

# **Statistical Methods in the Analysis of Randomized Control Trials: Applications to Pre-Exposure Prophylaxis for HIV Prevention and a Community Engagement Intervention in East Baltimore**

by

Claire Frances Ruberman

A dissertation submitted to The Johns Hopkins University  
in conformity with the requirements for the degree of  
Doctor of Philosophy

Baltimore, Maryland

July 2018

# Abstract

This work focuses on two examples of randomized control trials- a set of Phase II and Phase III clinical trials evaluating pre-exposure prophylaxis for HIV prevention and a cluster randomized trial assessing a community engagement intervention developed to improve health outcomes in Baltimore. We discuss the unique statistical challenges each trial raises, and we present our application of different methods that address them. We conclude by discussing the commonalities between the different trials, ongoing methodological challenges, and future directions for this work.

# Thesis Committee

## Primary Readers

Michael A. Rosenblum (Primary Advisor)  
Associate Professor  
Department of Biostatistics  
Johns Hopkins Bloomberg School of Public Health

Craig W. Hendrix  
Professor  
Departments of Medicine & Pharmacology and Molecular Sciences  
Johns Hopkins University School of Medicine

Caitlin E. Kennedy  
Associate Professor  
Departments of International Health & Health, Behavior and Society  
Johns Hopkins Bloomberg School of Public Health

Elizabeth L. Ogburn  
Assistant Professor  
Department of Biostatistics  
Johns Hopkins Bloomberg School of Public Health

## **Alternate Readers**

Ciprian M. Crainiceanu

Professor

Department of Biostatistics

Johns Hopkins Bloomberg School of Public Health

Danielle German

Associate Professor

Department of Health, Behavior and Society

Johns Hopkins Bloomberg School of Public Health

# Acknowledgments

To the Department of Biostatistics, faculty, staff, students, and postdocs, thank you for making this incredible experience possible. I cannot imagine a more supportive, collaborative, and stimulating environment in which to do a PhD.

To Michael Rosenblum, thank you for your guidance, patience, and encouragement. Words cannot express how grateful I am for your generosity with your time, kindness, and enthusiasm for research. Thank you for being an incredible role model of both a statistician and person.

To Craig Hendrix, Betsy Ogburn, Caitlin Kennedy, Danielle German, and Ciprian Crainiceanu, thank you for serving on my thesis committee. I truly appreciate your availability and valuable perspectives and suggestions. Thank you as well to Bryan Lau and Jeff Leek for being on my preliminary oral exam committee and asking insightful questions that have challenged my thinking.

I am so thankful to have been able to work with and learn from so many incredible statisticians and scientists. Thank you to Michael, Craig, Jon Steingrimsen, Gary Rosner, Rada Savic, and Katarina Vucicevic for all of the collaboration on the PrEP projects. Thank you for patiently helping me learn about your many areas

of expertise. To Albert Wu, Chidinma Ibe, Christine M. Weston, Lee Bone, Tony Boonyasai, Ja Alah-Ai Heughan, Sandra Hwang, Yanyan Lu, and Shuwen Liang, thank you for welcoming me onto the Baltimore CONNECT team and for all the meaningful work you're doing for our city. To Jeff Leek and Andrew Jaffe, thank you for the most enthusiastic introduction to genomics imaginable.

To Marie Diener-West, John McGready, Karen Bandeen-Roche, Leah Jager, Michael Griswold, and Daniel Obeng, thank you for the many teaching opportunities that have deeply enriched my time at Hopkins. Thank you for being fantastic models for both effective and caring teaching.

To Galit Alter, thank you for taking a chance on bringing a high schooler into your lab, for introducing me to the world of HIV research, for your continued mentorship, and for your inimitable energy and passion for science.

To my friends, thank you for making Baltimore a home to me. To Jack Fu and Leslie Myint, thank you for sharing this experience with me from start to finish. I am truly grateful for and in admiration of your enthusiasm and fearlessness, Jack, and your levelheadedness and strength, Leslie. To the Hopkins Marathon Team and Team Christopher Place, thank you for building communities that have extended far beyond running and modeling relentless endurance.

To Vitaly Lorman, thank you for your ceaseless confidence in me, your humor, and your perspective. Thank you for being an incredibly loving and supportive partner. Thank you, of course, to Banjo for all the joy you bring.

I would like to give my deepest thanks to my family. To my father, thank you for always answering my math questions with more questions and for listening to

far more than your fair share of practice presentations. To my mother, thank you for being my biggest cheerleader and pushing me to trust myself. To Sarah, thank you for continuing to ground and encourage me, for reining in my competitiveness, and for being the one I look up to the most. This work is dedicated to the loving memory of my grandfathers, Kenneth J. Borst and Dr. William Ruberman.

# Table of Contents

<b>Table of Contents</b>	<b>viii</b>
<b>List of Tables</b>	<b>xi</b>
<b>List of Figures</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	2
1.1.1 Pre-Exposure Prophylaxis for HIV Prevention . . . . .	2
1.1.2 Baltimore Community Organizations Neighborhood Network: Enhancing Capacity Together (CONNECT) partnership	4
1.2 Contributions . . . . .	5
<b>2 Estimating the Protective Effect of Longitudinal Drug Concentration in Pre-Exposure Prophylaxis for HIV Prevention</b>	<b>12</b>
2.1 Introduction . . . . .	12
2.2 Two Pre-Exposure Prophylaxis Randomized Trials . . . . .	17



2.3	Longitudinal Data Structure and Target of Estimation . . . . .	19
2.4	Targeted Maximum Likelihood Estimation for Longitudinal Data Structures in Nested Case-Cohort and Case-Control Studies. . . . .	22
2.5	Results . . . . .	25
2.5.1	Estimates using Longitudinal TMLE . . . . .	25
2.5.2	Comparisons to Previous Analyses . . . . .	27
2.6	Discussion . . . . .	30
2.7	Acknowledgements . . . . .	35
<b>3</b>	<b>Pharmacokinetic and Pharmacodynamic Modeling of Tenofovir</b>	<b>54</b>
3.1	Introduction . . . . .	54
3.1.1	Pharmacodynamic Modeling . . . . .	56
3.1.2	Pharmacokinetic Studies of PrEP . . . . .	57
3.2	Methods . . . . .	58
3.2.1	The Pharmacokinetic Model . . . . .	59
3.2.2	Longitudinal Targeted Maximum Likelihood Estimation . . . . .	63
3.3	Results . . . . .	64
3.3.1	Pharmacokinetic Model Results . . . . .	64
3.3.2	LTMLE Analysis Results . . . . .	65
3.3.3	Comparison of Results Using Raw vs. Modeled Concentrations . . . . .	66
3.4	Discussion . . . . .	67

3.4.1	Future Directions . . . . .	69
<b>4</b>	<b>Design and Analysis of a Balanced, Covariate-Adjusted Cluster Randomized Trial: the Baltimore CONNECT Project</b>	<b>79</b>
4.1	Introduction . . . . .	80
4.1.1	Cluster randomized trials for health services research . . . .	80
4.1.2	Baltimore CONNECT Project . . . . .	84
4.2	Methods . . . . .	88
4.2.1	Constrained Randomization . . . . .	88
4.2.2	Randomization Inference . . . . .	90
4.2.3	Notation and Justification for Randomization Inference . . .	93
4.3	Results . . . . .	95
4.4	Comparison with Model-Based Analysis . . . . .	96
4.4.1	Methods . . . . .	96
4.4.2	Results . . . . .	98
4.5	Discussion . . . . .	98
4.6	Acknowledgements . . . . .	100
4.7	Appendix: Confidence Interval Construction . . . . .	102
<b>5</b>	<b>Conclusion</b>	<b>109</b>
	<b>CV</b>	<b>115</b>

# List of Tables

2.1	Number of concentration substudy participants in the Partners PrEP and VOICE studies by treatment, concentration regimen followed, and seroconversion outcome. . . . .	41
2.2	Results from the longitudinal TMLE analyses of Partners PrEP and VOICE . . . . .	42
2.3	Results from the modified intention-to-treat analyses of Partners PrEP and VOICE . . . . .	43
2.4	Risk behaviors in the VOICE gel arm over the first 18 study months.	44
2.5	Comparison of the number of VOICE gel arm substudy participants by concentration regimen followed and seroconversion outcome using data from the first 6, 12, and 18 months. . . . .	45
2.6	Comparison of results from the longitudinal TMLE analyses of the VOICE gel arm using data from the first 6, 12, and 18 months. . . .	46

3.1	Results from the longitudinal TMLE analyses of Partners PrEP and VOICE using average daily concentrations from the PK model as the exposures. . . . .	75
4.1	Results from the difference-in-differences analyses using randomization inference. . . . .	101
4.2	Results from the difference-in-differences analyses using a two-stage model. . . . .	101

# List of Figures

2.1	Distribution of plasma concentration measurements in the first 18 months by study. . . . .	36
2.2	Comparison of results analyzing the Partners PrEP study using the longitudinal TMLE to results from previous methods. . . . .	37
2.3	Comparison of results using the longitudinal TMLE to results analyzing the VOICE oral arms from previous methods. . . . .	38
2.4	Comparison of results using the longitudinal TMLE to results analyzing the VOICE gel arm from previous methods. . . . .	39
2.5	Concentration measurements from the first three semi-annual visits by HIV status in the VOICE gel and oral arms. . . . .	40
3.1	The two compartment population-level PK model used to model tenofovir plasma concentrations from oral dosing. . . . .	72
3.2	Comparison of the distribution of raw plasma tenofovir concentrations from stored samples and estimated daily concentrations from the pharmacokinetic model. . . . .	73

3.3	Comparison of average daily tenofovir levels from the pharmacokinetic model to raw plasma tenofovir concentrations from stored samples. . . . .	74
-----	---	----

# Chapter 1

## Introduction

Randomized controlled trials have long been considered the “gold standard” for evaluating the effect of an exposure on an outcome of interest. The particularities of a trial design depend upon the questions the trial seeks to address, but, broadly speaking, either individuals or clusters of individuals are randomized to treatment groups. One (or more) treatment group serves as the control (or comparison); study participants in that group receive either a placebo, a standard-of-care treatment, or no intervention, depending upon the context. With proper implementation, participants are allocated to treatment groups in such a way that the groups vary only by their exposure, preventing selection bias. However, study participants may not receive the treatment to which they were assigned, or, treatment may be received at inconsistent levels within a study arm, for a number of reasons including low or non-adherence, treatment discontinuation due to study protocol, study dropout, or spillover between different treatment arms. These potential problems can all bias estimates of the treatment effect if not properly accounted for.

This thesis is centered around a collection of randomized control trials. In Chapter 2 and Chapter 3, we analyze of several clinical trials designed to assess the efficacy of pre-exposure prophylaxis (PrEP) for HIV prevention. In Chapter 4, we present the design and analysis of a cluster randomized trial that was part of the broader Baltimore Community Organizations Neighborhood Network: Enhancing Capacity Together (CONNECT) partnership (Wu, 2018). In the remainder of this chapter, we provide background for the two applications and summarize our main contributions to each. In Chapters 2 to 4, we provide the details of our analyses. Finally, we conclude in Chapter 5 with a discussion of the connections between the two projects and future directions for our work.

## **1.1 Background**

### **1.1.1 Pre-Exposure Prophylaxis for HIV Prevention**

Pre-exposure prophylaxis (PrEP) using the antiretroviral drugs tenofovir disoproxil fumarate (TDF) alone or in combination with emtricitabine (FTC) is a growing strategy to prevent HIV infection worldwide. Since 2015, the World Health Organization has recommended oral PrEP for all populations at “substantial risk” of HIV infection (World Health Organization, 2015). In addition to oral dosing of PrEP (either TDF alone or the combination TDF/FTC), vaginal gel containing tenofovir has been recognized as an important method of preventing HIV infection in women.

Placebo controlled, randomized clinical trials assessing the efficacy of PrEP, in



both oral and topical formulations, have had varying results. A number of trials have demonstrated a high protective effect of daily oral PrEP among the following at-risk populations: heterosexual couples where one partner is HIV positive and the other negative in Kenya and Uganda (Baeten et al., 2012); men who have sex with men in Peru, Ecuador, Brazil, the United States, South Africa and Thailand (Grant et al., 2010); injection drug users in Thailand (Choopanya et al., 2013); and heterosexual men and women in Botswana (Thigpen et al., 2012). Several trials have found “on-demand PrEP,” wherein drug is administered both shortly before and after sex, to be effective in preventing HIV infection. These studies have focused on tenofovir gel taken among South African women (Karim et al., 2010) and oral TDF/FTC among men who have sex with men (Molina et al., 2015).

In contrast, two studies of African women concluded no protective effect of PrEP. The first, FEM-PrEP, evaluated oral PrEP in South Africa, Kenya and Tanzania (Van Damme et al., 2012). The second, VOICE, included both oral and gel formulations in South Africa, Uganda, and Zimbabwe (Marrazzo et al., 2015). In both studies one or more treatment arms were stopped early due to futility. Low adherence to assigned study products is thought to be a large contributor as to why the studies failed to conclude any effectiveness of PrEP (Corneli et al., 2014; van der Straten et al., 2014; Dai et al., 2015).

Beyond clinical trials, PrEP implementation and effectiveness has been evaluated in a number of pragmatic trials. These include open-label extensions of many of the phase II and III clinical trials described above (Grant et al., 2014; Baeten et al., 2016). Additional demonstration projects, many of which are ongoing or planned,

have focused on PrEP use in transgender women and men who have sex with men (Hosek et al., 2013; Cohen et al., 2015; McCormack et al., 2016; Hojilla et al., 2016; Hosek et al., 2017; Mahon, 2018), female sex workers (Kyongo et al., 2016; Cowan et al., 2016), and adolescent girls and young women (Celum et al., 2015; Cowan et al., 2016; *EMPOWER Consortium Demonstration Project* 2016; *USAID Announces Microbicide Awards* 2016). An ongoing challenge in PrEP implementation is to understand which PrEP drug formulations and dosing strategies provide sufficiently high protection against HIV infection in a broad range of at-risk populations.

### **1.1.2 Baltimore Community Organizations Neighborhood Network: Enhancing Capacity Together (CONNECT) partnership**

The second application, although focused upon improving health quality in Baltimore, Maryland, is rooted in global health delivery. The Baltimore CONNECT partnership was developed using the World Health Organization's African Partnerships for Patient Safety Community Engagement (ACE) Approach (Syed et al., 2009; Ibe et al., 2018; Wu, 2018). Adopting the ACE framework, the goal of the study was to bridge existing social and medical services, between community-based organizations (CBOs) in East Baltimore and The Johns Hopkins Health System (JHHS) (Ibe et al., 2018). By facilitating referrals both between CBOs and between CBOs and the JHHS, the study sought to improve the health of East Baltimore residents, in particular those identified as at high-risk for hospitalization or trips to the emergency room. The details of the study framework, grounded in the concept of "reverse innovation," are given in Ibe et al. (2018).

## 1.2 Contributions

Despite the many clinical trials assessing the efficacy of PrEP, much uncertainty remains about the relationship between drug concentration in blood plasma and the reduction in HIV risk. Key challenges in estimating this relationship include the following: data on drug concentrations are relatively sparse and are collected via case-control or case-cohort sampling within the active treatment arm(s); adherence to assigned study drug may vary by study visit; and participants may miss study visits or be lost to follow up. To address these challenges, we apply targeted maximum likelihood estimation (TMLE) (Van der Laan and Rubin, 2006; Rose and van der Laan, 2011; Van der Laan and Gruber, 2012) to estimate the protective effect of drug concentration against HIV infection using longitudinal data from two randomized, placebo-controlled clinical trials of daily PrEP: Partners PrEP and VOICE. In Chapter 2 we present analyses using raw, measured concentration data. We extend our analysis in Chapter 3 by integrating pharmacokinetic models developed by Vucicevic, Savic, and Hendrix (In Preparation, 2018) using data from the same set of clinical trials.

In Chapter 4, we describe the design, implementation, and analysis of a cluster randomized trial with three key features: 1) constrained randomization to balance groups on key baseline variables, 2) randomization inference to handle a relatively small number of clusters with potential spillover between them, and 3) adjustment for prognostic baseline variables that were not included in 1) so as to improve precision. Despite the potential benefits of these three techniques, to the best of our

knowledge they have not been utilized simultaneously for a cluster randomized trial.

# References

- Wu A. W., Hwang S. Weston C. M. Ibe C. Boonyasai R. Bone L. Basu L. Lief I. Gentry J. Purnell L. Rosenblum M. for the Baltimore CONNECT study team. (In Press) (2018). "Baltimore CONNECT: a randomized trial to build partnership between community organizations and a local health system". In: *Progress in Community Health Partnerships: Research, Education, and Action*.
- World Health Organization (2015). "Treat all people living with HIV, offer antiretrovirals as additional prevention choice for people at 'substantial' risk". In: URL: <http://www.who.int/mediacentre/news/releases/2015/hiv-treat-all-recommendation/en>.
- Baeten, Jared M, Deborah Donnell, Patrick Ndase, Nelly R Mugo, James D Campbell, Jonathan Wangisi, Jordan W Tapper, Elizabeth A Bukusi, Craig R Cohen, Elly Katabira, et al. (2012). "Antiretroviral prophylaxis for HIV prevention in heterosexual men and women". In: *New England Journal of Medicine* 367.5, pp. 399–410.
- Grant, Robert M, Javier R Lama, Peter L Anderson, Vanessa McMahan, Albert Y Liu, Lorena Vargas, Pedro Goicochea, Martín Casapía, Juan Vicente Guanira-Carranza, Maria E Ramirez-Cardich, et al. (2010). "Preexposure chemoprophylaxis for HIV prevention in men who have sex with men". In: *New England Journal of Medicine* 363.27, pp. 2587–2599.
- Choopanya, Kachit, Michael Martin, Pravan Suntharasamai, Udomsak Sangkum, Philip A Mock, Manoj Leethochawalit, Sithisat Chiamwongpaet, Praphan Kitisin, Pitinan Natrujirote, Somyot Kittimunkong, et al. (2013). "Antiretroviral prophylaxis for HIV infection in injecting drug users in Bangkok, Thailand (the Bangkok Tenofovir Study): a randomised, double-blind, placebo-controlled phase 3 trial". In: *The Lancet* 381.9883, pp. 2083–2090.
- Thigpen, Michael C, Poloko M Kebaabetswe, Lynn A Paxton, Dawn K Smith, Charles E Rose, Tebogo M Segolodi, Faith L Henderson, Sonal R Pathak, Fatma

- A Soud, Kata L Chillag, et al. (2012). "Antiretroviral preexposure prophylaxis for heterosexual HIV transmission in Botswana". In: *New England Journal of Medicine* 367.5, pp. 423–434.
- Karim, Quarraisha Abdool, Salim S Abdool Karim, Janet A Frohlich, Anneke C Grobler, Cheryl Baxter, Leila E Mansoor, Ayesha BM Kharsany, Sengeziwe Sibeko, Koleka P Mlisana, Zaheen Omar, et al. (2010). "Effectiveness and safety of tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women". In: *Science* 329.5996, pp. 1168–1174.
- Molina, Jean-Michel, Catherine Capitant, Bruno Spire, Gilles Pialoux, Christian Chidiac, I Charreau, and JF Delfraissy (2015). "On demand PrEP with oral TDF-FTC in MSM: results of the ANRS Ipergay trial". In: *Conference on Retroviruses and Opportunistic Infections*. Vol. 2015.
- Van Damme, Lut, Amy Corneli, Khatija Ahmed, Kawango Agot, Johan Lombaard, Saidi Kapiga, Mookho Malahleha, Fredrick Owino, Rachel Manongi, Jacob Onyango, et al. (2012). "Preexposure prophylaxis for HIV infection among African women". In: *New England Journal of Medicine* 367.5, pp. 411–422.
- Marrazzo, Jeanne M, Gita Ramjee, Barbra A Richardson, Kailazarid Gomez, Nyaradzo Mgodi, Gonasagrie Nair, Thesla Palanee, Clemensia Nakabiito, Ariane van der Straten, Lisa Noguchi, et al. (2015). "Tenofovir-based preexposure prophylaxis for HIV infection among African women". In: *New England Journal of Medicine* 372.6, pp. 509–518.
- Corneli, Amy L, Jennifer Deese, Meng Wang, Doug Taylor, Khatija Ahmed, Kawango Agot, Johan Lombaard, Rachel Manongi, Saidi Kapiga, Angela Kashuba, et al. (2014). "FEM-PrEP: adherence patterns and factors associated with adherence to a daily oral study product for pre-exposure prophylaxis". In: *Journal of Acquired Immune Deficiency Syndromes* (1999) 66.3, p. 324.
- van der Straten, Ariane, Ashley Mayo, Elizabeth R Brown, K Rivet Amico, Helen Cheng, Nicole Laborde, Jeanne Marrazzo, and Kristine Torjesen (2014). "Perceptions and Experiences with the VOICE Adherence Strengthening Program (VASP) in the MTN-003 Trial". In: *AIDS and Behavior* 19.5, pp. 770–783.
- Dai, James Y, Craig W Hendrix, Barbra A Richardson, Cliff Kelly, Mark Marzinke, Z Mike Chirenje, Jeanne M Marrazzo, and Elizabeth R Brown (2015). "Pharmacological Measures of Adherence and Risk of HIV Acquisition in the VOICE Study". In: *Journal of Infectious Diseases*, jiv333.
- Grant, Robert M, Peter L Anderson, Vanessa McMahan, Albert Liu, K Rivet Amico, Megha Mehrotra, Sybil Hosek, Carlos Mosquera, Martin Casapia, Orlando

- Montoya, et al. (2014). "Uptake of pre-exposure prophylaxis, sexual practices, and HIV incidence in men and transgender women who have sex with men: a cohort study". In: *The Lancet Infectious Diseases* 14.9, pp. 820–829.
- Baeten, J, Renee Heffron, Lara Kidoguchi, Nelly Mugo, Elly Katabira, Elizabeth Bukusi, S Asiimwe, J Morton, K Ngure, N Bulya, et al. (2016). "Integrated delivery of PrEP and ART results in sustained near elimination of HIV transmission in African HIV serodiscordant couples: final results from The Partners Demonstration Project". In: *21st International AIDS Conference (AIDS 2016)*, pp. 18–22.
- Hosek, Sybil, George Siberry, Margo Bell, Michelle Lally, Bill Kapogiannis, Keith Green, M Isa Fernandez, Brandy Rutledge, Jaime Martinez, Robert Garofalo, et al. (2013). "Project PrEPare (ATN082): The acceptability and feasibility of an HIV pre-exposure prophylaxis (PrEP) trial with young men who have sex with men (YMSM)". In: *Journal of Acquired Immune Deficiency Syndromes (1999)* 62.4.
- Cohen, Stephanie E, Eric Vittinghoff, Oliver Bacon, Susanne Doblecki-Lewis, Brian S Postle, Daniel J Feaster, Tim Matheson, Nikole Trainor, Robert Wilder Blue, Yannine Estrada, et al. (2015). "High interest in pre-exposure prophylaxis among men who have sex with men at risk for HIV-infection: baseline data from the US PrEP demonstration project". In: *Journal of Acquired Immune Deficiency Syndromes (1999)* 68.4, p. 439.
- McCormack, Sheena, David T Dunn, Monica Desai, David I Dolling, Mitzy Gafos, Richard Gilson, Ann K Sullivan, Amanda Clarke, Iain Reeves, Gabriel Schembri, et al. (2016). "Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial". In: *The Lancet* 387.10013, pp. 53–60.
- Hojilla, J Carlo, Kimberly A Koester, Stephanie E Cohen, Susan Buchbinder, Deawodi Ladzekpo, Tim Matheson, and Albert Y Liu (2016). "Sexual behavior, risk compensation, and HIV prevention strategies among participants in the San Francisco PrEP demonstration project: a qualitative analysis of counseling notes". In: *AIDS and Behavior* 20.7, pp. 1461–1469.
- Hosek, Sybil G, Bret Rudy, Raphael Landovitz, Bill Kapogiannis, George Siberry, Brandy Rutledge, Nancy Liu, Jennifer Brothers, Kathleen Mulligan, Gregory Zimet, et al. (2017). "An HIV preexposure prophylaxis demonstration project and safety study for young MSM." In: *Journal of Acquired Immune Deficiency Syndromes (1999)* 74.1, pp. 21–29.

- Mahon, Caitlin (2018). 'PrEP Brazil' a success for gay men and trans women. URL: <https://www.avert.org/news/'prep-brazil'-success-gay-men-and-trans-women>.
- Kyongo, Jordan, Michael Kiragu, Serah Nduta, Robinson Karuga, Reuben Mulwa, Peris Mwimali, Mercy Seii, George Kimathi, Jane Thiomi, Dorothy Njeru, et al. (2016). "Oral PrEP for HIV Prevention: Early Lessons from a PrEP Demonstration Project in Kenya". In: *AIDS Research and Human Retroviruses*. Vol. 32, pp. 286–286.
- Cowan, Frances M, Sinead Delany-Moretlwe, Eduard J Sanders, Nelly R Mugo, Fernand A Guedou, Michel Alary, Luc Behanzin, Owen Mugurungi, and Linda-Gail Bekker (2016). "PrEP implementation research in Africa: what is new?" In: *Journal of the International AIDS Society* 19.7S6.
- Celum, Connie L, Sinead Delany-Moretlwe, Margaret McConnell, Heidi Van Rooyen, Linda-Gail Bekker, Ann Kurth, Elizabeth Bukusi, Chris Desmond, Jennifer Morton, and Jared M Baeten (2015). "Rethinking HIV prevention to prepare for oral PrEP implementation for young African women". In: *Journal of the International AIDS Society* 18.4S3.
- EMPOWER Consortium Demonstration Project (2016). URL: <https://www.avac.org/trial/empower-consortium-demonstration-project>.
- USAID Announces Microbicide Awards (2016). URL: <https://www.usaid.gov/what-we-do/global-health/hiv-and-aids/technical-areas/microbicides-providing-hiv-prevention-option/2015-microbicide-awards>.
- Syed, Shamsuzzoha B, R Gooden, J Storr, JD Hightower, P Rutter, S Nejad Bagheri, A Lardner, E Kelley, and D Pittet (2009). "African partnerships for patient safety: a vehicle for enhancing patient safety across two continents.[corrected]." In: *World Hospitals and Health Services* 45.4, pp. 24–27.
- Ibe, Chidinma A, Lopa Basu, Rachel Gooden, Shamsuzzoha B Syed, Viva Dadwal, Lee R Bone, Patti L Ephraim, Christine M Weston, Albert W Wu, J Barron, et al. (2018). "From Kisiizi to Baltimore: cultivating knowledge brokers to support global innovation for community engagement in healthcare". In: *Globalization and Health* 14.1, p. 19.
- Van der Laan, Mark J and Daniel Rubin (2006). "Targeted maximum likelihood learning". In: *The International Journal of Biostatistics* 2.1.
- Rose, Sherri and Mark J van der Laan (2011). "A targeted maximum likelihood estimator for two-stage designs". In: *The International Journal of Biostatistics* 7.1, pp. 1–21.



- Van der Laan, Mark J and Susan Gruber (2012). "Targeted minimum loss based estimation of causal effects of multiple time point interventions". In: *The International Journal of Biostatistics* 8.1.
- Vucicevic, Katarina, Radojka M Savic, and Craig W. Hendrix (In Preparation, 2018). "A pharmacokinetic-pharmacodynamic model of tenofovir for HIV pre-exposure prophylaxis".

## Chapter 2

# Estimating the Protective Effect of Longitudinal Drug Concentration in Pre-Exposure Prophylaxis for HIV Prevention

Reproduced with permission from:

Ruberman, Claire F., Jon A. Steingrimsdottir, Craig W. Hendrix, and Michael Rosenblum (2018). “Estimating the Protective Effect of Longitudinal Drug Concentration in Pre-Exposure Prophylaxis for HIV Prevention.” In preparation.

### 2.1 Introduction

Pre-exposure prophylaxis (PrEP) using the antiretroviral drugs tenofovir disoproxil fumarate (TDF) alone or in combination with emtricitabine (FTC) is a growing strategy for preventing HIV infection in vulnerable populations. In 2015, the World

Health Organization recommended that PrEP be used by all individuals at high risk of HIV infection as part of a comprehensive prevention strategy (World Health Organization, 2015). Large PrEP evaluation and demonstration projects have been and continue to be implemented worldwide (Krakower and Mayer, 2015; Baeten et al., 2016a; Lal et al., 2017; Hoagland et al., 2017). Clinical trials evaluating the prevention effect of PrEP have had varying, though mostly positive, results (Van der Straten et al., 2012; Baeten and Grant, 2013; Fonner et al., 2016; Spinner et al., 2016).

We focus on the following two randomized, placebo-controlled clinical trials of daily PrEP: The Partners Preexposure Prophylaxis (PrEP) Study (Baeten et al., 2012) and Vaginal and Oral Interventions to Control the Epidemic (VOICE) (Marrazzo et al., 2015). Both trials evaluate oral TDF and/or combination TDF/FTC in preventing HIV seroconversion. The VOICE study also includes a tenofovir-containing vaginal microbicide gel treatment arm. The Partners PrEP trial demonstrated effectiveness of PrEP based on intention to treat analyses, reporting relative risk reductions (one minus relative risk) of 67% (95% confidence interval: 44% to 81%) and 75% (95% CI: 55% to 87%) in the TDF and TDF/FTC arms, respectively, compared to the placebo arm (Baeten et al., 2012). In contrast, the VOICE trial reported a relative risk reduction in the harmful direction of  $-4.4\%$  (95% CI:  $-49\%$  to  $27\%$ ) for the oral TDF/FTC arm. Both the oral TDF and tenofovir gel arms were stopped early for futility with estimated relative risk reductions of  $-49\%$  (95% CI:  $-129\%$  to  $-3\%$ ) and  $15\%$  (95% CI:  $-21\%$  to  $39\%$ ), respectively (Marrazzo et al., 2015). The differences in PrEP effectiveness across trials have received substantial interest and

have partially been attributed to differences in medication adherence, population, and mode of transmission (Van der Straten et al., [2012](#); Dai et al., [2015](#)).

We estimate the protective effect of setting plasma drug concentrations of tenofovir above or below different thresholds over a period of eighteen months. Learning about these effects may help inform the formulation of effective dosage strategies for PrEP. It may also help determine when an adherence intervention is needed to boost drug concentrations to a protective level. Identifying a threshold for protective concentration levels may also be useful in setting target concentrations for new modes of PrEP delivery, such as tenofovir gel rectal microbicides (McGowan, [2014](#)), dapivirine based vaginal rings (Baeten et al., [2016b](#)), and long-acting injectable formulations of antiretroviral drugs (Margolis et al., [2017](#)).

Our analyses focus on substudies within the active treatment arms (the oral TDF and TDF/FTC arms of Partners PrEP and VOICE and the 1% tenofovir gel arm of VOICE), in which PrEP drug concentrations were measured from stored plasma samples. In both studies, plasma samples were taken quarterly from every study participant and stored. Additionally samples were taken and stored for all cases at their time of seroconversion. Study participants were selected for the concentration substudies through case-cohort designs. In each of the active treatment arms, plasma concentrations were analyzed from the stored samples at multiple visits for all participants who seroconverted (cases) up through the visit where seroconversion was detected. Concentrations were also measured from a randomly selected subset of participants (cohort) at the same set of visits until dropout or study termination. In the Partners PrEP substudy, plasma samples from

the 1, 3, and 6 month visits and every 6 months thereafter were assayed. All of the plasma samples from the VOICE substudy were assayed. We analyze the Partners PrEP and VOICE studies separately, but in each study we treat the oral TDF and TDF/FTC arms as a single arm in our analyses because the dose of TDF given in each arm was the same (at 300 mg daily).

Recent work has estimated the causal effect of drug plasma concentration using data from the above trials, including Dai et al. (2013) and Murnane et al. (2015). These analyses infer the protective effect of assignment to PrEP among the group of participants who would have a specified level of drug concentration (at either a fixed study visit or at any study visit) if assigned to one of the active arms. We use the longitudinal TMLE of Van der Laan and Gruber (2012) to estimate the protective effect against HIV infection of setting plasma drug concentration levels to be above the limit of quantification (0.31 ng/mL) at each of the 6, 12, and 18 month study visits. Rather than using one measure to characterize a participant's concentration level throughout the study, our analyses incorporate time-varying concentration levels for study participants.

We account for reported instances of unprotected sex as a potential time-varying confounder of the relationship between drug concentration and risk of HIV infection. Golub, Operario, and Gorbach (2010), Liu et al. (2013), and Calabrese and Underhill (2015) found suggestive evidence of such an association; they identified groups of individuals who adhere to PrEP regimens and do not partake in high risk behaviors. Additionally, decreases were observed in self-reported sexual risk

behaviors over the course of the Partners PrEP trial (Baeten et al., 2012). Furthermore, there was evidence in the VOICE study that levels of medication adherence may have changed over time (Van der Straten et al., 2014).

Bias due to time-dependent confounding cannot generally be removed by simple regression adjustment (Hernán, Brumback, and Robins, 2000). The longitudinal TMLE analysis accounts for measured, time-varying confounders. The longitudinal TMLE also accounts for informative censoring. It builds upon ideas from Van der Laan and Rubin (2006), the sequential regression estimators of Robins (2000) and Bang and Robins (2005), and the general semiparametric efficiency theory of Robins and Rotnitzky, 1992. The assumptions are discussed in Section 4.2.

We implement our analyses using the LTMLE (Longitudinal TMLE) R package (Schwab et al., 2016). We apply weighting techniques from Rose and van der Laan (2011) to account for the concentration substudy designs. Longitudinal TMLE has previously been used to analyze the causal effects of, for example, antiretroviral therapies on drug resistance and survival among HIV-infected patients (Stitelman, De Gruttola, and van der Laan, 2012), warfarin on stroke or death (Brooks et al., 2013), hepatitis C virus clearance on end-stage liver-disease free survival (Schnitzer et al., 2014b), breastfeeding on gastrointestinal infection in infants (Schnitzer et al., 2014a), early interventions on preventing child obesity (Decker et al., 2014), and task shifting with HIV treatment programs on patient outcomes in East Africa (Tran et al., 2016). Rose (2011) applied the method of Rose and van der Laan (2011) in conjunction with the superlearner machine learning algorithm to estimate a regression function. However, the quantity estimated was not a causal effect as

is the goal of our estimation problem, which has additional challenges such as handling time-dependent confounding. To the best of our knowledge, this is the first application of the Rose and van der Laan (2011) method to estimate a causal effect. This is also the first application to combine this method with the longitudinal TMLE of Van der Laan and Gruber (2012).

Section 2.2 describes the two PrEP trials. Section 2.3 details the structure of the data and our target of estimation, and Section 4.2 discusses the TMLE procedure for estimating longitudinal effects of plasma concentration on HIV risk using data from case-cohort substudies. Our results from the two trials and comparisons to previous results are given in Section 2.5. Limitations of our methods and potential reasons for differences between our results and those from related work are discussed in Section 2.6.

## **2.2 Two Pre-Exposure Prophylaxis Randomized Trials**

Partners PrEP was a phase III, randomized, double-blinded, placebo-controlled trial that assessed the use of daily oral TDF or TDF/FTC to prevent HIV infection among 4747 heterosexual serodiscordant couples (one HIV positive the other HIV negative) in Kenya and Uganda (Baeten et al., 2012). Beginning in 2009 and lasting through 2011, the HIV negative partner from each couple was randomized to one of daily oral TDF, TDF/FTC, or placebo and followed monthly for three years or until seroconversion. We restrict our attention to a case-control study consisting of 17 and 13 cases in the TDF and FTC/TDF arms and a randomly selected cohort of 200 participants (100 from each arm), for whom plasma concentrations were

measured. Two of the study participants selected into the cohort seroconverted over the course of the study and two more were lost to followup. Additionally, one case in the FTC/TDF arm, who seroconverted at 23 months, did not have plasma samples available and was excluded from the analysis.

The VOICE study was a phase IIB, randomized, placebo-controlled assessing daily oral TDF, oral TDF/FTC, or tenofovir 1% vaginal gel among 5029 heterosexual women in South Africa, Uganda, and Zimbabwe. The primary endpoint was HIV-1 infection, with HIV testing performed monthly (Marrazzo et al., 2015). The study began in September 2009 and continued to enroll women through June 2011. Although follow-up was planned until June 2012, the Data and Safety Monitoring Board (DSMB) determined on September 16, 2011 that oral TDF tablets were safe but not effective in VOICE and recommended discontinuing their use (*MTN Statement on Decision to Discontinue Use of Oral Tenofovir Tablets in VOICE, a Major HIV Prevention Study in Women 2011*). The DSMB made a similar determination for tenofovir gel on November 17, 2011 (*MTN Statement on Decision to Discontinue Use of Tenofovir Gel in VOICE, a Major HIV Prevention Study in Women 2011*). We analyze data from a case-cohort concentration substudy within the oral TDF, oral TDF/FTC, and TFV gel arms consisting all seroconverters who do not have acute HIV infection at enrollment and returned for at least one visit in the six months after enrollment and a randomly sampled cohort. There were 741 participants in the case-cohort substudy assigned to either oral TDF or oral TDF/FTC, 113 of whom were cases. Additionally, there were 669 participants in the case-cohort substudy assigned to the TFV gel arm, including 61 seroconverters.



We restricted our analyses to the first 18 months of each study due to data sparsity at later visits. Only 3 of the 29 seroconversions in the Partners PrEP analysis data and 3 of 113 seroconversions in the VOICE analysis data occurred after 18 months. The median follow-up times for participants in the substudies were 18 and 13 months for Partners PrEP and VOICE, respectively, so we considered the 18 month cutoff to be reasonable. Similar adjustments for sparsity are made when analyzing the effect of hepatitis C virus clearance using longitudinal TMLE in Schnitzer et al. (2014b). We included baseline data and semiannual concentration measurements in our analyses.

## 2.3 Longitudinal Data Structure and Target of Estimation

We analyzed concentrations measured from the 6, 12, and 18 month study visits. If a study participant missed one of these semiannual visits, we used, if available, their plasma concentration measurement from the visit three months prior. Participants who missed a semiannual visit and did not have a concentration measurement from the three months prior were right censored.

Our analyses use the longitudinal TMLE of Van der Laan and Gruber (2012) with case-cohort weighting as in Rose and van der Laan (2011), which we describe below. For each participant we observe a vector of baseline variables  $W$ . Visits after baseline are coded as  $j = 1, 2$ , and  $3$ , representing the 6 month, 12 month, and 18 month visits, respectively. Each observation at visit  $j$  consists of an indicator of censoring,  $C_j$ , and, if the participant has not been censored by that visit, an indicator,

$A_j$ , that plasma concentration exceeds a threshold, an indicator,  $L_j$ , of reported unprotected sex in the six months prior, and an indicator,  $Y_j$ , that the participant is HIV positive. We specify the following data ordering for each participant, which reflects the time ordering by visit:

$$(W, L_1, A_1, Y_1, C_2, L_2, A_2, Y_2, C_3, L_3, A_3, Y_3).$$

Note that setting concentration to follow a certain regimen may potentially affect time dependent covariates. For example a participant may be less likely to have unprotected sex if they do not take the drug.

Using the potential outcomes framework described in Rubin (1974) let  $Y^{(\bar{a})}$  denote a study participant's HIV status (not necessarily observed) at 18 months were they to attend their 6, 12, and 18 month study visits and follow treatment regimen  $\bar{a}$ . For example,  $Y^{(\bar{a}=1)}$  and  $Y^{(\bar{a}=0)}$  correspond to the participant's HIV statuses had their concentrations at each of their 6, 12, and 18 month study visits been maintained above and below the fixed threshold, respectively.

We define and compare the probability of seroconversion by 18 months under two treatment regimens: maintaining a plasma tenofovir concentration above or below a threshold at every visit (up to the final visit or until seroconversion is detected, whichever happens first) except for visits where drug was contraindicated due to safety concerns arising from pregnancy, breast-feeding, or laboratory abnormalities (Baeten et al., 2012). We estimate the marginal probability  $P(Y^{(\bar{a})} = 1)$  of seroconversion by 18 months in the treatment arm setting a treatment regimen  $\bar{a}$  of above or below a concentration threshold throughout the study (Van der Laan

and Gruber, 2011; Van der Laan and Gruber, 2012). This estimand differs from the target of estimation in principal stratification analysis by Dai et al. (2015): the average causal effect among compliers (the population whose concentration would exceed a threshold if assigned to the treatment arm). In contrast, our estimand is for the whole treatment arm population.

We use the limit of tenofovir quantification, 0.31 ng/mL, as our concentration threshold in these analyses. Table 2.1 describes how many participants in our data sets followed either of the two regimens described above. This is important because if few participants follow a given regimen, then it is more challenging to reliably estimate the corresponding probabilities  $P(Y^{(\bar{a})})$ . Study participants were right censored if more than six months passed between study visits. A relatively small percentage of participants followed either treatment regimen in its entirety.

The meaning of setting concentration above the limit of quantification depends on the distribution of concentration measurements conditioned on being above this limit and on the observed history. In particular, setting concentration to above the limit of quantification means setting concentration to a random draw from this conditional distribution. Though we are not able to give plots of this conditional distribution (which involves multiple variables and would be difficult to display), we present the marginal distributions of the observed concentrations in the Partners PrEP oral, VOICE oral, and VOICE gel arms in Figure 2.1. The distributions of quantifiable concentrations (74% of the observed concentrations in Partners PrEP and 28% and 20% of concentration measurements in the VOICE oral and gel arms, respectively) differ by study. Quantifiable concentration levels were highest

in the Partners PrEP oral arms, with a mean (median) of 93 ng/mL (81 ng/mL), as compared to 75 ng/mL (63 ng/mL) in the VOICE oral arms and 2 ng/mL (1 ng/mL) in the VOICE gel arm. The differences in concentrations observed in the VOICE oral and gel arms are consistent with pharmacokinetic studies examining different routes of dosing (Hendrix et al., 2013).

## 2.4 Targeted Maximum Likelihood Estimation for Longitudinal Data Structures in Nested Case-Cohort and Case-Control Studies.

Assumptions are needed in order to identify the parameter  $P(Y^{(\bar{a})} = 1)$  i.e., the probability of seroconversion by 18 months if all participants were set to follow  $\bar{a}$ . We make the following assumptions:

- Consistency: A participant's potential outcome under their observed plasma drug concentration history is their observed outcome.
- Positivity: At every visit, given any participant's covariate history up to that visit, a participant has a positive probability of continuing to follow a given treatment regimen.
- Time ordering: Each variable does not have a causal effect on those measured before.
- I.I.D. data: Each participant's longitudinal data vector (from Section 2.3) is an

independent, identically distributed draw from an unknown joint distribution  $P$ .

- No unmeasured confounders: There are no variables missing from the analysis that affect both (i) censoring and/or plasma drug concentration and (ii) seroconversion.

For a detailed description of the aforementioned assumptions see Stitelman, De Gruttola, and van der Laan (2012). As part of the time ordering assumption, we assume that a participant's concentration measurement at a given semiannual visits is representative of their drug concentration over the last six months. Additionally, in order to apply the two-stage sampling method of Rose and van der Laan (2011), we make the assumption that all study participants who seroconverted and were not lost to follow-up by 18 months are included as cases and that each cohort is a simple random sample of the remaining participants. This holds by design.

Fitting the TMLE estimator requires specifying models for censoring, drug concentration, time dependent covariates and seroconversion at each visit. All models used for the TMLE were main effects logistic regression models. In the Partners PrEP study, we adjusted for age and gender as baseline variables. In the VOICE study, we adjusted for age, marital status, and HSV-2 status as baseline variables. In both studies, we also adjusted for reported unprotected sex during the six months prior to each study visit as a potential time dependent confounder. These variables were selected because they were associated with plasma concentration levels and/or risk of seroconversion (Kiser et al., 2008; Haberer et al., 2013;

Donnell et al., 2014; Murnane et al., 2015; Burns, Hendrix, and Chaturvedula, 2015; Lu et al., 2016). Event sparsity limited the number of baseline variables that we were able to adjust for.

The TMLE estimator has several useful properties. The estimator is consistent if either (1) the models for concentration and censoring are correct, or (2) both the models for seroconversion and time dependent covariates are correct. This property is referred to as double robustness. In contrast, the principal stratification methods used in previous analyses are not doubly robust. In addition, if all the models required for the implementation of the TMLE estimator are correctly specified the estimator is asymptotically efficient (that is, has the smallest variance among a large class of estimators). Finally, the TMLE estimator is a substitution estimator, so it is guaranteed to produce estimators for probabilities that fall between zero and one.

For nested two-stage sampling designs, Rose and van der Laan (2011) show that incorporating appropriate weights to the TMLE fitting process leads to consistent estimators of treatment effects that are doubly robust and locally efficient. The weights are defined as the inverse probability of selection into the concentration substudy given the data from the main study. All cases in oral TDF and TDF/FTC arms are included in the concentration substudies, so all cases, even those who seroconverted after 18 months, have a weight of 1. In the Partners PrEP, VOICE oral, and VOICE gel substudies, the cohort weights are 15, 3, and 1.5, respectively. Each weight is calculated as the ratio of the number of study participants in the active treatment arm who test negative for HIV at all study visits or until censoring

to the the number of study participants in the concentration substudy who test negative for HIV at all study visits or until censoring.

## 2.5 Results

### 2.5.1 Estimates using Longitudinal TMLE

We present analyses of Partners PrEP and VOICE studies using the `ltmle` package (Schwab et al., 2016) in R (R Core Team, 2015). We set the concentration threshold used to define the regimens to be the limit of tenofovir plasma quantification, which was 0.31 ng per milliliter. This concentration is consistent with oral dosing within the last week (Donnell et al., 2014; Hendrix et al., 2016) and with gel dosing within the last two to three days (Hendrix et al., 2013).

Table 2.2 presents the estimated HIV incidences (infections per 100 person-years) in each of the Partners PrEP oral, VOICE oral, and VOICE gel arms under two different concentration regimens. The relative risk reduction,  $(1 - \text{relative risk}) \times 100\%$ , quantifies the effectiveness against HIV infection of setting concentration at the first three bi-annual visits to above the limit of quantification as compared to below the limit of quantification. Each 95% confidence interval was calculated using the bias-corrected and accelerated (BCa) bootstrap with 2,000 bootstrap replicates. The estimated probabilities of seroconversion vary between studies, which was expected due to differing baseline risks among the study populations (Hugonnet et al., 2002; Beyrer et al., 2012). In both studies, the estimates were similar whether we adjusted for only baseline variables or for baseline variables

and measures of time-varying risk. However, the confidence intervals for the estimated incidences and relative risk reductions were much wider when adjusting for time-varying risk. In the VOICE gel analysis, the confidence interval for the relative risk reduction when setting concentration to be quantifiable does not overlap zero when we adjust only for baseline variables; however, it does overlap zero when we adjust for baseline variables and time-varying risk.

We also considered concentration thresholds of 10 and 40 ng/mL, consistent with dosing in the last 2 to 3 days and 24 hours, respectively (Donnell et al., 2014). In the Partners PrEP study, the results of our analyses were very similar regardless of concentration threshold. It is likely that our analyses have low power to distinguish between the effects of setting concentration to be quantifiable versus setting concentration to be above 10 or 40 ng/mL because almost all of the concentration measurements above the limit of quantification exceeded 40 ng/mL. Of the quantifiable 381 concentration measurements in the Partners PrEP analysis, only 19 (5%) were between 0.31 and 10 ng/mL, and only 38 (10%) were between 10 and 40 ng/mL.

In the VOICE analyses, when we used as thresholds 10 or 40 ng/mL, the bootstrap procedure detected potential instability in the BCa confidence intervals. The instability likely occurred because very few study participants, in particular cases, maintained concentration regimens above the higher thresholds. As a consequence, the results for VOICE at thresholds of 10 and 40 ng/mL may be unreliable, and we do not report them.



## 2.5.2 Comparisons to Previous Analyses

The results of the modified intention-to-treat (ITT) analyses, which compared the risk of infections between different treatment arms after excluding study participants determined to be HIV positive at enrollment, are presented in Table 2.3. In each of the Partners PrEP oral, VOICE oral, and VOICE gel analyses, our estimates of the incidence setting concentration to consistently below the limit of quantification were very similar to the observed HIV incidences in their corresponding placebo arms. Figures 2.2, 2.3, and 2.4 compare estimates of the effectiveness of PrEP between our longitudinal TMLE analyses, the modified ITT analyses, and previous secondary analyses by Murnane et al. (2015) and Dai et al. (2015) for the Partners PrEP and VOICE studies, respectively. We note that the target of estimation and assumptions differ for each approach.

The primary Partners PrEP analysis, a modified ITT analysis, concluded that treatment with TDF and TDF/FTC, as compared to placebo, conferred relative reductions of HIV incidence,  $(1 - \text{incidence rate ratio}) \times 100\%$ , of 67% (95% CI: 44% to 81%) and 75% (95% CI: 55% to 87%), respectively (Baeten et al., 2012). As a secondary analysis, Murnane et al. (2015) used principal stratification to estimate the effect of PrEP on risk of HIV infection among “high adherers.” They defined high adherers as those who, if assigned to TDF or TDF/FTC, would achieve a plasma concentration level of PrEP above 40 ng/mL at their six month visit. Using data from the concentration substudy, they built a logistic regression model for the probability that a study participant is a high adherer. This model was used to predict each study participant’s probability of high adherence. They

used Cox regression with the following terms to estimate the effectiveness of PrEP: randomization arm, the predicted probability of high adherence, and the interaction between the two. Among the strata of high adherers, Murnane et al. (2015) estimated a reduction in hazard of HIV acquisition by 81% (44% to 93%) with TDF and 88% (48% to 97%) with TDF/FTC. Although the causal questions differ between the longitudinal TMLE and principal stratification analyses, both analyses concluded similar levels of PrEP effectiveness. We note that the above principal stratification approach relies on the parametric prediction models for both plasma drug concentration and risk of seroconversion being correctly specified. The method also assumes that plasma drug concentration stays constant among study participants between their 6 month visit and end of follow up (up to 36 months).

In the VOICE trial, the modified ITT analysis compared HIV incidence in each of the treatment arms to their corresponding placebo arms using a proportional hazards model stratified by site. The effectiveness of oral TDF, oral TDF/FTC, and TFV gel,  $(1 - \text{hazard ratio}) \times 100\%$ , was estimated to be  $-49\%$  (95% CI:  $-129$  to  $3\%$ ),  $-4.4\%$  (95% CI:  $-49\%$  to  $27\%$ ), and  $15\%$  (95% CI:  $-21$  to  $39\%$ ), respectively (Marrazzo et al., 2015). Additionally, Dai et al. (2015) used principal stratification to estimate the effect of PrEP on the risk of HIV infection among adherers. They defined adherers in two ways: first, as those with a quantifiable level of plasma tenofovir in their three month sample; second, as those with a quantifiable level of plasma tenofovir in at least one plasma sample during follow-up. For each of the oral TDF, oral TDF/FTC, and gel arms, they used a weighted Poisson

model to compare HIV incidence between adherers in the treatment arm and all participants in the placebo arm, while adjusting for potential confounding. For the first definition of adherer, Dai et al. (2015) estimated relative risk reductions of 2% (95% CI: -71% to 43%), -2% (95% CI: -116% to 52%), and 47% (95% CI: 3% to 71%) with assignment to oral TDF, oral TDF/FTC, and TFV gel arms, respectively. For the second definition of adherer, Dai et al. (2015) estimated relative risk reductions of -7% (95% CI: -142% to 53%), -26% (95% CI: -140% to 34%), and 60% (95% CI: 2% to 84%) with oral TDF, oral TDF/FTC, and TFV gel, respectively.

In contrast, using the longitudinal TMLE, we estimated the relative risk of setting drug concentration to above the limit of quantification in the combined VOICE oral TDF and TDF/FTC arms to be 42% (95% CI: -31% to 80%). In the TFV gel arm we estimated a relative risk reduction of 88% (95% CI: -48% to 100%). Unlike previous work, we accounted for varying levels of drug concentration over time. Dai et al. (2015) used a homogeneous Markov chain model to estimate the probability, given a quantifiable plasma concentration at one visit, of having plasma concentration below the limit of quantification at the next visit. They estimated non-negligible probabilities of 39% for the combined oral TDF and TDF/FTC arms and 56% for the TFV gel arm, demonstrating that the assumption of plasma concentration not changing over time was likely violated in the VOICE trial.

To highlight the importance of accounting for longitudinal concentration, we compared the results from our analyses with a similar analysis using the longitudinal TMLE method but with the assumption, made in previous analyses, that concentration is constant over time. We repeated the longitudinal TMLE analysis

after replacing the observed 12 and 18 month concentrations (for study participants who neither had been censored nor had seroconverted) with their 6 month measurements. In the VOICE study, repeating the analysis using only concentration measurements from the 6 month study visit, we estimated in the oral arms a  $-1\%$  relative risk reduction (95% CI:  $-65\%$  to  $43\%$ ), which highly overlaps the confidence intervals in Dai et al. (2015). In the gel arm, we estimated a much lower relative risk reduction of  $17\%$  (95% CI:  $-127\%$  to  $79\%$ ) compared to that in the original analysis (Figure 2.4). The estimated effectiveness using only the 6 month concentration measurements was closer to the estimate by Dai et al. (2015) than our original estimate.

In contrast, in the Partners PrEP study, the estimated relative risk reduction changed very little (from  $82\%$  to  $81\%$ ) when we used only the 6 month concentration measurements (Figure 2.2). Our results for the Partners PrEP study are likely consistent with previous analyses because most study participants maintained fairly consistent concentration levels over time (Donnell et al., 2014). In contrast, concentration measurements among individual study participants in both the VOICE oral and gel arms varied considerably over the course of the study, typically decreasing over time, as illustrated in Figure 2.5.

## 2.6 Discussion

We used TMLE to estimate the causal effect of setting plasma concentration measures to be consistently above a specified threshold on the risk of HIV infection in the Partners PrEP and VOICE studies. We add to a growing body of research in

this area by using targeted maximum likelihood estimation to account for the longitudinal nature of the data and correcting for potential time varying confounding.

We used weighting to account the for the case-cohort sampling schema for concentration measurements. An alternative approach to account for the sampling schema would be to use multiple imputation; however, this requires the imputation model, in addition to the outcome model, to be correct. Additionally, imputation requires the predictors of drug concentration levels, such as participant reported adherence and counts of returned pills, to be accurate. These measures have been found to be unreliable in several PrEP studies, including VOICE (Marrazzo et al., 2015; Dai et al., 2015).

Our results for the Partners PrEP study are similar to previous analyses. Unlike previous analyses, our point estimates from the combined oral arms of the VOICE study are in the direction of benefit. We estimate a higher protective effect in the VOICE gel arm than previous analyses. These differences are not surprising given that the target parameters differ between these analyses and that we account for time-varying levels of drug concentration. Additionally, unlike previous analyses that considered the two oral arms separately, we estimated the effect of tenofovir in the combined TDF and TDF/FTC oral arms in each of the Partners PrEP and VOICE studies. All of our results are consistent with claims that sustaining high concentration levels is a key factor in the potential success of any future PrEP trials.

As with previous analyses, caution must be taken with our results. As in any analysis aimed at estimating causal effects from observational data, a causal interpretation of the estimated parameters requires the strong assumptions listed

in Section 4.2, including that of no unmeasured confounding. Several groups have recently reported that bacteria associated with bacterial vaginosis, can metabolize tenofovir by an, as yet, uncertain mechanism, resulting in lower concentrations of tenofovir in cervicovaginal fluid and, likely, in tissue and plasma (Heffron et al., 2017; Hillier et al., 2017; Klatt et al., 2017; Velloza and Heffron, 2017). This has resulted in some loss of tenofovir's protective effect in the CAPRISA 004 clinical trial (Klatt et al., 2017; Velloza and Heffron, 2017). Because we do not know which, if any, women in the Partners PREP or VOICE trials had bacterial vaginosis at the time of plasma sampling, we cannot rule this out as an additional unmeasured variable that may impact our findings. Additionally, there may be sources of measurement error in the data, such as a white-coat effect, wherein concentration levels increase just prior to a clinic visit since the participant knows he/she will have concentration assessed. If there were a white-coat effect, then participant-visits might be classified as exceeding the concentration threshold, even if they were below the threshold for most of the time at risk between visits.

The main limitations in our longitudinal analyses were that the outcome, sero-conversion, was rare and was (like blood concentration of drug) measured at visits separated by multiple months. The problem of event sparsity was also present when Schnitzer et al. (2014b) used longitudinal TMLE to estimate the causal effect of hepatitis C virus clearance on end-stage liver disease free survival. They used exposure and censoring models that depended only on baseline covariates and those from the previous time point, rather than the full history and restricted the followup period. Similarly, our exposure and censoring models depended only on

covariates from the previous time point. We also restricted our analyses to the first 18 months of the study.

An important limitation in the VOICE analyses is that the majority of concentration measurements (72% and 80% in the oral and gel arms, respectively) were below the limit of quantification. As indicated in Table 2.1, there were very few study participants who followed a high concentration regimen. In particular, in the concentration substudy of the VOICE gel arm, only 11 of 669 participants followed a quantifiable concentration regimen. As a result, these analyses extrapolate from a small number of observations.

In Table 2.4 we compare self-reported risk behaviors over time between the 11 participants and the overall VOICE gel arm to investigate whether the participants following a quantifiable concentration regimen were systematically less risky than the rest of the study arm. If that were the case, the reduction in risk of HIV infection we estimated from setting participants to follow a quantifiable concentration regimen might be due to confounding by risk factors rather than a biological protective effect of tenofovir. We do not observe any changes in risk behavior over time, and behavior among the 11 study participants following the quantifiable concentration regimen appears to be similar or slightly riskier to the overall behavior in the VOICE gel arm.

We repeated our LTMLE analyses of the VOICE gel arm using data from only the first 6 and 12 months of the study. Using data from the first 6 months, we estimated the protective effect of setting plasma tenofovir concentration to quantifiable at the 6 month visit, and using data from the first 12 months, we estimated the protective

effect of setting plasma tenofovir concentration to quantifiable at both the 6 and 12 month study visits. Although in practice we are interested in the protective effect of maintaining drug concentrations over a longer period of time, these analyses have the advantage that a larger number of study participants follow a quantifiable concentration regimen when only considering concentration data from the first 6 or 12 months as compared to 18 months. As shown in Table 2.5, restricting our analyses to the first 6 or 12 months increases the number of study participants following a quantifiable concentration regimen to 148 and 31 study participants, respectively. The results from the LTMLE analyses using data from only the first 6 and 12 months, presented in table 2.6 are qualitatively similar.

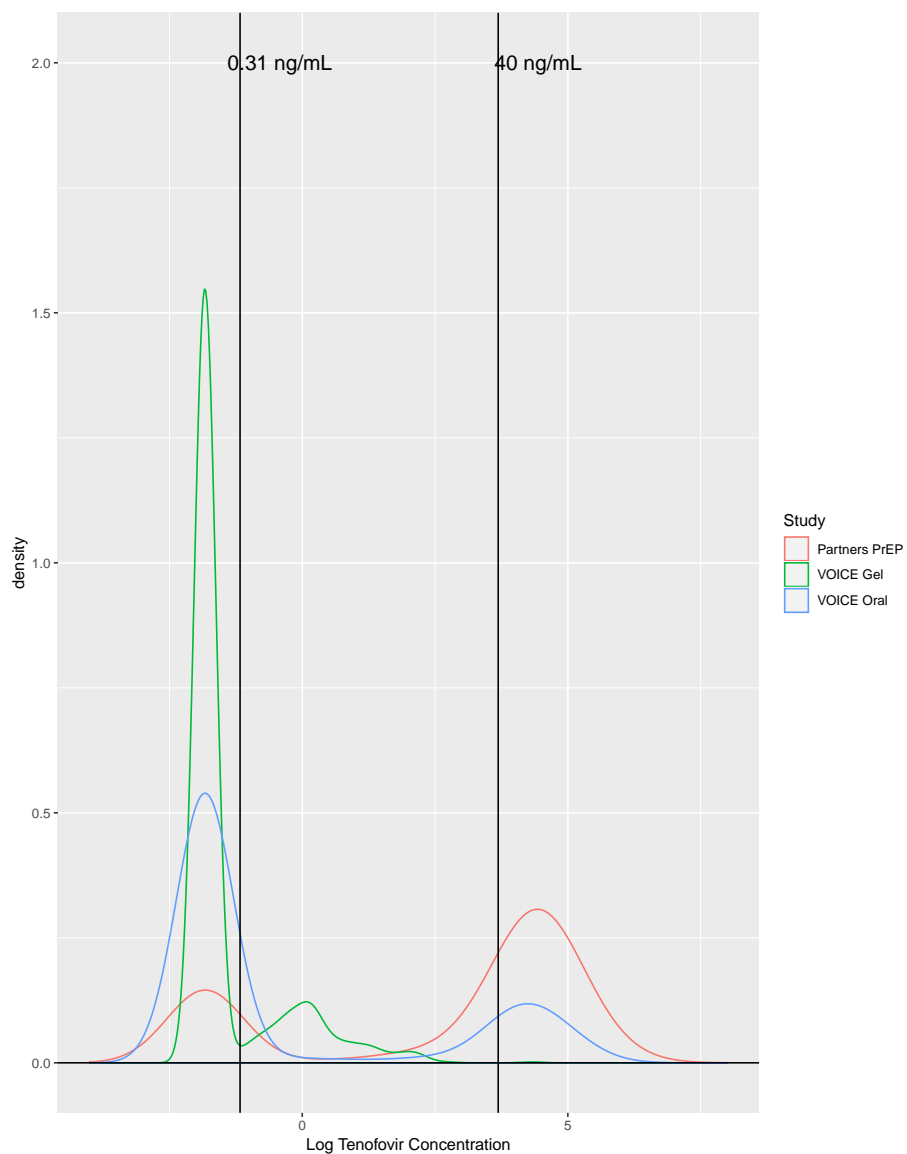
A final limitation is that our analyses of focus exclusively on estimating the effect of plasma tenofovir concentration on risk of HIV infection. Serum concentrations have been found to be considerably higher after oral dosing as compared to vaginal gel dosing, (Hendrix et al., 2013) meaning detection of plasma tenofovir from topical dosing requires a higher level of adherence than from oral dosing. Consequently, the women following a quantifiable tenofovir regimen in the VOICE gel arm are likely all highly adherent, while the women following a quantifiable tenofovir regimen in the oral arms are a mixture of moderate and highly adherent participants. Because moderately adherent participants in the VOICE gel arm may have been classified as following a below tenofovir regimen, the true efficacy might be higher than estimated for the topical dosing arm. In contrast, vaginal gel dosing achieves much higher drug concentrations in vaginal tissue, an important site of consideration for HIV infection, than oral dosing (Hendrix et al., 2013), achieving



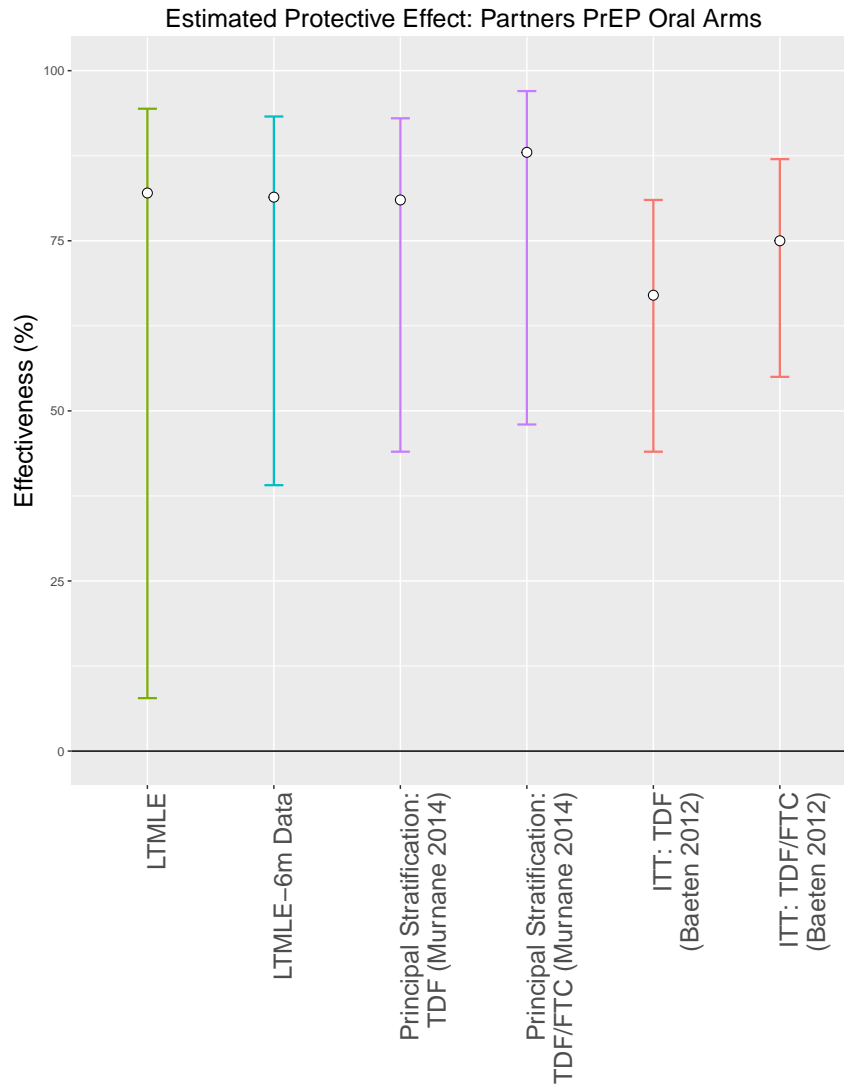
quantifiable levels of plasma tenofovir from oral and gel dosing likely has different implications for both adherence levels and tenofovir concentrations in different compartments relevant for HIV infection. In particular, when analyzing only plasma concentration data, detection of tenofovir from topical dosing indicates a higher level of adherence than oral dosing. As a result, the true efficacy might be higher than estimated for the topical dosing arm.

## **2.7 Acknowledgements**

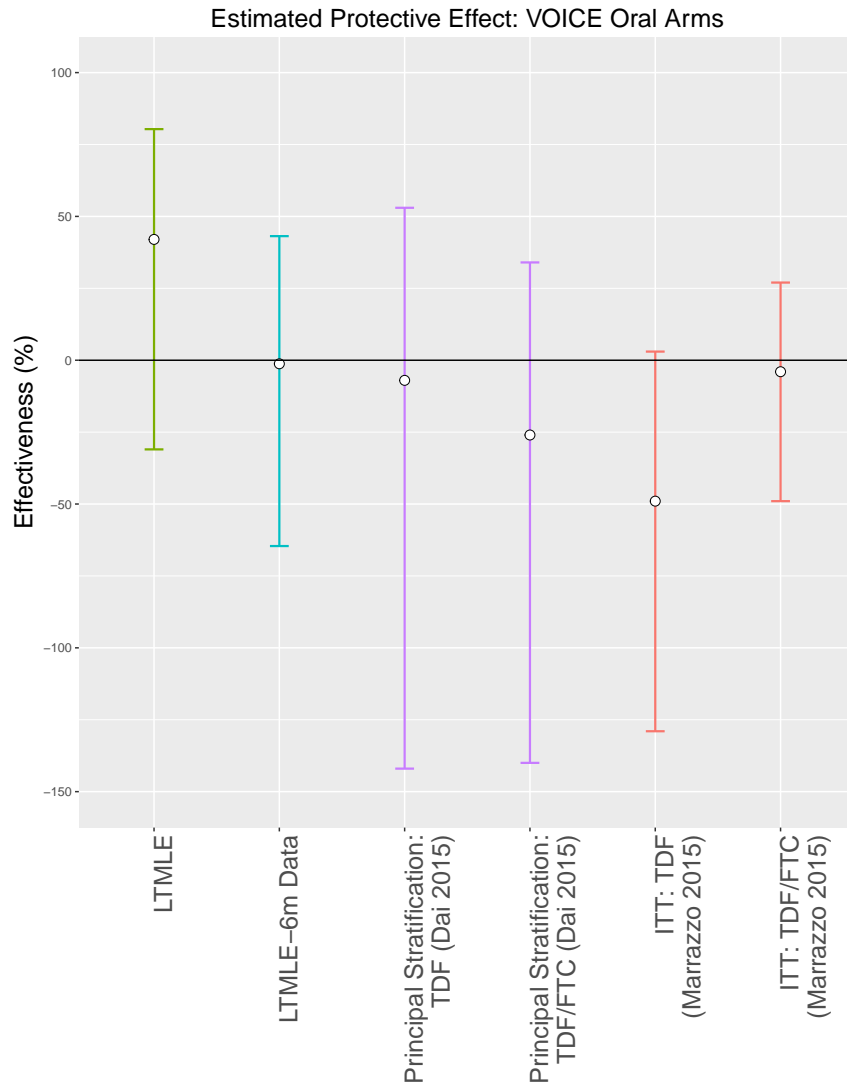
This work was supported by the Johns Hopkins University Center for AIDS Research (1P30AI094189).



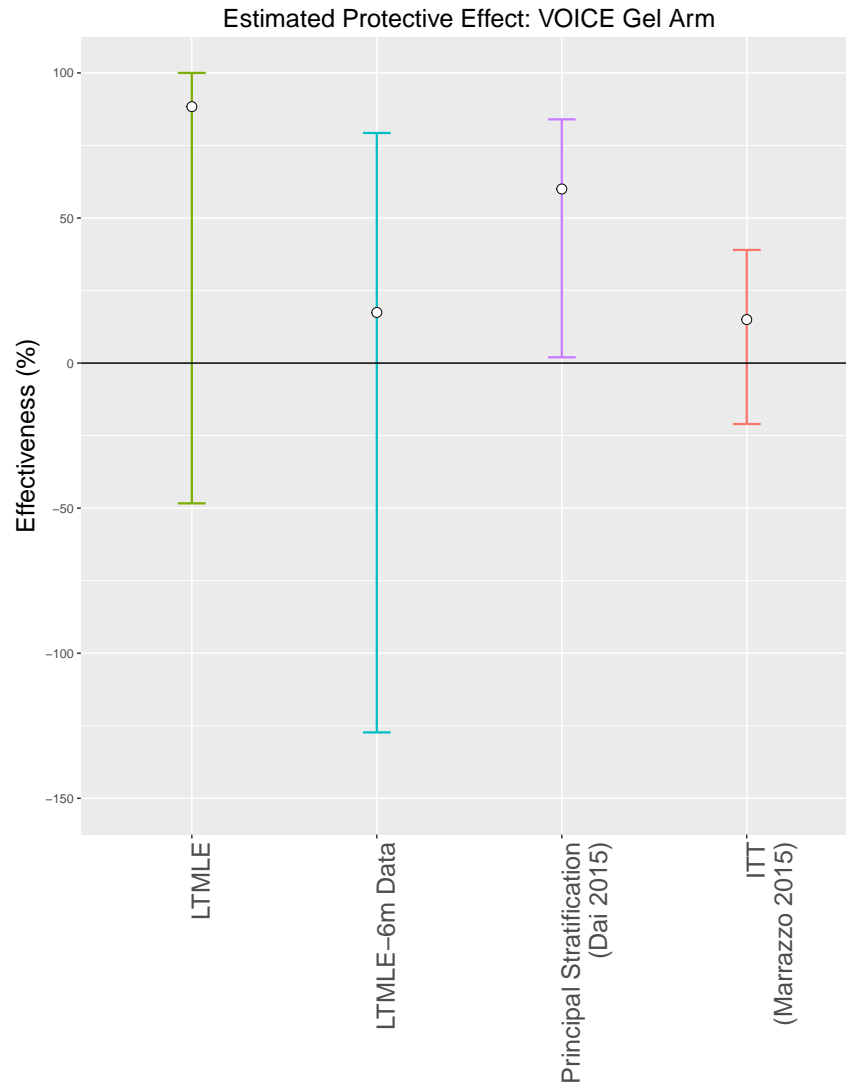
**Figure 2.1: Distribution of plasma concentration measurements in the first 18 months by study.** The limit of quantification is 0.31 ng/mL for both studies. In Partners PrEP, 74% of concentration measurements were above the limit of quantification. In the VOICE study, 28% and 20% of concentration measurements from the oral and gel arms were above the limit of quantification, respectively.



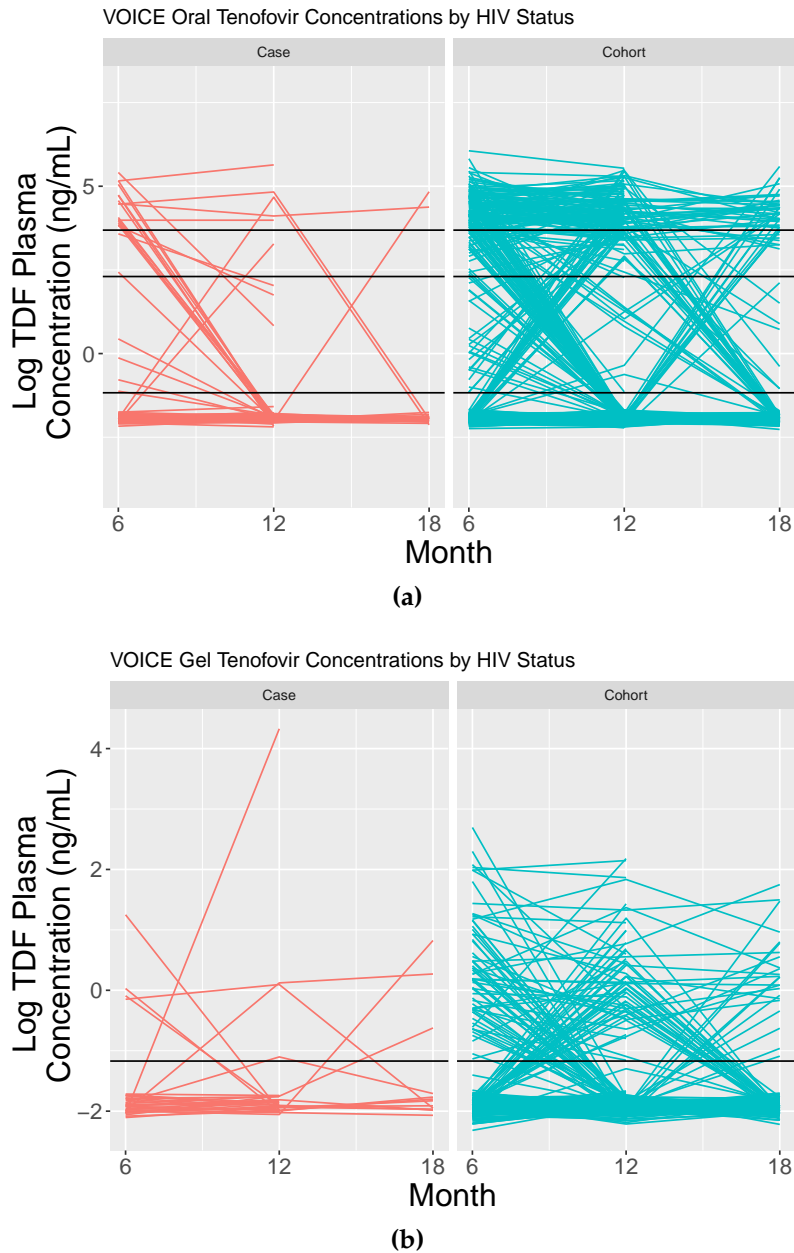
**Figure 2.2: Comparison of results analyzing the Partners PrEP study using the longitudinal TMLE to results from previous methods.** We compare the protective effect of PrEP estimated by our longitudinal TMLE analyses (“LTMLE”), intention-to-treat analyses (“ITT”), secondary analyses using concentration data by Murnane et al. (2015) (“Principal Stratification”), and repeating our longitudinal TMLE analysis when we only use each study participant’s 6 month concentration measurement (“LTMLE-6m Data”). All point estimates (dots) are given with 95% confidence intervals (bars).



**Figure 2.3: Comparison of results analyzing the VOICE oral arms using the longitudinal TMLE to results from previous methods.** We compare the protective effect of PrEP estimated by our longitudinal TMLE analyses (“LTMLE”), intention-to-treat analyses (“ITT”), secondary analyses using concentration data by Dai et al. (2015) (“Principal Stratification”), and repeating our longitudinal TMLE analysis when we only use each study participant’s 6 month concentration measurement (“LTMLE-6m Data”). All point estimates (dots) are given with 95% confidence intervals (bars).



**Figure 2.4: Comparison of results analyzing the VOICE gel arm using the longitudinal TMLE to results from previous methods.** We compare the protective effect of PrEP estimated by our longitudinal TMLE analyses (“LTMLE”), intention-to-treat analyses (“ITT”), secondary analyses using concentration data by Dai et al. (2015) (“Principal Stratification”), and repeating our longitudinal TMLE analysis when we only use each study participant’s 6 month concentration measurement (“LTMLE-6m Data”). All point estimates (dots) are given with 95% confidence intervals (bars).



**Figure 2.5: Concentration measurements from the first three semi-annual visits by HIV status in the VOICE gel and oral arms.** Each line shows the 6, 12, and 18 month concentration measurement for an individual study participant. Lines that end before 18 months are due to seroconversion, missed study visits, or participant drop-out. Participants are separated by their HIV status at the end of the study. Concentration levels vary considerably among individuals in both the VOICE oral (a) and gel (b) arms, typically with the trend of decreasing concentrations over time.

		Partners PrEP	VOICE Oral	VOICE Gel
Quantifiable Concentration Regimen	All	84 (37%)	45 (6%)	11 (2%)
	HIV+	8 (31%)	8 (7%)	2 (4%)
	HIV-	76 (37%)	37 (6%)	9 (1%)
BLQ Concentration Regimen	All	31 (14%)	257 (35%)	140 (21%)
	HIV+	12 (46%)	82 (75%)	42 (78%)
	HIV-	19 (9%)	175 (28%)	98 (16%)
Switched	All	19 (8%)	81 (11%)	55 (8%)
	HIV+	1 (4%)	17 (15%)	9 (17%)
	HIV-	18 (9%)	64 (10%)	46 (7%)
Censored	All	91 (40%)	358 (48%)	463 (69%)
	HIV+	5 (19%)	3 (3%)	1 (2%)
	HIV-	86 (42%)	355 (56%)	462 (75%)
Total	All	225	741	669
	HIV+	26	110	54
	HIV-	204	634	616

**Table 2.1: Number of concentration substudy participants in the Partners PrEP and VOICE studies by treatment, concentration regimen followed, and seroconversion outcome.** Number and percent of concentration substudy participants in the Partners PrEP and VOICE oral TDF and TDF/FTC and VOICE TFV gel arms who maintained quantifiable or below quantifiable concentration levels over the first eighteen months of the study or until serconversion, were censored at or before eighteen months, or switched between concentration levels. Participants who did not follow a treatment regimen in its entirety contributed to the analysis until they switched concentration groups or were right censored. The first row, "Quantifiable Concentration Regimen," enumerates the number of study participants who maintained quantifiable concentration levels at each of their 6, 12, and 18 month visits, or until serconversion. The second row, "BLQ Concentration Regimen," does the same for those with concentration measurements below the limit of quantification at all visits.

Study	Adjustment Variables	Estimated Incidence Setting Concentration to an Undetectable Regimen	Estimated Incidence Setting Concentration to a Detectable Regimen	% Relative Risk Reduction $100 \times (1\text{-relative risk})$
Partners PrEP	Baseline	2.3 (0.8, 5.4)	0.4 (0.2, 1.1)	81 (27, 94)
Partners PrEP	Baseline + Risk	2.4 (0.7, 5.4)	0.4 (0.2, 1.2)	82 (8, 94)
VOICE Oral	Baseline	4.8 (3.7, 6.3)	2.8 (1.1, 5.7)	42 (-22, 79)
VOICE Oral	Baseline + Risk	4.8 (3.8, 6.3)	2.8 (1, 5.9)	42 (-31, 80)
VOICE Gel	Baseline	6.8 (4.5, 10.9)	0.7 (0, 5.5)	89 (21, 100)
VOICE Gel	Baseline + Risk	6.8 (4.4, 10.7)	0.8 (0, 7.9)	88 (-48, 100)

**Table 2.2: Estimates and 95% confidence intervals based on the longitudinal TMLE from Section 4 of the incidence of HIV in the Partners PrEP and VOICE study populations when concentration is set to above or below the limit of quantification at each of the 6, 12, and 18 months study visits.** Incidence is measured on the scale of seroconversions per 100 person-years. The final column gives estimates of the associated relative risk reduction,  $100 \times (1\text{-relative risk})$ , comparing setting concentration above versus below the limit of quantification. Analyses were first performed with adjustment only for baseline variables ("Baseline") including age and gender for Partners PrEP and age, marital status, and HSV-2 status for VOICE. These variables were chosen because they were associated with plasma concentration levels and/or risk of seroconversion. Analyses were then performed with adjustment for both baseline variables and self-reported risk levels at each of the study visits ("Baseline + Risk"). The "Risk" variables for both Partners PrEP and VOICE are indicators at each of the 6, 12, and 18 month study visits of reported unprotected sex during the 6 months prior.



Study	Treatment Arm	Incidence Placebo (Seroconversions per 100 person- years)	Incidence Active Drug	Percent Relative Risk Reduction
Partners PrEP	Oral TDF Oral TDF/FTC	2.0 (1.4, 2.5)	0.7 (0.3, 1.0) 0.5 (0.2, 0.8)	67 (44, 81) 75 (55, 87)
VOICE	Oral TDF Oral TDF/FTC	4.2 (2.9, 5.8) 4.6 (3.5, 5.9)	6.3 (4.7, 8.3) 4.7 (3.6, 6.1)	-49 (-129, -3) -4 (-49, 27)
	Tenofovir Gel	6.8 (5.3, 8.6)	6.0 (4.6, 7.6)	15 (-21, 39)

**Table 2.3: Results from the modified intention-to-treat analyses from Partners PrEP and VOICE**(Baeten et al., 2012; Marrazzo et al., 2015). Incidence is defined as the number of cases per 100 person-years. Note that there was only one oral placebo arm in the VOICE study. However, because the oral TDF arm was discontinued due to futility, incidence in the oral TDF arm is compared to the incidence in the oral placebo prior to TDF discontinuation. In contrast, incidence in the oral TDF/FTC arm is compared to incidence in the oral placebo arm for the whole length of the study. Percent relative risk reduction is defined as  $100 \times (1 - \text{relative risk})$ , and was calculated in both studies using site stratified Cox proportional-hazards models. All estimates are reported with 95% confidence intervals.

	Vaginal Sex Past 3m (% yes)	Vaginal Sex Past 4w (% yes)	Vaginal Sex Past 7d n acts (mean)	Vaginal Sex Past 7d (% condom)	Vaginal Sex Last Act (% condom)	Anal Sex Acts Past 3m Mean (Median)	Anal Sex Last Act (% condom)
6m Overall	97	90	2.5	76	84	0.4, 0	70
12m Overall	96	88	2.4	76	84	0.3, 0	72
18m Overall	95	88	2.4	76	83	0.3, 0	71
6m Quant	100	91	2.5	56	84	0.5, 0	0
12m Quant	95	86	2.4	58	84	0.5, 0	0
18m Quant	90	84	2.5	55	84	0.4, 0	0

**Table 2.4: Risk behaviors in the VOICE gel arm over the first 18 study months..** We compare risk behaviors in the overall VOICE gel concentration substudy (“Overall”) to the 11 study participants who followed a quantifiable concentration regimen (“Quant”).

		VOICE Gel 6 months	VOICE Gel 12 months	VOICE Gel 18 months
Quantifiable Concentration Regimen	All	148 (22%)	31 (5%)	11 (2%)
	HIV+	2 (8%)	2 (5%)	2 (4%)
	HIV-	146 (23%)	29 (5%)	9 (1%)
BLQ Concentration Regimen	All	521 (78%)	235 (35%)	140 (21%)
	HIV+	22 (92%)	38 (88%)	42 (78%)
	HIV-	499 (77%)	197 (31%)	98 (16%)
Switched	All	0	68 (10%)	55 (8%)
	HIV+	0	3 (7%)	9 (17%)
	HIV-	0	65 (10%)	46 (7%)
Censored	All	0	335 (50%)	463 (69%)
	HIV+	0	0	1 (2%)
	HIV-	0	335 (54%)	462 (75%)
Total	All	669	669	669
	HIV+	24	43	54
	HIV-	645	626	616

**Table 2.5: Comparison of the number of VOICE gel arm substudy participants by concentration regimen followed and seroconversion outcome using data from the first 6, 12, and 18 months.** Number and percent of concentration substudy participants who maintained quantifiable or below quantifiable concentration levels over the first 6, 12, or 18 months of the study or until seroconversion, were censored, or switched between concentration levels. Participants who did not follow a treatment regimen in its entirety contributed to the analysis until they switched concentration groups or were right censored. The first row, "Quantifiable Concentration Regimen," enumerates the number of study participants who maintained quantifiable concentration levels each of their semiannual visits over the first 6, 12, or 18 months of the study, or until seroconversion. The second row, "BLQ Concentration Regimen," does the same for those with concentration measurements below the limit of quantification at all visits.

Months	Adjustment	Incidence BLQ	Incidence Quant	Relative Risk Reduction
6	Baseline	5.4 (3.4, 8.1)	2.2 (0, 8.3)	58 (-64, 100)
6	Baseline + Risk	5.4 (3.4, 8.3)	2.4 (0, 8.5)	56 (-82, 100)
12	Baseline	7.3 (5, 10.9)	1.1 (0, 4)	85 (42, 100)
12	Baseline + Risk	7.2 (4.9, 10.8)	1.2 (0, 4.3)	84 (36, 100)
18	Baseline	6.8 (4.5, 10.9)	0.7 (0, 5.5)	89 (21, 100)
18	Baseline + Risk	6.8 (4.4, 10.7)	0.8 (0, 7.9)	88 (-48, 100)

**Table 2.6: Estimates and 95% confidence intervals based on the longitudinal TMLE from Section 4 of the incidence of HIV in VOICE gel study populations when concentration is set to above or below the limit of quantification at each semi-annual visit using data from the first 6, 12, and 18 months of the study.** Incidence is measured on the scale of seroconversions per 100 person-years. The final column gives estimates of the associated relative risk reduction,  $100 \times (1 - \text{relative risk})$ , comparing setting concentration above versus below the limit of quantification. Analyses were first performed with adjustment only for baseline variables ("Baseline") including age, marital status, and HSV-2 status. These variables were chosen because they were associated with plasma concentration levels and/or risk of seroconversion. Analyses were then performed with adjustment for both baseline variables and self-reported risk levels at each of the study visits ("Baseline + Risk"). The "Risk" variables are indicators at each semiannual visit of reported unprotected sex during the 6 months prior.

# References

- World Health Organization (2015). "Treat all people living with HIV, offer antiretrovirals as additional prevention choice for people at 'substantial' risk". In: URL: <http://www.who.int/mediacentre/news/releases/2015/hiv-treat-all-recommendation/en>.
- Krakower, Douglas S and Kenneth H Mayer (2015). "Pre-exposure prophylaxis to prevent HIV infection: current status, future opportunities and challenges". In: *Drugs* 75.3, pp. 243–251.
- Baeten, J, Renee Heffron, Lara Kidoguchi, Nelly Mugo, Elly Katabira, Elizabeth Bukusi, S Asimwe, J Morton, K Ngure, N Bulya, et al. (2016a). "Integrated delivery of PrEP and ART results in sustained near elimination of HIV transmission in African HIV serodiscordant couples: final results from The Partners Demonstration Project". In: *21st International AIDS Conference (AIDS 2016)*, pp. 18–22.
- Lal, Luxi, Jennifer Audsley, Dean Murphy, Christopher K Fairley, Mark Stoove, Norm Roth, Richard Moore, BK Tee, Nalagafiar Puratmaja, Peter L Anderson, et al. (2017). "Medication adherence, condom use and sexually transmitted infections in Australian PrEP users: interim results from the Victorian PrEP Demonstration Project." In: *AIDS*.
- Hoagland, Brenda, Ronaldo I Moreira, Raquel B De Boni, Esper G Kallas, José Valdez Madruga, Ricardo Vasconcelos, Silvia Goulart, Thiago S Torres, Luana Marins, Peter L Anderson, et al. (2017). "High pre-exposure prophylaxis uptake and early adherence among men who have sex with men and transgender women at risk for HIV Infection: the PrEP Brasil demonstration project". In: *Journal of the International AIDS Society* 20.1.
- Van der Straten, Ariane, Lut Van Damme, Jessica E Haberer, and David R Bangsberg (2012). "Unraveling the divergent results of pre-exposure prophylaxis trials for HIV prevention". In: *AIDS* 26.7, F13–F19.

- Baeten, Jared M and Robert Grant (2013). "Use of antiretrovirals for HIV prevention: what do we know and what don't we know?" In: *Current HIV/AIDS Reports* 10.2, pp. 142–151.
- Fonner, Virginia A, Sarah L Dalglish, Caitlin E Kennedy, Rachel Baggaley, Kevin R O'Reilly, Florence M Koechlin, Michelle Rodolph, Ioannis Hodges-Mameletzis, and Robert M Grant (2016). "Effectiveness and safety of oral HIV preexposure prophylaxis for all populations". In: *AIDS (London, England)* 30.12, p. 1973.
- Spinner, Christoph D, Christoph Boesecke, Alexander Zink, Heiko Jessen, Hans-Jürgen Stellbrink, Jürgen Kurt Rockstroh, and Stefan Esser (2016). "HIV pre-exposure prophylaxis (PrEP): a review of current knowledge of oral systemic HIV PrEP in humans". In: *Infection* 44.2, pp. 151–158.
- Baeten, Jared M, Deborah Donnell, Patrick Ndase, Nelly R Mugo, James D Campbell, Jonathan Wangisi, Jordan W Tappero, Elizabeth A Bukusi, Craig R Cohen, Elly Katabira, et al. (2012). "Antiretroviral prophylaxis for HIV prevention in heterosexual men and women". In: *New England Journal of Medicine* 367.5, pp. 399–410.
- Marrazzo, Jeanne M, Gita Ramjee, Barbra A Richardson, Kailazarid Gomez, Nyaradzo Mgodhi, Gonasagrie Nair, Thesla Palanee, Clemensia Nakabiito, Ariane van der Straten, Lisa Noguchi, et al. (2015). "Tenofovir-based preexposure prophylaxis for HIV infection among African women". In: *New England Journal of Medicine* 372.6, pp. 509–518.
- Dai, James Y, Craig W Hendrix, Barbra A Richardson, Cliff Kelly, Mark Marzinke, Z Mike Chirenje, Jeanne M Marrazzo, and Elizabeth R Brown (2015). "Pharmacological Measures of Adherence and Risk of HIV Acquisition in the VOICE Study". In: *Journal of Infectious Diseases*, jiv333.
- McGowan, Ian (2014). "The development of rectal microbicides for HIV prevention". In: *Expert Opinion on Drug Delivery* 11.1, pp. 69–82.
- Baeten, Jared M, Thesla Palanee-Phillips, Elizabeth R Brown, Katie Schwartz, Lydia E Soto-Torres, Vaneshree Govender, Nyaradzo M Mgodhi, Flavia Matovu Kiweewa, Gonasagrie Nair, Felix Mhlanga, et al. (2016b). "Use of a vaginal ring containing dapivirine for HIV-1 prevention in women". In: *New England Journal of Medicine* 375.22, pp. 2121–2132.
- Margolis, David A, Juan Gonzalez-Garcia, Hans-Jürgen Stellbrink, Joseph J Eron, Yazdan Yazdanpanah, Daniel Podzamczek, Thomas Lutz, Jonathan B Angel, Gary J Richmond, Bonaventura Clotet, et al. (2017). "Long-acting intramuscular cabotegravir and rilpivirine in adults with HIV-1 infection (LATTE-2): 96-week

- results of a randomised, open-label, phase 2b, non-inferiority trial". In: *The Lancet* 390.10101, pp. 1499–1510.
- Dai, James Y, Peter B Gilbert, James P Hughes, and Elizabeth R Brown (2013). "Estimating the efficacy of preexposure prophylaxis for HIV prevention among participants with a threshold level of drug concentration". In: *American Journal of Epidemiology* 177.3, pp. 256–263.
- Murnane, Pamela M, Elizabeth R Brown, Deborah Donnell, R Yates Coley, Nelly Mugo, Andrew Mujugira, Connie Celum, Jared M Baeten, Partners PrEP Study Team, Andrew Mujugira, et al. (2015). "Estimating efficacy in a randomized trial with product nonadherence: application of multiple methods to a trial of preexposure prophylaxis for HIV prevention". In: *American Journal of Epidemiology* 182.10, pp. 848–856.
- Van der Laan, Mark J and Susan Gruber (2012). "Targeted minimum loss based estimation of causal effects of multiple time point interventions". In: *The International Journal of Biostatistics* 8.1.
- Golub, Sarit A, Don Operario, and Pamina M Gorbach (2010). "Pre-exposure prophylaxis state of the science: empirical analogies for research and implementation". In: *Current HIV/AIDS Reports* 7.4, pp. 201–209.
- Liu, Albert Y, Eric Vittinghoff, Kata Chillag, Kenneth Mayer, Melanie Thompson, Lisa Grohskopf, Grant Colfax, Sonal Pathak, Roman Gvetadze, Brandon Oâ€™Hara, et al. (2013). "Sexual risk behavior among HIV-uninfected men who have sex with men (MSM) participating in a tenofovir pre-exposure prophylaxis (PrEP) randomized trial in the United States". In: *Journal of Acquired Immune Deficiency Syndromes* (1999) 64.1, p. 87.
- Calabrese, Sarah K and Kristen Underhill (2015). "How Stigma Surrounding the Use of HIV Preexposure Prophylaxis Undermines Prevention and Pleasure: A Call to Destigmatize 'Truvada Whores'". In: *American Journal of Public Health* 105.10, pp. 1960–1964.
- Van der Straten, Ariane, Ashley Mayo, Elizabeth R Brown, K Rivet Amico, Helen Cheng, Nicole Laborde, Jeanne Marrazzo, and Kristine Torjesen (2014). "Perceptions and Experiences with the VOICE Adherence Strengthening Program (VASP) in the MTN-003 Trial". In: *AIDS and Behavior* 19.5, pp. 770–783.
- Hernán, Miguel Ángel, Babette Brumback, and James M Robins (2000). "Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men". In: *Epidemiology* 11.5.

- Van der Laan, Mark J and Daniel Rubin (2006). "Targeted maximum likelihood learning". In: *The International Journal of Biostatistics* 2.1.
- Robins, James M (2000). "Robust estimation in sequentially ignorable missing data and causal inference models". In: *Proceedings of the American Statistical Association Section on Bayesian Statistical Science*. Vol. 1999, pp. 6–10.
- Bang, Heejung and James M Robins (2005). "Doubly robust estimation in missing data and causal inference models". In: *Biometrics* 61.4, pp. 962–973.
- Robins, J. M. and A. Rotnitzky (1992). "Recovery of information and adjustment for dependent censoring using surrogate markers". In: *AIDS Epidemiology-Methodological Issues* 297–331.
- Schwab, Joshua, Samuel Lendle, Maya Petersen, and Mark van der Laan (2016). *ltmle: Longitudinal Targeted Maximum Likelihood Estimation*. URL: <https://CRAN.R-project.org/package=ltmle>.
- Rose, Sherri and Mark J van der Laan (2011). "A targeted maximum likelihood estimator for two-stage designs". In: *The International Journal of Biostatistics* 7.1, pp. 1–21.
- Stitelman, Ori M, Victor De Gruttola, and Mark J van der Laan (2012). "A general implementation of TMLE for longitudinal data applied to causal inference in survival analysis". In: *The International Journal of Biostatistics* 8.1.
- Brooks, Jordan C, Mark J van der Laan, Daniel E Singer, and Alan S Go (2013). "Targeted minimum loss-based estimation of causal effects in right-censored survival data with time-dependent covariates: Warfarin, stroke, and death in atrial fibrillation". In: *Journal of Causal Inference* 1.2, pp. 235–254.
- Schnitzer, Mireille E, Erica EM Moodie, Mark J van der Laan, Robert W Platt, and Marina B Klein (2014b). "Modeling the impact of hepatitis C viral clearance on end-stage liver disease in an HIV co-infected cohort with targeted maximum likelihood estimation". In: *Biometrics* 70.1, pp. 144–152.
- Schnitzer, Mireille E, Mark J van der Laan, Erica EM Moodie, and Robert W Platt (2014a). "Effect of breastfeeding on gastrointestinal infection in infants: a targeted maximum likelihood approach for clustered longitudinal data". In: *The Annals of Applied Statistics* 8.2, p. 703.
- Decker, Anna L, Alan Hubbard, Catherine M Crespi, Edmund YW Seto, and May C Wang (2014). "Semiparametric Estimation of the Impacts of Longitudinal Interventions on Adolescent Obesity using Targeted Maximum-Likelihood: Accessible Estimation with the ltmle Package". In: *Journal of Causal Inference* 2.1, pp. 95–108.



- Tran, Linh, Constantin T Yiannoutsos, Beverly S Musick, Kara K Wools-Kaloustian, Abraham Siika, Sylvester Kimaiyo, Mark J van der Laan, and Maya Petersen (2016). "Evaluating the Impact of a HIV Low-Risk Express Care Task-Shifting Program: a case study of the targeted learning roadmap". In: *Epidemiologic Methods* 5.1, pp. 69–91.
- Rose, Sherri (2011). "Causal Inference for Case-Control Studies". PhD thesis. University of California, Berkeley.
- MTN Statement on Decision to Discontinue Use of Oral Tenofovir Tablets in VOICE, a Major HIV Prevention Study in Women (2011). URL: <http://www.mtnstopshiv.org/node/3619>.
- MTN Statement on Decision to Discontinue Use of Tenofovir Gel in VOICE, a Major HIV Prevention Study in Women (2011). URL: <http://www.mtnstopshiv.org/node/3909>.
- Rubin, Donald B (1974). "Estimating causal effects of treatments in randomized and nonrandomized studies." In: *Journal of Educational Psychology* 66.5, p. 688.
- Van der Laan, Mark J and Susan Gruber (2011). "Targeted minimum loss based estimation of an intervention specific mean outcome". In: *U.C. Berkeley Division of Biostatistics Working Paper Series*.
- Hendrix, Craig W, Beatrice A Chen, Vijayanand Guddera, Craig Hoesley, Jessica Justman, Clemensia Nakabiito, Robert Salata, Lydia Soto-Torres, Karen Patterson, Alexandra M Minnis, et al. (2013). "MTN-001: randomized pharmacokinetic cross-over study comparing tenofovir vaginal gel and oral tablets in vaginal tissue and other compartments". In: *PloS One* 8.1, e55013.
- Kiser, Jennifer J, Courtney V Fletcher, Patricia M Flynn, Coleen K Cunningham, Craig M Wilson, Bill G Kapogiannis, Hanna Major-Wilson, Rolando M Viani, Nancy X Liu, Larry R Muenz, et al. (2008). "Pharmacokinetics of antiretroviral regimens containing tenofovir disoproxil fumarate and atazanavir-ritonavir in adolescents and young adults with human immunodeficiency virus infection". In: *Antimicrobial Agents and Chemotherapy* 52.2, pp. 631–637.
- Haberer, Jessica E, Jared M Baeten, James Campbell, Jonathan Wangisi, Elly Katabira, Allan Ronald, Elioda Tumwesigye, Christina Psaros, Steven A Safren, Norma C Ware, et al. (2013). "Adherence to antiretroviral prophylaxis for HIV prevention: a substudy cohort within a clinical trial of serodiscordant couples in East Africa". In: *PLoS Medicine* 10.9, e1001511.
- Donnell, Deborah, Jared M Baeten, Namandjé N Bumpus, Justin Brantley, David R Bangsberg, Jessica E Haberer, Andrew Mujugira, Nelly Mugo, Patrick Ndase,

- Craig Hendrix, et al. (2014). "HIV protective efficacy and correlates of tenofovir blood concentrations in a clinical trial of PrEP for HIV prevention". In: *Journal of Acquired Immune Deficiency Syndromes* (1999) 66.3, p. 340.
- Burns, Rebecca N, Craig W Hendrix, and Ayyappa Chaturvedula (2015). "Population pharmacokinetics of tenofovir and tenofovir-diphosphate in healthy women". In: *The Journal of Clinical Pharmacology* 55.6, pp. 629–638.
- Lu, Yanhui, Vineet Goti, Ayyappa Chaturvedula, Jessica E Haberer, Michael J Fossler, Mark E Sale, David Bangsberg, Jared M Baeten, Connie L Celum, and Craig W Hendrix (2016). "Population pharmacokinetics of tenofovir in HIV-1 uninfected members of sero-discordant couples and effect of dose reporting methods". In: *Antimicrobial Agents and Chemotherapy*, AAC–00559.
- R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: <https://www.R-project.org/>.
- Hendrix, Craig W, Adriana Andrade, Namandjé N Bumpus, Angela D Kashuba, Mark A Marzinke, Ayana Moore, Peter L Anderson, Lane R Bushman, Edward J Fuchs, Ilene Wiggins, et al. (2016). "Dose frequency ranging pharmacokinetic study of tenofovir-emtricitabine after directly observed dosing in healthy volunteers to establish adherence benchmarks (HPTN 066)". In: *AIDS Research and Human Retroviruses* 32.1, pp. 32–43.
- Hugonnet, Stephane, Frank Mosha, James Todd, Kokugonza Mugeye, Arnoud Klokke, Leonard Ndeki, David Ross, Heiner Grosskurth, and Richard Hayes (2002). "Incidence of HIV infection in stable sexual partnerships: a retrospective cohort study of 1802 couples in Mwanza Region, Tanzania." In: *Journal of Acquired Immune Deficiency Syndromes* (1999) 30.1, pp. 73–80.
- Beyrer, Chris, Stefan D Baral, Frits van Griensven, Steven M Goodreau, Suwat Chariyalertsak, Andrea L Wirtz, and Ron Brookmeyer (2012). "Global epidemiology of HIV infection in men who have sex with men". In: *The Lancet* 380.9839, pp. 367–377.
- Heffron, Renee, R Scott McClelland, Jennifer E Balkus, Connie Celum, Craig R Cohen, Nelly Mugo, Elizabeth Bukusi, Deborah Donnell, Jairam Lingappa, James Kiarie, et al. (2017). "Efficacy of oral pre-exposure prophylaxis (PrEP) for HIV among women with abnormal vaginal microbiota: a post-hoc analysis of the randomised, placebo-controlled Partners PrEP Study". In: *The Lancet HIV* 4.10, e449–e456.

- Hillier, SL, LA Meyn, K Bunge, M Austin, BJ Moncla, CS Dezzutti, and LC Rohan (2017). "Impact of vaginal microbiota on genital tissue and plasma concentrations of tenofovir". In: *Conference on Retroviruses and Opportunistic Infections (CROI)*. Vol. 2017.
- Klatt, Nichole R, Ryan Cheu, Kenzie Birse, Alexander S Zevin, Michelle Perner, Laura Noël-Romas, Anneke Grobler, Garrett Westmacott, Irene Y Xie, Jennifer Butler, et al. (2017). "Vaginal bacteria modify HIV tenofovir microbicide efficacy in African women". In: *Science* 356.6341, pp. 938–945.
- Velloza, Jennifer and Renee Heffron (2017). "The Vaginal Microbiome and its Potential to Impact Efficacy of HIV Pre-exposure Prophylaxis for Women". In: *Current HIV/AIDS Reports* 14.5, pp. 153–160.

## Chapter 3

# Pharmacokinetic and Pharmacodynamic Modeling of Tenofovir

### 3.1 Introduction

This chapter extends the work of Chapter 2 by integrating pharmacokinetic models into our causal analysis of the protective effect of PrEP drug concentration against HIV infection. Treating output from the PK model, rather than raw concentration levels, as the exposure allows us to account for both the time between a study participant's last PrEP dose and when their plasma sample was taken and the known variability of drug concentrations in the body over time. The work in this chapter largely supports the conclusions from Chapter 2, but it also allows us to analyze the effect of sustaining higher levels of tenofovir drug concentration in the VOICE oral arms than was possible using the raw concentration data. In the remainder of this section, we provide background into pharmacokinetic modeling.

In Section 3.2 we discuss the methods for modeling both PrEP drug concentrations and their effect on the the the risk of HIV infection. In Section 3.3 we present our results and compare them to the results from Chapter 2. We conclude in Section 3.4 with a discussion of ongoing and future areas of work.

Pharmacokinetics (PK) is the study of the behavior of a drug in the body over time. Drug characteristics typically reported include absorption, distribution, metabolism, and excretion (Nelson, 1961; Dudley, 1995; Zhang et al., 2006; Mould and Upton, 2013; Shargel, Andrew, and Wu-Pong, 2015). PK modeling uses drug concentration data sampled at known times from one or more individuals with known dosage regimens. Population PK modeling uses data from multiple individuals and distinguishes between population and individual level characteristics of drug behavior in the body (Mould and Upton, 2013). Often, a mixed-effects model is used to characterize a population model wherein fixed effects represent structural population level typical values and random effects variability between individuals and between occasions. A second level of random effects is typically included as well to reflect residual variability, or error in model predictions (Bonate and Steimer, 2006).

Typically, non-linear mixed effects models are used for pharmacological modeling of drug concentrations in the body. The NON-linear Mixed-Effects Modeling (NONMEM) software is frequently used to estimate parameters from a mixed-effects model (Beal et al., 2009). Parameters of interest include the following: absorption rate constant, volumes in the central and peripheral compartments, bioavailability, the proportion of drug administered that reaches circulation, and

clearance, the rate at which a drug is completely removed from some biologic matrix (Wakefield, Aarons, and Racine-Poon, 1999; Mould and Upton, 2013; Upton and Mould, 2014). The mean values of the PK parameters (typical values) are fixed effects. Variations around these means (both within and between individuals) are described by random effects. Additionally, individual covariates such as age, sex, and creatinine clearance modify the fixed effects PK parameters. Residual variability may be characterized by additive and/or proportional error parameters. Endpoints for PK models include area under the concentration versus time curve (AUC), peak (maximum) concentration, time to peak concentration, half-life, trough (minimum) concentration, and time above a predefined threshold. Both likelihood and Bayesian approaches may be used to estimate the parameters of a PK model (Mallet, 1986; Racine-Poon and Wakefield, 1998; Wakefield, Aarons, and Racine-Poon, 1999).

### **3.1.1 Pharmacodynamic Modeling**

In contrast to PK modeling, pharmacodynamic (PD) modeling examines the effects of the drug on the body, in terms of both drug safety and effectiveness. Often, the two are distinguished by the following: “what the body does to the drug” (pharmacokinetic) versus “what the drug does to the body” (pharmacodynamic). Typically, one or more of the endpoints of the PK model is used as a measure of exposure in the PD model and related to some outcome of interest in the body. The analysis conducted in Chapter 2 using longitudinal TMLE is one such example of a PD analysis.

### 3.1.2 Pharmacokinetic Studies of PrEP

A number of PK studies have been conducted to better understand the relationship between different PrEP dosage strategies and drug concentrations in the body in a range of populations. PK models developed from these studies quantify levels of TDF and FTC and their active metabolites, tenofovir diphosphate (TFV-DP) and emtricitabine triphosphate (FTC-TP) in different compartments, including blood plasma, peripheral blood mononuclear cells, and tissue.

PK models have been developed from small studies with directly observed dosing, including the HIV Prevention Trials Network (HPTN) 066 Study of TDF and FTC in healthy men and women. Hendrix et al. (2016) established steady-state concentrations of TFV and FTC in the plasma and TFV-DP and FTC-TP in PBMCs under different oral dosing regimens. Patterson et al. (2011) developed a PK model in healthy men and women collecting concentration measurements over the course of two weeks after one oral dose of tenofovir in various compartments including blood plasma, genital secretions, and mucosal tissues. Importantly, they characterized the decay of TFV, which is highly relevant for alternative dosing strategies than daily dosing. They also determined that concentrations of both TFV and TFV-DP were 100 times higher in colorectal tissues compared to the female genital tract (Patterson et al., 2011). This finding indicates that target drug concentrations and dosing strategies may differ between populations at risk of HIV and by sites of exposure. Louissaint et al. (2013) developed a PK model using concentration measurements over time in plasma, tissues, blood CD4 cells, and PBMCs after a single dose of oral PrEP. This work has further contributed to the

understanding of the rise, accumulation, and decay of tenofovir levels throughout the body.

Other PK models have been developed using data from clinical trials without directly observed dosing. For example, (Burns, Hendrix, and Chaturvedula, 2015) developed a population PK model of tenofovir using data from the MTN-001 open label clinical trial. In this trial, healthy women were randomized to either or both of oral tenofovir or tenofovir vaginal gel. An important contribution of the PK model developed from the MTN-001 trial was that it accounted for varying levels of study participant adherence. A PK model developed from the Cell-PrEP study established steady-state concentrations of PrEP in rectal mononuclear cells, an important site for men who have sex with men (Seifert et al., 2014).

## 3.2 Methods

We integrate the PK model developed by Vucicevic, Savic, and Hendrix (In Preparation, 2018) into the longitudinal TMLE framework in Chapter 2. Instead of using raw plasma concentrations as the drug exposures, we use the average daily steady-state concentrations at each participant’s study visit generated from the PK model. The raw concentrations used in the analysis in Chapter 2 are taken at one point in time and may not represent the level of drug in a study participant’s body over a period of time. In contrast, the PK model accounts for the time between when a concentration measurement was taken and when the study participant reported taking their most recent dose.



### 3.2.1 The Pharmacokinetic Model

The pharmacokinetic model that we use is a two-compartment nonlinear mixed-effects combined additive and proportional residual error model developed by Vucicevic, Savic, and Hendrix ([In Preparation, 2018](#)). The model integrates PK data from the oral TDF and FTC/TDF arms of three PrEP clinical trials. It has been built using the largest known clinical trial database of oral tenofovir measurements. Data from the Partners PrEP and VOICE studies, both of which are described in Chapter 2, were used to build the model. Additionally, the model integrated pharmacokinetic data from the Pre-exposure Prophylaxis Initiative (iPrEx) trial, a multinational, randomized, placebo controlled trial of daily TDF/FTC in 2499 men and transgender women who have sex with men; (Grant et al., [2010](#); Anderson et al., [2012](#)). As with the Partners PrEP and VOICE studies, plasma drug concentrations in the iPrEx trial were measured from stored samples for a subset of study participants in the active treatment arms. A nested case-control substudy design was used for concentration sampling. In all three of the PrEP studies, the oral dose in the active treatment arms (whether TDF or FTC/TDF) included 300 mg of tenofovir, given every 24 hours.

The PK model includes two main compartments: the blood plasma (a central compartment) and a peripheral compartment. They are related by a distribution parameter  $Q$ , the intercompartmental clearance between the central and peripheral compartments. Absorption of the drug from the oral dosage into the plasma is characterized by the parameter  $k_A$ , and clearance of the drug from the plasma by  $CL$ . Figure 3.1 illustrates the structure of the model and the parameters that relate

the movement of drug into and between the different compartments.

Let  $A_1$ ,  $A_2$ , and  $A_3$  denote the amount of drug in the dosage, oral compartment, and peripheral compartment, respectively. Additionally, let  $V_2$  and  $V_3$  denote the volumes in the central and peripheral compartments and  $C_2$  and  $C_3$  their respective concentrations. The parameters are related by the following set of differential equations:

$$\frac{dA_1}{dt} = -k_A A_1;$$

$$\frac{dA_2}{dt} = k_A A_1 - \frac{CL}{V_2} A_2 - \frac{Q}{V_2} A_2 + \frac{Q}{V_3} A_3;$$

$$\frac{dA_3}{dt} = \frac{Q}{V_2} A_2 - \frac{Q}{V_3} A_3.$$

Let  $C_2 = \frac{A_2}{V_2}$  and  $C_3 = \frac{A_3}{V_3}$  denote the concentrations in the central and peripheral compartments, respectively. When a uniform dose  $D$  is given every  $\tau$  hours, the concentrations in the central and peripheral compartments as a function of hours since the last dose,  $t$ , are given by the following two equations, whose general form are given in Wagner (1975):

$$\begin{aligned}
(C_2)_\eta &= \frac{k_A D}{V_2} \left[ \left( \frac{1 - e^{-\eta \alpha \tau}}{1 - e^{-\alpha \tau}} \right) \left( \frac{Q/V_3 - \alpha}{(k_A - \alpha)(\beta - \alpha)} \right) e^{-\alpha t} \right. \\
&\quad + \left( \frac{1 - e^{-\eta \beta \tau}}{1 - e^{-\beta \tau}} \right) \left( \frac{Q/V_3 - \beta}{(k_A - \beta)(\alpha - \beta)} \right) e^{-\beta t} \\
&\quad \left. + \left( \frac{1 - e^{-\eta k_A \tau}}{1 - e^{-\beta \tau}} \right) \left( \frac{Q/V_3 - k_A}{(\alpha - k_A)(\beta - k_A)} \right) e^{-k_A t} \right]; \\
(C_3)_\eta &= \frac{Q/V_2 k_A D}{V_3} \left[ \left( \frac{1 - e^{-\eta \alpha \tau}}{1 - e^{-\alpha \tau}} \right) \frac{e^{-\alpha t}}{(k_A - \alpha)(\beta - \alpha)} \right. \\
&\quad + \left( \frac{1 - e^{-\eta \beta \tau}}{1 - e^{-\beta \tau}} \right) \frac{e^{-\beta t}}{(k_A - \beta)(\alpha - \beta)} \\
&\quad \left. + \left( \frac{1 - e^{-\eta k_A \tau}}{1 - e^{-\beta \tau}} \right) \frac{e^{-k_A t}}{(\alpha - k_A)(\beta - k_A)} \right].
\end{aligned}$$

where

$$\begin{aligned}
\alpha &= \frac{1}{2} \left[ (Q/V_2 + CL/V_2 + Q/V_3) + \sqrt{(Q/V_2 + CL/V_2 + Q/V_3)^2 - 4Q/V_3 CL/V_2} \right]; \\
\beta &= \frac{1}{2} \left[ (Q/V_2 + CL/V_2 + Q/V_3) - \sqrt{(Q/V_2 + CL/V_2 + Q/V_3)^2 - 4Q/V_3 CL/V_2} \right].
\end{aligned}$$

The parameters  $(k_A, V_2, TVCL, V_3, Q)$  are fixed effects estimated by NONMEM.

Additionally, NONMEM estimates the random effect  $\eta_1$ , which describes individual level clearance:  $CL = TVCL e^{\eta_1}$ . The population level concentrations in the two compartments (as functions of time) can be estimated by setting the random effect  $\eta_1$  to its mean value, 0. An individual concentration curve is generated by using the person-specific random effect estimated.

Finally, residual error is characterized by two population-level parameters: an additive and proportional error constant,  $W_a$  and  $W_p$ , respectively. Let  $Y^{PK}$  be an individual predicted concentration based on the PK model and  $Y^{TRUE}$  the true concentration. The additive and proportional residual variability structure defines:  $Y^{TRUE} = Y^{PK} + W_a + Y^{PK} \times W_p$ .

The PK model incorporates varying levels of participant adherence and resulting concentration measurements below the limit of quantification. Although plasma concentration samples were used from participants with a reported dosing within 100 hours, 43% of all plasma measurements across the studies were below the limit of quantification (which was 0.31 ng/mL in Partners PrEP and VOICE and 10 ng/mL in iPrEx), and 50% of participants had at least one observation below the limit of quantification. To account for below quantifiable concentration measurements, Vucicevic, Savic, and Hendrix ([In Preparation, 2018](#)) maximized the likelihood for observations above the limit of quantification and treated observations below the limit of quantification as censored (Ahn et al., [2008](#); Bergstrand and Karlsson, [2009](#)).

Participant non-adherence was accounted for by estimating the probability that a study participant took their assigned drug at a particular study visit. This

subject-visit specific parameter is denoted by  $p_{\text{adher}}$ . A fully adherent individual would have  $p_{\text{adher}} = 1$ , and decreasing values of  $p_{\text{adher}}$  reflect lower levels of adherence. A mixture model that identifies adherent ( $p_{\text{adher}} = 1$ ) and non-adherent ( $p_{\text{adher}} = 0$ ) subpopulations was used.

An empirical Bayes approach was used to estimate the subject-specific pharmacokinetic parameters in NONMEM with a First Order Conditional Estimation (FOCE) method (Beal et al., 2009). The standard errors of the empirical Bayes estimates of the random effects were estimated using the method of Kang et al. (2012). Average daily tenofovir concentrations at steady-state for a particular study participant at each of their visits were estimated using the fixed effects estimates, their random effect estimates, and their individual probabilities of taking the drug at each study visit.

The exposure we will use in our PD analysis is the average steady-state concentration at each participant-visit. Steady-state is the period in which the rates of drug intake into and elimination from a given compartment are equal. The average daily steady-state plasma concentration at that visit,  $C_p^{ss}$ , is calculated as:

$$C_p^{ss} = \frac{D * p_{\text{adher}}}{CL * \tau}$$

### 3.2.2 Longitudinal Targeted Maximum Likelihood Estimation

We repeat the longitudinal TMLE analysis from Chapter 2 using the average daily tenofovir concentrations at each participant-visit from the PK model in place of the raw concentrations. We use a threshold of 40 ng/mL to define a high concentration

regimen. The PK model was developed using only the oral TDF and oral FTC/TDF arms of PrEP trials, so we omit the gel arm of VOICE in our analyses.

## 3.3 Results

### 3.3.1 Pharmacokinetic Model Results

The parameters of the pharmacokinetic model from this data were similar to those determined from smaller, more intensive PK sampling studies for PrEP. Notably the subject-visit specific probability of adherence,  $p_{\text{adher}}$ , have a bimodal distribution for all of the studies.

We compare the distributions of the raw and modeled average tenofovir concentrations for both the Partners PrEP and VOICE studies in Figure 3.2. PK visits that occurred more than 100 hours after the last reported oral drug dose are excluded from the raw drug concentration data so that the raw and modeled tenofovir concentrations are from the same set of participant visits. Both the raw and modeled concentrations have multimodal distributions, differentiating study participants into several adherence subpopulations. There are a substantial number of raw concentrations below the limit of quantification (0.31 ng/mL), particularly among the VOICE oral study arms. In contrast, all of the average daily steady-state tenofovir concentrations generated by the PK model are above the limit of quantification. This likely reflects the considerable shrinkage of the parameter  $p_{\text{adher}}$  (by 43%), so that all subject-visit probabilities of adherence are nonzero. As a result, average steady-state concentrations for non-adherent participants may be inflated.

In Figure 3.3 we compare the raw and steady-state average daily concentrations from the PK model by participant visit. Both the raw and modeled concentration measurements are clustered into different subpopulations of adherers, occasional adherers, and non-adherers. However, their values are not directly comparable because the raw concentration measurements are from a point in time (typically shortly before the next dose is taken) and approximate trough concentrations, while the modeled values are average steady-state concentrations over the course of a day. The modeled concentrations also include considerable shrinkage towards typical (high) levels of adherence.

### **3.3.2 LTMLE Analysis Results**

We present the results of the LTMLE analysis looking at the estimated risk of setting average daily steady-state concentration to above or below 40 ng/mL at each of the first three semi-annual visits in Table 3.1. The estimated HIV incidences and relative risk reductions are very similar in each of the studies comparing the analyses adjusting only for baseline variables to those adjusting for both baseline variables and time-varying risk behaviors. In both studies, setting participants to follow a concentration regimen of above 40 ng/mL at each of the 6, 12, and 18 month study visits suggests a protective effect against HIV infection. However, the confidence intervals in the Partners PrEP analysis when adjusting for time-varying risk and in both of the VOICE analyses are very wide and overlap zero.

### 3.3.3 Comparison of Results Using Raw vs. Modeled Concentrations

The results of the longitudinal TMLE analyses are similar whether using the raw or modeled concentrations. In the Partners PrEP study, when adjusting for time-varying risk behaviors, the estimated relative risk reduction setting average daily steady-state concentration (based on the PK model output) to above 40 ng/mL at each of the 6, 12, and 18 month visits is 91% (95% CI: -13, 98). Similarly, the estimated relative risk reductions (based on the raw concentration) setting concentration to above 0.31 ng/mL and to above 40 ng/mL are 82% (95% CI: 8, 94) and 84% (95% CI: 10, 95), respectively. The estimated relative risk reductions are also similar when adjusting only for baseline variables, although the confidence intervals are narrower.

In the VOICE study, using the modeled concentrations we estimate a 62% (95% CI: -9, 90) when setting average daily steady-state concentration to above 40 ng/mL and adjusting for both baseline variables and time-varying risk. As with the analyses of the raw concentration data, we estimate the effect of longitudinal PrEP drug concentration to be in the direction of benefit, but the confidence interval for the estimated effectiveness overlaps zero. As discussed in Chapter 2, our estimate of the risk of seroconversion setting (raw) concentration to above 40 ng/mL at each of the 6, 12, and 18 month visits in the VOICE oral arm was unstable, so we can only compare setting observed concentration to above versus below the limit of quantification (0.31 ng/mL) to setting the modeled average concentrations to above versus below 40 ng/mL.



As previously discussed, study visits where the time from last reported dose taken to time of specimen collected exceeded 100 hours, as well as those where the time was unknown, were excluded from the PK model. As a result, more study participants were artificially censored when performing the LTMLE analysis on the PK output as compared to the raw plasma concentrations. The rationale for excluding participants after 100 hours was that based on its half life, by 100 hours after the most recent dose, all of the tenofovir has been completely eliminated from the body. Therefore, there may have been fewer observations where a study participant had no drug when they were at risk of seroconversion in this analysis as compared to the analysis using raw tenofovir concentrations. We repeated the longitudinal TMLE analysis using the raw concentration data restricted to PK measurements taken within 100 hours of the last reported dose. Our estimates of HIV incidences and effectiveness were almost identical to those from the analyses using all of the raw concentration measurements.

### **3.4 Discussion**

We utilize the methodology of Chapter 2 in combination with pharmacokinetic models developed from the oral TDF and FTC/TDF arms of the Partners PrEP and VOICE studies, with additional pharmacokinetic data from a third clinical trial. Using modeled rather than raw concentrations reduces the influence of the time of recent dose, especially if within the last day or two, on measured concentration. This is important because plasma tenofovir has a short enough half-life that concentration levels will vary substantially over the course of a one

day dosing window. The population PK model borrows information between individuals both within the same study and across studies while still accounting for variability among study participants and among visits for an individual participant. This is particularly beneficial given the sparsity of concentration measurements. Although not directly comparable, the results of the longitudinal TMLE analysis using average daily concentrations generated by the PK model are consistent with the results using raw concentrations. The raw concentrations are systematically lower than the steady-state concentrations from the PK model, so there were too few raw concentration measurements in the VOICE oral arm exceeding 40 ng/mL to estimate this effect.

An important limitation is that we do not know when study participants are exposed to HIV. The PK model has the strength of accounting for the length of time between when the last dose and the plasma sample were taken, thus allowing us to capture a study participant's average concentration over a period of time. However, the true exposures of interest, a study participant's drug concentration(s) at their time(s) of HIV exposure are unattainable. Additionally, this analysis requires the same set of assumptions as in Chapter 2. Further, the average daily steady-state concentrations generated from the PK model require the strong assumption that average daily steady-state concentration is proportional to the probability of adherence at a study visit.

### 3.4.1 Future Directions

Further development of the pharmacokinetic model will largely guide areas of potential future research. Thus far, the PK model has focused on drug concentrations from oral tenofovir dosing. An expansion of the model to include gel dosing will allow us to extend our analyses in Chapter 2 of the VOICE gel arm. As discussed in Hendrix (2013), tenofovir concentrations in multiple sites of the body are relevant for protecting against HIV infection. The most important compartments differ by at-risk population due to both biological differences and varying sites of infection (Hendrix, 2013). Understanding their relative contributions to HIV protection remains an ongoing challenge. An extension of the PK model to quantify the movement of drug in additional compartments, such as tissues and peripheral blood mononuclear cells, may allow us to estimate the protective effect of drug concentrations in different compartments of the body that may be more relevant than blood plasma. This may be particularly useful for comparing different routes of dosing, such as oral versus vaginal gel tenofovir. Mediation analyses may be a useful tool for discerning the protective effects of concentration levels in different compartments against HIV risk.

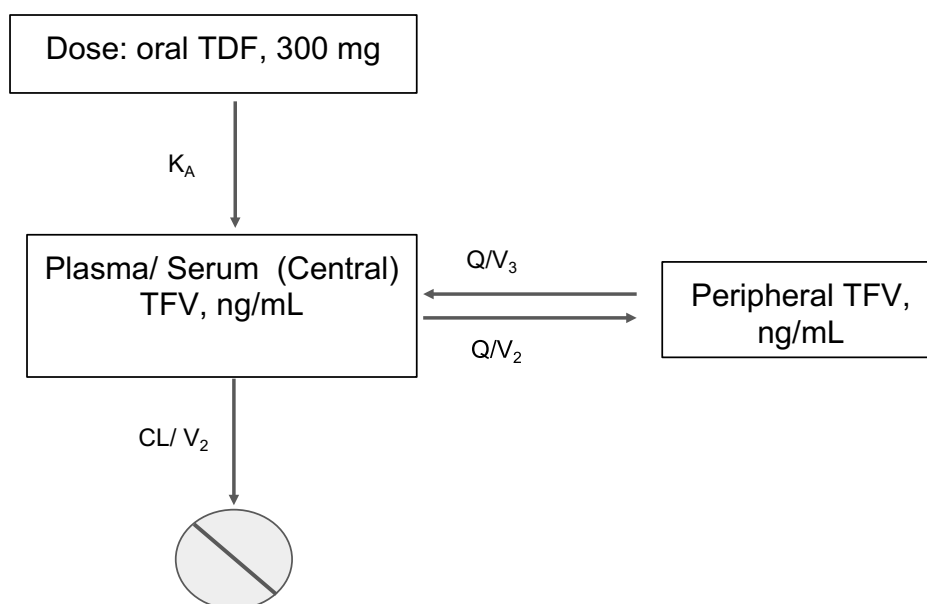
Another important development may be modeling additional patient level covariates. In the PK model so far, adherence has been categorized using pharmacological measures (drug concentrations). As a result, it is not possible to model concentrations for a study participant not included in a concentration substudy (as the PK measures needed to characterize their levels of adherence are unmeasured). Incorporating additional study participant level characteristics may allow for the

prediction of drug concentration levels in study participants in the treatment arms who were not included in the concentration substudies.

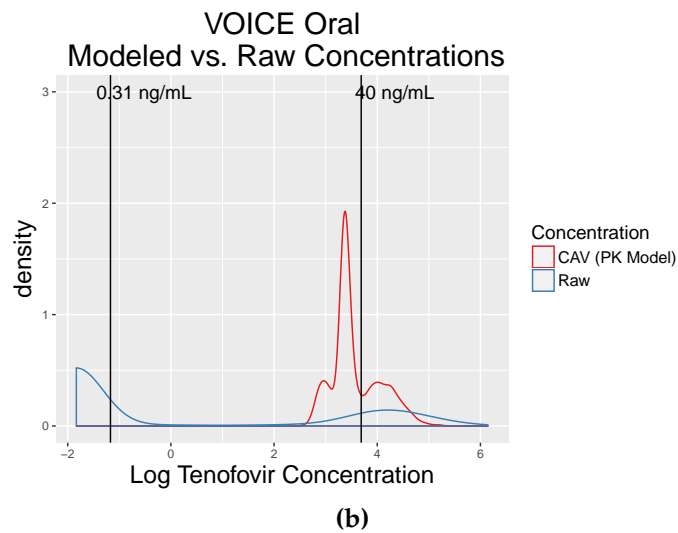
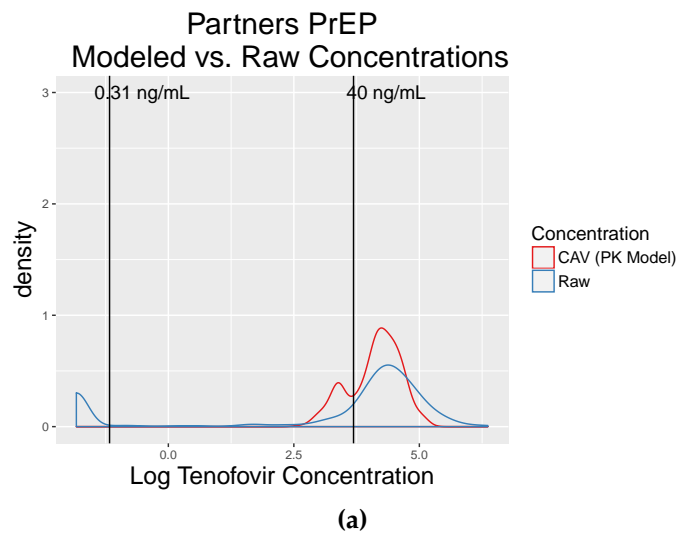
Finally, we are exploring the possibility of developing a Bayesian hierarchical model as a method to integrate the output from the PK model into an analysis of the effect of tenofovir drug concentration on risk of HIV infection. This framework has the potential advantage of better utilizing the characterizations of variability output from the PK model in the form of parameter distributions for both the fixed and random effects. In particular, the PK model can be used to generate for each study participant a collection of possible concentration-time curves and associated average daily tenofovir concentrations from the individual level parameter distribution. These generated concentrations can represent the bottom level in the hierarchical population PK model.

An ongoing challenge with the Bayesian model is accounting for the case-cohort and nested case-control sampling schema from the trials. In the LTMLE analyses, both with the observed and modeled concentrations, we used weighting. We do not know of a natural extension of the inverse weighting approach in the Bayesian analysis. One promising method, which has been used in other PrEP analyses, including a PK/PD analysis by Anderson et al. (2012) is multiple imputation for all study participants not included in the concentration substudies. However, as discussed in Chapter 2, adherence as reported by study participants and measured from returned pills often is only weakly correlated with adherence levels estimated from plasma samples, especially in the VOICE study (Van der Straten et al., 2014; Marrazzo et al., 2015; Dai et al., 2015). We also found that estimated levels of

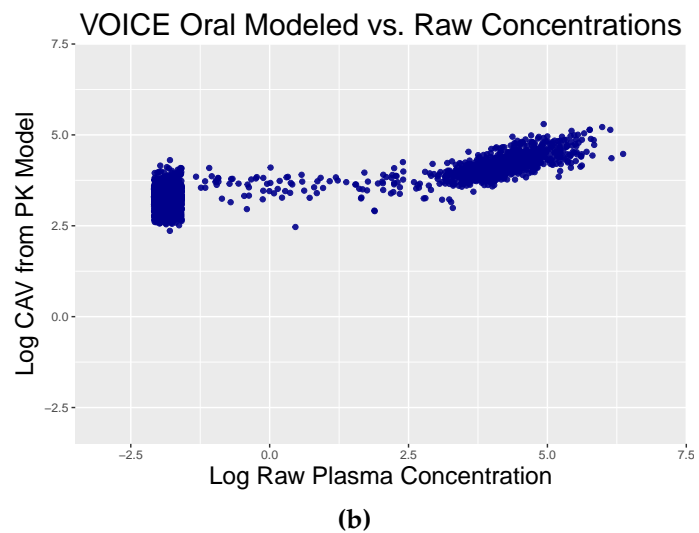
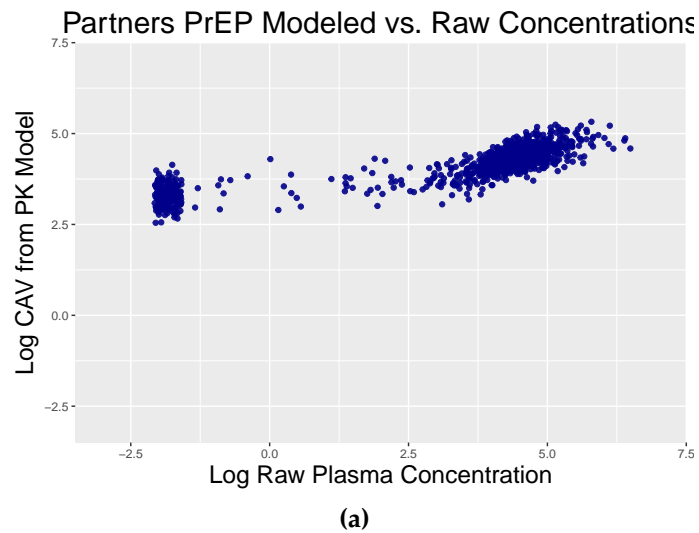
adherence from the the population PK model differed substantially from self reported levels of adherence and adherence calculated using pill counts. As a result, the reliability of any predictions of concentration levels for study participants lacking any pharmacological measurements seems limited. Incorporating the sampling schema in a Bayesian framework may be an alternative approach to pursue, but it may involve many computational challenges and remains an open problem.



**Figure 3.1: The two compartment population-level PK model used to model tenofovir plasma concentrations from oral dosing.**  $K_A$ ,  $Q/V_2$  and  $Q/V_3$  are first order rate constants that characterize the movement of tenofovir from the oral dosage to the plasma (compartment 2) to the peripheral compartment (compartment 3). This figure is adapted from Scheme 4 of *Fundamentals of Clinical Pharmacokinetics* (p. 198) by J. G. Wagner, 1975, Drug Intelligence Publications, which illustrates the basic structure of the two-compartment model we used.



**Figure 3.2: Comparison of the distribution of raw plasma tenofovir concentrations from stored samples and estimated daily concentrations from the pharmacokinetic model in the Partners PrEP (top) and VOICE (bottom) oral arms. The two vertical lines show the values of 0.31 ng/mL, the limit of quantification of tenofovir for both studies, and 40 ng/mL, associated with typical daily tenofovir dosing trough (pre-dose) plasma concentrations.**



**Figure 3.3: Comparison of average daily tenofovir levels from the pharmacokinetic model to raw plasma tenofovir concentrations from stored samples.**



Study	Adjustment Variables	Estimated Incidence Setting Concentration to Consistently Below 40 ng/mL	Estimated Incidence Setting Concentration to Consistently Above 40 ng/mL	% Relative Risk Reduction $100 \times (1\text{-relative risk})$
Partners PrEP	Baseline	2 (0.4, 7.6)	0.2 (0.1, 0.4)	91 (38, 98)
Partners PrEP	Baseline + Risk	2.1 (0.2, 6.2)	0.2 (0.1, 0.4)	91 (-13, 98)
VOICE Oral	Baseline	2.9 (1.9, 4.4)	1.1 (0.4, 3.1)	63 (-20, 89)
VOICE Oral	Baseline + Risk	2.9 (2, 4.4)	1.1 (0.3, 3.1)	62 (-9, 90)

**Table 3.1: Estimates and 95% confidence intervals based on the longitudinal TMLE of the incidence of HIV in the Partners PrEP and VOICE oral study populations when concentration (average daily tenofovir based on the PK model) is set to above or below 40 ng/mL at each of the 6, 12, and 18 months study visits.** Incidence is measured on the scale of seroconversions per 100 person-years. The final column gives estimates of the associated relative risk reduction (1-relative risk) comparing setting concentration above versus below 40 ng/mL. We perform the longitudinal analyses first adjusting for only baseline variables ('Baseline') and then adjusting both for baseline variables and reported unprotected sex at each of the semi-annual visits during the six months prior to the study visit ('Baseline + Risk'). Baseline variables for the Partners PrEP study were age and gender, and baseline variables for the VOICE study were age, marital status, and HSV-2 status.

# References

- Nelson, Eino (1961). "Kinetics of drug absorption, distribution, metabolism, and excretion". In: *Journal of Pharmaceutical Sciences* 50.3, pp. 181–192.
- Dudley, Michael N (1995). "Clinical pharmacokinetics of nucleoside antiretroviral agents". In: *Journal of Infectious Diseases* 171.Supplement\_2, S99–S112.
- Zhang, Liping, Vikram Sinha, S Thomas Forgue, Sophie Callies, Lan Ni, Richard Peck, and Sandra RB Allerheiligen (2006). "Model-based drug development: the road to quantitative pharmacology". In: *Journal of Pharmacokinetics and Pharmacodynamics* 33.3, pp. 369–393.
- Mould, DR and Richard Neil Upton (2013). "Basic concepts in population modeling, simulation, and model-based drug development part 2: introduction to pharmacokinetic modeling methods". In: *CPT: Pharmacometrics & Systems Pharmacology* 2.4, pp. 1–14.
- Shargel, Leon, BC Andrew, and Susanna Wu-Pong (2015). *Applied biopharmaceutics & pharmacokinetics*. McGraw-Hill Medical Publishing Division.
- Bonate, Peter L and Jean-Louis Steimer (2006). *Pharmacokinetic-Pharmacodynamic Modeling and Simulation*. Springer.
- Beal, SSLBBABRJ, LB Sheiner, A Boeckmann, and RJ Bauer (2009). "NONMEM user's guides (1989–2009)". In: *Icon Development Solutions, Ellicott City, MD*.
- Wakefield, Jon, Leon Aarons, and Amy Racine-Poon (1999). "The Bayesian approach to population pharmacokinetic/pharmacodynamic modeling". In: *Case Studies in Bayesian Statistics*, pp. 205–265.
- Upton, RN and DR Mould (2014). "Basic concepts in population modeling, simulation, and model-based drug development part 3: Introduction to pharmacodynamic modeling methods". In: *CPT: Pharmacometrics & Systems Pharmacology* 3.1, pp. 1–16.
- Mallet, AA (1986). "A maximum likelihood estimation method for random coefficient regression models". In: *Biometrika* 73.3, pp. 645–656.

- Racine-Poon, Amy and Jon Wakefield (1998). "Statistical methods for population pharmacokinetic modelling". In: *Statistical Methods in Medical Research* 7.1, pp. 63–84.
- Hendrix, Craig W, Adriana Andrade, Namandjé N Bumpus, Angela D Kashuba, Mark A Marzinke, Ayana Moore, Peter L Anderson, Lane R Bushman, Edward J Fuchs, Ilene Wiggins, et al. (2016). "Dose frequency ranging pharmacokinetic study of tenofovir-emtricitabine after directly observed dosing in healthy volunteers to establish adherence benchmarks (HPTN 066)". In: *AIDS Research and Human Retroviruses* 32.1, pp. 32–43.
- Patterson, Kristine B, Heather A Prince, Eric Kraft, Amanda J Jenkins, Nicholas J Shaheen, James F Rooney, Myron S Cohen, and Angela DM Kashuba (2011). "Penetration of tenofovir and emtricitabine in mucosal tissues: implications for prevention of HIV-1 transmission". In: *Science Translational Medicine* 3.112, 112re4–112re4.
- Louissaint, Nicolette A, Ying-Jun Cao, Paul L Skipper, Rosa G Liberman, Steven R Tannenbaum, Sridhar Nimmagadda, Jean R Anderson, Stephanie Everts, Rahul Bakshi, Edward J Fuchs, et al. (2013). "Single dose pharmacokinetics of oral tenofovir in plasma, peripheral blood mononuclear cells, colonic tissue, and vaginal tissue". In: *AIDS Research and Human Retroviruses* 29.11, pp. 1443–1450.
- Burns, Rebecca N, Craig W Hendrix, and Ayyappa Chaturvedula (2015). "Population pharmacokinetics of tenofovir and tenofovir-diphosphate in healthy women". In: *The Journal of Clinical Pharmacology* 55.6, pp. 629–638.
- Seifert, Sharon M, David V Glidden, Amie L Meditz, Jose R Castillo-Mancilla, Edward M Gardner, Julie A Predhomme, Caitlin Rower, Brandon Klein, Becky J Kerr, L Anthony Guida, et al. (2014). "Dose response for starting and stopping HIV preexposure prophylaxis for men who have sex with men". In: *Clinical Infectious Diseases* 60.5, pp. 804–810.
- Vucicevic, Katarina, Radojka M Savic, and Craig W. Hendrix (In Preparation, 2018). "A pharmacokinetic-pharmacodynamic model of tenofovir for HIV pre-exposure prophylaxis".
- Grant, Robert M, Javier R Lama, Peter L Anderson, Vanessa McMahan, Albert Y Liu, Lorena Vargas, Pedro Goicochea, Martín Casapía, Juan Vicente Guanira-Carranza, Maria E Ramirez-Cardich, et al. (2010). "Preexposure chemoprophylaxis for HIV prevention in men who have sex with men". In: *New England Journal of Medicine* 363.27, pp. 2587–2599.

- Anderson, Peter L, David V Glidden, Albert Liu, Susan Buchbinder, Javier R Lama, Juan Vicente Guanira, Vanessa McMahan, Lane R Bushman, Martín Casapía, Orlando Montoya-Herrera, et al. (2012). "Emtricitabine-tenofovir concentrations and pre-exposure prophylaxis efficacy in men who have sex with men". In: *Science Translational Medicine* 4.151, 151ra125–151ra125.
- Wagner, John G (1975). *Fundamentals of Clinical Pharmacokinetics*. Drug Intelligence Publications.
- Ahn, Jae Eun, Mats O Karlsson, Adrian Dunne, and Thomas M Ludden (2008). "Likelihood based approaches to handling data below the quantification limit using NONMEM VI". In: *Journal of Pharmacokinetics and Pharmacodynamics* 35.4, pp. 401–421.
- Bergstrand, Martin and Mats O Karlsson (2009). "Handling data below the limit of quantification in mixed effect models". In: *The AAPS journal* 11.2, pp. 371–380.
- Kang, Dongwoo, Kyun-Seop Bae, Brett E Houk, Radojka M Savic, and Mats O Karlsson (2012). "Standard error of empirical bayes estimate in NONMEM® VI". In: *The Korean Journal of Physiology & Pharmacology* 16.2, pp. 97–106.
- Hendrix, Craig W (2013). "Exploring concentration response in HIV pre-exposure prophylaxis to optimize clinical care and trial design". In: *Cell* 155.3, pp. 515–518.
- Van der Straten, Ariane, Jonathan Stadler, Elizabeth Montgomery, Miriam Hartmann, Busiswe Magazi, Florence Mathebula, Katie Schwartz, Nicole Laborde, and Lydia Soto-Torres (2014). "Women's experiences with oral and vaginal pre-exposure prophylaxis: the VOICE-C qualitative study in Johannesburg, South Africa". In: *PLoS One* 9.2, e89118.
- Marrazzo, Jeanne M, Gita Ramjee, Barbra A Richardson, Kailazarid Gomez, Nyaradzo Mgodli, Gonasagrie Nair, Thesla Palanee, Clemensia Nakabiito, Ariane van der Straten, Lisa Noguchi, et al. (2015). "Tenofovir-based preexposure prophylaxis for HIV infection among African women". In: *New England Journal of Medicine* 372.6, pp. 509–518.
- Dai, James Y, Craig W Hendrix, Barbra A Richardson, Cliff Kelly, Mark Marzinke, Z Mike Chirenje, Jeanne M Marrazzo, and Elizabeth R Brown (2015). "Pharmacological Measures of Adherence and Risk of HIV Acquisition in the VOICE Study". In: *Journal of Infectious Diseases*, jiv333.

## **Chapter 4**

# **Design and Analysis of a Balanced, Covariate-Adjusted Cluster Randomized Trial: the Baltimore CONNECT Project**

Reproduced with permission from:

Ruberman, Claire F., Albert W. Wu, Yanyan Lu, Shuwen Liang, Chidinma Ibe, Christine M. Weston, Lawrence H. Moulton, Sarah Kachur, and Michael Rosenblum (2018). "Design and Analysis of a Balanced, Covariate-Adjusted Cluster Randomized Trial: the Baltimore CONNECT Project." In preparation.

## 4.1 Introduction

### 4.1.1 Cluster randomized trials for health services research

Community engagement interventions may improve health outcomes in populations with complex needs, and their implementation and evaluation are of interest for health services research. Evaluation of such multi-component interventions can be challenging, and published evaluations are often threatened by confounding factors. For example, a large body of literature exists on the promotion of breast-feeding to reduce diarrhoeal diseases in infants, such as Béhar (1975), Victora et al. (1987), Dewey, Heinig, and Nommsen-Rivers (1995), Arifeen et al. (2001), and Lamberti et al. (2011). However, sources of confounding including infant age and mothers' socioeconomic status, education level, reliance on child care, self-perception, and maternal attitude raise methodological challenges (Sauls, 1979; Feachem and Koblinsky, 1984). Similarly, community-based interventions to control the spread of dengue face a range of confounders, including climate and adherence (Heintze, Garrido, and Kroeger, 2007).

Cluster randomized trial designs are useful for evaluating the effectiveness of community-based interventions to reduce issues of confounding and bias (Hayes and Moulton, 2017). Such trials randomize social units, such as classrooms, schools, athletic teams, villages, hospitals, or workplaces (Donner and Klar, 2000). As we will discuss in more detail, the design and analysis of cluster randomized trials require special considerations because, although clusters are randomized to study

arms, individual-level outcomes within a cluster are often of interest to investigators (Donner and Klar, 2000; Hayes and Bennett, 1999). Correlations between individuals within a cluster must be accounted for in both sample size calculations and the analysis of outcomes (Donner, Birkett, and Buck, 1981; Raudenbush, 1997; Hayes and Bennett, 1999).

Many of the advantages of cluster randomized trials (CRTs), are relevant to health services research and community based interventions. In some cases, feasibility calls for designing and implementing randomization on a cluster level (Donner, 1998). Examples include the influential study by Sommer et al. (1986) on the effect of Vitamin A supplementation in Indonesia on reducing childhood mortality and in the mass education intervention in the Community Intervention Trial for Smoking Cessation (Fisher Jr, 1995). In others, the outcome of interest may be on the cluster level, such as when studying herd immunity through vaccination (Smith, Morrow, and Ross, 2015). Cluster randomization may also be appropriate when the primary goal is to demonstrate efficacy at a group or community level (Smith, Morrow, and Ross, 2015). Additionally, ethical considerations may motivate cluster randomization (Fairhurst and Dowrick, 1996; Hussey and Hughes, 2007), although additional challenges, such as informed consent, may arise from cluster randomization (Taljaard et al., 2011; Sim and Dawson, 2012; Weijer et al., 2012). Finally, cluster randomization may be utilized for practical considerations, such as accounting for individual-level contamination and peer-motivated compliance (Hussey and Hughes, 2007).

CRTs present unique statistical challenges. Factors such as subject selection,

influential cluster-level covariates, and the spread of infectious diseases among people in regular contact with one another cause correlations among individuals within a cluster (Donner and Klar, 2000). Identifying clusters by geographic location can result in spatial correlation (Taljaard et al., 2011). These correlations almost always inflate the variance of effect estimates and reduce power and efficiency relative to a trial that randomizes the same number of independent individuals (Donner, Birkett, and Buck, 1981; Donner, 1998; Hayes and Bennett, 1999; Klar and Donner, 2001; Donner and Klar, 2004). Additionally, for a fixed number of individuals, the number of randomized units will be much smaller in a CRT, so the likelihood of chance imbalance on features increases (Moulton, 2004).

The concern about imbalances between treatment groups may be partially addressed by imposing constraints on the randomization schema. Constrained (also referred to as restricted) randomization may be implemented in a CRT, whereby a set of balance criteria on certain covariates are determined prior to the randomization (Moulton, 2004). As detailed in Hayes and Moulton (2017), such criteria may include upper bounds on the difference in means (or proportions) across treatment arms for numerical (or categorical) variables.

In conjunction with constrained randomization, statistical adjustment for chance imbalances in other baseline covariates between treatment groups may be used to improve precision and power in the analysis of CRTs (Gail et al., 1996; Braun and Feng, 2001; Small, Ten Have, and Rosenbaum, 2008). A statistical method for leveraging baseline variables in the analysis, as developed by Small, Ten Have, and Rosenbaum (2008), is described in Section 4.2.2.



Standard analyses of cluster randomized trials are based upon a statistical model to describe the data generating process. Analyses may be performed in one stage as an individual-level analysis, or in two stages as a cluster-level analysis; however, both have several drawbacks. Although Hayes and Moulton (2017) recommend using a cluster-based analysis when the treatment groups contain 15-20 or fewer cluster each, both approaches may lack robustness to the model assumptions when there are very few clusters in each treatment arm.

Additionally, model-based analyses may not be suitable when spillover (interference) occurs between clusters. Spillover, wherein the treatment received by one unit affects the outcomes of another unit, may attenuate or inflate the true effect of a treatment, depending upon its mechanism (Sobel, 2006; Tchetgen and VanderWeele, 2012; Aronow and Samii, 2013; VanderWeele, Tchetgen, and Halloran, 2014). Often, analyses of cluster randomized trials, while accounting for spillover within a cluster, assume no spillover between different clusters. Model-based analyses that do account for spillover between clusters must make additional assumptions about the mechanism of spillover.

Randomization inference offers an alternative to model-based analyses. It uses only the randomization process itself for making inferences by comparing the value of the observed test statistic to its collection of possible values from all permutations of the treatment assignment labels (Fisher, 1935; Rosenbaum, 2002). Randomization inference tests the null hypothesis of no primary treatment effect, meaning “the treatment confers no more benefit or harm to treated units than it confers to untreated controls” (Rosenbaum, 2007). In particular, the null hypothesis

includes the both the scenarios where the treatment has no effect on any cluster and where the treatment affects all clusters equally. Randomization inference has the added benefit of addressing the challenge of spillover effects. The data-generating process of the outcomes does not affect the distribution of the test statistic under the null hypothesis of no primary treatment effect, even if it includes interference between units (Rosenbaum, 2007).

Although CRTs may be strengthened by the use of constrained randomization, randomization inference, and adjustment for prognostic baseline variables (at both the individual and cluster level) (Li et al., 2015), this combination of all three tools has, to the best of our knowledge, not yet been applied to a community-based intervention. In this chapter, we present the design and analysis of a CRT developed to strengthen connections between community based organizations in East Baltimore and improve the health of the clients they serve. We aim to address the following question: what are the strengths and weaknesses of using a balanced, covariate-adjusted, cluster randomized trial design for evaluating effectiveness of a community engagement intervention for strengthening connections among hospitals and clinics, community based organizations, and community?

### **4.1.2 Baltimore CONNECT Project**

The Baltimore Community Organizations Neighborhood Network: Enhancing Capacity Together (CONNECT) Project was a community engagement and research partnership between the Johns Hopkins Health Systems (JHHS), community-based

organizations (CBOs) in East Baltimore, and residents of East Baltimore neighborhoods. The study, funded by the Patient Centered Outcomes Research Institute (PCORI), aimed to improve the health of East Baltimore residents by leveraging already available resources and social structures, and fostering a sustainable network of community and health care organizations. A community engagement approach was used to co-develop and apply a capacity building intervention among a group of CBO partners and the Baltimore CONNECT study team (Wu et al., 2018).

A systematic process was used for identifying all candidate community based organizations. First, the IRS Master file of Tax Exempt Organizations was consulted; second, all non-profit organizations that resided in or near the following J-CHiP (Berkowitz et al., 2016) zip codes: 21202, 21205, 21213, 21219, 21222, 21224, 21231 were identified; third, the National Taxonomy of Exempt Entities Classification System was used to further include only organizations that provided health and human services, and those that engaged in the direct delivery of services to adults. The organizations included were limited to those with 501(c) 3 status, and excluded social services run by city, state or federal government agencies. After creating a tentative list of CBOs that fit the minimum inclusion and exclusion criteria, a list of these organizations, along with their contact information, a description of the type of services provided, and mission statements was presented to the stakeholders at our first stakeholder meeting for further input. After refining the list, all eligible community organizations were invited to an information session to explain the details of the study. Those who expressed interest in the study were invited to participate; of those invited, twenty-two organizations accepted the invitation for

enrollment into the study.

The methods for constrained randomization to assign CBOs to the intervention or control study arms are delineated in Section 4.2. We refer to CBOs assigned to the intervention and control study arms as iCBOs and cCBOs, respectively. One of the iCBOs dropped out of the study prior to baseline data collection, and one of the cCBOs closed before follow-up. Our analyses were based on the twenty CBOs that remained in the study.

A community engagement intervention was developed and implemented among the intervention CBOs. Components of the intervention, detailed by Wu et al. (2018), included the following: monthly meetings among the intervention CBOs, an online toolkit of resources, a subscription to the search engine Healthify, a research assistant, and meet-and-greet sessions between staff members from the intervention CBOs and the Johns Hopkins Health System. The goal of the intervention was to improve the health of East Baltimore residents.

The effectiveness of the intervention was evaluated using administrative claims data made available through the John Hopkins Community Health Partnership (J-CHiP) (Berkowitz et al., 2016). Funded by the Center for Medicare and Medicaid Innovation, J-CHiP was a large initiative conducted at Johns Hopkins Hospital and Johns Hopkins Bayview Medical Center to improve the coordination and quality of care in Johns Hopkins Health Care inpatient and outpatient settings (Berkowitz et al., 2016). The J-CHiP population consisted of all Priority Partners (Medicaid) Managed Care Organization and Medicare patients who: 1) were identified for the J-CHiP program by either a risk prediction model or referral; 2) passed a

screening process; and 3) had monthly utilization data available, including number of emergency department (ED) visits and days spent in the hospital. Records were analyzed for 4917 patients.

Because of CBO client confidentiality, we were unable to determine which, if any, of the twenty CBOs in the study each J-CHiP patient utilized. We assumed that individuals seeking services from CBOs are more likely to utilize the services of CBOs that are geographically closer to where they live compared with CBOs that are farther away. We used Google API to determine the distance between the home address of each patient enrolled in J-CHiP and each CBO and determined the closest CBO to each patient; this is called the “proximal CBO” for each patient.

The primary analysis used a difference-of-differences approach. We compared the changes in the number of emergency department (ED) visits and days spent in the hospital from before and after the capacity building intervention; this control was done between J-CHiP patients proximal to the iCBOs and J-CHiP patients proximal to the cCBOs. Though it would have been preferable to directly measure the impact of the intervention on CBO clients, it was not feasible to do this.

Spillover was a major concern in the analysis of this study. If a CBO’s set of clients includes a J-CHiP patient geographically closer to another CBO, the treatment assignment of the former CBO may affect outcomes measured at the latter CBO. CBO staff and client surveys indicated a network of communication and referrals between the different organizations, which the intervention sought to foster (Wu et al., 2018). As a result, spillover was very plausible. We used randomization inference for the analysis because it is still valid in the presence of

spillover.

## 4.2 Methods

### 4.2.1 Constrained Randomization

As previously discussed, cluster randomized trials risk imbalance on important variables, motivating the use of constrained randomization (Moulton, 2004). We summarize the steps of a constrained cluster randomization schema based on variables of interest (Moulton, 2004), and we provide examples of each step from the Baltimore CONNECT project.

1. Determine cluster-level characteristics that may be prognostic of the outcome of interest. These may be characteristics of the clusters themselves or cluster-specific summary measures of individual characteristics (Raab and Butcher, 2001). We applied the following randomization criteria. For any zip-code serviced by more than one CBO, at least one CBO in that zip-code must be randomized to each of the control and intervention arms. For any service type provided by more than one CBO in the study, at least one CBO of that type must be randomized to each of the control and intervention arms.
2. Enumerate all the possible allocations of treatment and control to the clusters. Supposing there are  $K$  clusters included in the randomization, with  $K_T$  assigned to the treatment and  $K_C = K - K_T$  assigned to the control, there will be  $\binom{K}{K_T}$  total allocations. In the Baltimore CONNECT study, we randomized

11 of the 22 CBOs to the intervention arm and the other 11 to the control arm, for a total of 705,432 possible treatment assignments.

3. Select the allocations that fit the criteria determined in (1). Out of the 705,432 possible treatment allocations, 2636 obeyed both the zip code and service type constraints.
4. Construct a square matrix with columns and rows enumerated by the clusters, where each entry gives the number of times the associated pair is assigned the same study arm (treatment or control) from the list of allocations in (3).
5. Examine the matrix for pairs of clusters that are (nearly) always or (nearly) never in the same study arm; these are indicators of over-restriction in the randomization constraints. If the allocation list does not have these problems, continue to (6). If the list is unacceptable, relax (or tighten) criteria and return to (3). Our validity matrix had six zeroes (pairs of CBOs always assigned to different study arms) and one pair of CBOs always assigned to the same arm, which we considered acceptable.
6. Among the list of allocations in (3) deemed acceptable, randomly select one to proceed with.

An important decision in implementing constrained randomization is how many constraints to include. More constraints can lead to better balance, but can impact both the power of the trial and the validity of the randomization. Validity refers to the lack of linkage among clusters in the randomization scheme,

so that the more uniform the distribution of the elements of the matrix in step (4), the greater the validity (Bailey and Rowley, 1987). An invalid design can cause statistical inference to have incorrect Type I error and confidence interval coverage (Moulton, 2004). One can avoid an overly constrained randomization by checking, among the acceptable treatment assignments, for pairs of clusters that are always or never assigned to the same treatment arm. If such pairs are identified or the number of acceptable treatment assignments is very small relative to the number of possible unconstrained treatment assignments, it may be beneficial to relax the randomization criteria (Moulton, 2004; Hayes and Moulton, 2017).

#### 4.2.2 Randomization Inference

We used randomization inference to test the null hypothesis of no effect of the intervention on any individuals in the study population. We compared the estimated treatment effect under the actual randomization compared to all the other ways the randomization could have occurred (Murray, 1998; Small, Ten Have, and Rosenbaum, 2008). In our constrained randomization, there were 2636 allowable treatment allocations, each of which had a  $\frac{1}{2636}$  probability of being implemented. We recalculated the difference-in-differences test statistic for each of the 2635 other possible allocations of the treatment assignments that met the randomization criteria. The p-value is defined as the proportion of treatment assignments yielding a difference-in-differences statistic of magnitude greater or equal to the observed statistic.



The primary outcome for J-CHiP patients was the difference between the average monthly sum of ED visits and days spent in the hospital in the period after as compared to before the intervention. The test statistic was the difference-in-differences in the average number of monthly ED visits and hospital days from the post intervention period as compared to the pre-intervention period between iCBOs and cCBOs. We adjusted for the following baseline individual-level variables that may be prognostic of outcomes: baseline utilization, age, gender, and insurance type.

We present methods for a covariate adjusted, permutation-based, randomization inference analysis developed by Rosenbaum (2002) and Small, Ten Have, and Rosenbaum (2008). Suppose cluster  $k$  contains  $n_k$  individuals, for a total of  $N = \sum_{k=1}^K n_k$  individuals. Let  $Z_k$  be an indicator that cluster  $k$  is assigned to the intervention study arm. The algorithm for the analysis is as follows.

1. For each eligible patient, average the monthly sum of ED visits and days spent in the hospital in the pre-intervention period, from 9/1/12 to 2/28/14. For individual  $i$  in cluster  $k$ , call this value  $R_{1ki}$ .
2. For each eligible patient, average the monthly sum of ED visits and days spent in the hospital in the post-intervention period, from 4/1/14 to 9/30/15, and denote this value  $R_{2ki}$ .
3. Calculate for each patient a change score over the course of the study, from the average monthly sum of ED visits and hospital days in the pre-intervention period to the post-intervention period. The change score for individual  $i$  in

cluster  $k$  is  $R_{ki} := R_{2ki} - R_{1ki}$ .

4. Regress the change score outcomes from all study participants on a set of individual characteristics observed prior to randomization  $\mathbf{X}$  (baseline utilization, age, gender, and insurance type), using a linear model. Calculate the residuals from the regression, and for individual  $i$  in cluster  $k$ , denote their adjusted response (residual)  $e_{0ki}$ .

5. Calculate an average adjusted change score among patients proximal to  $\text{CBO}_k$  :

$$e_{0k} = \frac{1}{n_k} \sum_{i=1}^{n_k} e_{0ki}.$$

6. Determine the difference between the weighted average of the adjusted change scores among the intervention CBOs and the control CBOs. For both the control and intervention groups, weights should be proportional to the number of J-CHiP patients assigned to each CBO,  $n_k$ . The weighted averages for the control and intervention (treatment) arms are:

$$e_C = \frac{1}{\sum_{k=1}^K (1 - Z_k) n_k} \sum_{\{k: Z_k=0\}} n_k e_{0k};$$

$$e_T = \frac{1}{\sum_{k=1}^K Z_k n_k} \sum_{\{k: Z_k=1\}} n_k e_{0k}.$$

The difference-in-differences statistic,  $\hat{S}$ , is defined as:

$$\hat{S} := e_T - e_C$$

7. Recompute the test statistic in (6) under all of the 2635 other possible allocations of the treatment assignments. Call these statistics  $\hat{S}^*$ .
8. Compare the distribution of the permuted  $\hat{S}^*$  in (7) to the observed  $\hat{S}$  in (6) from the true treatment assignments. The proportion of the  $\hat{S}^*$  with values as or more extreme (higher in magnitude) than that of  $\hat{S}$  is the p-value. The null hypothesis of no treatment effect is rejected at level  $\alpha = 0.05$  if the p-value is less than  $\alpha$ .

For the *unadjusted* analysis, omit Step 4 and proceed in the remaining steps using the  $R_{ki}$  in place of the  $e_{0ki}$ . In Section 4.7, we describe how to construct confidence intervals using randomization inference. Unlike the hypothesis testing, the confidence interval construction requires the assumptions of constant treatment effect and no spillover.

### 4.2.3 Notation and Justification for Randomization Inference

Randomization inference employs a potential outcomes framework (Rosenbaum, 2002; Small, Ten Have, and Rosenbaum, 2008). The following justification for using randomization inference in the presence of potential spillover of the treatment impact uses key ideas from Rosenbaum (2007). We define a potential outcome (change in ED visits plus days in the hospital)  $r_{ki\pi}$  for each individual  $i$  with

proximal cluster  $k$  under each treatment assignment  $\pi$  to all clusters (i.e.,  $\pi$  is one of the 2636 allowed treatment permutations to all clusters). The observed outcome  $R_{ki}$  is the potential outcome  $r_{ki\Pi}$  where  $\Pi$  is the selected treatment assignment (to all clusters).

The null hypothesis of no treatment effect is that  $r_{ki\pi_1} = r_{ki\pi_2}$  for any assignments (permutations)  $\pi_1, \pi_2$ . That is, the null hypothesis is that there is no impact on any individual of assigning any set of clusters to the intervention versus the control. This null hypothesis implies no effect on a participant of assigning her/his cluster to treatment versus control; it also implies more: that there is no impact from assigning any set of clusters (among those allowed by the constrained randomization) to the intervention versus control.

We based our inference upon the difference-in-differences statistic, which is a function of the observed responses. Under the null hypothesis of no treatment effect, the test statistic constructed from any possible permutation of the study assignments will be equally likely because every individual will have the same potential outcome regardless of their cluster's assignment to treatment or intervention.

We adjusted for baseline variables to try to remove variability in the outcome explained by the baseline variables as follows. Denote by  $\mathbf{R}$  the  $N$ -dimensional vector of the observed individual responses and by  $\mathbf{X}$  an  $N \times p$  matrix of  $p$  baseline variables measured on each individual before the intervention. Calculate the difference (i.e., the residual, denoted by  $e_{0ki}$ ) between the observed outcomes  $\mathbf{R}$  and the projection of the outcomes  $\mathbf{R}$  onto the space spanned by the covariates  $\mathbf{X}$ .

This residual is a function of the observed data (the outcomes  $\mathbf{R}$  and covariates  $\mathbf{X}$ ), and its calculation makes no assumptions about the data generating process itself (Small, Ten Have, and Rosenbaum, 2008). Similarly, the adjusted difference-in-differences statistic is a function of the observed data, and only the study arm assignments are random. As a result, we can characterize the distribution of the adjusted difference-in-differences statistic under the null hypothesis of no treatment effect without any distributional assumptions about how outcomes were generated.

## 4.3 Results

Overall health care utilization stayed constant from the pre- to post- intervention periods. Both the monthly number of ED visits and days in the hospital had constant means of 0.19 and 0.39, respectively, and medians of zero.

The results of the difference-in-differences analyses are summarized in Table 4.1. The unadjusted and adjusted difference-in-differences statistics are very close to zero, and we cannot reject the null hypothesis that the intervention had no effect on the before-after change in total ED visits and hospital days for J-CHiP patients in the study. The width of confidence interval for the difference-in-differences statistic is reduced by 15% adjusting for baseline variables.

Because the confidence intervals were determined using randomization inference, as detailed in Section 3.3, their interpretation assumes a constant treatment

effect. For example, in the second row of Table 4.1, we may conclude 95% confidence that if the treatment effect is constant across individuals, after adjusting for baseline variable it will be between -0.115 and 0.108.

In contrast, confidence intervals determined through model-based analyses do not require the assumption of constant treatment effects but instead employ other strong assumptions. The next section compares our analyses with a model-based approach.

## **4.4 Comparison with Model-Based Analysis**

### **4.4.1 Methods**

Model-based analyses are typically conducted in one or two stages. In a one-stage approach, outcomes are modeled on the individual level. Within cluster correlations are typically accounted for by using mixed effects models or generalized estimating equations (Hayes and Moulton, 2017). Additionally, terms may be included in the model to adjust for pre-intervention variables (Hayes and Moulton, 2017).

The two-stage approach is as follows. In the first stage, a summary statistic is obtained for each cluster based on individual-level data; in the second stage, the summaries are compared, for example with a two-sample t-test. Pre-randomization individual and/or cluster-level variables can be adjusted for by regressing individual outcomes on them; the individual-level residuals, rather than outcomes, are then used to calculate cluster-level summaries (Hayes and Moulton, 2017).

One-stage analyses are typically more efficient. However, Hayes and Moulton (2017) recommend performing two-stage analyses when there are 15-20 or fewer clusters in each study arm because they typically are more robust to model assumptions than one-stage analyses are when the number of clusters is small. Because our study had only 10 cCBOs and iCBOs each, we compared our results from the previous section to a two-stage, model-based analysis.

In the first stage of the model-based analysis, we determined a summary measure for each CBO based upon individual-level data, following steps 1-5 of the analysis detailed in Section 4.2.2 to generate an adjusted change score  $e_{0k}$  for each of the CBOs. In the second stage, we compared the adjusted change scores between the cCBOs and iCBOs using an unpaired t-test. Following the recommendation in (Hayes and Moulton, 2017), we did not incorporate cluster-level weights. The average difference-in-differences for patients assigned to the cCBOs and iCBOs,  $\bar{e}_C$  and  $\bar{e}_T$ , respectively, are:

$$\bar{e}_C = \frac{\sum_{k=1}^K (1 - Z_k) e_{0k}}{\sum_{k=1}^K (1 - Z_k)};$$

$$\bar{e}_T = \frac{\sum_{k=1}^K Z_k e_{0k}}{\sum_{k=1}^K Z_k}.$$

The test statistic is defined as:

$$t = \frac{\bar{e}_T - \bar{e}_C}{s \sqrt{\frac{1}{K/2} + \frac{1}{K/2}}},$$

where

$$s^2 := \frac{\sum_{k:Z_k=1} (e_{0k} - \bar{e}_T)^2 + \sum_{k:Z_k=0} (e_{0k} - \bar{e}_C)^2}{K - 2}.$$

In this model-based analysis, we assumed that the cluster-level outcomes (average differences in patient utilization from baseline to follow-up) are normally distributed and that the variance of the cluster-level outcomes is constant. However, these two assumptions are difficult to assess in a study with only twenty clusters in total.

#### 4.4.2 Results

The unadjusted and adjusted difference-in-differences statistics from the two-stage, model-based analysis are given in Table 4.2. The difference-in-differences statistics (unadjusted and adjusted) varied slightly between the permutation and model-based analyses because in the latter the clusters are unweighted. However, the conclusion of no treatment effect remained the same.

### 4.5 Discussion

We presented an application of a constrained cluster randomized design in health services research and detailed the process of study design and analysis. This approach may be of use for future community-based participatory research. We



illustrated that adjusting for baseline variables in the difference-in-differences analysis using methods developed by Rosenbaum (2002) can substantially improve efficiency in a randomization inference analysis. No model assumptions are needed in the randomization inference approach, meaning the test is valid regardless of whether the linear regression model used to adjust for baseline variables was correctly specified or not. A key part of the intervention in the study was the creation of a network between the CBOs, so we expected spillover effects, and benefited from the fact that when “randomization forms the basis for inference,” such as a permutation test of a cluster randomized trial, the level of significance will be valid regardless of whether there is spillover between units in the intervention and control arms (Fisher, 1935; Rosenbaum, 2007).

Our analysis of the Baltimore CONNECT study had several limitations. The confidence intervals we constructed for the difference-in-differences statistics using randomization inference (see Section 4.7 for details) are valid only if the effect of the intervention is constant, meaning the difference in change scores if an individual were assigned to an iCBO as compared to a cCBO is the same among all study participants. Further, we were only able to test the sharp null hypothesis that the intervention did not have an effect on any of the units. We could not determine whether there was a population level average benefit, or, for example, how many days in the hospital and ED visits could be saved through the intervention on average (Small, Ten Have, and Rosenbaum, 2008; Rosenbaum, 2007).

Finally, it was observed that the iCBOs did not all participate in the intervention to the same extent. One possible extension may be adopting methods developed by

Rosenbaum (1996) for exact randomization inference using instrumental variables, which has since been employed for a range of applications (Greevy et al., 2004; Rosenbaum, 2002; Imbens and Rosenbaum, 2005).

## 4.6 Acknowledgements

This work was supported by the Patient-Centered Outcomes Research Institute (CD-12-11-4948). This work is solely the responsibility of the authors and does not represent the views of PCORI.

	<b>Difference-in-Differences Statistic</b>	<b>p-value</b>	<b>Confidence Interval</b>
Unadjusted utilization	0.005	0.926	(-0.130, 0.127)
Adjusted utilization	0.018	0.627	(-0.115, 0.108)

**Table 4.1: Results from the difference-in-differences analyses using randomization inference.** Utilization is defined as a patient’s monthly number of emergency department visits and days spent hospitalized. In both the unadjusted and adjusted analyses, we fail to reject the null hypothesis of no treatment effect.

	<b>Difference-in-Differences Statistic</b>	<b>p-value</b>	<b>Confidence Interval</b>
Unadjusted utilization	0.002	0.982	(-0.144, 0.141)
Adjusted utilization	0.037	0.652	(-0.205, 0.132)

**Table 4.2: Results from the difference-in-differences analyses using a two-stage model.**

## 4.7 Appendix: Confidence Interval Construction

We present the methods for confidence interval construction under randomization inference (a continuation of Section 4.2.2). We note that unlike the hypothesis tests, these confidence intervals incorporate assumptions of both a constant treatment effect and no spillover.

First, construct a function that adjusts the change score for baseline characteristics by extracting the residuals from a linear regression of the outcomes  $\mathbf{R}$  on the individual-level baseline variables  $\mathbf{X}$  :

$$g : (X_{ki}, R_{ki}) \mapsto e_{0ki}.$$

The function  $g$  calculates the difference (i.e., the residual, denoted by  $e_{0ki}$ ) between the observed outcomes  $\mathbf{R}$  and the projection of the outcomes  $\mathbf{R}$  onto the space spanned by the covariates  $\mathbf{X}$ . It is a function of the observed data that makes no assumptions about the data generating process itself (Small, Ten Have, and Rosenbaum, 2008). After following steps 1 through 8 of Section 4.2.2, the endpoints for a 95% confidence interval can be calculated using the following method detailed in Rosenbaum (2002) and Small, Ten Have, and Rosenbaum (2008) and extended to handle arbitrary interference as in Aronow and Samii, 2017. Under the assumption of a constant treatment effect  $\tau$ , for all treatment assignments  $\pi_1, \pi_2$  such that cluster  $k$  is assigned to treatment under  $\pi_1$  but to control under  $\pi_2$ :

$$r_{ki\pi_1} = r_{ki\pi_2} + \tau.$$

We define  $r_{Cki}$  ( $r_{Tki}$ ) to be the potential outcome for participant  $i$  under any assignment that sets cluster  $k$  to control (treatment).

To determine a confidence interval for the treatment effect  $\tau$ , construct hypothesis tests for

$$H_0 : \tau = \tau_0$$

over a grid of potential values  $\tau_0$ . For each value of  $\tau_0$ , recompute the difference-in-differences statistic under the null hypothesis  $H_0 : \tau = \tau_0$ .

- a. For each cluster assigned to the control arm,  $\{k : Z_k = 0\}$ , the  $n_k$  dimensional vector of potential responses (change scores from baseline to followup)  $\mathbf{r}_{Ck}^{\tau_0}$  under the control is simply their vector of observed changes  $\mathbf{R}_k$ .
- b. For clusters assigned to the intervention arm,  $\{k' : Z_{k'} = 1\}$ , the  $n_{k'}$  dimensional vector of potential responses (change scores from baseline to followup)  $\mathbf{r}_{Ck'}^{\tau_0}$  under the control is

$$\mathbf{r}_{Ck'}^{\tau_0} = \mathbf{R}_{k'} - \tau \mathbf{1},$$

where  $\mathbf{1}$  is an  $n_{k'}$  dimensional vector of 1's.

- c. Letting  $\mathbf{r}_C^{\tau_0}$  be the vector of potential outcomes for all patients, calculate the residuals after adjusting for baseline characteristics using the model described in (4), and denote the residuals by:

$$\mathbf{e}^{\tau_0} := g(\mathbf{X}, \mathbf{r}_C^{\tau_0}).$$

- d. Repeat steps (5)-(8) using  $\mathbf{e}^{\tau_0}$  in place of the observed vector of patient specific residuals  $\mathbf{e}^0$  to determine a difference-in-differences statistic  $\hat{S}^{\tau_0}$  and associated p-value.
- e. The set of values  $\tau_0$  under which one would fail to reject the null hypothesis  $H_0 : \tau = \tau_0$  in a two-sided 0.05-level test forms the 95% confidence interval for the treatment effect  $\tau$  under a hypothesis of a constant treatment effect  $\tau_0$  for every individual.

For the *unadjusted* analysis, omit Step c. and proceed using the  $\mathbf{r}_C^{\tau_0}$  in place of the  $\mathbf{e}^{\tau_0}$ .

## References

- Béhar, Moisés (1975). "The role of feeding and nutrition in the pathogeny and prevention of diarrheic processes". In: *Bull Pan Am Health Organ* 9.1, pp. 1–9.
- Victora, Cesar G, J Patrick Vaughan, Cintia Lombardi, Sandra M C Fuchs, Luciana P Gigante, Peter G Smith, Leticia C Nobre, AnaMaria B Teixeira, Leila B Moreira, and Fernando C Barros (1987). "Evidence for protection by breast-feeding against infant deaths from infectious diseases in Brazil". In: *The Lancet* 330.8554, pp. 319–322.
- Dewey, Kathryn G, M Jane Heinig, and Laurie A Nommsen-Rivers (1995). "Differences in morbidity between breast-fed and formula-fed infants". In: *The Journal of Pediatrics* 126.5, pp. 696–702.
- Arifeen, Shams, Robert E Black, Gretchen Antelman, Abdullah Baqui, Laura Caulfield, and Stan Becker (2001). "Exclusive breastfeeding reduces acute respiratory infection and diarrhea deaths among infants in Dhaka slums". In: *Pediatrics* 108.4, e67–e67.
- Lamberti, Laura M, Christa L Fischer Walker, Adi Noiman, Cesar Victora, and Robert E Black (2011). "Breastfeeding and the risk for diarrhea morbidity and mortality". In: *BMC public health* 11.3, p. 1.
- Sauls, Henry S (1979). "Potential effect of demographic and other variables in studies comparing morbidity of breast-fed and bottle-fed infants". In: *Pediatrics* 64.4, pp. 523–527.
- Feachem, Richard G and Marge A Koblinsky (1984). "Interventions for the control of diarrhoeal diseases among young children: promotion of breast-feeding." In: *Bulletin of the World Health Organization* 62.2, p. 271.
- Heintze, Carl, M Velasco Garrido, and Axel Kroeger (2007). "What do community-based dengue control programmes achieve? A systematic review of published evaluations". In: *Transactions of the Royal Society of Tropical Medicine and Hygiene* 101.4, pp. 317–325.

- Hayes, Richard J. and Lawrence H. Moulton (2017). *Cluster randomized trials*. 2nd ed. Chapman and Hall/CRC Biostatistics.
- Donner, Allan and Neil Klar (2000). *Design and analysis of cluster randomization trials in health research*.
- Hayes, RJ and S Bennett (1999). "Simple sample size calculation for cluster-randomized trials." In: *International journal of epidemiology* 28.2, pp. 319–326.
- Donner, Allan, Nicholas Birkett, and Carol Buck (1981). "Randomization by cluster sample size requirements and analysis". In: *American Journal of Epidemiology* 114.6, pp. 906–914.
- Raudenbush, Stephen W (1997). "Statistical analysis and optimal design for cluster randomized trials." In: *Psychological Methods* 2.2, p. 173.
- Donner, Allan (1998). "Some aspects of the design and analysis of cluster randomization trials". In: *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 47.1, pp. 95–113.
- Sommer, Alfred, Edi Djunaedi, AA Loeden, Ignatius Tarwotjo, Keith P West, Robert Tilden, Lisa Mele, Aceh Study Group, et al. (1986). "Impact of vitamin A supplementation on childhood mortality: a randomised controlled community trial". In: *The Lancet* 327.8491, pp. 1169–1173.
- Fisher Jr, Edwin B (1995). "The results of the COMMIT trial. Community Intervention Trial for Smoking Cessation." In: *American Journal of Public Health* 85.2, pp. 159–160.
- Smith, Peter G, Richard H Morrow, and David A Ross (2015). *Field Trials of Health Interventions*. Oxford University Press.
- Fairhurst, Karen and Christopher Dowrick (1996). "Problems with recruitment in a randomized controlled trial of counselling in general practice: causes and implications". In: *Journal of Health Services Research* 1.2, pp. 77–80.
- Hussey, Michael A and James P Hughes (2007). "Design and analysis of stepped wedge cluster randomized trials". In: *Contemporary Clinical Trials* 28.2, pp. 182–191.
- Taljaard, Monica, Andrew D McRae, Charles Weijer, Carol Bennett, Stephanie Dixon, Julia Taleban, Zoe Skea, Martin P Eccles, Jamie C Brehaut, Allan Donner, et al. (2011). "Inadequate reporting of research ethics review and informed consent in cluster randomised trials: review of random sample of published trials". In: *BMJ* 342, p. d2496.
- Sim, Julius and Angus Dawson (2012). "Informed consent and cluster-randomized trials". In: *American Journal of Public Health* 102.3, pp. 480–485.



- Weijer, Charles, Jeremy M Grimshaw, Martin P Eccles, Andrew D McRae, Angela White, Jamie C Brehaut, and Monica Taljaard (2012). "The Ottawa statement on the ethical design and conduct of cluster randomized trials". In: *PLoS Medicine* 9.11, e1001346.
- Klar, Neil and Allan Donner (2001). "Current and future challenges in the design and analysis of cluster randomization trials". In: *Statistics in Medicine* 20.24, pp. 3729–3740.
- Donner, Allan and Neil Klar (2004). "Pitfalls of and controversies in cluster randomization trials". In: *American Journal of Public Health* 94.3, pp. 416–422.
- Moulton, Lawrence H (2004). "Covariate-based constrained randomization of group-randomized trials". In: *Clinical Trials* 1.3, pp. 297–305.
- Gail, Mitchell H, Steven D Mark, Raymond J Carroll, Sylvan B Green, and David Pee (1996). "On design considerations and randomization-based inference for community intervention trials". In: *Statistics in Medicine* 15.11, pp. 1069–1092.
- Braun, Thomas M and Ziding Feng (2001). "Optimal permutation tests for the analysis of group randomized trials". In: *Journal of the American Statistical Association* 96.456, pp. 1424–1432.
- Small, Dylan S, Thomas R Ten Have, and Paul R Rosenbaum (2008). "Randomization inference in a group-randomized trial of treatments for depression: covariate adjustment, noncompliance, and quantile effects". In: *Journal of the American Statistical Association* 103.481, pp. 271–279.
- Sobel, Michael E (2006). "What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference". In: *Journal of the American Statistical Association* 101.476, pp. 1398–1407.
- Tchetgen, Eric J Tchetgen and Tyler J VanderWeele (2012). "On causal inference in the presence of interference". In: *Statistical Methods in Medical Research* 21.1, pp. 55–75.
- Aronow, Peter M and Cyrus Samii (2013). "Estimating average causal effects under interference between units". In: *arXiv preprint: 1305.6156*.
- VanderWeele, Tyler J, Eric J Tchetgen Tchetgen, and M Elizabeth Halloran (2014). "Interference and sensitivity analysis". In: *Statistical science: a Review Journal of the Institute of Mathematical Statistics* 29.4, p. 687.
- Fisher, Ronald A (1935). "The logic of inductive inference". In: *Journal of the Royal Statistical Society* 98.1, pp. 39–82.
- Rosenbaum, Paul R et al. (2002). "Covariance adjustment in randomized experiments and observational studies". In: *Statistical Science* 17.3, pp. 286–327.

- Rosenbaum, Paul R (2007). "Interference between units in randomized experiments". In: *Journal of the American Statistical Association* 102.477, pp. 191–200.
- Li, Fan, Yuliya Lokhnygina, David M Murray, Patrick J Heagerty, and Elizabeth R DeLong (2015). "An evaluation of constrained randomization for the design and analysis of group-randomized trials". In: *Statistics in Medicine*.
- Wu, Albert W, Sandra Hwang, Christine M Weston, Chidinma Ibe, Romsai Boonyasai, Lee Bone, Lopa Basu, Isaac Lief, Janice Gentry, Leon Purnell, and for the Baltimore CONNECT study team. (In Press) Rosenblum Michael (2018). "Baltimore CONNECT: a randomized trial to build partnership between community organizations and a local health system ". In: *Progress in Community Health Partnerships: Research, Education, and Action*.
- Berkowitz, Scott A, Patricia Brown, Daniel J Brotman, Amy Deutschendorf, Linda Dunbar, Anita Everett, Debra Hickman, Eric Howell, Leon Purnell, Carol Sylvester, et al. (2016). "Case Study: Johns Hopkins Community Health Partnership: A model for transformation". In: *Healthcare*. Vol. 4. 4. Elsevier, pp. 264–270.
- Raab, Gillian M and Izzy Butcher (2001). "Balance in cluster randomized trials". In: *Statistics in Medicine* 20.3, pp. 351–365.
- Bailey, RA and CA Rowley (1987). "Valid randomization". In: *Proc. R. Soc. Lond. A* 410.1838, pp. 105–124.
- Murray, David M (1998). *Design and analysis of group-randomized trials*. Vol. 29. Oxford University Press, USA.
- Rosenbaum, Paul R (1996). "Identification of causal effects using instrumental variables: comment". In: *Journal of the American Statistical Association* 91.434, pp. 456–458.
- Greevy, Robert, Jeffrey H Silber, Avital Cnaan, and Paul R Rosenbaum (2004). "Randomization inference with imperfect compliance in the ACE-inhibitor after anthracycline randomized trial". In: *Journal of the American Statistical Association* 99.465, pp. 7–15.
- Imbens, Guido W and Paul R Rosenbaum (2005). "Robust, accurate confidence intervals with a weak instrument: quarter of birth and education". In: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 168.1, pp. 109–126.
- Aronow, Peter M, Cyrus Samii, et al. (2017). "Estimating average causal effects under general interference, with application to a social network experiment". In: *The Annals of Applied Statistics* 11.4, pp. 1912–1947.

# Chapter 5

## Conclusion

We have presented the analysis of two types of randomized control trials: a collection of clinical trials assessing the use of pre-exposure prophylaxis for HIV prevention and a cluster randomized trial evaluating the use of a community engagement intervention designed to improve health outcomes. Our analyses addressed a range of statistical challenges, including sparse data and rare outcomes, certain data available for only a subset of study participants through nested substudies, low (and at times uncertain) levels of adherence, and small numbers of and interference between study units. Although the designs and substantive areas of the trials differ, the different applications offer a number of parallels.

A substantial challenge in both projects is uncertainty in the level of exposure each study participant received to their assigned treatment. In the PrEP trials, participant adherence to was highly varied and was estimated using unreliable measures (Abaasa et al., [2017](#)). We used participant concentration levels estimated from a pharmacokinetic model to better account for variability in dosing than raw

concentration measurements could. An important extension of the PrEP project will be to quantify how the uncertainty in the average concentrations impacts our estimates of the protective effect of different levels of PrEP concentration. In Chapter 3, we discussed a potential Bayesian framework for propagating the prediction uncertainty from PK modeling to the second stage modeling of HIV risk.

In the Baltimore CONNECT study, we did not know if the patients from whom we had health care utilization data actually visited their local community based organization. To address this challenge, we inferred the causal effect of assigning one's geographically closest CBO to the intervention on hospital utilization among Medicare and Medicaid recipients in Baltimore. It would be preferable, however, to analyze the effect of the intervention on hospital patients known to be served by the CBOs. Developing a such a study design while maintaining both client and patient confidentiality remains an open challenge.

A natural extension of both projects is to consider future trial design. As mentioned in Chapter 1, there are a number of ongoing PrEP clinical trials. The continued development of different PrEP formulations and importance of determining effective dosing strategies for a wide range of at-risk populations motivates future trial development. Cutrell et al. (2017) discuss a number of considerations and propose innovations for future trial design, in particular for non-inferiority and superiority studies that compare new PrEP formulations to oral TDF/FTC. PrEP implementation and delivery will also guide future trial design (Norton, Larson, and Dearing, 2013; Baeten et al., 2013; Krakower et al., 2014; Mayer, Krakower,

and Boswell, 2016). Coordination of resources and care between community and clinical settings, as was the goal of Baltimore CONNECT, will likely feature heavily in this work (Norton, Larson, and Dearing, 2013). Lessons from the Baltimore CONNECT study may be beneficial in considering future trials that assess the utility of leveraging existing social structures to implement PrEP for reducing HIV infection.

Additionally, the Baltimore CONNECT partnership has since expanded to form a network of nonprofits as a 501(c)(3) organization. Evaluation of the sustainability of this network and its broader impact will motivate future study design considerations.

Although this dissertation centers around statistical challenges methods in the analysis of randomized control trials, in both of the applications we presented, the complex social structures, needs, and interactions within these trials affect our analyses. Integrating qualitative research through mixed methods for developing and supplementing analyses is integral. In PrEP clinical trials, a number of factors, such as trust in one's partner, power dynamics in relationships, and social stigma, affected both levels and accurate reporting of adherence (Ware et al., 2012; Van der Straten et al., 2014; Straten et al., 2016). These factors were largely discerned from interviews and discussions with study participants. Carballo-Diéguez et al. (2017) discuss the effectiveness of a mixed methods approach to determining study participant adherence. The following point made by Van der Straten et al. (2014) in their analysis of a qualitative, supplementary study to the VOICE trial is critical:

Clinical trials are more than biomedical enterprises to test new drugs:

they are social phenomena that create new social relations within the household, the clinical trial setting, the local community, and translocally with donor organizations and research agencies. These social relations will shape and reshape local knowledge and the meaning of participation in clinical trials and of testing experimental drugs. Additionally, drugs are not mere active pharmaceutical ingredients, they are social innovations that require commensurability within the lives of their adopters and their social network [51,52], whether the adopters are clinical trial participants or real world users. (Van der Straten et al., 2014)

A similar point is expressed by Ibe et al. (2018) in discussing the framework to the Baltimore CONNECT study that Ibe et al. (2018)

...the structure of the trial itself stimulated knowledge brokerage at multiple levels. This positioned the study's stakeholders to emerge as community knowledge brokers and placed them squarely on the pathway of global innovation flow. (Ibe et al., 2018)

These studies not only evaluate the impact of exposures and interventions but also directly impact the populations they study. The analysis of randomized control trials exists within a broad public health framework. To be impactful, our statistical analyses should not be conducted in isolation, but rather consider the complicated social structures and interactions in play.

## References

- Abaasa, Andrew, Craig Hendrix, Monica Gandhi, Peter Anderson, Anatoli Kamali, Freddie Kibengo, Eduard J Sanders, Gaudensia Mutua, Namandjé N Bumpus, Frances Priddy, et al. (2017). "Utility of Different Adherence Measures for PrEP: Patterns and Incremental Value". In: *AIDS and Behavior*, pp. 1–9.
- Cutrell, Amy, Deborah Donnell, David T Dunn, David V Glidden, Anneke Grobler, Brett Hanscom, Britt S Stancil, R Daniel Meyer, Ronnie Wang, and Robert L Cuffe (2017). "HIV prevention trial design in an era of effective pre-exposure prophylaxis". In: *HIV Clinical Trials* 18.5-6, pp. 177–188.
- Norton, Wynne E, R Sam Larson, and James W Dearing (2013). "Primary care and public health partnerships for implementing pre-exposure prophylaxis". In: *American Journal of Preventive Medicine* 44.1, S77–S79.
- Baeten, Jared M, Jessica E Haberer, Albert Y Liu, and Nirupama Sista (2013). "Pre-exposure prophylaxis for HIV prevention: where have we been and where are we going?" In: *Journal of Acquired Immune Deficiency Syndromes (1999)* 63.0 2, S122.
- Krakower, Douglas, Norma Ware, Jennifer A Mitty, Kevin Maloney, and Kenneth H Mayer (2014). "HIV providers' perceived barriers and facilitators to implementing pre-exposure prophylaxis in care settings: a qualitative study". In: *AIDS and Behavior* 18.9, pp. 1712–1721.
- Mayer, Kenneth H, Douglas S Krakower, and Stephen L Boswell (2016). "Antiretroviral preexposure prophylaxis: opportunities and challenges for primary care physicians". In: *JAMA* 315.9, pp. 867–868.
- Ware, Norma C, Monique A Wyatt, Jessica E Haberer, Jared M Baeten, Alexander Kintu, Christina Psaros, Steven Safren, Elioda Tumwesigye, Connie L Celum,

- and David R Bangsberg (2012). "What's love got to do with it? Explaining adherence to oral antiretroviral pre-exposure prophylaxis (PrEP) for HIV serodiscordant couples". In: *Journal of Acquired Immune Deficiency Syndromes* (1999) 59.5.
- Van der Straten, Ariane, Jonathan Stadler, Elizabeth Montgomery, Miriam Hartmann, Busiswe Magazi, Florence Mathebula, Katie Schwartz, Nicole Laborde, and Lydia Soto-Torres (2014). "Women's experiences with oral and vaginal pre-exposure prophylaxis: the VOICE-C qualitative study in Johannesburg, South Africa". In: *PloS One* 9.2, e89118.
- Straten, Ariane, Elizabeth R Brown, Jeanne M Marrazzo, Michael Z Chirenje, Karen Liu, Kailazarid Gomez, Mark A Marzinke, Jeanna M Piper, and Craig W Hendrix (2016). "Divergent adherence estimates with pharmacokinetic and behavioural measures in the MTN-003 (VOICE) study". In: *Journal of the International AIDS Society* 19.1.
- Carballo-Diéguez, Alex, Ivan C Balán, William Brown III, Rebecca Giguere, Curtis Dolezal, Cheng-Shiun Leu, Mark A Marzinke, Craig W Hendrix, Jeanna M Piper, Barbra A Richardson, et al. (2017). "High levels of adherence to a rectal microbicide gel and to oral Pre-Exposure Prophylaxis (PrEP) achieved in MTN-017 among men who have sex with men (MSM) and transgender women". In: *PLoS One* 12.7, e0181607.
- Ibe, Chidinma A, Lopa Basu, Rachel Gooden, Shamsuzzoha B Syed, Viva Dadwal, Lee R Bone, Patti L Ephraim, Christine M Weston, Albert W Wu, J Barron, et al. (2018). "From Kisiizi to Baltimore: cultivating knowledge brokers to support global innovation for community engagement in healthcare". In: *Globalization and Health* 14.1, p. 19.



## **Curriculum Vitae**

Date of Birth: December 16, 1990

Place of Birth: Boston, MA

# Claire Ruberman

Department of Biostatistics  
Johns Hopkins School of Public Health  
615 N. Wolfe Street, E3038  
Baltimore, MD 21205

Phone: (617) 512-7965  
Email: [claireruberman@gmail.com](mailto:claireruberman@gmail.com)

## Education

- 2013- Present    **PhD Candidate in Biostatistics** (Expected to graduate August 2018)  
Johns Hopkins Bloomberg School of Public Health, Baltimore, MD  
Thesis advisor: Dr. Michael Rosenblum  
Dissertation title: "Statistical Methods in the Analysis of Randomized Controlled Trials"
- 2009-2013        **B.A. in Mathematics**, Magna Cum Laude  
Pomona College, Claremont, CA  
Thesis advisor: Dr. Ghassan Sarkis  
Thesis title: "Nowhere-Zero Flows in the Linear Lattice"

## Research Experience

- 2014- Present    Graduate Research Assistant  
Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health  
Supervisors: Drs. Craig Hendrix, Jeffrey Leek, Michael Rosenblum, Albert Wu
- June-Aug 2012   Summer Research Student  
Department of Chemical Engineering, Massachusetts Institute of Technology  
Supervisor: Dr. Arup Chakraborty
- June-Aug 2011   Summer Research Student  
The Ragon Institute of Harvard, MIT, and Massachusetts General Hospital  
Supervisor: Dr. Galit Alter
- June-Aug 2010   Research Assistant  
Department of Biology and Neuroscience, Brandeis University  
Supervisor: Dr. Sacha Nelson

## Papers

### *Peer-Reviewed Publications*

Manimaran S., Selby H.M., Okrah K., **Ruberman C.**, Leek J.T., Quackenbush J., Johnson, W.E. (2016). BatchQC: interactive software for evaluating sample and batch effects in genomic data. *Bioinformatics*, btw538.

Shekhar K., **Ruberman C.F.**, Ferguson A.L., Barton J.P., Kardar M., Chakraborty A.K. (2013). Spin models inferred from patient-derived viral sequence data faithfully describe HIV fitness landscapes. *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics* 88, no. 6: 062705.

Sarkis G., Shahriari S., Breese D., Fish B., Juarez I., Kothari U., Ragain S., **Ruberman C.**, Sarkar A., Siegel Z. (2015). Zero-sum Flows of the Linear Lattice. *Finite Fields and their Application*, 31, pp. 108-120.

Sarkis G., Shahriari S., Barnett Z., Breese D., Fish B., Frick W., Khatutsky A., McGuinness D., Rodrigues D., **Ruberman C.** (2013). Diamond Free Subsets in the Linear Lattices. *Order*, pp. 1-13.

### *Submitted Manuscripts*

Parsana P.\*, **Ruberman C.\***, Jaffe A.E., Schatz M.C., Battle A., Leek J.T. Addressing confounding artifacts in reconstruction of gene co-expression networks. (\* Equal Contribution)

Wu A.W., Weston C.M., Purnell L., Ibe C., Bone L., Boonyasai R., Heughan J.A., Hwang S., **Ruberman C.**, Rosenblum M., Gentry J., Lu Y. The Baltimore CONNECT Study: Community Organization Neighborhood Network: Enhancing Capacity Together

### *Manuscripts in Preparation*

**Ruberman C.**, Steingrimsson J.A., Hendrix C., Rosenblum M. Estimating the Protective Effect of Longitudinal Drug Concentration in Pre-Exposure Prophylaxis for HIV Prevention.

**Ruberman C.**, Rosenblum M., Wu A.W., Lu Y., Liang S., Ibe C., Weston C.M., Moulton L., Kachur S. Design and Analysis of a Cluster Randomized Trial with Spillover Effects for Health Services Research: The Baltimore CONNECT Project.

## Presentations

"Design and Analysis of a Cluster Randomized Trial with Spillover Effects for Health Services Research: the Baltimore CONNECT Project." Contributed session on pragmatic cluster randomized trials. International Conference on Health Policy Statistics. January 11, 2018, Charleston, SC.

"Estimating the Protective Effect of Longitudinal Drug Concentration in Pre-exposure Prophylaxis for HIV Prevention." Poster presentation. HIV Research for Prevention: AIDS Vaccine, Microbicide and ARV-based Prevention Science. October 17, 2016, Chicago, IL; also presented at the ENAR 2017 Spring Meeting and Joint Statistical Meetings 2017

"Addressing Batch Effects and Latent Variables in Gene Co-Expression Network Analyses." Topic contributed paper presentation. Joint Statistical Meetings. August 4, 2016, Chicago, IL.

"Estimating the Protective Effect of Longitudinal Drug Concentration in Pre-exposure Prophylaxis for HIV Prevention." Webinar. May 3, 2016, Center for Prevention Implementation and Methodology at Northwestern University, Department of Psychiatry and Behavioral Sciences.

## Teaching Experience

October 2017	The American University of Armenia, Gerald and Patricia Turpanjian School of Public Health Instructor for Biostatistics: Modeling and Sampling
Summer 2017	JHSPH, Graduate Summer Institute of Epidemiology and Biostatistics Teaching Assistant for Statistical Reasoning in Public Health II Teaching Assistant for Longitudinal Data Analysis
Sept. 2016- Present	JHSPH, Department of Biostatistics Lead TA and Lab Instructor for Statistical Methods in Public Health I-III
Summer 2016	JHSPH, Graduate Summer Institute of Epidemiology and Biostatistics Teaching Assistant for Statistical Reasoning in Public Health I-II Teaching Assistant for Longitudinal Data Analysis
2015-2017	Co-organizer of Johns Hopkins Department of Biostatistics TA Training Day
Summer 2015	JHSPH, Graduate Summer Institute of Epidemiology and Biostatistics Teaching Assistant for Data Analysis Workshop II
Sept 2014-May2016	JHSPH, Department of Biostatistics Teaching Assistant for Statistical Methods in Public Health I-IV
Sept 2010-Dec 2011	Pomona College, Department of Mathematics Teaching Assistant for Vector Calculus, Linear Algebra, Combinatorics

## Awards and Scholarship

2016	<b>June B. Culley Award:</b> Honors outstanding achievement by a Biostatistics student on his or her school-wide oral examination paper
2013-2015	Doctoral Training Grant in Environmental Biostatistics
2009-2013	<b>Undergraduate Awards:</b> Phi Beta Kappa, Pomona College Scholar, Llewellyn Bixby Sophomore Prize in Mathematics, Jaeger Freshman Prize in Mathematics

## Programming and Software

R, Stata, MATLAB, LaTeX, shell scripting