Effects of Last Dosing Time on Population Pharmacokinetic Analysis of Tacrolimus Performed using Extracted Data from Electronic Health Records

Michael L. Williams[1], Hannah L. Weeks[1], Cole Beck[1], Leena Choi[1*]

[1]Department of Biostatistics Vanderbilt University Medical Center, Nashville, TN.

Abstract

A tacrolimus population pharmacokinetic (PK) study with 440 subjects was performed using real world data extracted from electronic heath records (EHRs) to identify patient factors affecting variability in tacrolimus PK parameters. Using last-dose times extracted by our own natural language processing system, *medExtractR*, we assessed the effects of incorporating last-dose times in the data on tacrolimus PK parameter estimates. The estimates of population level PK parameters were Cl = 24.9 L/h [23.5, 26.3] and V = 6310 L [5461, 7159]. There was no appreciable difference in parameters estimates with *vs.* without last-dose time incorporated in the data. We also investigated the effects of absorption rate constants that are often fixed at a published value in tacrolimus population PK analysis. Our sensitivity analysis revealed little difference between parameters estimated assuming a range of absorption rate constants.

***Corresponding Author**

Leena Choi, PhD
2525 West End, Ste. 11000
Nashville, TN 37203
E-mail: leena.choi@vanderbilt.edu
Phone: 615.343.3497, Fax 615.343.4924

Key Words: Population pharmacokinetics, electronic health records, natural language processing system, last dosing time, pragmatic research, opportunistic sampling

Introduction

Tacrolimus is an immunosuppressive calcineurin inhibitor widely used to prevent rejection following organ transplantation. Tacrolimus has a narrow therapeutic window, yielding less efficacy (i.e., an increased probability of implant rejection)[1] or toxicity (e.g., nephrotoxicity, neurotoxicity, diabetogenicity)[2,3] if it falls outside of this window. The therapeutic window could be considered differently depending on timeline since transplantation (e.g., 10-20ng/ml in the first month, 5-15ng/ml in the second and third months, 5-10ng/ml in subsequent months post-transplantation).[4] The challenges of maintaining this therapeutic range is exacerbated by individual variability in tacrolimus disposition. Several patient characteristics affecting clearance have been reported: examples include liver function,[5] time since transplantation,[6] and age.[7] Especially, the effect of single nucleotide polymorphisms (SNPs), rs776746, on CYP3A5 gene coding the major metabolizing enzyme for tacrolimus, has been reported from multiple studies including our own.[8] As patients carrying the CYP3A5*3 SNP, a loss-of-function allele, reduce clearance, these patients require significantly smaller tacrolimus doses to maintain the same tacrolimus concentration level.[9-13] This SNP alone was reported to be associated with 39% of the variability in dose requirement between patients.[8]

Tacrolimus population pharmacokinetic (PK) studies have traditionally relied on prospective observational studies performed during routine care. While such studies are not as costly or restrictive as randomized controlled trials, they still impose costs associated with data collection and restrictions on patient enrollment.[14] Postmarketing drug studies have been increasingly performed using real-world data sources such as electronic health records (EHRs) which are particularly valuable as a source of longitudinal clinically relevant data. The adoption of EHRs has made large-scale retrospective studies feasible with rapid cohort generation from patients enrolled in routine care, and there has been increasing interest in using EHRs as real word data.

Due to the narrow therapeutic window with high variability in tacrolimus disposition, it is one of the drugs that require therapeutic drug monitoring (TDM).[15] Tacrolimus blood concentrations are routinely collected as part of standard of care, making retrospective studies with EHR data more appealing. Several tacrolimus PK studies have been conducted using such data.[16-18] As tacrolimus population PK studies are performed retrospectively using trough drug concentrations, they often assume the absorption rate constant, k_a, at a fixed value in the PK models. Although some studies may perform a sensitivity analysis, the effect of this assumption has not been thoroughly investigated.

We recently developed a natural language processing system, *medExtractR*,[19] to extract medication information from free-text clinical notes as part of a system to enable the use of EHRs in retrospective studies of drugs.[20,21] The system, once finalized, should relieve the primary burden in data generation which often involves the manual extraction of medication data. In addition to drug dosing information, *medExtractR* is designed to extract explicit last dosing times if present in the notes. Tacrolimus population PK studies typically use trough concentrations that are assumed to be measured before taking a morning dose and assumed to be preceded by a previous dose at a regularly scheduled time (i.e., twice a day). Any deviation of this assumed schedule from the true dosing time may result in a different estimate of the PK parameters. To the best of our knowledge, no studies have investigated the sensitivity of tacrolimus PK parameter estimates to deviations in last-dosing time.

The goals of this study were threefold: (1) to perform tacrolimus population PK analysis using real world data generated with EHRs and identify the important factors affecting PK profile; (2) to investigate the effects of last dosing time on the tacrolimus PK parameter estimates using the data extracted from the EHRs; and (3) to investigate the effects of assuming the absorption rate constant, k_a, at a fixed value in a model on the other PK parameter estimates.

Methods

Study Design and Data Source

This study was approved by the Vanderbilt Institutional Review Board. We used data obtained from the previous study, which was described in detail in Birdwell *et al.*[8] The data included in the tacrolimus PK modeling were medication dose, drug concentration levels, demographics, laboratory, and genotype data. Additionally, we extracted the last dosing time from the same clinical notes that yielded the original dataset.

Times of last dose were extracted using *medExtractR* (see Weeks *et al.* 2020 for details).[19] Extracted last dose times appeared in various formats, for example using am or pm (e.g., "10 am", "9:30 pm"), military time (e.g., 2200, 20:30), or using a modifying word or phrase (e.g., "8 last night", "yesterday morning at 7"). **Figure 1** outlines the algorithm to process these extractions. All time expressions were initially converted into the same format of HH:MM:SS. For example, the phrase "8:30pm" would become "20:30:00". Concentration measurements were generally assumed to be trough levels taken at a morning appointment. For this reason, PM last dose times were assumed to have occurred on the previous day and AM times were assumed to have occurred on the same day as the laboratory value. Conflicts could occur if different last dose times were extracted within the same note or on the same date. Assuming concentration measurements were intended to be trough values, time differences within 2 hours were considered close enough to be equivalent with respect to impact on the PK parameter estimates and the earlier extracted time was kept. Cases where discrepancies still existed after removing differences within 2 hours were manually reviewed with a clinical expert to determine which last dose time was correct. Ambiguous cases where the correct last dose time could not be determined were treated as missing.

Data Process

We used the original data (which do not have NLP-extracted last-dose times) and the last-dose times we extracted from the clinical notes for the same cohort to build four datasets: two datasets for the entire cohort with and without last-dose time, and two for a reduced cohort with and without last-dose time. The reduced cohort was defined based on the number of extracted last-dose times for each subject. Subjects having at least 4 extracted last-dose times were included in the reduced cohort. Otherwise, the PK datasets for both cohorts were built with the same algorithm implemented as a function in the *EHR* R package. A brief description of the PK data building method with and without last-dose time is as follows.

When last-dose time is not available, a regular dosing interval is assumed. Specifically, we assume that a dose of tacrolimus is taken 30 minutes (a reasonable timeframe confirmed by a clinician) after blood is drawn for a concentration observation. We further assume that the drug is taken every 12 hours following that initial dose. As most drug level measurements are being taken in the morning to provide trough concentration level, subjects are assumed to take their

morning dose 30 minutes after having the blood drawn and continue to take their doses every 12 hours from that point until their next concentration measurement, the timing of which will dictate their next proceeding dose timings. This fully describes the generation of the dataset without last-dose times.

When last-dose time is available, the dataset with last-dose times can be generated. For example, if a last-dose time is found matched with a concentration measurement, the preceding dose is set to be given at the time reflected in the clinical note. This dataset better represents the dosing times, among which the last-dose time is the most informative to estimating PK parameters. Note that not all extracted concentrations need to be associated with a last-dose time. For concentrations missing last-dose times, the record will be the same as the record without last-dose times.

Population PK Analysis

We performed population PK analysis of tacrolimus using a nonlinear mixed-effects model implemented by Monolix® (version 2020) with the stochastic approximation of expectation and maximization (SAEM) method. A one-compartment PK model was chosen as the base model, assuming a combined additive and proportional residual error model and lognormal distribution for the random effects PK parameters. A model with random effects with unstructured covariance for all main PK parameters except for the absorption rate constant, k_a, was assumed in the final model. As k_a cannot be reliably estimated without drug concentrations measured during the absorption phase, it was fixed at the previously published value of 4.5.[22]

Covariate model building was performed using individual specific PK parameters estimated from the base model. Both graphical and statistical methods were considered with the following candidate covariates, which we chose *a priori* based on previous research and biological plausibility: weight, age, sex, hemoglobin, albumin, race, and a CYP3A SNP (rs776746). Model selection was performed based on the objective function (-2 log likelihood) along with the number of parameters, which would approximately follow χ^2 distribution. The χ^2 statistics of 3.84 and 6.64 with 1 degree of freedom correspond to p values of 0.05 and 0.01, respectively. Thus, we considered the objective function value decrease of 6.64 to be significant model improvement. Variables were added to the base model one-by-one and evaluated for a decrease in the objective function value of > 6.64. Variable selection was performed only with the dataset of the entire cohort with last-dose times; selected variables were then used to build the same model from the remaining three datasets. The model was qualitatively assessed through visual examination using goodness-of-fit plots such as the observed vs. predicted concentrations, the individual weighted residuals, and the visual predictive check using Monolix® (version 2020).

Sensitivity Analysis

The k_a in our final models was assumed to be 4.5 as reported previously reported.[22] In order to assess whether our findings are sensitive to this selected value, we refit the model at another published k_a value of 3.09.[12] In addition, we refit the model with the published k_a value for the extended-release formulation, 0.375.[23]

Results

Population Characteristics for the Two Cohorts

The study cohort was described previously by Birdwell *et al*.[8] The previously described cohort of 446 subjects was reduced to 440 after removing participants with only a single concentration observation (N=4) and participants with no known SNP (N=2). The characteristics of study population for the two cohorts are described in **Table 1**. Each participant in the cohort has a maximum of 10 tacrolimus blood concentration measurements. These observations constitute either the first 10 concentrations or every concentration if less than 10 measurements for each patient beyond the 1-month post-transplant period. Of these concentration measurements, 48% were accompanied by an extracted last-dose time. Median tacrolimus dose across all subjects was 3 mg twice daily, and median blood concentration measurements was 7.0 mcg/L. The reduced cohort consists of 257 subjects. The percentage of concentration measurements associated with a last-dose time is increased to 70% as designed, while median tacrolimus dose and median blood concentration remain almost the same.

Population PK Model

The results of the primary analysis, including the base model and the final covariate model based on datasets from the entire cohort and the reduced cohort, are presented in **Table 2**. PK parameters such as clearance (CL, L/hr) and volume of distribution (V, L) were first estimated from the base model without covariates. Note that we denote CL/F by CL for simplicity, where F represents relative bioavailability and is omitted elsewhere. The PK parameters varied substantially among individual participants; the coefficients of variation (CV%) for CL and V were 42% and 192%, respectively.

CYP3A SNP, weight, and age all substantially improved the model fit. Their inclusion reduced the objective function value by 161, 7.28, and 18.3, respectively. Other *a priori* defined covariates (serum albumin, hemoglobin, and race) explained little of the inter-individual variability for PK parameters and hence minimally improved model fit (i.e., decreased the objective function by < 6.64) and were excluded from the final model. The final model is presented as follows:

$$CL_i=\theta_1 \times (age_i/46)^{\theta_2} \times \exp[\theta_3 \times CYP3A5*3/*1_i] \times \exp[\theta_4 \times CYP3A5*1/*1_i] \times \exp[\eta_i^{CL}],$$

and

$$V_i= \theta_5 \times (wt_i/70)^{\theta_6} \times \exp[\eta_i^{V_1}].$$

where CL_i and V_i are the individual-specific PK parameters for subject *i*, corresponding to CL and V, wgt_i is participant weight in kilograms (kg), age_i is participant age at study entry in years, $CYP3A5*3/*1_i$ and $CYP3A5*1/*1_i$ are indicators for genotype for *3/*1 and *1/*1, respectively, where the reference is genotype for *3/*3. η_i^{CL} and $\eta_i^{V_1}$ are random effects explaining inter-individual variability for CL and V, which follow a normal distribution with mean zero and variance of ω^2_{CL} and ω^2_V. The θs in the equations denote model parameters as typically used in statistical models. Age and weight are standardized by dividing by their population median.

In the final covariate model for the entire cohort dataset with last-dose times, the estimates of typical values of CL and V were 24.9 L/hr and 6310 L for a 70 kg, 46-year-old

subject with CYP3A5*3/*3. Subjects with *1/*3 and *1/*1 increased clearance 1.77 (i.e., $e^{0.57}$=1.77) and 2.25 times (i.e., $e^{0.81}$=2.25) compared to clearance for those with *3*3, respectively. The CV for CL and V were 18% and 369%, respectively, reduced from the base model.

PK parameters estimated in the without last-dose time dataset do not fall outside of the Wald confidence intervals constructed about the relevant parameters estimated from the dataset with last-dose times. This is true in both the entire cohort and the reduced cohort selected for greater prevalence of last-dose times.

Model diagnostics

Figure 2 shows model diagnostics in the model built from the entire cohort dataset with last-dose times. Overall, the goodness-of-fit plots present reasonable model fit although some deviation from normality was noticed in the observed vs. predicted concentrations plots. The weighted residuals also demonstrate slight deviation from normality, but 95% of the residuals fall within the range of 2 to -2. The visual predictive check shows that predicted median and lower quantile sometimes overestimate true values across follow-up time.

Sensitivity Analysis of k_a in Population PK Model

Table 3 shows parameter estimates from the entire cohort dataset with last-dose times when k_a is either assumed to be 4.5, 3.09, or 0.375. The estimates (± S.E.) of typical values for Cl were 24.9 ± 0.69, 24.1 ± 0.70, and 25.5 ± 0.70, and those for V were 6310 ± 433, 6130 ± 722, and 6040 ± 719, for k_a being 4.5, 3.09, and 0.375 models, respectively. None of the estimates for k_a = 3.09 and k_a = 0.375 models fall outside of the 95% Wald confidence intervals for the corresponding parameters of the k_a = 4.5 model.

Discussion

Our tacrolimus population PK study performed with real world data using solely EHR data reproduced the well-established relationship between CYP3A SNP and some of the PK characteristics of tacrolimus. Our base model estimate of CL of 30.8 L/h is large compared to published estimates (with 95% CI) of 22.1 [19.3, 24.0][11] and 22.7 [21.0, 24.4].[12] Our base model estimate of V of 6550 L is also large compared to base published estimates of 653 [418, 888][11] and 1090 [911.8, 1268.2].[12] However, the significant association of CYP3A5 SNP published previously is replicated. Hematocrit did not affect model performance enough to be included in our model yet was found to have a significant effect by Zuo *et al.*[12] Future work will investigate the sensitivity of our estimates to model specification to attempt to account for these discrepancies.

The difference in last-dosing time did not meaningfully alter the estimation of PK parameters or covariate effects in the model in either the entire cohort dataset or the reduced dataset which oversampled subjects for which more last-dosing time information was available. The lack of a difference can be attributed to the following reasons. First, the estimated timing of the last dose may be close enough to the actual time, which would be a reasonable assumption for many patients as the compliance of organ transplant patient population is known to be high because of the serious consequence of organ rejection if they are not compliant to taking medication at scheduled intervals, or missing doses. Second, tacrolimus concentration levels are

likely in the steady-state after 1-month, from which our data collection was started. In addition, as tacrolimus has a long half-life, trough concentrations would well approximate the average steady-state concentration,[24] which would allow for the PK modeling to be less sensitive to deviation of true dosing time. Our findings support the general modeling approach used in several tacrolimus population PK studies,[11,12] where k_a was fixed at a published value. Our study results will be useful for PK studies performing using observational data such as EHRs for many medications having similar PK characteristics, such as a long half-life.[20]

We are interested in whether our estimates for clearance and covariate effects are sensitive to our selection for k_a. The choice of k_a in a range of 12-fold difference (from 0.375 to 4.5) had little impact on parameter or covariate estimates in the entire dataset; all parameters in the $k_a = 3.09$ or 0.375 models were within the 95% confidence interval of the corresponding parameters in the $k_a = 4.5$ model. This is likely due to the use of trough concentrations, which are taken well after the absorption phase and are therefore less impacted by the absorption process. Given the twice-daily dosing schedule and clinical goals of tacrolimus blood concentration maintenance via TDM, any reasonable selection for k_a is not likely to impact PK modeling. These k_a values were reported for immediate-release formulations for tacrolimus, which is the formulation used in our study. As the extended-release formulation for tacrolimus has been on the market for a while and will likely be increasingly used in the future, our findings would be also useful for tacrolimus population PK studies performed with the extended-release formulations.

Future work will be to replicate these results using our complete medication information pipeline and compare it to the results obtained through clinically validated data. Reproduction of PK parameters and covariate effects will indicate that our system is viable to replace costly manual information extraction to build PK data.

Funding

LC is supported by NIH/NIGMS (R01-GM124109).

Conflict of Interest / Disclosure

The authors have no conflicts of interest to disclose.

Author Contributions

All authors participated in critical review and revision of the final manuscript and approved the final manuscript draft.
MLW: Wrote Manuscript, Performed Research, and Analyzed Data.
HLW: Performed Research and Contributed Analytical Tools.
CB: Contributed Analytical Tools.
LC: Wrote Manuscript, Designed Research, Performed Research, and Analyzed Data.

References

1. Ringe, B., Braun, F., Lorf, T., Canelo, R., Schütz, E., Sattler, B., & Ramadori, G. (1998). Tacrolimus and Mycophenolate Mofetil in Clinical Liver Transplantation: Experience With a Steroid-Sparing Concept. *Transplantation Proceedings*, *30*(4), 1415–1416. https://doi.org/10.1016/S0041-1345(98)00296-6
2. Mor, E., Yussim, A., Chodoff, L., & Schwartz, M. E. (1997). New Immunosuppressive Agents for Maintenance Therapy in Organ Transplantation. *BioDrugs*, *8*(6), 469–488. https://doi.org/10.2165/00063030-199708060-00007
3. Staatz, C. E., & Tett, S. E. (2004). Clinical Pharmacokinetics and Pharmacodynamics of Tacrolimus in Solid Organ Transplantation. *Clinical Pharmacokinetics*, *43*(10), 623–653. https://doi.org/10.2165/00003088-200443100-00001
4. Busuttil, R. W., Klintmalm, G. B. G., Lake, J. R., Miller, C. M., & Porayko, M. (1996). GENERAL GUIDELINES FOR THE USE OF TACROLIMUS IN ADULT LIVER TRANSPLANT PATIENTS. [Letter]. *Transplantation*, *61*(5), 845–847.
5. Fung, J. J., & Starzl, T. E. (1995). FK506 in Solid Organ Transplantation. *Therapeutic Drug Monitoring*, *17*(6), 592–595.
6. Hu, R.-H., Lee, P.-H., & Tsai, M.-K. (2000). Clinical influencing factors for daily dose, trough level, and relative clearance of tacrolimus in renal transplant recipients. *Transplantation Proceedings*, *32*(7), 1689–1692. https://doi.org/10.1016/S0041-1345(00)01413-5
7. Jain, A. B., Fung, J. J., Tzakis, A. G., Venkataramanan, R., Abu-Elmagd, K., Alessiani, M., Reyes, J., Irish, W., Warty, V., Mehta, S., Todo, S., & Starzl, T. E. (1991). Comparative Study of Cyclosporine and FK 506 Dosage Requirements in Adult and Pediatric Orthotopic Liver Transplant Patients. *Transplantation Proceedings*, *23*(6), 2763–2766.
8. Birdwell, K. A., Grady, B., Choi, L., Xu, H., Bian, A., Denny, J. C., Jiang, M., Vranic, G., Basford, M., Cowan, J. D., Richardson, D. M., Robinson, M. P., Ikizler, T. A., Ritchie, M. D., Stein, C. M., & Haas, D. W. (2012). The use of a DNA biobank linked to electronic medical records to characterize pharmacogenomic predictors of tacrolimus dose requirement in kidney transplant recipients: *Pharmacogenetics and Genomics*, *22*(1), 32–42. https://doi.org/10.1097/FPC.0b013e32834e1641
9. Fukudo, M., Yano, I., Yoshimura, A., Masuda, S., Uesugi, M., Hosohata, K., Katsura, T., Ogura, Y., Oike, F., Takada, Y., Uemoto, S., & Inui, K. (2008). Impact of MDR1 and CYP3A5 on the oral clearance of tacrolimus and tacrolimus-related renal dysfunction in adult living-donor liver transplant patients: *Pharmacogenetics and Genomics*, *18*(5), 413–423. https://doi.org/10.1097/FPC.0b013e3282f9ac01
10. Jacobo-Cabral, C. O., García-Roca, P., Romero-Tejeda, E. M., Reyes, H., Medeiros, M., Castañeda-Hernández, G., & Trocóniz, I. F. (2015). Population pharmacokinetic analysis of tacrolimus in Mexican paediatric renal transplant patients: Role of CYP3A5 genotype and formulation: Role of CYP3A5 genotype and formulation type on tacrolimus PK. *British Journal of Clinical Pharmacology*, *80*(4), 630–641. https://doi.org/10.1111/bcp.12649
11. Li, D., Lu, W., Zhu, J.-Y., Gao, J., Lou, Y.-Q., & Zhang, G.-L. (2007). Population pharmacokinetics of tacrolimus and CYP3A5, MDR1 and IL-10 polymorphisms in adult liver transplant patients: Tacrolimus population pharmacokinetics and CYP3A5. *Journal of Clinical Pharmacy and Therapeutics*, *32*(5), 505–515. https://doi.org/10.1111/j.1365-2710.2007.00850.x
12. Zuo, X., Ng, C. M., Barrett, J. S., Luo, A., Zhang, B., Deng, C., Xi, L., Cheng, K., Ming, Y.,

Yang, G., Pei, Q., Zhu, L., Yuan, H., Liao, H., Ding, J., Wu, D., Zhou, Y., Jing, N., & Huang, Z. (2013). Effects of CYP3A4 and CYP3A5 polymorphisms on tacrolimus pharmacokinetics in Chinese adult renal transplant recipients: A population pharmacokinetic analysis. *Pharmacogenetics and Genomics*, *23*(5), 251–261. https://doi.org/10.1097/FPC.0b013e32835fcbb6

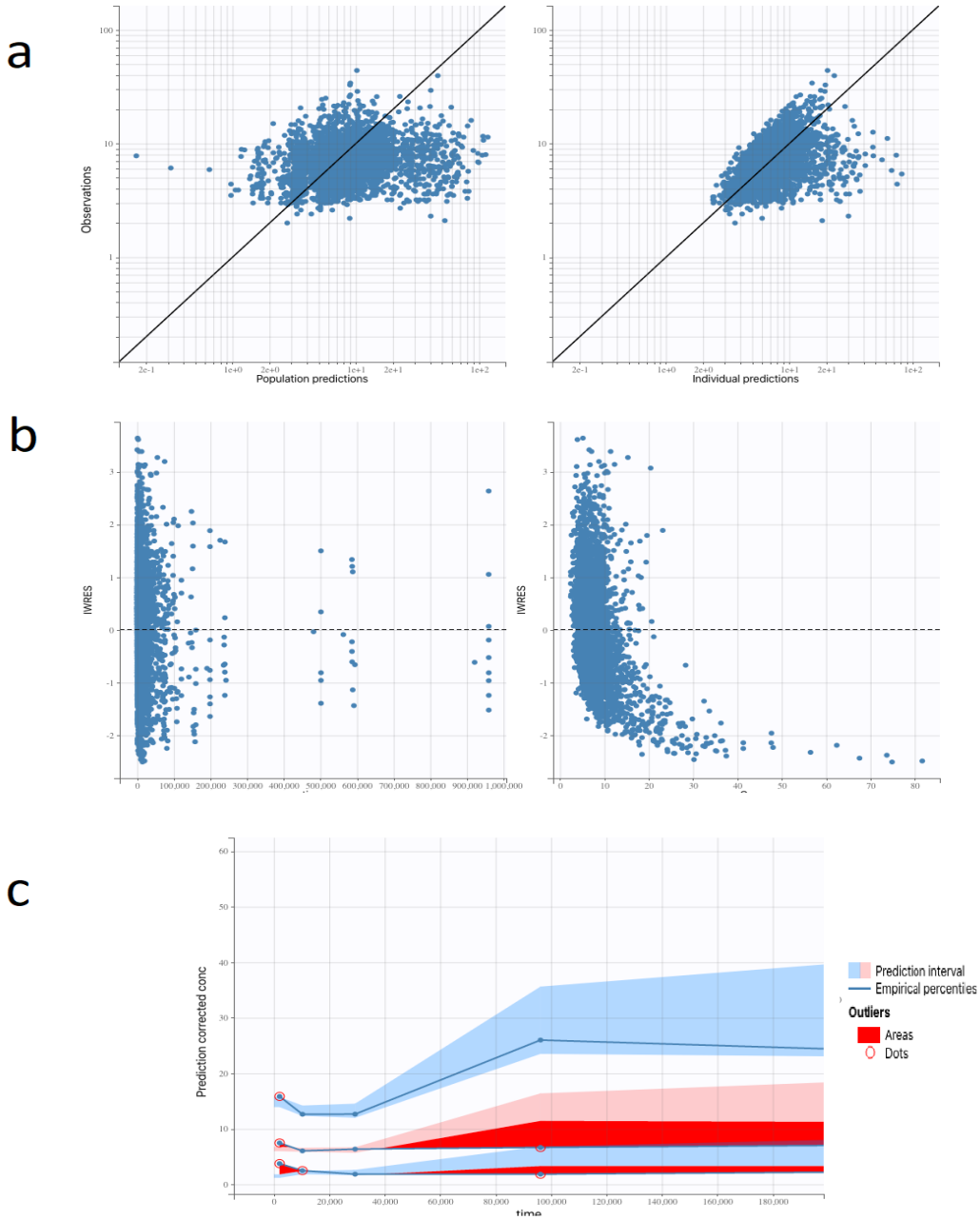
13. Chen, J. S., Li, L. S., Cheng, D. R., Ji, S. M., Sun, Q. Q., Cheng, Z., Wen, J. Q., Sha, G. Z., & Liu, Z. H. (2009). Effect of CYP3A5 Genotype on Renal Allograft Recipients Treated With Tacrolimus. *Transplantation Proceedings*, *41*(5), 1557–1561. https://doi.org/10.1016/j.transproceed.2009.01.097
14. Sheiner, L. B., & Ludden, T. M. (1992). Population Pharmacokinetics/Dynamics. *Annu. Rev. Pharmacol. Toxicol.*, *32*, 185–209.
15. Brunet, M., van Gelder, T., Åsberg, A., Haufroid, V., Hesselink, D. A., Langman, L., Lemaitre, F., Marquet, P., Seger, C., Shipkova, M., Vinks, A., Wallemacq, P., Wieland, E., Woillard, J. B., Barten, M. J., Budde, K., Colom, H., Dieterlen, M.-T., Elens, L., … Bergan, S. (2019). Therapeutic Drug Monitoring of Tacrolimus-Personalized Therapy: Second Consensus Report. *Therapeutic Drug Monitoring*, *41*(3), 261–307. https://doi.org/10.1097/FTD.0000000000000640
16. Gervasini, G., Garcia, M., Macias, R. M., Cubero, J. J., Caravaca, F., & Benitez, J. (2012). Impact of genetic polymorphisms on tacrolimus pharmacokinetics and the clinical outcome of renal transplantation. *Transplant International*, *25*(4), 471–480. https://doi.org/10.1111/j.1432-2277.2012.01446.x
17. Oteo, I., Lukas, J. C., Leal, N., Suarez, E., Valdivieso, A., Gastaca, M., Ortiz de Urbina, J., & Calvo, R. (2013). Tacrolimus pharmacokinetics in the early post-liver transplantation period and clinical applicability via Bayesian prediction. *European Journal of Clinical Pharmacology*, *69*(1), 65–74. https://doi.org/10.1007/s00228-012-1300-z
18. Vadcharavivad, S., Praisuwan, S., Techawathanawanna, N., Treyaprasert, W., & Avihingsanon, Y. (2016). Population pharmacokinetics of tacrolimus in Thai kidney transplant patients: Comparison with similar data from other populations. *Journal of Clinical Pharmacy and Therapeutics*, *41*(3), 310–328. https://doi.org/10.1111/jcpt.12396
19. Weeks, H. L., Beck, C., McNeer, E., Williams, M. L., Bejan, C. A., Denny, J. C., & Choi, L. (2020). medExtractR: A targeted, customizable approach to medication extraction from electronic health records. *Journal of the American Medical Informatics Association*, *27*(3), 407–418. https://doi.org/10.1093/jamia/ocz207
20. Choi, L., Beck, C., McNeer, E., Weeks, H. L., Williams, M. L., James, N. T., Niu, X., Abou-Khalil, B. W., Birdwell, K. A., Roden, D. M., Stein, C. M., Bejan, C. A., Denny, J. C., & Driest, S. L. V. (2020). Development of a System for Postmarketing Population Pharmacokinetic and Pharmacodynamic Studies Using Real-World Data From Electronic Health Records. *Clinical Pharmacology & Therapeutics*, *107*(4), 934–943. https://doi.org/10.1002/cpt.1787
21. McNeer, E., Beck, C., Weeks, H. L., Williams, M. L., James, N. T., & Choi, L. (2019). *A post-processing algorithm for building longitudinal medication dose data from extracted medication information using natural language processing from electronic health records* [Preprint]. bioRxiv. https://doi.org/10.1101/775015
22. Jusko, W. J., Piekoszewski, W., Klintmalm, G. B., Shaefer, M. S., Hebert, M. F., Piergies, A. A., Lee, C. C., Schechter, P., & Mekki, Q. A. (1995). Pharmacokinetics of tacrolimus in liver transplant patients*. *Clinical Pharmacology & Therapeutics*, *57*(3), 281–290.

https://doi.org/10.1016/0009-9236(95)90153-1
23. Lu, Z., Bonate, P., & Keirns, J. (2019). Population pharmacokinetics of immediate- and prolonged-release tacrolimus formulations in liver, kidney and heart transplant recipients. *British Journal of Clinical Pharmacology*, *85*(8), 1692–1703. https://doi.org/10.1111/bcp.13952
24. Passey, C., Birnbaum, A. K., Brundage, R. C., Oetting, W. S., Israni, A. K., & Jacobson, P. A. (2011). Dosing equation for tacrolimus using genetic variants and clinical factors: Tacrolimus dosing equation. *British Journal of Clinical Pharmacology*, *72*(6), 948–957. https://doi.org/10.1111/j.1365-2125.2011.04039.x

Figure 1. Last-Dose Time Extraction. Extraction and processing examples for time at which a patient took their last medication dose. In step 2, PM times are assumed to occur the day prior to the note date. AM times are assumed to occur on the same day at the note date. Final dataset includes one last dose time for each ID-date pair.

Step	ID 101, 2000-01-01, Note 1	ID 101, 2000-01-01, Note 2	ID 201, 2001-06-30, Note 1
0. Clinical note excerpt	*…on tacrolimus 1.5mg bid – last dose at 2130…*	*…Pt taking Prograf 3mg once a day at 7am. pt took last dose of Prograf at 9:30 last night instead …*	*…Pt took last dose of tacro at 6:30pm instead of 8 pm…*
1. Extract last dose time with *medExtractR*	**DrugName**: tacrolimus **LastDose**: 2130	**DrugName**: Prograf **LastDose**: 7am **DrugName**: Prograf **LastDose**: 9:30 last night	**DrugName**: tacro **LastDose**: 6:30pm **LastDose**: 8 pm
2. Process into standardized format	"2130" → 1999-12-31 21:30:00	"7am" → 2000-01-01 07:00:00 "9:30 last night" → 1999-12-31 21:30:00	"6:30pm" → 2000-06-29 18:30:00 "8 pm" → 2000-06-29 20:00:00
3. Resolve time conflicts	Unique times: 1999-12-31 21:30:00, 2000-01-01 07:00:00 - Time difference > 2 hours requires manual review - Note review indicates that 1999-12-31 21:30:00 is correct (2000-01-01 07:00:00 is the usual dose time)		Unique times: 2000-06-29 18:30:00 2000-06-29 20:00:00 - Time difference < 2 hours - Keep earlier time (no manual review)
4. Final formatted time	**ID**: 101, Date: 2000-01-01 **Last Dose Time**: 1999-12-31 21:30:00		**ID**: 201, Date: 2001-06-30 **Last Dose Time**: 2001-06-29 20:30:00

Figure 2: Model Diagnostics. Model diagnostic plots for the model fit to the entire cohort dataset with last-dose times. Observed vs. predicted concentration plots are shown on the log scale for both population (left) and individual (right) level predictions. Individual weighted residuals (b) are plotted against time (left) and predicted concentration (right). The corrected visual predictive check (c) is shown up to the follow-up time encompassing 99% of observed concentrations.



	Entire Cohort (N=440)	Reduced Cohort (N=257)
Age (Year)	45.6 (13.3)	45.8 (12.5)
	46 [36, 56]	47 [37, 55]
Sex = Female	170 (0.39)	87 (0.34)
Race		
White	322 (0.73)	195 (0.76)
Black	101 (0.23)	53 (0.21)
Hispanic	4 (0.01)	3 (0.01)
Asian	5 (0.01)	3 (0.01)
Other	3 (0.01)	3 (0.01)
Undisclosed	5 (0.01)	0 (0.00)
Weight (kg)	79.7 (19.7)	79.1 (20.2)
	77.7 [64.4, 91.6]	75.8 [63.5, 90.7]
Hemoglobin (g/dl)	12.5 (2.1)	12.6 (2.1)
	12.6 [11.1, 14.0]	12.6 [11.1, 14.1]
Albumin (g/dl)	4.1 (0.5)	4.1 (0.1)
	3.8 [4.1, 4.4]	4.1 [3.9, 4.4]
Proportion of concentrations with last-dose time	0.48	0.70
Daily tacrolimus dose (mg/day)	3.3 (1.8)	3.2 (1.7)
	3 [2, 4]	3 [2, 4]
Tacrolimus blood concentration (mcg/L)	7.5 (3.2)	7.4 (3.1)
	7.0 [5.3, 8.9]	7.0 [5.4, 8.8]
Number of concentration levels per subject	9.3 (1.6)	9.7 (1.0)
	10 [10, 10]	10 [10, 10]

Table 1. Demographic and Clinical Characteristics. Population characteristics of the entire dataset (left) and the reduced dataset (right). Values are presented as count (proportion) for categorical variables and mean (S.E.) median[interquartile range] for continuous variable.

Base Model		Covariate Models				
Entire Cohort (N=440)			Entire Cohort (N=440)		Reduced Cohort (N=257)	
With Last-Dose Time			With Last-Dose Time	Without Last-Dose Time	With Last-Dose Time	Without Last-Dose Time
Parameter	Estimate (SE)	Parameter	Estimate (SE)			
		$CL_i=\theta_1 \times (age_i/46)^{\theta 2} \times \exp[\theta_3 \times CYP3A5*3*1_i] \times \exp[\theta_4 \times CYP3A5*1*1_i]$				
CL	30.8 (0.85) [29.1, 32.5]	θ_1	24.9 (0.69) [23.5, 26.3]	25.1 (0.69) [23.7, 26.5]	23.5 (0.81) [21.9, 25.1]	23.9 (0.82) [22.3, 25.5]
		θ_2	-0.22 (0.07) [-0.35, -0.09]	-0.22 (0.07) [-0.35, -0.09]	-0.25 (0.09) [-0.42, -0.08]	-0.25 (0.09) [-0.42, -0.08]
		θ_3	0.57 (0.06) [0.46, 0.68]	0.58 (0.05) [0.47, 0.69]	0.54 (0.07) [0.41, 0.67]	0.54 (0.07) [0.41, 0.67]
		θ_4	0.81 (0.07) [0.67, 0.95]	0.81 (0.07) [0.68, 0.94]	0.93 (0.10) [0.74, 1.12]	0.93 (0.10) [0.74, 1.12]
		$V_i= \theta_5 \times (wt_i/70)^{\theta 6}$				
V	6550 (721) [5137, 7963]	θ_5	6310 (433) [5461, 7159]	6390 (737) [4945, 7835]	3980 (524) [2953, 5007]	4400 (568) [3287, 5513]
		θ_6	0.709 (0.433) [-1.55, 0.13]	0.81 (0.43) [-0.03, 1.65]	-0.20 (0.48) [-1.14, 0.74]	-0.27 (0.47) [-1.19, 0.65]
ω_{CL} (%CV)	0.27 (0.02) [0.23, 0.31]	ω_{CL} (%CV)	0.18 (0.01) [0.15, 0.21]	0.18 (0.01) [0.15, 0.21]	0.16 (0.02) [0.12, 0.19]	0.17 (0.02) [0.14, 0.20]
ω_V (%CV)	3.80 (0.29) [3.23, 4.37]	ω_V (%CV)	3.69 (0.30) [3.10, 4.28]	3.61 (0.30) [3.02, 4.20]	2.69 (0.28) [2.14, 3.24]	2.59 (0.28) [2.04, 3.14]
$\sigma^2_{additive}$ (ng/ml)	2.22e-16 (0.11) [-0.22, 0.22]	$\sigma^2_{additive}$ (ng/ml)	2.22e-16 (0.11) [-0.25, 0.25]	2.22e-16 (0.11) [-0.22, 0.22]	2.22e-16 (0.13) [-0.25, 0.25]	2.22e-16 (0.12) [-0.24, 0.24]
$\sigma^2_{proportional}$ (%CV)	0.38 (0.01) [0.35, 0.41]	$\sigma^2_{proportional}$ (%CV)	0.38 (0.01) [0.35, 0.41]	0.38 (0.01) [-0.35, 0.41]	0.36 (0.01) [0.33, 0.39]	0.36 (0.01) [0.33, 0.39]

Table 2. Estimated PK Parameters. PK parameter estimates for both the entire and reduced datasets either with or without last-dose time. Estimates are presented as mean (S.E.) [Wald 95% confidence interval]. CYP3A5*3/*1_i and CYP3A5*1/*1_i are indicators for genotype for *3/*1 and *1/*1, respectively, and wt_i represent weight for subject *i*.

	k_a = 4.5	k_a = 3.09	k_a = 0.375
Parameter		**Estimate (SE) [95% CI]**	
$CL_i=\theta_1 \times (age_i/46)^{\theta_2} \times \exp[\theta_3 \times CYP3A5*3*1_i] \times \exp[\theta_4 \times CYP3A5*1*1_i]$			
θ_1	24.9 (0.69)	24.1 (0.70)	25.5 (0.70)
	[23.5, 26.3]	[22.7, 25.5]	[24.1, 26.9]
θ_2	-0.22 (0.07)	-0.26 (0.07)	0.21 (0.07)
	[-0.35, -0.09]	[-0.39, -0.13]	[0.08, 0.34]
θ_3	0.57 (0.06)	0.57 (0.06)	0.58 (0.06)
	[0.46, 0.68]	[0.46, 0.68]	[0.47, 0.69]
θ_4	0.81 (0.069)	0.79 (0.070)	0.79 (0.069)
	[0.67, 0.95]	[0.65, 0.93]	[0.65, 0.93]
$V_i= \theta_5 \times (wt_i/70)^{\theta_6}$			
θ_5	6310 (433)	6130 (722)	6040 (719)
	[5461, 7159]	[4715, 7545]	[4631, 7449]
θ_6	0.709 (0.433)	0.85 (0.44)	0.70 (0.45)
	[-1.55, 0.13]	[-0.01, 1.71]	[-0.18, 1.58]
ω_{CL} (%CV)	0.18 (0.01)	0.18 (0.01)	0.18 (0.01)
	[0.15, 0.21]	[0.15, 0.21]	[0.15, 0.21]
ω_V (%CV)	3.69 (0.30)	3.80 (0.30)	3.92 (0.31)
	[3.10, 4.28]	[3.21, 4.39]	[3.31, 4.53]
$\sigma^2_{additive}$ (ng/ml)	2.22e-16 (0.11)	2.22e-16 (0.10)	2.22e-16 (0.11)
	[-0.25, 0.25]	[-0.19, 0.19]	[-0.22, 0.22]
$\sigma^2_{proportional}$ (%CV)	0.38 (0.01)	0.38 (0.01)	0.38 (0.01)
	[0.35, 0.41]	[-.36, 0.40]	[0.35, 0.41]

Table 3. Sensitivity Analysis. PK parameter estimates for models assuming k_a to be 4.5, 3.09, 0.375 (left to right). Estimates are presented as mean (S.E.) [Wald 95% confidence interval]. $CYP3A5*3/*1_i$ and $CYP3A5*1/*1_i$ are indicators for genotype for *3/*1 and *1/*1, respectively, and wt_i represent weight for subject *i*.