POPULATION PHARMACOKINETIC AND MACHINE LEARNING METHODS TO INCREASE PRECISION OF ADHERENCE ASSESSMENTS FOR HIV PREVENTION

by

MUSTAFA E. IBRAHIM

BS, University of Cincinnati, 2010

MS, Northeastern University, 2016

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Pharmaceutical Sciences Program

2020
This thesis for the Doctor of Philosophy degree by

Mustafa E. Ibrahim

has been approved for the

Pharmaceutical Sciences Program

by

Laura Saba, Chair

Peter L. Anderson, Advisor

Jose R. Castillo-Mancilla

Jennifer J. Kiser

Samantha MaWhinney

Date: 08/14/2020
ABSTRACT

Adherence to a prescribed medication regimen is an important consideration in both clinical trial conduct and real-world treatment plans, and it is increasingly evident that adherence predicts better treatment outcomes. Adequate adherence is the determining factor of treatment efficacy for HIV pre-exposure prophylaxis with daily tenofovir disoproxil fumarate/emtricitabine (F/TDF), and objective and precise methods to assess adherence are therefore needed. A previous study evaluated the pharmacokinetics and dose proportionality of tenofovir-diphosphate, the active intracellular anabolite of tenofovir, in red blood cells and established gradient adherence benchmarks based on these concentrations. Subsequent analyses indicated four doses/week as corresponding to the concentration at which the relative risk reduction of HIV acquisition is 100% (86%-100%). However, several limitations exist with these interpretations that can impede generalized implementation in clinical trials and clinical practice. The focus of this work was to improve upon the precision of the use of these concentrations as an objective measure of adherence by population pharmacokinetic/pharmacodynamic modeling and machine learning methods to characterize inter-individual variability of tenofovir-diphosphate concentrations and develop tailored adherence benchmarks based on individual variations in concentration. We found that current interpretations generalize to overly conservative estimates of adherence and that individual variations in drug concentration are dependent on body weight...
and platelet count. Incorporation of these variables into adherence assessments decreased misclassification rates by almost half and identified a sub-population of those weighing less than 70 kg as having the highest baseline concentrations. When these tailored benchmarks were applied to a “real-world” population of individuals who participated in the Open Label Extension of the iPrEx trial, 24/28 observed HIV infections occurred in individuals weighing less than 70 kg, indicating that these individuals were even less adherent than previous interpretations implied. These results indicate the probability of increased pharmacological forgiveness of F/TDF for PrEP and provide further support for intermittent PrEP dosing. In addition, four distinct adherence patterns were identified in these individuals, which provided an additional characteristic for consideration in the allocation of adherence and risk-reduction counseling.

The form and content of this abstract are approved. I recommend publication.

Approved: Peter L. Anderson
DEDICATION

I would like to dedicate this work to every individual who found the courage to carve their own path, to find and embrace their voice and to relentlessly believe in themselves through every challenge and obstacle that came their way. I could not have done that myself without the support, love and friendship of my older siblings, Rania, Nasrien, and Omar, or without the encouragement from my biggest fans, my mom and dad. I am grateful for my best friends, Shaska and Morgan, who listened to me, had fun with me, and our time spent together was always a reminder to not sweat the small stuff. I learned something different from each of these people and I know that those lessons will propel me forward in every milestone and adventure to come.
ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Peter Anderson, for his mentorship, guidance and support over the last four years. He believed in my potential, trusted my work, and encouraged and challenged me to reach above the status quo in all aspects of my education. I am grateful to have had the opportunity to work in his lab and to feel as if I added value to the output of the lab. I would also like to thank my thesis committee for serving on the committee and being available for guidance and advice as I encountered roadblocks in my work: Dr. Jose Castillo-Mancilla for providing the clinical standpoint of my projects, teaching me about the patient population he treats and the feasibility of the implementation of my ideas in the clinic; Dr. Jennifer Kiser for her clinical pharmacology perspective, her emphasis on the clinical aspect of my work and always asking excellent and relevant questions that required me to re-think my approach to my research questions; Dr. Samantha MaWhinney for always being available to discuss and guide me in the biostatistical portions of my work, encouraging me to learn more about biostatistics, and for providing recommendations of additional course work that proved to be invaluable in understanding the foundations of biostatistics and mixed models; and Dr. Laura Saba for serving as Chair of my committee, for providing the foundations of machine learning that I needed to begin diving into the field and for always keeping the committee meetings on track and relevant to the objectives. I also would like to thank the lab manager, Lane Bushman, and all the Professional Research Assistants who tirelessly and diligently worked to enroll participants and perform the analytical analyses that enabled my work. The University of Colorado is a Certara Center of Excellence school. The Center or Excellence program supports leading institutions with Certara’s state of the art model-informed drug development software.
# TABLE OF CONTENTS

**CHAPTER**

I. SCOPE OF WORK ................................................................................................................. 1

II. REVIEW OF LITERATURE ............................................................................................... 14

III. HANDLING DATA BELOW THE LIMIT OF QUANTIFICATION IN POPULATION PK MODELING  ............................................................................................................................. 32

IV. INDIVIDUALIZED ADHERENCE BENCHMARKS FOR HIV PRE-EXPOSURE PROPHYLAXIS ....................................................................................................................... 50

V. REVISITING THE PHARMACOLOGICAL FORGIVENESS OF HIV PRE-EXPOSURE PROPHYLAXIS ....................................................................................................................... 72

VI. SUMMARY AND FUTURE DIRECTIONS ....................................................................... 75

REFERENCES ......................................................................................................................... 80

APPENDIX

A. POPULATION PHARMACOKINETICS OF ETRAVIRINE IN HIV-POSITIVE CHILDREN AGES 1-<6 YEARS ................................................................................................................. 93

B. COMPLETE R CODE ........................................................................................................ 111
**LIST OF ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIDS</td>
<td>Acquired Immunodeficiency Syndrome</td>
</tr>
<tr>
<td>AK2</td>
<td>Adenylate Kinase</td>
</tr>
<tr>
<td>ART</td>
<td>Antiretroviral Treatment</td>
</tr>
<tr>
<td>CK</td>
<td>Creatine Kinase</td>
</tr>
<tr>
<td>CNT</td>
<td>Concentrative Nucleoside Transporter</td>
</tr>
<tr>
<td>CrCL</td>
<td>Creatinine Clearance</td>
</tr>
<tr>
<td>ENT</td>
<td>Equilibrative Nucleoside Transporter</td>
</tr>
<tr>
<td>FTC</td>
<td>Emtricitabine</td>
</tr>
<tr>
<td>F/TDF</td>
<td>Tenofovir Disoproxil Fumarate and Emtricitabine</td>
</tr>
<tr>
<td>HIV</td>
<td>Human Immunodeficiency Virus</td>
</tr>
<tr>
<td>iPrEx</td>
<td>Preexposure Prophylaxis Initiative Trial</td>
</tr>
<tr>
<td>MRP</td>
<td>Multidrug Resistance-associated Protein</td>
</tr>
<tr>
<td>NDPK</td>
<td>Nucleoside Diphosphate Kinases</td>
</tr>
<tr>
<td>NRTI</td>
<td>Nucleos(t)ide Reverse Transcriptase Inhibitors</td>
</tr>
<tr>
<td>PBMC</td>
<td>Peripheral Blood Mononuclear Cells</td>
</tr>
<tr>
<td>PopPK</td>
<td>Population Pharmacokinetics</td>
</tr>
<tr>
<td>PEP</td>
<td>Post-exposure Prophylaxis</td>
</tr>
<tr>
<td>PGK</td>
<td>Phosphoglycerate Kinase</td>
</tr>
<tr>
<td>PK</td>
<td>Pyruvate Kinase</td>
</tr>
<tr>
<td>PK-PD</td>
<td>Pharmacokinetic-Pharmacodynamic</td>
</tr>
<tr>
<td>PrEP</td>
<td>Pre-exposure Prophylaxis</td>
</tr>
<tr>
<td>RBC</td>
<td>Red Blood Cell</td>
</tr>
<tr>
<td>TAF</td>
<td>Tenofovir Alafenamide</td>
</tr>
<tr>
<td>TFV</td>
<td>Tenofovir</td>
</tr>
<tr>
<td>TFV-MP</td>
<td>Tenofovir-monophosphate</td>
</tr>
<tr>
<td>TFV-DP</td>
<td>Tenofovir-diphosphate</td>
</tr>
</tbody>
</table>
CHAPTER I

SCOPE OF WORK

Adherence in Clinical Trials and Therapeutic Care

Adherence to a medication regimen refers to the degree to which patients and clinical trial participants take prescribed medications as directed by their healthcare provider. In other words, adequate adherence by an individual consists of taking medication at the right dose, the right time, the right way (e.g., with or without food) and for the right duration. It is increasingly evident that higher adherence predicts better treatment outcomes, such as for viral suppression during antiretroviral treatment for HIV infection and for cardiovascular outcomes [1-3]. For example, patients with ≥80% adherence to antihypertensives were 45% more likely to achieve blood pressure control in one study and another study estimated a 25% increased hazard in mortality following nonadherence to statins after a myocardial infarction [4, 5]. In terms of HIV infection, a recent meta-analysis of the association between antiretroviral therapy (ART) adherence and virologic outcomes reported that the odds of virologic failure were three times higher for nonadherent patients and also estimated ≥80% as an optimal adherence threshold for virologic suppression [3]. These studies utilized subjective methods of adherence, such as Medication Event Monitoring System (MEMS), self-report and pill count, which do not directly confirm medication ingestion. One of the goals of this thesis was to develop quantitative and objective assessments of adherence using measured drug concentrations, which provide the most accurate evidence of drug ingestion [6]. In addition to negative treatment outcomes such as these, suboptimal adherence can also adversely affect the conduct and interpretation of results of clinical trials.
While first-in-human studies are typically done under stringent directly observed dosing conditions, later phases of clinical trials are typically conducted in unmonitored adherence settings or with unreliable methods for assessing adherence [7]. After analyzing approximately 17,000 participants across 95 clinical trials of various disease states, a meta-analysis reported progressively declining adherence over time. For daily dosing at day 100, 20% of participants had discontinued treatment and that number rose to 40% by the end of the first year [8]. The consequences of nonadherence in clinical trials can be deleterious, including effects such as failure to confirm efficacy; underestimated efficacy; underestimation of risk of harm; drug resistance; and treatment failure [7]. Unrecognized nonadherence can also increase variance of response due to the variable underdosing, therefore weakening the study’s statistical power and requiring larger sample sizes [9, 10]. Consequently, a potentially effective drug can be deemed ineffective and increasing the doses studied due to unrecognized nonadherence can cause toxicity to those individuals who are fully adherent to the medication. In an analysis of 354 new molecular entities approved between 1980 and 1999, 58/73 (79%) that underwent post-approval dosage changes were safety-motivated and resulted in net dosage decreases [11]. The impact of nonadherence cannot be overlooked as an important variable contributing to these dosage decreases. In summary, adherence is a critical component of successful clinical trial conduct and “real-world” treatment plans. One area that emphasizes the importance of adherence on treatment outcomes is in Human Immunodeficiency Virus (HIV) prevention, which will be the focus of this work.

**HIV Overview**

HIV belongs to a class of viruses known as retroviruses, an RNA virus that inserts a DNA copy of its genome into the host cell genome for replication. The main mode of HIV
transmission is through sexual contact, although it can also be transmitted through contact with infected blood after reusing or sharing needles for intravenous drug use or occupational hazards such as accidental needlesticks. The main targets of HIV are CD4+ T-cells, although the virus can infect other immune cells, such as macrophages and immature dendritic cells [12]. After initial transmission, the first step in viral replication is the attachment of the viral particle to the CD4+ receptor and coreceptors (CXCR4 and CCR5) of the host cell [13]. This initial fusion is critical for the release of the virion contents into the cytoplasm of the cell. Once inside the cytoplasm, HIV reverse transcriptase converts viral RNA into DNA. This DNA then translocates into the host cell’s nucleus, where the HIV integrase enzyme splices the viral DNA into the host cell’s genome. Once this cell is activated, viral replication occurs and can indefinitely infect more cells, ultimately leading to CD4+ T cell depletion, increased susceptibility to opportunistic infections, and development of Acquired Immunodeficiency Syndrome (AIDS) when the CD4+ cell count drops below 200 cells per cubic millimeter of blood, without the use of antiretroviral therapy. In addition, if the host cell remains inactivated, HIV can remain latent for many years, further complicating the complete eradication of the virus [14].

Several classes of antiretroviral therapy have been developed that have significantly reduced mortality rates and led to sustained viral suppression with continued use. These drug classes target HIV at different phases of the viral replication cycle, including infusion inhibitors, nucleos(t)ide reverse transcriptase inhibitors (NRTI), non-nucleoside reverse transcriptase inhibitors, integrase inhibitors and protease inhibitors [15]. The focus of this work centers on the use of tenofovir disoproxil fumarate and emtricitabine (F/TDF), which belong to the NRTI class of antiretroviral drugs. NRTIs are analogues of naturally occurring purine and pyrimidine deoxynucleotides, which are needed for viral DNA elongation and production [15, 16]. However,
NRTIs lack a 3’-hydroxyl group on the deoxyribose moiety, which prevents the formation of the phosphodiester bond needed for the incoming nucleoside triphosphate, thereby terminating the growth of the viral DNA. NRTIs require phosphorylation inside the cell by cellular kinases to exhibit their antiviral effect. NRTIs were the first drug class to be approved by the Food and Drug Administration for the treatment of HIV and have since continued to be a fundamental component of combination highly active antiretroviral treatment [17]. The high efficacy of NRTI regimens and several other favorable characteristics of F/TDF such as once-daily dosing, lack of contraindicated medications, and a favorable safety profile, led to the continued study of this drug for HIV pre-exposure prophylaxis.

**HIV Prevention**

The rationale for a pharmacological basis for HIV prevention, known as HIV pre-exposure prophylaxis (PrEP), emerged from previous studies confirming the risk reduction of HIV acquisition after occupational and nonoccupational exposure to HIV, known as post-exposure prophylaxis (PEP), and from studies of mother-to-child transmission. Administration of zidovudine was shown to reduce the risk of perinatal HIV transmission by 67.5% (95% confidence interval: 40.7%-82.1%) in a placebo-controlled trial [18]. Due to ethical and logistical reasons, no prospective randomized controlled trials of PEP have been conducted and most of the evidence for PEP after occupational exposure has been linked to one retrospective study of zidovudine, which demonstrated an 81% risk reduction of HIV acquisition in healthcare workers [19]. One meta-analysis reported that there have been at least 24 cases of HIV seroconversion after occupational PEP initiation, suggesting that this regimen is not 100% effective [20]. More studies of PEP have been conducted after nonoccupational exposure, such as sexual encounters or intravenous drug use [21-25].
The United States Department of Health and Human Services issued recommendations regarding nonoccupational PEP treatment in 2005, based on animal studies, stating that a 28-day highly active antiretroviral therapy regimen is recommended for individuals seeking care ≤72 hours after exposure to potentially infectious bodily fluids from a person known to be infected with HIV [26]. PEP was not recommended if the exposure occurred more than 72 hours after seeking care. These recommendations require that a person recognize when they have been potentially exposed to HIV and begin therapy within 72 hours, both of which are significant limitations to PEP. This led to the investigation of other preventative methods, namely pre-exposure prophylaxis (PrEP) [27].

Tenofovir disoproxil fumarate and emtricitabine (F/TDF) was investigated for PrEP due to its favorable characteristics such as once-daily dosing, lack of contraindicated medications, and a favorable safety profile. Initial animal studies in mice and nonhuman primates provided evidence of the efficacy of PrEP and encouraged transition into human studies [28-32]. For example, a study of bone marrow-liver-thymus (BLT) mice transplanted with human CD4\(^+\) T-cells showed that PrEP with F/TDF was highly effective at preventing intravaginal HIV transmission [28]. Importantly, studies conducted in non-human primates showed that the combination of both emtricitabine and TDF was more effective for PrEP than either drug alone, and that the effectiveness depended on drug administration before and after SHIV (Simian-HIV) challenge, rather than only after SHIV exposure [31, 32]. These encouraging results prompted the investigation of F/TDF for PrEP in humans.

The first clinical trial evaluating the feasibility and efficacy of PrEP was conducted via a topical administration of 1% tenofovir gel to 445 South African women, which showed that the tenofovir gel reduced HIV acquisition by 39% [33]. Studies of oral PrEP regimens also
demonstrated an acceptable safety profile and additional benefit of the prevention of HIV. In a placebo-controlled trial of 936 women at a high risk of HIV infection, there were no significant differences between treatment groups in clinical or laboratory adverse effects [34]. Effectiveness could not be definitively concluded in this study due to early closure of two study sites, which reduced the statistical power needed to test effectiveness. Additional studies confirmed the tolerability and acceptable safety profile of F/TDF and reported declining or stable changes in high risk sexual behavior among those receiving F/TDF, a phenomena known as risk compensation [35, 36].

In 2010, the results of the first clinical trial to evaluate oral PrEP were reported [37]. The iPrEx trial enrolled 2,499 men and transgender women who have sex with men from six countries in North and South America, Africa and Asia to receive daily F/TDF or placebo. The trial concluded an overall 44% risk reduction of HIV acquisition with F/TDF compared to placebo (95% CI 15-63%). In 2012, positive results were published for the TDF2 and Partners PrEP trials in heterosexual men and women and serodiscordant couples [38, 39]. Although the TDF2 study was concluded early due to low retention, a significant 62% protective effect of F/TDF was reported (95% CI 22-83% risk reduction). In addition, the Partners PrEP trial in 4,747 HIV-serodiscordant couples indicated a 67% (95% CI 44-81%) relative reduction in HIV incidence with TDF alone and 75% (95% CI 55-87%) relative reduction with F/TDF. These results convey the protective benefit of oral F/TDF for PrEP in three distinct populations: men and transgender women who have sex with men, serodiscordant couples and heterosexual men and women at high risk for HIV infection.

Nonetheless, trials in women conducted via both tenofovir gel and oral F/TDF were stopped early due to futility. In the Microbicide Trials Network VOICE trial, where 5,029
women received tenofovir gel, TDF or F/TDF, the use of the tenofovir gel was discontinued early due to futility because of no differences found between the tenofovir gel and placebo gel. Subsequent analyses revealed low adherence rates among these women, which will be addressed in more detail in the next section. This was also observed in the FEM-PrEP trial of 2,120 women from Kenya, South Africa and Tanzania, in which oral PrEP with daily F/TDF did not significantly reduce the rate of HIV infection compared to placebo [40].

The overall results reported from these trials were positively modified after assessing and accounting for participant adherence during these trials, which is now considered to be the determining factor of treatment efficacy for oral PrEP. These trials led to FDA approval of oral daily F/TDF in 2012 for daily use in individuals at high risk for HIV acquisition.

Adherence and Efficacy of PrEP

The overall range of the protective efficacy of oral F/TDF versus placebo in the initial studies ranged from 0-75%. Adherence to the prescribed daily regimen is a critical component of the wide range of efficacy and has since been recognized as the major determinant of efficacy of PrEP. In the clinical trial evaluating 1% tenofovir gel, overall incidence of HIV was reduced by 39%, but that estimate rose to 54% among highly adherent participants (>80% adherence as assessed by the number of gel applicators returned), compared to 38% and 28% in intermediate and low adherers, respectively [33]. More objective methods of adherence have also been evaluated through the use of plasma tenofovir concentrations. Notably, plasma drug detection can only indicate recent dosing up to the last seven days given the shorter half-life of these moieties.

In the Partners PrEP trial, only 31% of 29 participants who became infected with HIV had detectable plasma tenofovir at the seroconversion visit, compared to 82% of 902 random
seronegative samples [39]. High adherence in this population was also corroborated by high pill-count measures, leading to the estimated 86% risk reduction of HIV. In the iPrEx trial, only 3 out of 34 individuals who became infected with HIV had detectable drug concentrations, compared to 22 of 43 matched seronegative controls [37]. The relative risk reduction increased from 44% to 92% among those with detectable drug levels and was even higher (95%) after controlling for high-risk sexual behavior.

The use of intracellular drug concentrations as objective measures of adherence was also first investigated in the iPrEx trial [41]. The active intracellular moiety of TDF, tenofovir-diphosphate (TFV-DP), exhibits a longer half-life in PBMCs (~4 days) than the parent plasma tenofovir (~17 hours), which allows for cumulative dosing assessments, reported as the average number of doses per week taken. When TFV-DP was analyzed in PBMCs, only three of the 34 individuals (9%) who acquired HIV had detectable TFV-DP concentrations, which were in the range of less than two doses per week. A large fraction of iPrEx participants showed evidence of less than two doses per week. Additional evidence of the significant effects of adherence on PrEP treatment outcomes was generated in post hoc analyses of the VOICE and FEM-PrEP trials, both conducted in women [40, 42]. In these discontinued trials, <40% of women in the seroconversion and control arms had detectable plasma drug concentrations, indicating no recent use of PrEP.

Taken together, these results highlight the critical component of maintaining adequate adherence to F/TDF to prevent HIV, especially during periods of high risk of acquisition. In addition, these results also imply that less than daily dosing produced the protective benefit of PrEP observed, as evidenced by the large number of moderately adherent participants across these trials. Objective and precise methods to assess adherence are needed to further evaluate and
refine adherence-efficacy relationships, as well as to support future PrEP studies and aid health care providers in identifying individuals at-risk for acquiring HIV while on PrEP.

**Objective Measures of PrEP Adherence**

Several subjective and objective methods – which can also be classified as direct or indirect assessments – exist for assessing adherence to PrEP, each carrying its own strengths and limitations [6]. Methods for indirect assessments include pharmacy refill records, Medication Events Monitoring Systems (MEMS), pill count, self-reported adherence via interview and questionnaires, and clinician assessments. Several studies have compared the consistency between different measures of adherence to HIV antiretroviral therapy, PrEP regimens, as well as across various other therapeutic areas [43-48]. Based on these and a multitude of other comparisons, it is apparent that large discrepancies exist between subjective measures of adherence, with self-report consistently identified as the weakest measure. For example, in the discontinued VOICE and FEM-PrEP trials, overall self-reported adherence was >90% although <40% of the participants showed evidence of recent PrEP use by assessment of detectable drug concentrations in plasma [40, 42]. In a study comparing adherence measures to HIV protease inhibitors, self-reported adherence was 93%, compared to 83% and 63% via pill count and MEMS assessments, respectively [48]. These discrepancies highlight the need for objective and direct measures for assessing adherence.

Direct measures of adherence include measurement of the drug and/or its metabolites in blood or urine, and is considered to be the most accurate evidence of drug ingestion [6]. Certainly, analysis of detectable drug concentrations in the initial PrEP trials demonstrated the importance of adherence and its relationship with efficacy at the trial level, a relationship that could not be discerned with self-reported adherence. Nonetheless, just as with indirect measures,
limitations exist with collecting drug concentrations. These include cost; the possibility of “white coat adherence” in that improved adherence is observed close to clinic visits; and the simple dichotomous Yes/No categorization that short half-life moieties specify without more information of cumulative patterns of nonadherence, including with the use of plasma tenofovir. The focus of this work addresses each of these limitations through the use of tenofovir diphosphate (TFV-DP) concentrations, the active intracellular anabolite of tenofovir, in red blood cells (RBCs).

The clinical pharmacology of F/TDF and the conversion from TDF to intracellular TFV-DP will be covered in detail in the next chapter. Concentrations of TFV-DP were found to exhibit a remarkably long median half-life (interquartile range) of 17.1 (15.7-20.2) days in RBCs [49]. This long half-life and resulting 25-fold accumulation (Accumulation Ratio = $1/(1-e^{-ke\cdot\text{tau}})$), where $ke$ is the elimination rate constant and tau is the dosing interval), from first dose to steady-state after daily dosing offers an opportunity to measure cumulative adherence in the preceding 1-2 months, reported as the average number of doses per week ingested. The shorter half-life of tenofovir in plasma (~17 hours) does not allow for this type of cumulative assessment, but rather a dichotomous Yes/No result of whether a recent dose was ingested up to the preceding 7 days, limitations of which were described above. Other recent dosing markers of F/TDF adherence include sample detection in urine and saliva. Urine tenofovir concentrations generally mirror plasma concentrations and can dichotomously assess adherence in the previous 7 days. Concentrations above 1,000 ng/ml indicate dosing in the previous 2-3 days, 10-1000 ng/ml in the preceding 3-7 days and less than 10 ng/mL as >7 days [50]. Both tenofovir and emtricitabine have low penetration ratios in saliva (0.02 and 0.17, respectively), and thus drug can only be detected in the preceding 24-48 hours [51].
The half-life of TFV-DP in PBMCs is approximately 4 days, which translates to an 8-fold accumulation from first dose to steady state. TFV-DP concentrations in PBMCs can estimate cumulative adherence in the preceding 14 days. However, costly, specialized, and time-consuming processing of PBMC samples preclude their widespread implementation. Hair collection is also a viable marker of cumulative adherence that can estimate gradient levels of adherence in the preceding 4-6 weeks, given the slow hair growth rate over time.

The focus of this thesis was to further develop the use of intracellular TFV-DP concentrations in RBCs. A recent study evaluated the pharmacokinetics and dose-proportionality of TFV-DP in RBCs after directly observed dosing of F/TDF, sharing a study design similar to that of the STRAND study that evaluated adherence benchmarks of TFV-DP in PBMCs for use in the iPrEx trial \[52\]. Participants were randomized to receive two of the three directly observed dosing regimens of 33%, 67% or 100% daily dosing, which translated to an average of 2.3, 4.7 and 7 doses per week, respectively. The time-concentration plots of the accumulation of TFV-DP in RBCs for each dosing regimen is shown in Figure 1. Dose-proportionality was confirmed and thereby enabled the generation of adherence gradients (<2, 2-3, 4-6, 7 doses/week) based on interquartile ranges of these concentrations. However, several limitations are evident in this study, which is the primary motivation of the current work. A 30% coefficient of variation was observed for TFV-DP concentrations, which can lead to misclassification of adherence, decrease the precision of personalized assessments and adherence-efficacy relationships, and over- or under-estimate the true adherence thresholds of each dosing classification. The study only evaluated three fixed doses (2.3, 4.7 and 7 doses/week), necessitating additional simulation-based analyses of the full range of adherence patterns. Finally, these thresholds were also
Figure 1 TFV-DP accumulation in Dried Blood Spots following 100% (blue), 67% (green) and 33% (red) daily dosing for 12 weeks, followed by washout. Open circles are observed data and filled squares are model fits. Smoothed curves show data with population fitted values over 12 weeks on drug (Adapted from Anderson PL, et al. *Antimicrobial agents and chemotherapy*, 62(1), pp.e01710-17)
developed by inter-quartile ranges, indicating at least a 25% misclassification rates by definition at each decision boundary. Given the importance of adherence to PrEP treatment outcomes, the widespread use of these thresholds to support current and future trials of PrEP, and ongoing evaluations of the pharmacological forgiveness and intermittent or reduced dosing regimens of F/TDF, it is essential to address the existing limitations of this objective measure of adherence in novel ways and to apply this new understanding to “real-world” populations of individuals receiving PrEP.

**Thesis Aims and Clinical Significance**

The major aims of this work were to:

1) Identify gaps in the current knowledge of pharmacometric modeling, particularly methods to handle data that are below the limit of quantification (BLQ), that will improve the accuracy of model development in this work and establish guidelines for the larger pharmacometric community

2) Improve the precision of the current use of TFV-DP concentrations as an objective measure of adherence to PrEP through novel methods that incorporate individual characteristics

3) Apply the findings of this work to a “real-world” population of men who have sex with men enrolled in the Open Label Extension of the iPrEx trial to characterize tailored adherence in this population, evaluate the pharmacological forgiveness of F/TDF, and to better understand patterns and trajectories of adherence to enable identification and effective allocation of adherence counseling and monitoring resources
CHAPTER II

REVIEW OF LITERATURE

Tenofovir Disoproxil Fumarate Cellular and Clinical Pharmacology

This dissertation work focused on potential sources of inter-individual variability of tenofovir disoproxil fumarate and emtricitabine (F/TDF; Truvada©; Gilead Sciences), which belong to the nucleotide reverse transcriptase inhibitor (NRTI) class of antiretroviral drugs. As such, it is critical to have an in-depth understanding of the cellular and clinical pharmacology of this drug and the pharmacokinetic processes from absorption to elimination that can impact these processes.

Emtricitabine (FTC) is a nucleoside analog of cytidine, which is phosphorylated intracellularly to the active form, emtricitabine triphosphate [53]. FTC has potent activity against HIV and hepatitis B infection and was approved for HIV treatment in combination with other antiretroviral drugs. The half-life of emtricitabine in plasma (8-10 hours) and longer intracellular half-life of approximately 39 hours, favored its use for once-daily dosing. The second component of F/TDF is TDF, which is the focus of this thesis.

Tenofovir (TFV), which contains a stable phosphonate bond, is structurally similar to the purine deoxyadenosine monophosphate (dAMP). Several studies demonstrated the enhanced antiviral effects of TDF that led to its approval in 2004 and has since become a fundamental staple of many highly active combination antiretroviral regimens [54-58]. The clinical pharmacology of TDF will be described in detail in the following sections.

TDF is a prodrug of TFV, which is the parent form of the drug responsible for the ultimate antiviral effect. A prodrug formulation was necessary for the successful development and optimization of F/TDF because of the ionic nature (i.e., two negative charges on the
phosphonate) of TFV that significantly limited its permeability across intestinal mucosa. The oral bioavailability of TFV was only 5-18% in rats, dogs and cynomolgus monkeys [59]. However, after the addition of two ester groups to the phosphonate moieties of TFV, bioavailability increased to 30% with increased chemical and intestinal *in vitro* stability, prompting the use of the prodrug form in subsequent clinical trials [59, 60]. Subsequent studies have investigated factors that could potentially increase the bioavailability of TDF. Oral bioavailability of TDF increased to 39% when administered with a high-fat meal (approximately half of the macronutrient composition as fats), concurrent with a 40% increase in the area under the concentration curve (AUC∞), a 14% increase in the Cmax, and slower absorption (Tmax of 2 hours with high-fat vs. 1 hour with fasted conditions) [61]. However, these results were deemed to be clinically insignificant given the wide therapeutic index of TDF. Other studies have evaluated esterase inhibition and transporter-inhibition mediated by protease inhibitors on the bioavailability of TDF [62-64].

After oral administration, TDF undergoes sequential ester hydrolysis by carboxylesterases and phosphodiesterases to yield the monoester TFV first, then the parent TFV moiety. Esterases are ubiquitously found throughout the body, including in intestinal cells, hepatocytes, plasma, and other tissues of the body. As such, TDF is efficiently cleaved upon first pass and TFV is the predominate form circulating in the blood. Overall, protein binding is low, with less than 1% and 7% bound in plasma and serum, respectively [65]. TDF and TFV are not substrates of CYP enzymes and it is believed that TFV is excreted unchanged in urine by both filtration and tubular secretion [66].

While not a focus of this thesis, a second prodrug of tenofovir, tenofovir alafenamide (TAF), has recently been approved for HIV treatment and prevention [67]. TAF has been shown
to have greater stability in plasma compared to TDF, which results in higher capacity for cellular loading into PBMCs. Greater stability and cellular loading also results in approximately 90% lower plasma tenofovir concentrations compared to TDF administration, which significantly reduces the risk of kidney dysfunction associated with chronic TDF dosing. TAF also results in higher intracellular TFV-DP concentrations in PBMCs and higher antiviral potency. The metabolism of TAF, however, is distinct from that of TDF. Once inside the cell, TAF is metabolized by lysosomal protease Cathepsin A to tenofovir-alanine. This conversion to tenofovir-alanine is mediated by carboxylesterase 1 (CES1) in liver cells. Acidic hydrolysis in lysosomes facilitates the conversion of tenofovir-alanine to tenofovir, which can then be phosphorylated into TFV-DP by cellular kinases, similar to TDF.

All NRTIs require intracellular relocation and phosphorylation to the active triphosphate anabolite form to exert their antiviral effects. After TDF cleavage to generate TFV, TFV is believed to enter the cell by endocytosis. This process was not believed to be transporter-mediated or saturable. It is also hypothesized that TDF can also be directly loaded into target cells. More recent evidence also illustrates an important contribution of the monoester to the cellular loading and formation of tenofovir-diphosphate (TFV-DP) [68]. Importantly, it has been shown that TFV can be phosphorylated in both active and resting cells, which likely results in better antiviral activity in cells with low proliferative capabilities such as macrophages, monocytes and dendritic cells [69]. This observation is hypothesized to be due to the fact that TFV does not require the initial phosphorylation step, which is the rate-limiting step of other nucleoside analogs.

Once inside the cell, cellular kinases are responsible for two distinct phosphorylation steps to form tenofovir-monophosphate (TFV-MP) and then the active anabolite, TFV-DP. It has
been demonstrated that specific kinases are responsible for each phosphorylation step and that these steps may occur in a compartment-specific manner. It is hypothesized that adenylate kinase (AK2), a ubiquitous enzyme found in mitochondria of lymphocytes, is responsible for the initial phosphorylation of TFV to TFV-MP and nucleoside diphosphate kinases (NDPK) are responsible for the formation of TFV-DP, including NDPK1, NDPK2, pyruvate kinase (PK), phosphoglycerate kinase (PGK), and creatine kinase (CK) [69, 70]. It was also noted that differential expression and activity of these enzymes in tissues susceptible to HIV infection may play a role in differential pharmacology of TDF. It was observed that pyruvate kinase, muscle (PKM) and liver and red blood cell (PKLR), phosphorylated TFV-MP to TFV-DP in PBMCs and vaginal tissue, while creatinine kinase (CKM) was responsible for the phosphorylation in the colon [70]. Differing routes of TFV-DP activation highlights the possibility of tissue-specific pharmacokinetic-pharmacodynamic (PK-PD) relationships. In terms of pharmacogenomics, low frequency of variants in these genes in this study and the confounding issue of adherence prevented conclusive evidence of their impact, highlighting the need for further pharmacogenomic studies of these enzymes. This dissertation will explore pharmacogenomic associations between variants in genes relevant to TDF pharmacology and PK processes.

Several transporters are also involved in the disposition of TDF and a thorough understanding of these processes is important to elucidate the potential sources of variability along the TDF to TFV-DP conversion spectrum. At the absorption phase, TDF absorption can be influenced by P-gp mediated efflux, as assessed by the diminished transport polarity of TDF across Caco-2 cells incubated with P-gp inhibitors [60]. MRP2, which is heavily localized in the liver, was also found to affect hepatobiliary elimination of TDF, TFV monoester and TFV, while OAT1/3 and MRP4 are implicated in TFV-associated renal proximal tubule toxicity [71, 72].
Just as with kinase expression described above, differential expression of these transporters can impact TDF pharmacology and PK-PD relationships. For example, high expression of nucleoside transporters (e.g., CNT2 and ENT2) were found in colorectal tissue and were higher in colorectal CD4+ T cells compared to circulating CD4+ T cells [73, 74]. In contrast, CNT and ENT transporters were expressed in ectocervix tissue, while none of the CNT transporters or ENT2 were expressed in vaginal tissue [75]. Differential expression such as this in tissues susceptible to HIV infection provide further evidence that TDF dosing may not be a one-size-fits all and that additional demographic and physiological factors should be considered for dose recommendations.

Several population pharmacokinetic (PopPK) and statistical models have been developed to explain inter-individual variability in pharmacokinetic parameters of TDF disposition. In doing so, tailored dosing recommendations can be made, and within the context of the present work, more accurate models of TDF and TFV-DP disposition can be developed. Individuals with hepatic impairment exhibited similar systemic exposure to individuals without impairment, as expected for a renally eliminated drug such as tenofovir. However, creatinine clearance <50 ml/min indicated a need for reduced TDF dosing [76].

Several factors can also influence the variability of response of TDF administration for both HIV prevention and treatment. These factors include variable adherence, differential tissue and cell type penetration of tenofovir and cellular activation states [77]. In addition, differences in TDF disposition have been previously observed between genders. A trend of higher TFV-DP concentrations in PBMCs among females was observed after the same dose of F/TDF, in addition to approximately 20% higher TFV-DP concentrations in red blood cells measured with dried blood spots [52, 78]. Nonetheless, additional insight is needed to explain these gender
differences. Age is also considered a potential source of variability, as those <25 years old showed faster plasma TFV clearance, despite higher intracellular TFV-DP concentrations [79]. African-Americans also exhibited lower TFV-DP concentrations in RBCs, which also needs further evaluation, as the association of hemoglobinopathies or genetic differences in relevant transporters and enzymes according to race cannot be discounted [77].

As expected, measures of renal function, such as creatinine clearance (CrCL) and the ratio body weight/serum creatinine, are related to plasma oral clearance and significantly explain some of the variability in plasma TFV concentrations [80, 81]. Variants in ABCC2 (MRP2) are associated with kidney tubular dysfunction, which can be explained by the increased accumulation of TFV in the tubules with reduced MRP2 transporter function [82, 83]. Other studies identified variants in CNT2 and MRP7 as covariates on plasma TFV clearance, while variants in MRP4 were associated with 35% higher TFV-DP concentrations [80, 84]. The high expression of these transporters in liver and kidney are consistent with their effects on toxicity and tenofovir systemic exposure.

Despite these known sources of variability, a large amount of inter-individual remains unexplained in PopPK models of tenofovir, and even higher variability is observed in models describing the kinetics of intracellular TFV-DP. After including covariates, inter-individual variability remained high on plasma tenofovir clearance, ranging from approximately 30-40% in previous PopPK models [81, 85-88]. In addition, inter-individual variability of TFV-DP kinetics described by indirect response models or a first-order constant linking TFV to TFV-DP formation showed inter-individual variabilities as high as 160% CV in various parameters [86, 88]. Of note, known adherence is a major confounding factor in several of these studies, which this present work will also address. Indeed, unrecognized adherence or ignoring true dosing
information can yield biased PK parameter estimates and significantly over-estimate inter-individual and residual variability [89, 90]. Interindividual variability in plasma tenofovir clearance was only 16% when electronic based health records were incorporated into the model [91]. These results summarize the need for improved models of plasma and intracellular kinetics of TDF through identification of additional covariates and use of directly observed dosing to eliminate the confounding factor of adherence.

Large inter-individual and residual variability have adverse consequences in interpretations, simulations and applications of PopPK models. It is important to identify sources of variability from a comprehensive list of potential sources, including pharmacogenomics, pharmacometabolomics, demographics and blood laboratory values. With the feasibility of collection of big data such as this, so too comes the need for improved methods to predict outcomes and identify a variable’s importance in those predictions. Machine learning algorithms offer a viable avenue for such investigations.

**Biostatistics**

**Machine Learning**

In an era where “big data” is abundant and becoming increasingly feasible to collect, machine learning plays an important role in learning patterns within the data to accurately classify or predict future observations. Machine learning can broadly be defined as “a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty” [92]. Machine learning algorithms can be thought of as analogous to human learning, for example in the area of facial recognition. Just as a child learns to differentiate a picture of a dog from a cat from a human face as more pictures are presented to him or her, machine learning algorithms
learn distinct patterns within dog, cat and human face pictures that allow it to predict and classify future pictures as one of these species. This is an example of supervised learning, where the target outcome is known during the training/learning phase. Unsupervised learning, based mainly on clustering methods without known target outcomes, are beyond the scope of this work.

Machine learning can be used for many reasons, including for learning associations, classification, pattern recognition, knowledge extraction, regression and data reduction [93]. Many supervised learning algorithms exist for both regression- and classification-based tasks, and each carry their own strengths and limitations with respect to metrics such as accuracy and interpretability. A general process or workflow for conducting machine learning analyses begins by data collection and preparation, followed by feature choice, algorithm choice, parameter and model tuning, training and finally evaluation [92]. For the purposes of the present work, tree based and Least Absolute Shrinkage and Selection Operator (LASSO) regression methods were used for both variable selection and for evaluating and comparing the models’ predictive abilities, which will be described in more detail in the following sections.

**Decision Trees**

Decision trees and random forests offer several advantages that warrant an in-depth investigation of their utility in covariate selection and predictive accuracy. Decision trees are easy to interpret graphically, robust to outliers, evaluate all possible outcomes of the decision, can automatically detect nonlinearities and interactions between covariates and are nonparametric. Their limitations - mainly an increased risk of overfitting with a single tree and model instability due to small changes in the data (e.g., removing a variable) - are partly addressed by the random forest [94, 95]. To understand random forests, one must first fully comprehend the methodology of decision trees. **Figure 2** illustrates a graphical representation of
Figure 2 Example Illustration of A Decision Tree To Classify The Type of Vehicle to Purchase

(Adapted from https://medium.com/greyatom/decision-trees-a-simple-way-to-visualize-a-decision-dc506a403aeb last accessed 05 March 2020)
an example decision tree, highlighting the terms used in this algorithm, including root, internal and leaf (or terminal) nodes. Decision trees are predictive models that perform recursive binary splitting of the data into smaller, more homogenous units with respect to the outcome. At each step, the decision tree must determine which variable to use and at which value of that variable to split the node, when to continue splitting or stop, and the final classification (or regression) prediction at each terminal node. Several factors are involved in each of these determinations, as well as for the subsequent growth (i.e., increasing the number of internal nodes) or pruning of the tree (i.e., removing internal nodes that are likely “noise” and do not improve model predictiveness of validation datasets). Pruning via cost-complexity will be explained in detail in the following section.

**Decision Tree Model Development**

Decision trees learn from the input data, known as the training dataset. Typically, a portion of the training dataset, known as the test dataset, is set aside and not used in the model building process and used only as an evaluation of final model’s predictive performance. The construction of a decision tree depends on partitioning (or splitting) the training data into smaller, homogenous subsets based on the splits of the covariates that increase the homogeneity of the outcome at each step of the tree-building process. In the example shown in Figure 2, an example of a split would be the individual’s marriage status, where being married leads to a decision of a minivan for car purchase, and not being married leads to a decision of a sports car. In this example, marriage status is the variable (among men over 30 years old) that led to the best separation (or increase in homogeneity) between car types in the training data in that step of the tree building process.
As such, a decision tree produces a flowchart-like structure where a series of binary Yes/No “questions” (if-then structure) are developed from each covariate and the impurity of the resulting node is calculated. Impurity is defined as the degree of heterogeneity of the outcome variable with each binary split of the covariate. For example, if the dataset that generated the decision tree in Figure 1 included a range of ages 20-50 years old, every possible split would be tested, such that “<21 years old?”,”<22 years old?”,”<23 years old?” and so on, would be tested. The impurity of the resulting nodes of each split are compared to the impurity produced from a similar test of all covariates, and the covariate that generates the largest decrease in impurity is chosen as the root node, which was “<30 years old?” in Figure 2.

Several methods of quantifying this impurity have been characterized. For classification outcomes, the Gini index or entropy are commonly used [96]. For continuous outcomes, reduction of variance, such as that measured by the residual sum of squares, is used. The equations for Gini index and Entropy are shown in Equations 1 and 2, respectively, where \( p_i \) refers to the proportion of cases among samples belonging to class \( i \) of \( m \) classes. Because of the added computational complexity of entropy, the Gini index is preferred. Values of 0 indicate perfect classification.

Equation 1: \[ 1 - \sum_{i=1}^{m} p_i^2 \]

Equation 2: \[ -\sum_{i=1}^{m} p_i \log p_i \]

After the root node is chosen, the variable selection procedure is repeated for the newly generated subsets of data, hence the recursive nature of decision tree algorithms. If the Gini index is 0 for one of the nodes, that becomes the terminal node – the node with the final prediction classification – and a prediction is made for that branch. If impurity still exists in the
resulting node(s), the same procedure (i.e., test every covariate then calculate the impurity) is performed and the covariate split that produces a node with the lowest Gini index (or conversely, the highest Information Gain) is chosen for that smaller subset of the data. This procedure continues indefinitely until the decision tree is 100% accurate on the training data, often resulting in trees with hundreds of branches and nodes. This inevitably will not translate to high accuracy of new data presented to the tree, as it has perfectly learned only the training data and specifically all of the noise associated with that dataset [97]. Several approaches can be taken to address this over-fitting limitation of decision trees.

To deal with overfitting, one can set stopping criteria for the decision tree, such as the minimum number of samples per node, the maximum depth (i.e., the length of the longest path from the root node to a terminal node) that tree can grow to, or the maximum size (i.e., the number of nodes in the tree). Cost-complexity pruning and cross-validation are also critical tools in optimal model development and performance. This process repeats with each fold appearing in the validation dataset once. An evaluation of the model metrics, such as accuracy for classification or Root Mean Square Error (RMSE) for continuous variables, informs the analyst about the generalizability of the model to unseen data. With cost-complexity pruning, a regularization parameter is added to the cost function (e.g., the residual sum of squares) that penalizes the tree according to the number of terminal nodes of the tree – larger trees incur a larger penalty - and a regularization parameter value of 0 is equivalent to the full tree. Similar to LASSO methods that will be expanded upon in a later section, the penalty parameter ($\alpha$) must be tuned via cross-validation to obtain the tree with the optimal number of branches that minimize the cross-validation error. [98]. Cross-validation is an approach where the training data is split into $k$-folds (e.g., $k=5$ or 10). The model is trained on $k-1$ folds and the remaining dataset serves
as the validation dataset on which the model obtained from the reduced training data is applied. For each training fold of the cross-validation procedure, first the full tree is grown ($\alpha=0$). Next, various values of $\alpha$ are tested, where higher values lead to smaller trees and a very high value of $\alpha$ results in only the root node. Each tree generated from the various $\alpha$ values is tested on the validation dataset. This process is repeated with each training fold of the cross-validation, and the $\alpha$ value that produces the lowest sum of squares (or highest accuracy in classification) is the $\alpha$ value used for the original tree.

**Random Forests**

Despite cross-validation and cost-complexity pruning, a single decision tree can rarely provide consistently accurate predictions of future data, especially if the distribution of covariates in these data are different than the training model. Ensemble methods of decision trees, namely random forests, are well-studied algorithms and have been shown to have high predictive accuracy, but at the expense of ease of interpretability [99]. Two fundamental concepts define this algorithm, bagging and random feature selection. Bagging, which is short for bootstrap aggregation, is the process by which a different training dataset is selected for each tree. The training data undergoes bootstrap with replacement, such that the same sample can appear twice, and the original sample size is retained. The remaining samples become the out-of-bag sample and the model is evaluated with these samples. The final prediction of each sample is generated based on a “majority vote” of the classification generated by each decision tree in the random forest where that particular sample was not used (i.e., the sample was an out-of-bag sample). For regression problems, the mean value of all the out-of-bag predictions is used for the final prediction.
In this way, bagging eliminates the need for cross-validation of random forests since the out-of-bag samples similarly represent the $k$-folds. Nonetheless, cross-validation can still take place and cross-validation error can be compared with out-of-bag error estimates. The second hallmark of random forests is that at each node of a tree, only a random subset of features is selected to evaluate the optimal split at that node. For example, if there are 16 covariates tested, each node of the tree will only evaluate a subset of these covariates (usually the square root of the total number of covariates) to produce the node. The subsequent node will then evaluate a new set of random covariates to grow the tree, and so on. A diagram illustrating bagging and random feature selection of the random forest is shown in Figure 3.

The result of bootstrap with replacement and random feature selection within the random forest ensure that each tree is decorrelated, which strengthens the overall prediction when the outcomes of the ensemble of trees are taken together. The degree of correlation between each tree is a strong indicator of the predictive performance of the random forest [99-101]. It is also important to note that no pruning is performed on each tree, again to ensure decorrelated trees. Therefore, each tree is fully grown and evaluated on the out-of-bag samples. The average prediction of each out-of-bag sample for continuous outcomes, or the majority vote for classification tasks, becomes the final prediction. Thus, while each individual tree fits the training data with 100% accuracy, which is undesirable as previously described with a single decision tree, the average of many trees provide more accurate predictions.

Another important advantage of random forests is the identification of variable importance rankings [101]. Variable importance rankings make use of the out-of-bag samples. First, the total number of correct classifications of the out-of-bag samples of the original model are calculated for every tree in the random forest. Next, one covariate is randomly permuted, and
Figure 3 Example Illustration of Bagging and Random Feature Selection in the Random Forest Algorithm (Adapted from https://nititek.wordpress.com/2013/12/10/bootstrapping/ last accessed 11 May 2020)
the out-of-bag samples are then re-run down the perturbed tree. The number of correct
classifications from the permuted out-of-bag samples are then subtracted from the original out-
of-bag data. The average of this number over all trees in the forest is the importance score for
that particular variable. Low importance scores imply that permuting the variable did not result
in substantial differences in the number of correct classifications and that variable can therefore
be excluded from future analyses for reasons such as reducing model complexity. This process is
repeated for every covariate to rank the variables by their predictive power of the target outcome.

**Least Absolute Shrinkage and Selection Operator Regression**

Another useful model for predictions and variable selection is LASSO regression,
particularly when collinearity exists between variables or there is a large number of predictors
relative to the sample size [102]. LASSO is a regularization technique in which the residual sum
of squares (RSS) is minimized as an Ordinary Least Squares regression would, but is also subject
to a regularization parameter, as shown in Equation 3. While simultaneously minimizing the
RSS, large β coefficients push the magnitude of the loss function in the other direction. Because
of this, the regularization parameter, which can be viewed as a parameter responsible for
determining the weight of the penalty term, penalizes parameter coefficients for being too large
by shrinking their estimate towards zero. The LASSO then reduces the likelihood of over-fitting
and also performs covariate selection, as some of the coefficients are shrunk to exactly zero.

\[
\frac{\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2}{RSS} + \lambda \sum_{j=1}^{p} |\beta_j| \quad \text{(penalty term)}
\]

**Equation 3:**

When the regularization parameter is zero, the problem becomes an Ordinary Least
Squares regression and as it approaches infinity, all covariate coefficients become zero. Similar
to the cost-complexity tuning parameter described above with reducing the size of decision trees,
the value of the regularization parameter in LASSO regression must also be optimized by cross-validation. Choosing various values of the parameter and testing the resultant model on the validation folds can help assess which model is best and the number of optimal covariates. Thus, LASSO reduces the risk of over-fitting by reducing model complexity and concurrently provides variable selection and the identification of influential covariates in response predictions. The premise behind LASSO is that regression models with many covariates tend to result in coefficients that are much larger than the true coefficients, hence over-fitting the model. By shrinking the parameter coefficients, predicting test or unseen data is improved at the expense of an increase of unsubstantial bias in the training data.

Pharmacometric Modeling

Pharmacokinetics (PK) is the quantitative study of the processes that govern the time course of drug concentrations in the body. These processes are divided into absorption, distribution, metabolism and excretion. Pharmacodynamics (PD) is the investigation of drug effects and response. This dissertation research focused on describing the PK of plasma tenofovir and the accumulation of TFV-DP served as a proxy for the PD “response” in which the PK/PD model was developed. PK/PD modeling relies on the use of nonlinear mixed effects modeling [103]. Mixed effects modeling allows for the estimation of between- and within-subject variability and the exploration of covariates that can decrease between-subject variability. Indeed, if between-subject variability is large, conclusions made from the model are not reliable and cannot be generalized to the target population. For example, a drug with a narrow therapeutic index but with large between-subject variability is not ideal, as individuals can widely vary in the toxic or sub-therapeutic range due to the inherent variability. Therefore, it is critical to use population PK/PD modeling to explore sources and correlates of variability within
the various PK parameters. This was also the goal of this dissertation work. An illustration of PK modeling development and application will also be demonstrated in Appendix A through the development of a population model of the antiretroviral drug etravirine in children under six years old.

Traditional covariate search methods in population PK modeling involve forward/backward stepwise selection based on statistical inference [104]. However, the large number of covariates explored in this work requires novel methods of covariate identification and selection. In addition, studies have reported significant selection bias with smaller datasets (<50 subjects) and concluded that significance-based covariate selection should not be used for smaller datasets. The focus of this work centers on integrating the machine learning algorithms described in this section to identify variable importance rankings as they relate to each PK parameter. These machine learning models can then be optimized to predict future data, or the identified variables can be incorporated into the PK model. This dissertation research will identify novel predictors of F/TDF pharmacokinetic processes to reduce the between-subject variability observed in the data. The final covariates incorporated in the model will be incorporated into simulations to provide tailored adherence interpretations, thereby allowing for more precise exploration of adherence-efficacy relationships for HIV PrEP. One of the lingering challenges in population PK modeling is handling data that are below the limit of quantification (BLQ). This was the first aim of the thesis to summarize current knowledge of how to handle BLQ data and to provide comprehensive recommendations regarding different methods to incorporate BLQ data in pharmacometric models.
CHAPTER III

HANDLING DATA BELOW THE LIMIT OF QUANTIFICATION IN POPULATION PK MODELING

Background

A commonly encountered problem during pharmacokinetic data analysis is concentration measurements that are below the limit of quantification (BLQ). In 2001, Beal described seven methods for handling BLQ observations, listed below in Table 1, which have since been widely adopted as standards of practice in pharmacometric modeling [105]. Despite widespread acceptance of the M3 likelihood-based method, which treats BLQ data as censored and maximizes the likelihood that a concentration observed to be BLQ is also predicted to be BLQ, as generating the least biased results, less information is available about when it is critical to use M3 (i.e. at which percentage of observations being recorded as BLQ requires this approach), the impact of both the number of samples per subject or overall number of subjects in the dataset, the impact of BLQ observations on parameter estimation for compounds that exhibit nonlinear PK, and the impact of BLQ observations on structural model selection (e.g. 2 compartment vs 1 compartment). In addition, the influence of BLQ data on structural model features, such as the relative magnitudes of clearance and volume, has not been fully explored.

Several publications have examined various approaches to handling BLQ values (Table 1) including deleting all BLQ observations (M1), the YLO approach (M2) (where BLQ data is excluded and a maximum likelihood conditional estimation method is applied to the remaining observations, conditioned on the fact that all observations are above the lower limit of quantification), setting all BLQ observations to 0 (M7), setting all BLQ observations to \( \frac{1}{2} \) the lower limit of quantification (LLOQ/2) (M5), setting the first BLQ
Table 1 Summary of Methods to Handle Data Below the Limit of Quantification in Population

PK Modeling

<table>
<thead>
<tr>
<th></th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>M1</td>
<td>Ignore missing values</td>
</tr>
<tr>
<td>M2</td>
<td>Exclude BLQ data and a maximum likelihood conditional estimation method is applied to the remaining observations, conditioned on the fact that all observations are above the lower limit of quantification</td>
</tr>
<tr>
<td>M3</td>
<td>BLQ data treated as censored and maximizes the likelihood that a concentration observed to be BLQ is also predicted to be BLQ</td>
</tr>
<tr>
<td>M4</td>
<td>Similar to M3 but conditioned on the assumption that concentrations are ≥0</td>
</tr>
<tr>
<td>M5</td>
<td>Replace all BLQ with LLOQ/2</td>
</tr>
<tr>
<td>M6</td>
<td>Replace first BLQ with LLOQ/2, ignore the rest</td>
</tr>
<tr>
<td>M7</td>
<td>Replace all BLQ with zero</td>
</tr>
</tbody>
</table>
observation in a subject to LLOQ/2 and removing the rest (M6), and likelihood-based use of M3 or M4 methods; these results are summarized in Table 2 [106-111]. Differences are evident between these published studies regarding the point at which ignoring/deleting BLQ observations was found to cause bias and impact parameter precision, which can be partly attributed to methodological differences, such as differing percentages of BLQ tested, use of the M5 vs. M6 method for incorporating LLOQ/2, approaches to simulating datasets and statistical methods used to evaluate the impact of the various methods employed for handling BLQ observations. These dissimilarities can prevent generalizable interpretability. We conducted a comprehensive, consistent analysis of the main BLQ data handling methods (i.e. M1, M3, M5 and M6 methods) to summarize current knowledge and to address apparent discrepancies between previously published studies.

Methods of BLQ Simulation and Parameter Estimation

A population of 400 individuals was created. Using the 400-subject dataset, four dataset sizes were generated: 200, 100, 50, and 25 subject datasets. These five datasets were then used to simulate pharmacokinetic profiles.

Pharmacokinetic profiles were simulated following a single intravenous (IV) bolus or an oral dose (50 mg) using a one-compartment, two-compartment or nonlinear model with first order elimination using NONMEM 7.3. Inter-individual variability was set at 30% for clearance and volume, with covariance of 5% using a 2-block omega. A log transform both sides proportional error model was used with residual variability set to 20%. Clearance and volume values were used to simulate four different pharmacokinetic scenarios; slow clearance/small volume, slow clearance/large volume, rapid clearance/small volume, and rapid clearance/large volume. An overview of the simulations and analyses are presented in Figure 4.
<table>
<thead>
<tr>
<th>Title</th>
<th>Methods</th>
<th>Conclusions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hing et al, 2001 [109] NONMEM V</td>
<td>One-compartment Rat study with one sample per animal</td>
<td>M1 and M7 led to biased clearance (CL) and inter-individual variability (IIV) estimates</td>
</tr>
<tr>
<td></td>
<td>Methods Tested: M1, M7 and 4 substitution methods</td>
<td>Loss of precision for all methods occurred at BLQ &gt;25%</td>
</tr>
<tr>
<td>Keizer et al, 2015 [110] NONMEM VI</td>
<td>IV one-compartment</td>
<td>IV one-compartment: no bias was observed for any of the methods when BLQ &lt;20%; considerable bias occurs with M1 and M6 at 40%; IIV was comparable at 10% and 20%, except for M3, which showed higher variation</td>
</tr>
<tr>
<td></td>
<td>IV two-compartment</td>
<td>IV two-compartment: bias in fixed parameters was observed after M1 and M6 at 10% BLQ, except for the estimation of V and IIV in V (central volume)</td>
</tr>
<tr>
<td></td>
<td>Oral one-compartment</td>
<td>One-compartment oral model: At &lt;20% BLQ, all methods except M1 provided reasonable and broadly similar performance for both fixed effects and IIV</td>
</tr>
<tr>
<td></td>
<td>BLQ: 10%, 20%, 40%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Methods Tested: M1, M6, and M3</td>
<td></td>
</tr>
<tr>
<td>Xu et al, 2011 [111] NONMEM VI</td>
<td>One-compartment</td>
<td>One-compartment: The impact of ignoring BLQ &lt;10% was minimal</td>
</tr>
<tr>
<td></td>
<td>Two-compartment</td>
<td>Two-compartment: When BLQ &lt;5%, M1 did not create bias in the fixed-effect parameters, whereas more pronounced bias in the estimates IIV was observed. The greatest impact was on Vp</td>
</tr>
<tr>
<td></td>
<td>BLQ: 1%, 2.5%, 7%, 10%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Methods Tested: M1, M3</td>
<td></td>
</tr>
</tbody>
</table>
### Table 2 Cont’d Summary of published literature evaluating methods for handling concentration data below the limit of quantification (BLQ)

<table>
<thead>
<tr>
<th>Authors</th>
<th>Model A:</th>
<th>Model B:</th>
<th>Model A:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bergstrand et al, 2009 [107]</td>
<td>one-compartment, transit compartments, BLQ in absorption phase</td>
<td>two-compartment</td>
<td>CL, Vc, and IIV on CL and Vc were not biased by presence of BLQ samples and similar for each method. Ka, mean transit time and number of transit compartments were biased with M1</td>
</tr>
<tr>
<td>NONMEM VI</td>
<td>BLQ: 10-30%</td>
<td></td>
<td>M3 generated the best performance</td>
</tr>
<tr>
<td></td>
<td>Methods Tested: M1, M2, M3</td>
<td></td>
<td>Model B: M1 led to substantial bias in CL, Q, and Vp. The M3 method was the least biased</td>
</tr>
<tr>
<td>Ahn et al, 2008 [106]</td>
<td>Two-compartment with first-order absorption</td>
<td></td>
<td>M3 and M4 produced similar results without log transformation</td>
</tr>
<tr>
<td>NONMEM VI</td>
<td>BLQ: 10-40%</td>
<td></td>
<td>Parameter estimates were biased with M1, especially when BLQ was 40%</td>
</tr>
<tr>
<td></td>
<td>Methods Tested: M1, M2, M3, M4</td>
<td></td>
<td>Clearance was more negatively biased as %BLQ increased. Vp and Q were more positively biased</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>The most accurate and precise estimates were obtained with M3</td>
</tr>
<tr>
<td>Duval et al, 2002 [108]</td>
<td>Two-compartment</td>
<td></td>
<td>A bias on CL of &gt;20% was observed with M1 at BLQ ≥20%</td>
</tr>
<tr>
<td>NONMEM VI</td>
<td>Datasets based on: (a) the ratio of the area under the curve (AUC) of the distribution phase to the total AUC and (b) the ratio of the half-life of the distribution phase to the half-life of the elimination phase</td>
<td></td>
<td>No major trends were observed for Vc and Q between M1 and M6 substitution</td>
</tr>
<tr>
<td></td>
<td>BLQ: 5-50%</td>
<td></td>
<td>IIV on CL is improved with M6, while loss of information on IIV was observed for all other parameters, regardless of method</td>
</tr>
<tr>
<td></td>
<td>Methods Tested: M1 and M6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 4 Simulation and analysis of BLQ data with single dose (SD) datasets from one-compartment (1c) models with intravenous or oral (O1) dosing, two-compartment (2c) or nonlinear models (NL). Estimation was completed with rich (R1) or sparse (S1) sampling designs. Methods for handling BLQ data included the M1, M3, M5 and M6 methods. The ratio of BLQ samples in each scenario was 0.05, 0.1, 0.2, 0.3, and 0.5.
Rich and sparse schemes were used to simulate pharmacokinetic profiles for the dosing scenarios. Simulations were performed to evaluate the impact of BLQ observations on IV one-compartment, oral one-compartment, IV two-compartment and IV nonlinear models. In each model, the percent of observations that were BLQ in each dataset was 5%, 10%, 20%, 30% or 50%. These datasets were generated by removing the tail end of the dataset, consequently increasing the LLOQ with each BLQ percentage increase.

The methods tested in each scenario included deleting observations (M1), setting all BLQ observations to ½ the lower limit of quantification (LLOQ/2) (M5), setting the first BLQ observation in a subject to LLOQ/2 and removing the rest (M6), and the M3 likelihood method.

**Results of BLQ Simulation and Parameter Estimation**

A flow diagram summarizing the results of BLQ simulations are presented in Figure 5. Across multiple scenarios, common BLQ data handling methods were compared. In the simplest case - a densely sampled (5 samples/subject), IV one compartment model with CL=20 L/hr and V=70 L - the M3 method did not provide any added benefit for accurately estimating CL and its associated population parameter variability (PPV) for CL (PPVCL) as compared to M1. However, between 30% and 50% BLQ observations, M3 achieved less biased estimates of V and PPV for V (PPVV). When V remained unchanged (V=70 L) and clearance was reduced (decreased from CL=20 to CL=10 L/hr), the M3 method resulted in less biased PPVV and PPVCL estimates at 30% or more BLQ. No differences were detected for estimates of CL and V between M3 and M1 methods, between 5% and up to 30% BLQ. At more than 30% BLQ, the M3 produced significantly less biased results compared to M1.

The M5 method produced biased results at scenarios 10-50% BLQ. With M6, bias on CL and PPVCL was similar to M1 and M3 up to 20% BLQ, after which significant bias was
Figure 5 Summary Decision Flow Diagram of Handling BLQ Data
observed. Similar trends were observed for V. A previous evaluation of M6 method for a one-compartment model showed considerable bias and imprecision on CL at 40% BLQ, while no systematic bias was observed at 10% and 20% [110]. Of note, 30% BLQ was not assessed in this study. An additional study reported unpredictable results when using M6 for BLQ observations in the elimination phase and M5 for BLQ observations in the absorption phase, sometimes inflating or sometimes reducing the bias [107]. Nonetheless, the M3 method was determined to be the best overall method. The size of the dataset did not appear to influence the results, as similar trends were observed for dataset sizes of 400, 100, 50 and 25 subjects.

After simulations of an IV one-compartment model with sparse sampling (2 samples/subject), bias was observed for V between 20-50% BLQ for M1, while bias on CL was similar between M1 and M3 for all BLQ percentages. However, in smaller datasets (i.e. n=25 subjects), bias on clearance was apparent between 20-50% BLQ with M1. These results are in agreement with previous assessments of one-compartment models [109-111]. Collectively, these data indicate for a simple IV, one-compartment model, either the M1 or the M3 methods are approximately equivalent when up to 30% of the observations are BLQ. However, the M3 method is superior when the portion of BLQ observations is >30%. Furthermore, use of the M6 method generated biased parameter estimates of CL or PPVCL at BLQ percentages as low as 20% and as low as 10% for M5, suggesting this data handling method should not be used for datasets with moderate to large percentages of BLQ.

The one-compartment extravascular dosing scenario was similar to the IV dosing scenario. Accuracy of CL estimates was similar between the M3 and M1 up to 30% BLQ. At 50% BLQ however, M3 produce less biased CL estimates. Not surprisingly, the time at which the BLQ observations occurred was important. When the percent of BLQ in the absorption phase
was 5%, M1 generated similar results to M3. When the percentage of BLQ in the absorption phase was ≥10%, the use of M3 was superior to M1. Using RMSE as the primary criteria for comparison, M3 is recommended for less biased estimates of V, PPVV and PPVCL regardless of overall percentage of BLQ. Using nRMSE criteria (normalized RMSE values to 0%), when the percent of BLQ was 5% or greater in the absorption phase and the overall BLQ percentage was ≥20%, M3 generated the least biased results of the other evaluated methods. Previous oral one-compartment studies suggest that excluding observations produces bias on absorption parameters (e.g., mean transit time, number of transit compartments) at >30% BLQ. Similar to our findings, Bergstrand and colleagues also observed accuracy and precision of CL estimates were comparable between M3 and M1, up to 30% [107].

In an IV two-compartment model with CL=70 L/hr, bias in CL and V were similar between M1 and M3 at BLQ percentages up to 30%. However, estimates for peripheral compartments (V2 and Q) were more sensitive to missing data. Using RMSE, M3 yielded less bias in these parameters at ≥5% BLQ. Using nRMSE, M3 was superior for V2 and Q when BLQ percentage was 30% or more. PPVCL was less biased when using M3 when BLQ was 20% or more. When simulating slower clearance values (CL=20 L/hr), results were similar except for 50% BLQ, in which the M3 method generated less biased estimates for CL than the other methods evaluated. The size of the dataset had no influence on these results. Previous evaluations of two-compartment models have observed bias with M1 and M6 with BLQ percentages as low as 5% or 10% [106-108, 110, 111].

Finally, nonlinear PK models were evaluated. M3 was superior at reducing bias on Vmax and Km at all BLQ percentages. Volume estimates were stable throughout all percentages between M1 and M3, while M5 and M6 resulted in biased parameters at ≥5%. These results
suggest M3 should be used for achieving less biased estimates of $V_{\text{max}}$, $K_m$ and PPV for $V_{\text{max}}$ and PPV for $V$ at BLQ 5% or more.

In summary, selection of an appropriate method to handle BLQ data is an important consideration in PK model development. Based on previous studies and our current analysis, when developing a simple IV one compartment model with rich sampling, the M3 method should be used in cases with $\geq 30\%$ BLQ observations. Otherwise, excluding observations when BLQ percentage is $<30\%$ may not excessively bias CL and $V$ estimates, though the M3 method would still be tested. Smaller datasets (less than 50 subjects) with sparse sampling should be implemented with the M3 method at approximately $\geq 20\%$ BLQ observations. In an oral one-compartment model, the percentage of BLQ in the absorption phase is a determinant of using M3 or excluding observations. If the percentage of BLQ in the absorption phase is $>5\%$, M3 should be implemented for less biased absorption parameters. Nonetheless, bias on CL is stable through 30% BLQ. When using a two-compartment model, CL and $V$ tend to have similar bias up to 30% BLQ, regardless of the method used. However, ignoring all BLQ observations creates bias in $V_2$ and $Q$ parameters, and the M3 method would therefore be warranted for BLQ percentages greater than 5%. Finally, the M3 method is necessary for less biased estimates of $V_{\text{max}}$ and $K_m$ in nonlinear models when BLQ values are present. Most studies report bias when replacing all BLQ observations with LLLOQ/2. If this method is implemented, replacing the first observation with LLOQ/2 and discarding the remaining ones generates less bias than replacing all BLQ observations with LLOQ/2.

Model misspecification is an additional concern when handling BLQ observations. A previous investigation of a one-compartment IV model misspecification reported steadily increasing Type I error rates (lower OFV for two-compartment models compared to one-
compartment at the alpha 0.05 and 0.01 levels) using M1 as the BLQ percentage increased. For example, when approximately 50% of the data was BLQ, 96% of 500 simulations had lower OFV for two-compartment models. When the M2 method was implemented, this percentage significantly decreased to nominal levels, leading to recommendations of M2 implementation when BLQ >10% [112]. For nonlinear models, 25% of the simulations had lower OFV when modeled as linear elimination at 10% BLQ. At 30% BLQ, approximately 50% of the simulations would be incorrectly classified as linear elimination. Similar to one- vs. two-compartment models, these percentages decreased to nominal levels after M3 implementation. Taken together, model misspecification and parameter estimate bias warrant the use of likelihood-based approaches to handling BLQ in most modeling scenarios.

**Summary of Recommendations for Handling BLQ Data.**

- When the percentage of BLQ data is low (5% or less), all methods (M1, M3, M5, M6) should perform similarly, regardless of the model used and regardless of sampling density.

- If the percentage of BLQ data exceeds 5%, M5 should not be used.

- M6 generates biased CL, V, PPVCL, and PPVV at BLQ percentages greater than 20% for one-compartment models with rich sampling.

- In IV one-compartment models, with 30% or less BLQ observations, the M3 method does not appear to provide a major advantage over M1 and M6 when rich sampling data is available.

- For IV two-compartment models with rich sampling, the M3 method should be implemented at 5% or more BLQ for less biased estimates of V2 and Q. CL and V tend to have similar bias up to 30% BLQ between M3 and M1. M6 generates biased results.
For drugs that exhibit nonlinear PK, the M3 method should be implemented at 5% or more BLQ.

**Application of BLQ Results**

After development of these BLQ recommendations, they were implemented in a bioequivalence analysis of F/TDF, co-encapsulated with an ingestible sensor. The total percent of BLQ samples was 22%, which as the previous recommendations suggest, do not influence PK parameters when using the M1, M3 or M6 methods. M1 and M6 were both tested in this analysis, with no difference between results. The details for this analysis follow.

**Bioequivalence of Tenofovir And Emtricitabine After Coencapsulation With the Proteus Ingestible Sensor**

In 2012, the U.S. Food and Drug Administration (FDA) approved daily oral tenofovir disoproxil fumarate plus emtricitabine (TDF/FTC, Truvada®) for HIV pre-exposure prophylaxis (PrEP), primarily based on the results of two large clinical trials. The iPrEx and Partners PrEP studies [37, 42] reported 44% (95% confidence interval [CI], 15%–63%) and 75% (95% CI, 55%–87%) HIV risk reduction with daily TDF/FTC relative to placebo. Both studies showed that quantifiable drug concentrations in plasma were strongly associated with prophylactic effect. For example, in the iPrEx trial, the HIV risk reduction increased from 44% to 92% in those with quantifiable plasma drug concentrations. Low adherence rates (21%–37%) were later discovered in the VOICE and FEM-PrEP trials, which failed to demonstrate efficacy [113]. Adherence, therefore, plays a key role in the efficacy of TDF/FTC for PrEP.

Several subjective and objective measures exist for assessing adherence, such as self-report, pill counts, clinician assessment, pharmacy refill records, and drug concentration measurements. Although these methods have been long-standing, several drawbacks limit their
use, including lack of accuracy and precision, expense, “white coat” effects, and difficult implementation into clinical practice.

To address some of these challenges, a unique approach to objectively and rapidly measure drug ingestion has been developed using Proteus Discover (Proteus Digital Health, Inc.). This digital medicine program consists of an FDA-approved ingestible sensor pill that transmits a small electrical signal after interaction with gastric fluid, an adhesive sensor patch—worn on the torso and replaced weekly—that detects the signal from the ingestible sensor, and an application on a smartphone or tablet device that captures all transmitted data. In this way, time-stamped and longitudinal records of drug adherence can be measured in a precise manner. The ingestible sensor pills can be coencapsulated with medications to create digital medicines. The sensor is the size of a grain of sand and contains minute quantities of silicon, copper, and magnesium that pass through the body naturally. Several studies have found that coencapsulated versions of medicine had similar dissolution or pharmacokinetics profiles as the unencapsulated versions. Abilify® MyCite, a fully integrated digital medicine with the ingestible sensor inside of the drug tablet, was approved for use by the FDA, encouraging further investigation in other fields, including PrEP. Proteus Discover® has also been extensively evaluated in other therapeutic areas, including hypertension, hypercholesterolemia, diabetes, heart failure, tuberculosis, HIV, and other diseases. The objective of this study was to confirm that the rate and extent of absorption of TDF/FTC coencapsulated with Proteus Discover are unchanged relative to unencapsulated drug.

The study took place on the University of Colorado–Anschutz Medical Campus. The study protocol was approved by the Institutional Review Board and all participants signed an informed consent. The study design was a single-dose randomized (1:1) cross-over study
conducted in healthy volunteers. Key exclusion criteria included a positive HIV EIA or suspected acute HIV infection, creatinine clearance <60 mL/(min · 1.73 m²) (modification of diet in renal disease criteria for impaired kidney function), and contraindicated concomitant medications that could interfere with TDF/FTC disposition.

Participants were randomized to begin with a single dose of the unencapsulated or coencapsulated TDF/FTC. Tablets were encapsulated with Capsugel DBcaps® (Aael) and microcrystalline cellulose was used to fill empty space. TDF/FTC (300/200 mg) was administered with 250 mL of water after an overnight fast of at least 10 h. A nonstandardized meal was provided 4 h after the dose. Blood was collected at predose and at 0.25, 0.5, 1, 2, 4, 6, 10, 24, 48, and 72 h postdose. The rationale behind these sampling times was to capture the maximum concentration ($C_{\text{max}}$) of tenofovir (TFV) in plasma, and to include a total collection time exceeding 3.3 half-lives of TFV (~17 h). A 14-day washout (range 13–21 days; >7 half-lives) separated each period.

A noncompartmental analysis was carried out using a built-in function in Phoenix® WinNonlin® (Certara). Values less than the lower limit of quantitation (BLQ) were evaluated by imputing half of the lower limit of quantitation (LLOQ). Sensitivity analyses treated BLQ values as missing. Noncompartmental parameters that were estimated included $C_{\text{max}}$, area under the concentration–time curve (AUC) from time 0 to the last measured time point ($AUC_{\text{last}}$), AUC extrapolated to infinity ($AUC_{\text{inf}}$), and half-life. Linear up/log down trapezoidal methods were utilized to calculate AUC. $C_{\text{max}}$ was observed from the concentration–time curves and half-life was calculated from the elimination constant, which was derived from points in the terminal declining slope. Outcomes were log-transformed to normalize distributions for statistical analysis, then were back-transformed to the original scale. Geometric mean ratios
(GMRs) were calculated for each parameter and bioequivalence was defined as the 90% CI of each ratio being within 80%–125%. GMR comparisons for AUC\text{last}, AUC\text{inf}, and C\text{max} were modeled using a mixed effects model with treatment, treatment sequence, and period as fixed effects and patient nested in treatment sequence as the random effect, using the built-in bioequivalence function in Phoenix WinNonlin.

Twenty-four participants (11 males; 19 Caucasian, 3 African American, and 2 Hispanic) completed both visits. Mean ± SD age was 28 ± 4 years and weight was 74 ± 14 kg. From a total of 477 samples, 105 (22%) were BLQ, primarily at the 48- and 72-h time points. Geometric mean concentration vs. time plots for plasma TFV and FTC are shown in Figure 6. TFV geometric mean AUC\text{inf} (%CV) for the unencapsulated and coencapsulated formulations was 1,978 (27) and 2,042 (26) ng × h/mL, respectively. Geometric mean FTC AUC\text{inf} (%CV) estimates were 9,342 (23) and 9,512 (20) ng × h/mL for the unencapsulated and coencapsulated formulations, respectively. TFV geometric mean C\text{max} (%CV) for the unencapsulated and coencapsulated drugs was 222 (37) and 229 (32) ng/mL, respectively. For FTC, geometric mean C\text{max} for the unencapsulated and coencapsulated formulations were 1,567 (33) and 1,684 (29) ng/mL, respectively.

When imputing half the LLOQ, the 90% CIs for TFV C\text{max}, AUC\text{last}, and AUC\text{inf} ratios were 89%–119%, 94%–111%, and 96%–111%, respectively. The half-life comparison also satisfied the CI requirement (95%–108%). The 90% CIs for FTC C\text{max}, AUC\text{last}, and AUC\text{inf} were 96%–120%, 96%–108%, and 96%–108%, respectively. The 90% CI for FTC half-life was 83%–112%. CI estimates also satisfied bioequivalence requirements when BLQ results were considered missing (<5% change in CI estimates for AUC and C\text{max} calculations and a 12%
Figure 6 Geometric mean concentration vs time plots of coencapsulated (●) and unencapsulated (□) TFV (left) and FTC (right). FTC, emtricitabine; TFV, tenofovir.
increase in the upper limit of FTC half-life CI). Sequence and period effects did not have a significant impact on bioequivalence results, as assessed by the mixed effects model.

In general, TDF/FTC and the sensor tablet were safe and well tolerated. Fourteen adverse events were recorded and reviewed by the study team, approximately equal numbers after the encapsulated versus nonencapsulated formulations (11/14 Grade 1—cold symptoms, nausea, vomiting; 2/14 Grade 2—facial swelling, respiratory infection; 1/14 Grade 3—headache). The headache and facial swelling were subjective reports from two different participants. The relationship with study drugs was unclear and both resolved during the study. Other AEs were as expected. One subject became pregnant and was removed from the study before her second visit and was not included in the pharmacokinetic analysis.

Taken together, the results of this study demonstrate bioequivalence for TDF/FTC coencapsulation with the Proteus ingestible sensor pill. These results support future clinical research that seeks to over-encapsulate TDF/FTC with the Proteus ingestible sensor. As with all adherence measures, Proteus Discover has some limitations, including the added component of adherence to the patch that must be replaced weekly, and potential cost and patient privacy concerns. These considerations, along with patient and provider satisfaction, should also be assessed in future studies.

In summary, handling data that are BLQ is an important consideration in population PK modeling to estimate the most unbiased parameters. These recommendations can be applied to all population PK models, including in the development of models of plasma tenofovir and intracellular tenofovir-diphosphate.
CHAPTER IV

INDIVIDUALIZED ADHERENCE BENCHMARKS FOR HIV PRE-EXPOSURE PROPHYLAXIS

Introduction

Adherence to a prescribed medication regimen is an important consideration in both clinical trial conduct and real-world treatment plans. Inadequate adherence by clinical trial participants, such as medication discontinuation after randomization or intermittent ingestion, can compromise the results of the trial and lead to inaccurate conclusions and wasted costs [114]. It has also been estimated that 33-69% of all medication-related hospital admissions are due to poor adherence, costing $100 billion each year [115, 116]. Collecting adherence data from participants is now a routine part of clinical trials, ranging from self-report to more objective measures of drug ingestion [1]. One example of the significance of adherence on treatment outcomes is in the field of HIV pre-exposure prophylaxis (PrEP) [133].

The use of daily oral emtricitabine plus tenofovir disoproxil fumarate (F/TDF, Truvada®) has demonstrated a significant reduction in the risk of HIV acquisition, leading to the drug’s approval for HIV PrEP in 2012 [37, 39]. Nevertheless, two trials evaluating the efficacy of F/TDF for PrEP in women did not show similar risk reductions, leading to the trials’ discontinuation [40, 42]. Subsequent analyses revealed low adherence rates among these women, as assessed by the low proportion of individuals with detectable plasma tenofovir concentrations in the active arm [117]. In addition, in the iPrEx trial in men who have sex with men [37], the rate of HIV risk reduction rose from 44% to 92% in individuals with detectable plasma tenofovir concentrations. Adherence to F/TDF is therefore a critical determinant of the efficacy of PrEP,
and objective and precise analysis methods that incorporate individual characteristics to more precisely quantify adherence are needed.

Concentrations of tenofovir diphosphate (TFV-DP), the intracellular anabolite of TDF, exhibit a half-life of approximately 17 days in red blood cells [49]. This long half-life and resulting 25-fold accumulation from first dose to steady-state allow for an objective estimate of cumulative adherence over the preceding 1-2 months. The DOT-DBS study evaluated the pharmacokinetics and dose-proportionality of TFV-DP after directly observed dosing of F/TDF and established adherence gradients (<2, 2-3, 4-6, 7 doses/week) based on TFV-DP concentrations [52]. When applied to a clinical cohort, four or more doses/week, corresponding to TFV-DP concentrations greater than 700 fmol/punch, was associated with 100% (86-100%) HIV risk reduction [118]. However, a 30% coefficient of variation (CV) was observed for TFV-DP concentrations, which can lead to misclassification of adherence interpretations, decrease the precision of individualized assessments and adherence-efficacy relationships, and over- or underestimate the true adherence thresholds of each dosing classification. These thresholds were also based on inter-quartile ranges, suggesting at least a 25% misclassification rate, by definition, at each classification boundary. To address these gaps, this analysis aimed to improve the precision of current TFV-DP DBS adherence thresholds through the development of tailored TFV-DP concentration benchmarks corresponding to categorical adherence interpretations (doses/week). Emphasis was placed on clinical utility without the need for additional data collection (i.e., metabolites or genotyping) and so analyses were first conducted on clinical variables alone. Pharmacometabolomic and pharmacogenomic associations were investigated separately to explore novel biological mechanisms of variability and to evaluate their added benefit to the outcomes of the study.
Methods

Study Design

This study was a sub-analysis of a previously published prospective, pharmacokinetic, cross-over study of F/TDF (NCT02022657) [119]. Inclusion criteria included individuals between 18-50 years, without HIV, and able to comply with study-specific procedures, including directly observed dosing. Exclusion criteria included an HIV+ or HBV+ EIA test, eGFR less than 60 ml/min and medical diagnoses that alter red blood cell kinetics. IRB approval was obtained, and all participants provided informed consent before enrolling into the study.

Each participant was randomized to two of five dosing regimens of 200/300mg of F/TDF for 12 weeks each, separated by a 12-week washout period (Figure 7). Each dosing regimen was 33%, 67% or 100% of daily dosing (intermittent or holiday). 33% intermittent dosing entailed one day on-drug, two days off-drug and 67% dosing entailed two days on-drug, one day off-drug, repeated for 12 weeks. 33% holiday dosing entailed daily dosing for 1 week followed by 2 weeks off-drug, repeated for 12 weeks, and 67% holiday dosing entailed daily dosing for 2 weeks followed by 1 week off-drug, repeated for 12 weeks.

Pharmacokinetic Sampling

Blood was collected for plasma and dried blood spot (DBS) analysis at baseline and approximately every two weeks afterwards and every three weeks during the washout period. Samples were collected by convenience without regard to the time or day since the last dose.

Analytical Measurements

For DBS, 25 µl of whole blood was transferred to a Whatman 903 card. Cards were air dried for at least 3 hours or overnight in -80°C until analysis. TFV-DP was quantified from a 3mm punch using validated methodology [120].
Figure 7 DOT-DBS Study Dosing Schema
The TFV-DP in DBS assay was linear from 25-6000 fmol/sample. Plasma tenofovir was assayed within a range of 10-1,500 ng/ml and ultra-low samples within a range of 0.25-25 ng/ml.

**Pharmacokinetic Modeling**

Pharmacokinetic analysis was accomplished by nonlinear mixed effects modeling with Phoenix NLME (Version 8.2, Certara®). Two classes of models were developed and compared to describe the formation of intracellular TFV-DP: a plasma tenofovir model driving the formation of TFV-DP via first-order or saturable kinetics, and a one-compartment constant input model that mimics the accumulation of TFV-DP, given the 17-day half-life relative to short dosing intervals. Model fits were compared as described below, and the best fitting model was selected for further analyses.

Because samples were taken by convenience and there were few samples between 2-12 hours post-dose, plasma tenofovir concentrations were leveraged with data from a single-dose, 72-hour intensive PK study of F/TDF, which was described in detail in Chapter III [121]. Various structural and absorption models were tested to characterize plasma tenofovir concentrations. First-order, saturable, and hybrid first order/saturable formation (tolerance) models were assessed for modeling TFV-DP concentrations. These models were compared to a constant input model of TFV-DP, ignoring the contribution of plasma tenofovir. Significant changes in AIC/BIC, improvement of residual and standard errors, predicted vs. observed concentration plots, residual plots, model parsimony, and biological plausibility aided in selecting the best model. Visual prediction checks were used to validate the models.

**Random Forests for Covariate Selection**

Covariate selection for pharmacokinetic model parameters (clearance in the constant-input model and apparent clearance and volume of distribution in the plasma tenofovir model)
was primarily guided and justified by physiological plausibility and clinical relevance. Covariates screened included gender, age, weight, race, creatinine clearance and relevant complete blood count laboratory values, including hematocrit, mean corpuscular volume (MCV), bilirubin, platelets, red cell distribution width (RDW) and liver functions enzymes (ALT/AST). Individual post hoc estimates of each parameter of the base PK model were modelled separately using a random forest algorithm, with the intent to identify predictive covariates associated with each parameter. Model performance was assessed by the root mean squared error (RMSE) and mean absolute error (MAE) of the cross-validation procedure. The random forests were also optimized for maximal predictive performance by adjusting various input factors, such as the maximum number of decision trees in the forest and the number of covariates considered in each tree. Variable importance scores, which rank the importance of the covariates by their predictiveness of the outcome variable (i.e., each PK parameter), were generated to select physiologically and clinically relevant variables to be included in the final pharmacokinetic model.

A full covariate selection model approach was used, where all selected covariates from the random forest were entered into the pharmacokinetic model simultaneously [122]. The population estimates and variances of the PK parameter and of the covariate estimates were used to simulate a random distribution of 1,000 values, using the coveffectsplot R package designed to visualize covariate effects in pharmacometric analyses. The effects on PK parameters of extreme values of each covariate (5\textsuperscript{th} and 95\textsuperscript{th} percentile values for continuous variables), relative to the “reference” median value, were visualized with forest plots. 90\% confidence intervals were calculated and only covariates with effects that did not include one in the interval were included in the model.
Categorical Adherence Predictions

For each dosing frequency (1-7 doses/week), a simulation dataset was created with 1,000 random values of each covariate within the range of observed values. Each dosing frequency was modeled by varying the inter-dose interval. For example, one dose/week corresponded to one dose every 168 hours, two doses/week as one dose every 84 hours, and so on.

Simulated concentrations at Week 12 (steady-state) were then grouped into 1 (<2), 2-3, 4-5 and 6-7 doses/week. A cross-validated decision tree classified dosing frequency regimens based on Week 12 concentrations and optimal splits of covariate values. The depth of the decision tree was optimized by cost-complexity pruning [123]. Decision trees were selected for classification instead of random forests, which are a collection of decision trees, because of the superior interpretability of decision trees compared to the “black-box” output of random forests. Receiver operating characteristic (ROC) analyses, a test commonly used to assess the performance of a binary classifier, were conducted to compare the results of the decision tree model at each classification boundary.

Expected Steady-State Predictions

Random forest and LASSO regression models were developed to evaluate their ability to predict an individual’s expected steady-state concentration prior to drug administration, a secondary outcome of the study designed as an alternative approach to support the predictive findings obtained from pharmacokinetic modeling.

Data from both dosing regimens were used for this analysis, in that each individual contributed two steady-state observations to the analysis. Dose proportionality was confirmed (β (90% CI): 1.01 (0.92-1.1)) in the DOT-DBS study and so 33% and 67% steady-state TFV-DP
concentrations were dose-adjusted to estimate 100% adherent TFV-DP concentrations (e.g., steady-state TFV-DP concentrations for 33% dosing were multiplied by 3).

Seventy percent of the TFV-DP concentrations were used for model training and cross-validation and 30% for test data, which was not incorporated in the model development process and only used as a final evaluation of model performance. Concentrations from the same individual were not split between the training and test dataset, which would have generated overly optimistic results since the model would have already been trained on that particular individual.

A random forest was performed on the training data as described above, but also taking into account the repeated measures nature of the data [124, 125]. The same individual was not included in each fold of the cross-validation procedure, again to prevent overly optimistic results. The final model was evaluated on the test dataset and RMSE, MAE and percent error scores were calculated between predicted and observed concentrations. To reduce model complexity, only the top 5 predictors (based on the variable importance scores) from the full random forest model were incorporated in a separate model and RMSE/MAE scores between the full and reduced models were compared. Results were also compared to a linear model using LASSO regression.

Results

Forty-eight healthy participants completed the DOT-DBS study. Baseline characteristics of the 48 participants are shown in Table 3. There were 25 females and 10 African American participants, two of whom were also Hispanic. Median (range) age was 29 (21-49) years and weight was 77 (51-155) kg. Median eGFR was 98 (69-156) ml/min. Key CBC lab values included a hematocrit of 43% (35-49%) and platelet count of 247 (153-430) *10^9 per liter.
Table 3 Baseline Characteristics of DOT-DBS study participants

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Overall (n=48)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>23 (48%)</td>
</tr>
<tr>
<td>Female</td>
<td>25 (52%)</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>26 (54%)</td>
</tr>
<tr>
<td>Black</td>
<td>10 (17%)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>14 (29%)(^b)</td>
</tr>
<tr>
<td><strong>Age (yr)</strong></td>
<td>29 (21-49)</td>
</tr>
<tr>
<td><strong>Weight (kg)</strong></td>
<td>77 (51-155)</td>
</tr>
<tr>
<td><strong>BMI (kg/m(^2))</strong></td>
<td>26 (17-54)</td>
</tr>
<tr>
<td><strong>eGFR (ml/min)</strong></td>
<td>98 (69-156)</td>
</tr>
<tr>
<td><strong>Hematocrit (%)</strong></td>
<td>43 (35-49)</td>
</tr>
<tr>
<td>*<em>Platelet Count (<em>10(^9) per Liter)</em></em></td>
<td>247 (153-430)</td>
</tr>
<tr>
<td><strong>Total Bilirubin (mg/dL)</strong></td>
<td>0.5 (0.2-1.7)</td>
</tr>
</tbody>
</table>

\(^a\)Results are reported as median (range) or number (percent)

\(^b\)Two Hispanic individuals were black
Population Pharmacokinetics

In total, 716 samples each of plasma tenofovir and TFV-DP concentrations were available for analysis, including samples from a single-dose, intensive PK study used to leverage plasma data. Time post-dose ranged from 3 minutes to ~28 days.

Plasma tenofovir was best described by a three-compartment model with transit absorption, linked to the formation of TFV-DP via a first-order rate constant. Nonetheless, a less complex one-compartment constant input model – without the influence of plasma tenofovir - demonstrated similar model fits and error estimates of TFV-DP and was therefore chosen for further analyses. Random forests indicated weight, platelet count, hematocrit, bilirubin, eGFR, race (Black vs. non-Black) and gender as potential covariates for clearance, the parameter responsible for determining steady-state concentrations. Gender was associated with both weight (91 ± 24 versus 71 ± 18 kg, P=0.003 males versus females) and hematocrit (45% ± 2 versus 40% ± 3, P<0.0001 males versus females) and was considered redundant to the effects of these two variables and therefore excluded from the full covariate model. Simultaneous inclusion of weight and gender in the constant input model diminished the effect of weight by 33% and resulted in lower precision and non-significance of both estimates, supporting the consequences of including closely related variables in the same model [126]. Gender was therefore evaluated by univariate analysis.

The effects of extreme values (5th and 95th percentiles of continuous covariates) of each covariate, relative to the median clearance value, are shown in Figure 8. The point estimates of the effects of low and high weight values were 14% and 19%, while the effects of low and high platelet counts were 8% and 15%. The distribution of the effects of hematocrit and bilirubin
Figure 8 Effects of 5th and 95th Percentile and Categorical Covariate Values on Constant Input Clearance. Gender effects were evaluated in a univariate model.
included one - further supported by low precision of the parameter estimates – and were excluded from further analysis. The addition of eGFR to a model with weight and platelet count did not improve model fit or further reduce inter-individual variability (<1% change) and was also removed for reasons of model parsimony. A low number of black participants and a wide distribution of the effect of race (Figure 8) also precluded the ability to precisely identify its influence on clearance.

Clearance increased by 4% (95% CI 2-6%) for every 10kg increase in weight, while an increase of $10 \times 10^9$ platelets/L decreased clearance by 2% (95% CI 1-3%). In terms of steady-state concentrations, the “typical” individual with the median weight and platelet count (77kg and of $247 \times 10^9$ platelets/L) resulted in a steady-state concentration (calculated by the dosing rate divided by clearance) of approximately 1700 fmol/punch (95% CI 1600-1800 fmol/punch), while a 10-unit increase in both variables resulted in a population-average steady-state concentration of approximately 1600 fmol/punch (95% CI 1550-1700 fmol/punch). In summary, the final model included weight and platelet count on estimated steady-state TFV-DP, which decreased inter-individual variability from 18% in the base model to 13%. Residual variability was 17%.

Categorical Adherence Predictions

Using the constant input model, simulations of varying platelet and weight values were generated for dosing frequencies of 1-7 doses/week. Overall descriptive statistics of each dosing group are shown in Table 4. Receiver operating curve (ROC) analysis indicated a cut-off of approximately 580 fmol/punch for ≥3 doses/week and 830 fmol/punch for ≥4 doses/week, with >94% positive and negative predictive value. To better understand the exact dosing frequency of 700 fmol/punch – a key protective concentration corresponding to the 25th percentile of ≥4 doses/week with previous interpretations [123] – simulations of 3, 3.25, 3.5 and 3.75 doses/week...
Table 4 Descriptive Statistics of Simulated TFV-DP Concentrations (fmol/punch) by Dosing Frequency

<table>
<thead>
<tr>
<th>Doses/Week</th>
<th>Mean</th>
<th>SD</th>
<th>Median</th>
<th>25th Per</th>
<th>75th Per</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>272</td>
<td>69</td>
<td>264</td>
<td>222</td>
<td>313</td>
</tr>
<tr>
<td>2</td>
<td>509</td>
<td>133</td>
<td>492</td>
<td>414</td>
<td>586</td>
</tr>
<tr>
<td>3</td>
<td>748</td>
<td>198</td>
<td>725</td>
<td>605</td>
<td>864</td>
</tr>
<tr>
<td>4</td>
<td>994</td>
<td>253</td>
<td>964</td>
<td>813</td>
<td>1139</td>
</tr>
<tr>
<td>5</td>
<td>1223</td>
<td>316</td>
<td>1186</td>
<td>995</td>
<td>1405</td>
</tr>
<tr>
<td>6</td>
<td>1485</td>
<td>387</td>
<td>1435</td>
<td>1210</td>
<td>1709</td>
</tr>
<tr>
<td>7</td>
<td>1724</td>
<td>441</td>
<td>1673</td>
<td>1408</td>
<td>1981</td>
</tr>
</tbody>
</table>
were conducted without the influence of covariates. 3.5 doses/week (modeled as one dose every 48 hours) was the minimum dosing frequency that produced TFV-DP concentrations ≥700 fmol/punch at the 25th percentile. ROC analyses indicated a cut-off of approximately 800 fmol/punch to differentiate adherence levels ≥3.5 doses/week, although positive and negative predictive value were lower (67% and 62%, respectively), probably due to a greater overlap in concentrations given the narrower inter-dose intervals between 3-3.75 doses/week.

Dosing frequencies were then grouped into <2, 2-3, 4-5 and 6-7 doses/week and a cross-validated decision tree generated an optimal categorization based on TFV-DP concentration cut-offs, weight and platelet count. These results are summarized in Table 5. Concentrations were rounded to the nearest 50 fmol/punch. The proportion of correctly classified observations for each categorization was >80%. Misclassification rates were compared between current and these tailored adherence interpretations of the raw data of the 48 participants. Using previous interpretations, 28/94 (30%) of the assessments would be misclassified, while 16/94 (17%) misclassifications occurred after incorporating weight and platelet counts, predominantly improving the precision of individuals who weighed less than 70kg.

A second outcome of the study was to test the ability to predict an individual’s expected steady-state concentration prior to drug administration. Mixed random forests and LASSO regression models were developed using the original set of covariates and subsequently repeated with only the Top 5 predictors. This was done to reduce model complexity and eliminate extraneous/noise variables. The top 5 covariates of the random forest were gender, weight, platelet count, hematocrit and age. The percent error between predicted and observed concentrations, RMSE and MAE values of the test dataset are shown in Table 6. Predictions from the population PK model are also shown for comparison. Random forests
Table 5 Adherence Benchmarks of Simulated TFV-DP Concentrations (fmol/punch) Compared to Current Interpretations

<table>
<thead>
<tr>
<th>Weight (kg)</th>
<th>Previous</th>
<th>&lt;70</th>
<th>70-110</th>
<th>70-110</th>
<th>&gt;110</th>
<th>&gt;110</th>
</tr>
</thead>
<tbody>
<tr>
<td>Platelet Count (*10⁹/Liter)</td>
<td>Previous</td>
<td>&lt;200, 200-315, 315</td>
<td>&gt;200</td>
<td>&lt;200</td>
<td>&gt;200</td>
<td>&lt;200</td>
</tr>
<tr>
<td>Doses/Week</td>
<td>&lt;2</td>
<td>&lt;350</td>
<td>≤300</td>
<td>≤300</td>
<td>≤300</td>
<td>≤300</td>
</tr>
<tr>
<td></td>
<td>2-3</td>
<td>350-699</td>
<td>301-849</td>
<td>301-849</td>
<td>301-849</td>
<td>301-699</td>
</tr>
<tr>
<td></td>
<td>4-5</td>
<td>700-1049</td>
<td>850-1499</td>
<td>850-1299</td>
<td>850-1149</td>
<td>700-1299</td>
</tr>
<tr>
<td></td>
<td>6-7</td>
<td>&gt;1050</td>
<td>≥1500</td>
<td>≥1300</td>
<td>≥1150</td>
<td>≥1300</td>
</tr>
</tbody>
</table>
Table 6 Percent Error and Model Performance Statistics of Expected TFV-DP Concentration Predictions

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>25th</th>
<th>Median</th>
<th>Mean</th>
<th>75th</th>
<th>Max</th>
<th>MAE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pop PK</td>
<td>0.81</td>
<td>10.3</td>
<td>23.1</td>
<td>26.9</td>
<td>36.5</td>
<td>100.9</td>
<td>388</td>
<td>474</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.3</td>
<td>7.4</td>
<td>13.5</td>
<td>17.0</td>
<td>24.6</td>
<td>60.8</td>
<td>311.1</td>
<td>418.7</td>
</tr>
<tr>
<td>Random Forest Top 5</td>
<td>1.3</td>
<td>6.3</td>
<td>14.4</td>
<td>17.1</td>
<td>22.5</td>
<td>54.9</td>
<td>304.2</td>
<td>395.2</td>
</tr>
<tr>
<td>LASSO</td>
<td>0.1</td>
<td>4.6</td>
<td>16.4</td>
<td>19.2</td>
<td>28.1</td>
<td>85.9</td>
<td>325.8</td>
<td>442.2</td>
</tr>
<tr>
<td>LASSO Top 5</td>
<td>0.8</td>
<td>8.6</td>
<td>16.4</td>
<td>21.0</td>
<td>28.8</td>
<td>79.4</td>
<td>370.8</td>
<td>485.4</td>
</tr>
</tbody>
</table>

MAE: Mean Absolute Error

RMSE: Root Mean Squared Error
produced the lowest MAE and RMSE values (approximately 300 fmol/punch), although the differences between LASSO regression and population PK predictions were not substantial.

**Pharmacometabolomic and Pharmacogenomic Associations**

One hundred seven RBC-specific metabolites and 1,671 SNPs in genes relevant to F/TDF pharmacology were explored for their additional benefit to improving the precision of TFV-DP concentrations. Random forests identified a SNP in equilibrative nucleoside transporter 2 (ENT2) and one SNP in ABCB1 (P-glycoprotein) and ABCC4 (MRP4) as strong predictors of clearance in the constant input model. Although graphical inspection revealed differences between genotype ([Figure 9](#)), addition of either of these SNPs into the final constant input model (with weight and platelet count included) did not further decrease inter-individual variability (<1% change).

**Discussion**

The purpose of this analysis was to identify individual characteristics that improve the precision of the current use of TFV-DP concentrations as an objective measure of adherence to F/TDF for HIV PrEP. This was accomplished by applying the most parsimonious constant input model to describe the accumulation of TFV-DP, followed by a comprehensive search of demographic and clinical covariates. Estimating both between- and within-subject variability via population PK modeling offered an alternative and more informative method of characterizing TFV-DP variability. We found that incorporation of platelet count and body weight into adherence evaluations improved the delineation between boundaries of current TFV-DP thresholds and decreased the misclassification rate by almost half (17% versus 30%), thereby providing tailored assessments and enabling more precise adherence-efficacy relationships.
Figure 9 Box plot relationships of significant SNPs (20= wild-type, 11= heterozygous, 0= variant) versus constant-input clearance
Further, we used ROC and decision tree analyses to identify TFV-DP benchmarks instead of relying on 25th percentiles. Interestingly, these analyses did not dramatically change the previous thresholds. Nonetheless, our results suggest that previous interpretations were overly conservative, and that the pharmacological forgiveness of F/TDF for PrEP may be higher than currently recognized.

Previous interpretations indicated four or more doses/week as corresponding to concentrations of 700 fmol/punch, which was associated with 100% (86-100%) relative risk reduction of HIV acquisition [127]. Our analyses suggest that this adherence interpretation is dependent on body weight and platelet count. Concentrations above 700 fmol/punch indicate 2-3 doses/week for individuals ≤110 kg and ≥4 doses/week for those >110 kg. Higher rates of misclassification in women using previous interpretations were largely mitigated after incorporating body weight (31% to 13% misclassification rate), as most of the women in the study weighed <70 kg. Lower thresholds to attain the adherence-efficacy plateau (i.e., 700 fmol/punch) align with intermittent F/TDF dosing of the IPERGAY trials, where participants experienced high efficacy (86%, 95% CI: 40-99%) following a median of 15 and 18 doses/month (3.75 and 4.5 doses/week) [128, 129].

Several pharmacokinetic models were evaluated to describe the formation of TFV-DP. The development of these models allowed for novel insights into the cellular pharmacology of F/TDF, quantification of between- and within-subject variability, and the identification of sources of variability. Although previous plasma tenofovir models report a two-compartment model, our analysis included samples up to 28 days post-dose and an ultra-low lower limit of quantification of 0.25 ng/ml. This extended sampling time provided evidence of a third compartment, which we hypothesize represents the catabolism of intracellular TFV-DP back to
tenofovir and its subsequent release back from within the cell into the plasma. Previous models suggest saturable uptake of tenofovir in PBMCs, although these studies did not assess RBC uptake and were conducted in people living with HIV. Uptake in PBMCs in healthy individuals has also been modeled with first-order kinetics [88]. Future research should investigate the saturating PK of F/TDF in healthy individuals, as well as the influence of tenofovir-monoester on cellular uptake [130].

Inter-individual and residual variability was significantly lower than previously reported, which can partly be attributed to our directly observed dosing methods. Unrecognized partial adherence or ignoring actual dosing information can yield biased PK parameter estimates and significantly overestimate inter-individual and residual variability [131, 132]. In addition, previous plasma tenofovir models report 30-40% inter-individual CV on oral clearance after inclusion of covariates [133-137], while incorporation of electronic monitoring-based dosing methods resulted in only 16% CV on clearance, similar to the 17% CV observed in our data [138]. Parameter estimates and errors of the current models can be used for future pharmacokinetic simulations of F/TDF, such as alternative dosing regimens or modeling dose forgiveness designs.

Another potential and novel source of inter-individual variability can be attributed to pharmacometabolomic and pharmacogenomic associations [139, 140]. Previous studies have identified polymorphisms in various genes related to the cellular pharmacology of F/TDF and their association with TFV-DP concentrations or development of renal dysfunction [84, 141]. Although they did not further reduce variability, we also identified several SNPs that enabled novel insight into the cellular pharmacology of F/TDF. SNPs were associated with different phases of the conversion of TDF to TFV-DP, including elimination and absorption and
intracellular relocation at the transporter level. Low inter-individual variability after inclusion of demographic and clinical covariates may have decreased the additional value of metabolites and genotypes. Future studies will continue to explore pharmacometabolomic and pharmacogenomic influences on the cellular pharmacology of F/TDF.

While this study had strengths, there were also limitations. Redundancy between the effects of covariates was an important consideration during model development. Covariates found to be associated with each other – such as relationships found between gender, hematocrit and weight - were excluded from entering the model simultaneously. Covariate selection between the associated variables was justified by biological plausibility and the intended application of the model to men who have sex with men. Although we selected weight in the final model and its inclusion decreased misclassification rates in women, it is possible that gender still influences F/TDF pharmacology. It is also possible that different models can have similar predictive accuracy. For example, creatinine clearance was a significant covariate but was not included in the final model. Because body weight – a variable associated with creatinine clearance – was already included in the model, the inclusion of platelet count offered unique information for explaining individual variation in steady-state TFV-DP concentrations. In addition, absence of additional directly observed F/TDF studies prevented external validation of these results. Finally, participants in this study were adults recruited from the United States and the majority were Caucasian. Unique physiological differences in various populations, such as pregnant women, diverse ethnicities and pediatric patients limit the use of these adherence interpretations for these groups.

In summary, our current analysis provided new insight into the sources of variability associated with TFV-DP concentrations, leading to updated adherence benchmarks. These
updated benchmarks, which take into account weight and platelet count, can be used for a more precise approach to assess adherence and adherence-response relationships. Future studies will apply these benchmarks in real-world PrEP trials to characterize adherence patterns within individuals, factors associated with adherence, and confirm relationships between adherence and relative risk reduction of HIV acquisition. After development of these individualized adherence assessments, the next step was to apply these benchmarks to the iPrEx Open Label Extension trial, with the goal of refining adherence-efficacy relationships.
CHAPTER V

REVISITING THE PHARMACOLOGICAL FORGIVENESS OF HIV PRE-EXPOSURE PROPHYLAXIS

The evidence is clear that increasing gradients of adherence result in increasing protective benefit of emtricitabine plus tenofovir disoproxil fumarate (F/TDF) or tenofovir alafenamide (F/TAF) for HIV pre-exposure prophylaxis (PrEP) in men who have sex with men (MSM) [142, 143]. Intracellular concentrations of tenofovir-diphosphate (TFV-DP) in red blood cells, measured with dried blood spots, can be used as an objective measure of cumulative adherence to F/TDF and F/TAF, given the ~17-day half-life observed in red blood cells [52, 144]. The evidence suggests that the “adherence-efficacy” curve reaches a plateau at an average of four or more doses per week for MSM. For F/TDF, current understanding suggests that TFV-DP >700 fmol/punch in dried blood spots (DBS) is consistent with four or more doses per week, resulting in the maximal benefit in MSM (100% (95% CI 86-100%) relative risk reduction) [118]. However, current adherence interpretations for F/TDF are influenced by inter-individual variability. For example, a 30% coefficient of variation (%CV) was observed in TFV-DP concentrations following directly observed therapy. Benchmarks were developed by setting thresholds at 25th percentiles (e.g., 700 fmol/punch is the 25th percentile of 4 doses per week) resulting in a 25% misclassification rates by definition. More precise methods of interpreting adherence thresholds would refine adherence-efficacy relationships.

A recent analysis identified weight and platelet count as significant predictors of TFV-DP concentrations. After incorporation of these two variables, a sub-population of individuals weighing less than 70kg emerged who exhibited the highest intrinsic TFV-DP concentrations. Within these individuals, 850 and 900 fmol/punch corresponded to four or more doses/week by
decision tree and receiver operating characteristic (ROC) analyses, respectively. Three or more doses per week were associated with TFV-DP concentrations ≥640 fmol/punch. Individuals weighing more than 110kg and with a platelet count <200 *10⁹/Liter were identified as having the lowest intrinsic concentrations. Within these individuals, 700 and 650 fmol/punch were associated with four or more doses per week, as assessed by decision tree and ROC analyses, respectively. 500 fmol/punch was associated with three or more doses per week in this sub-population.

When these tailored interpretations were applied to the iPrEx OLE cohort study [118], we found that 24/28 (86%) of the observed infections occurred in individuals weighing less than 70 kg. It is therefore likely that these individuals were even less adherent than previous interpretations would imply. TFV-DP concentrations of these 28 individuals ranged from below the limit of quantification (i.e., off-drug) to 416 fmol/punch. Three of the four individuals weighing >70kg had concentrations that were below the limit of quantification, while one individual’s concentration at the seroconversion visit was 416 fmol/punch, consistent with two doses per week using tailored interpretations. These tailored results indicate that the iPrEx OLE efficacy plateau occurred at three or less doses per week, and not four doses per week as previously considered.

The implications of these findings support substantial pharmacological forgiveness of F/TDF for PrEP and are consistent with the high efficacy of “on-demand” PrEP regimens before and after sexual activity for MSM. They also align with pharmacokinetic models of F/TDF that predict almost complete protection with two or more doses/week, as assessed by attainment of the 90% effective (EC₉₀) active metabolite to endogenous nucleotide ratio [145]. Similar investigations to these are needed to understand PrEP efficacy in women according to gradients
of adherence. These findings favor continued investigation of intermittent PrEP dosing, especially in individuals at high-risk for HIV infection with concurrent low adherence to daily PrEP [128].
CHAPTER VI

SUMMARY AND FUTURE DIRECTIONS

The efficacy of daily F/TDF for HIV PrEP is heavily dependent upon adherence to the prescribed regimen. The wide variability in efficacy observed in the initial PrEP trials (0-75%) can partly be attributed to variable adherence in each trial. Thus, there is a critical clinical need for objective measures of adherence to F/TDF. These objective assessments can be implemented both in the context of clinical trials and in real-world treatment plans. In clinical settings, individuals are asked to come into the clinic approximately every three months for HIV testing and blood work for evaluating kidney function. It is during these visits that the objective adherence assessments can also be evaluated, and the appropriate steps can be taken with respect to adherence counseling, increased monitoring, discussion of alternative treatments or treatment discontinuation. Feedback from providers to patients of drug level assessments and the importance of adherence can increase patient acceptance and adherence in the long-term. Moreover, objective measures of F/TDF are critically needed to support ongoing clinical trials. Accurate and continuous assessments of each individual’s adherence allow for more precise adherence-efficacy relationships and prevent the possibility of futile interventions, such as those observed in women in the initial PrEP trials.

We have shown that concentrations of TFV-DP in red blood cells, measured with dried blood spots, can be used for objective adherence assessments, given the 17-day half-life in these cells. This long half-life results in a 25-fold accumulation from first dose to steady-state, allowing for a cumulative assessment of adherence over the preceding 1-2 months. Given that individuals should be going into the clinic every three months for HIV testing, the use of TFV-DP concentrations provide an accurate depiction of an individual’s average adherence in-
between visits. However, a limitation of TFV-DP cumulative assessments is that they cannot
discern specific patterns of adherence, but rather a summary of average adherence. Using plasma
tenofovir and FTC-TP concentrations are also available in the toolbox of adherence measures for
evaluating recent dosing. Plasma tenofovir and FTC-TP concentrations can assess whether a
dose was taken in the preceding seven days. The use of FTC-TP is advantageous, as both TFV-
DP and FTC-TP can be measured in the same assay. Future studies should evaluate point of care
testing for these moieties which would enhance clinician-patient conversations about adherence.

Despite the development and widespread use of TFV-DP benchmarks for various dosing
frequencies (<2, 2-3, 4-6, and 7 doses/week), several limitations existed that must be addressed.
These limitations include the use of inter-quartile ranges for establishing benchmarks -
suggesting 25% misclassification rates – and a 30% coefficient of variation observed in TFV-DP
concentrations. Moreover, high rates of misclassification were observed in women using
previous interpretations. Because of the aforementioned clinical applications of objective
measures of PrEP adherence, there is a critical need to improve the precision of the use of these
concentrations to provide tailored adherence interpretations.

To address these limitations, our first task was to address lingering challenges in
population PK modeling. A common issue encountered in pharmacokinetic modeling is the
presence of data below the limit of quantification (BLQ). Chapter III of this dissertation
proposed guidelines to the pharmacometric community on how to handle BLQ data, which can
be applied to all population models. We found that for simpler models and moderate percentage
of BLQ samples, such as intravenous and oral one-compartment models with 30% or less BLQ,
the use of the likelihood-based M3 method does not appear to provide additional benefit to
decreasing the bias of estimated PK parameters, compared to treating the BLQ samples as
missing. For two-compartment and nonlinear models, the M3 method should be implemented at 5% or more BLQ, although clearance and central volume were not biased in two-compartment models up to 30% BLQ when BLQ was treated as missing. Given that M3 implementation sometimes results in complex models with unsuccessful convergence, these guidelines propose scenarios where the simpler M1 (treat as missing) approach can be utilized. Overall, these results provided a comprehensive and consistent evaluation of methods to handle BLQ, which were applied to other portions of this dissertation and can be applied to all population models.

We then began the development of a structural and statistical model for describing TFV-DP accumulation in red blood cells (Chapter IV). This was accomplished by population pharmacokinetic modeling, which offered an alternative and more informative approach to quantifying the variability associated with TFV-DP concentrations, as compared to inter-quartile ranges. We first explored a model linking plasma tenofovir to the formation of TFV-DP. Because of convenience sampling, we needed to leverage plasma data from a previous, intensive PK study of F/TDF to better describe the absorption kinetics of tenofovir. The best-fitting model was a three-compartment model for plasma tenofovir linked to the formation of TFV-DP via a first-order constant. Saturable and tolerance models were not supported by the data, indicating the $E_{\text{max}}$ was not attained and the formation of TFV-DP was still in the linear range. Nonetheless, conclusions of the saturation of tenofovir uptake into red blood cells cannot be confirmed without additional data at both lower and higher doses. The linearity of tenofovir uptake and the small rate constant observed raises questions about the influence of plasma tenofovir on red blood cell uptake versus the pro-drug and plasma tenofovir-monoester intermediate, as we have previously described [68]. Future in vitro or ex vivo studies should characterize the pharmacokinetics of the monoester in relation to its effects on TFV-DP uptake. In addition,
another limitation of this model was that the first-order constant was assumed to account for all steps of uptake and conversion of tenofovir to TFV-DP. Future studies should explore the utility of physiologically-based PK models that incorporate *in vitro* kinase enzyme kinetics into the model.

The second class of pharmacokinetic models developed was a one-compartment constant input model. This parsimonious model ignored plasma tenofovir contributions and described the accumulation of TFV-DP as mimicking an IV infusion model. The 17-day half-life relative to the short dosing intervals allows for the development of such a model. Ultimately, this less complex constant input model was chosen for further analysis. However, given the large number of variables investigated in this dissertation, in addition to a smaller sample size which precludes the use of stepwise and significance-based testing of covariates, novel approaches for covariate selection needed to be considered. Machine learning approaches, specifically decision trees and random forests, were used to select the most predictive covariates to enter the final constant input model. Ultimately, this led to selection of body weight and platelet count as the most influential covariates, and tailored interpretations were developed based on various values of each covariate. The development of this model allowed for simulations of the full range of adherence patterns, without the need for conducting additional experiments. Future studies should continue to explore the utility of machine learning in clinical pharmacology analyses. The ability to predict a future individual’s PK parameters based on established machine learning models can significantly reduce the time and cost associated with conducting large PK studies in later phases of clinical trials. These interpretations can be readily implemented in the clinic, can support ongoing clinical trials, and can be used to explore the benefit of adherence-guided PrEP treatment versus standard of care.
When we applied these tailored adherence interpretations to the iPrEx OLE clinical study in men who have sex with men (Chapter V), we were able to refine the relationships studied between adherence and efficacy. We found that most infections (24/28) occurred in individuals who weighed less than 70kg, which was the sub-population with the highest intrinsic TFV-DP concentrations using tailored interpretations. This suggests that these individuals were even less adherent than current interpretations would imply, because higher concentrations were associated with lower adherence. Therefore, these results support substantial pharmacological forgiveness of F/TDF for PrEP and align with pharmacokinetic models of F/TDF that predict almost complete protection with two or more doses/week, as assessed by attainment of the 90% effective (EC$_{90}$) active metabolite to endogenous nucleotide ratio. This supports reducing dosing regimens, such as those investigated in the IPERGAY “on-demand” trials. The fact that most infections occurred in men weighing less also raises the question of a biological basis for higher rates of infection in lighter individuals. While adherence is likely the determining factor, additional investigations should be conducted to further probe the influence of weight on the risk of HIV infection.

Daily PrEP may not be a feasible or practical choice for all individuals considering PrEP. Higher costs and risk of toxicity associated with chronic F/TDF dosing may prevent individuals from choosing this highly effective preventative method. It is also widely recognized that adherence declines over time, which was also observed in the iPrEx OLE study, which suggests that commitment to daily PrEP may lead to increased burden and increased risk of HIV infection after declining adherence. This dissertation supports alternative PrEP approaches, where individuals can assess their risk of HIV acquisition and initiate PrEP during those periods of high risk.
REFERENCES


APPENDIX A

POPULATION PHARMACOKINETICS OF ETRAVIRINE IN HIV-POSITIVE CHILDREN AGES 1-<6 YEARS

Introduction

According to the UNAIDS, 1.7 million children under 15 years were living with HIV in 2018 [146]. Despite increased and robust efforts to expand access to antiretroviral (ARV) therapy globally, fewer ARV are available for pediatric use compared with adults [147]. Several challenges exist in the development of ARV for pediatric populations [148]. Differences in the pharmacokinetic profile of drugs occur between children and adults, such that the ongoing maturation of organs involved in drug disposition, in addition to high inter-individual variability observed in children, may make direct adult dose scaling unpredictable [149]. Other factors such as adherence, drug palatability, formulation and handling impose other practical challenges [150].

The non-nucleoside reverse transcriptase inhibitors (NNRTIs) have been components of recommended ARV therapy for treatment-naïve individuals, including children, for many years [151, 152]. They generally demonstrate simple dosing regimens and favorable efficacy and tolerability [153]. However, nevirapine and efavirenz, the most widely administered drugs of this class, have low genetic barriers for development of drug resistant mutations, even after a single dose. Among Thai children who failed an NNRTI-based regimen, 97% had NNRTI-related resistance [154]. In addition, across 19 studies, the prevalence of pretreatment drug resistance was 42.7% (95% CI 26.2%–59.1%) among children exposed to prevention of mother-to-child transmission (PMTCT) and was most commonly observed in children <3 years and with the use
of NNRTIs [155]. Therefore, there is a need to develop alternative therapeutic options to overcome resistance in infants and children failing ART therapy.

Etravirine (ETR) was shown to exhibit potent \textit{in vitro} activity against HIV-1 strains that are resistant to efavirenz and nevirapine [156, 157]. Three large clinical trials, DUET1/2 and TRIO, demonstrated that addition of ETR improved virologic and immunological outcomes compared to placebo in adults [158-160]. Studies in children and adolescents 6-17 years old also demonstrated exposure and safety profiles comparable to adults, encouraging the investigation of weight-based dosing in children younger than 6 years [161] [162].

International Maternal Pediatric Adolescent AIDS Clinical Trials (IMPAACT) P1090 was a dose finding trial of ETR in children \( \geq 1 \) to \( < 6 \) years of age. ETR met protocol-defined pharmacokinetic targets in P1090, but exposures were lower overall compared with historical data. The objectives of this analysis were to (1) develop a population pharmacokinetic model to identify covariates associated with ETR pharmacokinetics in P1090, (2) to use the pharmacokinetic measures to assess ETR adherence and (3) to explore the ETR pharmacokinetic-viral response relationship.

\textbf{Methods}

\textit{Study Design}

IMPAACT P1090 was a Phase I/II, open-label study designed to determine the pharmacokinetics, optimal dosage, safety, and tolerability of ETR in treatment-experienced children with HIV ages \( \geq 1 \) to \( < 6 \) years. Detailed study design and methods have been previously described. Children with \( \leq 10 \)-fold loss in ETR sensitivity based on phenotypic testing were eligible if they were currently failing a combination ARV regimen (containing at least three ARVs from at least 2 drug classes) for at least eight weeks, or on a treatment interruption of at
least 4 weeks with a history of virologic failure on a combination ARV regimen. After enrollment, the protocol team approved an individualized optimized background regimen (OBR) for each participant, which included a ritonavir-boosted protease inhibitor (PI/r) and at least one other active drug. The study was reviewed and approved by the Institutional Review Boards at the clinical sites and all parents or guardians provided written informed consent.

Participants were stratified by age in two cohorts (Cohort I: ≥ 2 years to < 6 years; Cohort II: ≥ 1 year to < 2 years). Enrollment began with Cohort I and after pharmacokinetic and safety criteria were met for the first six children enrolled in the cohort (the “mini-cohort”), Cohort II opened. An intensive pharmacokinetic was performed 7-18 days after initiating ETR. ETR doses were adjusted in participants that failed to achieve an ETR area-under-the-curve from 0 to 12 hours (AUC_{12h}) above 2,350 ng*hr/mL (the 10th percentile in adults). A repeat intensive pharmacokinetic assessment was performed 7-14 days after the ETR dose adjustment. Sparse (convenience) samples were obtained at Weeks 4, 8, 12, 24 and 48, and at any visits to confirm virologic failure to quantify ETR in plasma.

Initially, ETR was evaluated at 5.2 mg/kg twice daily, based on the PIANO study in adolescents [10]. However, ETR exposures in the mini-cohort failed predefined pharmacokinetic thresholds. Thus, modeling and simulation were used to develop a weight-banded dosing regimen that improved the probability of achieving the target exposures, not exceeding the adult dose of 200mg twice daily. Etravirine was administered either as the tablet swallowed whole or dispersed in liquid.

Model Development

Phoenix NLME (Version 8.2, Certara) was used to develop a population pharmacokinetic model of ETR. A naïve pooled approach was first performed to obtain initial
parameter estimates. From these estimates, a base model was developed using first-order conditional estimation. Additive, log-additive, multiplicative and mixed residual error models were evaluated. A one- versus two- compartment model was compared utilizing goodness-of-fit plots and -2LL/AIC criteria. Covariance between model parameters was also assessed. Covariates tested in the model included age, weight, body surface area (BSA), form of dosage administration (swallowed whole vs. dispersed in liquid), concomitant PI/r (lopinavir, darunavir, or atazanavir), and country of residence (US, South Africa, or Brazil). Each covariate was tested separately for individual effects on relative bioavailability, the absorption constant, clearance or volume of distribution. For those covariates that decreased the objective function by $>3.84\ (P = 0.05)$, a forward-backward stepwise procedure (forward/backward significant $P=0.05/0.01$) was then performed to evaluate multiple covariate significance to generate a final model. Goodness-of-fit was evaluated by inspecting residual and observed vs. predicted concentration plots, the objective function, visual predictive check (VPC) (N=1000) and biological plausibility.

Model Simulations

Using the final covariate model, simulations (N=1000) were performed to obtain expected ETR exposure metrics for the lowest FDA-approved dose (100 mg), administered as either swallowed whole or dispersed in liquid. These simulations provided additional information regarding the likelihood of attaining the study’s target exposure and its relationship with virological success.

Adherence Assessments

The contribution of adherence to lower ETR exposures in P1090 was investigated using the integrated pharmacokinetic adherence measure (iPAM) approach developed by Brundage et al (11). In this approach, a model of intensive pharmacokinetic samples is developed and posthoc individual estimates of pharmacokinetic parameters are used to evaluate the ratio of observed to
predicted concentrations. For this analysis, a model using only intensive pharmacokinetic data from pharmacokinetic profiles that met protocol-defined individual pharmacokinetic-criteria was developed using the same structural components and covariate effects as the final population pharmacokinetic model. The sparse samples collected served as indicators of adherence throughout the study. The ratio of observed to predicted concentration was calculated and children were deemed nonadherent if the observed concentration was below 30% of the expected concentration (i.e., ratio <0.7). Of note, analytical variability and pharmacokinetic model misspecification cannot be discerned from these calculated deviations. However, if a participant was nonadherent, large discrepancies (i.e. >30%) between observed and predicted concentrations could potentially delineate overall adherence.

**Pharmacodynamic Assessments**

Virologic failure was defined as a viral load $\geq$400 copies/mL at Week 48. Associations between demographic variables and pharmacokinetic parameters and virologic status at Week 48 were assessed with Student’s t-tests. Adherence comparisons were evaluated between virologic successes and failures.

**Results**

Twenty-six children enrolled into the study. However, one child whose intensive pharmacokinetic was deemed unevaluable had an extremely low AUC and suspected adherence challenges and subsequently discontinued the study at Day 16. Baseline characteristics of the 25 children with evaluable pharmacokinetic are shown in Table 7. Subjects were recruited for the study from Brazil, South Africa and the United States. Median age was 4 years old, ranging from 1 to 6 years. Median weight was 15 kg (8-24 kg). The optimized background regimen for participants included ritonavir and either lopinavir, darunavir or atazanavir.
<table>
<thead>
<tr>
<th>Table 7 Baseline Demographics of 25 Children Receiving Etravirine</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Participants (n=25)</strong></td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td><strong>Race/Ethnicity, n (%)</strong></td>
</tr>
<tr>
<td>Black or Black African</td>
</tr>
<tr>
<td>Hispanic</td>
</tr>
<tr>
<td><strong>Country, n (%)</strong></td>
</tr>
<tr>
<td>South Africa</td>
</tr>
<tr>
<td>Brazil</td>
</tr>
<tr>
<td>United States</td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
</tr>
<tr>
<td>Median (Range)</td>
</tr>
<tr>
<td><strong>Weight (kg)</strong></td>
</tr>
<tr>
<td>Median (Range)</td>
</tr>
<tr>
<td><strong>Body Surface Area (m²)</strong></td>
</tr>
<tr>
<td>Median (Range)</td>
</tr>
<tr>
<td><strong>Dose, mg, n(%)</strong></td>
</tr>
<tr>
<td>75</td>
</tr>
<tr>
<td>100</td>
</tr>
<tr>
<td>125</td>
</tr>
<tr>
<td><strong>Dosage Administrationa, n (%)</strong></td>
</tr>
<tr>
<td>Dispersed</td>
</tr>
<tr>
<td>1-2 years</td>
</tr>
<tr>
<td>2-6 years</td>
</tr>
<tr>
<td>Swallowed whole</td>
</tr>
<tr>
<td>1-2 years</td>
</tr>
<tr>
<td>2-6 years</td>
</tr>
<tr>
<td><strong>Concomitant Protease Inhibitor n(%)</strong></td>
</tr>
<tr>
<td>Ritonavir/atazanavir</td>
</tr>
<tr>
<td>Ritonavir/darunavir</td>
</tr>
<tr>
<td>Ritonavir/lopinavir</td>
</tr>
</tbody>
</table>

aOne participant took a combination of ETR (100 mg swallowed and 25 mg dispersed)
Noncompartmental Analysis

Time vs. concentration curves for etravirine for all intensive pharmacokinetic visits (Day 8-17 and after dose adjustment, if applicable) are shown in Figure 10. A noncompartmental analysis was performed to estimate AUC\textsubscript{12hr}, maximum concentration (C\text{max}) and time of C\text{max} (t\text{max}) (Table 8). Geometric AUC\textsubscript{12hr} mean was <2,350 ng*hr/mL in 8 children. A dose adjustment was performed and a repeat intensive pharmacokinetic was conducted in these children.

AUC\textsubscript{12hr} was significantly higher in those who swallowed the tablet whole compared to those who took ETR dispersed in liquid (7,118 ± 5,191 ng*hr/mL vs. 3,438 ± 1,461 ng*hr/mL, P=0.003) (Figure 11). Significant differences were also found between the form of administration for C\text{max} (787 ± 536 ng/ml vs. 442 ± 240 ng/ml, P=0.01) and C\text{12hr} (522 ± 380 ng/ml vs. 243 ± 222 ng/ml, P=0.01). Body weight was positively correlated with AUC\textsubscript{12hr} (r =0.42, P=0.01), C\text{max} (r =0.41, P =0.01) and C\text{12hr} (r =0.41, P =0.02). Differences in AUC\textsubscript{12hr} between the form of administration remained significant after adjusting for weight. Age and type of PI/r were not correlated with any of these parameters. Half-life estimates were similar between children who swallowed the pill whole and those who dispersed etravirine in liquid (9 ± 5 vs. 14 ± 5 hr, P=0.06).

Population Pharmacokinetics

A population pharmacokinetic analysis was performed on all intensive and sparse sampling data. A one-compartment model with a multiplicative residual error best described the data. Transit absorption compartments significantly improved the fit of the model, consistent with previous models that incorporated a lag time. A full variance-covariance matrix was included between model parameters.
Figure 10 Concentration vs. Time Plots of 25 Children Receiving ETR. The solid black line is the LOESS smoothed line.
Table 8 Mean (SD) Pharmacokinetic Parameters for All Intensive Pharmacokinetic Visits Compared to Historical Data

<table>
<thead>
<tr>
<th></th>
<th>Historical Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 - &lt; 6 years (n=35)</td>
</tr>
<tr>
<td>( \text{AUC}_{12h}, \text{ng*hr/mL} )</td>
<td>4,483 (3382)</td>
</tr>
<tr>
<td>( C_{\text{max}}, \text{ng/mL} )</td>
<td>542 (374)</td>
</tr>
<tr>
<td>( T_{\text{max}}, \text{h} )</td>
<td>4 (range 1-9)</td>
</tr>
<tr>
<td>Individual ETR dose increase required (AUC &lt; 2350 ng*hr/mL)</td>
<td>8 (32%)</td>
</tr>
</tbody>
</table>
Figure 11 ETR AUC$_{12h}$ (median) by Dosage Administration (Children represented more than once)
The addition of the form of administration on relative bioavailability significantly decreased the objective function (ΔOBJ=4.32, P<0.05). Relative to the bioavailability of swallowing the tablet whole (fixed at 100%), the bioavailability of taking etravirine dispersed in liquid was 59%. Country of origin also significantly affected apparent clearance in univariate analyses, although this relationship disappeared after inclusion of form of administration into the model. Notably, six of the seven children who took etravirine whole were from South Africa, which was also the country with the strongest relationship to apparent clearance (40% lower CL/F). Age, weight, body surface area and PI/r did not influence model parameters. Population pharmacokinetic estimates in this population compared to historical data are shown in Table 9.

Final model parameter estimates (%CV, 95% CI) for \( \frac{V}{F} \), CL/F and \( k_a \) were 213 L (19%, 134-293), 17 L/hr (11%, 13-21) and 0.99 hr\(^{-1}\) (15%, 0.7-1.28). Residual variability was estimated to be 38%.

A visual predictive check (VPC) was done to validate the model (Figure 12). The majority of observed concentrations fell within the 5\(^{th}\) and 95\(^{th}\) percentiles of the predicted model. Goodness-of-fit plots, including CWRES vs. time, CWRES vs. predicted concentration and observed vs. predicted concentrations, also further validated the model.

**Dosing Simulations**

All children (n=8) who required a dose adjustment based on the initial intensive pharmacokinetic (AUC\(_{12\text{hr}}\) < 2,350 ng*hr/ml) received 75 or 100 mg of etravirine. Seven of these children also took the drug dispersed in liquid. To determine the model-predicted AUC\(_{12\text{hr}}\) for each dosing scenario, simulations were performed for a 75 and 100 mg dose of etravirine with either form of administration. The median (IQR) AUC\(_{12\text{hr}}\) when 75 mg etravirine was swallowed whole was 3,952 (2835 – 5801) ng*hr/mL and 2,389 (1,624 - 3,371) ng*hr/mL when dispersed
Table 9 Population Pharmacokinetic Parameters of ETR in 25 children Compared to Historical Data

<table>
<thead>
<tr>
<th></th>
<th>1 - &lt;6 years (%CV, 95%CI)</th>
<th>6 – 17 years (PIANO)</th>
<th>Adults (DUET 1/2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_d/F$ (L)</td>
<td>213 (19%, 134-293)</td>
<td>597 (8%)</td>
<td>422 (36%)</td>
</tr>
<tr>
<td></td>
<td>IIV: 37%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CL/F (L/hr)</td>
<td>17 (11%, 13-21)</td>
<td>46.3 (11%)</td>
<td>43.7 (3%)</td>
</tr>
<tr>
<td></td>
<td>IIV: 52%</td>
<td>IIV: 67%</td>
<td>IIV: 60%</td>
</tr>
<tr>
<td>$F_{1\text{whole}}$</td>
<td>1.0 (fixed)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>$F_{1\text{dispersed}}$</td>
<td>0.59 (26%, 0.55-0.64)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_a$ (hr$^{-1}$)</td>
<td>0.99 (15%, 0.7-1.28)</td>
<td>1.07 (34%)</td>
<td>0.88 (46%)</td>
</tr>
<tr>
<td></td>
<td>IIV: 65%</td>
<td>IIV: 174%</td>
<td></td>
</tr>
<tr>
<td>$\sigma$ (%CV)$^2$</td>
<td>38%</td>
<td>31%</td>
<td>20%</td>
</tr>
<tr>
<td>Half-life (hr)$^3$</td>
<td>9.7 ± 5.6</td>
<td>8.9</td>
<td>6.7</td>
</tr>
</tbody>
</table>

$^1$IIV: inter-individual variability (%CV)

$^2$σ: proportional residual error

$^3$Half-life estimates for PIANO and DUET were calculated by population estimates of CL/F and $V_d/F$
Figure 12 Concentration vs. Time Plots of 25 Children Receiving ETR. The solid black line is the LOESS smoothed line.
in liquid. For 100 mg etravirine, median (IQR) AUC$_{12hr}$ when swallowed whole was 5,581 (3,930 – 7,854) ng*hr/mL and was 3,149 (2,123 – 4,456) ng*hr/mL when dispersed in liquid.

**Adherence Assessments**

Similar to the final structural model, a one-compartment model with form of administration on relative bioavailability was developed using only the intensive pharmacokinetic sampling. Posthoc estimates of CL, V$_d$ and k$_a$ were obtained for each individual, and were used to calculate predicted time-post-dose concentrations. The ratio of observed sparse sampling concentrations to the predicted concentration (iPAM score) was calculated. If the score was less than 0.7, overall non-adherence was suspected. From the 45 available concentrations in the children who took the pill dispersed in liquid, 14 (31%) were below 0.7 (range 0.11-0.68). Only three of the 20 (15%) available concentrations were below 0.7 in the group who swallowed the tablet(s) whole (range 0.62-0.67) (Figure 13).

**Pharmacodynamics**

Eight participants had an HIV-1 RNA >400 copies/mL at week 48. Three children were in Cohort I (≥2 years to <6 years) and five were in Cohort II (≥1 year to <2 years). Those who did not reach suppression were younger in age (2.8 ± 1.4 vs. 4.5 ± 1.1 years, P = 0.005). BSA (P=0.06) trended towards an association with Week 48 viral virologic status, with children with higher BSA achieving virological success. Pharmacokinetic parameters (AUC$_{12hr}$, C$_{max}$, C$_{12hr}$, C$_{avg}$) did not differ in those with HIV-1 RNA <400 vs. ≥400 copies/mL at Week 48.

Seven of eight (88%) of the virologic failures took etravirine dispersed in liquid. Of these seven children, four of seven (57%) took a final dose of 100mg. Significant differences in adherence assessments were found between virologic failures and suppressed children. When the ratios of observed to predicted concentrations were compared, virologic failures had 28% lower
Figure 13 Non-adherent iPAM Scores by Form of Administration at Sparse Sampling Visits. Solid line is the mean iPAM Score and dashed line is the non-adherent threshold.
iPAM scores compared with virologic successes (0.43 ± 0.17 vs. 0.71 ± 0.20, P=0.001). Seven of eight (88%) of the failures achieved viral loads <400 copies/ml for at least two weeks through week 48 (range 2-8 weeks).

**Discussion**

The objectives of this analysis were to develop a population pharmacokinetic model of ETR in combination with an optimized background regimen in children living with HIV ages ≥ 1 year to < 6 years, and to evaluate significant sources of inter-individual variability. Although AUC\textsubscript{12h} and C\textsubscript{max} estimates were 1.2- and 1.5-fold higher in adults compared to children, we found that the pharmacokinetic profile of etravirine in this population was similar to that observed in adolescents and adults in the PIANO and DUET 1/2 studies, likely due to large inter-individual variability observed across all populations (7,8,10). The relative bioavailability of etravirine was 41% lower when administered by dispersion in liquid, which was coupled with suspected nonadherence in children who took the tablet dispersed. Etravirine exhibits considerable inter-individual variability across all populations, although no significant differences were found between the age groups with respect to AUC, t\textsubscript{max} and C\textsubscript{max} comparisons.

Pharmacokinetic parameters were moderately lower in these children, which would be expected in a younger population, although large inter-individual was observed in all parameters, as was also observed in estimates in adolescents and adults [163]. In children and adolescents, weight was significantly associated with ETR exposure, similar to associations found in the present study [16]. However, weight did not significantly affect ETR exposure in adults [21]. Across all populations, age, gender, and race did not affect ETR exposure. In children and adolescents, use of darunavir/ritonavir or lopinavir/ritonavir resulted in the lowest ETR exposure (5513 ± 5154 and 4317 ± 3106 ng*hr/mL, respectively), compared to use of
atazanavir/saquinavir/nelfinavir/tipranavir (7172 ± 2526 ng*hr/mL). Although our study was likely not powered enough to detect differences in PI/r, exposures in children <6 years using darunavir and lopinavir were similar to these mean estimates in children and adolescents. Despite concurrent decreases in volume and clearance estimates, half-lives were similar across all age groups.

Children who received etravirine dispersed in liquid had 41% lower bioavailability compared to those who took etravirine whole. We hypothesize that this is potentially driven by incomplete absorption/dosing due to palatability issues or incomplete dosing if etravirine was not fully dissolved in the liquid. Adherence is also likely a factor in this finding, as a larger proportion of children who took etravirine dispersed had greater than 30% lower concentrations throughout the study than predicted by the population model. Because half-life comparisons were similar between both groups, we do not suspect that the form of administration affected systemic clearance or volume of distribution, further validating bioavailability and adherence concerns.

The country of residence (South Africa, Brazil or United States) was also a significant factor on CL/F, with children in South Africa demonstrating a 40% reduction in CL/F. Although this is likely because the majority of South African children also received etravirine whole, it is possible that country of origin can also influence etravirine disposition. Etravirine is metabolized by CYP2C9 and CYP2C19 enzymes, and induces CYP3A4 and potentially CYP3A5, and is very highly bound to plasma proteins (99%) (10). Genetic polymorphisms leading to differences in expression of these enzymes across diverse populations, as well as in the abundances of albumin and alpha(1)-acid glycoprotein, could potentially alter the pharmacokinetics of etravirine [164,
We hypothesize that this could also be a product of bioavailability, such that different foods in each country would alter bioavailability of etravirine.

Eight children required a dose adjustment after an initial dose of 75 or 100 mg. Seven of these children were taking etravirine dispersed in liquid. The FDA has approved a minimum dose of 100 mg in children weighing at least 10kg based on the results from P1090. Through model simulations, we found that with 100mg dispersed, the 25th percentile of AUC$_{12hr}$ was estimated to be 2,123 ng*hr/ml. The dispersed form of etravirine dosed at 100 mg twice daily may require additional monitoring of adherence and viral suppression in children taking these doses, as the 25th percentile of AUC$_{12hr}$ is near the 10th percentile of values in adults (2,350 ng*hr/mL), which was the target AUC in the current study. We also found that the majority of virologic failures occurred in children taking the dispersed form of etravirine and within these failures, more than half were taking 100mg. In congruence with our finding of lower adherence in children taking etravirine dispersed, we also found lower adherence in virologic failures. Because the majority of virologic failures showed evidence of suppression prior to viral rebound, we suspect that nonadherence versus resistance was the driving factor.

In summary, the pharmacokinetic profile of etravirine in children ages 1-6 years who swallowed the tablet whole is similar to that of previously reported studies in adolescents and adults, encouraging the drug’s use in this population. The form of administration of etravirine was found to significantly affect relative bioavailability. Adherence is a significant concern in this population, especially in those who took the tablet dispersed in liquid. This model can aid in simulations of alternative dosing regimens, such as once-daily dosing, and can inform future studies in pediatric populations.
APPENDIX B

COMPLETE R CODE

#########################################################################
Random Forest to Identify Top Covariates of PK Parameters
#########################################################################

set.seed(12345)
dat

#Set outcome parameter

train<-dat
y=train$CL

#Create dataset with dummy variables

dummies_model<- dummyVars(CL~., data=train)
trainData_mat<-predict(dummies_model, newdata=train)
train<-data.frame(trainData_mat)
train$CL<-y

#construct train control object
#5-fold, repeated cross-validation

fit_control=trainControl(method="repeatedcv", number=5, repeats=5)
subsets <- c(1:14)

rfe_control2 = rfeControl(functions = rfFuncs,
method="repeatedcv",
number=5, repeats = 5)

model_rf <- rfe(x=train, train$CL,
sizes = subsets,
rfeControl = rfe_control2, trControl=fit_control)

model_rf
model_rf$resample

#Variable importance rankings
varImp(model_rf, conditional=TRUE)
Simulations of Covariate Effects and Forest Plot Development

```r
set.seed(657687)
df <- data.frame(
  MASS::mvrnorm(n = 1000,
    mu = c(3.40,-0.308,0.340,0.362,0.18, 0.157, 0.03, 0.17),
    Sigma=matrix(c((3.40*0.026)^2,
      0,0,0,0,0,0,(-0.308*0.375)^2,
      0,0,0,0,0,0,(0.340*0.24)^2,
      0,0,0,0,0,0,(0.362*0.41)^2,
      0,0,0,0,0,0,(0.18*3.9)^2,
      0,0,0,0,0,0,(0.157*0.54)^2,
      0,0,0,0,0,0,(0.03*1.5)^2,
      0,0,0,0,0,0,(0.17*0.31)^2),
     8,8,byrow = TRUE)))

names(df) <-
c("POPCL","dPlateletdCL","dWeightdCL","dCrCLdCL","dHCTdCL","dBlackdCL","dBilirubindCl", "dGenderdCL")

knitr::kable(head(df,5))

CLBSVdistribution <- data.frame(CL= 3.40*exp(rnorm(10000,0,sd=0.026^0.5)))

CLBSVdistribution$CLBSV<- CLBSVdistribution$CL/3.40

dfeffects <- df
dfeffects$REF <- dfeffects$POPCL/ median(dfeffects$POPCL)
dfeffects$WT_50 <- dfeffects$REF*(50/77)^dfeffects$dWeightdCL
dfeffects$WT_130 <- dfeffects$REF*(130/77)^dfeffects$dWeightdCL
dfeffects$Platelet_160 <- dfeffects$REF*(160/250)^dfeffects$dPlateletdCL

dfeffects$Platelet_330 <- dfeffects$REF*(330/250)^dfeffects$dPlateletdCL

dfeffects$CrCL_70 <- dfeffects$REF*(70/100)^dfeffects$dCrCLdCL
dfeffects$CrCL_130 <- dfeffects$REF*(130/100)^dfeffects$dCrCLdCL
dfeffects$HCT_30 <- dfeffects$REF*(30/43)^dfeffects$dHCTdCL

dfeffects$HCT_50 <- dfeffects$REF*(50/43)^dfeffects$dHCTdCL
dfeffects$Bilirubin_0.2 <- dfeffects$REF*(0.2/0.5)^dfeffects$dBilirubindCl

dfeffects$Bilirubin_1.5 <- dfeffects$REF*(1.5/0.5)^dfeffects$dBilirubindCl
dfeffects$Black <- dfeffects$REF*exp(dfeffects$dBlackdCL)
dfeffects$Male <- dfeffects$REF*exp(dfeffects$dGenderdCL)

dfeffects$BSV<-  CLBSVdistribution$CLBSV
```

112
dfeffects <- dfeffects[, c("WT_50", "WT_130", "Platelet_160", "Platelet_330", "CrCL_70", "CrCL_130", "HCT_30", "HCT_50", "Bilirubin_0.2", "Bilirubin_1.5", "Black", "Male", "REF", "BSV")]

dflong <- tidyr::gather(dfeffects)
ggplot2::ggplot(dflong, aes(x=value, y=key, fill=factor(..quantile..))) +
ggridges::stat_density_ridges(geom = "density_ridges_gradient", calc_ecdf = TRUE,
quintile_lines = TRUE, rel_min_height = 0.01,
quantiles = c(0.05, 0.5, 0.95)) +
ggplot2::scale_fill_manual(name = "Probability", values = c("#FF0000A0", "white", "white", 
"#0000FFA0"),
labels = c("(0, 0.05]", "(0.05, 0.5]", "(0.5, 0.95]", "(0.95, 1]")
)+
ggplot2::annotate("rect",
xmin = 0.8,
xmax = 1.25,
ymin = -Inf,
ymax = Inf,
fill = "gray", alpha = 0.4
)+
ggplot2::geom_vline(
  ggplot2::aes(xintercept = 1),
  size = 1
)+
ggplot2::theme_bw() +
ggplot2::labs(x="Effects Relative to parameter reference value", y="")

dfeffectslong <- tidyr::gather(dfeffects)
dfeffectslong <- dplyr::group_by(dfeffectslong, key)
dfeffectslongsummaries <- dplyr::summarise(dfeffectslong, mid = quantile(value, 0.5),
lower = quantile(value, 0.05),
upper = quantile(value, 0.95))

dfeffectslongsummaries

dfeffectslongsummaries$paramname <- "CL"

### FOREST PLOT #######
library(forestplot)
forestdat <-
structure(list(
  mean = c(NA, 0.86, 1.19, 1.15, 0.92, 0.88, 1.10, 1.19, 1.17, 0.94, 1.03, 0.97, 1.03),
  lower = c(NA, 0.80, 1.10, 1.04, 0.86, 0.8, 1.02, 1.08, 1.01, 0.62, 0.86, 0.90, 0.94),
  upper = c(NA, 0.93, 1.3, 1.26, 0.98, 0.97, 1.19, 1.31, 1.35, 1.42, 1.23, 1.05, 1.13),
  .Names = c("mean", "lower", "upper"),
  row.names = c(NA, -11L),
  class = "data.frame")

tabletext <- cbind(
  c("Covariate", "Weight 50kg", "Weight 130kg", "Platelet 160 per L", "Platelet 330 Per L",
   "eGFR 70 ml/min", "eGFR 130 ml/min", "Male", "Black", "Hematocrit 30%", "Hematocrit 50%",
   "Bilirubin 0.2 mg/dl", "Bilirubin 1.5 mg/dl"))

forestplot(tabletext,
  forestdat,
  clip = c(0.2, 2),
  boxsize = 0.1,
  zero = 1,
  xlog = FALSE,
  xlab = "Effect Relative to Median",
  col = fpColors(lines = "black", box = "black"),
  ci.vertices = TRUE,
  xticks = c(0.5, 0.75, 1, 1.25, 1.5),
  colgap = unit(0.03, "npc"),
  hrzl_lines = TRUE,
  lineheight = unit(1.1, "cm"),
  graphwidth = unit(7, "cm"),
  txt_gp = fpTxtGp(label = gpar(cex = 1.2),
    title = gpar(cex = 1.2),
    ticks = gpar(cex = 1.2),
    xlab = gpar(cex = 1.2)))
Receiver Operating Characteristic (ROC) Analysis

```
optimal.cutpoint.Youden <- optimal.cutpoints(X = "Median", status = "ThreeDoses",
tag.healthy = 0, methods = "Youden", data = data, pop.prev = NULL,
control = control.cutpoints(), ci.fit = FALSE, conf.level = 0.95, trace = FALSE)

summary(optimal.cutpoint.Youden)
plot(optimal.cutpoint.Youden)
```

Decision Tree Development

```
caret.control <- trainControl(method = "repeatedcv",
   number = 5,
   repeats = 3)

rpart.grid <- expand.grid(.cp=0.008)

rpart.cv <- train(GROUP ~ Median + WEIGHTper + PLATELETper,
data = data,
method = "rpart",
trControl = caret.control, tuneGrid=rpart.grid,
tuneLength = 15)

rpart.best <- rpart.cv$finalModel
rpart.best
rpart.rules(rpart.best)
rpart.cv$sresample
prp(rpart.best, type = 0, extra = 1, under = TRUE)

#Conditional Inference Tree
ct<-.ctree(GROUPnew~.,data = train.data, control=ctree_control(maxdepth=4, minbucket = 50))
plot(ct)
```