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 363 subjects was performed using real-world data extracted from electronic heath records (EHRs) to identify 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 base model estimates of population level PK parameters were Cl = 40.0 L/h [36.6, 43.4] and V = 4697 L [3478, 5916]. 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 medication widely used to prevent rejection following organ transplantation. Tacrolimus has a narrow therapeutic window, which makes it challenging for drug concentration levels to be appropriately maintained. Changes in dose, concomitant medications, or co-morbidities can yield less efficacy (i.e., an increased probability of allograft rejection)[1] or more toxicity (e.g., nephrotoxicity, neurotoxicity, diabetogenicity).[2,3] The therapeutic window also varies depending on time since 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, time since transplantation,[5] and age.[6] In particular, the effect of a single nucleotide polymorphism (SNP), rs776746, on the CYP3A5 gene coding the major metabolizing enzyme for tacrolimus has been reported from multiple studies, including our own.[7] Individuals carrying the CYP3A5*3, a loss-of-function allele, have reduced clearance and require significantly smaller tacrolimus doses to maintain the same tacrolimus concentration level.[8–12] This SNP alone was estimated to be associated with 39% of the variability in dose requirement between subjects.[7]

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 participant enrollment.[13] 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, leading to increased interest in using EHRs as real-word data.

Due to the narrow therapeutic window with high variability in tacrolimus disposition, tacrolimus is one of the drugs that require therapeutic drug monitoring (TDM).[14] Tacrolimus blood concentrations are routinely checked as part of standard of care, making retrospective studies with EHR data more appealing with automatically deposited drug level data. Several tacrolimus population PK studies have been conducted using such data.[15–17] 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 investigate this assumption by comparing models fit at varying assumed k_a levels.

We recently developed a natural language processing (NLP) system, *medExtractR*,[18] 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.[19,20] The system, once finalized, should relieve the primary burden in data generation and manual extraction of medication data. In addition to drug dosing information, *medExtractR* is designed to extract explicit last dosing times (timing of the dose prior to a recorded blood concentration) if present in the notes. Tacrolimus population PK studies often assume trough concentrations measured before taking a morning dose that is 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 from assumed 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.*[7] 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.

Extraction of Last-Dose Times

Times of last dose were extracted using *medExtractR* (see Weeks *et al.* 2020 for details).[18] 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. We observed a small percentage of conflicts (6.04%), about half of which (3.89%) were within 2 hours. 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 (1.90%) 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 (0.90%).

Data Process

From the set of subjects (N = 399) described previously by Birdwell *et al.*,[7] we refined our study cohort, called "entire cohort" (N = 363), as follows. We excluded all drug concentration data measured beyond three years post-surgery and then excluded subjects with fewer than 4 remaining concentrations. A subject with no known genotype for rs776746 (N=1) was also excluded. We used the original data (which do not have NLP-extracted last-dose times) for the entire cohort, defined above. Last-dose times were extracted from the clinical notes for this 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; it included subjects having at least 4 concentration measurements with extracted last-dose times. Otherwise,

the PK datasets for both cohorts were built with the same algorithm implemented as a function in an R package, *EHR*. A brief description of the PK data building method with and without last-dose time is as follows.

PK Data Building

When last-dose time is not available within the clinical notes, 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 the drug level checking. We further assume that the drug is taken every 12 hours following that initial dose, as all subjects took immediate-release formulation and no extended-release formulation was prescribed. As almost all drug measurements are being taken in the morning to provide trough concentration levels, 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 each concentration measurement determines the dose timing pattern until the next measurement is taken.

When last-dose time is available, the extracted time is used. 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 have an associated last-dose time available in the clinical notes. The datasets with and without last-dose time will be identical for the subset of concentration measurements missing last-dose time.

Population PK Analysis

We performed population PK analysis of tacrolimus using a nonlinear mixed-effects model implemented by NONMEM®[21] version VII with the first order conditional estimation method with interaction. 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] A correlation structure was placed on the estimates for clearance and volume.

The dataset from the entire cohort with the last-dose time was considered as the primary dataset for developing a population PK model. 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 values (-2 log likelihood), and their difference between the models along with the number of parameters, which would approximately follow χ^2 distribution. The χ^2 statistics of 3.84 with 1 degree of freedom correspond to a p value of 0.05. Thus, we considered the objective function value decrease of 3.84 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 at least 3.84.

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.

Sensitivity Analysis

The k_a in our final models was assumed to be 4.5 as previously reported.[22] In order to assess whether our findings are sensitive to this selected value, we refit the model using the entire cohort at another published k_a value of 3.09.[11] 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 population characteristics for the two cohorts are described in **Table 1**. Each subject in the cohort has a maximum of 10 tacrolimus blood concentration measurements. These observations constitute either the first 10 concentrations or every concentration if there are fewer than 10 measurements for each subject 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 7.5 ng/mL. The reduced cohort consists of 223 subjects. The percentage of concentration measurements associated with a last-dose time is increased to 73% 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 subjects; the between-subject variation in coefficient of variation (%CV) for CL and V in the base model are 60.9% and 72.0%, respectively.

Inclusion of the SNP rs776746 alone decreased the objective function by 108 from 11768 for the base model; hence the SNP was always included in the following covariate model selection considering strong evidence for its causal effect on clearance in the literature as well as our own results. Inclusion of age, albumin level, and hemoglobin level all improved the model fit from the SNP only model – decreasing the objective function value by 11, 4, and 16, respectively. The final covariate model with these 3 additional covariates improved the model fit by 36 from the SNP only model. Other covariates did not yield significantly improved model fit. The final model is presented as follows:

$$CL_{ij}=\theta_1 \times (SNP_i)^{\theta_2} \times (age_{ij}/47)^{\theta_3} \times (alb_{ij}/4.1)^{\theta_4} \times (hgb_{ij}/12.5)^{\theta_5} \times \exp[\eta_i^{CL}],$$

and

$$V_{ij} = \theta_6 \times \exp[\eta_i^V],$$

where CL_{ij} and V_{ij} are the subject-specific CL and V for subject *i* at time *j*. The age_{ij} is subject age in years, alb_{ij} is blood albumin level in g/dL, and hgb_{ij} is blood hemoglobin level in g/dL for subject *i* at time *j*, each of which is standardized by dividing by its population median. The genotype for SNP, rs776746, is continuously coded as 1, 2, 3 corresponding to the number of CYP3A5*1 alleles of 0, 1, and 2. The η_i^{CL} and η_i^V are random effects explaining between-subject variability for CL and V, which follow a bivariate normal distribution with mean zeros and covariance, where $\omega_{CL}/100$ and $\omega_V/100$ are standard deviations corresponding to CL and V, respectively. The θs in the equations denote model parameters as typically used in statistical models.

In the final covariate model for the entire cohort dataset with last-dose times, the estimates of typical values of CL and V were 31.9 L/hr and 3735 L for a 47-year-old with no CYP3A5*1 allele, albumin 4.1 g/dL, and hemoglobin 12.5 g/dL. Subjects with *1/*3 and *1/*1 increased clearance 1.6 (i.e., $2^{0.70}$=1.6) and 2.2 times (i.e., $3^{0.70}$=2.2) compared to clearance for those with *3*3, respectively. The between-subject variation in %CV for CL and V were 49.4% and 79.7%, respectively. %CV for CL was reduced from the base model due to the inclusion of covariates on clearance.

PK parameters estimated in the without last-dose time dataset do not fall outside of the 95% Wald confidence intervals (CI) 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 for the final 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 Tacrolimus 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 (± SE) of typical values for Cl at median covariate values and no CYP3A5*1 allele were 31.9 ± 1.4, 32.0 ± 1.4, and 32.4 ± 1.4, and those for V were 3735 ± 637, 3752 ± 635, and 3723 ± 623 for k_a assumed to be 4.5, 3.09, and 0.375, respectively. None of the estimates for k_a = 3.09 and k_a = 0.375 models fall outside of the 95% Cis 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 the effect of rs776746 on CYP3A5 and tacrolimus clearance. Our base model CL estimate was 40.0 L/h [36.6, 43.4] compared to

published estimates of 22.1 [19.3, 24.0][10] and 22.7 [21.0, 24.4].[11] Although the typical value of CL is not completely comparable due to different covariate models used across studies, our final covariate model estimated CL of 31.9 [29.2-34.7] while the other studies reported 15.9 [13.2 – 18.6][10] and 26.6 [18–35.2].[11] Our estimates are larger than those published values. This may be due to the difference in the study populations, especially different post-transplantation days. Our cohort included data from 30 days of post- transplantation up to 1,095 days (3 years) with median of 482 days, while the median (range) post-transplantation days were 14 (1 – 175) and 9 (0–95) from these two studies, which are much shorter than ours. In addition, the study population for Li *et al.* was liver transplant subjects, which would have additional impact on clearance as poor liver function reduces tacrolimus clearance.[24,25]

The differential frequency distribution of CYP3A5*1 genotype between populations of African ancestry (AA) and Northern European ancestry (NE) is well established and is apparent in our study population (**Figure 3**). We observed greater CYP3A5*1 genotype prevalence in the AA population compared to the NE population. Thus, clearance for AA subjects is much higher than for NE subjects if CYP3A5*1 genotype is not adjusted. However, once genotype is accounted for, the difference between PK parameters in NE and AA subjects is negligible; inclusion of a covariate for race did not sufficiently impact the objective function to be included in our final covariate model. AA patients are known to require higher doses of tacrolimus, but it is unknown to what extent this is due to greater prevalence of CYP3A5*1 rather than differences in bioavailability or absorption rates.[26] Our results support that this difference is more attributable to the differential distribution of CYP3A5*1 genotype than any other factors.

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. There is more similarity within cohorts regardless of the last-dose time group status than between the entire and the reduced cohort given the last-dose time group. 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, this is a reasonable assumption as the compliance of organ transplant subject 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. Second, tacrolimus concentration levels are likely in the steady-state after 1 month of organ transplantation, the time 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,[27] which would allow for the PK modeling to be less sensitive to deviation of true dosing time. V is, however, consistently estimated to be larger in the without last-dose time group; perhaps this can be attributed to the increased uncertainty due to incorrect last-dose times. These results can inform the design of PK studies performed using observational data such as EHRs for medications having similar PK characteristics.[19]

Our findings support the general modeling approach used in several tacrolimus population PK studies,[10,11] where k_a was fixed at a published value – our estimates for clearance and covariate effects were not sensitive to our selection for k_a. The choice of k_a in a range of a 12-fold difference (from 0.375 to 4.5) had little impact on parameter estimates in the entire dataset; all parameter estimates from the $k_a = 3.09$ or 0.375 models were within the 95% CIs 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 tacrolimus PK modeling. These k_a values were reported for immediate-release formulations for tacrolimus, which is the formulation used in our study. Notably, the models with k_a set to 4.5 and 3.09 are nearly identical while the model with k_a set to 0.375 reveals slight, negligible differences. As the extended-release formulation for tacrolimus is now on the market and will likely be increasingly used in the future, our findings would also be useful for tacrolimus population PK studies performed with the extended-release formulations.

Our study has some limitations. First, the last-dose time information was not available for all drug concentration level measurements. However, based on the results from the reduced datasets that included the last-dose time information for 73% of drug levels, it does not appear that including last-dose times in PK modeling will result in significant changes as long as medications share similar PK profile with tacrolimus, such as a long half-life. Second, we did not specifically study medications with different PK profiles to investigate the effects of the last-dose time and the fixed k_a on PK parameter estimates. Thorough investigation on these topics with diverse medications may be needed, potentially including a detailed simulation-based study. Future work will be also needed 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: Designed research, wrote manuscript, performed research, and analyzed data.

References

1. Ringe, B. *et al.* Tacrolimus and Mycophenolate Mofetil in Clinical Liver Transplantation: Experience With a Steroid-Sparing Concept. *Transplant. Proc.* **30**, 1415–1416 (1998).
2. Mor, E., Yussim, A., Chodoff, L. & Schwartz, M. E. New Immunosuppressive Agents for Maintenance Therapy in Organ Transplantation. *BioDrugs* **8**, 469–488 (1997).
3. Staatz, C. E. & Tett, S. E. Clinical Pharmacokinetics and Pharmacodynamics of Tacrolimus in Solid Organ Transplantation. *Clin. Pharmacokinet.* **43**, 623–653 (2004).
4. Busuttil, R. W., Klintmalm, G. B. G., Lake, J. R., Miller, C. M. & Porayko, M. General Guidelines for the Use of Tacrolimus in Adult Liver Transplant Patients. [Letter]. *Transplantation* **61**, 845–847 (1996).
5. Hu, R.-H., Lee, P.-H. & Tsai, M.-K. Clinical influencing factors for daily dose, trough level, and relative clearance of tacrolimus in renal transplant recipients. *Transplant. Proc.* **32**, 1689–1692 (2000).
6. Jain, A. B. *et al.* Comparative Study of Cyclosporine and FK 506 Dosage Requirements in Adult and Pediatric Orthotopic Liver Transplant Patients. *Transplant. Proc.* **23**, 2763–2766 (1991).
7. Birdwell, K. A. *et al.* The use of a DNA biobank linked to electronic medical records to characterize pharmacogenomic predictors of tacrolimus dose requirement in kidney transplant recipients: *Pharmacogenet. Genomics* **22**, 32–42 (2012).
8. Fukudo, M. *et al.* Impact of MDR1 and CYP3A5 on the oral clearance of tacrolimus and tacrolimus-related renal dysfunction in adult living-donor liver transplant patients: *Pharmacogenet. Genomics* **18**, 413–423 (2008).

9. Jacobo-Cabral, C. O. *et al.* Population pharmacokinetic analysis of tacrolimus in Mexican paediatric renal transplant patients: role of CYP3A5 genotype and formulation. *Br. J. Clin. Pharmacol.* **80**, 630–641 (2015).
10. Li, D. *et al.* Population pharmacokinetics of tacrolimus and CYP3A5, MDR1 and IL-10 polymorphisms in adult liver transplant patients: Tacrolimus population pharmacokinetics and CYP3A5. *J. Clin. Pharm. Ther.* **32**, 505–515 (2007).
11. Zuo, X. *et al.* Effects of CYP3A4 and CYP3A5 polymorphisms on tacrolimus pharmacokinetics in Chinese adult renal transplant recipients: a population pharmacokinetic analysis. *Pharmacogenet. Genomics* **23**, 251–261 (2013).
12. Chen, J. S. *et al.* Effect of CYP3A5 Genotype on Renal Allograft Recipients Treated With Tacrolimus. *Transplant. Proc.* **41**, 1557–1561 (2009).
13. Sheiner, L. B. & Ludden, T. M. Population Pharmacokinetics/Dynamics. *Annu Rev Pharmacol Toxicol* 185–209 (1992).
14. Brunet, M. *et al.* Therapeutic Drug Monitoring of Tacrolimus-Personalized Therapy: Second Consensus Report. *Ther. Drug Monit.* **41**, 261–307 (2019).
15. Gervasini, G. *et al.* Impact of genetic polymorphisms on tacrolimus pharmacokinetics and the clinical outcome of renal transplantation. *Transpl. Int.* **25**, 471–480 (2012).
16. Oteo, I. *et al.* Tacrolimus pharmacokinetics in the early post-liver transplantation period and clinical applicability via Bayesian prediction. *Eur. J. Clin. Pharmacol.* **69**, 65–74 (2013).
17. Vadcharavivad, S., Praisuwan, S., Techawathanawanna, N., Treyaprasert, W. & Avihingsanon, Y. Population pharmacokinetics of tacrolimus in Thai kidney transplant patients: comparison with similar data from other populations. *J. Clin. Pharm. Ther.* **41**, 310–328 (2016).

18. Weeks, H. L. *et al.* medExtractR: A targeted, customizable approach to medication extraction from electronic health records. *J. Am. Med. Inform. Assoc.* **27**, 407–418 (2020).
19. Choi, L. *et al.* Development of a System for Postmarketing Population Pharmacokinetic and Pharmacodynamic Studies Using Real-World Data From Electronic Health Records. *Clin. Pharmacol. Ther.* **107**, 934–943 (2020).
20. McNeer, E. *et al.* A post-processing algorithm for building longitudinal medication dose data from extracted medication information using natural language processing from electronic health records. *bioRxiv* (2019).doi:10.1101/775015
21. Beal, S. & Sheiner, L. B. NONMEM Users Guide - Part I. (1989).
22. Jusko, W. J. *et al.* Pharmacokinetics of tacrolimus in liver transplant patients. *Clin. Pharmacol. Ther.* **57**, 281–290 (1995).
23. Lu, Z., Bonate, P. & Keirns, J. Population pharmacokinetics of immediate- and prolonged-release tacrolimus formulations in liver, kidney and heart transplant recipients. *Br. J. Clin. Pharmacol.* **85**, 1692–1703 (2019).
24. Jain, A. B. *et al.* Effect of Hepatic Dysfunction and T Tube Clamping on FK 506 Pharmacokinetics and Trough Concentrations. *Transplant. Proc.* **22**, 57–59 (1990).
25. Bekersky, I., Dressler, D., Alak, A., Boswell, G. W. & Mekki, Q. A. Comparative Tacrolimus Pharmacokinetics: Normal versus Mildly Hepatically Impaired Subjects. *J. Clin. Pharmacol.* **41**, 628–635 (2001).
26. Uber, P. A., Mehra, M. R., Scott, R. L., Prasad, A. K. & Park, M. H. Ethnic disparities in the pharmacologic characteristics of tacrolimus in heart transplantation. *Transplant. Proc.* **33**, 1581–1582 (2001).

27. Passey, C. *et al.* Dosing equation for tacrolimus using genetic variants and clinical factors: Tacrolimus dosing equation. *Br. J. Clin. Pharmacol.* **72**, 948–957 (2011).

	Entire Cohort (N=363)	Reduced Cohort (N=223)
Age (Year)	45.9 (12.7) 47.0 [37.0, 55.0]	47.2 (12.0) 48.0 [38.0, 55.0]
Sex = Female	134 (0.37)	76 (0.34)
Race		
Caucasian American	266 (0.73)	164 (0.74)
African American	83 (0.23)	50 (0.22)
Hispanic	4 (0.01)	3 (0.01)
Asian	5 (0.01)	3 (0.01)
Other	3 (0.01)	3 (0.01)
Undisclosed	2 (0.01)	0 (0.00)
Weight (kg)	82.9 (20.8) 81.2 [66.7, 95.3]	83.0 (20.7) 80.0 [66.4, 94.9]
Hemoglobin (g/dl)	12.5 (2.1) 12.5 [11.1, 14.0]	12.6 (2.0) 12.5 [11.1, 14.0]
Albumin (g/dl)	4.1 (0.4) 4.1 [3.9, 4.4]	4.1 (0.4) 4.1 [3.8, 4.4]
Proportion of concentrations with last-dose time	0.48	0.73
Daily tacrolimus dose (mg/day)	3.4 (1.8) 3 [2, 4]	3.1 (1.7) 3 [2, 4]
Tacrolimus blood concentration (ng/mL)	7.5 (3.1) 7.1 [5.5, 9.0]	7.5 (3.0) 7.1 [5.5, 8.9]
Number of concentration levels per subject	9.0 (1.8) 9.0 [10.0, 10.0]	9.5 (1.2) 10.0 [10.0, 10.0]
Total cohort concentrations	3258	2116
Total days of follow-up	588 (295) 557 [361, 854]	525 (268) 482 [324, 686]

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 (SE) median [interquartile range] for continuous variable.

Base Model		Covariate Models				
Entire Cohort (N=363)			Entire Cohort (N=363)		Reduced Cohort (N=223)	
With Last-Dose Time			With Last-Dose Time	Without Last-Dose Time	With Last-Dose Time	Without Last-Dose Time
Objective Function	**11768**	**Objective Function**	**11633**	**11638**	**7347**	**7350**
Parameter	**Estimate (SE) [95% CI]**	**Parameter**	**Estimate (SE) [95% CI]**			
		$CL_{ij}=\theta_1 \times SNP^{\theta_2} \times (age_{ij}/47)^{\theta_3} \times (alb_{ij}/4.1)^{\theta_4} \times (hgb_{ij}/12.5)^{\theta_5}$				
CL	40.0 (1.7) [36.6, 43.4]	θ_1	31.9 (1.4) [29.2, 34.7]	31.9 (1.4) [29.1, 34.7]	28.2 (1.2) [25.8, 30.5]	28.4 (1.1) [26.2, 30.6]
		θ_2	0.70 (0.07) [0.56, 0.84]	0.71 (0.07) [0.57, 0.85]	0.69 (0.07) [0.55, 0.83]	0.69 (0.07) [0.55, 0.83]
		θ_3	-0.19 (0.08) [-0.35, -0.03]	-0.19 (0.08) [-0.35, -0.03]	-0.25 (0.10) [-0.45, -0.05]	-0.23 (0.10) [-0.43, -0.03]
		θ_4	0.29 (0.15) [0.00, 0.58]	0.29 (0.16) [-0.02, 0.60]	0.23 (0.18) [-0.12, 0.58]	0.22 (0.18) [-0.13, 0.57]
		θ_5	-0.29 (0.10) [-0.49, -0.09]	-0.28 (0.10) [-0.48, -0.08]	-0.25 (0.13) [-0.50, 0.00]	-0.26 (0.13) [-0.51, -0.01]
		$V_i= \theta_6$				
V	4697 (622) [3478, 5916]	θ_6	3735 (637) [2486, 4984]	4023 (607) [2834, 5212]	3034 (582) [1893, 4175]	3585 (567) [2475, 4696]
ω_{CL} (%CV)	60.9 (3.9) [53.3, 68.5]	ω_{CL} (%CV)	49.4 (4.2) [41.2, 57.6]	48.6 (4.2) [40.4, 56.7]	43.2 (2.4) [38.5, 47.8]	42.7 (2.2) [38.4, 47.1]
ω_V (%CV)	72.0 (11.0) [50.4, 93.6]	ω_V (%CV)	79.7 (16.0) [48.3, 111.1]	73.9 (12.4) [49.7, 98.2]	81.7 (12.0) [58.1, 105.2]	63.7 (12.6) [39.0, 88.3]
$\sigma^2_{additive}$ (ng/ml)	1.99 (0.32) [1.36, 2.62]	$\sigma^2_{additive}$ (ng/ml)	1.87 (0.34) [1.20, 2.54]	1.84 (0.37) [1.11, 2.57]	0.50 (0.89) [-1.24, 2.24]	0.42 (0.99) [-1.52, 2.36]
$\sigma^2_{proportional}$ (%CV)	31.5 (3.4) [24.8, 38.1]	$\sigma^2_{proportional}$ (%CV)	32.7 (3.2) [26.3, 39.0]	33.1 (3.4) [26.5, 39.7]	38.3 (2.1) [34.2, 42.3]	38.6 (2.1) [34.6, 42.6]

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 (SE) [Wald 95% confidence interval]. *SNP* is coded continuously as 1, 2, 3 for 0, 1, and 2 mutations, respectively. *age*, *alb*, and *hgb* represent covariate values for subject *i* at time *j*.

	k_a = 4.5	k_a = 3.09	k_a = 0.375
Objective Function	**11633**	**11632**	**11618**
Parameter		**Estimate (SE) [95% CI]**	
$CL_{ij}=\theta_1 \times SNP^{\theta 2} \times (age_{ij}/47)^{\theta 3} \times (alb_{ij}/4.1)^{\theta 4} \times (hgb_{ij}/12.5)^{\theta 5}$			
θ_1	31.9 (1.4) [29.2, 34.7]	32.0 (1.4) [29.2, 34.7]	32.4 (1.4) [29.7, 35.2]
θ_2	0.70 (0.07) [0.56, 0.84]	0.71 (0.07) [0.57, 0.85]	0.72 (0.07) [0.58, 0.86]
θ_3	-0.19 (0.08) [-0.35, -0.03]	-0.19 (0.08) [-0.35, -0.03]	-0.19 (0.08) [-0.35, -0.03]
θ_4	0.29 (0.15) [0.00, 0.58]	0.29 (0.15) [0.00, 0.58]	0.30 (0.16) [-0.01, 0.61]
θ_5	-0.29 (0.10) [-0.49, -0.09]	-0.29 (0.10) [-0.49, -0.09]	-0.30 (0.10) [-0.50, -0.10]
$V_i = \theta_6$			
θ_6	3735 (637) [2486, 4984]	3752 (635) [2508, 4997]	3723 (623) [2503, 4943]
ω_{CL} (%CV)	49.4 (4.2) [41.2, 57.6]	49.4 (4.2) [41.3, 57.6]	48.8 (4.0) [41.0, 56.5]
ω_V (%CV)	79.7 (16.0) [48.3, 111.1]	79.3 (15.9) [48.1, 110.5]	77.2 (15.7) [46.4, 108.0]
$\sigma^2_{additive}$ (ng/ml)	1.87 (0.34) [1.20, 2.54]	1.88 (0.34) [1.21, 2.55]	1.86 (0.33) [1.21, 2.51]
$\sigma^2_{proportional}$ (%CV)	32.7 (3.2) [26.3, 39.0]	32.7 (3.2) [26.3, 39.0]	32.8 (3.1) [26.6, 38.9]

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 (SE) [Wald 95% confidence interval]. *SNP* is coded continuously as 1, 2, 3 for 0, 1, and 2 mutations, respectively. *age*, *alb*, and *hgb* represent covariate values for subject *i* at time *j*.

Figure 1. Last-Dose Time Extraction. Extraction and processing examples for time at which a subject 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. (A) Observed vs. predicted concentration plots on the log scale for both population (left) and individual (right) level predictions; (B) Individual weighted residuals plots against time (left) and population predicted concentration (right); (C) the visual predictive check.

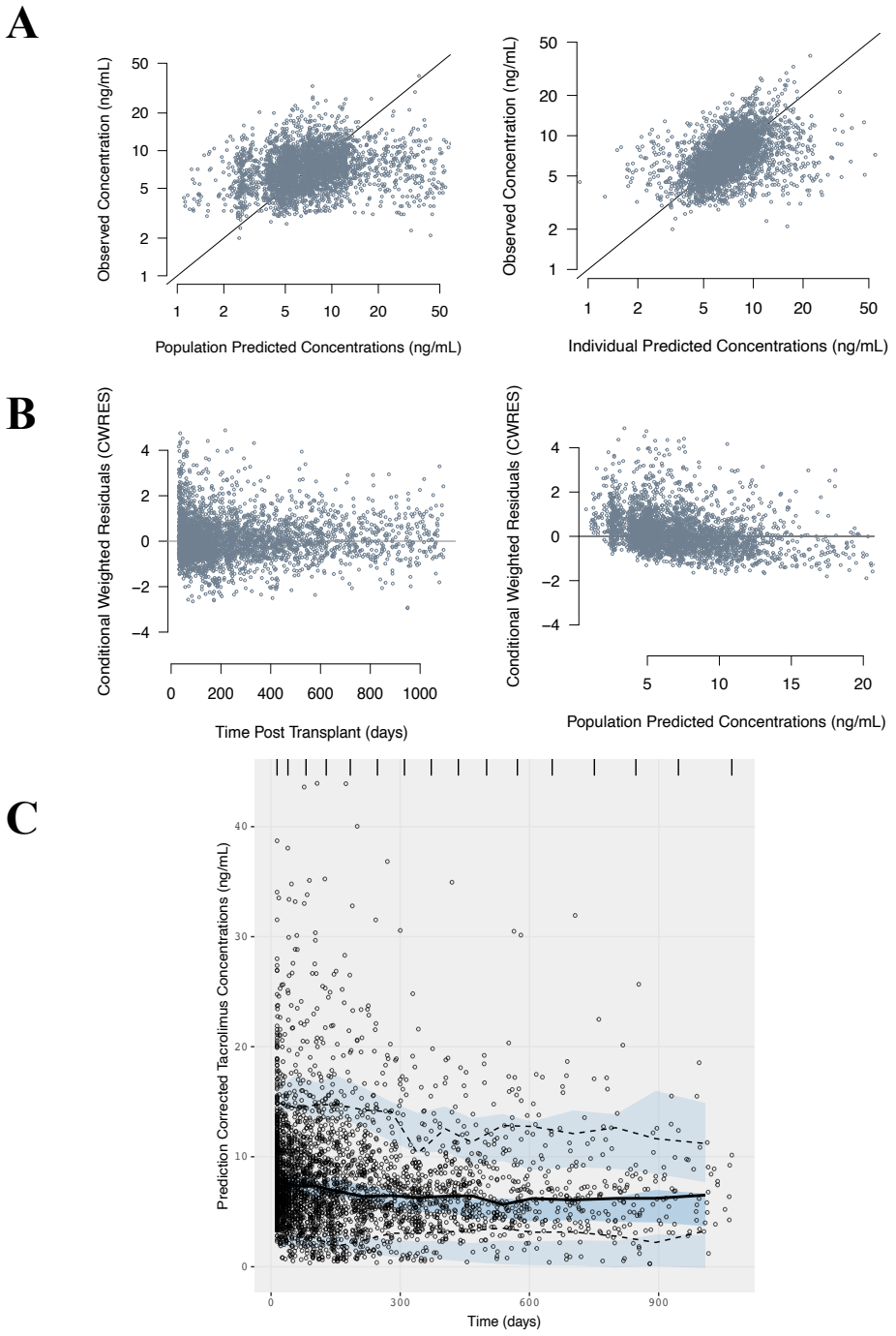
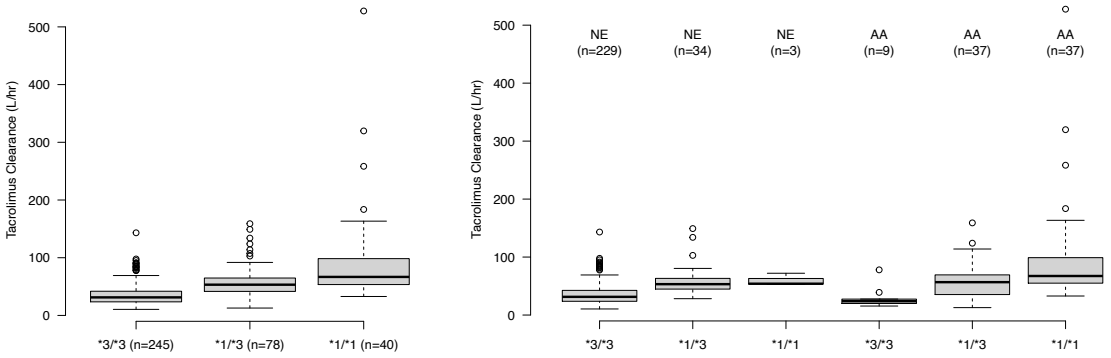
Figure 3: Estimated Clearance Among CYP3A5 SNPs. Box and whisker plots for clearance by CYP3A5 SNP (left) and clearance by CYP3A5 SNP among African ancestry (AA) and Northern European ancestry (NE) subjects.