Title

Transporting Subgroup Analyses of Randomized Controlled Trials for Planning Implementation of New Interventions.

Journal

American Journal of Epidemiology, 190(8)

ISSN

0002-9262

Authors

Mehrotra, Megha L
Westreich, Daniel
Glymour, M Maria
et al.

Publication Date

2021-08-01

DOI

10.1093/aje/kwab045

Peer reviewed

American Journal of Epidemiology

Vol. 190, No. 8
https://doi.org/10.1093/aje/kwab045
Advance Access publication:
February 22, 2021

Practice of Epidemiology

Transporting Subgroup Analyses of Randomized Controlled Trials for Planning Implementation of New Interventions

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

* Correspondence to Dr. Megha L. Mehrotra, Department of Epidemiology and Biostatistics, University of California, San Francisco, 550 16th Street 2nd Floor #2660, San Francisco, CA 94158. (e-mail: megha.mehrotra@ucsf.edu).

Initially submitted April 2, 2019; accepted for publication February 18, 2021.

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; however, this approach has not yet been used for subgroup analyses of trials. Through an example from the Iniciativa Profilaxis Pre-Exposición (iPrEx) study (multiple countries, 2007–2010) of preexposure prophylaxis for human immunodeficiency virus, we illustrate how transporting subgroup analyses can produce target-specific subgroup effect estimates and numbers needed to treat. This approach could lead to more tailored and accurate guidance for resource allocation and cost-effectiveness analyses.

external validity; implementation; number needed to treat; subgroup analyses; target-specific subgroup analyses; transportability

Abbreviations: HIV, human immunodeficiency virus; iPrEx, Iniciativa Profilaxis Pre-Exposición; ITT, intention to treat; NNT, number needed to treat; PrEP, preexposure prophylaxis; RCT, randomized controlled trial; MSM, cisgender men who have sex with men; ncRAI, condomless receptive anal intercourse; TGW, transgender women who have sex with men.

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 (1, 2). Policy-makers then prioritize those groups with the lowest numbers needed to treat (NNTs) (3, 4)—that is, the number of individuals needed to be offered the intervention to prevent 1 incident outcome—to receive the intervention. For example, Iniciativa Profilaxis Pre-Exposición (iPrEx) (5) was a placebo-controlled RCT that evaluated the safety and effectiveness of combination daily oral tenofovir disoproxil fumarate/emtricitabine for human immunodeficiency virus (HIV) chemoprophylaxis (preexposure prophylaxis: PrEP) in transgender women and cisgender men who have sex with men (TGW and MSM). The study found that randomization to the active arm was associated with a 44% reduction in HIV incidence compared with the placebo arm (5). 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 (6). These results have subsequently informed policy recommendations and cost-effectiveness analyses of PrEP implementation (7–9).

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 (10), these differences also mean that the effect sizes estimated for a subgroup of the trial with a particular characteristic might be poor indicators of the expected effect sizes in target populations similar on that characteristic (11–16). Indeed, even if the overall trial population resembles, on average, a particular target population, differences might still exist within subgroups 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 whom 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 (17). 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 (10, 18, 19). 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 (17, 18, 20). When differences between populations are limited to pretreatment (baseline) covariates, transportability conceptually coincides with standardization across several characteristics (19).

To date, transportability has been applied to transport average treatment effects to new target populations (21–23) or to understand observed heterogeneity between sites (24) or groups (25) 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 (5) to illustrate how to apply transportability theory and estimators to transport subgroup effect estimates and NNTs to 2 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 2,499 MSM and transgender women in Brazil, Peru, Ecuador, United States, South Africa, and Thailand (2007–2010). All participants were HIV-negative at enrollment, reported risk behavior for HIV, and were assigned male sex at birth. The median age at enrollment was 25 years, and most participants had not received a college education (5). 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 2 clinics that serve young Latino TGW and MSM in San Francisco, California, and Chicago, Illinois (2004). 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; recent sexual behavior, including any ncRAI in the prior 3 months, 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) 1-year HIV risk differences and NNTs would have been had the iPrEx trial population shared the same distribution of baseline characteristics as was observed in each clinic population.

Data and measurements

The iPrEx study randomized 2,499 HIV-negative MSM and transgender women 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 ($n = 2{,}441$).

To represent our 2 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 (27, 28). 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 (28).

In both the iPrEx and the Latino MSM Community Involvement studies, participants were asked about their sexual behavior, demographic characteristics, sexually transmitted infections history, and alcohol and drug use via a computer-assisted structured interview (5, 27, 28).

Notation, target parameters, and identification

Our goal was to estimate the subgroup-specific ITT HIV risk difference at 1 year between those randomized to the PrEP arm and those randomized to the placebo arm, and the corresponding NNTs to prevent 1 infection per year in iPrEx, San Francisco, and Chicago. Our subgroups variables of interest were MSM, TGW, report of any ncRAI in the prior 3 months, primary sexual role (top, bottom, versatile), and report of using cocaine in the prior 6 months.

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 2 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:

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

and the NNT (4)—1 over the absolute value of the ITT—for each subgroup $G = g$ in population $S = s$ as:

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

For simplicity, we assume there was no measurement error. To identify the target parameters within the iPrEx study population, we must assume conditional treatment exchangeability and treatment positivity. 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. Treatment positivity: $P(Z = z\,|G = g) > 0$ for all g for which $P(G = g) > 0$. That is, there must be a nonzero probability of being assigned each treatment for each subgroup (29). Randomized treatment assignment in the iPrEx trial guarantees that there are no structural positivity violations, but it does not guarantee the absence of practical positivity violations, which are more likely to occur in smaller samples in subgroups.

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

$$\begin{aligned}\psi_g^0 &\equiv E\left(HIV^{Z=1} - HIV^{Z=0} \middle| G = g, S = 0\right) \\ &= E\left[HIV \middle| Z = 1, G = g, S = 0\right] \\ &\quad - E\left[HIV \middle| Z = 0, G = g, S = 0\right] \end{aligned} \quad (3)$$

and:

$$\begin{aligned}\xi_g^0 &= \frac{1}{\left|\psi_g^0\right|} \\ &= \frac{1}{\left|E\left[HIV \middle| Z=1, G=g, S=0\right] - E\left[HIV \middle| Z=0, G=g, S=0\right]\right|}.\end{aligned} \quad (4)$$

To identify the transported target parameters we must meet the following additional criteria (23):

- Conditional population exchangeability: $E(HIV|S = 0, \mathbf{W_g}, Z, G = g) = E(HIV|S = s', \mathbf{W_g}, Z, G = g)$. 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 ($\mathbf{W_g}$) and treatment assignment.
- Population positivity: $P(S = 0, Z = z|\mathbf{W_g}, G = g) > 0$ for all $P(\mathbf{W_g}|G = g, S = s') > 0$. That is, every combination of $\mathbf{W_g} = \mathbf{w_g}$ that could be drawn from the distribution of $\mathbf{W_g}$ in each strata $G = g$ within each target population is represented in the iPrEx study population in $G = g$ and has a nonzero probability of being assigned $Z = z$.

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

Selection diagrams are augmented directed acyclic graphs (30, 31) introduced by Pearl and Bareinboim (18) 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 (18, 32). 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 (33)) from the outcome variable (18). 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 Web Appendix 1 and Web Figure 1 (available at https://doi.org/10.1093/aje/kwab045).

Figure 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 ($\mathbf{W_g}$) 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 $HIV \perp S \mid Z, G = g, \mathbf{W_g}$):

- Gender identity (MSM and TGW): age, education, number of partners, ncRAI, cocaine use, and alcohol consumption.
- ncRAI: age, education, number of partners, cocaine use, alcohol consumption.
- Primary sexual role (top, bottom, versatile): age, education, number of partners, ncRAI, cocaine use, alcohol consumption, and gender identity.

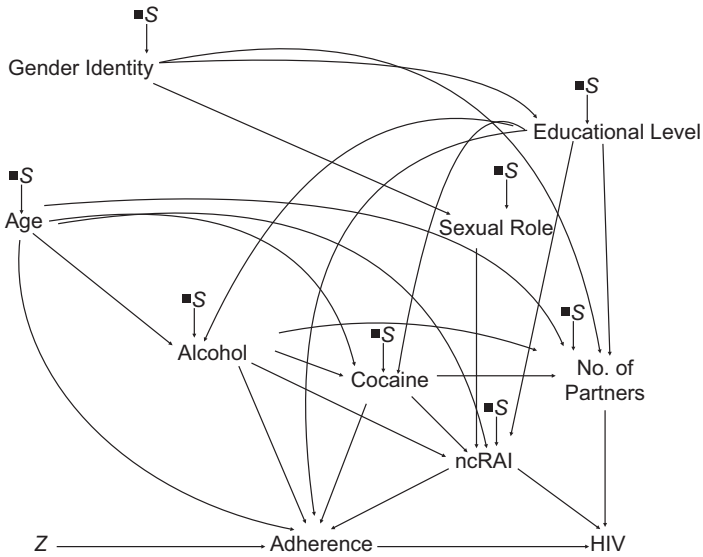

Figure 1. Proposed selection diagram representing the assumed causal model within the Iniciativa Profilaxis Pre-Exposición (iPrEx) study (multiple countries, 2007–2010) and assumed differences between the study population and each target population (Latino MSM Community Involvement Study: San Francisco and Chicago, 2004). Age is age at baseline; gender identity refers to transgender women who have sex with men or cisgender men who have sex with men; educational level is highest level attained; 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. *S*, selection node; *Z*, treatment assignment.

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

Under these assumptions, the transported target parameters are the subgroup-specific ITT effects and NNTs had the iPrEx study population had the same distribution of $\mathbf{W_g}$ as was observed in each subgroup of each target population (San Francisco or Chicago) and are identified by:

$$\begin{aligned}\psi_g^{s'} &\equiv E\left(HIV^{Z=1} - HIV^{Z=0} \middle| G=g, S=s'\right) \\ &= E\left(E\left[HIV \middle| Z=1, \mathbf{W_g}, S=0\right] \middle| G=g, S=s'\right) \\ &\quad - E\left(E\left[HIV \middle| Z=0, \mathbf{W_g}, S=0\right] \middle| G=g, S=s'\right) \quad (5)\end{aligned}$$

and the transported NNTs are:

$$\begin{aligned}\xi_g^{s'} &= \frac{1}{\left|\psi_g^{s'}\right|} \\ &= \frac{1}{\left|E\left(E\left[HIV \middle| Z=1, \mathbf{W_g}, G=g, S=0\right] \middle| G=g, S=s'\right) \right.} \\ &\quad \left. -E\left(E\left[HIV \middle| Z=0, \mathbf{W_g}, G=g, S=0\right] \middle| G=g, S=s'\right)\right| \end{aligned} \quad (6)$$

Estimation

To estimate the subgroup-specific ITT effects in iPrEx, we fitted a log-binomial regression model with main terms for treatment assignment and the subgroup variable as well as a term for interaction between treatment assignment and subgroup. Using this model, we predicted the marginal incidence risk difference at 1 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, although doing so might improve efficiency (34). Note that because this model does not adjust for additional covariates and is fully saturated, these estimates are identical to simply comparing the raw proportions within each treatment subgroup (35).

The identifiability result in equation 5 can be rearranged to equal the subgroup-specific expected value of HIV infections after 1 year under each treatment assignment in the study population weighted by the inverse odds of selection weights (36):

$$IOSW_i = \begin{cases} \frac{P(S_i=0|\mathbf{W_g},G_i)}{P(S_i=s'|\mathbf{W_g},G_i)} * \frac{P(S_i=s',G_i)}{P(S_i=0,G_i)}, & S_i = 0, \\ 0, & S_i = s' \end{cases}$$

See the Appendix of Westreich et. al. (36) for the complete derivation of the inverse odds of selection weights formula.

Each component of the inverse odds of selection weights 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

Table 1. Demographic Characteristics According to Population, Multiple Locations[a], 2004–2010

Characteristic	iPrEx (*n* = 2,499)		San Francisco (*n* = 210)		Chicago (*n* = 263)	
	No.	%	No.	%	No.	%
No. of male partners in prior 3 months[b]	16.9 (35.6)		8 (10.2)		7.3 (10)	
Age at baseline, years						
18–25	1,374	55	37	17.6	76	28.9
26–35	730	29.2	78	37.1	113	43
36–45	270	10.8	66	31.4	50	19
>45	125	5	29	13.8	24	9.1
Highest level of education						
Less than high school	524	21	57	27.1	63	24
High school	884	35.4	38	18.1	68	25.9
College	1,091	43.7	115	54.8	132	50.2
Gender identity						
Cisgender man	2,174	87	172	81.9	249	94.7
Transgender woman	325	13	38	18.1	14	5.3
Alcohol consumption in prior month						
None or less than once a month	496	19.8	90	42.9	68	25.9
1–4 drinks per day	634	25.4	72	34.3	85	32.3
5 or more drinks per 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	2,368	94.8	187	89	215	81.7
Yes	131	5.2	23	11	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	1,024	41	87	41.4	94	35.7
ncRAI in prior 3 months						
No	1,014	40.6	159	75.7	185	70.3
Yes	1,485	59.4	51	24.3	78	29.7

Abbreviations: iPrEx, Iniciativa Profilaxis Pre-Exposición; ncRAI, condomless receptive anal intercourse.
[a] iPrEx (multiple countries, 2007–2010) and the Latino MSM Community Involvement Study (San Francisco, California, and Chicago, Illinois, 2004).
[b] Values are expressed as mean (standard deviation).

probability weights would be an appropriate analogous estimator (19, 36).

The inverse odds weights were used to fit weighted saturated log-binomial regressions with a term for interaction between treatment assignment and subgroup. The regressions did not adjust for any additional covariates. We used this model to predict the number of incident HIV infections at 1 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 (37) with 2,000 resamples. The bootstrap resampled both the iPrEx study population and target populations, and then we calculated new weights and fitted 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 NNT was estimated as the inverse of the difference in risk of HIV infection at 1 year of follow-up (38) giving the number of individuals who need to be offered PrEP to avert 1 infection in 1 year.

All analyses were conducted using R, version 3.4.1 (R Foundation for Statistical Computing, Vienna, Austria) (39), and Stata, version 15.1 (StataCorp LP, College Station, Texas) (40). The code for these analyses can be found on GitHub (41).

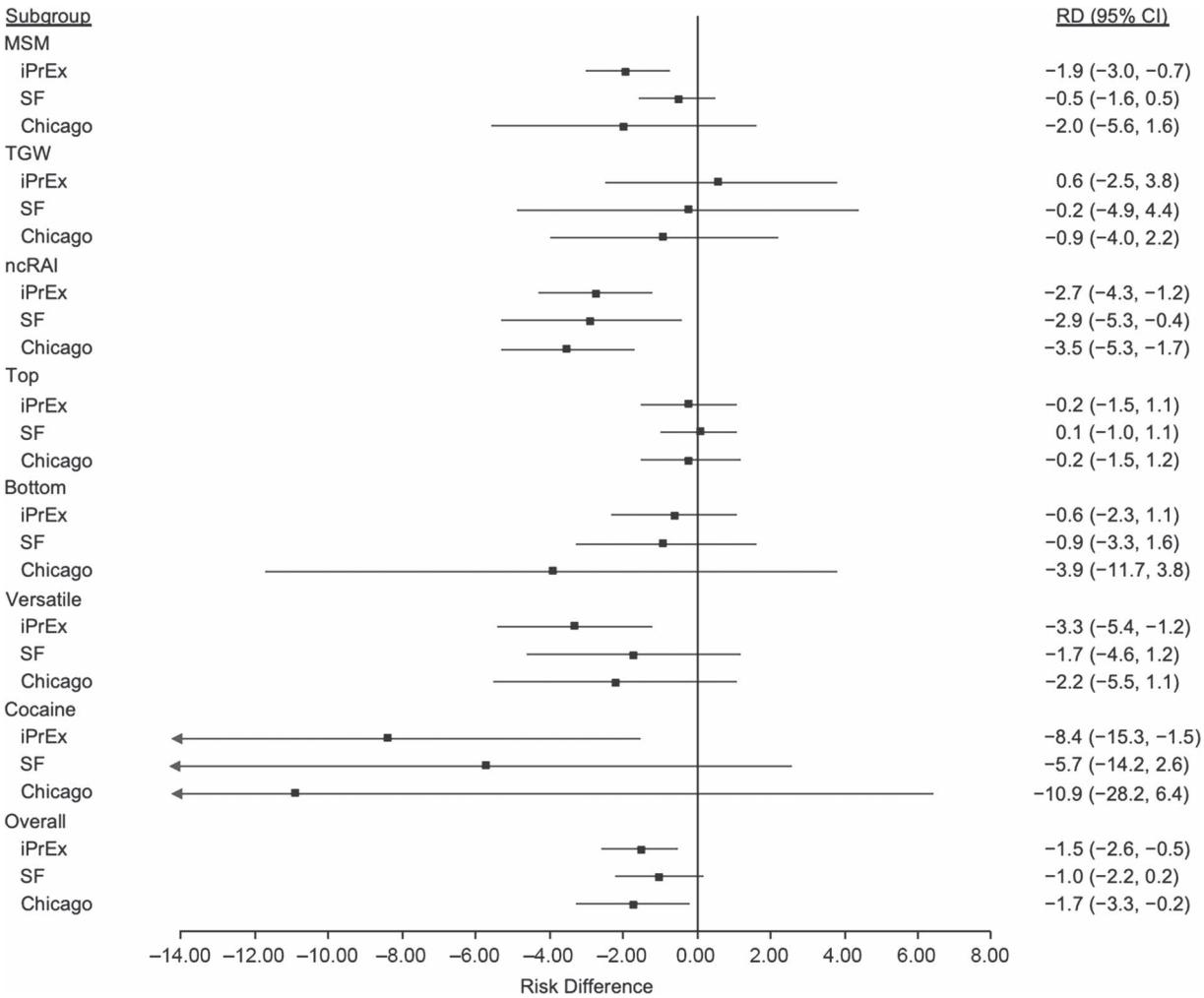

Figure 2. Subgroup-specific risk differences (RDs) in the Iniciativa Profilaxis Pre-Exposición (iPrEx) study (multiple countries, 2007–2010) and the Latino MSM Community Involvement Study (San Francisco (SF) and Chicago, 2004). CI, confidence interval; MSM, cisgender men who have sex with men; ncRAI, condomless receptive anal intercourse; TGW, transgender women who have sex with men.

RESULTS

There were differences in most baseline characteristics across settings (Table 1), and in particular, the iPrEx study population had on average more recent sexual partners and more individuals reporting recent ncRAI.

Figure 2 shows the subgroup-specific intention-to-treat risk differences at 1 year, and Figure 3 shows the numbers needed to treat to prevent 1 infection in each population. In all settings, cocaine users had the lowest NNT. In Chicago, the NNT was next lowest among those whose primary sexual role was "bottom," whereas in iPrEx the sexual role with the next lowest NNT was "versatile." In all settings, PrEP was not expected to be beneficial for those whose primary sexual position was "top." Finally, although 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 were diminished.

DISCUSSION

Subgroups with the lowest NNTs in trial populations might 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 might 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 specific target populations. In iPrEx, those who indicated that their primary sexual role was "versatile"

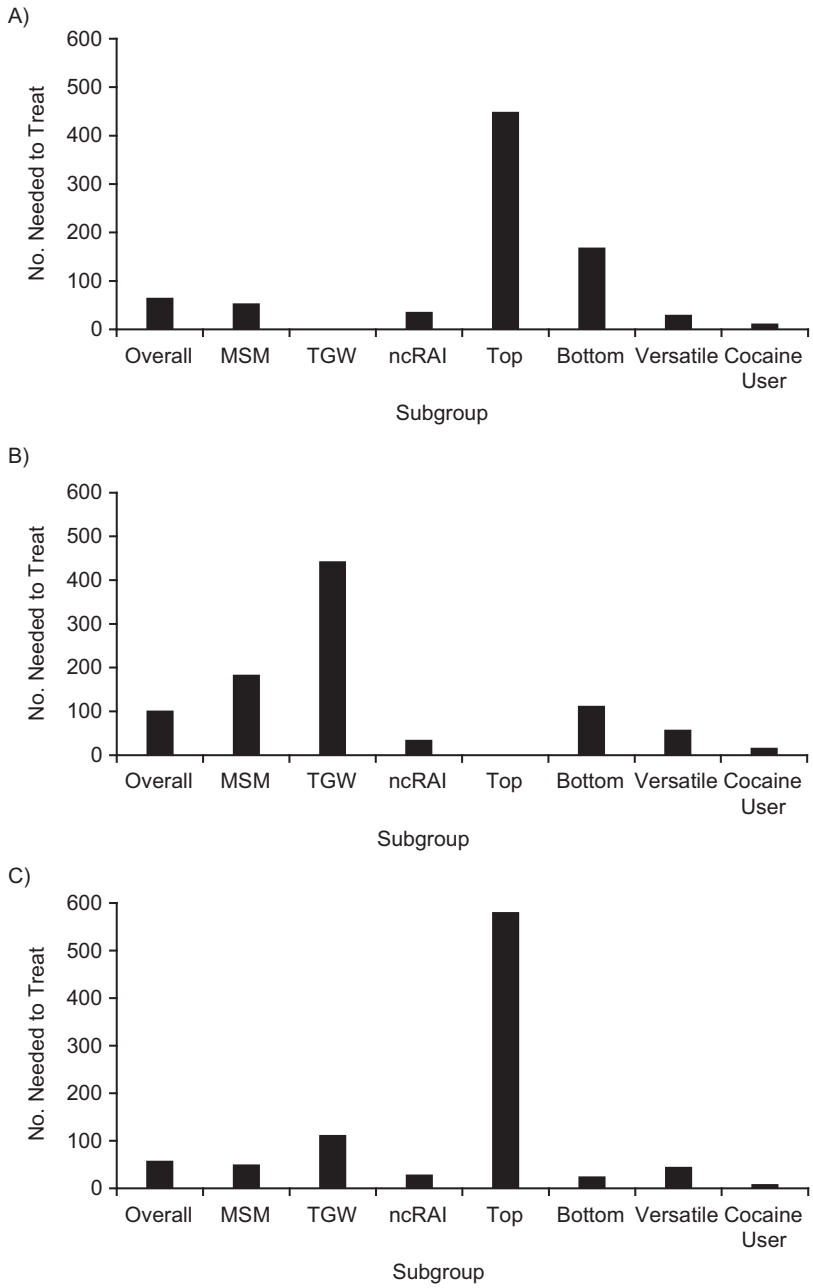

Figure 3. Subgroup-specific numbers needed to treat in Iniciativa Profilaxis Pre-Exposición (iPrEx) study (multiple countries, 2007–2010) and the Latino MSM Community Involvement Study (San Francisco (SF) and Chicago, 2004). A) iPrEx. B) San Francisco, California. C) Chicago, Illinois. The number needed to treat is the number of individuals who need to be offered preexposure prophylaxis to prevent 1 incident infection in 1 year. When the calculated number needed to treat was less than zero, the value was set to "not applicable" indicating that there was no benefit of offering preexposure prophylaxis to participants in this subgroup. MSM, cisgender men who have sex with men; ncRAI, condomless receptive anal intercourse; TGW, transgender women who have sex with men.

had 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 ncRAI have a lower NNT than any specific sexual role. Prioritizing PrEP according to self-reported sexual role is appealing, because 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 data set 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 might 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.

Although 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 might not reflect the current needs of these populations. Next, given that 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 population shared the same distribution of baseline characteristics as was observed in each clinic population. However, these estimates do not necessarily reflect what would be observed if the study were to be repeated in these target populations today. 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 2. The NNTs, 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 it also underscores an important challenge in transporting subgroup analyses more broadly. Trials are often underpowered to detect subgroup differences, and transport estimators could reduce the precision of subgroup estimates. Here, we elected to use inverse-odds-weighting estimators because we believe these estimators are easy to understand and implement given their similarity to other common tools used in causal inference (36). Other estimation approaches, including g-computation and targeted maximum likelihood estimation, can also be used to transport subgroup results from trials and might be more statistically efficient. We encourage researchers who are planning on transporting their results to target populations to consider tradeoffs in bias and variance when selecting an estimator for their particular application (42).

Finally, a central issue that researchers will face when employing transportability methods is that 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 the implementation of new interventions.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology and Biostatistics, University of California, San Francisco, San Francisco, California, United States (Megha L. Mehrotra, M. Maria Glymour, David V. Glidden); Department of Epidemiology, UNC Gillings School of Global Public Health, University of North Carolina, Chapel Hill, Chapel Hill, North Carolina, United States (Daniel Westreich); and Institute for Public Health, Washington University in St. Louis, St. Louis, Missouri, United States (Elvin Geng).

This work was supported by the National Institute of Mental Health (grant 5F31 MH111346 to M.L.M.). This work was also supported by the National Institutes of Health (grants DP2 HD084070 to D.W., K24 AI134413 to E.G., CFAR SWG Implementation Science Working Group to E.G., and AI 126597 to D.V.G.). The Iniciativa Profilaxis Pre-Exposición (iPrEx) study (NCT00458393) was supported by the National Institutes of Health (grant AI064002 to R.M.G.) and the Bill & Melinda Gates Foundation.

D.V.G. has accepted fees from Gilead Sciences. The other authors report no conflicts.

REFERENCES

1. Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet*. 2005;365(9454):176–186.
2. VanderWeele TJ, Knol MJ. On the interpretation of subgroup analyses in randomized trials: heterogeneity versus secondary interventions. *Ann Intern Med.* 2011;154(10):680–683.

3. 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.
4. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *BMJ*. 1995; 310(6977):452–454.
5. 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.
6. 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.
7. 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.
8. 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* (electronic article). 2015;18(4):19973.
9. 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): e25096.
10. Westreich D, Edwards JK, Lesko CR, et al. Target validity and the hierarchy of study designs. *Am J Epidemiol*. 2019; 188(2):438–443.
11. Kennedy-Martin T, Curtis S, Faries D, et al. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials*. 2015;16:495.
12. Gonzalez LE, Sutton SK, Pratt C, et al. The bottleneck effect in lung cancer clinical trials. *J Cancer Educ*. 2013;28(3): 488–493.
13. Susukida R, Crum RM, Stuart EA, et al. Assessing sample representativeness in randomized controlled trials: application to the National Institute of Drug Abuse clinical trials network. *Addiction*. 2016;111(7): 1226–1234.
14. Eisenberg Y, Mohiuddin H, Cherukupally K, et al. 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(1):306.
15. Isaacs T, Hunt D, Ward D, et al. 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):e256.
16. Curno MJ, Rossi S, Hodges-Mameletzis I, et al. A systematic review of the inclusion (or exclusion) of women in HIV research: from clinical studies of antiretrovirals and vaccines to cure strategies. *J Acquir Immune Defic Syndr*. 2016;71(2): 181–188.
17. Pearl J, Bareinboim E. External validity: from do-calculus to transportability across populations. *Stat Sci*. 2014;29(4): 579–595.
18. Pearl J, Bareinboim E. Transportability across studies: a formal approach. 2011. https://apps.dtic.mil/sti/citations/ADA557437. Accessed August 11, 2015.
19. Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: the ACTG 320 trial. *Am J Epidemiol Engl*. 2010;172(1):107–115.
20. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. Lippincott Williams & Wilkins; 2008.
21. Lesko CR, Buchanan AL, Westreich D, et al. Generalizing study results: a potential outcomes perspective. *Epidemiology*. 2017;28(4):553–561.
22. Balzer LB. "All generalizations are dangerous, even this one."–Alexandre Dumas. *Epidemiology*. 2017;28(4):562–566.
23. Rudolph KE, van der Laan MJ. Robust estimation of encouragement-design intervention effects transported across sites. *J R Stat Soc Ser B Stat Methodol*. 2017;79(5): 1509–1525.
24. 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. *Epidemiology*. 2018;29(2):199–206.
25. 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.
26. Wilson D, Halperin DT. "Know your epidemic, know your response": a useful approach, if we get it right. *Lancet*. 2008; 372(9637):423–426.
27. Ramirez-Valles J, Garcia D, Campbell RT, et al. 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.
28. Ramirez-Valles J. Latino MSM Community Involvement: HIV protective effects. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. 2014. (http://doi.org/10.3886/ICPSR34385.v2). Accessed May 4, 2021.
29. Petersen ML, Porter KE, Gruber S, et al. Diagnosing and responding to violations in the positivity assumption. *Stat Methods Med Res*. 2012;21(1):31–54.
30. 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(suppl 2): 139S–161S.
31. Pearl J. *Causality: Models, Reasoning, and Inference*. New York, NY: Cambridge University Press; 2000.
32. Petersen ML. Compound treatments, transportability, and the structural causal model: the power and simplicity of causal graphs. *Epidemiology*. 2011;22(3):378–381.
33. Pearl J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Francisco, CA: Morgan Kaufmann Publishers Inc.; 1988.
34. Rosenblum M, van der Laan MJ. Simple, efficient estimators of treatment effects in randomized trials using generalized linear models to leverage baseline variables. *Int J Biostat Ger*. (electronic article). 2010;6(1):13.
35. Lumley T, Shaw PA, Dai JY. Connections between survey calibration estimators and semiparametric models for incomplete data. *Int Stat Rev*. 2011;79(2):200–220.
36. Westreich D, Edwards JK, Lesko CR, et al. Transportability of trial results using inverse odds of sampling weights. *Am J Epidemiol*. 2017;186(8):1010–1014.
37. Efron B, Tibshirani R. *An Introduction to the Bootstrap*. New York, NY: Chapman & Hall; 1993.
38. 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.
39. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2017.

40. StataCorp. *Stata Statistical Software: Release 15*. StataCorp LLC: College Station, TX; 2017.
41. Mehrotra ML. Transport-subgroups. 2019. https://github.com/megtron9/transport-subgroups. Accessed May 4, 2021.
42. Dahabreh IJ, Robertson SE, Steingrimsson JA, et al. Extending inferences from a randomized trial to a new target population. *Stat Med*. 2020;39(14):1999–2014.