Title
From Trials to Public Health Impact: Transportability of Causal Effects to Inform Implementation of HIV Pre-exposure Prophylaxis

Author
Mehrotra, Megha Lynne

Publication Date
2019

Peer reviewed|Thesis/dissertation

From Trials to Public Health Impact: Transportability of Causal Effects to Inform Implementation of HIV Pre-exposure Prophylaxis

by

Megha Lynne Mehrotra

DISSERTATION

Submitted in partial satisfaction of the requirements for degree of

DOCTOR OF PHILOSOPHY

in

Epidemiology and Translational Science

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:

759D048A0C4A4F7...

David Glidden

Chair

DocuSigned by:

Maria Glymour

Maria Glymour

DocuSigned by:

Elvin Geng

DocuSigned by:

F80CC8225AAD43C...

Daniel Westreich

Committee Members

Dedication

I am grateful for the endless support, encouragement, and kindness of the Department of Epidemiology and Biostatistics at UCSF. The faculty, staff, and students that I've been lucky to interact with have helped to shape me into the epidemiologist that I am today. I'm grateful for the opportunity I've had to learn from all of you.

To my dissertation committee: David Glidden, Maria Glymour, Elvin Geng and Daniel Westreich: I always tell people that I'm probably the happiest doctoral student they've ever met, and there's no doubt in my mind that a huge reason for my success and happiness over the past 5 years is because of you. I'm lucky to have found such an inspiring and fun group of mentors, and I could not have reached this point without your investment in me. Thank you all for providing me the perfect balance of support, guidance, and freedom to grow and learn, while making the whole process truly enjoyable. Dr. Glidden, thank you for being the best possible mentor to me every step of the way over the past 10 years. I've learned so much and am so grateful for your wisdom and support. Dr. Glymour, thank you for your incredible leadership—not only on my committee but as a program director. Thank you for being a reliable source of motivation, advice, and sanity-checks, and thank you for pushing me to explore careers away from the Bay Area. I learned a lot in the process. Dr. Geng, thank you for the endless opportunities to work on fascinating projects, for pushing me to be a better writer, and for introducing me to the world of rigorous implementation science research. Your group meetings are always the highlight of my week. Dr. Westreich, thank you for being such a great mentor and advocate and being willing to serve on my committee from afar. You've consistently challenged me and pushed me to have a deeper understanding of causal inference and transportability.

To Robert Grant: Thank you for taking a chance on me in 2009 and giving me the freedom to learn and evolve over all these years. My career grew up with the iPrEx study, and you showed me how to simultaneously conduct state of the art science while still being a strong

advocate for the people and communities who make our research possible. I learn something new every time we talk. Thank you for always pushing boundaries and standing up for what's right.

To Patricia Defechereaux, Vanessa McMahon, Carlo Hojilla, Robert Hance, and the rest of the Grant Lab and iPrEx Study Team: I cannot think of a more fun and inspiring group of people to have worked with over the past 10 years. Thank you for showing me how to get things done.

To my peers in the PhD program, particularly my cohort, Stephen Asiimwe, Joshua Demb, Natalie Engmann, and Alyssa Mooney: Thank you for being so brilliant, encouraging, and fun. I've learned so much from each of you and really value you all as colleagues and friends.

To Scott Zimmerman: I've learned more from bouncing ideas off of you than from any class or textbook. Thank you for learning causal inference with me, for always being willing to indulge my methodological tangents, and for making me use good notation.

To my family, Sunil, Pravina, and Dhruv: thank you for your unconditional love and support, and for constantly reminding me that you believed in me every step of the way.

Finally, to Ryan: thank you for being my partner through everything. Over the past 5 years, I worked harder than ever but also had the most fun I've ever had. Thank you for the fun. Thank you for making sure we spent as much time as possible in the mountains. Thank you for reminding me that life isn't just about work—no matter how much I may enjoy the work.

Acknowledgment of Previously Published Materials

A version of Chapter 1 in this dissertation was published in *Journal of Acquired Immune Deficiency Syndromes* in 2019.[1] The Dissertation Committee members supervised this research that forms the basis of this dissertation chapter, the published material is substantially the product of Megha Mehrotra's period of study at UCSF and was primarily conducted and written by her. The work she completed for this published manuscript is comparable to a standard dissertation chapter.

Approved:

David Glidden, PhD, Dissertation Chair

[1] Mehrotra ML, Westreich D, McMahan VM, et al. Baseline Characteristics Explain Differences in Effectiveness of Randomization to Daily Oral TDF/FTC PrEP Between Transgender Women and Cisgender Men Who Have Sex with Men in the iPrEx Trial. *J Acquir Immune Defic Syndr*. 2019;81(3):e94-e98. doi:10.1097/QAI.0000000000002037

From Trials to Public Health Impact: Transportability of Causal Effects to Inform Implementation of HIV Pre-exposure Prophylaxis

Megha L. Mehrotra

ABSTRACT

With the support of several successful randomized placebo-controlled trials, the FDA approved Truvada for daily oral pre-exposure prophylaxis of HIV (PrEP) in the United States in 2012, and shortly thereafter the CDC and WHO released guidelines for widespread PrEP use by all at-risk individuals around the world. However, PrEP rollout is still is infancy, and there are several important questions regarding PrEP implementation that cannot be addressed by randomized trials. The causal transportability theory developed by Pearl and Bareinboim is a mathematically-grounded framework used to consider how effects observed in one setting might be applied to another. This dissertation proposes novel ways transportability can be applied to improve how trial results are used to inform implementation of PrEP. applies transportability to address some of these lingering questions about PrEP implementation.

The first chapter uses transportability to better understand why randomization to PrEP was effective in preventing HIV among cisgender men but not effective among transgender women in the iPrEx study. We find that after transporting the results of the trial from cisgender men to transgender women, differences in measured baseline characteristics between the populations were sufficient to explain the observed effect heterogeneity in the trial. The second chapter demonstrates how transportability can be applied to subgroup analyses of randomized controlled trials to produce target-specific guidance for how to most efficiently implement new interventions. To illustrate this approach, we transport subgroup analyses of the iPrEx trial to two hypothetical target populations and show that the subgroups with the lowest number needed to treat differs depending on the composition of the target population. The third and final chapter addresses a common practical challenge faced in applying transportability theory to real-world data: how to decide which variables to include in a transport estimator. In this

chapter, we discuss the various types of unnecessary variables that may inadvertently be included in transport estimators. We use a Monte Carlo simulation study to identify what types of variables should be included to maximize the performance (with respect to mean-squared error) of the parametric g-computation transport estimator.

Together these projects highlight how transportability theory can be applied to improve translation of study results to real-world populations.

Table of Contents

List of Figures

List of Tables

INTRODUCTION

Results of randomized controlled trials are critically important in the early stages of implementing a new intervention. Trial results 1) demonstrate if the intervention was effective, 2) highlight which groups it was most effective for, and 3) provide early insight into some potential challenges in how an intervention should be implemented. However, trial populations are rarely representative of real-world populations planning on implementing a new intervention, and differences between populations may affect how useful trial results can be for planning implementation. Until recently, this issue of external validity of trial results was an intractable problem. However, recent developments in the causal inference literature provide a theoretical solution to many of these challenges and can improve the interpretation of trial results for broader populations.

THE TRANSPORTABILITY FRAMEWORK

Why interventions might vary in effectiveness across different settings is intuitive: if there are certain characteristics that modify the effect of an intervention, and the distribution of those characteristics varies from setting to setting, then the intervention's effectiveness would similarly vary. Further, if we can measure and account for all the characteristics that both modify the effect of the intervention and differ between two settings, then we should be able to predict how effective an intervention would be if it were to be implemented in the new setting.

Transportability formalizes this intuition by building on the theoretical foundations of observational causal inference. In doing so, simple modifications of existing tools and statistical estimators that are widely used in the causal inference literature can be applied to predict how well an intervention might work when implemented in a new setting where it was not formally tested. That is, transportability provides tools to (1) formally evaluate whether findings in one

setting could be used to generate valid estimates in another, and (2) if so, estimate what the effect would have been had the study been conducted in the new setting.

Much in the same way that observational causal inference uses directed acyclic graphs (DAGs)[9] to identify the variables needed to control for confounding, transportability employs similar causal graphs—called selection diagrams-- to assist in isolating the important characteristics that determine whether and how the effectiveness of an intervention might differ between the study population and the population to which we wish to apply the results (the target population). Selection diagrams encode formal assumptions about the underlying causal relationships and mechanisms through which an intervention is believed to operate in the study population as well as assumptions about how the target and study populations differ from one another.

Selection diagram begin with a traditional DAG representing the study population—paying special attention to the mechanisms through which the intervention is hypothesized to affect the outcome and to any characteristics that may affect the outcome or modify the effectiveness of the intervention (**Figure 1.1**). Selection diagrams explicitly consider how a target population differs from the study population by including selection nodes indicating these potential differences (**Figure 1.2**). Unlike standard random variables that are usually included in DAGs, selection nodes do not have probability distributions and cannot be influenced by other variables in the graph. Instead, they function as indicators that point to where the data generating processes may differ between the two settings. In other words, they indicate where, if we were to draw a separate DAG for the target population, we might expect the processes that give rise to the data might differ between contexts. Importantly, the absence of a selection node on a variable indicates that we assume there are no differences in that variable's

distribution between the two populations given its parents (ie. the most proximal causes of the variable explicitly represented on the DAG).

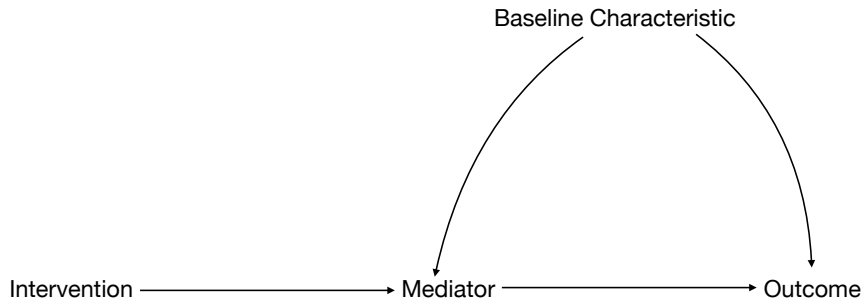


Figure 1.1 Directed Acyclic Graph

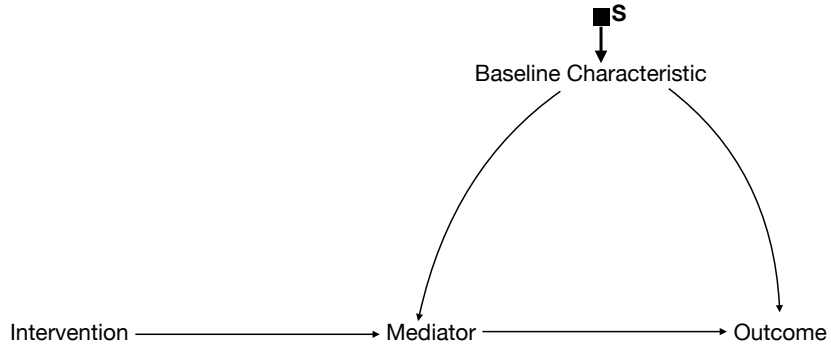


Figure 1.2 Selection Diagram

Selection diagrams reveal first whether the effect of the intervention which was estimated in the study population can be transported to a specific alternative target population given the data available from both the study population and the target population: if we can measure (and thereby adjust) for enough variables such that all the selection nodes are rendered independent of the outcome, the observed estimate could be used to produce a valid estimate in the target population, and we would deem that the observed estimate is

transportable to the target population.[1] Further, we can derive the specific transport formula to predict the effect in the target population based on the selection diagram.

HIV PRE-EXPOSURE PROPHYLAXIS

In 2010, the iPrEx study published the first results from a randomized controlled study showing a 44% reduction of HIV incidence in those randomized to receive daily oral Truvada for pre-exposure prophylaxis (PrEP) compared to placebo.[2] In 2012, with the support of additional randomized controlled trials (RCTs) in serodiscordant couples[3] and heterosexual men and women,[4] the US Food and Drug Administration approved the use of Truvada for prevention of HIV infection.[5] In 2015, PrEP was included as a key component of President Obama's National HIV/AIDS Strategy,[6] and recently the WHO released updated guidelines for widespread PrEP use globally.[7] Given the tremendous promise of PrEP thus far and the demonstrated effectiveness of widespread HIV testing and treatment,[8,9] researchers, advocates, and policy-makers are beginning to recognize that we now have the necessary tools to dramatically reduce – if not completely halt—HIV transmission.[10–12] However, despite calls from the CDC, WHO, and International AIDS Society for global PrEP roll-out, the United States, France, and South Africa remain the only countries to have approved Truvada for prevention of HIV thus far. In order for PrEP to reach its full potential, it must be efficiently and widely implemented around the globe. The urgent need to address the challenges in PrEP implementation is reflected in the recent strategic plans and research priorities of PEPFAR, the National Institute of Mental Health, and the Office of AIDS Research [13–17].

PrEP effectiveness is strongly tied to adherence, [2–4,18–22] and those who are unable to achieve sufficient adherence will not benefit from PrEP. Thus, adherence support through counseling and monitoring will play an important role in PrEP programs. However, these tools are often costly and may require substantial investments in health-systems infrastructure. Thus, particularly in resource-limited settings, policy makers will need to efficiently target adherence

support tools towards those who need them most. Unfortunately, identifying priority populations for adherence support based on clinical studies can be challenging. For example, transgender women (TGW) had much lower adherence compared to men who have sex with men (MSM) in the iPrEx study, and therefore did not benefit from PrEP randomization according to an intention-to-treat analysis.[2,23] However, because the TGW and MSM populations in the iPrEx study differed significantly in a number of important demographics (including age, education, and race), it is possible that the observed differences in adherence between MSM and TGW can be fully explained by the other demographic disparities between the populations. Because of the large number of differences between MSM and TGW, standard regression approaches to attempt to answer this question are not practical.[24]

Since 2007, UNAIDS has promoted the "Know your epidemic. Know your response." campaign.[25,26] This strategy highlights the need for tailored HIV prevention policies to match the heterogeneous nature of the HIV epidemic; no single prevention strategy will work in all contexts. PrEP implementation is no different, and policy makers will need to know how PrEP fits in to the response to their own HIV epidemic. Successful PrEP implementation will require interpreting and synthesizing the clinical trial data to (1) estimate how well PrEP will work in each context, and (2) identify populations who would most benefit from additional adherence support.

OBJECTIVES OF DISSERTATION

The overall goal of this dissertation is to apply transportability methods to address several of these important issues surrounding implementation of HIV pre-exposure prophylaxis. This will be addressed over three different chapters: 1) Using transportability to determine whether population compositional differences between cisgender men and transgender women in the iPrEx study were sufficient to explain the observed effect heterogeneity in the trial; 2)

applying transportability to subgroup analyses of the iPrEx study to produce target-specific implementation guidance of PrEP; and 3) conducting a simulation experiment to guide variable selection strategies for applied transport estimators.

Chapter 1 is an application of transportability that addresses a key question about PrEP effectiveness in transgender women. Chapter 2 applies transportability to the subgroup analyses of the iPrEx study, but also illustrates how this approach can be useful more broadly in other subgroup analyses of clinical trials. Chapter 3 is a methods paper that aims to inform how transportability theory can best be applied in practice.

Individually, each chapter evaluates an important question about PrEP implementation or application of transportability theory to real-world questions. Together, these projects highlight the broad utility of transportability theory and provide a guide for maximizing the public health relevance of randomized trial results.

CHAPTER 1: Baseline characteristics explain differences in effectiveness of randomization to daily oral TDF/FTC PrEP between transgender women and cisgender men who have sex with men in the iPrEx trial

Megha L. Mehrotra, Daniel Westreich, Vanessa M. McMahan, M. Maria Glymour, Elvin Geng, Robert M. Grant, David V. Glidden

ABSTRACT

Background: The iPrEx trial found that randomization to daily oral tenofovir disproxil fumarate/emtracitabine pre-exposure prophylaxis (PrEP) reduced HIV incidence by half in cisgender men who have sex with men (MSM) but found no benefit for transgender women who have sex with men (TGW). This unexplained difference is a barrier to PrEP implementation in TGW. We assess whether measured baseline participant characteristics can account for the difference in effectiveness of randomization to PrEP between TGW and MSM.

Methods: With data for 2,160 MSM and 339 TGW iPrEx participants who were HIV negative at baseline, we used the transportability framework to estimate what the intention to treat (ITT) effect of randomization would have been in MSM participants, had they shared the same distribution of baseline characteristics as TGW participants. We used a generalization of the parametric g-formula to transport the ITT incidence rate ratio (IRR) from MSM to TGW.

Results: The observed IRR in TGW was 1.29 (95%CI [0.24, 2.35]) and 0.53 (95%CI [0.36, 0.77]) in MSM. The final transport estimator included condomless receptive anal intercourse, number of partners, history of STIs, history of transactional sex, living situation, and baseline depressive symptoms. The transported estimate for MSM, i.e., the effect anticipated if MSM had the same distribution of these 6 characteristics as TGW, was IRR=1.28 (95%CI [0.12, 40.04]).

Conclusions: Population composition differences between MSM and TGW in iPrEx fully explained the observed effect heterogeneity in the trial.

INTRODUCTION

Daily oral tenofovir disproxil fumarate/emtracitabine (TDF/FTC) for HIV pre-exposure prophylaxis (PrEP) nearly eliminates the risk of HIV infection in certain populations when taken consistently.[27–29] However, PrEP will not live up to its full potential unless those at greatest risk of HIV infection use it. In particular, PrEP has the potential to be particularly impactful for transgender women-- a key population carrying one of the highest HIV burdens globally.[30] However, uptake of PrEP in this population has been low, and this may be in part due to a lack of high quality evidence about PrEP in transgender women.[31]

iPrEx was the only placebo-controlled randomized study of daily oral PrEP that included any transgender women who have sex with men (TGW), and consequently, the trial's results play an outsized role in our understanding of PrEP's efficacy in this key population.[2] Though randomization to the active arm was associated with a 44% reduction in HIV incidence in the sample overall, stratified analyses found no benefit for TGW (hazard ratio 1.1, 95%CI [0.5, 2.7]).[23] A closer look at measured drug levels in iPrEx found that tenofovir concentrations were generally lower in TGW compared to cisgender men who have sex with men (MSM), and drug was not detected at the seroconversion visit in any TGW on the active arm who became HIV positive.[23]

There are at least two possible explanations for the iPrEx results. First, there were a multitude of measured baseline differences between TGW and MSM. If these population composition differences occurred across characteristics that were important effect modifiers of PrEP's effectiveness—either by affecting adherence to PrEP or by modifying HIV risk—then even in the absence of any biological differences in TDF/FTC's efficacy, the intention-to-treat

(ITT) estimate of PrEP's effectiveness in iPrEx might differ between the two groups.[32] Second, there may be other unknown or unmeasured differences between TGW and MSM that might impact TDF/FTC's effectiveness. For example, recent small pharmacological studies suggest that feminizing hormones might interfere with the ability of tenofovir to block HIV infection by lowering the available blood concentration of tenofovir diphosphate.[33,34] However, whether these potential drug interactions would be sufficient to affect the overall efficacy of TDF/FTC in TGW is unclear. Understanding why randomization to PrEP was not effective in TGW in iPrEx may have useful implications for PrEP implementation.

In this manuscript, we aim to better understand to what extent population composition differences between MSM and TGW could explain the observed effect heterogeneity in iPrEx. We frame this issue as a transportability[1] question, and estimate what the ITT effect of randomization to PrEP would have been in MSM had they shared the same population composition as TGW in the study.[24,35] If this transported estimate is similar to the observed ITT estimate in TGW, then we can conclude that the effect heterogeneity in iPrEx might be due to measured population composition differences alone. If, on the other hand, the transported estimate is not similar to what was observed in iPrEx, then this suggests that unique contextual or biological factors (or unmeasured differences in population composition) were the sources of the effect heterogeneity in the study.

METHODS

Study population and procedures

iPrEx was a placebo-controlled randomized trial of daily oral TDF/FTC PrEP conducted between 2007 and 2011 in Brazil, Peru, Ecuador, the United States, South Africa, and Thailand. iPrEx enrolled 2499 cisgender men and transgender women who have sex with men.[2] All participants were HIV-negative at enrollment, reported risk behavior for HIV, and were assigned male sex at birth. Gender identity was recorded via a computer assisted structured interview

(CASI) where participants were asked how they identified, and any participant who selected "trans", "woman", or "travesti" (in Brazil, Peru, or Ecuador) was considered to be a TGW for the purposes of this analysis. In addition, we included any participant who reported taking feminizing hormones (irrespective of gender identity) as a TGW. This is consistent with prior subgroup analyses of the iPrEx trial.[23]

The same baseline CASI questionnaire also asked participants questions about demographics, living situation, relationship status, recent and lifetime sexual history, and substance use. Depressive symptoms were measured via an interviewer-administered Center for Epidemiologic Studies Depression Scale (CES-D). Detailed study procedures can be found in Grant et al, 2010.[2]

Variable selection and statistical methods

We first estimated the observed intention to treat incidence rate ratio in MSM (IRR_{msm}) and TGW (IRR_{tgw}) using a Poisson regression that included an offset for follow-up time. We excluded individuals who were HIV positive at enrollment or who did not return after their enrollment visit.

We estimated what the incidence rate ratio would have been in MSM had they shared the same baseline characteristics as the TGW study participants ($\widetilde{IRR}_{msm}$). We identified 15 candidate baseline characteristics that we hypothesized were both associated with HIV incidence and differed in distribution between MSM and TGW in iPrEx: age; total number of partners in the prior 3 months; any condomless receptive anal sex in the prior 3 months; sexual role (top, bottom, or versatile); race; ethnicity (Hispanic/Latino or non-Hispanic/Latino); country of residence; highest level of education; marital status; living situation ("With whom do you live primarily?"); past month alcohol consumption; history of transactional sex in the past 6 months; any history of a sexually transmitted infection in the past 6 months; past month cocaine use; and

past week depressive symptoms. We used a data-driven variable selection algorithm to narrow this list of 15 potential covariates down to include only those that are necessary and sufficient to transport the ITT incidence rate ratio from MSM to TGW. In order for a variable to be selected, it must both modify the ITT incidence rate ratio among MSM and differ in distribution between MSM and TGW.[36]

Using this reduced set of variables ($\boldsymbol{W}^*$), we applied a generalization of the g-formula[37] to estimate $\widetilde{IRR}_{msm}$.[37,38] This approach is analogous to model-based direct standardization in which the MSM population is standardized to resemble the distribution of covariates observed in TGW.[39] Assuming correct model specification, $\widetilde{IRR}_{msm}$ estimates the ITT incidence rate ratio MSM would have had if they had the same distribution of baseline covariates as TGW in iPrEx. We also estimate the percent of the observed effect heterogeneity between MSM and TGW that can be accounted for by measured population composition differences as $\left(\frac{\log\left(IRR_{msm}\right)-\log\left(\widetilde{IRR}_{msm}\right)}{\log\left(IRR_{msm}\right)-\log\left(IRR_{tgw}\right)} * 100\right)$. All analyses were performed using R v3.4.1 and STATA 15.1.[40,41]

RESULTS

Of the 2499 participants enrolled in iPrEx, 10 were HIV positive at enrollment and 44 did not return for follow-up visits. Of the remaining 2445 participants, 290 identified as trans, 29 identified as women, and 14 identified as men but reported using feminizing hormones. Together, these participants comprised the TGW group for this analysis (N=333/2445 (14%)). 67 (20%) of the 333 TGW participants reported using feminizing hormones.[23]

Table 2.1 compares the 15 candidate baseline characteristics between MSM and TGW. All but 3 of these variables differed significantly between MSM and TGW (mean baseline CESD score, ethnicity, and cocaine use in the past month). In addition, treatment assignment was balanced in both groups.

The variable selection algorithm identified 6 of these 15 baseline characteristics as being necessary and sufficient for transporting the incidence rate ratio: CES-D score; number of partners in the prior 3 months; any condomless receptive anal intercourse in the prior 3 months; living situation; any history of transactional sex in the prior 6 months; and any STI diagnoses in the prior 6 months.

In MSM in iPrEx, there were 77 incident HIV infections in the placebo arm and 41 infections in the active arm; in TGW, there were 10 infections in the placebo arm and 13 in the active arm. The ITT incidence rate ratio in MSM (IRR_{msm}) was 0.53 (95%CI [0.36, 0.77]), and in TGW the IRR_{tgw} was 1.29 (95%CI [0.24, 2.35]). After standardizing the MSM population according to the 6 selected baseline characteristics, the transported incidence rate ratio ($\widetilde{IRR}_{msm}$) was 1.28 (95%CI [0.12, 40.04]) (**Table 2.2**). This corresponds to nearly complete (99%) reduction in the observed effect heterogeneity. Overall, after accounting for baseline characteristics, the transported ITT incidence rate ratio closely resembles what was observed in transgender women in iPrEx.

DISCUSSION

Differences in population composition between MSM and TGW in iPrEx explained the observed effect heterogeneity in the trial results. This finding should allay concerns that biological differences in TDF/FTC's efficacy in TGW or other unmeasured factors were major drivers of the effect heterogeneity in iPrEx.

Whether using feminizing hormones reduces the absorption of tenofovir diphosphate enough to produce clinical differences in PrEP's efficacy remains an important question. Only 20% of TGW in iPrEx reported taking feminizing hormones, and there were no HIV infections among any of the participants taking feminizing hormones who were assigned to the placebo arm. Thus, we were unable to rule out the possibility that hormones reduce PrEP effectiveness using the iPrEx study data. Nonetheless, our results suggest that differences in a handful of

other measured characteristics between MSM and TGW in iPrEx could fully account for the effect heterogeneity observed in the trial.

The small number of transgender women included in iPrEx is a major obstacle for understanding PrEP in this key population. By using a transportability approach, we were able to better describe the effect heterogeneity in iPrEx after accounting for numerous differences between transgender women and cisgender men. Given the limited sample size, this would have been impossible using traditional regression adjustment. Additionally, we could have also estimated what the ITT result would have been had TGW in the trial had the same baseline characteristics as MSM to confirm our findings. Doing so would require fitting a conditional model adjusting for baseline characteristics in TGW alone, which was impossible given the small sample of TGW. Our findings are valuable despite wide confidence intervals, given that iPrEx is the only placebo-controlled randomized trial of PrEP that included any transgender women. Any insights about the effects of PrEP in this population are valuable even if substantial uncertainty remains.

The six baseline variables identified as necessary for transporting the incidence rate ratio between MSM and TGW were: number of partners in the prior 3 months; any condomless receptive anal intercourse in the prior 3 months; history of transactional sex in the prior 6 months; any STI diagnoses in the prior 6 months; current living situation; and CES-D score. Upstream structural and social factors that disproportionately affect TGW likely contribute to differences across these variables, so it is possible that these differences will persist in real-world contexts.[30,42–46] Consequently, our results do not imply that effectiveness of TDF/FTC PrEP implementation in the general population will necessarily be the same for both transgender women and MSM. Generalization to external settings requires knowledge about the population compositions of TGW and MSM in the specific target population of interest.[1,32,47,48]

Moving forward, there remains an urgent need for high-quality trans-specific research on HIV prevention strategies.[49] The effect heterogeneity in iPrEx exemplifies why transgender women should not be aggregated with cisgender men when conducting research, and future studies should ensure that enough transgender women are included in studies to provide adequate power to analyze these groups separately.[50] Additionally, further research is needed on PrEP for transgender men or non-binary individuals to ensure that PrEP implementation programs meet the needs of everyone who could benefit from PrEP.

Overall, our study--along with others from iPrEx and iPrEx OLE-- suggests TDF/FTC PrEP works similarly for MSM and transgender women when accounting for other characteristics. PrEP should be offered to anyone at risk of HIV infection regardless of gender identity.[23]

Table 2.1 Baseline characteristics by gender

		TGW (N=333)	**MSM** (N=2112)	**p-value**
Age at baseline, mean (SD)		26 (7)	27 (9)	0.030
CESD Score, mean (SD)		17 (8)	17 (8)	0.63
Number of partners in prior 3 months, median (IQR)		15 (5, 55)	5 (3, 13)	<0.001
Any condomless receptive anal intercourse in the prior 3 months		286 (86%)^	1172 (55%)	<0.001
Country	*US*	6 (2%)	217 (10%)	<0.001
	Peru	184 (55%)	1192 (56%)	
	Ecuador	60 (18%)	228 (11%)	
	Brazil	37 (11%)	327 (15%)	
	South Africa	4 (1%)	77 (4%)	
	Thailand	42 (13%)	71 (3%)	
Treatment assignment	*Placebo*	165 (50%)	1056 (50%)	0.88
	Active Arm	168 (50%)	1056 (50%)	
Ethnicity	*Non Hispanic/Latino*	84 (25%)	597 (28%)	0.25
	Hispanic/Latino	249 (75%)	1515 (72%)	
Race	*Black/African American*	19 (6%)	186 (9%)	<0.001
	White	38 (11%)	386 (18%)	
	Mixed/Other	234 (70%)	1452 (69%)	
	Asian	42 (13%)	88 (4%)	
Marital Status	*Single*	237 (71%)	1594 (75%)	0.005
	w/Partner	95 (29%)	455 (22%)	
	Married	0 (0%)	33 (2%)	
	Divorced	1 (<1%)	28 (1%)	
	Widowed	0 (0%)	2 (<1%)	
Living Situation	*With family/friends*	226 (68%)	1628 (77%)	<0.001
	w/ Male partner	26 (8%)	120 (6%)	
	Alone	75 (23%)	299 (14%)	
	w/ Female partner	1 (<1%)	30 (1%)	
	other	5 (2%)	35 (2%)	
Education Level	*Less than Secondary*	125 (38%)	385 (18%)	<0.001
	Completed Secondary	122 (37%)	744 (35%)	
	Post-Secondary	84 (25%)	960 (45%)	
	No Answer/Missing	2 (1%)	23 (1%)	
Sexual Role	*Top*	14 (4%)	609 (29%)	<0.001
	Bottom	238 (71%)	587 (28%)	
	Versatile	75 (23%)	858 (41%)	
	Don't know	6 (2%)	58 (3%)	
Any transactional sex in prior 6 months		214 (64%)	790 (37%)	<0.001
Any STI diagnosis in prior 6 months		126 (38%)	515 (24%)	<0.001
Alcoholic drinks per day in the past month	*None/< once a month*	63 (19%)	427 (20%)	0.008
	1-4 per day	67 (20%)	557 (26%)	
	>=5 per day	150 (45%)	756 (36%)	
	Refused/Missing/Don't know	53 (16%)	372 (18%)	
Any cocaine use in the past month		25 (8%)	105 (5%)	0.055

^ All variables are N (%) except where noted

Table 2.2 Observed and transported intention to treat incidence rate ratios

	Incidence Rate Ratio (95% CI)	Observed Effect Heterogeneity[@]	Remaining Effect Heterogeneity[^]	Percent Explained[!]
MSM (observed)[#]	0.53 [0.33, 0.73]	-0.89		
TGW (observed)[$]	1.29 [0.24, 2.35]		-0.01	99%
MSM (transported)[&]	1.28 [0.12, 40.04]			

[#] IRR_{msm}: Intention to treat incidence rate ratio observed in MSM in iPrEx
[$] IRR_{tgw}: Intention to treat incidence rate ratio observed in TGW in iPrEx
[&] $\widehat{IRR}_{msm}$: Incidence rate ratio that would have been observed in MSM had MSM shared the same population composition as TGW in iPrEx
[@] $\log(IRR_{msm}) - \log(IRR_{tgw})$
[^] $\log(\widehat{IRR}_{msm}) - \log(IRR_{tgw})$
[!] Percent effect heterogeneity explained by population composition: $\left(\frac{\log(IRR_{msm}) - \log(\widehat{IRR}_{msm})}{\log(IRR_{msm}) - \log(IRR_{tgw})} * 100\right)$

CHAPTER 2: Target-specific subgroup analyses for implementation of new interventions

Megha L. Mehrotra, Daniel Westreich, M. Maria Glymour, Elvin Geng, David V. Glidden

ABSTRACT

Subgroup analyses of randomized controlled trials guide resource allocation and implementation of new interventions by identifying groups of individuals who are likely to benefit most from the intervention. Unfortunately, trial populations are rarely representative of the target populations of public health or clinical interest; unless the relevant differences between trial and target populations are accounted for, subgroup results from trials might not reflect which groups in the target population will benefit most from the intervention. Transportability provides a rigorous framework for applying results derived in potentially highly selected study populations to external target populations. The method requires that researchers measure and adjust for all variables that (1) modify the effect of interest and (2) differ between the target and trial populations. To date, applications of transportability have focused on the external validity of overall study results and understanding within-trial heterogeneity; but this approach has not yet been used for subgroup analyses of trials. Through an example from the iPrEx study of HIV pre-exposure prophylaxis, we illustrate how transporting subgroup analyses can produce target-specific subgroup effect estimates and numbers needed to treat. This approach may lead to more tailored and accurate guidance for resource allocation and cost-effectiveness analyses.

INTRODUCTION

Researchers regularly use subgroup analyses of randomized controlled trials (RCTs) to find groups within the overall trial population that benefitted most from randomization to the intervention.[51,52] Policy-makers then prioritize those groups with the lowest numbers needed to treat (NNTs)[53,54]—that is, the number of individuals needed to be offered the intervention to prevent one incident outcome-- to receive the intervention. For example, iPrEx[2] was a placebo controlled RCT that evaluated the safety and effectiveness of combination daily oral tenofovir disproxil fumarate/emtracitabine for HIV chemoprophylaxis (PrEP) in transgender women (TGW) and cisgender men who have sex with men (MSM). The study found that randomization to the active arm was associated with a 44% reduction in HIV incidence compared to the placebo arm.[2] A subsequent post-hoc subgroup analysis of the trial found that the lowest NNTs were among those participants who reported condomless receptive anal intercourse (ncRAI), cocaine use, or a sexually transmitted infection.[55] These results have subsequently informed policy recommendations and cost-effectiveness analyses of PrEP implementation.[56–58]

Using results from subgroup analyses to prioritize implementation relies on the often-unspoken assumption that the strata-specific effect sizes estimated in the trial accurately reflect expected effect sizes in real-world target populations. However, this assumption is unlikely to be met in most applications; with the exception of large, pragmatic, cluster-randomized trials, trial populations are highly selected and rarely representative of real-world target populations that ultimately implement new interventions. Just as differences between trial and target populations undermine the external validity of the overall study findings,[32] these differences also mean that the effect sizes estimated for a subgroup of the trial with a particular characteristic may be poor indicators of the expected effect sizes in target populations similar on that characteristic.[59–64] Indeed, even if the overall trial population resembles, on average, a particular target population, within subgroups, differences may still exist between the trial and the target populations.

Consider a simple example of a clinic deciding whether to adopt a new blood pressure therapy based on evidence from an RCT that enrolled individuals at high cardiovascular risk. The hypothetical RCT found that men benefitted more from the new therapy than women; cost-effectiveness analyses based on these results suggested that the clinic should only offer men the new therapy but keep women on the previous standard of care. Because individuals at high cardiovascular risk were differentially recruited for the study, the proportion of women in the trial who smoked was much higher than in the clinic population. If the new therapy is not effective among tobacco users, this could account for the lackluster results among women in the trial. If the trial had been conducted in the clinic population, where smoking is less common among women, the new therapy would have been deemed cost-effective for men and women alike. In this simple example, using the subgroup analyses from the RCT without accounting for differences in the trial and target populations would lead to incorrect decisions about who to prioritize to receive the new therapy.

Recent developments in causal inference provide a principled approach for extending—or transporting-- the results of a study to an external target population.[65] This approach sets forth the principles and conditions that enable using the results of a study to infer what those results would have been had the study been conducted in an external target population.[1,32,66] To do so, all variables that (1) modify the effect of the intervention and (2) differ in distribution between the study and target populations must be measured and accounted for.[1,65,67] When differences between populations are limited to pre-treatment (baseline) covariates, transportability conceptually coincides with standardization across several characteristics.[68]

To date, transportability has previously been applied to transport average treatment effects to new target populations[47,69,70] or to understand observed heterogeneity between sites[35] or groups[71] in a trial. The theory also presents a promising solution for producing target-specific guidance for how to prioritize new interventions, but to our knowledge this framework has not yet been employed for these purposes. Here, we use an example from the iPrEx study of HIV

chemoprophylaxis[2] to illustrate how to apply transportability theory and estimators to transport subgroup effect estimates and NNTs to two specific external target populations. We discuss the necessary assumptions and data that are required for this approach to be successful in practice.

METHODS

Motivating example

The iPrEx study population comprised a heterogeneous group of 2499 MSM and TGW in Brazil, Peru, Ecuador, the United States, South Africa, and Thailand. All participants were HIV-negative at enrollment, reported risk behavior for HIV, and were assigned male sex ant birth. The median age at enrollment was 25 and most participants had not received a college education.[2] In aggregate, the iPrEx study population is unlikely to be representative of other target populations planning to roll out PrEP. Moreover, the populations who are at highest risk of HIV vary across the world, and guidance for how to prioritize PrEP should be tailored accordingly to each specific setting.[26]

Suppose we are interested in implementing PrEP in two clinics that serve young Latino TGW and MSM with men in San Francisco and Chicago. The clinics have limited resources, and each would like to target outreach and marketing of PrEP to those who are most likely to benefit from it. Here, we focus on subgroups that can easily be measured via survey or self-report: gender identity, including cisgender men or transgender women (MSM and TGW); recent sexual behavior, including any condomless receptive anal intercourse in the prior 3 months ($ncRAI$) and primary sexual role ($top, bottom, versatile$); and any cocaine use in the prior 6 months. To generate customized recommendations for each clinic based on these subgroups, we estimate what the subgroup-specific intention-to-treat (ITT) one-year HIV risk differences and NNTs would have been had the iPrEx trial been conducted in each clinic population.

Data and measurements

The iPrEx study randomized 2499 HIV-negative MSM and TGW to receive either daily oral PrEP or placebo, and participants were followed from 2007-2010. We included all participants from the iPrEx trial who were HIV-negative at their enrollment visit and who had contributed any follow-up time (N=2441).

To represent our two target populations, we used all HIV-negative participants in the San Francisco (N=210) and Chicago (N=263) study sites of the Latino MSM Community Involvement Study.[72,73] The study was a cross-sectional survey conducted in 2004 of Latino gay or bisexual cisgender men or transgender women that aimed to collect information about the participants' experiences in their community, sexual behavior, and substance use. Data from the Latino MSM Community Involvement study were accessed through the Inter-university Consortium for Political and Social Research.[73]

In both the iPrEx and the Latino MSM Community Involvement studies, participants were asked about their sexual behavior, demographics, STI history, and alcohol and drug use via a computer assisted structured interview (CASI).[2,72,73]

Notation, target parameters and identification

Our goal was to estimate the subgroup-specific ITT HIV risk difference at one year between those randomized to the PrEP arm and those randomized to the placebo arm and the corresponding numbers needed to treat to prevent one infection per year in iPrEx, San Francisco, and Chicago. Our subgroups variables of interest were cisgender men who have sex with men (MSM); transgender women who have sex with men (TGW); people who reported any condomless receptive anal intercourse in the prior 3 months ($ncRAI$); primary sexual role ($top, bottom, versatile$); and people who reported using cocaine in the prior 6 months ($cocaine$). We use random variable Z to denote treatment assignment where $Z = 1$ indicates assignment to receive PrEP and $Z = 0$ indicates assignment to the placebo arm. We use HIV^Z to represent the

counterfactual outcome that would have been observed if $Z = z$ were assigned. S indicates the population of interest where $S = 0$ is the iPrEx study population; $S = s'$ is one of the two target populations where $s' \in \{Chicago, San\ Francisco\}$. $G = g$ indicates the subgroup of interest where $g \in \{MSM, TGW, ncRAI, top, bottom, versatile, cocaine\}$.

We define the ITT effect in subgroup $G = g$ in population $S = s$ as:

$$\boldsymbol{\psi_g^s = E\left(HIV^{Z=1} - HIV^{Z=0} \middle| G = g, S = s\right)} \quad \textbf{(Eq. 1)}$$

and the NNT[54] for each subgroup $G = g$ in population $S = s$ as:

$$\boldsymbol{\xi_g^s = \frac{1}{|\psi_g^s|} = \frac{1}{|E(HIV^{Z=1} - HIV^{Z=0} | G=g, S=s)|}} \quad \textbf{(Eq. 2)}$$

For simplicity, we assume there was no measurement error. To identify the target parameters within the iPrEx study population, we must assume:

1. Conditional treatment exchangeability: Z is independent of (HIV^0, HIV^1) given $G = g$, and $S = 0$. That is, there is no confounding of the association between treatment assignment and HIV incidence in the iPrEx study population within subgroup $G = g$. This assumption is met by randomization of treatment assignment in the iPrEx trial.
2. Treatment positivity: $P(Z = z \,|G = g) > 0$ for all g for which $P(G = g) > 0$. That is, there must be a non-zero probability of being assigned each treatment for each subgroup.[74] Randomized treatment assignment in the iPrEx trial guarantees that there are no structural positivity violations, but does not guarantee the absence of practical positivity violations, which are more likely to occur in smaller samples in subgroups.

In addition to the above assumptions, to identify the transported target parameters we must also meet the following criteria[24]:

3. Conditional population exchangeability: $E(HIV|S=0, \boldsymbol{W_g}, Z, G=g) = E(HIV|S=s', \boldsymbol{W_g}, Z, G=g)$. Within subgroup $G=g$, the iPrEx study population and target population are exchangeable with respect to HIV incidence conditional on some set of measured characteristics ($\boldsymbol{W_g}$) and treatment assignment.
4. Population positivity: $P\left(S=0, Z=z \middle| \boldsymbol{W_g}, G=g\right) > 0$ $P_{\boldsymbol{W_g}|G=g,S=s'}$ $a.e.$ That is, every combination of $\boldsymbol{W_g} = \boldsymbol{w_g}$ that could be drawn from the distribution of $\boldsymbol{W_g}$ in each strata $G=g$ within each target population is represented in the iPrEx study population in $G=g$ and has a non-zero probability of being assigned $Z=z$.

Thus, for each subgroup $G=g$ we must condition on the set of variables $\boldsymbol{W_g}$that ensures that assumption 3 is met.

Selection diagrams are augmented directed acyclic graphs[75,76] introduced by Pearl and Bareinboim that assist in identifying a set of variables that satisfies assumption 3 above. In these graphs, selection nodes are not standard random variables. Instead, they indicate where differences in the causal model might exist between the trial and target populations.[1,77] An effect can be transported across the populations if there exists a set of variables that, if conditioned on, will make all the selection nodes independent (or d-separated[78]) from the outcome variable.[1]This set of variables, called the s-admissible set, satisfies the conditional population exchangeability assumption given above. For rules on how to evaluate d-separation in selection diagrams, please see Appendix A.

Figure 3.1 depicts our proposed selection diagram representing the assumed causal model within the iPrEx study and assumed differences between the study population and each target population. Based on our selection diagram, we identified the s-admissible set of variables ($\boldsymbol{W_g}$) for each subgroup analysis, i.e. the set of variables that is sufficient to d-separate all the selection nodes from the outcome conditional on the subgroup of interest such that $HIV \perp S \mid Z, G=g, \boldsymbol{W_g}$**:**

Gender identity ($\boldsymbol{MSM}$ and $\boldsymbol{TGW}$): age, education, number of partners, $ncRAI$, cocaine use, and alcohol consumption.

Condomless receptive anal intercourse ($\boldsymbol{ncRAI}$): age, education, number of partners, cocaine use, alcohol consumption.

Primary sexual role ($\boldsymbol{top}$, $\boldsymbol{bottom}$, $\boldsymbol{versatile}$): age, education, number of partners, $ncRAI$, cocaine use, alcohol consumption, and gender identity.

Cocaine use: age, education, number of partners, $ncRAI$, and alcohol consumption

Given the above assumptions, the target parameters within the iPrEx study population are identified by:

$$\psi_g^0 \equiv E\left(HIV^{Z=1} - HIV^{Z=0} \middle| G = g, S = 0\right)$$

$$= E[HIV|Z = 1, G = g, S = 0] - E[HIV|Z = 0, G = g, S = 0] \qquad \textbf{(Eq. 3)}$$

and:

$$\xi_g^0 = \frac{1}{|\psi_g^0|} = \frac{1}{|E[HIV|Z=1,G=g,S=0] - E[HIV|Z=0,G=g,S=0]|} \qquad \textbf{(Eq. 4)}$$

The transported target parameters are the subgroup-specific ITT effects and NNTs had the study been conducted in each target population (San Francisco or Chicago) and are identified by:

$$\psi_g^{s'} \equiv E(HIV^{Z=1} - HIV^{Z=0} | G = g, S = s')$$

$$= \boldsymbol{E}\big(\boldsymbol{E}[\boldsymbol{HIV}|\boldsymbol{Z}=\boldsymbol{1},\boldsymbol{W_g},\boldsymbol{G}=\boldsymbol{g},\boldsymbol{S}=\boldsymbol{0}]|\boldsymbol{S}=\boldsymbol{s'}\big)$$

$$-\boldsymbol{E}\big(\boldsymbol{E}[\boldsymbol{HIV}|\boldsymbol{Z}=\boldsymbol{0},\boldsymbol{W_g},\boldsymbol{G}=\boldsymbol{g},\boldsymbol{S}=\boldsymbol{0}]|\boldsymbol{S}=\boldsymbol{s'}\big) \qquad \textbf{(Eq. 5)}$$

And the transported NNTs are:

$$\xi_g^{s'} = \frac{1}{\left|\psi_g^{s'}\right|}$$

$$= \frac{\boldsymbol{1}}{\boldsymbol{E}\big(\boldsymbol{E}[\boldsymbol{HIV}|\boldsymbol{Z}=\boldsymbol{1},\boldsymbol{W_g},\boldsymbol{G}=\boldsymbol{g},\boldsymbol{S}=\boldsymbol{0}]|\boldsymbol{S}=\boldsymbol{s'}\big)-\boldsymbol{E}\big(\boldsymbol{E}[\boldsymbol{HIV}|\boldsymbol{Z}=\boldsymbol{0},\boldsymbol{W_g},\boldsymbol{G}=\boldsymbol{g},\boldsymbol{S}=\boldsymbol{0}]|\boldsymbol{S}=\boldsymbol{s'}\big)} \qquad \textbf{(Eq. 6)}$$

Estimation

To estimate the ITT risk difference for each subgroup in iPrEx, we used the parametric g-formula. We fit a log-binomial regression model with main terms for treatment assignment and the subgroup variable as well as an interaction term between treatment assignment and subgroup. Using this model, we predicted the marginal incidence risk difference at one year within each subgroup. Because treatment was randomly assigned, we did not adjust for any additional covariates in each subgroup analysis in the iPrEx study population.

To transport the ITT effects, we first generated stabilized inverse odds of selection weights[79] using the following formula:

$$IOSW_i = \begin{cases} \dfrac{P\big(S_i=0|\boldsymbol{W_g},G_i\big)}{P\big(S_i=s'\big|\boldsymbol{W_g},G_i\big)} * \dfrac{P(S_i=s',G_i)}{P(S_i=0,G_i)}, & S_i=0, \\ 0 \quad , & S_i=s' \end{cases}$$

Each component of the IOSW was estimated using logistic regression. Note that because the iPrEx study population is not a subset of either target population, inverse odds weights were used instead of inverse probability weights. In settings where the study population is fully nested within the target population, inverse probability weights would be an appropriate analogous estimator.[68,79]

The inverse odds weights were used to fit weighted log-binomial regressions with an interaction between treatment assignment and subgroup. We used this model to predict the

number of incident HIV infections at one year by treatment assignment within each subgroup in each target population, and we calculated the transported marginal risk difference. Standard errors and 95% confidence intervals were calculated using a bias corrected and accelerated bootstrap[80] with 2000 resamples. The bootstrap resampled both the iPrEx study population and target populations, and then calculated new weights and fit the weighted log-binomial regression on each bootstrap sample. This ensured that the variability in the target population was also incorporated into the standard errors.

The number needed to treat was estimated as the inverse of the difference in risk of HIV infection at one year of follow-up[81] giving the number of individuals who need to be offered PrEP needed to avert one infection in one year.

All analyses were conducted using R version 3.4.1[40] and STATA version 15.1.[41]

RESULTS

There were differences in most baseline characteristics across settings (**Table 3.1**), and in particular, the iPrEx study population had on average more recent sexual partners and more individuals reporting recent condomless receptive anal intercourse.

Figure 3.2 shows the subgroup-specific intention-to-treat risk differences at one year, and **Figure 3.3** shows the numbers needed to treat to prevent one infection in each population. In all settings, cocaine users had the lowest number needed to treat. In Chicago, the NNT was next lowest among those whose primary sexual role was “bottom,” whereas in iPrEx the sexual role with the lowest NNT was “versatile.” In all settings, PrEP is not expected to be beneficial for those whose primary sexual position is “top.” Finally, though there were apparent differences in the effectiveness of randomization to PrEP between cisgender men and transgender women in iPrEx, after transporting the results to San Francisco these gender differences are diminished.

DISCUSSION

Subgroups with the lowest NNTs in trial populations may not be the same groups that would have the lowest NNTs in target populations. Without formally accounting for differences between trial and target populations, subgroup-specific effect sizes and NNTs from trials may not provide useful guidance for implementing new interventions in target populations. The transportability framework is a rigorous solution for generating target-specific subgroup results and tailored implementation guidance.

Assuming that we have adequately measured and accounted for all of the characteristics that both modify the effectiveness of randomization to PrEP and differ between the study and target populations, our worked example demonstrates how subgroup analyses might give meaningfully different guidance regarding resource allocation if they are transported to the specific target populations. In iPrEx, those who indicate that their primary sexual role is "versatile" have much lower NNTs than other sexual roles. However, after transporting the results to Chicago, we see that the sexual role with the lowest NNT is "bottom," and in San Francisco we find that those who report recent condomless receptive anal intercourse have a lower NNT than any specific sexual role. Prioritizing PrEP according to self-reported sexual role is appealing, as the information can easily be gathered in a clinic-setting through a single question. To use sexual role as a means to prioritize PrEP efficiently however, the iPrEx results must be transported to each target population with distinct covariate distributions.

The application of transportability relies on the availability of high-quality individual-level data in both the trial and target populations. The outcome itself does not need to be measured in target populations—which is particularly helpful for rare or hard to measure outcomes like HIV incidence-- but in order for the transportability assumptions to reasonably be met, there needs to be a rich dataset of characteristics that are associated with the outcome gathered in the target population. Which characteristics need to be measured depends on the intervention and outcome of interest; simply gathering basic demographic information may not always be

sufficient for a particular outcome. Similarly, individual-level trial data including all relevant effect modifiers need to be available to generate policy-relevant recommendations. These data requirements are not trivial, but as increasingly more studies make their data available for secondary analyses, and as more data are collected and aggregated on individuals in real-world target populations, transportability will likely soon become more feasible in applied research.

Though the particular examples presented here are helpful for illustrating how transportability can be used to improve subgroup analyses, there are several important limitations that preclude interpreting these findings substantively. First, the Latino MSM Community Involvement Study was conducted in 2004, so the characteristics and behaviors described in these data may not reflect the current needs of these populations. Next, as PrEP has become more widely adopted around the world, the characteristics of those individuals who are likely to adhere to PrEP has undoubtedly changed. This means that, assuming we have met all the assumptions necessary for transport and that our models were correctly specified, our transported estimates could only be interpreted as the effects we would have observed had the iPrEx trial been conducted in each target population at the time it was conducted (2007-2010). This limitation is not unique to our example. Unless trial results are transported immediately at the end of the study, factors that affect uptake, adherence, and effectiveness of a new intervention are likely to change, and the transported results will become less relevant over time.

The results of our illustrative example were uncertain, as demonstrated by the wide confidence intervals in **Figure 3.2**. The numbers needed to treat, which are derived from the risk differences, are similarly uncertain--particularly for those subgroups that included few individuals (cocaine users, for example). This uncertainty reflects the fact that both the study and target populations included relatively small samples, and also underscores an important challenge in transporting subgroup analyses more broadly. Trials are often underpowered to detect subgroup differences, and transport estimators may reduce the precision of subgroup estimates. While

other transport estimators such as the parametric g-computation transport estimator and targeted maximum likelihood might be more efficient than the IOSW estimator, researchers applying these tools in practice should be careful in weighing the bias-variance tradeoff for their particular application.

Finally, a central issue that researchers will face when employing transportability methods is results are likely to be sensitive to the assumptions made in the selection diagram, and many of these assumptions are untestable. Selection diagrams, as with any other causal graph, are typically built using a combination of prior knowledge, subject matter expertise, and previously published literature. Usually there will still be considerable uncertainty about the accuracy of these diagrams. In practice, quantitative bias analyses that put reasonable bounds on the transported estimates are merited and further work should explore how best to implement these analyses for transportability.

Transportability is a transparent framework for describing, evaluating, and testing the assumptions needed to produce target-specific subgroup effect estimates and NNTs. Moving forward, researchers publishing trial results should ensure that all important variables that might be relevant for transporting findings to target populations are made available so that local health departments, policy-makers, and other researchers can generate tailored recommendations for how to implement new interventions.

Table 3.1. Baseline characteristics by population.

	iPrEx (N=2499)	San Francisco (N=210)	Chicago (N=263)
Number of male parters in prior 3 months	16.9 (35.6)	8.0 (10.2)	7.3 (10.0)
Age at baseline			
18-25	1374 (55.0%)	37 (17.6%)	76 (28.9%)
26-35	730 (29.2%)	78 (37.1%)	113 (43.0%)
36-45	270 (10.8%)	66 (31.4%)	50 (19.0%)
>45	125 (5.0%)	29 (13.8%)	24 (9.1%)
Highest level of education			
Less than HS	524 (21.0%)	57 (27.1%)	63 (24.0%)
HS	884 (35.4%)	38 (18.1%)	68 (25.9%)
College	1091 (43.7%)	115 (54.8%)	132 (50.2%)
Gender Identity			
Cisgender Man	2174 (87.0%)	172 (81.9%)	249 (94.7%)
Transgender Woman	325 (13.0%)	38 (18.1%)	14 (5.3%)
Alcohol consumption in prior month			
None/< once a month	496 (19.8%)	90 (42.9%)	68 (25.9%)
1-4/day	635 (25.4%)	72 (34.3%)	85 (32.3%)
>=5/day	931 (37.3%)	47 (22.4%)	108 (41.1%)
Don't Know	437 (17.5%)	1 (0.5%)	2 (0.8%)
Cocaine use in prior month			
No	2368 (94.8%)	187 (89.0%)	214 (81.7%)
Yes	131 (5.2%)	23 (11.0%)	48 (18.3%)
Primary sexual position			
Top	641 (25.7%)	34 (16.2%)	37 (14.1%)
Bottom	834 (33.4%)	89 (42.4%)	132 (50.2%)
Versatile	1024 (41.5%)	87 (41.4%)	94 (35.7%)
ncRAI in prior 3 months^			
No	1014 (40.6%)	159 (75.7%)	185 (70.3%)
Yes	1485 (59.4%)	51 (24.3%)	78 (29.7%)

^ any condomless receptive anal intercourse in the prior 3 months

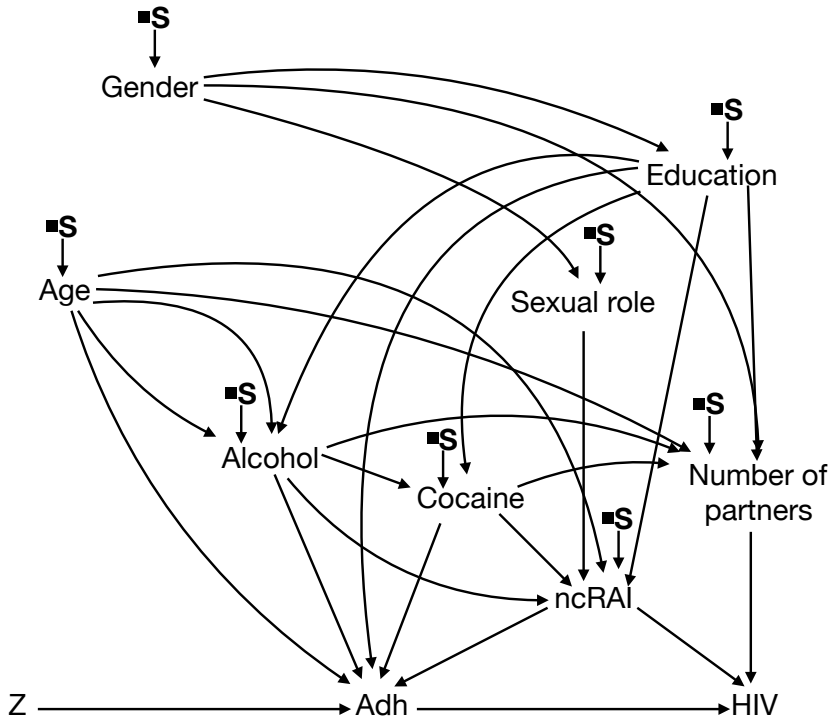


Figure 3.1 Z is treatment assignment; Adh is adherence; Age is age at baseline; Gender is gender identity; Education is highest level of education; Sexual role is primary sexual position ("top", "bottom", "versatile"); Alcohol is prior month alcohol consumption; Cocaine is prior month cocaine use; ncRAI is any condomless receptive anal intercourse in the 3 months prior to baseline; Number of partners is total number of male partners in the 3 months prior to baseline.

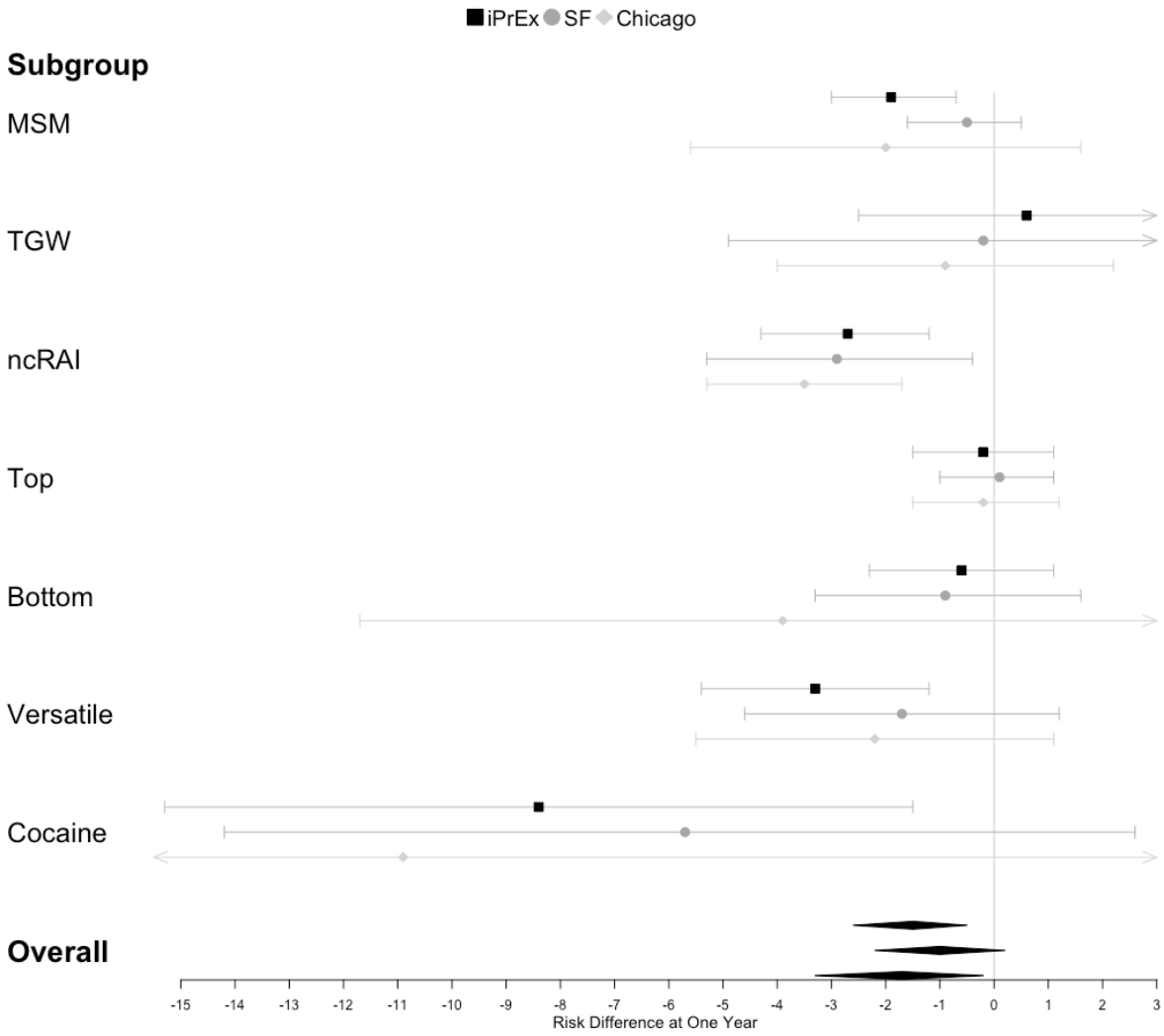


Figure 3.2. Subgroup-specific risk differences in iPrEx, San Francisco, and Chicago

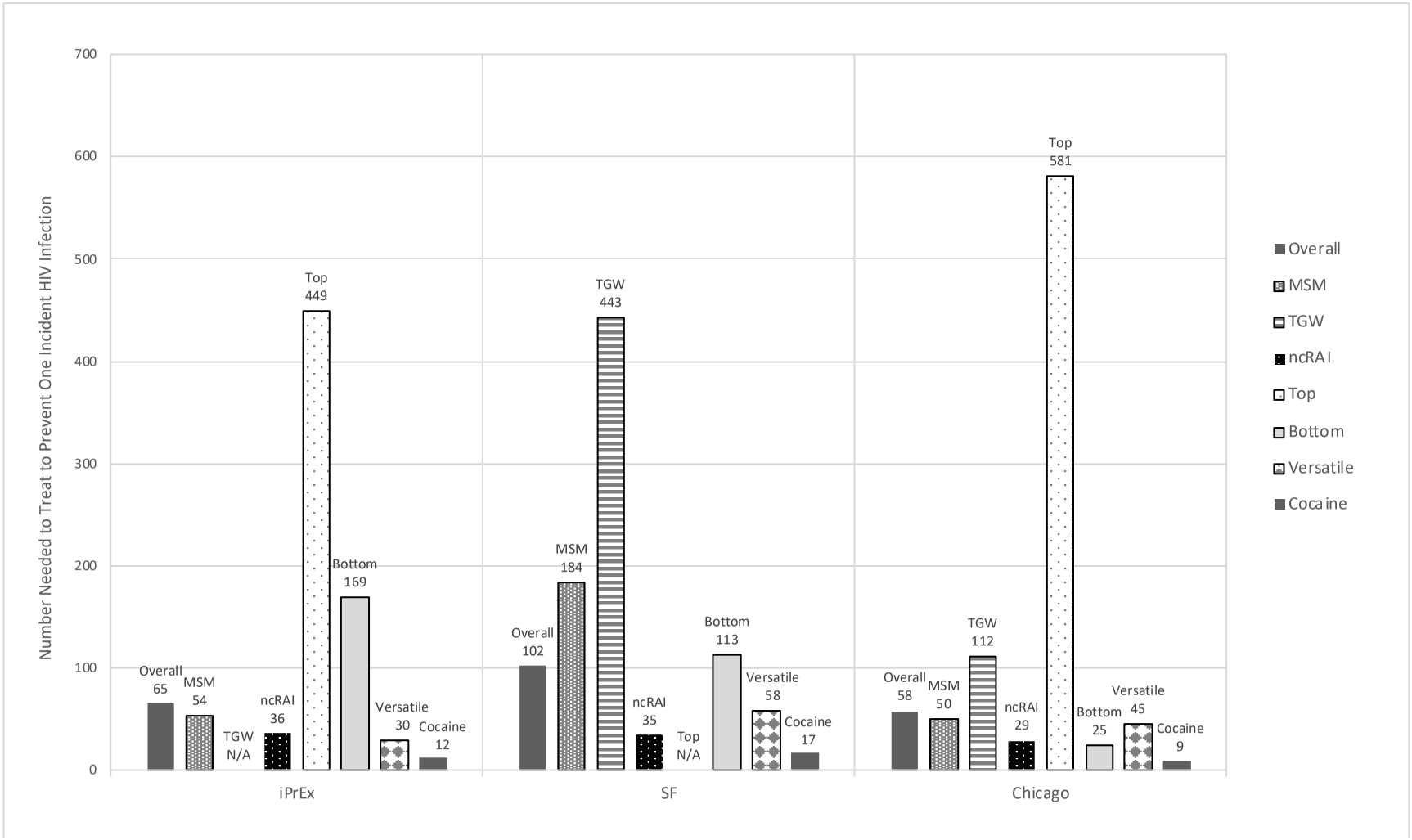


Figure 3.3. Subgroup-specific numbers needed to treat in iPrEx, San Francisco, and Chicago.

Appendix

d-separation rules for selection diagrams

The d-separation rules as described by Pearl, 1988[78] that are traditionally applied to directed acyclic graphs and other causal diagrams can also be applied to determine if there exists a set of variables such that conditional on that set, the conditional population exchangeability assumption is fulfilled.

Consider the selection diagram given in **Supplementary Figure 3.1**. In this example, A is the exposure, Y is the outcome, W1, W2, and W3 are other covariates, and the S nodes are the selection nodes. To transport the causal effect of A on Y from the study population to the target population, we must be able to identify a set of variables that makes all of the selection nodes independent of the outcome—or d-separates all of the selection nodes from the outcome.

To test whether all the selection nodes in a selection diagram are independent of the outcome, we need to test whether all of the paths from the selection nodes to the outcome can all be blocked by conditioning on other measured variables. A path between S and Y is blocked if:

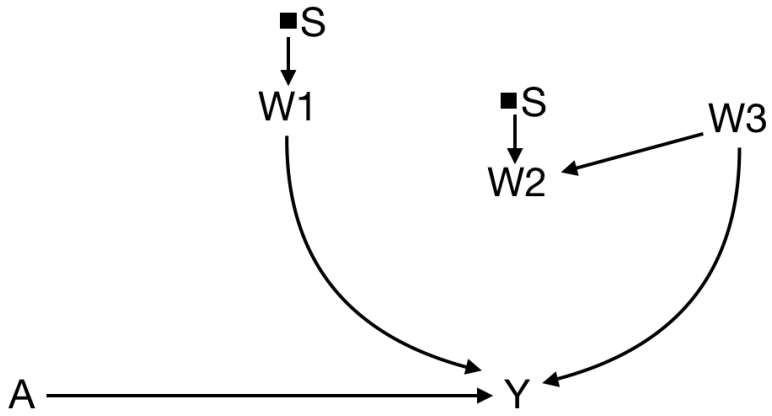
1) S→W1→Y and W1 is conditioned on

   or

2) S→W2←Y and W2 is **not** conditioned on.

W2 is a collider, and conditioning on a collider opens a path between the parents of the collider. This could be corrected by simultaneously conditioning on W3, thus blocking the path.

In Supplementary Figure 1, we would need to condition on W1 to block all the paths between the selection nodes and Y. If we condition on W2, we would open a path between the selection node and Y through W3.



Supplementary Figure 3.1. Selection Diagram

CHAPTER 3: Variable Selection for Transportability

Megha L. Mehrotra, M. Maria Glymour, Elvin Geng, Daniel Westreich, David V. Glidden

ABSTRACT

Transportability provides a principled framework to address the previously intractable problem of applying study results to new populations. Several transport estimators have been developed to use alongside the formal transportability theory. Here, we consider the problem of selecting variables for these transport estimators. We provide a brief overview of the transportability framework and illustrate that though selection diagrams are a vital first step in variable selection, these graphs alone may not identify the optimal set of variables for generating an unbiased transport estimate. Next, we conduct a simulation experiment assessing the impact of including unnecessary variables on the performance of the parametric g-computation transport estimator. Our results highlight that the types of variables included can affect the bias, variance, and mean squared error of the estimates. We find that addition of variables that differ between the source and target populations but that don't cause the outcome can increase the variance and mean squared error of the estimates, while inclusion of causes of the outcome—regardless of whether they modify the causal contrast of interest—reduces the variance of the estimates without increasing the bias. Exclusion of variables that are causes of the outcome but are not modifiers of the causal contrast does not increase bias. These findings suggest that variable selection approaches for transport should prioritize identifying and including all causes of the outcome in the study population rather than focusing on differences between the populations.

INTRODUCTION

The transportability framework, which builds on the theoretical foundations of causal inference, outlines the necessary rules and assumptions for determining when and how a causal effect estimated in one study population can be applied to an external target population.[65] These tools present a promising solution to the long-standing challenge of assessing the external validity of research findings and can also be used to better understand observed effect heterogeneity within a study.[35,38,82]

However, applying transportability methods to real-world problems is not always straightforward. A central challenge that researchers face in using these tools is deciding what variables need to be measured and included in their transport estimators. Similar to variable selection for confounding adjustment,[83] these decisions can greatly affect the bias and variance of the transported effect estimate, but to our knowledge, little has been written about different variable selection strategies for transport.

Ideally, subject matter expertise and a clear understanding of the study and target populations would be the primary guide for variable selection decisions. Selection diagrams—the causal graphs used for transport-- facilitate using prior knowledge of the underlying causal mechanisms to identify a set of variables that would be sufficient to transport an effect from a study population to a given target population. However, because of uncertainty about the underlying causal structure or mechanisms in real-world applications, using selection diagrams alone may be insufficient to narrow down the list of candidate variables to include only those that are necessary for a given application. Ultimately, even after careful use of selection diagrams, an applied researcher working with a finite sample of data will likely have to decide which variables she thinks are essential from an extensive list.

Here, we provide a practical guide to variable selection for transportability. We begin by briefly reviewing the transport framework and graphical approach to variable selection. Next, we introduce the *minimally sufficient transport set,* and illustrate why transporting causal contrasts

may require fewer variables than a standard selection diagram may indicate. Finally, we categorize the different types of variables (according to causal structure) that might be included in transport estimators and use Monte Carlo simulation experiments to evaluate how inclusion or exclusion of different variable types affects bias, variance, and mean squared error of the parametric g-computation transport estimator.

NOTATION AND DEFINITIONS

- Source population—population you are transporting results *from* (ie. study population)
- Target population—population you are transporting results *to*
- $P(Y^Z)$-- the distribution of counterfactual outcome Y if exposure Z is assigned value z.
- Φ -- a causal quantity that is a function of the counterfactual outcome distribution. For example, a causal contrast, i.e. a causal risk difference ($E(Y^{Z=1} - Y^{Z=0})$).
- S – selection node indicating population membership where $s \in \{0,1\}$ and $S = 1$ indicates the source population and $S = 0$ indicates the target population. These nodes are not standard random variables, but instead indicate where the data generating mechanisms may differ between the two populations.
- $\mathbf{TS_s}$ – an s-admissible transport set defined as a set of variables that d-separates all selection nodes from the outcome variable ($Y \perp S \mid \boldsymbol{TS_s}$). There may be more than one s-admissible set for a given graph.
- **TS** – a transport set defined as the set of variables included in a transport estimator. The transport set may or may not be an s-admissible set.
- **MSTS** – a minimally sufficient transport set. The smallest possible s-admissible transport set. There may be more than one **MSTS** for a given problem.

TRANSPORTABILITY

The goal of transportability is to extend or apply the results of a study conducted in one population (the *source population*) to another population (the *target population*). Why the results

of a study conducted one population may not apply to another is intuitive: if there are characteristics or factors that modify the effectiveness of the intervention under study and the distribution of these characteristics differs between the source and target populations, then we would expect that the results of a study would similarly vary depending on which population it was conducted in. If we are able to measure and account for those characteristics that both a) modify the effectiveness of the intervention and b) differ between two populations, we should be able to apply study results gathered in the source population to an external target population without having to repeat the entire study.

The transportability framework formalizes this intuition and sets forth formal mathematical rules and conditions under which the results of a study can be transported from a source population to a target population.[65] We define $P(Y^Z)$ as the counterfactual distribution of outcome Y if exposure Z is assigned value z for all possible values of Y and Z. This quantity can be thought of as the most general definition of a causal effect, as any causal contrast (ie. the causal risk difference $E(Y^{Z=1} - Y^{Z=0})$) is a function of this counterfactual distribution. $P(Y^Z)$ can be transported from a source population ($S = 1$) to a target population ($S = 0$) if the following assumptions are met:

1) S-admissibility (or conditional population exchangeability): $Y \perp S \mid \boldsymbol{TS_s}$ where $\boldsymbol{TS_s}$ is an s-admissible set.
2) Population positivity: $P(S = 1|\boldsymbol{TS_s} = \boldsymbol{ts_s}) > 0$ for every $\boldsymbol{ts_s}$ that has a positive density in the target population. That is, all values of the s-admissible set

Selection Diagrams for Variable Selection

To illustrate the transportability framework in action, we use a simple toy example loosely motivated by the Finnish Geriatric Intervention Study to Prevent Cognitive Impairment and Disability (FINGER)[84]. Suppose we conducted a randomized controlled trial evaluating

whether a multicomponent behavioral intervention was effective in reducing the 2-year risk of cognitive decline compared to standard of care among participants in Finland. The study found that randomization to a multicomponent behavioral intervention was effective in reducing the 2-year risk of cognitive decline, but we want to know what the results of this trial would have been had it been conducted in a US-based target population.

Figure 4.1 represents the true data generating mechanism for this toy example. For simplicity and without loss of generality, we assume that there are only two additional variables that might affect cognitive decline: systolic blood pressure >140 mmHg; and being a carrier of the apolipoprotein E-$\epsilon 4$ (APOE-$\epsilon 4$) variant. We define our outcome as risk of a 10% reduction in neurocognitive test battery (NTB) score after 2 years of follow-up. We define our source population $S = 1$ as the Finnish study population and our target population $S = 0$ as the US-based target population; Z is randomized treatment assignment; $B = 1$ (systolic blood pressure > 140mmHg) and $G = 1$ (APOE-$\epsilon 4$ carrier) both affect the risk of NTB score reduction by year 2. B and G differ in distribution between the study and target populations.

Akin to how directed acyclic graphs (DAGs) are used to select variables to control for confounding, selection diagrams are causal graphs used to determine which variables satisfy the s-admissibility criteria for transportability. To create a selection diagram, we begin by drawing a DAG that represents the data-generating model for the source population. Next, we add selection nodes that indicate where there might be differences in the data-generating models between the source and target populations (**Figure 4.2**). Selection nodes are not standard random variables; instead, they are indicators that point to the portions of the data-generating model that might differ between the two populations.

Any set of variables that d-separates all of the selection nodes from the outcome is an s-admissible set ($\boldsymbol{TS_s}$) for transporting $P(Y^Z)$. Note that a given graph may reveal more than one

s-admissible set. Based on the selection diagram given in **Figure 4.3**, the s-admissible set for this example is systolic blood pressure >140mmHg and APOE-$\epsilon 4$ ($\boldsymbol{TS_s} = \{B, G\}$).

Throughout this manuscript we restrict our discussion to scenarios in which selection nodes are only directed at pre-treatment variables. Transporting results in situations where there are selection nodes directed at mediating variables requires additional measurements and assumptions that are beyond the scope of this manuscript. For background on transporting causal effects when selection nodes are directed at mediating variables, we point readers to Appendix 3 of Pearl and Bareinboim, 2011[1] and Bareinboim and Pearl, 2012[85] for more details.

MINIMALLY SUFFICIENT TRANSPORT SET

A minimally sufficient transport set ($MSTS$) is the smallest possible s-admissible set that would satisfy assumption 1 for transporting a particular causal quantity from a source population to a target population. Though selection diagrams are useful for identifying s-admissible sets, in practice, they may not be able to isolate the $MSTS$ for two key reasons.

Transportability of causal contrasts

The transportability framework gives the assumptions and criteria for transporting the full counterfactual distribution of outcomes $P(Y^Z)$ from the source population to the target population. However, in many applications, researchers may only be interested in transporting a particular causal quantity (e.g. a causal contrast or mean outcome value). If the causal quantity of interest (Φ) is a function of $P(Y^Z)$, then any set of variables that is s-admissible for transporting $P(Y^Z)$ would also be s-admissible for transporting Φ. However, there may be some variables that are necessary to transport $P(Y^Z)$ that would be unnecessary for transporting Φ. For example, according to the selection diagram given in **Figure 4.2**, the s-admissible set to transport $P(Y^Z)$ includes both B and G. This is also apparent from the structural equations in **Figure 4.1**: $P(Y = 1)$ depends on both B and G. However, suppose we are only interested in

transporting the causal risk difference between those assigned to the intervention arm and those assigned to the treatment arm:

$$\Phi = P(Y^{Z=1} = 1) - P(Y^{Z=0} = 1) = P(Y = 1|Z = 1) - P(Y = 1|Z = 0)$$

From the structural equations in **Figure 4.1**, we see that this quantity only depends on B:

$$P(Y = 1|Z = 1) - P(Y = 1|Z = 0) = -.4B - .001(1 - B)$$

We can re-draw the selection diagram to reflect that we only want to transport this risk difference (**Figure 4.3**) and to indicate that the risk difference does not depend on G; only B is required to d-separate the risk difference from the selection nodes.

The transport formula for transporting $P(Y^Z)$ from the source to the target population using the transport set $\{B, G\}$ is:

$$P(Y^Z|S = 0) = \sum_g \sum_b P(Y = 1|Z, G, B, S = 1)P(B, G|S = 0)$$

And the transport formula for transporting Φ using the transport set $\{B\}$ is:

$$\begin{aligned}\Phi = P(Y^{Z=1} = 1\,|S = 0) - P(Y^{Z=0} = 1|S = 0) \\ = \sum_b P(Y = 1|Z = 1, B, S = 1)P(B|S = 0) \\ - \sum_b P(Y = 1|Z = 0, B, S = 1)P(B|S = 0)\end{aligned}$$

Figure. 4.1

$$S = \begin{cases} 1 & \text{if in FINGERS study population,} \\ 0 & \text{if in US target population} \end{cases}$$

$$P(G = 1) = \begin{cases} .4 & \text{if } S = 0, \\ .16 & \text{if } S = 1 \end{cases}$$

$$P(B = 1) = \begin{cases} .3 & \text{if } S = 0, \\ .55 & \text{if } S = 1 \end{cases}$$

$$P(Z = 1) = \begin{cases} .5 & \text{if } S = 0, \\ \text{Undefined} & \text{if } S = 1 \end{cases}$$

$$P(Y = 1) = -0.4BZ - 0.001(1 - B)Z + .1B + .2G + .6$$

$$P(Y = 1|Z = 1) - P(Y = 1|Z = 0) = -0.4B - 0.001(1 - B)$$

Figure 4.2

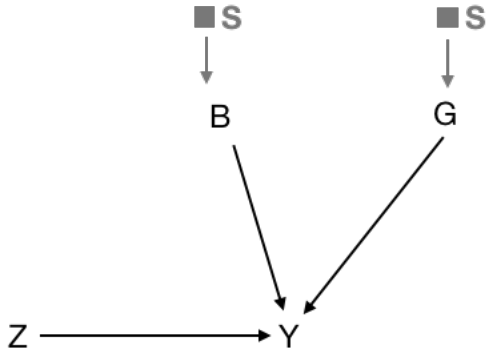
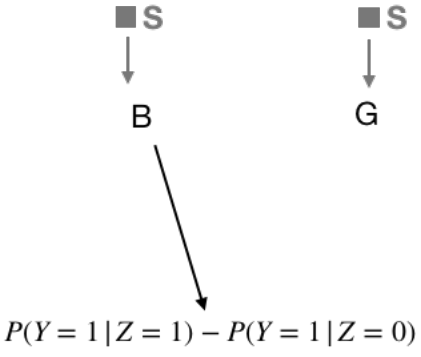


Figure 4.3



Figures 4.1- 4.3 Structural causal model, standard selection diagram, and modified selection diagram illustrating that fewer variables might be needed to transport a causal contrast than for transporting the full counterfactual outcome distribution.

Table 4.1 shows the results of using each transport set to transport the mean outcome in each arm; the risk difference between arms; and the risk ratio between arms. We see that **Eq. 4,** which includes both causes of the outcome, allows us to accurately transport all 3 quantities, but **Eq. 5** is sufficient to transport our causal quantity of interest (the risk difference).

Table 4.1 shows the transported risk difference and risk ratio adjusting for APOE-$\epsilon 4$ (G) and systolic blood pressure (B) or systolic blood pressure alone. If the target parameter is the risk difference, we see that adjusting for systolic blood pressure alone is sufficient.

		P(Y=1)	Risk Difference	Risk Ratio
Truth	$\boldsymbol{P(Y=1\|Z=0,S=0)}$	0.680	-0.121	0.823
	$\boldsymbol{P(Y=1\|Z=1,S=0)}$	0.559		
Transported using $\{B, G\}$	**Transported** $\boldsymbol{P(Y=1\|Z=0,S=0)}$	0.680	-0.121	0.823
	Transported $\boldsymbol{P(Y=1\|Z=1,S=0)}$	0.559		
Transported using $\{B\}$	**Transported** $\boldsymbol{P(Y=1\|Z=0,S=0)}$	0.632	-0.121	0.809
	Transported $\boldsymbol{P(Y=1\|Z=1,S=0)}$	0.511		

This toy example illustrates that transporting a specific causal quantity may require fewer variables than would be necessary for transporting the full counterfactual distribution. In practice—when the true data-generating model is unknown—knowing which variables from the s-admissible set for transporting $P(Y^Z)$ are unnecessary for transporting Φ requires parametric assumptions on the outcome-generating function that may be difficult to justify. As a result, researchers may reasonably choose to use the s-admissible set for $P(Y^Z)$ to avoid making these types of parametric assumptions even if doing so increases the chance that unnecessary variables are included in the transport estimators.

Uncertainty in causal diagrams

Even if a researcher intends to transport the entire counterfactual outcome distribution, uncertainty about the data generating processes in each population will lead to including extraneous variables in transport estimators. For selection diagrams to be most effective, they need to adequately and honestly reflect our prior knowledge and assumptions about the causal

mechanisms underlying the observed (and unobserved) data. As with all causal graphs, excluding edges or selection nodes from a selection diagram is a stronger assumption than including them.[86] That is, the most conservative graph would include edges between all random variables and would include selection nodes directed at every random variable. This graph would indicate that we think there is a *possibility* that each variable may cause the others (or be associated with through a common ancestor) and there may be differences anywhere in the data-generating model between the source and target populations.

In most (if not all) applied settings, there is often considerable uncertainty about the true data-generating model. Further, we often have little knowledge about how two populations might differ from one another. As a result, selection diagrams that honestly capture this uncertainty are likely to include more selection nodes or edges than are present in truth. The s-admissible sets based on these graphs will therefore likely include many more variables than necessary.

Overall, selection diagrams are an important tool to guide variable selection, but in most applications, selection diagrams may not be able to reveal a minimally sufficient transport set and transport estimators based on selection diagrams are likely to include additional unnecessary variables. How these extraneous variables might affect the performance of transport estimators is unclear.

SIMULATION EXPERIMENT

We conducted a Monte Carlo simulation experiment to examine the variable selection problem in transport estimators. Specifically, we explored how the inclusion of 5 different types of *unnecessary* variables (in addition to the MSTS) affect the bias, mean square error, and confidence interval coverage of the parametric g-formula transport estimator.[38] We limit our experiment to only consider variables that are not on the causal path from the exposure to the outcome (ie. pre-treatment variables).

Classification of transport variables according to causal structure

Figure 4.4 shows the 5 different variable types that might be unnecessarily included in an s-admissible set ($\boldsymbol{TS_s}$). In this example, all variable types are subsets of the s-admissible set and are (by definition) not part of the $MSTS$; all types are mutually exclusive. The unnecessary variables are categorized according to their relationships to the selection nodes, outcome variable, and causal quantity conditional on a specific $MSTS$. Note that if a variable is not a cause of the outcome, it also cannot be a cause of Φ. Similarly, if a variable is a cause of Φ it must also be a cause of the outcome.

1. $\boldsymbol{W_a}$**:** differ in distribution between the source and target populations; cause the outcome; but do not affect Φ.
2. $\boldsymbol{W_b}$: differ in distribution between the source and target populations and are not causes of the outcome.
3. $\boldsymbol{W_c}$: do not differ between the source and target populations and cause Φ.
4. $\boldsymbol{W_d}$: do not differ between the source and target populations; cause the outcome; but do not affect Φ.
5. $\boldsymbol{W_e}$: do not differ between the source and target populations and do not cause the outcome.

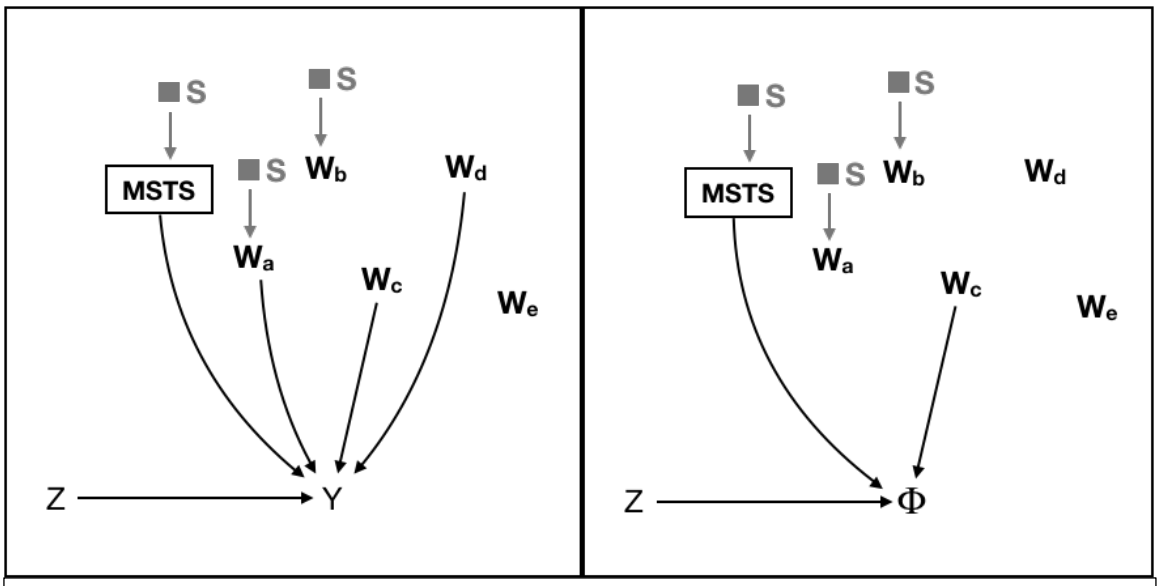


Figure 4.4. Selection diagrams showing the 5 categories of unnecessary variables that may be included in an s-admissible set. $\boldsymbol{MSTS}$ is the minimally sufficient transport set. After conditioning on this set, $\boldsymbol{W_a, W_b, W_c, W_d}$, and $\boldsymbol{W_e}$ are all unnecessary to transport Φ across the populations.

Simulation Experiments

We generated data according to the following data-generating processes. The magnitude and likelihood of practical positivity violations was highest in data-generating model 3.

$$S \sim Ber(0.5)$$

$$Z \sim Ber(0.5)$$

$$MSTS, W_a, W_b \sim N(1 + 3S, sd_m)$$

$$W_c, W_d \sim N(1, 1)$$

$$W_e \sim N(0, 1)$$

$$Y \sim N(100 + 20Z + 10(MSTS)Z + 10(W_a) + 10(W_c)Z + 10(W_d), 5)$$

Where for each data-generating model $M = m$:

$$sd_m = \begin{cases} 1 + 5S, & m = 1 \\ 1 + 3S, & m = 2 \\ 1 + S, & m = 3 \end{cases}$$

We aim to transport the causal quantity $\Phi = E(Y^{Z=1}) - E(Y^{Z=0})$ from the study population $(S = 1)$ to the target population $(S = 0)$. In all 3 data-generating models, the true value of Φ in the target population is 40 and the true value of Φ in the study population is 70.

For each data-generating model, we simulated 5000 datasets with a total N=5000 (with approximately 50% in $S = 1$ and 50% in $S = 0$). For each dataset, we fit a parametric g-computation transport estimator[38] for each of the following transport adjustment sets:

Table 4.2. List of transport adjustment sets used with the parametric g-formula transport estimator for each simulation. TS_7 includes any variables that differ between the two populations; TS_8 includes all causes of Y; TS_9 includes the $MSTS$ and all causes of Y that don't differ between the two populations; TS_{10} includes the full set of variables; and TS_{11} does not meet the s-admissibility criterion and serves as a negative control.

Transport Adjustment Set (TS_i)
$TS_1 = \{\text{MSTS}\}$
$TS_2 = \{\text{MSTS}, W_a\}$
$TS_3 = \{\text{MSTS}, W_b\}$
$TS_4 = \{\text{MSTS}, W_c\}$
$TS_5 = \{\text{MSTS}, W_d\}$
$TS_6 = \{\text{MSTS}, W_e\}$
$TS_7 = \{\text{MSTS}, W_a, W_b\}$
$TS_8 = \{\text{MSTS}, W_a, W_c, W_d\}$
$TS_9 = \{\text{MSTS}, W_c, W_d\}$
$TS_{10} = \{\text{MSTS}, W_a, W_b, W_c, W_d, W_e\}$
$TS_{11} = \{W_c\}$

To fit the g-formula transport estimators, we first fit a conditional linear regression in the source population ($S = 1$) regressing Y on Z and all variables in the transport set including all possible interaction terms. We then used this model to predict the values of Y in the target population setting $Z = 1$ and $Z = 0$ and calculated the difference in mean outcomes under each treatment assignment.[38] We used a non-parametric bootstrap with 1000 bootstrap samples to estimate the standard error.[87]

We report the estimated bias, variance, mean square error (MSE), and confidence interval coverage for each transport set. All analyses were conducted using R version 3.5.2.[88]

RESULTS

Across all 3 data-generating models, all transport sets that included the $MSTS$ (and therefore met the s-admissibility criterion) were unbiased (**Table 4.3**). However, using the $MSTS$ alone was not the optimal transport set in terms of MSE; TS_8 had the lowest MSE across all data-generating models. Among the s-admissible sets (all except TS_{11}), TS_3 had the highest MSE in each of the 3 data-generating models.

Excluding variables that were causes of the outcome but did not modify the causal quantity of interest (W_a and W_d) did not negatively affect the bias of the estimators, and including unnecessary variables that were not causes of the outcome and that did not differ between the populations (W_e) did not increase the MSE compared to the MSTS alone.

Because of the smaller standard deviations for $MSTS, W_a,$ and W_b in data-generating model 3, this model was most likely to produce practical positivity violations. However, the parametric models in the estimators were correctly specified and could therefore accurately extrapolate beyond the bounds of the source data, so these practical positivity violations did not induce bias in the transport estimators. Additionally, because the standard errors were smaller in the source population in this data-generating model compared to models 1 and 2, the estimates were

generally more precise. However, the pattern of relative performance between the transport sets differed in this data generating model. TS_8, which included all causes of Y but no other unnecessary variables, performed substantially better than the other transport sets, while TS_3 and TS_7 had markedly higher MSEs.

DISCUSSION

The impact of including unnecessary variables in the parametric g-formula transport estimator varies depending on the type of extraneous variable that is included. As expected, excluding variables that cause the outcome but don't modify the causal quantity of interest does not increase bias. However, including all causes of the outcome—regardless of whether the variable modifies the causal quantity of interest or varies between the populations—improves the MSE by reducing the standard error. Additionally, including variables that differ between the populations but don't cause the outcome tends to increase the MSE.

There are several practical implications uncovered by this study. When faced with a variable selection problem for transport, it's best to focus on including as many causes of the outcome as possible. This is perhaps counterintuitive. Obvious differences between source and target populations are often the impetus for applying transportability methods in the first place, and these types of differences are often easier to detect. However, the strategy of including all variables that differ between two populations increases the chance that some variables that are not causes of the outcome will be included in the transport set, and the inclusion of these variables will likely increase the MSE of the estimators.

Because we intended to highlight the impact of including different types of extraneous variables for the most common types of transport questions researchers are likely to face, we restricted our experiment to only include data-generating models with selection nodes on pre-treatment variables, and we only evaluated the parametric g-formula transport estimator. Other

commonly used transport estimators include the inverse odds of selection weights and doubly robust targeted maximum likelihood transport estimators. We expect to find similar patterns across the different estimators, but future work should explore variable selection under other data-generating conditions and with other estimation approaches.

Our simulation experiment also only included correctly specified parametric models in the g-formula transport estimators. As a result, the estimates were unbiased in spite of the practical positivity violations in data-generating model 3. If the models used in the estimators are not correctly specified, there is no guarantee that the estimates would be unbiased.

Based on our findings, a potential practical approach to variable selection for transportability would be to use the study data alone to determine what variables should be measured in target populations to transport the results. For example, after completion of a trial, researchers could conduct a careful analysis to identify all the characteristics that modified the effect of interest. Researchers looking to transport the trial's results to a specific target population would then know what characteristics they would need to measure to do so. So long as the study measured all effect modifiers, this approach would ensure that the s-admissibility criteria is met and that any unnecessary variables included transport set (because they don't differ between the populations) would improve the precision of the estimates. Of course, trial results can only be transported if they enroll populations that are heterogeneous with respect to the effect modifiers and if all those effect modifiers are measured. Future work should explore data-driven approaches for identifying optimal transport sets to further improve the accuracy and precision of transport estimators.

Table 4.3 Simulation results

	Data-Generating Model 1					Data-Generating Model 2					Data-Generating Model 3				
Transport Set	Mean	Std. Err	Bias	MSE	CI cover	Mean	Std. Err	Bias	MSE	CI cover	Mean	Std. Err	Bias	MSE	CI cover
1	39.95	7.59	-0.05	7.5	0.95	39.96	4.48	-0.04	4.43	0.95	40	3.03	0	3.15	0.94
2	39.94	0.48	-0.06	0.5	0.94	39.99	0.73	-0.01	0.73	0.95	39.98	3.03	-0.02	2.95	0.95
3	39.9	9.51	-0.1	8.9	0.96	39.94	7.01	-0.06	6.93	0.95	39.97	9.83	-0.03	10.02	0.95
4	39.95	7.55	-0.05	7.34	0.95	39.95	4.41	-0.05	4.38	0.95	40.01	2.82	0.01	2.91	0.95
5	39.99	7.42	-0.01	7.33	0.95	39.96	4.25	-0.04	4.16	0.95	39.99	2.52	-0.01	2.59	0.95
6	39.93	7.6	-0.07	7.55	0.95	39.96	4.5	-0.04	4.44	0.95	40	3.04	0	3.16	0.95
7	39.95	0.59	-0.05	0.65	0.93	39.98	1.12	-0.02	1.14	0.95	39.97	10.02	-0.03	9.88	0.95
8	39.98	0.14	-0.02	0.14	0.95	39.99	0.18	-0.01	0.18	0.95	40	0.52	0	0.52	0.95
9	40	7.43	0	7.19	0.96	39.99	4.2	-0.01	4.17	0.95	40.01	2.32	0.01	2.39	0.94
10	39.98	0.17	-0.02	0.16	0.95	40	0.26	0	0.25	0.95	39.99	1.93	-0.01	1.8	0.96
11	69.94	8.89	29.94	905.46	0	69.96	4.1	29.96	901.84	0	70	1.2	30	901.22	0

CONCLUSIONS

The three studies described within this dissertation demonstrate how causal transportability can improve how results of clinical trials are used to inform implementation of new interventions. The first study found that population compositional differences between transgender women and cisgender men in iPrEx were sufficient to explain differences in the effectiveness of randomization to the active arm of the trial. The second study demonstrated how to generate target-specific guidance about how best to implement new interventions by transporting subgroup analyses of randomized trials to each target population. The third study considered how best to select variables for transport estimators to maximize the performance of the parametric g-computation transport estimator.

Overall, causal transportability theory provides a rigorous solution to a wide range of previously intractable problems surrounding external validity of studies. However, there are important challenges in implementing these methods. First, transportability requires individual-level measurements of the transport adjustment set in both the study and target populations. This means that trials need to consider what characteristics are likely to impact the effectiveness of the intervention under study prior to collecting data. Additionally, it may require that these characteristics are measured in a representative sample of each target population. Another challenge in implementing transportability is that as time passes, it may become less possible to account for all the relevant differences between populations. This means that transporting results of a trial should occur as quickly as possible to maximize the chance that the s-admissibility criteria have been met. Overall, for transportability to be most useful, researchers need to plan on implementing these tools early in the design of the study so that they can be used as quickly as possible.

In the pursuit of maximizing external validity, many researchers have championed pragmatic trials as a means of evaluating the effectiveness of implementation strategies in usual

care settings.[89] However, in the face of heterogeneous effects of implementations, the quest for a single effect that applies universally is hopeless even if the trial design conforms to all the established features of a pragmatic trial (e.g., no recruitment restrictions, flexible interventions). So, unless the trial population is a random sample of all future target populations (an implausible concept given that even in the same location, the data generating process may evolve over time), even results of pragmatic trials will need to be formally transported to produce evidence that is relevant for different settings (or the same setting in the future). However, to avoid altering standard care practices, pragmatic trials often minimize the number of measurements taken over the course of the study, but formal transport requires individual-level measurements of variables that modify the effectiveness of the implementation strategy. By minimizing the number of characteristics measured, pragmatic trials preclude the ability to transport their results to external settings and are undermining their own goals of generating study results that could be applied to a range of target populations. Instead, if the objective of pragmatic trials is to produce more generalizable knowledge, it is essential that these studies understand and measure the mechanisms and modifiers of the implementation strategies being evaluated.

Overall, this dissertation demonstrates a few areas in which transportability can greatly improve implementation of new interventions. There remains much work to be done in this area, and future researchers should consider how the transportability framework might affect their approach to designing and analyzing randomized trials.

REFERENCES

1. Pearl J, Bareinboim E. *Transportability across Studies: A Formal Approach*. DTIC Document; 2011. http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA557437. Accessed August 11, 2015.

2. Grant RM, Lama JR, Anderson PL, et al. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. *N Engl J Med*. 2010;363(27):2587-2599. doi:10.1056/NEJMoa1011205

3. Baeten JM, Donnell D, Ndase P, et al. Antiretroviral Prophylaxis for HIV-1 Prevention among Heterosexual Men and Women. *N Engl J Med*. 2012;367(5):399-410. doi:10.1056/NEJMoa1108524

4. Thigpen MC, Kebaabetswe PM, Paxton LA, et al. Antiretroviral Preexposure Prophylaxis for Heterosexual HIV Transmission in Botswana. *N Engl J Med*. 2012;367(5):423-434. doi:10.1056/NEJMoa1110711

5. Commissioner O of the. Press Announcements - FDA approves first drug for reducing the risk of sexually acquired HIV infection. http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm312210.htm. Accessed August 17, 2015.

6. Office of National AIDS Policy. *National HIV/AIDS Strategy for the United States: Updated to 2020*. ONAP; 2015. https://www.whitehouse.gov/sites/default/files/docs/national_hiv_aids_strategy_update_2020.pdf. Accessed October 9, 2015.

7. World Health Organization, Department of HIV/AIDS, World Health Organization. *Guideline on When to Start Antiretroviral Therapy and on Pre-Exposure Prophylaxis for HIV.*; 2015. http://apps.who.int/iris/bitstream/10665/186275/1/9789241509565_eng.pdf. Accessed October 9, 2015.

8. Cohen MS, Chen YQ, McCauley M, et al. Prevention of HIV-1 Infection with Early Antiretroviral Therapy. *N Engl J Med*. 2011;365(6):493-505. doi:10.1056/NEJMoa1105243

9. Cohen MS, McCauley M, Gamble TR. HIV treatment as prevention and HPTN 052. *Curr Opin HIV AIDS*. 2012;7(2):99-105. doi:10.1097/COH.0b013e32834f5cf2

10. Can San Francisco Become the First AIDS-Free City? | Advocate.com. http://www.advocate.com/health/hiv-aids/2014/03/03/can-san-francisco-become-first-aids-free-city. Accessed October 9, 2015.

11. Park A. The End of AIDS. *Time*. November 2014. http://time.com/3596979/the-end-of-aids/. Accessed October 9, 2015.

12. The end of AIDS is within our grasp. San Francisco AIDS Foundation. http://sfaf.org/hiv-info/hot-topics/from-the-experts/end-of-aids-is-within-our-grasp.html. Accessed October 9, 2015.

13. Department of Health and Human Services, National Institutes of Health, Office of AIDS Research. *FY 2016: Trans-NIH Plan for HIV-Related Research*. http://www.oar.nih.gov/strategicplan/fy2016/pdf/FY-2016-Trans-NIH-Plan-for-HIV-Related-Research.pdf. Accessed November 15, 2015.

14. National Institutes of Health, Office of AIDS Reserach. NOT-OD-15-137: NIH HIV/AIDS Reserach Priorities and Guidelines for Determining AIDS Funding. August 2015. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-15-137.html. Accessed November 15, 2015.

15. U.S. Department of Health and Human Services, National Institutes of Health, National Institute of Mental Health. *NIMH Strategic Plan for Research (NIH Publication No. 02-2650).* National Institute of Mental Health; 2015. http://www.nimh.nih.gov/about/strategic-planning-reports/index.shtml. Accessed November 15, 2015.

16. U. S. Department of Health and Human Services, National Institutes of Health, President's Emergency Plan for AIDS Relief. NIH-PEPFAR Collaboration on Implementation Science for HIV: Towards an AIDS-free Generation (R01). 2015. http://grants.nih.gov/grants/guide/rfa-files/RFA-AI-15-020.html. Accessed November 15, 2015.

17. Department of State, Office of the Global AIDS Coordinator. *PEPFAR 2015 Annual Report to Congress*. Office of the Global AIDS Coordinator; 2015. http://www.pepfar.gov/documents/organization/239006.pdf. Accessed November 15, 2015.

18. Marrazzo JM, Ramjee G, Richardson BA, et al. Tenofovir-Based Preexposure Prophylaxis for HIV Infection among African Women. *N Engl J Med*. 2015;372(6):509-518. doi:10.1056/NEJMoa1402269

19. Mutua G, Sanders E, Mugo P, et al. Safety and Adherence to Intermittent Pre-Exposure Prophylaxis (PrEP) for HIV-1 in African Men Who Have Sex with Men and Female Sex Workers. *PLoS ONE*. 2012;7(4). doi:10.1371/journal.pone.0033103

20. Van Damme L, Corneli A, Ahmed K, et al. Preexposure Prophylaxis for HIV Infection among African Women. *N Engl J Med*. 2012;367(5):411-422. doi:10.1056/NEJMoa1202614

21. Choopanya K, Martin M, Suntharasamai P, et al. 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. *The Lancet*. 2013;381(9883):2083-2090. doi:10.1016/S0140-6736(13)61127-7

22. Grant RM, Anderson PL, McMahan V, et al. Uptake of pre-exposure prophylaxis, sexual practices, and HIV incidence in men and transgender women who have sex with men: a cohort study. *Lancet Infect Dis*. 2014;14(9):820-829. doi:10.1016/S1473-3099(14)70847-3

23. Deutsch MB, Glidden DV, Sevelius J, et al. HIV pre-exposure prophylaxis in transgender women: a subgroup analysis of the iPrEx trial. *Lancet HIV*. November 2015. doi:10.1016/S2352-3018(15)00206-4

24. Rudolph KE, van der Laan MJ. Double Robust Estimation of Encouragement-design Intervention Effects Transported Across Sites. 2015. http://biostats.bepress.com/ucbbiostat/paper335/. Accessed August 11, 2015.

25. UNAIDS. *Practical Guidelines for Intensifying HIV Prevention: Towards Universal Access*. Joint United Nations Programme on HIV/AIDS (UNAIDS); 2007. http://data.unaids.org/pub/Manual/2007/20070306_Prevention_Guidelines_Towards_Universal_Access_en.pdf.

26. Wilson D, Halperin DT. "Know your epidemic, know your response": a useful approach, if we get it right. *The Lancet*. 2008;372(9637):423-426. doi:10.1016/S0140-6736(08)60883-1

27. Baeten J, Grant R. Use of Antiretrovirals for HIV Prevention: What Do We Know and What Don't We Know? *Curr HIV/AIDS Rep*. 2013;10(2):142-151. doi:10.1007/s11904-013-0157-9

28. Volk JE, Marcus JL, Phengrasamy T, et al. No New HIV Infections with Increasing Use of HIV Preexposure Prophylaxis in a Clinical Practice Setting. *Clin Infect Dis*. September 2015:civ778. doi:10.1093/cid/civ778

29. Fonner VA, Dalglish SL, Kennedy CE, et al. Effectiveness and safety of oral HIV preexposure prophylaxis for all populations. *AIDS Lond Engl*. 2016;30(12):1973-1983. doi:10.1097/QAD.0000000000001145

30. Baral SD, Poteat T, Strömdahl S, Wirtz AL, Guadamuz TE, Beyrer C. Worldwide burden of HIV in transgender women: a systematic review and meta-analysis. *Lancet Infect Dis*. 2013;13(3):214-222. doi:10.1016/S1473-3099(12)70315-8

31. Liu AY, Turner C, Araysirikul S, et al. Substantial Gaps in the PrEP continuum Among Transwomen Compared with MSM in San Francisco. Oral Abstract Presentation presented at the: HIV Research for Prevention (HIVR4P); October 2018; Madrid, Spain.

32. Westreich D, Edwards JK, Lesko CR, Cole SR, Stuart EA. Target Validity and the Hierarchy of Study Designs. October 2018:21.

33. Hiransuthikul A, Himmad K, Kerr S, et al. Transgender Hormonal Therapy Might Interfere with PrEP. In: *AIDS 2018*. Amsterdam, Netherlands: IAS; 2018. https://www.medpagetoday.com/meetingcoverage/iac/74222. Accessed July 27, 2018.

34. Shieh E, Marzinke M, Fuchs E, et al. Transgender Women on Estrogen Have Significantly Lower Tenofovir/Emtracitabine Concentrations During Directly Observed Dosing When Compared to Cis Men. In: *HIV Research For Prevention 2018*. Madrid, Spain: HIVR4P; 2018.

35. Rudolph KE, Schmidt NM, Glymour MM, et al. Composition or Context: Using Transportability to Understand Drivers of Site Differences in a Large-scale Housing Experiment. *Epidemiol Camb Mass*. 2018;29(2):199-206. doi:10.1097/EDE.0000000000000774

36. Mehrotra ML. The debigulator: a variable selection algorithm for transportability. Poster presented at the: Society for Epidemiologic Research Annual Meeting; June 2018; Baltimore, Maryland.

37. Robins J. A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Math Model*. 1986;7(9):1393-1512. doi:10.1016/0270-0255(86)90088-6

38. Lesko CR, Buchanan AL, Westreich D, Edwards JK, Hudgens MG, Cole SR. Generalizing study results: a potential outcomes perspective. *Epidemiol Camb Mass*. 2017;28(4):553-561. doi:10.1097/EDE.0000000000000664

39. Rosenbaum PR. Model-Based Direct Adjustment. *J Am Stat Assoc*. 1987;82(398):387-394. doi:10.1080/01621459.1987.10478441

40. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2017. https://www.R-project.org/.

41. StataCorp. *Stata Statistical Software: Release 15*. College Station, TX: StataCorp LLC; 2017.

42. Nemoto T, Bödeker B, Iwamoto M. Social Support, Exposure to Violence and Transphobia, and Correlates of Depression Among Male-to-Female Transgender Women With a History of Sex Work. *Am J Public Health*. 2011;101(10):1980-1988. doi:10.2105/AJPH.2010.197285

43. Lombardi EL, Wilchins RA, Priesing D, Malouf D. Gender violence: transgender experiences with violence and discrimination. *J Homosex*. 2001;42(1):89-101.

44. Mehrotra ML, Amico KR, McMahan V, et al. The Role of Social Relationships in PrEP Uptake and Use Among Transgender Women and Men Who Have Sex with Men. *AIDS Behav*. May 2018:1-8. doi:10.1007/s10461-018-2151-0

45. Poteat T, Reisner SL, Radix A. HIV Epidemics among Transgender Women. *Curr Opin HIV AIDS*. 2014;9(2):168-173. doi:10.1097/COH.0000000000000030

46. Poteat T, Wirtz AL, Radix A, et al. HIV risk and preventive interventions in transgender women sex workers. *The Lancet*. 2015;385(9964):274-286. doi:10.1016/S0140-6736(14)60833-3

47. Balzer LB. "All Generalizations Are Dangerous, Even This One."-Alexandre Dumas. *Epidemiol Camb Mass*. 2017;28(4):562-566. doi:10.1097/EDE.0000000000000665

48. Dahabreh IJ, Robertson SE, Stuart EA, Hernan MA. Transporting inferences from a randomized trial to a new target population. *ArXiv180500550 Stat*. May 2018. http://arxiv.org/abs/1805.00550.

49. Sevelius JM, Keatley J, Calma N, Arnold E. 'I am not a man': Trans-specific barriers and facilitators to PrEP acceptability among transgender women. *Glob Public Health*. 2016;11(7-8):1060-1075. doi:10.1080/17441692.2016.1154085

50. Poteat T, German D, Flynn C. The conflation of gender and sex: Gaps and opportunities in HIV data among transgender women and MSM. *Glob Public Health*. 2016;0(0):1-14. doi:10.1080/17441692.2015.1134615

51. Rothwell PM. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *The Lancet*. 2005;365(9454):176–186.

52. VanderWeele TJ, Knol MJ. On the interpretation of subgroup analyses in randomized trials: heterogeneity versus secondary interventions. *Ann Intern Med*. 2011;154(10). doi:10.7326/0003-4819-154-10-201105170-00008

53. Laupacis A, Sackett DL, Roberts RS. An Assessment of Clinically Useful Measures of the Consequences of Treatment. *N Engl J Med*. 1988;318(26):1728-1733. doi:10.1056/NEJM198806303182605

54. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *BMJ*. 1995;310(6977):452-454.

55. Buchbinder SP, Glidden DV, Liu AY, et al. HIV pre-exposure prophylaxis in men who have sex with men and transgender women: a secondary analysis of a phase 3 randomised controlled efficacy trial. *Lancet Infect Dis*. 2014;14(6):468-475. doi:10.1016/S1473-3099(14)70025-8

56. Hull M, Tan D. Setting the stage for expanding HIV pre-exposure prophylaxis use in Canada. *Can Commun Dis Rep*. 2017;43(12):272-278.

57. Hankins C, Macklin R, Warren M. Translating PrEP effectiveness into public health impact: key considerations for decision-makers on cost-effectiveness, price, regulatory issues, distributive justice and advocacy for access. *J Int AIDS Soc*. 2015;18(4 (Suppl 3)). doi:10.7448/IAS.18.4.19973

58. Luz PM, Osher B, Grinsztejn B, et al. The cost-effectiveness of HIV pre-exposure prophylaxis in men who have sex with men and transgender women at high risk of HIV infection in Brazil. *J Int AIDS Soc*. 2018;21(3). doi:10.1002/jia2.25096

59. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials*. 2015;16. doi:10.1186/s13063-015-1023-4

60. Gonzalez LE, Sutton SK, Pratt C, Gilbertson M, Antonia S, Quinn GP. The Bottleneck Effect in Lung Cancer Clinical Trials. *J Cancer Educ Off J Am Assoc Cancer Educ*. 2013;28(3):488-493. doi:10.1007/s13187-013-0491-z

61. Susukida R, Crum RM, Stuart EA, Ebnesajjad C, Mojtabai R. Assessing sample representativeness in randomized controlled trials: application to the National Institute of Drug Abuse Clinical Trials Network. *Addict Abingdon Engl*. 2016;111(7):1226-1234. doi:10.1111/add.13327

62. Eisenberg Y, Mohiuddin H, Cherukupally K, Zaidi H, Kukreja S, Barengolts E. Similarities and differences between patients included and excluded from a randomized clinical trial of vitamin d supplementation for improving glucose tolerance in prediabetes: interpreting broader applicability. *Trials*. 2015;16:306. doi:10.1186/s13063-015-0812-0

63. Isaacs T, Hunt D, Ward D, Rooshenas L, Edwards L. The Inclusion of Ethnic Minority Patients and the Role of Language in Telehealth Trials for Type 2 Diabetes: A Systematic Review. *J Med Internet Res*. 2016;18(9). doi:10.2196/jmir.6374

64. Curno MJ, Rossi S, Hodges-Mameletzis I, Johnston R, Price MA, Heidari S. A systematic review of the inclusion (or exclusion) of women in HIV research: from clinical studies of antiretrovirals and vaccines to cure strategies. *JAIDS J Acquir Immune Defic Syndr*. 2016;71(2):181–188.

65. Pearl J, Bareinboim E. External Validity: From Do-Calculus to Transportability Across Populations. *Stat Sci*. 2014;29(4):579-595. doi:10.1214/14-STS486

66. Cole SR, Stuart EA. Generalizing Evidence From Randomized Clinical Trials to Target Populations: The ACTG 320 Trial. *Am J Epidemiol*. 2010;172(1):107-115. doi:10.1093/aje/kwq084

67. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. Lippincott Williams & Wilkins; 2008. http://books.google.com/books?hl=en&lr=&id=Z3vjT9ALxHUC&oi=fnd&pg=PR7&dq=%22led+eventually+to+the+landmark+report,+Smoking+and+Health,+issued+by+the+Surgeon+General%22+%22replacement+therapy+and+heart%22+%22the+interpretation+of+data.+In+1978,+a+controversy+erupted+about+whether+exogenous+estrogens%22+&ots=aQJFfMRI4U&sig=3U-7pQT3-y7QvPXGUjv_h6wyl_o. Accessed September 21, 2015.

68. Cole SR, Stuart EA. Generalizing Evidence From Randomized Clinical Trials to Target Populations The ACTG 320 Trial. *Am J Epidemiol*. 2010;172(1):107-115. doi:10.1093/aje/kwq084

69. Lesko CR, Buchanan AL, Westreich D, Edwards JK, Hudgens MG, Cole SR. Generalizing study results: a potential outcomes perspective. *Epidemiology*. doi:10.1097/EDE.0000000000000664

70. Rudolph KE, Díaz I, Rosenblum M, Stuart EA. Estimating Population Treatment Effects From a Survey Subsample. *Am J Epidemiol*. September 2014:kwu197. doi:10.1093/aje/kwu197

71. Mehrotra ML, Westreich D, McMahan V, et al. Baseline characteristics explain differences in effectiveness of randomization to daily oral TDF/FTC PrEP between transgender women and cisgender men who have sex with men in the iPrEx trial. *JAIDS J Acquir Immune Defic Syndr*. In Press.

72. Ramirez-Valles J, Garcia D, Campbell RT, Diaz RM, Heckathorn DD. HIV Infection, Sexual Risk Behavior, and Substance Use Among Latino Gay and Bisexual Men and Transgender Persons. *Am J Public Health*. 2008;98(6):1036-1042. doi:10.2105/AJPH.2006.102624

73. Ramirez-Valles J. Latino MSM Community Involvement: HIV Protective Effects. 2014. http://doi.org/10.3886/ICPSR34385.v2.

74. Petersen ML, Porter KE, Gruber S, Wang Y, van der Laan MJ. Diagnosing and responding to violations in the positivity assumption. *Stat Methods Med Res*. 2012;21(1):31-54. doi:10.1177/0962280210386207

75. Robins J. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. *J Chronic Dis*. 1987;40:139S-161S. doi:10.1016/S0021-9681(87)80018-8

76. Pearl J. *Causality: Models, Reasoning, and Inference*. New York, NY, USA: Cambridge University Press; 2000.

77. Petersen ML. Compound Treatments, Transportability, and the Structural Causal Model: The Power and Simplicity of Causal Graphs. *Epidemiology*. 2011;22(3):378-381. doi:10.1097/EDE.0b013e3182126127

78. Pearl J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 1988.

79. Westreich D, Edwards JK, Lesko CR, Stuart E, Cole SR. Transportability of trial results using inverse odds of sampling weights. *Am J Epidemiol*. May 2017. doi:10.1093/aje/kwx164

80. Efron B, Tibshirani R. *An Introduction to the Bootstrap*. New York: Chapman & Hall; 1993.

81. Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. *BMJ*. 1999;319(7223):1492-1495.

82. Mehrotra ML, Westreich D, McMahan VM, et al. Baseline Characteristics Explain Differences in Effectiveness of Randomization to Daily Oral TDF/FTC PrEP Between Transgender Women and Cisgender Men Who Have Sex With Men in the iPrEx Trial. *J Acquir Immune Defic Syndr 1999*. 2019;81(3):e94-e98. doi:10.1097/QAI.0000000000002037

83. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable Selection for Propensity Score Models. *Am J Epidemiol*. 2006;163(12):1149-1156. doi:10.1093/aje/kwj149

84. Ngandu T, Lehtisalo J, Solomon A, et al. A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. *The Lancet*. 2015;385(9984):2255-2263. doi:10.1016/S0140-6736(15)60461-5

85. Elias Bareinboim, Judea Pearl. Transportability of Causal Effects: Completeness Results. In: *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*. https://www.aaai.org/ocs/index.php/AAAI/AAAI12/paper/viewFile/5188/5259. Accessed March 13, 2018.

86. Pearl J. *Causality: Models, Reasoning, and Inference*. New York, NY, USA: Cambridge University Press; 2000.

87. Mooney CZ, Duval RD, Duval R. *Bootstrapping: A Nonparametric Approach to Statistical Inference*. SAGE; 1993.

88. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2018. https://www.R-project.org/.

89. Ware JH, Hamel MB. Pragmatic Trials — Guides to Better Patient Care? *N Engl J Med*. 2011;364(18):1685-1687. doi:10.1056/NEJMp1103502

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

Author Signature

09/05/12

Date