Estimating human immunodeficiency virus (HIV) prevention effects in low-incidence settings

Jacqueline E. Rudolph[1], **Stephen R. Cole**[1], **Joseph J. Eron**[2], **Angela D. Kashuba**[3], and **Adaora A. Adimora**[1,2]

[1]Department of Epidemiology, University of North Carolina at Chapel Hill

[2]School of Medicine, University of North Carolina at Chapel Hill

[3]School of Pharmacy, University of North Carolina at Chapel Hill

Abstract

Background.—Randomized controlled trials (RCTs) for determining efficacy of pre-exposure prophylaxis (PrEP) in preventing human immunodeficiency virus (HIV) infection have not been conducted among US women because their lower HIV incidence requires impractically large studies. Results from higher incidence settings, like Sub-Saharan Africa, may not apply to US women due to differences in age, sexual behavior, coinfections, and adherence.

Methods.—We propose a novel strategy for evaluating PrEP efficacy in the US using data from both settings to obtain four parameters: (1) intention-to-treat (ITT) and (2) per-protocol effects in the higher incidence setting, (3) per-protocol effect generalized to the lower incidence setting, and (4) back-calculated ITT effect using adherence data from the lower incidence setting. To illustrate, we simulated two RCTs comparing PrEP against placebo: one in 4000 African women and another in 500 US women. We estimated all parameters using g-computation and report risk ratios averaged over 2000 simulations, alongside the 2.5th and 97.5th percentiles of the simulation results.

Results.—Twelve months post-randomization, the African ITT and per-protocol risk ratios were 0.65 (0.47, 0.88) and 0.20 (0.08, 0.34), respectively. The US ITT and per-protocol risk ratios were 0.42 (0.20, 0.62) and 0.17 (0.03, 0.38), respectively. These results matched well the simulated true effects.

Conclusions.—Our simple demonstration informs the design of future studies seeking to estimate the effectiveness of a treatment (like PrEP) in lower incidence settings where a traditional RCT would not be feasible.

Keywords

pre-exposure prophylaxis; HIV; intention-to-treat effect; per-protocol effect; generalization; simulation

Correspondence to: Jacqueline Rudolph, 2101 McGavran-Greenberg Hall CB# 7435, Chapel Hill, NC 27599, jerudolp@live.unc.edu, T: (919) 966-7430.

Introduction

Randomized controlled trials (RCTs) are the gold standard for determining the efficacy of a treatment, but variations in adherence can compromise the generalizability of RCT results to groups that differ from the study population in which the RCT was conducted. Adimora et al[1] discussed an approach to estimate the intention-to-treat (ITT) effect in a target population whose outcome incidence is substantially lower than the incidence in the population where the treatment was formally tested using a RCT. The authors outlined how to use key data from both samples, such as biologically determined adherence and baseline effect measure modifiers, and the per-protocol effect (i.e., the effect of remaining on trial protocol, usually of remaining adherent to the trial drug) as an intermediary step to obtain the desired estimates in the target population. The approach discussed by Adimora et al. is important when the event of interest is uncommon enough in the target population to render a local RCT infeasible. In brief, the steps of this approach are (1) to estimate both the ITT and (2) per-protocol parameters in the higher-incidence setting, then (3) to generalize the per-protocol parameter estimate to the lower-incidence setting, and finally and innovatively (4) to estimate the ITT in the lower-incidence setting by combining the generalized per-protocol estimate with information on observed compliance in the target setting. While we believe this approach has many potential applications, we describe here how it might be implemented for the example described by Adimora et al., namely determining the effectiveness of pre-exposure prophylaxis (PrEP) as prevention for human immunodeficiency virus (HIV) transmission.[1] We provide a summary of simulation experiments that detail how the proposed approach works in typical sample sizes, and we give examples of the types of variables one would need to measure and include when adapting our approach to real-world data.

The Motivating Example

PrEP has been shown in some populations to be effective at preventing transmission of HIV. For instance, among men who have sex with men (MSM), tenofovir/emtricitabine reduced incidence of HIV by 44% compared to placebo (hazard ratio (HR): 0.56, 95% CI: 0.37-0.85).[2] One trial in serodiscordant, heterosexual couples in Kenya and Uganda found that tenofovir/emtricitabine reduced HIV incidence by 75% compared to placebo (HR: 0.25, 95% CI: 0.13-0.45).[3] However, we currently have a limited understanding of how well PrEP prevents HIV among women in the US.

There are two potential reasons for the lack of information on the effectiveness of PrEP among US women. First, it would be infeasible to conduct a Phase III RCT in this population because the incidence of HIV is too low. Even among US women at exceptionally high risk, whose HIV incidence rate was found by one prior study to be 320/100,000 person–years,[4] one would need to enroll 10,000 women to have adequate statistical power to detect the same effectiveness observed in MSM.[1] This would be infeasible not just because of the required sample size but also the difficulty of enrolling women who are at highest risk for HIV in the US. Second, results from populations of women in Southern Africa, where incidence rates have been observed to be as high as

5700/100,000 person–years,[5] are likely not directly generalizable to US women. There is concern regarding generalizability because distributions of key demographic, socioeconomic, clinical, and behavioral factors are likely to differ between US women and African women at risk for HIV. Generalization of results is made even more difficult because the trials conducted in African women (e.g., the FEM-PrEP and VOICE trials) had to be stopped due to the lack of demonstrated effectiveness of tenofovir/emtricitabine.[5,6] In these trials, adherence to study protocol was only 30-40%, so the observed ITT analyses were essentially futile (i.e., undetectable regardless of actual efficacy). Additionally (and more importantly for our purposes), the differences between African and US women in demographic, clinical, and behavioral factors make it likely that the adherence patterns observed in the African trials differ from what would be seen among US women.

We can overcome the above challenges by combining key data (particularly a valid measure of adherence) from both the higher incidence African and lower incidence US settings with modern analytic methods to obtain both the per-protocol and ITT effects in the target population under stated assumptions.

Methods

The Proposed Approach

Our approach consists of four steps: (1) obtain the ITT and (2) per-protocol effects in the higher incidence African setting,[7,8] (3) generalize the per-protocol results to the lower incidence US setting, and (4) estimate the ITT effect in the US. While per-protocol results have been generalized to a target population,[9] to our knowledge the last step is novel.

Prior to conducting the analyses, we generated 2000 simulations of the two populations of interest. Our higher incidence sample mimicked a placebo-controlled RCT of the effect of tenofovir/emtricitabine on prevention of HIV transmission in 4000 women in Southern Africa (specifically, the FEM-PrEP and VOICE trials).[5,6,10] Our lower incidence sample mimicked a RCT of 500 women in the US (with variable distributions being informed by HPTN 064).[4,11,12] In both settings, the populations were women at risk of acquiring HIV, with the expectation that some but not all women would be in serodiscordant partnerships. In each trial, we randomized half the participants to receive tenofovir/emtricitabine and half to placebo and, if they were assigned to PrEP, determined their adherence (treated simply as a binary variable). In Africa half of the women randomized to PrEP were expected to take it at a level required by the study protocol, e.g. a stipulation that women take 80-90% of dispensed pills or have 10ng of tenofovir per milliliter of plasma.[6] In the US, 75% of those randomized to PrEP were expected to follow protocol, using the same definition as in Africa. Whether women adhered was dependent on the woman's age (defined as >21 years or not). We then simulated whether they acquired HIV 12 months post-randomization, based on adherence, age, and the interaction between adherence and effect measure modifier lack of a sexually transmitted infection (STI) at baseline. Lack of an STI was selected as our example modifier because the distribution of STIs differs between the two settings and having an STI increases risk of acquiring HIV.[13-16] Figure 1 illustrates the causal relationships between the example variables included in our simulation.[17]

Starting with the higher incidence African setting, our first parameter of interest was the ITT effect of randomization to PrEP versus placebo on HIV incidence. We estimated the ITT risk ratio (RR) in two ways: (1) using a log binomial generalized linear regression model and (2) using the Snowden adaptation of Robin's generalized computational algorithm formula (g-formula).[18] It was not necessary to include any covariates in the models to estimate the ITT.

To account for non-compliance in the PrEP arm, we then obtained an estimate of the effect of remaining adherent to trial protocol on HIV seroconversion in the African trials, i.e. the per-protocol parameter. This method corrects for adherence by analytically censoring participants when they no longer adhere to trial protocol, following a pre-determined rule such as described above. Since this process involves analysis of adherence to protocol, which is by definition a post-randomization variable (and is not guaranteed to be balanced across the trial arms), we must account for (but not stratify on) those variables that confound the relationship between protocol adherence and the outcome.[7,8,19]

The first method we used to estimate the per-protocol effect was a log binomial model among women who adhered to the protocol, weighted to represent the entire trial using stabilized inverse probability of censoring weights.[8] Variables included in the censoring weight models were those that affect adherence and the outcome, so our weight models included our example confounder age. The second method was g-computation. The key to using the g-formula to estimate the per-protocol effect is in specifying the counterfactuals being compared: (1) set all participants to receive PrEP and adhere to protocol versus (2) set all participants to receive placebo.[20] When modeling the outcome, the model must include confounders of compliance and the outcome. Using the parametric g-formula (g-computation), we estimated risk of the outcome under counterfactual 1 and under counterfactual 2 and then compared those risks to obtain the RR.

Following estimation of the per-protocol effect in the higher incidence African setting, one can standardize the results to the target population using the known distributions of effect measure modifiers that differ between the populations.[21] In our scenarios, lack of an STI was our example baseline effect measure modifier. We again estimated this parameter in two ways. First, we estimated stabilized inverse odds of selection weights in a combined data set of the African and US trials, where selection was defined as inclusion in the African trial. The only variable that needed to be included in the weight models was lack of an STI.[22] Those weights were multiplied by the censoring weights, and we ran inverse probability weighted log binomial models in the African data to obtain the RR for the per-protocol estimate generalized to the lower incidence US setting. Second, we adapted our g-computation approach for the per-protocol estimate to include an initial step where we took a weighted sample from the African women, where participants were selected with replacement and with a weight determined by our previously estimated inverse odds of selection weights. This ensured that, while the distribution of all other variables remained the same as in African, the distribution of any baseline modifiers would be (on average) the same as in the US. The per-protocol effect was then estimated in the same way as above but now within this weighted sample.

In the final step, we obtained the ITT effect generalized to the US, using the reweighted African sample. We did this by combining information on adherence in the target, lower incidence US setting with the g-computation estimators of the generalized per-protocol parameter. This involved adding a step in which we used the US data to model adherence among those randomized to PrEP. The coefficients from this US adherence model were then used when generating the outcomes, instead of using the coefficients from the African adherence model. Finally, we obtained RRs by comparing the risks under the scenarios where (1) all participants were set to receive PrEP but adherence was allowed to be what it would have been in the US and (2) all participants were set to receive placebo.

Several assumptions are required to identify the effects of interest in our approach. For all four steps, one must assume that the treatment is well defined (i.e., there is no interference and any variations in treatment are irrelevant),[23,24] there is no measurement error, and all models are correctly specified. Perhaps more importantly for the design of a study seeking to use our approach, one must make the assumption of conditional exchangeability for internal validity[25] and its external validity counterpart.[26] For the ITT effect, randomization of treatment grants exchangeability in expectation. When estimating the per-protocol effects in steps 2 and 3, one has to assume that all confounders of adherence to protocol and the outcome have been accounted for, to ensure that those who adhered to protocol in the treatment arm are conditionally exchangeable with those who adhered to protocol in the comparator arm. Analogously, when generalizing to the target setting in steps 3 and 4, one has to assume that all baseline effect measure modifiers which also affect selection into one setting versus another have been accounted for, to ensure conditional exchangeability between those selected into the source population (higher incidence setting) and those not selected (lower incidence setting). Lastly, one has to assume there is a positive probability of being observed in all strata defined by the treatment (or, for external validity, selection) and those variables necessary to achieve conditional exchangeability.[26-28]

Below we provide the results summarized across 2000 simulations. Key population variables are summarized using the median and interquartile range (IQR) of the percentages in each simulation. We then report the RR at each step averaged across the simulations, alongside the 2.5[th] and 97.5[th] percentiles (hereafter referred to as the 95% central mass) of the 2000 RRs. We additionally give Monte Carlo simulation errors for each parameter, estimated by the standard error of the simulation RRs.

All analyses were conducted in SAS version 9.4 (SAS Institute, Cary, NC). Technical details on how each step was carried out using g-computation and SAS code are given in the eAppendices.

Results

As shown in Table 1, across the 2000 simulations, 50.0% (IQR: 49.5%, 50.5%) of the 4000 women enrolled in the African trial were randomly assigned to PrEP, but only 24.9% (IQR: 24.4%, 25.4%) of the 4000 women actually received PrEP. One third (33.3%; IQR: 32.9%, 33.9%) were >21 years of age, and 79.0% (IQR: 78.6%, 79.4%) tested negative for an STI at baseline. At 12 months, 4.2% (IQR: 4.0%, 4.4%) of African women acquired HIV.

In the lower-incidence US setting, again half (50.0%; IQR: 48.6%, 51.4%) of the 500 women were randomized to receive PrEP, and 37.0% (IQR: 35.4%, 38.4%) of the 500 women took PrEP. The US women were older than those in the African setting (66.6% >21 years of age; IQR: 65.4%, 68.2%) and more likely to test negative for an STI at baseline (89.0%; IQR: 88.2%, 90.0%). The 12-month risk of HIV infection in US women was 0.4% (IQR: 0.2%, 0.8%).

Table 2 provides the average and 95% central mass of the 2000 simulation parameter estimates obtained using the g-computation approach, and Figure 2 displays the full distribution of 2000 simulation RRs. We found that the risk of HIV infection in African women randomized to PrEP was 0.65 (95% central mass: 0.47, 0.88) times the risk among women randomized to placebo. The risk of HIV infection among African women who remained adherent to study protocol and took PrEP was 0.20 (95% central mass: 0.08, 0.34) times the risk among women who stayed on study protocol but took placebo. As expected, the ITT effect was closer to the null than the per-protocol effect because low adherence in the PrEP arm attenuated PrEP's protective effect.

As for the generalized effect estimates, the risk of HIV infection in US women randomized to PrEP was 0.42 (95% central mass: 0.20, 0.62) times the risk among women randomized to placebo, and the risk among US women who remained adherent to study protocol and took PrEP was 0.17 (95% central mass: 0.03, 0.38) times the risk among women who stayed on protocol but took placebo. Both the US ITT and per-protocol results were further from the null than the African results because more of the US women tested negative for an STI at baseline, and lack of an STI was associated with a more protective treatment effect. Additionally, for the ITT effect, US women were more likely to adhere to PrEP (partly because they were older), which resulted in a stronger protective effect of PrEP than in Africa where adherence was lower. The Monte Carlo simulation errors (defined as standard errors of the simulation RRs) were <0.01 for all four parameters.

To assess the performance of our approach, we used several different comparators. First, we saw that the parameters estimated using log binomial models and, for the per-protocol effects, inverse-probability-weighting methods were comparable (eTable 1). The generalized US ITT parameter was not estimated using weighting methods (to our knowledge, no such methods currently exist) and could not be compared. Second, we generated counterfactual outcomes for each individual in each simulation under the ITT and the per-protocol. We then estimated the true RRs in Africa and in the US by dividing the average of the counterfactual outcomes. For Africa the true ITT RR was 0.64 and the true per-protocol RR was 0.20; for the US those RRs were 0.44 and 0.17, respectively. Thus, all four of our g-computation estimates were close to the true parameter values.

Last, we compared our generalized results to what would have been obtained had we enrolled 10,000 women in the US – the estimated sample necessary to have adequate statistical power to detect the effect size observed in trials among MSM. Consequently, this is a large enough population that effects could be estimated without generalization. We applied the same methods used to estimate the African ITT and per-protocol effects and found that the ITT and per-protocol effects in this large US trial were 0.41 and 0.13,

respectively. These results (given in Table 2) again matched closely those obtained using our generalization approach.

Discussion

Here, we proposed an approach to generalize ITT and per-protocol effects from a higher incidence setting to obtain estimates of those same effects in a lower incidence setting. In our example, HIV incidence in the US was not high enough to permit estimation of these effects without enrolling an impractically large sample, but, using key data from a trial of African women alongside modern statistical methods, we were able to estimate both the generalized per-protocol and ITT effects. Our results were consistent across different estimators, i.e. the g-computation results matched results obtained using inverse-probability weighting methods. Furthermore, our per-protocol results were comparable to the "true RR" estimated from the individual counterfactual outcomes, and our generalization results matched those we would have seen if we enrolled a trial of 10,000 US women.

An approach like the one described here is necessary if the goal is to estimate the ITT effect of a treatment in the lower incidence setting. This is because the ITT is a function of the level of adherence within a population. Therefore, if one generalized the ITT from the higher-incidence setting to the lower-incidence setting (which requires measuring in both settings and analytically controlling for the necessary effect measure modifiers), the result would be biased if the patterns of adherence were different. That adherence is a post-randomization variable on the pathway between randomization and the outcome means that it must be carefully handled during the analysis. Estimation of the per-protocol effect is one way to validly account for adherence, after which generalization can be carried out in the usual way. The generalized per-protocol effect can then be tweaked using measured data on adherence in the lower-incidence target population to back-calculate the ITT effect.

There were of course several limitations of this simulation. Primarily, our design was simplified, with only one time-point and an unrealistically small variable set. Future applications of our approach will likely have many time points at which they assess adherence and HIV incidence, and use of such data will require adequately accounting for time-varying adherence and time-varying confounders. However, while we do not show it here, the methods used at each step can be extended to handling the time-varying case, and we plan in future work to adapt our approach to these more complex settings.[29] Moreover, assessment of variables in real trials will generally not be as clean cut as shown here. For instance, adherence can be difficult to measure, being highly prone to measurement error, and may not always be a yes/no binary variable. We also constructed our example such that all women in both samples remained in the study until the outcome was assessed and that there were no missing data; no actual trial is likely to be so perfect. As above, though, our approach could be extended to account for censoring or for missing data. Additionally, we did not demonstrate here how to obtain confidence intervals for the estimated RRs. When adapting our approach, one way to obtain valid confidence intervals would be to use bootstrap.

Above we discussed the assumptions necessary for our approach. Whether these assumptions are reasonable will be context dependent, and will in part rely on careful study design and data collection. For instance, here we only included the minimal number of variables necessary to demonstrate the approach. When using our approach, a researcher will need to carefully consider the context and question of interest to determine the variable set sufficient to meet the exchangeability assumption. For instance, while we only examined baseline STI, a researcher using our approach might also attempt to capture whether a woman was in a sero-discordant partnership or frequency of condom use, so these could also be used as effect-measure modifiers. Our goal, though, as proof of concept, was simply to give examples of the types of variables one would need.

Furthermore, a particular implication of the no measurement error assumption worth highlighting is the requirement of using a valid measure of adherence in both populations when estimating the per-protocol effect. While we simply used an arbitrary, dichotomized indicator of adherence to protocol, an investigator would ideally have an accurate measure of drug concentration, by which they could ascertain whether a woman received an effective dose of the drug.[30] This information would then guide the formulation of the trial's protocol regarding adherence. If an effective concentration was not known, the investigator could examine different cut points for drug concentration and compare the results of our approach for each cut point.

Some practical implications of our approach should also be noted. In our motivating PrEP example, the high-incidence trials from which we would generalize to the US already exist. Thus, it would save a great deal of time and money to enroll the small US trial and apply our approach to estimate the desired parameters, in comparison to enrolling the 10,000 woman trial and estimate the parameters more directly. However, if our approach were to be used in scenarios where a high incidence study does not exist, this choice would not be so clean cut. The investigators would need to weigh the costs of designing and conducting studies in both settings sequentially or concurrently (as well as the "cost" of making the necessary assumptions for our statistical approach) against the cost of running one larger study in the low incidence setting. This decision will likely be context-dependent, hinging on factors such as how large the study in the low incidence setting would need to be (which would depend on the outcome incidence) and how well the investigators believe they could measure all the necessary variables to carry out our approach.

Despite its limitations, the proposed approach has the potential to be applied to a number of important public health questions. For instance, to return to our motivating example, understanding how well PrEP can prevent HIV incidence among at-risk women in the US is an understudied but critical area of research. Assessment of PrEP efficacy in this population would require an impractically large trial because incidence is as low as 320 infections per 100,000 person–years (compared to 5700/100,000 person–years in Sub-Saharan Africa).[4,5] However, our approach would allow researchers to estimate PrEP effectiveness while enrolling a much smaller trial of US women partly because it does not rely on measuring the outcome in the US.

Moreover, our approach need not be limited to placebo-controlled RCTs in both settings. The strategy could be adapted for studies using an active comparator; such situations would require measuring adherence to the active comparator and accounting for this adherence in the per-protocol effect estimation. Observational studies could also serve as the data source. In an observational study, the ITT effect might compare women who were prescribed PrEP by a physician to a similar group of women who were not prescribed PrEP. Estimation of the per-protocol effect would, as in our example, require an accurate measure of adherence to PrEP. The use of observational data, however, would mean PrEP was no longer randomized, and exchangeability would no longer be expected when estimating the ITT. A researcher would thus need to additionally control for the confounders of being prescribed PrEP and HIV incidence. Accurate measures of all required variables (adherence, confounders, and effect measure modifiers) might also be more difficult to attain than in an RCT.

Thus, the simulations described here, while limited, inform the design of future studies that seek to examine the effectiveness and efficacy of a treatment (like PrEP) not just in higher incidence settings but also in lower incidence settings where a traditional RCT would not be feasible.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Adimora AA, Cole SR, Eron JJ. US Black Women and Human Immunodeficiency Virus Prevention: Time for New Approaches to Clinical Trials. Clin Infect Dis 2017;65(2):324–327. [PubMed: 28383649]

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):2587–99. [PubMed: 21091279]

3. Baeten JM, Donnell D, Ndase P, et al. Antiretroviral prophylaxis for HIV prevention in heterosexual men and women. N Engl J Med 2012;367(5):399–410. [PubMed: 22784037]

4. Hodder SL, Justman J, Hughes JP, et al. HIV acquisition among women from selected areas of the United States: a cohort study. Ann Intern Med 2013;158(1):10–8. [PubMed: 23277896]

5. 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):509–18. [PubMed: 25651245]

6. Van Damme L, Corneli A, Ahmed K, et al. Preexposure prophylaxis for HIV infection among African women. N Engl J Med 2012;367(5):411–22. [PubMed: 22784040]

7. Cain LE, Cole SR. Inverse probability-of-censoring weights for the correction of time-varying noncompliance in the effect of randomized highly active antiretroviral therapy on incident AIDS or death. Stat Med 2009;28(12):1725–38. [PubMed: 19347843]

8. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. Biometrics 2000;56(3):779–88. [PubMed: 10985216]

9. Lu H. Generalizing the adjusted per-protocol treatment effecting using inverse probability weights; Atlantic Causal Inference Conference; Chapel Hill, NC.

10. Corneli AL, Deese J, Wang M, et al. FEM-PrEP: adherence patterns and factors associated with adherence to a daily oral study product for pre-exposure prophylaxis. J Acquir Immune Defic Syndr 2014;66(3):324–31. [PubMed: 25157647]

11. Adimora AA, Hughes JP, Wang J, et al. Characteristics of multiple and concurrent partnerships among women at high risk for HIV infection. J Acquir Immune Defic Syndr 2014;65(1):99–106. [PubMed: 24056163]

12. Justman J, Befus M, Hughes J, et al. Sexual Behaviors of US Women at Risk of HIV Acquisition: A Longitudinal Analysis of Findings from HPTN 064. AIDS Behav 2015;19(7):1327–37. [PubMed: 25626889]

13. Hughes JP, Baeten JM, Lingappa JR, et al. Determinants of per-coital-act HIV-1 infectivity among African HIV-1-serodiscordant couples. J Infect Dis 2012;205(3):358–65. [PubMed: 22241800]

14. Fleming DT, Wasserheit JN. From epidemiological synergy to public health policy and practice: the contribution of other sexually transmitted diseases to sexual transmission of HIV infection. Sex Transm Infect 1999;75(1):3–17. [PubMed: 10448335]

15. Galvin SR, Cohen MS. The role of sexually transmitted diseases in HIV transmission. Nat Rev Microbiol 2004;2(1):33–42. [PubMed: 15035007]

16. Atashili J, Poole C, Ndumbe PM, Adimora AA, Smith JS. Bacterial vaginosis and HIV acquisition: a meta-analysis of published studies. AIDS 2008;22(12):1493–501. [PubMed: 18614873]

17. Pearl J Causal diagrams for empirical research. Biometrika 1995;82(4):669–688.

18. Snowden JM, Rose S, Mortimer KM. Implementation of G-computation on a simulated data set: demonstration of a causal inference technique. Am J Epidemiol 2011;173(7):731–8. [PubMed: 21415029]

19. Hernan MA, Hernandez-Diaz S. Beyond the intention-to-treat in comparative effectiveness research. Clin Trials 2012;9(1):48–55. [PubMed: 21948059]

20. Lodi S, Sharma S, Lundgren JD, et al. The per-protocol effect of immediate versus deferred antiretroviral therapy initiation. AIDS 2016;30(17):2659–2663. [PubMed: 27782964]

21. Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: The ACTG 320 trial. Am J Epidemiol 2010;172(1):107–15. [PubMed: 20547574]

22. Westreich D, Edwards JK, Lesko CR, Stuart E, Cole SR. Transportability of Trial Results Using Inverse Odds of Sampling Weights. Am J Epidemiol 2017;186(8):1010–1014. [PubMed: 28535275]

23. VanderWeele TJ. Concerning the consistency assumption in causal inference. Epidemiology 2009;20(6):880–3. [PubMed: 19829187]

24. Cole SR, Frangakis CE. The consistency statement in causal inference: a definition or an assumption? Epidemiology 2009;20(1):3–5. [PubMed: 19234395]

25. Hernan MA, Robins JM. Estimating causal effects from epidemiological data. J Epidemiol Community Health 2006;60(7):578–86. [PubMed: 16790829]

26. Lesko CR, Buchanan AL, Westreich D, et al. Generalizing Study Results: A Potential Outcomes Perspective. Epidemiology 2017;28(4):553–561. [PubMed: 28346267]

27. Westreich D, Cole SR. Invited commentary: positivity in practice. Am J Epidemiol 2010;171(6):674–7; discussion 678-81. [PubMed: 20139125]

28. 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. [PubMed: 21030422]

29. Robins J, Hernan MA. Estimation of the causal effects of time-varying exposures. In: Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G, eds. Longitudinal Data Analysis. Boca Raton, FL: Chapman & Hall/CRC Taylor & Francis Group, LLC, 2009;553–597.

30. Cottrell ML, Yang KH, Prince HM, et al. A Translational Pharmacology Approach to Predicting Outcomes of Preexposure Prophylaxis Against HIV in Men and Women Using Tenofovir Disoproxil Fumarate With or Without Emtricitabine. J Infect Dis 2016;214(1):55–64. [PubMed: 26917574]
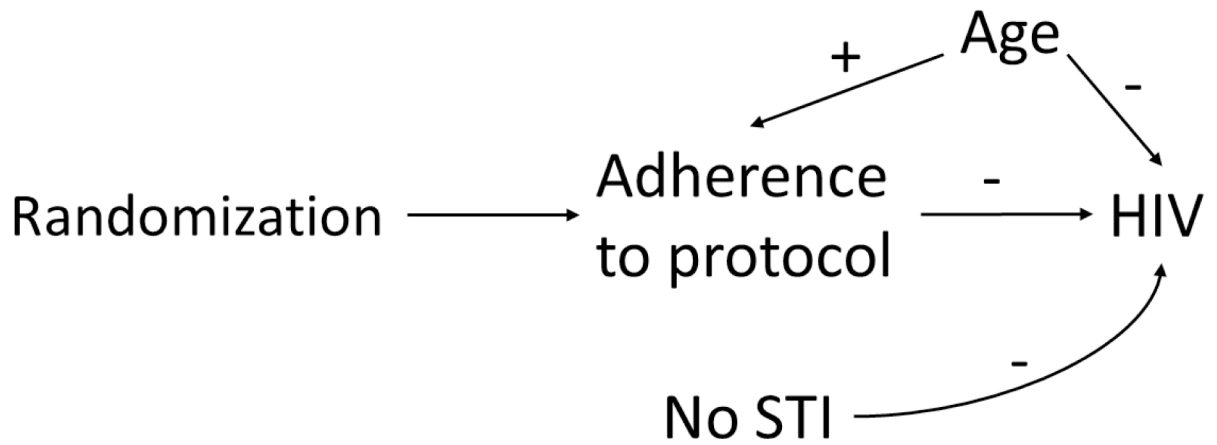
Figure 1.
Causal diagram used in simulations, showing relations between randomization to pre-exposure prophylaxis, adherence to study protocol, lack of sexually transmitted infection (STI) at baseline (an example effect measure modifier), age >21 years (an example confounder), and HIV seroconversion. In our simulation, older age was positively associated with adherence to protocol; older age, adherence, and the interaction between lack of an STI and adherence were negatively associated with HIV seroconversion.
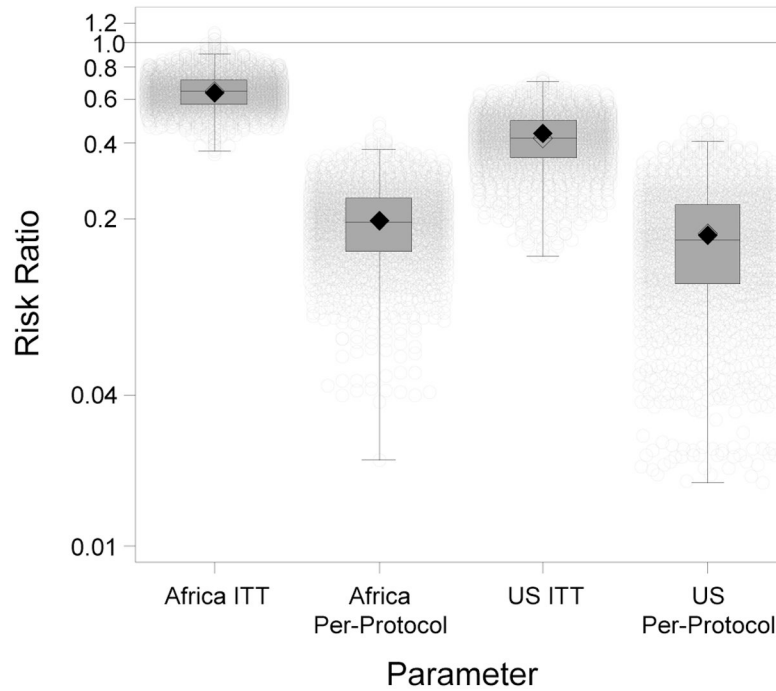
Figure 2.
Boxplots of the simulation risk ratios with individual simulation estimates for all 4 parameters estimated using g-computation. The filled diamonds are the true risk ratios, and the unfilled diamonds are the estimated risk ratios. The boxes represent the interquartile range, with the median risk ratio being the line within the box. Each of the light gray circles represents the result from a single simulation. The breadth of the whiskers is 1.5 times the interquartile range. Extreme outliers were removed from the picture (for the US ITT, 15 risk ratios $<1\times10^{-4}$ and, for the US per-protocol, 15 risk ratios $<1\times10^{-10}$).

Table 1.

Distributions of key variables across 2000 simulations in the higher HIV incidence African setting and lower incidence US setting

Variable	African Trials (trial n=4000)		US Trials (trial n=500)	
	Median (%)	IQR (%)	Median (%)	IQR (%)
Randomization to PrEP	50.0	49.5, 50.5	50.0	48.6, 51.4
Receipt of PrEP	24.9	24.4, 25.4	37.0	35.4, 38.4
Incident HIV	4.2	4.0, 4.4	0.4	0.2, 0.8
Age >21 years[a]	33.3	32.9, 33.9	66.6	65.4, 68.2
No STI at baseline[b]	79.0	78.6, 79.4	89.0	88.2, 90.0

Abbreviations: HIV, human immunodeficiency virus; US, United States; n: sample size; IQR, interquartile range; PrEP, pre-exposure prophylaxis; STI: sexually transmitted infection (other than herpes simplex virus)

[a] African age distribution based on references 6 and 10. US age distribution based on references 4 and 11.

[b] African distribution based on reference 10, and US distribution based on reference 12.

Table 2.

Intention-to-treat and per-protocol effects in the higher HIV incidence African setting and generalized to the lower incidence US setting, estimated using g-computation

Parameter	African Trials (trial n=4000)			US/African Composite Trials (trial n=500/4000)			US Trial (n=10,000)		
	RR	2.5th	97.5th	RR	2.5th	97.5th	RR	2.5th	97.5th
Intention to treat	0.65	0.47	0.88	0.42	0.20	0.62	0.41	0.19	0.68
Per protocol	0.20	0.08	0.34	0.17	0.03	0.38	0.13	0.00	0.31

Abbreviations: HIV, human immunodeficiency virus; US, United States; n: sample size; RR, average risk ratio; 2.5th: 2.5th percentile of the 2000 simulation RRs or, for the US trial, of 500 bootstrap resamples; 97.5th: 97.5th percentile of the 2000 simulation RRs or, for the US trial, of 500 bootstrap resamples