. At sufficiently high concentrations, tacrolimus acts as an immunosuppressor reducing the risk of graft rejection in solid organ transplant recipients.[2] High tacrolimus concentrations are strongly associated with a higher rate of drug-induced side effects, however, including acute kidney injury.[3] The acceptable range depends on transplanted organ, immunological risk, adjunct immunosuppressive therapies, and may be further adapted to physician discretion.[4,5]

The required doses to achieve concentrations at a given target are highly variable between patients. Tacrolimus exhibits a high interindividual PK variability,[6] with some studies reporting a 10-fold range in individual clearance.[7] To achieve safe and effective drug concentrations, regular follow-up of tacrolimus concentrations is therefore recommended.

However, translating an observed drug concentration into the required dose adaptation is not trivial. In theory, model-informed precision dosing (MIPD) can provide accurate dosing recommendations, yielding a high target concentration achievement. Many software packages offer dose adaptation of tacrolimus,[8] although only a single study has prospectively investigated whether MIPD improves this probability of target attainment (PTA).[9] Størset et al. managed to demonstrate MIPD increases PTA, as compared to the standard of care. Unfortunately, no improvement in early post-transplantation could be shown. The study may have been underpowered to show this, as only 80 patients were recruited. Enrollment was based on convenience, rather than a power calculation.

To evaluate the performance of MIPD, some in silico methods are available. Classical goodness of fit (GOF) metrics, such as mean prediction error or root mean squared prediction error (RMSE%), show how well a model can fit existing data. Recently, prospective evaluation was proposed to evaluate these metrics on future concentrations, based on only the concentration samples collected up to that point.[10] Unfortunately, none of these metrics directly translate into the expected individual PK, PD, or clinical outcomes when using MIPD.

Simulated data are preferred over GOF metrics for two reasons. First, predicted individual data can be condensed to clinical benefit (e.g., avoidance of rejection), and this benefit is weighed against the implementation cost. Should pharmacogenetic information be included, how many additional blood samples are needed, and should MIPD even be implemented at all? Quantifying the potential clinical benefit is key in justifying investment into MIPD.

Second, simulated data help to design a prospective clinical trial comparing MIPD to standard of care. Without an in silico prediction of the study end point, the only recourse is either to select the study sample size based on available resources (as large as we can), or on vague assumptions (we estimate X% improvement in PTA). Neither method is a solid way to design a clinical study, leading to inconclusive results. Even if there may be a theoretical and worthwhile benefit to MIPD, it may not be possible to demonstrate this effect in a reasonably funded clinical study. PTA predictions allow informed and realistic clinical study design.

This work shows how the impact of a Bayesian feedback MIPD tool optimizing PK outcomes can be predicted in tacrolimus dosing of renal transplant recipients the first 14 days post-transplant. These predictions are first used to evaluate whether the proposed MIPD achieves a sufficiently high improvement in the population, and indeed is a worthwhile investment. It is then used to design a sufficiently powered clinical trial to show this benefit. The developed simulation software is available as open source. We anticipate that this approach can be applied to many other drugs where the benefit or optimal modalities of implementing MIPD is uncertain.

METHODS

Source data

A retrospective study[11] of 315 kidney allograft recipients transplanted between 2004 and 2014 was repurposed for this work. For an in-depth description, we refer to the work by Vanhove et al. The dataset consisted of trough concentrations measured on days 0–14 post-transplantation under the standard of care (i.e., leading to a dose adaptation by experienced transplant physicians targeting trough concentrations between 12 and 15 ng/ml). Extensive data management was required to prepare this dataset for modeling (described in Supplementary Materials). Missing covariates were imputed as the population median.

Model development

Based on literature review, a one-compartment model with oral absorption was selected as the appropriate starting point for this sparse dataset. Following the approach from other studies,[12–14] the absorption rate constant (ka) was fixed to 4.5/h,[15] as it cannot be reliably estimated from trough data alone. A two-compartment model and addition of lag time were investigated as possible improvements. The use of a hematocrit-standardized model[16] was investigated, including concentration-dependent binding of tacrolimus to erythrocytes. The whole-blood concentration C_{wb} is related to hematocrit-standardized concentration C_{std} through a concentration-dependent proportionality factor R, with C_{stdmax} reflecting the maximum binding capacity, and C_{std50} the concentration associated with half maximum binding.

$$C_{wb} \approx C_b = C_{std} * R * \frac{Hct}{45\%} \tag{1}$$

$$R = C_{stdmax} * \frac{C_{std}}{C_{std} + C_{std50}} \tag{2}$$

As precision dosing targeting C_{wb} depends on future hematocrit values, a joint model was used predicting both tacrolimus whole-blood concentration and hematocrit. The time course of hematocrit was modeled using a sigmoid model. Interindividual variability (IIV) was applied to all parameters.

$$Hct = Baseline_{Hct} - Emax_{Hct} * \frac{t}{t + E50_{Hct}} \tag{3}$$

Random effects were modeled using lognormal IIV. Exploratory graphical analysis pointed to a potential increase in clearance over time. This was estimated using interoccasion variability (IOV) and investigated for correlation with available covariates and time since transplantation. The effect of time was modeled as either an exponential (Equation 4) or sigmoidal (Equation 5) function, with $T50$ the time at which 50% of maximum clearance was reached.

$$CL/F = CL0/F * \left(1 - e^{\frac{\ln(0.5)}{T50} t}\right) \tag{4}$$

$$CL/F = CL0/F * \frac{t}{T50 + t} \tag{5}$$

Covariates were selected using a stepwise covariate search, including covariates that improved the objective function value (OFV) by 3.84 or more ($p < 0.05$) in the forward step, and eliminating covariates resulting in less than 7.88 increase in OFV ($p > 0.005$). Continuous covariates (age and weight) were included as a power-model (Equation 6). Discrete covariates were included as an x-fold change (Equation 7). Last observation carried forward (LOCF) was used to interpolate time-varying covariates.

$$X = \theta_X * e^{\eta_X} * \left(COV/\mu_{COV}\right)^{\beta_{X,COV}} \tag{6}$$

$$X = \theta_X * e^{\eta_X} * \left(\beta_{X,COV}\right)^{COV} \quad (7)$$

Only covariates available in routine clinical practice were considered for covariate building. This was therefore limited to age, bodyweight, hematocrit, and formulation. Notably, *CYP3A5* genotype was excluded from covariate search, as it is not routinely measured in clinical practice.

Model evaluation

Models were evaluated using the likelihood ratio test, with a ΔOFV of −3.84 or more justifying the addition of a new parameter at $p < 0.05$. An evaluation of prediction-corrected visual predictive check (pcVPC), GOF plots, and biological plausibility of parameters were also performed, rejecting models with qualitatively poor results. The pcVPCs were generated from 500 simulated subsets of the original data. This was used to identify the appropriate model structure, covariates, IIV, and residual error model.

Models with good population fit were further evaluated on fitness for use in Bayesian feedback for MIPD. To this end, prospective evaluation with Bayesian feedback was used to evaluate predictive performance. Prediction error for observation $n + 1$ was calculated from an individual parameter estimate on observations 1..n. Only cases with n and $n + 1$ on consecutive days were included.

$$PE\,\% = \frac{IPRED - CONC}{CONC} \quad (8)$$

Prediction error *PE%* is based on the model prediction *IPRED* and the actual measured concentration *CONC*. It was used to characterize predictive performance of candidate models as RMSE, and models were compared using a *t*-test on (*PE%*).[2] This is especially relevant to assess fitness for use in MIPD, as prediction error *PE%* is directly related to the error in the resulting concentration C_{res} after applying the recommended MIPD dose targeting C_{target}. The intermediate steps are available in Supplementary Material.

$$PE\,\% = \frac{C_{target} - C_{res}}{C_{res}} \quad (9)$$

Based on Equation 9, an allowed C_{res} between 12 and 15 ng/ml, and C_{target} of 13.5 ng/ml resulted in an allowed *PE%* between −10% and 12.5%. This was used to derive a theoretical upper limit for MIPD PTA. We assumed an unlimited number of concentrations, allowing to identify individual parameters but not predict future within-subject variability.

$$\lim_{n->\infty} PTA = p\langle N\left(0, \sigma^2\right) \in [-10\%, 12.5\%]\rangle$$

To allow comparison, standard dosing by the physicians was also characterized in the form of *PE%*, by assuming physicians prescribed a dose that they think will hit the target. In other words, the in cerebro modeling of the physician predicts a trough concentration of 13.5 ng/ml at the dose they prescribed. The *PE%* for physicians was therefore calculated as

$$PE\,\% = \frac{13.5 - CONC}{CONC} \quad (10)$$

Bayesian feedback and model-predictive control

Tacrolimus exhibits high PK variability, not only between patients but also within a single subject. This variability can be described by the residual error model or through IOV, depending on whether rich individual data is available. To the best of our knowledge, all models previously used to describe tacrolimus PKs assume a random variability. This variability is not entirely random, however, and previous studies have reported a correlation of the error between subsequent occasions.[17] This seems intuitive: an increased clearance on day N should indeed carry over to the following day $N + 1$. This was quantified in the dataset by calculating the autocorrelation of the residual error from a standard Bayesian fit. For a detailed discussion, please see Supplementary Material.

To integrate this correlation between subsequent occasions, we opted to use a pragmatic approach, inspired by similar closed loop control systems in anesthesia.[18] On day 1, regular empirical Bayesian estimation (EBE) with a priori estimates θ and IIV Ω is used to estimate the most likely individual parameters η. On day i, the individual parameter estimates from day $i - 1$ are used as a priori estimates ($\theta' = \eta$), whereas the same IIV Ω is retained. This approach is reminiscent of model-predictive bioreactor control in chemical engineering and was therefore dubbed model-predictive control MIPD (MPC/MIPD). The predictive performance of this method was compared to classical EBE using the prediction error *PE%* described previously.

Dosing algorithm and simulation strategy

Once a method for predicting future concentrations was established, we could search the dose required to optimize resulting concentrations. The dosing algorithm is described by the following pseudo-code, with Y_i the measured concentration, t_i the associated sample time, η the individual parameter estimates, D_j the dose at administration j, t_j the associated dosing time, t_{j+1} the time of the subsequent dose, $f(\eta, t_{j+1}, D_j)$ the function

to predict a concentration at time t_{j+1} as a result of all doses up to and including dose j, and C_{target} the target concentration:

INITIALIZATION:

observed = []

regimen = [Loading Dose, Future Planned Doses]

WHEN A NEW CONCENTRATION (t_i, Y_i) BECOMES AVAILABLE:

add (t_i, Y_i) to observed

update regimen from the patient EHR system

fit η using MPC/MIPD

FOR all future doses at time t_j:

use a root finding algorithm to find dose $D_{rec,j}$

such that:

$$f(\eta, D_{rec,j}, t_{j+1}) = C_{target}$$

and adapt the regimen: $D_j = D_{rec,j}$

if simulating:

determine the next observed concentration

$$Y'_{i+1} := f(\eta_{orig}, t_{i+1}) + \epsilon$$

The dosing algorithm was designed to execute as new concentration samples become available, regardless of whether that sample is within the acceptable range. The full dosing history and concentrations at each iteration were used to fit individual parameters η, after which the PK model $f(\eta, D_{rec,j}, t_{i+1})$ was used to predict future trough concentrations. For each future administration, the dose was adapted such that the trough concentration was equal to the target concentration C_{target} of 13.5 ng/ml. This is visualized in Figure 1a.

We made sure computer dosing would require minimal changes to the current clinical workflow at UZ Leuven. Therefore, the loading dose was not adapted. Only dosing amounts were adapted. Planned dosing times and formulations (Advagraf or Prograft) were retained from the source dataset. Doses of 0 mg were not recorded in the source dataset, therefore a dummy dose of 0 mg was added at 08:00 if Advagraf was previously administered, or 08:00 and 20:00 if Prograft was previously administered. To ensure fair comparison with physician-based dosing, only doses 6 h after a concentration sample were considered for adaptation, as concentrations were generally only available in practice 3 h post-sampling and adapting a dose close to administration time was not deemed practical. Doses were rounded to 0.5 mg.

To simulate the resulting concentrations Y' after applying MIPD, the following procedure was used (visualized in Figure 1b). As we cannot go back in time and administer the recommended dose D_{rec} to the actual patient, we used the best prediction available: we simulated using a fitted η_{orig} on all observed concentrations for this patient in the

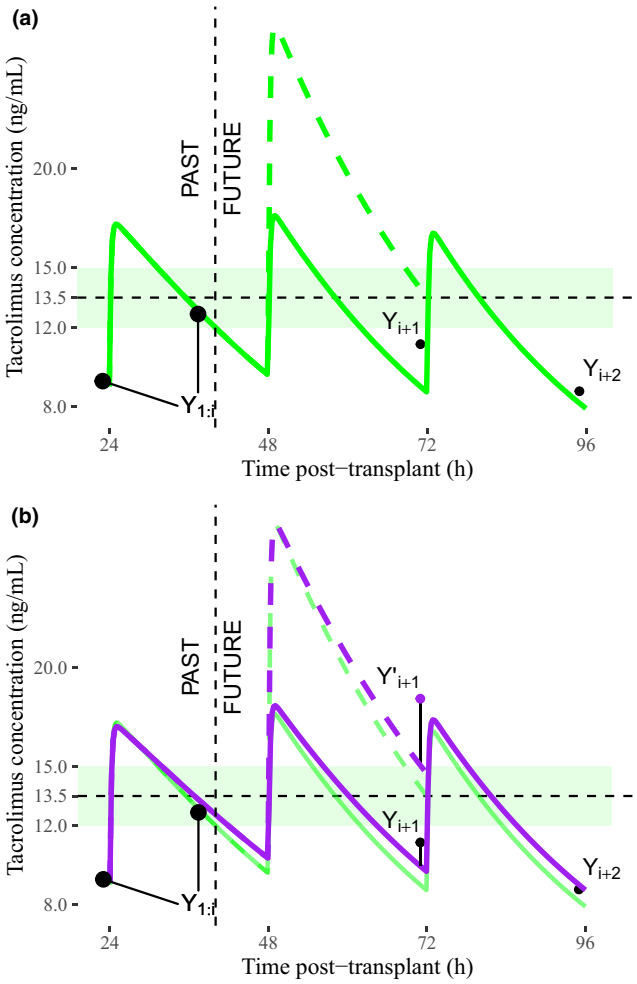

FIGURE 1 (a) Observations 1 to i were used to calculate fit $f(\eta)$ (solid green line). This was then used to find dose D_rec that makes $f(D_rec, t_(j+1)) = C_$target (dotted green line). (b) All observations 1:n were used to calculate fit $f(\eta\_orig)$ (solid purple line). This was then used to find $f(\eta\_orig, D\_rec, t\_(i+1))$ (dotted purple line). The final result $Y'\_(i+1)$ is calculated by adding the original residual error ε on top

historic dataset $f(\eta_{orig}, t_{i+1})$ and re-applied the original residual error ϵ.

$$Y'_{i+1} = f\left(\eta_{orig}, t_{i+1}\right) + \epsilon_{i+1} \tag{11}$$

This exercise was performed on all 315 patients for all available trough samples. Missing concentration samples were reused as missing data. This resulted in two parallel datasets: a first dataset of dose and resulting concentration per patient per day as originally performed in reality by physicians in standard of care, and a second hypothetical dataset where the dose was calculated through MIPD. Both arms could then be compared graphically and using statistical methods.

In collaboration with physicians, an *improvement* by MIPD was qualitatively defined as: more patients with

trough concentrations in the target window of 12–15 ng/ml, faster target attainment, and smaller deviations from target. This was quantitatively defined as (i) higher individual probability of target attainment, (ii) faster attainment of 1, 2, 3, ... cumulative days in target, and (iii) smaller overall distance to target on each day, defined in Equation 12:

$$C > 15 \Rightarrow D^2 = (\log C - \log 15)^2 \quad (12)$$

$$C < 12 \Rightarrow D^2 = (\log C - \log 12)^2$$

$$C \in [12, 15] \Rightarrow D^2 = 0$$

Statistical methods and power calculation

Based on the above population simulations, a good description of PK outcome for $N = 315$ individuals was available. This allowed defining statistical tests to quantify the effects in the population. Furthermore, a power calculation was performed to consider what effect size could be significantly proven in a trial.

Dose adaptation performance was expected to be time dependent. Any closed loop system requires some samples to reach the target, and overshoot, undershoot, or "lucky hits" are to be expected. The proposed statistical analysis accounted for these effects.

1. Improvement on individual PTA was evaluated as a Welch's *t*-test. A relative improvement of +33% was deemed clinically relevant.
2. Speed of target attainment was evaluated as a time-to-event (TTE) process with nonproportional hazards. A one-sided Mantel-Haenszel log-rank test on TTE greater than three concentrations in target was used. A minimum relative improvement of +33% fraction of patients reaching target on day 7 was deemed clinically relevant. Power for this test was calculated based on required difference in relative hazard ratio and the expected events over the accrual period of 14 days (see Supplementary Materials for more details).
3. We expected squared log-distance to target to decrease over time. Ideally, the mixed model repeated measurements (MMRMs) model detects a significant reduction of squared log-distance due to MIPD. Power for this test was not calculated, as no established method for power analysis of MMRM models with non-normal outcomes is available as of yet.

Based on the simulation, accurate estimates of the distribution of these statistics were available. These were subsequently used to determine required sampled size to detect the clinically relevant effect.

Clinical trial simulation

Finally, the candidate trial with $N = 200$ patients at 2:1 allocation was evaluated as a clinical trial simulation. Standard of care was not simulated, but rather sampled without replacement from the available $N = 315$ profiles in the retrospective dataset. The MIPD arm was similarly sampled from the profiles in the simulation previously performed. Dropout and missing data were considered as represented realistically in the retrospective dataset. This was repeated 1000 times to characterize the distribution of possible clinical trial outcomes and evaluate Probability of Study Success (PoSS).

Software

Monolix 2019[19] was used to perform modeling, using the Stochastic Approximation Expectation Maximization (SAEM) algorithm complemented with importance resampling to determine −2 log-likelihood. The R version 3.5.2 was used for all data management and simulation tasks, using tdmore version 1.1.[20] Tdmore can be freely downloaded at https://github.com/tdmore-dev/tdmore.

RESULTS

Model building

Base model

Key model parameter estimates are available in Table 1. Relevant diagnostic plots are available in the Supplementary Material. A one-compartment model with oral absorption showed considerable time-dependent bias on individual weighted residual (IWRES) versus time plots, overpredicting early (before day 4) concentrations and underpredicting late concentrations (days 7 and later). Inclusion of hematocrit-normalized concentration improved the fit ($\Delta OFV = -356.83$), but did not reduce time-dependent bias. Concentration-dependent binding was removed without any notable impact ($\Delta OFV = 0.14$). Estimation of IOV on clearance further showed a time-dependent trend, which was most appropriately modeled through Equation 4, yielding a T50 of 38.7 h and $\Delta OFV = -1875$. The pcVPC showed acceptable fit on median; the outer prediction interval improved by adding IIV on T50 ($\omega_{T50} = 125CV\%$, $\Delta OFV = -522$) and correlation between η_v and η_{CL} ($\rho = 0.681$, $\Delta OFV = -145.57$). Additive residual error was subsequently removed without impact to OFV or fit. The model did not further improve through absorption lag time or a two-compartment disposition.

TABLE 1 Parameter estimates for hematocrit-standardized, base, and full model

Parameter	Joint model (OFV = 1594.6*)	RSE	Base model (OFV = 19,669.68)	RSE	Full model (OFV = 19,560.72)	RSE
Typical values						
Ka [/h]	4.5	fix	4.5	fix	4.5	fix
V [L]	562	2.9%	767	3.2%	760	3.1%
*CL*0 [L/h]	17.8	2.5%	27.6	2.6%	27.2	2.5%
*T*50 [h]	19.5	6.7%	26.4	5.6%	25.7	6.3%
Hct baseline [%]	0.467	0.49%				
Hct E_{max} [%]	0.188	0.71%				
Hct *T*50 [h]	1.23	20%				
Covariate effects						
Hematocrit on CL					−0.461	0.33%
Weight on *CL*					0.571	NaN%
Weight on *V*					0.536	0.21%
Interindividual variability						
V	57.4%	4.8%	62.8%	4.9%	60.6%	5%
*CL*0	53.8%	4.2%	55.7%	4.2%	52.9%	4.3%
*T*50	136%	6.9%	117%	6.2%	123%	6.8%
Hct baseline	7.56%	4.4%				
Hct E_{max}	5.7%	11%				
Hct *T*50	710%	7.8%				
Correlations						
corr *V*, *CL*0	0.715	4.6%	0.671	5.5%	0.68	5.5%
corr HctT50, HctBaseline	−0.69	8.2%				
corr *T*50, HctBaseline	−0.375	17%				
corr *T*50, HctT50	0.5	15%				
Residual error						
Proportional	0.183	1.3%	0.187	1.3%	0.185	1.3%

Note: Interindividual CV% was calculated as exp(omega)-1. Relative standard error (RSE) was determined through importance resampling. RSE for the effect of bodyweight on *CL* could not be determined numerically. *Includes hematocrit observations.

Abbreviations: CL, clearance; CV%, percent coefficient of variation; Emax, maximum effect; Ka, absorption rate constant; OFV, objective function value; V, volume of distribution.

Predictive performance

Prospective evaluation showed high prediction error (RMSE of 0.361) due to bias introduced by using LOCF for hematocrit. Joint modeling of tacrolimus and hematocrit through Equation 3 retained good population fit and greatly improved predictive performance (RMSE of 0.307, p value 0.004). By moving from three to six estimated individual parameters, simulation time increased 40-fold. For further simulations, the hematocrit-standardized model was removed. This increased OFV by 213 points but did not significantly decrease predictive performance (RMSE of 0.325, $p = 0.172$).

Covariate search

Stepwise covariate modeling (SCM) is described in Supplementary Table S1. As significant covariates, we identified hematocrit on clearance ($\Delta OFV = -79.65$), bodyweight on clearance ($\Delta OFV = -13.48$), and age on clearance ($\Delta OFV = -10.04$).

Model-predictive control

Autocorrelation of the base model residual error was $E[\rho_k] = 0.52, 0.37, 0.24, 0.16, 0.13, 0.12, 0.10$, and 0.12 for

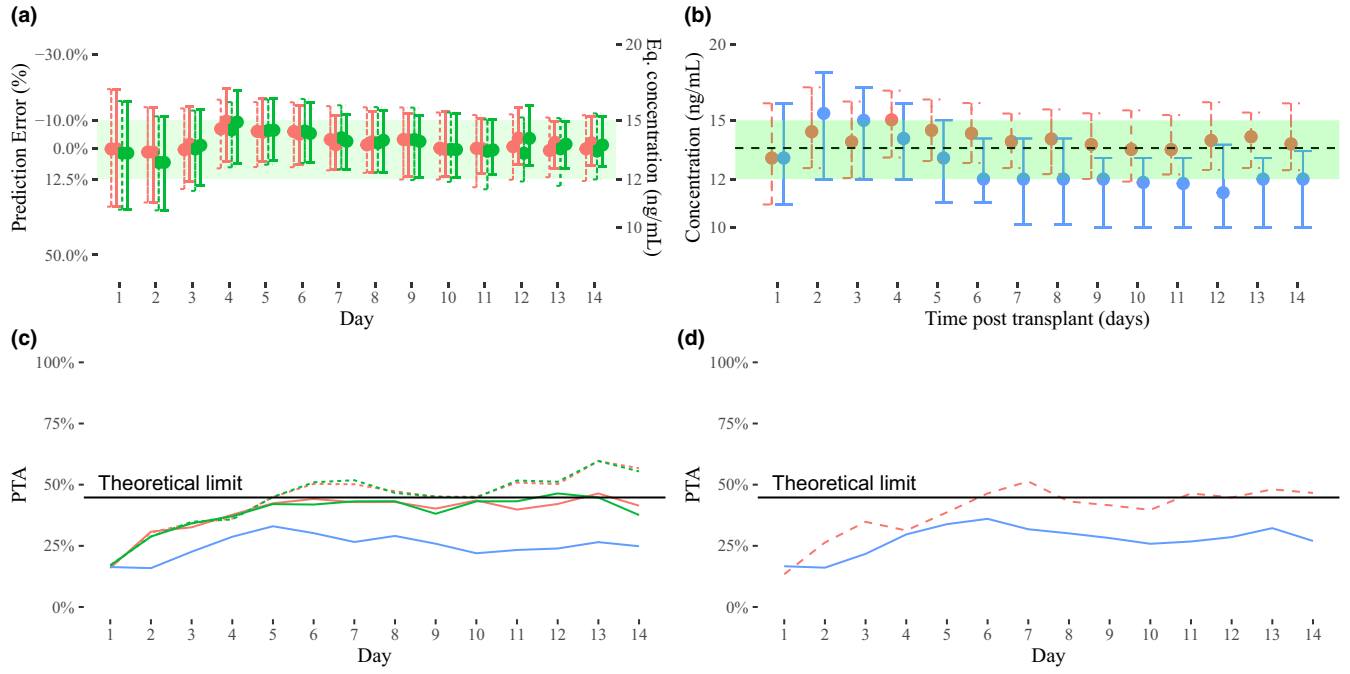

FIGURE 2 Relative prediction error (a) and predicted concentrations (b) median and 50% prediction interval for physician (blue), base (red), and full (green) model, using MPC (solid line) and EBE (dotted line). The target window is represented as green area. Probability of target attainment is shown for the relative prediction error (c) and predicted concentration (d). The theoretical limit is derived from the model residual error. EBE, empirical Bayesian estimation; MPC, model-predictive control; PoSS, Probability of Study Success

k-values of 1 (autocorrelation between subsequent observations) to 8 (autocorrelation between observations 8 days apart). This points to correlated consecutive residual errors. Predictive performance of MPC/MIPD is shown in Figure 2a. Compared to EBE, MPC/MIPD shows a significant improvement in predictive performance. This applies to both base and full models. Using these results, the base model with MPC/MIPD estimation was selected as the optimal approach, at RMSE of 0.304 ($p = 0.432$ vs. hematocrit-standardized model, $p = 0.148$ vs. base model with EBE estimation).

Based on the identified residual error, the theoretical upper limit for target attainment is 45.2%. Physician performance averaged 25% PTA, with clear underdosing visible in Figure 2b. For Bayesian estimation, PTA (Figure 2c) approached the theoretical upper limit on 6 out of 14 days, whereas MPC/MIPD exceeded this limit. There is no apparent bias visible in Figure 2a, as mean prediction error is close to 0. The full model did not outperform the base model. We opted to use the base model in further simulations, as collecting covariates was not worth the increased clinical workload.

Simulation of model-informed precision dosing

The base model was used with MPC/MIPD to simulate dose adaptation and resulting concentrations. The dose adaptations performed by physicians and MIPD are compared in Figure 3. Although physicians adapted conservatively, MIPD applied a temporary overcorrection of the dose in order to reach target concentration as fast as possible. Figure 2 shows a summary of concentration per day. MIPD resulted in a large PTA, as well as overall concentrations closer to the target window. Per-patient PTA was at 39% ± 15.8% for MIPD (mean ± SD), whereas physician PTA was at 28% ± 16.1%. The TTE curves for "X observations in the target window" are shown in Figure 4. The difference for reaching ">1 day in target" is quite small, with only a 1-day delay between arms on average. This delay grows larger, with ">3 days in target" being reached for 50% of the population on day 8 for the intervention arm, whereas only at day 10 for the control arm. On day 7, 48.2% ± 5.5% reached greater than 3 days in target for MIPD, whereas only 27.5 ± 3.51% reached this for the physician arm. Finally, the Kolmogorov-Smirnov (KS) test identified a significant reduction in squared log-distance to target window for every day. Log-squared distance to target was normally distributed after Box-Cox transformation, allowing the application of an MMRM analysis. This identified a significant treatment effect, yet only at a relative improvement of −13%. Squared log-distance to target window decreased from 0.080 ± 0.202 to 0.055 ± 0.191 (mean ± SD).

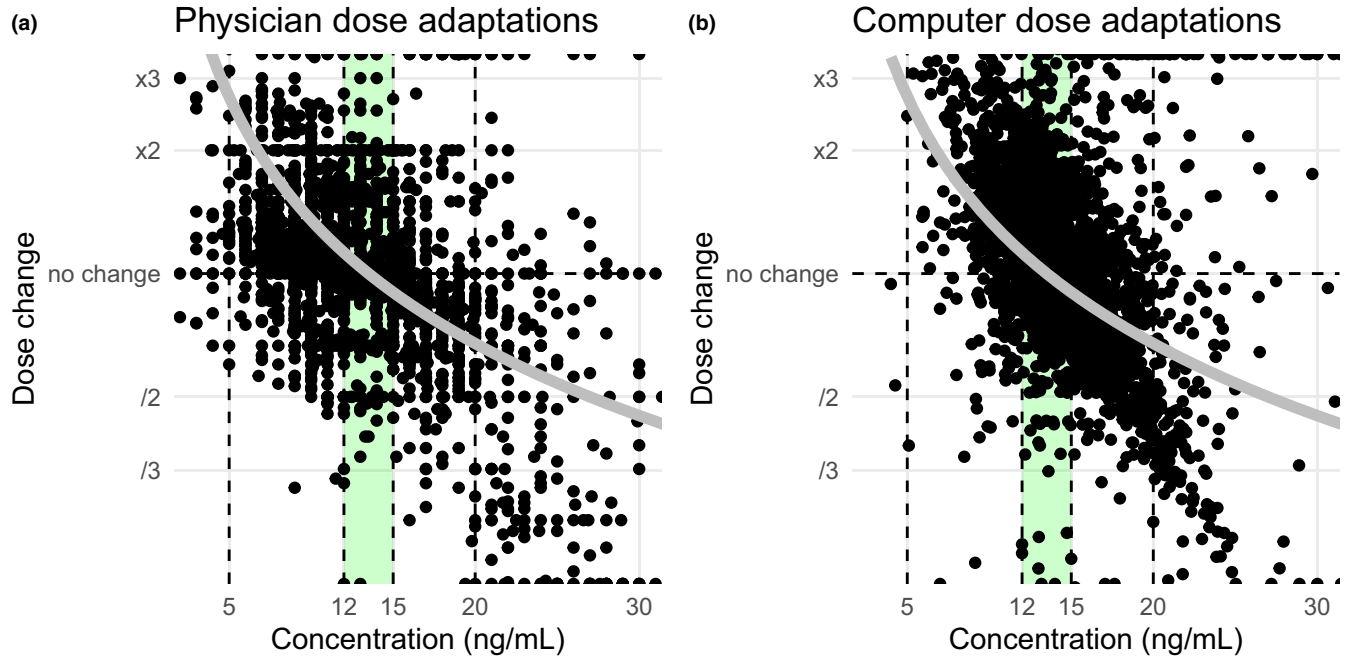

FIGURE 3 Physician dose adaptation (left) versus computer dose adaptation (right). The observed concentration (*x*-axis) results in a dose change (*y*-axis). The grey line shows the theoretical dose adaptation when following the rule of three in steady-state

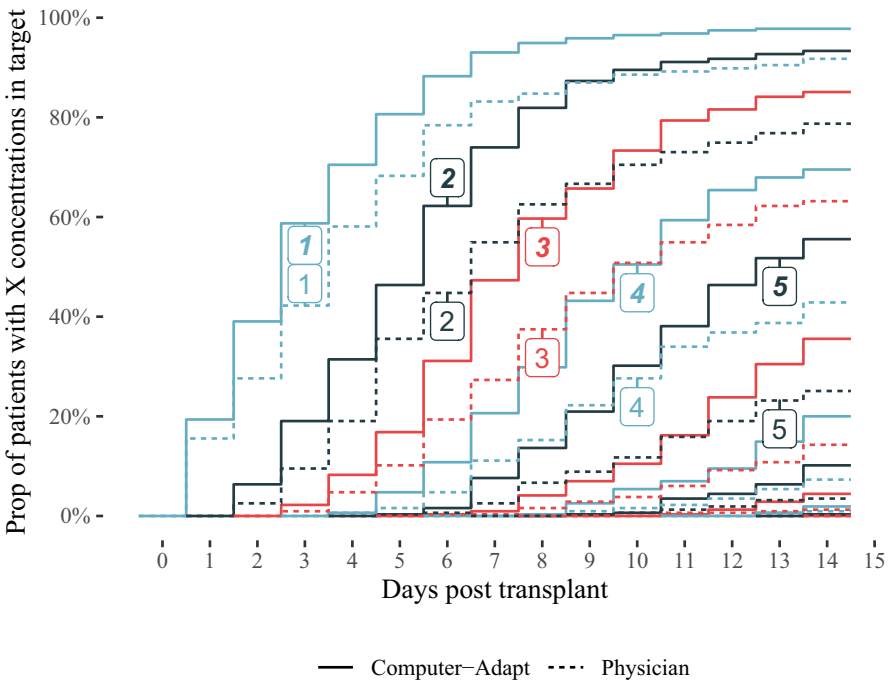

FIGURE 4 Proportion of patients with at least X concentrations in target, per day. Computer (solid line, bold font) versus physician (dotted line, normal font)

Power calculation

Based on the above estimates, study power and minimum detectable effect sizes are presented in Figure 5. The candidate trial of $N = 200$ will reliably detect a PTA improvement at $p < 0.01$. The clinically relevant PTA can be detected at $p < 0.01$ with $N = 145$ patients. An improvement on TTE less than +50% may not be reliably detected by a trial with $N = 200$ patients, yet the true effect will be detected even at $p < 0.01$. On log-squared distance to target, the population simulation showed a true effect size below the clinically relevant limit.

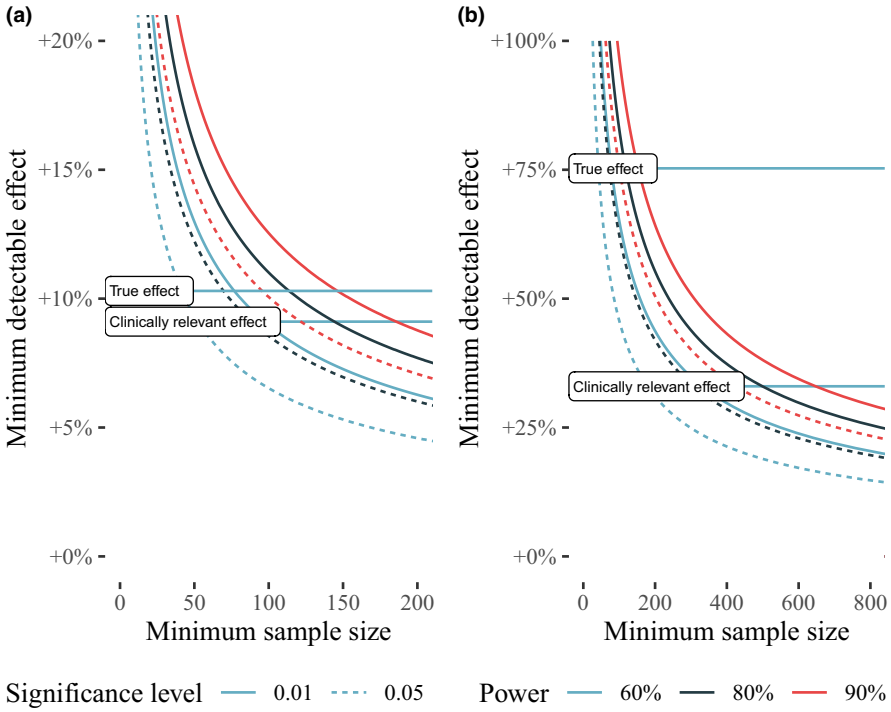

FIGURE 5 Minimum detectable effect size for difference in PTA (a) and difference in speed of reaching three concentrations in target (b). Horizontal lines show the true effect and minimum clinically relevant effect. PTA, probability of target attainment

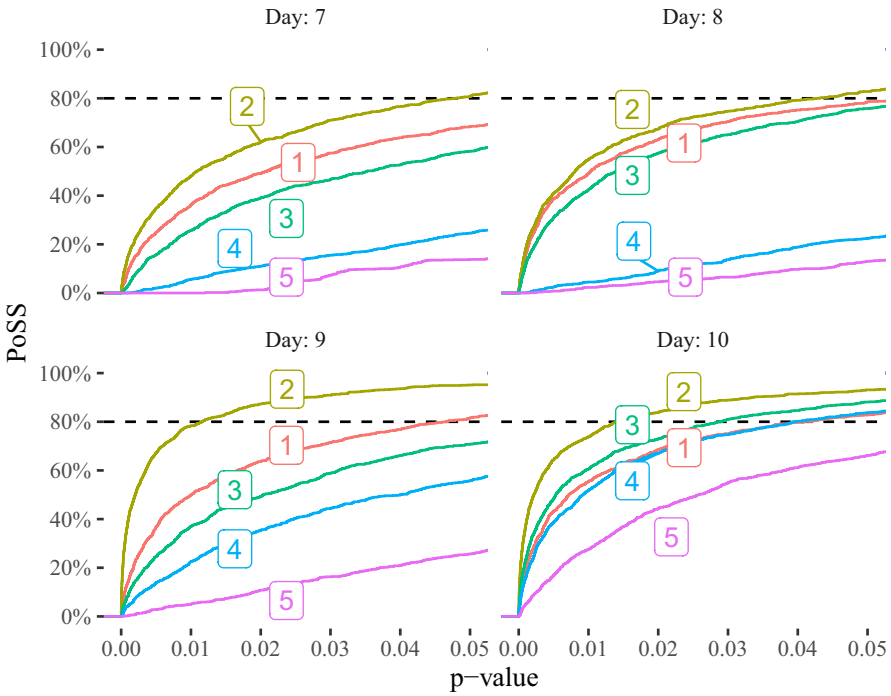

FIGURE 6 Power to detect an improvement in "time to reach >X concentrations in target" ($X = 1$–5) after 7, 8, 9, or 10 days, using the Mantel-Haenszel test. PoSS, probability of study success

Clinical trial simulation

Study power to detect these three aforementioned end points was simulated using a bootstrap of 1000 random trials of 200 patients each. It was trivial to show an improvement in average PTA per patient, with a 100% PoSS at $p < 0.01$. The average expected effect size was 11.3% (8%–14%), with per-trial 95% lower confidence limit of 7.5% (4%–10.5%). Study power for the TTE test is shown in Figure 6. PoSS depended on the day and the end

point at which both arms were compared. It was decided to evaluate "3-days-in-target" at day 8 post-transplant, which yielded a PoSS of 90% at $p < 0.01$. Looking at squared-log-distance-to-target, even the sensitive KS-test could only identify an improvement on days 3, 7, 8, and 10 post-transplant. Only at these days was PoSS greater than 80% for $p < 0.01$. Using MMRM analysis, an overall improvement could be reliably shown, although per-day effects could only be reliably shown on days 6, 7, 8, and 11 at $p < 0.05$.

DISCUSSION

To the best of our knowledge, this is the first example of predicting a clinical trial outcome comparing MIPD to standard of care, and the use of this prediction to optimize a future planned prospective clinical trial. To achieve this, a population PK model was first built and evaluated for GOF on the target population. A pragmatic approach was presented to incorporate unexplained variability in individual parameters. The full simulation code was implemented in a reusable R package. Finally, the predicted results were analyzed to describe the statistical power of a candidate clinical trial design, allowing optimization of said trial design.

Overall, the population PK model is in reasonable agreement with literature. Describing tacrolimus PK by a one-compartment model with oral absorption is common in the absence of rich concentration-time profiles.[21] Independent groups identified similar time-dependent clearance early post-transplantation.[12,22,23] Others identified a time-dependent increase over several weeks post-transplantation,[14,17,24,25] which is likely a different effect altogether. Parameter estimates are broadly in agreement with results from similar studies focusing on the first 14 days post-transplantation.[12] Identified covariates are also in agreement with previous studies,[21] although some identified a large difference in bioavailability between Prograft and Advagraf formulations. The identified power factor $\beta_{CL,WT}$ of 0.348 results in a 77% and 122% adjustment of clearance for the lightest and heaviest patients (33.5 and 125 kg, respectively) in the study, which in light of IIV 55.7% explains the minimal difference between base and full model predictive performance. The identified IIV is high, which we attribute to the poor PK stability of patients early post-transplantation. This agrees with other studies focusing on the same study period.[12,26,27]

Notably, the inclusion of hematocrit at first resulted in poor predictive performance. When the full profile of a time-varying covariate is not available during prospective evaluation, significant bias may be introduced. Joint modeling of both drug concentration and covariate is required to overcome this limitation. This markedly increased the computation times. Although model simplification resulted in a penalty to OFV, predictive performance was not significantly impacted. Notably, applying MPC/MIPD again resulted in low RMSE, rivaling the more complex model with more covariates, at feasible calculation times.

McDougall et al.[28] extensively explored the impact of model misspecification on precision dosing. They demonstrated that only severe model misspecification significantly impacts model-based precision dosing performance. This reasoning also applies to covariate models; covariates difficult to collect can be omitted without impact to model predictive performance. This further exemplifies the necessity to include prospective evaluation in the diagnostic toolset when developing models for precision dosing.

Tacrolimus PK has typically been described by a two-compartment model when rich data are used.[7] In this case, however, the use of two-compartment kinetics would not result in different results. The typical distribution phase is less than 12 h, and therefore no information on the distribution phase is present in daily trough concentrations, even with multiple dosing. There is an ongoing debate on appropriate PK targets for tacrolimus, with some evidence pointing to AUC as a superior metric.[4] Our trough dataset cannot be used to accurately predict AUC.[29] With rich data and an appropriate model, the presented approach may be applied to evaluate the accuracy (and improvement over standard of care) when using one, two, or more blood samples per day.

CYP3A5 genotype is missing from our current model, as it was not generally measured in transplant recipients at UZ Leuven hospital. It is worthwhile to evaluate the inclusion of this covariate, and to quantify the potential improvement in MIPD dosing accuracy. If this covariate is not available, a mixture model could be used to estimate individual CYP3A5 expression probability. However, due to the low number of CYP3A5 expressors in the target population, we expect the impact to be low and transient.

In this work, we argue that classical GOF evaluation is not appropriate when building a model for MIPD. Even though the full model is a significantly better description of the data as compared to the base model, this did not result in a significant improvement to predictive performance or PTA. Clinically, it is preferable to omit covariates that are cumbersome to measure, if they do not improve predictive performance significantly. In general, we argue that predictive performance assessment is a key step when building MIPD models.

However, contrary to the well-studied classical GOF evaluation, it is unclear how a model with poor predictive performance can be improved. As a first step, we suggest to include predictive performance evaluation in

model building software. The Perl-Speaks-Nonmem suite recently added the proseval tool, but lacks clear standard graphs to represent this data.

The MPC/MIPD method merits further discussion. Correlated residual errors in tacrolimus models were previously identified by Størset et al.,[17] who reported that bioavailability varied less between subsequent occasions. Correlated IOV, which could otherwise be classified as "parameter drift," has not been studied in detail. Pragmatic solutions include down-weighing older concentration samples or arbitrarily increasing ω during estimation. The novel idea of adapting the estimation method rather than the model resulted in a significant improvement to predictive performance in this dataset.

The predicted outcome, in the form of predicted tacrolimus trough concentrations for the MIPD arm, differs from the estimated model predictive performance for three reasons. First, an accurate prediction does not necessarily imply a future concentration in target. We can use the first dose recommendation as an example: the loading dose is too high in 170 out of 315 patients, and using the first trough concentration at 08:00 on day 1, the computer recommends a 0 mg evening dose. The computer predicts this dose will still result in too high concentrations the following morning. Second, the opportunity to adapt the dose may be far into the future, challenging the predictive performance of the model under high within-patient variability. When the dose adaptation is performed at 12:00, the Prograft administrations of 20:00 evening and 08:00 the following morning can be adapted, with the latest trough therefore at 20:00 the next day. For Advagraf, however, only the dose at 08:00 the following morning can be adapted, with trough at 08:00 2 days into the future. This increases prediction error and therefore reduced probability of target attainment. Finally, the prediction error does not directly translate to an error in resulting concentration after MIPD for non-steady-state. This highlights the importance of using PK models for dose adaptation: dose adaptation tables using dose-normalized concentration fail to capture the highly variable and non-steady-state nature of the first 2 weeks posttransplantation.

Figure 3 shows a computer algorithm performs aggressive dose adaptation, in stark contrast to conservative dose adaptation by physicians, who seem to be more cautious in this respect. We offer three explanations for this behavior. First, in cerebro modeling assumes steady-state, and therefore cannot correctly relate a wildly varying dosing history and concentrations to the required dose adaptation. Second, it is difficult for humans to capture PK dose-linearity. If the concentration is 50% below target, the dose should be doubled. Instead, we see slow up-titration by absolute steps, rather than, for example, doubling or halving the dose, contrary to current research advising against tacrolimus underexposure.[4] Finally, we identified time-dependent clearance during the first week post-transplantation. Even when gradually increasing the dose, doctors are chasing a moving goalpost. MIPD does not suffer from any of these shortcomings.

In contrast, MIPD even uses a corrective dose to ensure the target trough concentration is reached as soon as possible. It remains unclear whether this practice is beneficial in real life. First, there is an ongoing discussion on the validity of trough concentration as a PK target.[30] Targeting the area under the curve (AUC) may be more appropriate.[31] A recent study by Miano et al.[3] identified a PK/PD association for safety. A 54% increase in acute kidney injury was identified per 5 ng/ml increase in average tacrolimus trough concentrations over the previous 3 days. A similar association for efficacy could not be identified. Adding clinical utility (CU), a model integrating PK/PD/CU could focus on the true benefit for patients, rather than improvement on surrogate end points with only weak association to clinical benefit.

In this work, we demonstrated clearly that simulated MIPD concentrations can serve to refine the definition of trial end points. This allowed us to explore beyond mere "improvement in average PTA" and evaluate end points, such as "speed of target attainment" and "distance to target window." It was not possible to detect an improvement consistently on each separate day, even using advanced statistical techniques, such as MMRM. Only a consistent effect across all days could be shown reliably. All things considered, the proposed techniques dig deeper into MIPD performance than evaluating odds ratios of PTA. As of yet, the relevance of the presented surrogate end points and their relevance to clinical outcome is based on empiric evidence only. It is unfortunate that no PK/PD model predicting clinical outcomes has been identified. Such a model could serve to replace a naïve therapeutic drug monitoring approach targeting a therapeutic window, and instead directly find the appropriate dose to target a desired PD effect reaching optimum efficacy and toxicity. This model may also serve to design a concentration-controlled trial quantifying the clinical impact of MIPD.

A randomized controlled trial is always comparative in nature. Therefore, the presented results cannot easily be translated to other hospitals, as the standard of care differs widely between hospitals. As an example, steroid concomitant therapy was identified to influence tacrolimus PK,[12,13,32] but different treatments of high-dose steroids are in use at different hospitals. Furthermore, there is no clear evidence that retrospective data will be similar to standard of care performance in a comparative trial. Standards may have improved with increased experience, and a clinical trial setting may invite physicians to more

carefully perform dose adaptation to achieve the target window.

In conclusion, this work offers new insights into the use of simulation to predict and optimize MIPD for tacrolimus dose adaptation. We have shown the validity of predictive performance as a tool for model selection. MPC/MIPD was proposed as a method to incorporate unexplained but autocorrelated IOV. Retrospective data was used to fully simulate a hypothetical MIPD arm, which was then used to quantitatively analyze the improvement the technique offers. Finally, the simulated data was used to calculate trial power and optimize said trial. The simulation software was implemented as an open-source R package, allowing to repeat this exercise with any model. By making this software available, we hope quantitative predictions on MIPD become within reach, allowing to identify where this technology can benefit clinical care the most.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the contributions of Dr. Thomas Van Hove in collecting the dataset, Ir. Quentin Leirens in supporting the development of the software, and Dr. Daniel Röshammar for reviewing the manuscript. We recognize the support of Annouschka Laenen at the KU Leuven University Biostats department in providing feedback to the study statistical analysis plan.

CONFLICT OF INTEREST

R.F. previously worked for Astellas Pharma in 2017 as a paid consultant on an unrelated compound in clinical development. All other authors declared no competing interests for this work.

AUTHOR CONTRIBUTIONS

R.F., P.A., and D.K. wrote the manuscript. R.F., D.K., and T.B. designed the research. R.F. and T.B. performed the research. R.F. analyzed the data. R.F. and N.L. contributed new analytical tools.

REFERENCES

1. Holford NHG, Buclin T. Safe and effective variability—A criterion for dose individualization. *Ther Drug Monit*. 2012;34:565-568.
2. Vincenti F, Jensik SC, Filo RS, Miller J, Pirsch J. A long-term comparison of tacrolimus (FK506) and cyclosporine in kidney transplantation: evidence for improved allograft survival at five years1. *Transplantation*. 2002;73:775-782.
3. Miano TA, Flesch JD, Feng R, et al. Early tacrolimus concentrations after lung transplant are predicted by combined clinical and genetic factors and associated with acute kidney injury. *Clin Pharmacol Ther*. 2020;107:462-470.
4. Brunet M, van Gelder T, Åsberg A, et al. Therapeutic drug monitoring of tacrolimus-personalized therapy: second consensus report. *Ther Drug Monit*. 2019;41:47.
5. Wallemacq P, Armstrong VW, Brunet M, et al. Opportunities to optimize tacrolimus therapy in solid organ transplantation: Report of the European consensus conference. *Ther Drug Monit*. 2009;31:139-152.
6. Vanhove T, Annaert P, Kuypers DRJ. Clinical determinants of calcineurin inhibitor disposition: a mechanistic review. *Drug Metab Rev*. 2016;48:88-112.
7. Brooks E, Tett SE, Isbel NM, Staatz CE. Population pharmacokinetic modelling and Bayesian estimation of tacrolimus exposure: is this clinically useful for dosage prediction yet? *Clin Pharmacokinet*. 2016;55:1295-1335.
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;56:1491-1498.
9. Størset E, Åsberg A, Skauby M, et al. Improved tacrolimus target concentration achievement using computerized dosing in renal transplant recipients-a prospective, randomized study. *Transplantation*. 2015;99:2158-2166.
10. Lindbom L, Pihlgren P, Jonsson N. PsN-Toolkit—A collection of computer intensive statistical methods for non-linear mixed effect modeling using NONMEM. *Comput Methods Programs Biomed*. 2005;79:241-257.
11. Vanhove T, Hasan M, Annaert P, Oswald S, Kuypers DRJ. Pretransplant 4β-hydroxycholesterol does not predict tacrolimus exposure or dose requirements during the first days after kidney transplantation. *Br J Clin Pharmacol*. 2017;83:2406-2415.
12. Antignac M, Barrou B, Farinotti R, Lechat P, Urien S. Population pharmacokinetics and bioavailability of tacrolimus in kidney transplant patients. *Br J Clin Pharmacol*. 2007;64:750-757.
13. Velickovic-Radovanovic R, Catic-Djordjevic A, Milovanovic JR, Djordjevic V, Paunovic G, Jankovic SM. Population pharmacokinetics of tacrolimus in kidney transplant patients. *Int J Clin Pharmacol Ther*. 2010;48:375-382.
14. Han N, Yun H-Y, Hong J-Y, et al. Prediction of the tacrolimus population pharmacokinetic parameters according to CYP3A5 genotype and clinical factors using NONMEM in adult kidney transplant recipients. *Eur J Clin Pharmacol*. 2013;69:53-63.
15. Jusko WJ, Piekoszewski W, Klintmalm GB, et al. Pharmacokinetics of tacrolimus in liver transplant patients*. *Clin Pharmacol Ther*. 1995;57:281-290.
16. Størset E, Holford N, Midtvedt K, et al. Importance of hematocrit for a tacrolimus target concentration strategy. *Eur J Clin Pharmacol*. 2014;70:65-77.
17. Størset E, Holford N, Hennig S, et al. Improved prediction of tacrolimus concentrations early after kidney transplantation using theory-based pharmacokinetic modelling: Theory-based tacrolimus PK modelling. *Br J Clin Pharmacol*. 2014;78:509-523.
18. Krieger A, Pistikopoulos EN. Model predictive control of anesthesia under uncertainty. *Comput Chem Eng*. 2014;71:699-707.
19. Monolix version 2019R1. https://lixoft.com/products/monolix/.
20. Faelens R, Luyckx N, Leirens QT. (2020). https://github.com/tdmore-dev/tdmore
21. Campagne O, Mager DE, Tornatore KM. Population pharmacokinetics of tacrolimus in transplant recipients: what did we learn about sources of interindividual variabilities? *J Clin Pharmacol*. 2019;59:309-325.

22. Han N, Ha S, Yun H-Y, et al. Population pharmacokinetic-pharmacogenetic model of tacrolimus in the early period after kidney transplantation. *Basic Clin Pharmacol Toxicol*. 2014;114:400-406.
23. Zuo X, Ng CM, Barrett JS, et al. Effects of CYP3A4 and CYP3A5 polymorphisms on tacrolimus pharmacokinetics in Chinese adult renal transplant recipients: a population pharmacokinetic analysis. *Pharmacogenet Genomics*. 2013;23:251-261.
24. Bergmann TK, Hennig S, Barraclough KA, Isbel NM, Staatz CE. Population pharmacokinetics of tacrolimus in adult kidney transplant patients: impact of CYP3A5 genotype on starting dose. *Ther Drug Monit*. 2014;36:62-70.
25. Golubović B, Vučićević K, Radivojević D, Kovačević SV, Prostran M, Miljković B. Total plasma protein effect on tacrolimus elimination in kidney transplant patients – Population pharmacokinetic approach. *Eur J Pharm Sci*. 2014;52:34-40.
26. Staatz CE, Tett SE. Clinical pharmacokinetics and pharmacodynamics of tacrolimus in solid organ transplantation. *Clin Pharmacokinet*. 2004;43:623-653.
27. Musuamba FT, Mourad M, Haufroid V, et al. Statistical tools for dose individualization of mycophenolic acid and tacrolimus co-administered during the first month after renal transplantation. *Br J Clin Pharmacol*. 2013;75:1277-1288.
28. McDougall DA, Martin J, Playford EG, Green B. The impact of model-misspecification on model based personalised dosing. *AAPS J*. 2016;18:1244-1253.
29. Op den Buijsch RAM, van de Plas A, Stolk LML, et al. Evaluation of limited sampling strategies for tacrolimus. *Eur J Clin Pharmacol*. 2007;63:1039-1044.
30. Bouamar R, Shuker N, Hesselink DA, 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;13:1253-1261.
31. Kuypers DRJ, Claes K, Evenepoel P, Maes B, Vanrenterghem Y. Clinical efficacy and toxicity profile of tacrolimus and mycophenolic acid in relation to combined long-term pharmacokinetics in de novo renal allograft recipients. *Clin Pharmacol Ther*. 2004;75:434-447.
32. Press RR, Ploeger BA, den Hartigh J, et al. Explaining variability in tacrolimus pharmacokinetics to optimize early exposure in adult kidney transplant recipients. *Ther Drug Monit*. 2009;31:187-197.

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: Faelens R, Luyckx N, Kuypers D, Bouillon T, Annaert P. Predicting model-informed precision dosing: A test-case in tacrolimus dose adaptation for kidney transplant recipients. *CPT Pharmacometrics Syst Pharmacol*. 2022;11:348-361. doi:10.1002/psp4.12758