Estimating the Effect of PrEP in Black Men Who Have Sex with Men: A Framework to Utilize Data from Multiple Non-Randomized Studies to Estimate Causal Effects

Allison Meisner[1], Fan Xia[2], Kwun C.G. Chan[3], Kenneth Mayer[4,5,6,7], Darrell Wheeler[8], Sahar Zangeneh[9,10], Deborah Donnell[11,12]

[1]Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, WA, US
[2]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA, US
[3]Department of Biostatistics, University of Washington, Seattle, WA, US
[4]Harvard Medical School, Boston, MA, US
[5]Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, MA, US
[6]The Fenway Institute, Boston, MA, US
[7]Infectious Diseases Division, Beth Israel Deaconess Medical Center, Boston, MA, US
[8]State University of New York at New Paltz, New Paltz, NY, US
[9]RTI International, Research Triangle Park, NC, US
[10]School of Public Health, University of Washington, Seattle, WA, US
[11]Vaccine and Infectious Diseases Division, Fred Hutchinson Cancer Center, Seattle, WA, US
[12]Department of Global Health, University of Washington, Seattle, WA, US

Address:
1100 Fairview Ave N, M4-B829
Seattle, WA, US 98109
Email: ameisner@fredhutch.org

Running head: Causal Effect of PrEP in Black MSM

ABSTRACT

Black men who have sex with men (MSM) are disproportionately burdened by the HIV epidemic in the US. The effectiveness of pre-exposure prophylaxis (PrEP) in preventing HIV infection has been demonstrated through randomized placebo-controlled clinical trials in several populations. Importantly, no such trial has been conducted exclusively among Black MSM in the US, and it would be unethical and infeasible to do so now. To estimate the causal effects of PrEP access, initiation, and adherence on HIV risk, we utilized causal inference methods to combine data from two non-randomized studies that exclusively enrolled Black MSM. The estimated relative risks of HIV were: (i) 0.52 (95% confidence interval: 0.21, 1.22) for individuals with versus without PrEP access, (ii) 0.48 (0.12, 0.89) for individuals who initiated PrEP but were not adherent versus those who did not initiate, and (iii) 0.23 (0.02, 0.80) for individuals who were adherent to PrEP versus those who did not initiate. Beyond addressing the knowledge gap around the effect of PrEP in Black MSM in the US, which may have ramifications for public health, we have provided a framework to combine data from multiple non-randomized studies to estimate causal effects, which has broad utility.
Keywords: causal inference, non-randomized trials, HIV, pre-exposure prophylaxis (PrEP), Black men who have sex with men (MSM)

INTRODUCTION

The HIV epidemic, now in its fifth decade, has been concentrated in specific key populations since it began. One group that has been disproportionately affected since the very start of the epidemic is men who have sex with men (MSM). In the US, MSM accounted for 71% of all new HIV diagnoses in 2020 [1]. Among MSM in the US, Black men are at particular risk, accounting for 26% of all new HIV diagnoses and 39% of diagnoses among MSM in 2020 [1] despite Black people comprising only 13.6% of the US population [2]. It is believed that Black MSM are at high risk of HIV due to a higher prevalence of sexually transmitted infections (STIs), higher levels of unrecognized HIV in their sexual networks (which has downstream effects such as delayed initiation of anti-retroviral therapy, a treatment which can reduce the risk of transmission of HIV), sex partner demographics, and a constellation of marginalizing factors, including lower income, underemployment, educational inequalities, inadequate access to healthcare, incarceration, stigma, and discrimination [3-10].

The high risk and disproportionate burden of HIV among Black MSM have persisted even after the introduction of pre-exposure prophylaxis, or PrEP, which was approved by the Food and Drug Administration in 2012. In support of this approval, and in the years since, PrEP has been shown to be highly effective in preventing HIV infection in a variety of populations, including MSM [11-18]. However, while previous trials have included Black MSM in the US, they have not been sufficiently represented in randomized, placebo-controls trials to date. Consequently, little is known about the effectiveness of PrEP specific within the unique context of this key population. Given the extensive evidence that PrEP is effective in other populations, it is not ethical to randomize Black MSM to placebo. Consequently, a knowledge gap regarding the effectiveness of PrEP in this vulnerable population persists. Quantifying the effectiveness of PrEP in this population could provide insights into the generalizability of effectiveness estimates from other populations, highlighting factors which may modify the effect of PrEP. From a public health perspective, an estimate of effectiveness specific to Black MSM may motivate Black MSM to initiate and adhere to PrEP and may encourage their clinicians to support such behavior, leading to reductions in HIV incidence.

We sought to estimate the effect of PrEP on HIV risk in Black MSM in the US by combining data from two large studies conducted exclusively in Black MSM by the HIV Prevention Trials Network (HPTN), including a single-arm study and an observational study. More precisely, since engagement with PrEP is a multi-stage “cascade”, we estimated the effect of PrEP on HIV risk at three different stages of the cascade, namely, the effect of PrEP access, the effect of PrEP initiation without adherence, and the effect of PrEP adherence. In order to estimate a causal effect from non-randomized data while also leveraging data from the single-arm study, we utilized methods from causal inference, including inverse probability weighting and marginal structural models [19, 20]. Consequently, beyond providing estimates of the effectiveness of PrEP in reducing risk of HIV in Black MSM in the US, our work provides a more general framework to estimate causal effects using data from multiple, non-randomized studies.

METHODS

Data Sources

Data from two studies in HPTN were used in this analysis: HPTN 061 and HPTN 073. Full details on HPTN 061 and HPTN 073 can be found in [21] and [22], respectively.

Briefly, HPTN 061 was an observational study of Black MSM in several cities in the US which sought to evaluate the feasibility and acceptability of a multi-component HIV prevention intervention [21]. Men were enrolled between July 2009 and October 2010, with follow-up completed in 2011; thus, we assumed that there was functionally no access to PrEP for men in this study as the study finished prior to PrEP approval. While the study enrolled men with and without HIV, we focused on the 1,162 men who did not have HIV at enrollment and completed the baseline questionnaire. Twenty-eight participants were known to have acquired HIV infection during 12 months of follow-up [21]. The Supplement includes additional information on the design of HPTN 061.

HPTN 073 was a non-randomized, open-label PrEP study conducted among Black MSM in several US cities [22]. The study was conducted between 2013 and 2015, i.e., immediately after FDA approval of PrEP, but close in time to HPTN 061. Consequently, all men enrolled in HPTN 073 had access to PrEP through participation in the study (daily oral co-formulated emtricitabine and tenofovir disoproxil fumarate (FTC/TDF)) and were free to decide whether to initiate and/or adhere to PrEP at any point. Importantly, access to PrEP in HPTN 073 was of an “active” form: beyond PrEP being approved by the FDA and thus theoretically available (i.e., “passive access”), participants in HPTN 073 had frequent study visits wherein PrEP was offered and study staff were available to assist participants if they expressed an interest in initiating PrEP. Hereafter, this will only be referred to as “access” to PrEP, but it is important to be mindful of its active nature. A total of 226 men without HIV were enrolled and followed for one year. For those men in HPTN 073 who decided to initiate PrEP (nearly 79%), the date of first dose was recorded. Blood samples were subsequently taken to measure PrEP adherence and subjects were defined as adherent if they met the threshold for at least 4 doses of PrEP per week, i.e., the individual had either (a) tenofovir (TFV) levels of at least 4.2 ng/mL and FTC levels of at least 4.6 ng/mL in plasma or (b) TFV diphosphate levels of at least 9.9 fmol/10^6 and FTC triphosphate levels of at least 0.4 fmol/10^6 in peripheral blood mononuclear cells (PMBCs) [23]. Eight participants were known to have acquired HIV infection during 52 weeks of follow-up [22]. The Supplement includes additional information on the design of HPTN 073.

In order to use data from both HPTN 061 and HPTN 073, the cohorts, timing of measurements, and variables assessed needed to be as similar as possible. Measurements (including those of relevant covariates and HIV status) were more frequent in HPTN 073 than in HPTN 061 (every

three months vs. every six months), so we utilized the data from the enrollment, 6-month, and 12-month visits in both studies. We also restricted attention to only those variables measured in both studies, though the studies were conducted in a similar time period by the same trial network, so essentially all key variables were measured in both studies. A small number of participants (n = 28) from the HPTN 061 sample who did not satisfy the eligibility criteria for HPTN 073, either because they were transgender men or because they reported having exclusively transgender male partners in the previous 6 months at baseline, were excluded. Discrepancies in inclusion criteria between the two studies were expected to lead to differences in the distribution of some covariates between the two studies; such imbalances were typically accommodated by weighting [24, 25], as discussed below; the remaining discrepancies could not be resolved but were expected to have little effect (see the Discussion section and the Supplement).

Given that the estimates and inference resulting from our analysis will depend to some degree on the distribution of covariates when there is effect modification [26, 27] and/or unmeasured confounding, differences in covariate distributions may lead to issues with interpretability and/or generalizability if the relevant populations are not well-defined [28]. Using the nomenclature of Degtiar and Rose [27], we define our *target population*, i.e., that in which we would like to make inference, as Black MSM in the US. Our *study population*, in which the effect will actually be estimated and inference made, is the HPTN 073 population, i.e., the population from which the HPTN 073 study participants were (randomly) sampled. We chose the HPTN 073 population to be our study population since it is the more recent study. Our analysis is based on two study samples, the HPTN 073 sample, which is representative of the HPTN 073 population, and the HPTN 061 sample, which is representative of the HPTN 061 population. In general, study populations are defined by enrollment processes and eligibility criteria, e.g., all Black MSM living in certain US cities at the time of the study and satisfying all eligibility criteria; study samples are then (randomly) drawn from these populations [27].

Analysis

The goal of this analysis was to estimate the effects of PrEP access, initiation without adherence (“initiation”), and adherence on HIV risk in Black MSM in the US. Note that we were interested in estimating the marginal (total) effect of each of these exposures, i.e., not conditional on any covariates, consistent with the effect that would be estimated in a

randomized trial. Let A denote PrEP access ($A = 1$ indicates PrEP access and $A = 0$ indicates no PrEP access). We use the variable D to denote PrEP initiation and adherence: $D = 0$ if the participant has not initiated (and thus is not adherent), $D = 1$ if the participant has initiated but is not adherent, and $D = 2$ if the participant has initiated and is adherent. For all HPTN 061 study participants, $A = D = 0$. For all participants in HPTN 073, $A = 1$ and D was 0, 1, or 2. We further define Y to be HIV status ($Y = 1$ if the participant has HIV and $Y = 0$ if they do not) and $\boldsymbol{X}$ is a vector of baseline and time-varying covariates.

Recall that both studies have data collected at 0, 6, and 12 months; these time points are denoted by $t = 0, 1, 2$, respectively. We conceptualize the time between 0 and 6 months as the "first interval" and the time between 6 and 12 months as the "second interval". Access, which is time-invariant, was evaluated at the start of the study ($A = 0$ for HPTN 061 and $A = 1$ for HPTN 073). The covariates $\boldsymbol{X}_t$ denote the vector of covariate values measured at time t (including both time-varying and time-invariant covariates). Likewise, D_t denotes PrEP initiation/adherence assessed at time t, $t = 1, 2$; e.g., $D_t = 1$ if the participant initiated PrEP prior to time t but was not adherent at time t. Finally, HIV status was assessed at $t = 1$ and $t = 2$, giving rise to Y_1 and Y_2, respectively ($Y_0 = 0$ for all participants). Note that $D_1 > 0 \Rightarrow D_2 > 0$; in other words, if a participant initiated prior to the 6-month visit ($D_1 = 1$ or 2), then that participant also initiated before the 12-month visit ($D_2 = 1$ or 2). We will also use the subscript i to refer to values of the exposures, covariates, and outcomes for individual i, $i = 1, \ldots, n$, where $n = n_{061} + n_{073}$, the total number of participants in the HPTN 061 and HPTN 073 samples.

A directed acyclic graph (DAG) representing the relationships between the variables is given in Figure 1. In this figure, dashed lines are drawn from A to D_1 and D_2 because there is a (partially) deterministic relationship at play: if $A = 0$, then $D_1 = D_2 = 0$. Likewise, there are dashed lines drawn from Y_1 to D_2 and Y_2 because there is a (partially) deterministic relationship between these variables as well: if $Y_1 = 1$, then $Y_2 = 1$ and D_2 is not relevant (a person would not keep taking PrEP if they had previously tested positive for HIV). Importantly, we suppose that PrEP access (A) has no direct effect on HIV risk (Y_1 and Y_2); thus, we suppose that the effect of access is only through PrEP initiation and adherence. Any effect of PrEP access on other factors related to HIV risk is assumed to be extremely small. Likewise, there has been some debate as to whether taking PrEP influences the probability of engaging in higher risk behaviors (which may lead to an arrow from D_1 to $\boldsymbol{X}_1$ in Figure 1, since $\boldsymbol{X}_1$ likely includes some of these

behaviors); however, the preponderance of evidence indicates that any effect is of minimal magnitude [29, 30]. Thus, we are assuming there is no causal link between D_1 and $\boldsymbol{X}_1$.

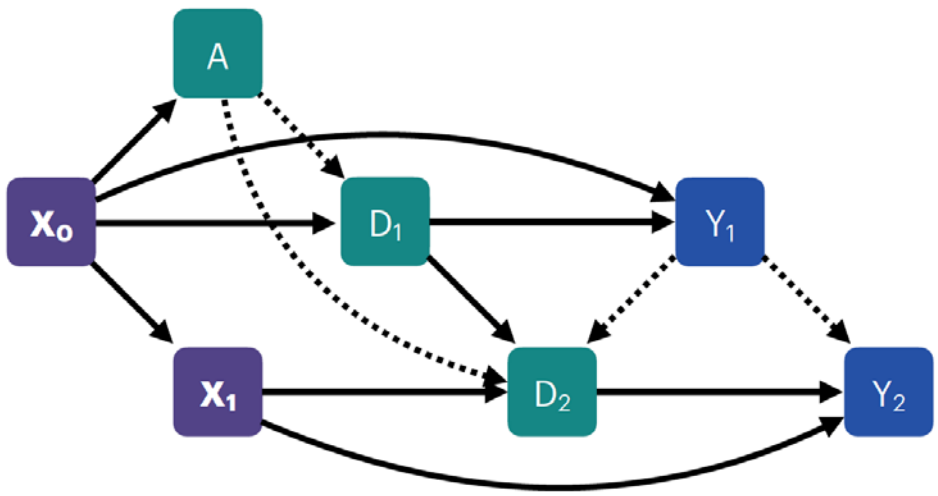


Figure 1: DAG representing the relationships among the exposures (A, D_1, D_2), the outcomes (Y_1, Y_2), and the covariates ($\boldsymbol{X}_0, \boldsymbol{X}_1$). Dashed lines indicate deterministic relationships.

<u>A Analysis</u>

We first describe the estimation of the effect of A on HIV risk. We are interested in estimating the effect of access on risk of HIV within 1 year (Y_2). Thus, we can consider the causal estimand

$$\frac{E[Y_2(a=1)|A=1]}{E[Y_2(a=0)|A=1]}, \tag{1}$$

where $Y_2(a)$ is the potential outcome of Y_2 under exposure $A = a$, i.e., the outcome that would have been observed had $A = a$ [31]. The estimand in (1) can be interpreted as the causal relative risk (RR) of HIV infection in one year comparing PrEP access versus no access in the HPTN 073 population, i.e., conditioned on $A = 1$, which can also be viewed as an average treatment effect in the treated (ATT) [32].

One approach to estimating the quantity in (1) is through inverse probability weighting, which uses weights to balance the distribution of covariates between the HPTN 061 sample and the HPTN 073 sample, thereby mimicking a randomized trial in which the only (systematic) difference between the exposure groups is the exposure itself [33]. While the ideal would be to make the HPTN 061 sample similar to the HPTN 073 sample on all covariates, in reality, we were not able to randomize and thus could only achieve similarity for the measured covariates [34]. Here, the goal of weighting was to make the distribution of covariates in the HPTN 061

sample similar to the distribution of covariates in the HPTN 073 sample because the HPTN 073 population was our target of inference. In other words, we used weights to obtain a sample from a "pseudopopulation" in which the distribution of the measured baseline covariates in both the exposed ($A = 1$) and the unexposed ($A = 0$) reflects the distribution of the covariates in the HPTN 073 population [20, 32, 35, 36], removing the potential for confounding by these variables [20, 36].

The estimand in (1) can be estimated using the HPTN 061 sample and the HPTN 073 sample with weights of the form [32, 37]

$$w_i^A = \begin{cases} \frac{\Pr(A = 1 | \boldsymbol{X}_0 = \boldsymbol{x}_{i0})}{\Pr(A = 0 | \boldsymbol{X}_0 = \boldsymbol{x}_{i0})} & \text{for HPTN 061} \\ 1 & \text{for HPTN 073} \end{cases},$$

$i = 1, \ldots, n$. Using the weights w_i^A, (1) was estimated by fitting a weighted log-linear model of the form

$$\log[\Pr(Y_2 = 1 | A)] = \beta_0 + \beta_1 A. \tag{2}$$

In order to use the inverse probability weights w_i^A to estimate (1), some key assumptions are required [38, 39] and were used to derive the weights w_i^A (see Supplement for details on the derivation):

(a) Conditional exchangeability: $Y_2(a = 0) \perp A \mid \boldsymbol{X}_0$ (assumption A1)

(b) Positivity: $\Pr(A = 0 | \boldsymbol{X}_0 = \boldsymbol{x}_0) > 0 \; \forall \, \boldsymbol{x}_0$ (assumption A2)

The assumption of conditional exchangeability (A1) means that conditional on the covariates $\boldsymbol{X}_0$, the distribution of $Y_2(a = 0)$ is the same in the HPTN 061 population and in the HPTN 073 population [40, 41]. This allows us to use the observed outcomes for $A = 0$ (i.e., Y_2 in the HPTN 061 population) in place of the potential outcomes $Y_2(a = 0)$ for $A = 1$, permitting estimation of (1) [42]. The idea behind (A1) is that it allows us to use the weights w_i^A to understand what would have happened to the HPTN 073 population (in terms of HIV risk) had there been no access to PrEP [41]. Assumption (A1) can also be thought of as an assumption of no unmeasured confounding of the association between PrEP access and HIV risk; essentially, we allowed for confounding by $\boldsymbol{X}_0$ but not by any other factors (see DAG in Figure 1) [41, 43, 44]. The idea is that the weights aimed to provide balance on the measured covariates, and (A1) implies unmeasured covariates were not relevant [38]. We do not need to make a similar assumption for $Y_2(a = 1)$ because these potential outcomes are observed in the HPTN 073 population, the target of inference [42].

The assumption of positivity (A2) makes $Y_2(a = 0)$ realizable for all participants and thus allows us to estimate $E(Y_2(a = 0)|A = 1)$ [40, 42]; note that $E(Y_2(a = 1)|A = 1)$ can be estimated using the HPTN 073 sample, provided consistency (discussed below) holds. As with (A1), we did not need to make an assumption about $\Pr(A = 1|\boldsymbol{X}_0)$ because (in theory) it would be permissible to include some HPTN 061 participants who were ineligible for HPTN 073 though they would not contribute to estimation [42] (though in reality the presence of such individuals would lead to difficulties in estimating $\Pr(A = 1|\boldsymbol{X}_0)$). The assumption (A2) essentially requires that all members of the HPTN 073 population are represented by individuals in the HPTN 061 population [27]. Near-violations of (A2) may also present difficulties, as this could mean that massive weights are given to some observations, which would then have an outsized influence on the results, potentially leading to bias, instability, and increased uncertainty [26, 34, 36, 45, 46]. Importantly, issues with positivity (violations of (A2)) are not anticipated as the HPTN 073 and HPTN 061 populations are quite similar.

While (A1) and (A2) were the key assumptions, we also assumed consistency ($A = a \Rightarrow Y_2 = Y_2(a)$ for a particular subject) [36, 39] and the stable unit treatment value assumption (SUTVA), which supposes that there were not multiple versions of exposure (i.e., the exposures A = 0 and A = 1 are well-defined and homogeneous) and that each individual's potential outcomes were independent of the exposure status of other individuals, i.e., there was no interference [47]. Consistency generally holds if the recorded exposure (in this case, A) reflects the exposure the participant actually experienced [48].

To obtain w_i^A, an estimate of $\Pr(A = 1|\boldsymbol{X}_0)$ is needed. Given that the goal of the weights was to make the HPTN 061 sample as similar as possible to the HPTN 073 sample in terms of the distribution of the covariates, we used the covariate-balancing propensity score (CBPS), which estimates $\Pr(A = 1|\boldsymbol{X}_0)$ by modeling treatment assignment while optimizing (mean) covariate balance [49, 50]. Our implementation relied on the CBPS package in R [51, 52].

<u>D Analysis</u>

Next, we describe the estimation of the effect of PrEP initiation and adherence on HIV risk. Recall that initiation and adherence (D) is a time-varying variable, where the value during the

first time interval is D_1 and the value during the second time interval is D_2. Thus, our causal estimand can be written as

$$\frac{E[Y_t(d_t)|A = 1, Y_{t-1} = 0]}{E[Y_t(d_t = 0)|A = 1, Y_{t-1} = 0]}, \tag{3}$$

where $t = 1, 2$, $d_t = 1, 2$. Here, the potential outcome $Y_t(d_t)$ is the outcome an individual would have experienced (in this case HIV infection by visit t) had they received exposure d_t during the prior time interval. The estimand in (3) yields the RR of HIV infection during the next six months for (i) individuals who had initiated PrEP but were not adherent at time t vs. those who had not yet initiated PrEP ($d_t = 1$ vs. 0) and (ii) individuals who were adherent to PrEP at time t vs. those who had not yet initiated PrEP ($d_t = 2$ vs. 0) [53]. We conditioned on $Y_{t-1} = 0$ because individuals who already tested positive for HIV were no longer at risk, and thus excluded from the population (i.e., the risk set). The estimand in (3) conditions on $A = 1$ because the HPTN 073 population was our target of inference.

Given that we wanted to (i) use data from HPTN 073, a non-randomized study, and (ii) incorporate data from HPTN 061, a single-arm study in which PrEP was not available, two sets of weights were required. The first set of weights was used to make the distribution of the covariates in the HPTN 061 population similar to that in the HPTN 073 population:

$$w_{it}^{D,1} = \begin{cases} \frac{\Pr(A = 1|\boldsymbol{X}_0 = \boldsymbol{x}_{i0}, Y_{t-1} = 0)}{\Pr(A = 0|\boldsymbol{X}_0 = \boldsymbol{x}_{i0}, Y_{t-1} = 0)} & \text{for HPTN 061} \\ 1 & \text{for HPTN 073} \end{cases},$$

$t = 1, 2, i = 1, \ldots, n$. These weights are quite similar to w_i^A, but are time-varying since they condition on $Y_{t-1} = 0$ [36]. As with w_i^A, the (time-varying) weights $w_{it}^{D,1}$ yield a sample from a pseudopopulation in which the distribution of measured baseline covariates within both $A = 1$ and $A = 0$ reflects that in the HPTN 073 population (among individuals who have not yet tested positive for HIV at time t). By reweighting the data in the HPTN 061 sample to have the same distribution of baseline covariates as the HPTN 073 population, we (i) address confounding of the D_t-Y_t association due to the causal association between $\boldsymbol{X}_0$ and A and (ii) allow estimation of the effect of D in the HPTN 073 population, the target of inference (see DAG in Figure 1) [42].

The second set of weights was used to achieve balance in the distribution of measured covariates for $D = 0$, 1, and 2 in the HPTN 073 sample. These weights have the form

$$w_{it}^{D,2} = \begin{cases} 1 & \text{for HPTN 061} \\ \frac{1}{\Pr\left(D_t = d_{it}\middle|A = 1, \boldsymbol{X}_{t-1} = \boldsymbol{x}_{i,t-1}, Y_{t-1} = 0\right)} & \text{for HPTN 073} \end{cases},$$

where $i = 1, \ldots, n, t = 1, 2$. These inverse probability weights yield a sample from a pseudopopulation in which the distribution of the measured covariate history is independent of D_t, and the distribution of the covariates in each level of D_t is the same as that in the entire HPTN 073 population. The weights $w_{i2}^{D,2}$ for the HPTN 073 sample address confounding of the D_2-Y_2 association by $\boldsymbol{X}_1$ and do not need to include $\boldsymbol{X}_0$ because $Y_2(d_2)$ and D_2 are independent conditional on $\boldsymbol{X}_1$ and $Y_1 = 0$. Such an independence is implied by the DAG in Figure 1 [54] and explicitly described in assumption B3 below.

The final weights for the analysis of PrEP initiation and adherence were obtained by multiplying $w_{it}^{D,1}$ and $w_{it}^{D,2}$, yielding

$$w_{it}^{D} = \begin{cases} \dfrac{\Pr\left(A = 1 | \boldsymbol{X}_0 = \boldsymbol{x}_{i0}, Y_{t-1} = 0\right)}{\Pr\left(A = 0 | \boldsymbol{X}_0 = \boldsymbol{x}_{i0}, Y_{t-1} = 0\right)} & \text{for HPTN 061} \\ \dfrac{1}{\Pr\left(D_t = d_{it} | A = 1, \boldsymbol{X}_{t-1} = \boldsymbol{x}_{i,t-1}, Y_{i,t-1} = 0\right)} & \text{for HPTN 073} \end{cases},$$

where $i = 1, \ldots, n, t = 1, 2$, and $d_t = 0, 1, 2$. The weights create a sample from a pseudopopulation in which, at each time point $t = 1, 2$ and among individuals at risk for HIV, both (i) the distribution of the covariates $\boldsymbol{X}_0$ within $A = 0$ and (ii) the distribution of the covariates $\boldsymbol{X}_{t-1}$ within $A = 1$ for each level of D_t are equivalent to the relevant covariate distributions in the HPTN 073 population, the target of inference [42, 55, 56]. Thus, the weights address all sources of measured confounding of the association between D and HIV risk per the DAG in Figure 1. As above, we used CBPS to estimate $\Pr\left(A_i = 1 \middle| \boldsymbol{X}_{i0}, Y_{i,t-1} = 0\right), t = 1, 2$ and $\Pr\left(D_{it} = d_{it} \middle| A_i = 1, \boldsymbol{X}_{t-1}, Y_{i,t-1} = 0\right), d_t = 0, 1, 2, t = 1, 2$, using separate models for each time point (i.e., four models were fit). While the levels of D have a natural ordering, we modeled D as a nominal variable when implementing CBPS, which allows for greater flexibility with little loss of information given that D has only three levels.

Using the weights w_{it}^{D}, (3) was estimated by fitting a weighted log-linear model [57] among individuals who have not yet tested positive for HIV, similar to a pooled logistic regression model [36, 58]:

$$\log[\Pr(Y_t = 1 | D_t, Y_{t-1} = 0)] = \beta_{0t} + \beta_1 D_t. \tag{4}$$

The model presented in (4) is a simplified version of a marginal structural model, which parameterizes the marginal causal effect of a time-varying exposure on a particular outcome as a marginal contrast in potential outcomes [20, 59]. Marginal structural models are often used

alongside inverse probability weighting to account for time-varying covariates, addressing confounding while avoiding conditioning [20, 36, 60].

The weights w_{it}^{D} allowed us to estimate the causal effect of D in the HPTN 073 population, using data from both the HPTN 061 sample and the HPTN 073 sample, provided several key assumptions hold [38, 55] and were used to derive the weights w_{it}^{D} (see Supplement for details on the derivation):

(a) Conditional exchangeability for A: $\{Y_t(a=0) \perp A | \boldsymbol{X}_0, Y_{t-1}=0\}$ for $t=1,2$ (assumption B1)
(b) Positivity for A: $\Pr(A=0|\boldsymbol{X}_0=\boldsymbol{x}_0, Y_{t-1}=0) > 0 \; \forall \; \boldsymbol{x}_0$ for $t=1,2$ (assumption B2)
(c) Conditional exchangeability for D: $\{Y_t(d_t) \perp 1(D_t=d_t)|\boldsymbol{X}_{t-1}, A=1, Y_{t-1}=0\}$ for $d_t \in \{0,1,2\}, t=1,2$ (assumption B3)
(d) Positivity for D: $0 < \Pr(D_t=d_t|A=1, \boldsymbol{X}_{t-1}=\boldsymbol{x}_{t-1}, Y_{t-1}=0) < 1 \; \forall \; \boldsymbol{x}_{t-1}$, for $d_t \in \{0,1,2\}$, $t=1,2$ (assumption B4)

The assumption (B1) serves the same role as the assumption (A1) in the A analysis; that is, it allows us to use the observed outcomes for $A=0$ (i.e., Y_t in the HPTN 061 population) in place of the potential outcomes $Y_t(a=0)$ for $A=1$ (after weighting) [42]. As A has no direct effect on Y_t (see Figure 1) and $A=0 \Rightarrow D_t=0, t=1,2$, we have $Y_t(a=0)=Y_t(d_t=0)$, so we can write (B1) as $\{Y_t(d_t=0) \perp A|\boldsymbol{X}_0, Y_{t-1}=0\}$, allowing us to use the observed outcomes in the HPTN 061 sample as potential outcomes $Y_t(d_t=0)$ for $A=1$, that is, in the HPTN 073 population. This also implies that the denominator of (3) is the risk of HIV in the absence of PrEP initiation, regardless of whether there was access to PrEP. Likewise, the assumption of positivity for A (B2) makes $Y_t(d_t=0)$ realizable for all participants which allows estimation of $E\{Y_t(d_t=0)|A=1\}$ [42], mirroring assumption (A2).

We further assume exchangeability for D (B3), which allows us to use the observed outcomes for one exposure group (i.e., level of D) in place of the potential outcomes for another exposure group within the HPTN 073 population, thus making the causal effect identifiable [42, 61]. This is an assumption of no unmeasured confounding of the association between PrEP initiation/adherence and HIV risk [43, 44]. Lastly, we make an assumption of positivity for D (B4); in particular, we must assume that all individuals in the HPTN 073 population had the potential to have any value of D. This assumption makes each potential outcome $Y_t(d_t)$, $d_t=0,1,2, t=1,2$ realizable for all participants in the HPTN 073 population and so allows us to

estimate the causal effect (3) [40, 42]. As with (A2), if $\Pr(D_t = d_t) \approx 0$ for the observed value d_t for a given observation, then that observation will have an extremely large weight, and could unduly influence the results and decrease precision. While (B1-B4) were the key assumptions, as above we must also assume consistency ($D_t = d_t \Rightarrow Y_t = Y_t(d_t)$ for a particular subject) and SUTVA.

In addition to the assumptions enumerated above for the analyses of A and D, we must assume (i) no model misspecification in (2) and (4), (ii) no model misspecification in the exposure models for $\Pr(A = 0 | \boldsymbol{X}_0, Y_{t-1} = 0)$ and $\Pr(D_t = d_t | A = 1, \boldsymbol{X}_{t-1}, Y_{t-1} = 0)$, $t = 1, 2$, and (ii) no measurement error in the exposures, covariates, and outcome [36, 41, 62].

Covariate Selection, Missing Data, and Quantifying Uncertainty

For inverse probability weights, all covariates related to both the outcome and the exposure should be included [63], though the inclusion of covariates only related to the outcome may increase efficiency without incurring bias [64-66]. Furthermore, including too many covariates can increase the likelihood of positivity violations [26, 45, 67]. Thus, we focused only on covariates that are believed or known to be causes of HIV [43, 66]. As we were utilizing data from two well-designed and carefully conducted studies, we were not particularly limited by data availability.

The risk of infection with HIV in any population, including Black MSM in the US, is a function of an individual's sexual network and the frequency of condomless sexual encounters, and may be augmented by cofactors such as concomitant STIs. These features correspond to known risk factors for HIV [21, 68] and are plausible causal agents given what is known about HIV transmission in this population. Thus, we chose to include region, age, number of partners (0-1, 2-3, 4 or more), sexual orientation (gay vs. other), condomless receptive anal intercourse (URAI), and the presence of a sexually transmitted infection (STI), each of which captures a dimension of the above features. Number of partners, URAI, and STI were time-varying covariates, while region, age, and sexual orientation were time-invariant. Region was modeled as a categorical (factor) variable, age was modeled as a continuous variable, number of partners was modeled as an ordinal variable, and the remainder were binary variables.

Missing data were present in our sample, driven largely by missingness in the covariates listed above. Loss to follow-up led to missing data on HIV status for a small fraction of HPTN 073 participants and a larger proportion of HPTN 061 participants. PrEP adherence, which was measured at the six- and twelve-month visits, was available for most individuals who had initiated PrEP. For those for whom it was not available, we attempted to use information on PrEP holds and discontinuations, which were recorded in HPTN 073, to impute the missing values. For all remaining missing data, multiple imputation with chained random forests and predictive mean matching [69] with 10 imputations was implemented using the missRanger package in R [70] in order to include all enrolled participants in the analysis. This approach to missingness assumes the data are missing at random (that is, missingness is random, conditional on observed data) and that the imputation model is correctly specified [71, 72]. The imputation models included the exposures, outcome (HIV status), and covariates used in the main analysis, as well as other variables associated with the exposures and/or outcome (see the Supplement for more details).

To quantify uncertainty in our results, we generated 95% confidence intervals (CIs) using the bootstrap, as has been recommended in this setting [73-75]. In particular, we generated 1000 bootstrap samples, stratified by study (HPTN 061 vs. HPTN 073). We performed multiple imputation in each sample and averaged over the imputations (on the log scale) to obtain a single estimate in each bootstrap sample [76]. Based on these 1000 estimates, we estimated percentile bootstrap intervals (again on the log scale).

RESULTS

Effect of PrEP Initiation and Adherence

Table 1 summarizes the key characteristics of the HPTN 061 and HPTN 073 samples. There were 1,134 individuals in the HPTN 061 sample (after removing those ineligible for HPTN 073) and 226 individuals in the HPTN 073 sample. In HPTN 073, more than three-quarters of participants initiated PrEP in the first 6 months, with very few new initiators in the second interval. Nearly 44% of all participants (57.2% of those who initiated PrEP) were adherent to PrEP in the first six months; this decreased to 36.3% of all participants (46.1% of PrEP initiators) in the second six months. There was no missingness in the PrEP initiation variable

and very limited missingness in the PrEP adherence variable. In HPTN 061, 18 individuals tested positive for HIV by the six-month visit ($t = 1$); by the twelve-month visit ($t = 2$), a total of 28 individuals had tested positive (i.e., there were 10 new seroconversions between the six- and twelve-month visits). In HPTN 073, 5 individuals tested positive by the six-month visit ($t = 1$) and a total of 8 tested positive by the twelve-month visit ($t = 1$). The degree of missingness of HIV status was considerably higher in HPTN 061 compared to HPTN 073, particularly at the twelve-month time point. There were several important differences in the relevant covariates between HPTN 061 and HPTN 073 and marked differences in the extent of missingness. For both studies, the extent of missing covariate values was low at baseline, though this increased substantially at the 6-month time point ($t = 1$).

Table 1: Cohort summaries for the HPTN 061 sample and the HPTN 073 sample.

	HPTN 061 (n = 1134)			**HPTN 073 (n = 226)**		
	$t = 0$	$t = 1$	$t = 2$ †	$t = 0$	$t = 1$	$t = 2$ †
PrEP initiations, n (%)‡	--	0 (0%)	0 (0%)	--	173 (76.5%)	178 (78.8%)
Missing		0 (0%)	0 (0%)		0 (0%)	0 (0%)
PrEP adherence, n (%)	--	0 (0%)	0 (0%)	--	99 (43.8%)	82 (36.3%)
Missing		0 (0%)	0 (0%)		4 (1.8%)	4 (1.8%)
Covariates, n (%) or mean (SD)						
Age (years)	37.0 (12.2)	--	--	29.4 (9.9)	--	--
Missing	0 (0%)			0 (0%)		
Region						
Northeast	565 (49.8%)	--	--	75 (33.2%)	--	--
South	217 (19.1%)			75 (33.2%)		
West	352 (31.0%)			76 (33.6%)		
Missing	0 (0%)			0 (0%)		
Gay only	356 (31.4%)	--	--	158 (69.9%)	--	--
Missing	2 (0.2%)			0 (0%)		
Number of partners						
0 – 1	192 (16.9%)	353 (31.1%)	--	75 (33.2%)	88 (38.9%)	--
2 – 3	407 (35.9%)	297 (26.2%)		74 (32.7%)	51 (22.6%)	
≥ 4	532 (46.9%)	270 (23.8%)		73 (32.3%)	46 (20.4%)	
Missing	3 (0.3%)	214 (18.9%)		4 (1.8%)	41 (18.1%)	
URAI	459 (40.5%)	230 (20.3%)	--	95 (42.0%)	95 (42.0%)	--
Missing	35 (3.1%)	230 (20.3%)		5 (2.2%)	38 (16.8%)	
STI	126 (11.1%)	30 (2.6%)	--	32 (14.2%)	30 (13.3%)	--
Missing	80 (7.1%)	244 (21.5%)		8 (3.5%)	30 (13.3%)	
Outcome: HIV, n (%)	--	18 (1.6%)	28 (2.5%)	--	5 (2.2%)	8 (3.5%)
Missing		150 (13.2%)	263 (23.2%)		14 (6.2%)	21 (9.3%)

Note: URAI = unprotected (condomless) receptive anal intercourse; STI = sexually transmitted infection.
†People with $Y_1 = 1$ were not removed from the $t = 2$ column. ‡PrEP initiations: n = number of initiations prior to t.

Based on the weighted analysis, we estimated a 48% reduction in risk of HIV in 1 year due to having access to PrEP (0.520, 95% CI for the RR: 0.213, 1.217) (Table 2). For the analysis of PrEP initiation and adherence, PrEP initiation (without adherence) resulted in an estimated 52% reduction in risk of HIV in 6 months compared to individuals who did not initiate PrEP (95% CI for the RR: (0.122, 0.893)) and PrEP adherence resulted in an estimated 77% reduction in risk of HIV in 6 months compared to individuals who did not initiate PrEP (95% CI for the RR: 0.015, 0.798). Thus, we observed substantial reductions in estimated risk of HIV due to PrEP access,

initiation, and adherence, though only the latter two were significant at the 0.05 level. Importantly, all of these estimates pertain to the HPTN 073 population. Four bootstrap samples (0.4%) were removed because $D_1 = D_2 = 0$ for all HIV cases, which gave extreme and unstable effect estimates in these samples.

Table 2: Effect estimates for PrEP access, initiation, and adherence.

Comparison	Outcome	Estimated RR (95% CI)
PrEP access: $A = 1$ vs. $A = 0$	Risk of HIV in 1 year (Y_2)	0.520 (0.213, 1.217)
PrEP initiation without adherence: $D = 1$ vs. $D = 0$	Risk of HIV in 6 months (Y_t)	0.477 (0.122, 0.893)
PrEP adherence: $D = 2$ vs. $D = 0$	Risk of HIV in 6 months (Y_t)	0.227 (0.015, 0.798)

Note: RR = relative risk; CI = confidence interval.

Diagnostics & Sensitivity Analysis

Check of Balance

Recall that the utility of the weights w_i^A and w_{it}^D rests on the notion that they are functions of balancing scores and thus address confounding by the relevant (measured) covariates [38]; consequently, it was important to ensure that balance on the included covariates was achieved [77]. Balance is typically assessed by calculating standardized differences and ensuring these differences are small, typically less than 10% [35, 77]. For binary exposures, the standardized difference is obtained by dividing the difference in (weighted) means or proportions of each covariate for the two exposure levels is by the standard deviation [35]. For categorical exposures with more than two levels, standardized differences between each pair of exposure levels are calculated for each covariate [78] and typically summarized by taking the maximum over the exposure level pairs for each covariate [79]. Thus, we assessed whether balance between $A = 0$ and $A = 1$ on the covariates $\boldsymbol{X}_0$ was achieved by the weights w_i^A and whether balance between $D_t = 0$, 1, and 2 in HPTN 073 on the covariates $\boldsymbol{X}_{t-1}$ was achieved by the weights $w_{it}^{D,2}$, $t = 1, 2$. See the Supplement for further details. We used the cobalt package in R to compute balance measures [80].

The balance measures (Figure 2) correspond to the maximum standardized differences (i.e., the maximum imbalance) across all imputations. We considered the maximum across imputations as it is necessary to ensure balance in each imputed dataset [71]. The measures indicate the presence of covariate imbalance for all three exposures (A, D_1, and D_2) prior to weighting, and minimal covariate imbalance after weighting.

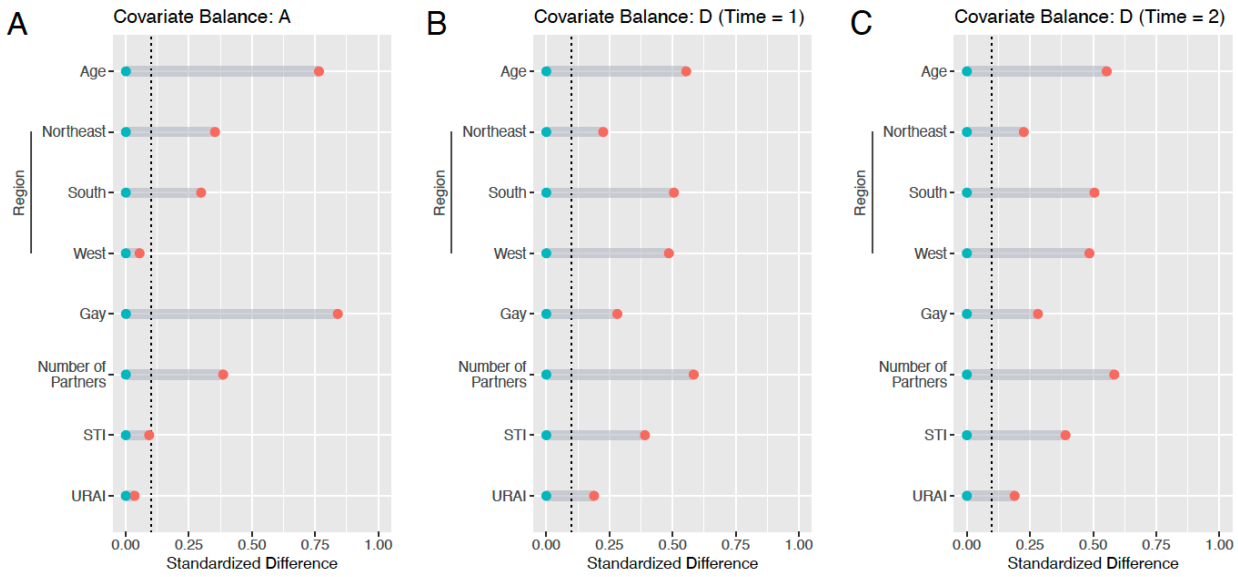


Figure 2: Measures of covariate balance before (orange) and after (blue) weighting. These plots summarize the standardized difference across exposure levels for the covariates included in the weight models for three exposures: A (panel A), D_1 (panel B), and D_2 (panel C). The vertical line at 0.1 highlights the level of imbalance that is generally considered acceptable (10%).
Note: STI = sexually transmitted infection; URAI = unprotected (condomless) receptive anal intercourse.

We also considered the maximum degree of covariate imbalance across imputations for each bootstrap sample, as balance is needed in each bootstrap sample in order to obtain valid effect estimates in each sample and thus valid confidence intervals. For the A analysis, covariate balance was achieved through weighting in all bootstrap samples. For D_1, in 18 bootstrap samples, covariate balance was not achieved after weighting in at least one imputation. For D_2, covariate imbalance was observed after weighting in 106 bootstrap samples. Consequently, effect estimates from these samples may be biased due to residual confounding by the imbalanced covariate(s), which may in turn affect the validity of the bootstrap confidence intervals. We re-estimated the confidence intervals after removing these 124 samples and observed little change (95% CI for $D = 1$ vs. $D = 0$: 0.130, 0.893; 95% CI for $D = 2$ vs. $D = 0$: 0.017, 0.798).

<u>Check of Positivity</u>

In our analysis, the maximum value of w_i^A across all bootstrap samples and all imputations was 8.05, while the maximum value of w_{it}^D was 159.75; however, consideration of only the weight distribution can mask positivity violations [75]. A better way to identify violations of positivity

assumptions is to consider the estimated exposure probabilities, $\widehat{\Pr}(A = 0|\boldsymbol{X}_0, Y_{t-1} = 0)$ and $\widehat{\Pr}(D_t = d_t|A = 1, \boldsymbol{X}_{t-1}, Y_{t-1} = 0)$, and the degree to which these are "extreme." We note that the estimates of $Pr(A = 0|\boldsymbol{X}_0, Y_1 = 0)$ were nearly identical to the estimates of $\widehat{Pr}(A = 0|\boldsymbol{X}_0)$, so only the latter are discussed here. We defined "extreme estimated probabilities" as $\widehat{Pr}(A = 0|\boldsymbol{X}_0) < 0.001$ and $\widehat{Pr}(D_t = d|A = 1, \boldsymbol{X}_{t-1}, Y_{t-1} = 0) < 0.001$ or $> 0.999, t = 1, 2, d = 0, 1, 2$. In the original sample, there were very few extreme estimates of exposure probabilities across the 10 imputations (Table 3), with the exception of $\widehat{Pr}(D_2 = 0|A = 1, \boldsymbol{X}_1, Y_1 = 0)$, for which we observed extreme values in up to 3.2% of individuals in each imputed dataset. These extreme values tended to be close to zero (Figure 2).

Considering the bootstrap samples, we observed that (near-)positivity violations were most common for $\widehat{Pr}(D_2 = 0|A = 1, \boldsymbol{X}_1, Y_1 = 0)$, with bootstrap samples having up to 49.8% of participants in each imputed dataset with extreme estimated probabilities (Table 3). There was a fairly high proportion of extreme values of $\widehat{Pr}(D_2 = 1|A = 1, \boldsymbol{X}_1, Y_1 = 0)$ and $\widehat{Pr}(D_2 = 2|A = 1, \boldsymbol{X}_1, Y_1 = 0)$, as well as $\widehat{Pr}(D_1 = 0|A = 1, \boldsymbol{X}_0, Y_1 = 0)$ and $\widehat{Pr}(D_1 = 1|A = 1, \boldsymbol{X}_0, Y_1 = 0)$. After excluding the most extreme 5% of bootstrap samples, the issues with extreme estimates generally only persisted for $\widehat{Pr}(D_2 = 0|A = 1, \boldsymbol{X}_1, Y_1 = 0)$ and $\widehat{Pr}(D_2 = 2|A = 1, \boldsymbol{X}_1, Y_1 = 0)$ (Table 3). In general, very small estimated probabilities were observed more often than very large estimated probabilities (Figure 3).

The extreme values observed for $\widehat{Pr}(D_2 = 0|A = 1, \boldsymbol{X}_1, Y_1 = 0)$ and $\widehat{Pr}(D_2 = 2|A = 1, \boldsymbol{X}_1, Y_1 = 0)$ were driven by men who were at quite high risk of HIV (and so were more likely to initiate/adhere to PrEP) and at very low risk of HIV (and thus perhaps less motivated to initiate/adhere to PrEP), respectively. In particular, the men in the former group tended to have 4 or more male partners, have an STI, and engage in URAI. The latter group, on the other hand, typically had fewer partners, did not engage in URAI, and did not have STIs. Taken together, these results suggest that for men in the HPTN 073 population with very high levels of risk, the counterfactual $Y_2(d_2 = 0)$ may not be realizable, and for men with very low levels of risk, the counterfactuals $Y_2(d_2 = 1)$ and $Y_2(d_2 = 2)$ might not be realizable. Consequently, our estimates of the effect of PrEP initiation and adherence in these very high and very low risk men (including the estimates from the bootstrap samples, which were used to estimate confidence intervals) may be based on extrapolation [81, 82].

Importantly, others have noted that near-violations of positivity generally must be extreme to have a practical impact on the causal effect estimate [83]. Furthermore, the possible impact of near-violations of positivity in our analysis mainly affects the bootstrap confidence intervals; given that extreme weights generally result in high variability, our confidence intervals are likely to be conservative.

Table 3: Proportion of extreme estimates for each exposure probability, across imputations in the original sample and across bootstrap samples and imputations.

Probability	Original sample Max.†	Bootstrap Max.‡	95th percentile‡
$\Pr(A=0\|\boldsymbol{X}_0)$	0%	0%	0%
$\Pr(D_1=0\|A=1,\boldsymbol{X}_0)$	0%	18.1%	1.3%
$\Pr(D_1=1\|A=1,\boldsymbol{X}_0)$	0%	9.3%	0%
$\Pr(D_1=2\|A=1,\boldsymbol{X}_0)$	0%	0.9%	0%
$\Pr(D_2=0\|A=1,\boldsymbol{X}_1,Y_1=0)$	3.2%	49.8%	37.7%
$\Pr(D_2=1\|A=1,\boldsymbol{X}_1,Y_1=0)$	0%	9.2%	0.5%
$\Pr(D_2=2\|A=1,\boldsymbol{X}_1,Y_1=0)$	0%	14.7%	6.4%

†Maximum proportion of extreme estimates across imputations. Extreme observations for $\Pr(A=0|\boldsymbol{X}_0)$ are those < 0.001; extreme observations for the remainder of the probabilities are those < 0.001 or > 0.999. The proportion of extreme observations is computed for each imputation, and the maximum across imputations is computed.
‡Maximum and 95th percentile proportion of extreme observations across bootstrap samples and imputations.

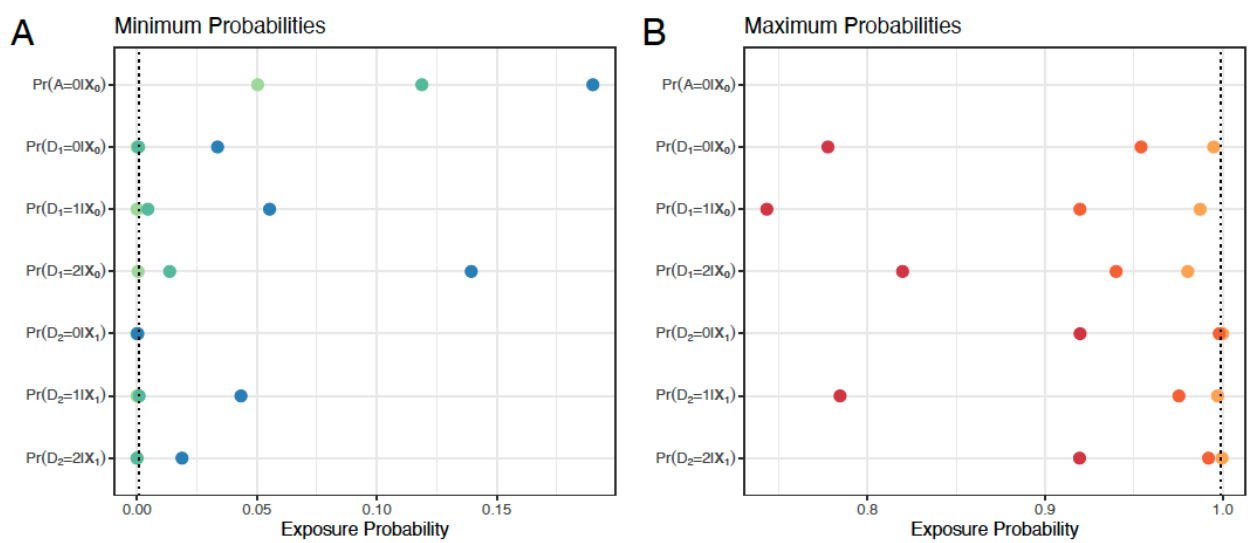


Figure 3: Summary of extreme values of estimated exposure probabilities across imputations and across bootstrap samples and imputations. Minimum values are show in panel A and maximum values are shown in panel B. In panel A, the minimum across 10 imputations in the original sample is shown in dark blue, the 5th percentile across bootstrap samples of the minimum probability among imputations in each bootstrap sample is shown in dark green, and the minimum across all bootstrap samples and all imputations is shown in light green. In panel B, the maximum across 10 imputations in the original sample is shown in dark red, the 95th percentile across bootstrap samples of the maximum probability among imputations in each bootstrap sample is shown in dark orange, and the maximum across all bootstrap samples and all imputations is shown in light orange. Note that $P(A=0|\boldsymbol{X}_0)$ are not shown in panel B

because no assumptions were made regarding the maximum value of $P(A = 0|\boldsymbol{X}_0)$. The vertical lines indicate extreme probabilities (defined as falling below 0.001 or above 0.999).

Unmeasured Confounding

Unmeasured or uncontrolled confounding (referred to as "unmeasured confounding" hereafter) arises from incomplete adjustment of one or more confounders, either because they are unmeasured, they are measured with error, or there is residual confounding due to improper adjustment. We evaluated the potential impact of unmeasured confounding on our casual effect estimates using methods similar to those of Robins [19], Brumback et al. [84], Chiba [57], and Klungsøyr et al. [85] (see details of our implementation in the Supplement), which can be used to estimate the true causal effect of an exposure given a certain degree of unmeasured confounding. The magnitude and direction of unmeasured confounding is quantified by the bias factor, here defined for A:

$$\alpha = \frac{\Pr(Y_2(a) = 1|A = 1, \boldsymbol{X}_0)}{\Pr(Y_2(a) = 1|A = 0, \boldsymbol{X}_0)},$$

which was assumed to be the same for $a \in (0, 1)$ and to not depend on $\boldsymbol{X}_0$ [57]. Chiba [57] proposed considering

$$\frac{1}{RR_c} \leq \alpha \leq RR_c,$$

where $RR_c > 1$ is the RR of Y_2 for a strong risk factor; we fixed $RR_c = 2$. We observed that for $\alpha = 0.71$, the estimated RR of HIV for $A = 1$ vs. $A = 0$ would be 0.77 (95% CI: 0.32, 1.78) (Figure 3). Conversely, for $\alpha = 1.4$, the estimated RR of HIV for PrEP access would be 0.35 (0.14, 0.82).

Likewise for the analysis of PrEP initiation/adherence (D), the bias factor is defined as

$$\alpha_t(d, d') = \frac{\Pr\left(Y_t(d) = 1|D_t = d', \boldsymbol{X}_{t-1}\right)}{\Pr\left(Y_t(d) = 1|D_t = d, \boldsymbol{X}_{t-1}\right)},$$

for $d, d' \in \{0, 1, 2\}$ and $t = 1, 2$. We assumed $\alpha_t(d, d')$ does not depend on $\boldsymbol{X}_{t-1}$ and $\alpha_1(d, d') = \alpha_2(d, d') \equiv \alpha(d, d')\ \forall\ d, d'$. In addition, we assumed $\alpha(0,1) = \alpha(1,2), \alpha(0,2) = \alpha(0,1)^2$, and $\alpha(d, d') = 1/\alpha(d', d)$. We considered

$$\frac{1}{RR_c} \leq \alpha(0,1) \leq RR_c,$$

and $RR_c = 2$; RR_c can be conceptualized as above. We found that for $\alpha(0,1) = 0.71$, the estimated RR of HIV for $D = 1$ vs. 0 was 0.66 (95% CI: 0.17, 1.15) and the estimated RR of HIV

for $D = 2$ vs. 0 was 0.43 (0.03, 1.33). For $\alpha(0,1) = 1.4$, these estimated RRs were 0.34 (0.09, 0.70) and 0.12 (0.01, 0.49), respectively.

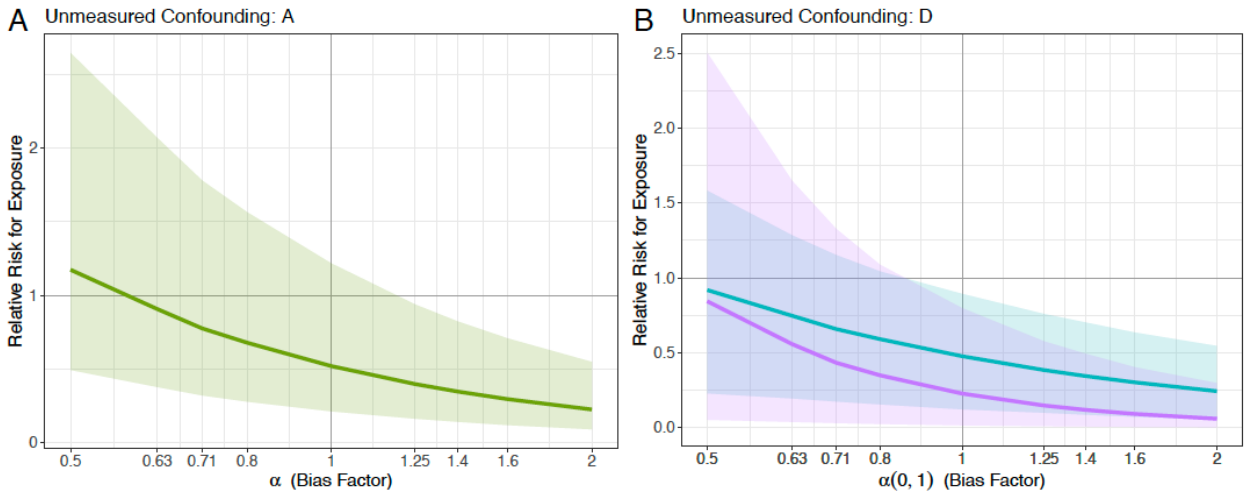


Figure 4: Results of the sensitivity analysis to assess the impact of unmeasured confounding. Estimates of the true causal effect of PrEP access (panel A) and PrEP initiation/adherence (panel B) and corresponding 95% confidence intervals were obtained using the bias factor (x-axis) to quantify the effect of unmeasured confounding. In panel B, blue indicates D = 1 vs. 0, while purple corresponds to D = 2 vs. 0.

DISCUSSION

We have provided estimates of the causal effects of PrEP access, initiation (without adherence), and adherence in Black MSM in the US based on a careful analysis combining data from two non-randomized studies, HPTN 061 and HPTN 073. Using a weighting approach and marginal structural models, we found that in the HPTN 073 population, access to PrEP reduced risk of HIV by an estimated 48% and, compared to Black MSM who did not engage with PrEP, PrEP initiation reduced HIV risk by an estimated 52% and PrEP adherence reduced HIV risk by an estimated 77%, though only the latter two estimates were statistically significant. These large estimated reductions in risk of HIV are not surprising, given the large protective effects of PrEP observed in other populations, particularly in settings associated with high adherence [11, 12, 14, 17, 86]. These results are of considerable public health importance as they provide estimates of the effect of PrEP specific to a population that bears a disproportionate burden of the HIV epidemic in the US. These estimates of the effect of PrEP among Black MSM may motivate some men to initiate and adhere to PrEP, substantially reducing their risk of HIV infection.

Importantly, beyond the specific goal of estimating the causal effect of PrEP in Black MSM in the US, our approach provides a framework for combining data from multiple non-randomized studies to estimate causal effects. This framework may be useful for estimating the effect of PrEP in other populations and, more broadly, for estimating causal effects of other exposures using existing datasets. Such strategies are of particular value in the context of large networks, like HPTN, which offer the opportunity to utilize data from multiple studies to answer causal questions without a randomized trial. Furthermore, when datasets like these exist, there is an ethical obligation to utilize the data as fully as possible, using robust and rigorous methods. Such stewardship is particularly necessary when the data come from minoritized communities, as these communities are frequently called upon by researchers to participate in such studies and may be overburdened or wary of participation altogether due to past encounters with the medical community.

Alternative Approaches

Our approach, based on marginal structural models with probability weighting, is quite common, though alternatives exist. Perhaps the most obvious is a standard regression model, i.e., a generalized linear model adjusting for the suspected confounding factors. This would have provided an estimate of the conditional effects of PrEP access, initiation, and adherence [60], which may differ from the marginal effect in the presence of effect modification [87, 88]. Furthermore, such an analysis would be computationally difficult given our very rare outcome and the need to adjust for a non-trivial number of presumed confounders. Conversely, estimating the weights in our analysis was straightforward as the exposures were quite common. Additionally, while our analysis relies on strong assumptions, similar assumptions (including no measurement error, no unmeasured confounding, no model misspecification, and no interference) are needed for the estimates from a standard regression model to have a causal interpretation [36, 41, 62]. Our analytic approach also allowed us to be very deliberate about our target of inference. A regression model, on the other hand, would provide estimates for a population that is an ambiguous weighted average of the HPTN 061 and HPTN 073 populations [89].

We have alluded to the fact that a randomized, placebo-controlled trial may be useful in estimating the causal effect of PrEP in Black MSM, but such a trial would be unethical in our

setting. Furthermore, it is not feasible to randomize individuals to levels of PrEP adherence, or even to PrEP adherence or non-adherence: at most, participants can be provided with pills and instructions and their adherence monitored, but they cannot be forced to take medication (though there are biobehavioral interventions that can improve PrEP adherence, e.g., [90]). Marginal structural models, which can be thought of as seeking to approximate an unblinded randomized trial [73, 91], are thus particularly useful here. Importantly, however, as discussed in the Methods, our approach can provide balance between exposure levels only on measured variables [34]. This is in contrast with randomized trials, which provide balance on all variables (in expectation). On the other hand, while randomization yields balance in expectation on measured and unmeasured variables, allowing estimation of the causal effect, the presence of non-representative sampling [27, 92, 93], non-compliance [48, 91], chance imbalances [27, 93], and/or loss to follow-up [41] in randomized trials can threaten both internal and external validity. In addition, given the challenges related to recruiting stigmatized populations, like Black MSM, for research, it may be difficult to conduct a randomized trial in this population [28] and the careful use of existing data and multiple data sources (e.g., external controls) can yield valuable insights.

It is common to assume a parsimonious dose-response relationship [20], for a non-binary exposure, as we have done in the analysis of D. While it would have been preferable to fit a more flexible model, doing so was not possible as there were no cases of HIV among individuals with $D = 2$ in the original sample (before imputation). Since D was modeled as an ordinal variable, the estimate of the causal effects of $D = 2$ vs. $D = 0$ and $D = 1$ vs. $D = 0$ are constrained, given the presumed log-linear relationship. Given that previous research has found very large reductions in risk of HIV among individuals who were adherent to PrEP [13, 86, 94], there is evidence that our analysis, in which the relationship between D and HIV risk is constrained, may have overestimated the effect of PrEP initiation without adherence, and underestimated the effect of PrEP adherence.

Another possibility would have been to include datasets from other HPTN studies involving Black MSM in the US, e.g., HPTN 083, which was a randomized study comparing long-acting injectable cabotegravir to daily oral TDF-FTC [95]. The TDF-FTC arm, which was largely comprised of MSM and included 434 Black participants in the US, could have contributed to our analysis. However, we chose not to include data from this study for two reasons. First, HPTN 083 was a randomized study and is thus susceptible to the challenges described above related

to external validity. Second, HPTN 083 was conducted between December 2016 and March 2020, making it fairly far removed temporally from HPTN 061.

Strengths

Our analysis involved data from two carefully conducted observational studies managed by the same trial network, including one fairly large study. Since both studies were under the umbrella of HPTN and were conducted reasonably close together in time, data collection (variables measured, definitions used, and assessment mechanisms), participant recruitment and retention strategies, and outcome definitions were rigorous and similar between the two studies, making their combination more compelling and a clear strength of our analysis. While some challenges related to combining the data from HPTN 061 and HPTN 073 persist (as discussed below), this analysis would not have been possible if we had been limited to only one of these datasets: relying only on HPTN 073 to evaluate the effect of PrEP initiation and adherence would be infeasible given its relatively small size and very low incidence of HIV [28]. Importantly, individuals who enroll in observational studies, like HPTN 061 and HPTN 073, may be more representative of the general population than those who would enroll in a randomized trial, enhancing the generalizability of results.

Another clear strength of our analysis is the rigorous and robust approach taken to effect estimation, quantification of uncertainty, and evaluation of the plausibility of key assumptions. In so doing, we have provided a framework that can be used more generally to estimate causal effects when data from multiple non-randomized studies are available. Even when a randomized controlled trial is ethical and feasible, if data from non-randomized studies exists, it may be advisable to employ a framework like ours to either provide motivation for a randomized trial or, if the data are from well-conducted studies, to obviate the need for a trial altogether, preserving financial resources, reducing burden on would-be participants, and allowing the expedient estimation of causal effects.

Limitations

Several limitations of our analysis exist. The first set of limitations relates to the data sources used in our analysis. First, the HPTN 073 sample may not be representative of the HPTN 073 population (the population from which the HPTN 073 study participants were sampled) [27] due to selection bias [44]. In addition, the HPTN 073 population may differ from our target

population, Black MSM in the US, due to enrollment processes and eligibility criteria [27, 96]. Generalizing our results to the population of Black MSM in the US would require a careful consideration of selection bias [97] and, relatedly, the sampling schemes used. HPTN 073 attempted to recruit an equal number of participants over and under age 25 and aimed to enroll an equal number of individuals at each site and relied on informal recruitment methods, like peer referral, the ramifications of which are difficult to quantify as the sampling probabilities are unknown. Thus, some degree of generalizability may be lost, though this will only be a concern insofar as there is effect modification and/or unmeasured confounding [98].

Despite efforts to exclude individuals in HPTN 061 not eligible for HPTN 073, discrepancies between the two samples may persist because variables related to PrEP eligibility in HPTN 073 and specific clinical exclusions used in HPTN 073 were not measured in HPTN 061. However, the impact on our causal estimates is expected to be minimal as the discrepant exclusion criteria are not expected to lead to differences in HIV risk. Additionally, while we have achieved balance on measured confounders between the HPTN 061 and HPTN 073 samples, differences in the distribution of effect modifiers not included in our weight models between the HPTN 061 and HPTN 073 populations may lead to residual bias [27]. Finally, we assumed that individuals in the HPTN 061 population were exchangeable with those in the HPTN 073 population who did not initiate PrEP (conditional on covariates). However, all individuals enrolled in HPTN 073 had access to the client-centered care coordination (C4) program, which aimed to support participants' PrEP understanding, initiation, and adherence through risk counseling and case management activities [22]. It is reasonable to assume that neither C4, a light-touch program, nor PrEP access has a direct effect on HIV risk.

The two studies also differed in the frequency of data collection, which was every three months in HPTN 073 and every six months in HPTN 061. Given that we only utilized the data collected every six months, there is a possibility of some measurement error in the data from HPTN 073. We took steps to mitigate the impact of this source of measurement error (see the Supplement for details); consequently, it is likely to have little influence on the main results. Relatedly, the covariates measured in both studies were based on self-report, which could lead to measurement error due to psychological bias [21, 99]. Another important source of potential measurement error relates to PrEP initiation and adherence. Measures of PrEP adherence based on blood samples reflect PrEP-taking behavior over the previous several weeks, but our analysis applied these adherence measures to the entire six-month interval. Likewise, for PrEP

initiation, there may be some difference between people who initiate early versus late in a given time interval. Future analyses based on data with shorter time intervals would likely be informative.

The assumption of SUTVA may be suspect with small sexual networks, as PrEP offers benefits not just for individual men, but also for men in their sexual networks (as is generally the case for interventions designed to prevent transmission of infectious diseases). Importantly, though, networks for Black MSM are likely to be large enough to avoid issues of interference, particularly given that HPTN 061 and HPTN 073 were studies conducted in several different cities in the US. We also assumed functionally no access to PrEP in HPTN 061, but there was one individual who reported using PrEP at baseline. Given that this is likely to have almost no impact on our results and may in fact not be accurate since PrEP was not approved during HPTN 061 we did not remove this individual from the analysis. The positivity assumption appears to have been violated for a subset of participants in both the original sample and across many bootstrap samples, particularly for men at very high or very low risk of HIV. Consequently, the results of the analysis of PrEP initiation and adherence may rely to some degree on extrapolation in certain subsets of the covariate space. Future work should seek to estimate the effect of PrEP and provide inference in these groups, as the effects there may differ from those among men not at the extremes of the HIV risk distribution.

Conclusions

We have provided estimates of the causal effects of PrEP access, initiation, and adherence in a key population in the US HIV epidemic, Black MSM, finding large reductions in risk, particularly for initiation and adherence. More broadly, we have created a rigorous analytic framework for combining data from multiple studies to estimate causal effects. This framework may be generally useful when seeking to estimate causal effects using several data sources, particularly when a randomized control trial is unethical or infeasible.

DATA AVAILABILITY

Data and code to run this analysis are available at: https://github.com/meisnera/BMSM_PrEP.

ACKNOWLEDGEMENTS

We acknowledge funding from NIH U01 HL146242 and NIH P30 MH123248. Overall support for the HIV Prevention Trials Network (HPTN) is provided by the NIH under Award Numbers UM1AI068619 (HPTN Leadership and Operations Center), UM1AI068617 (HPTN Statistical and Data Management Center), and UM1AI068613 (HPTN Laboratory Center). Additional support for this study was provided by UM1AI069412. We extend our thanks the HPTN 061 and HPTN 073 study teams and study participants.

TABLE AND FIGURE CAPTIONS

Table 1: Cohort summaries for the HPTN 061 sample and the HPTN 073 sample.

Table 2: Effect estimates for PrEP access, initiation, and adherence.

Table 3: Proportion of extreme estimates for each exposure probability, across imputations in the original sample and across bootstrap samples and imputations.

Figure 1: DAG representing the relationships among the exposures (A, D_1, D_2), the outcomes (Y_1, Y_2), and the covariates ($\boldsymbol{X}_0, \boldsymbol{X}_1$). Dashed lines indicate deterministic relationships.

Figure 2: Measures of covariate balance before (orange) and after (blue) weighting. These plots summarize the standardized difference across exposure levels for the covariates included in the weight models for three exposures: A (panel A), D_1 (panel B), and D_2 (panel C). The vertical line at 0.1 highlights the level of imbalance that is generally considered acceptable (10%).

Figure 3: Summary of extreme values of estimated exposure probabilities across imputations and across bootstrap samples and imputations. Minimum values are show in panel A and maximum values are shown in panel B. In panel A, the minimum across 10 imputations in the original sample is shown in dark blue, the 5th percentile across bootstrap samples of the minimum probability among imputations in each bootstrap sample is shown in dark green, and the minimum across all bootstrap samples and all imputations is shown in light green. In panel B, the maximum across 10 imputations in the original sample is shown in dark red, the 95th percentile across bootstrap samples of the maximum probability among imputations in each bootstrap sample is shown in dark orange, and the maximum across all bootstrap samples and all imputations is shown in light orange. Note that $P(A = 0|\boldsymbol{X}_0)$ are not shown in panel B because no assumptions were made regarding the maximum value of $P(A = 0|\boldsymbol{X}_0)$. The vertical lines indicate extreme probabilities (defined as falling below 0.001 or above 0.999).

Figure 4: Results of the sensitivity analysis to assess the impact of unmeasured confounding. Estimates of the true causal effect of PrEP access (panel A) and PrEP initiation/adherence (panel B) and corresponding 95% confidence intervals were obtained using the bias factor (x-axis) to quantify the effect of unmeasured confounding. In panel B, blue indicates D = 1 vs. 0, while purple corresponds to D = 2 vs. 0.

REFERENCES

1. HIV.gov. *Data & Trends: U.S. Statistics*. 2022 [cited 2023 August 15]; Available from: https://www.hiv.gov/hiv-basics/overview/data-and-trends/statistics/.
2. *US Census Quick Facts*. 2020 [cited 2023 September 14]; Available from: https://www.census.gov/quickfacts/fact/table/US/IPE120221.
3. Brewer, R.A., et al., *Exploring the relationship between incarceration and HIV among black men who have sex with men in the United States.* JAIDS: Journal of acquired immune deficiency syndromes 2014. **65**(2): p. 218.
4. Fields, E.L., et al., *"I always felt I had to prove my manhood": Homosexuality, masculinity, gender role strain, and HIV risk among young Black men who have sex with men.* American Journal of Public Health, 2015. **105**(1): p. 122-131.
5. Levy, M.E., et al., *Understanding structural barriers to accessing HIV testing and prevention services among black men who have sex with men (BMSM) in the United States.* AIDS and Behavior, 2014. **18**: p. 972-996.
6. Malebranche, D.J., et al., *Race and sexual identity: perceptions about medical culture and healthcare among Black men who have sex with men.* Journal of the National Medical Association, 2004. **96**(1): p. 97.
7. Mayer, K.H., et al., *Concomitant socioeconomic, behavioral, and biological factors associated with the disproportionate HIV infection burden among Black men who have sex with men in 6 US cities.* PLOS One, 2014. **9**(1): p. e87298.
8. Mena, L., R.A. Crosby, and A. Geter, *A novel measure of poverty and its association with elevated sexual risk behavior among young Black MSM.* International Journal of STD & AIDS, 2017. **28**(6): p. 602-607.
9. Millett, G.A., et al., *Comparisons of disparities and risks of HIV infection in black and other men who have sex with men in Canada, UK, and USA: a meta-analysis.* The Lancet, 2012. **380**(9839): p. 341-348.
10. Nelson, L.E., et al., *Economic, legal, and social hardships associated with HIV risk among black men who have sex with men in six US cities.* Journal of Urban Health, 2016. **93**: p. 170-188.
11. Baeten, J.M., et al., *Antiretroviral prophylaxis for HIV prevention in heterosexual men and women.* New England Journal of Medicine, 2012. **367**(5): p. 399-410.
12. Choopanya, K., 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): p. 2083-2090.
13. Donnell, D., et al., *HIV protective efficacy and correlates of tenofovir blood concentrations in a clinical trial of PrEP for HIV prevention.* JAIDS: Journal of Acquired Immune Deficiency Syndromes, 2014. **66**(3): p. 340.
14. Grant, R.M., et al., *Preexposure chemoprophylaxis for HIV prevention in men who have sex with men.* New England Journal of Medicine, 2010. **363**(27): p. 2587-2599.
15. Marrazzo, J.M., et al., *Tenofovir-based preexposure prophylaxis for HIV infection among African women.* New England Journal of Medicine, 2015. **372**(6): p. 509-518.
16. Murnane, P.M., et al., *Efficacy of pre-exposure prophylaxis for HIV-1 prevention among high risk heterosexuals: subgroup analyses from the Partners PrEP Study.* AIDS, 2013. **27**(13).
17. Thigpen, M.C., et al., *Antiretroviral preexposure prophylaxis for heterosexual HIV transmission in Botswana.* New England Journal of Medicine, 2012. **367**(5): p. 423-434.
18. Van Damme, L., et al., *Preexposure prophylaxis for HIV infection among African women.* New England Journal of Medicine, 2012. **367**(5): p. 411-422.

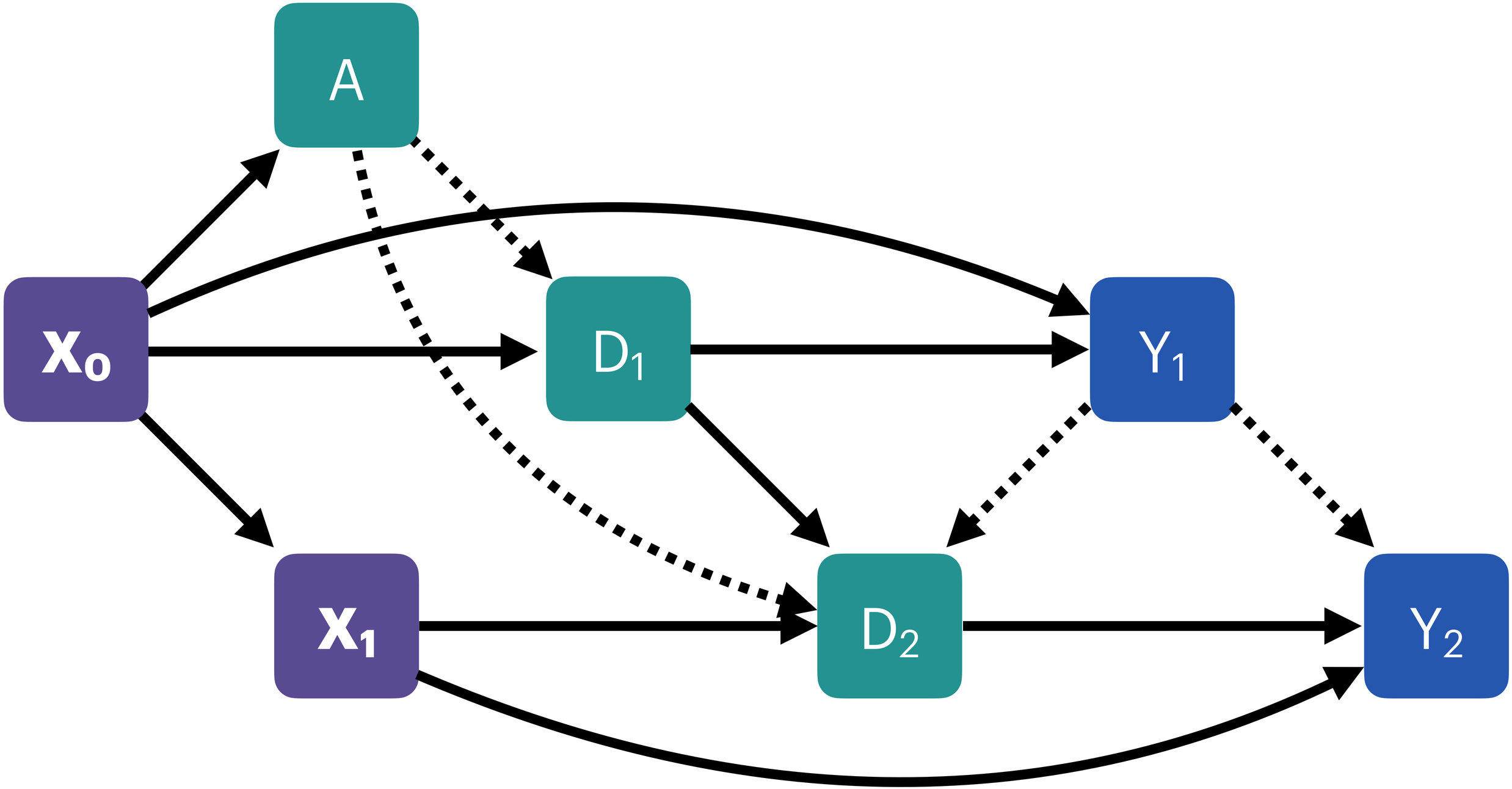
19. Robins, J.M., *Association, causation, and marginal structural models.* Synthese, 1999. **121**(1/2): p. 151-179.
20. Robins, J.M., M.A. Hernan, and B. Brumback, *Marginal structural models and causal inference in epidemiology.* Epidemiology, 2000: p. 550-560.
21. Koblin, B.A., et al., *Correlates of HIV acquisition in a cohort of Black men who have sex with men in the United States: HIV prevention trials network (HPTN) 061.* PLOS One 2013. **8**(7): p. e70413.
22. Wheeler, D.P., et al., *Pre – exposure prophylaxis initiation and adherence among Black men who have sex with men (MSM) in three US cities: results from the HPTN 073 study.* Journal of the International AIDS Society, 2019. **22**(2): p. e25223.
23. Hendrix, C.W., et al., *Dose frequency ranging pharmacokinetic study of tenofovir-emtricitabine after directly observed dosing in healthy volunteers to establish adherence benchmarks (HPTN 066).* AIDS Research and Human Retroviruses, 2016. **32**(1): p. 32-43.
24. Wang, C., et al., *Propensity score-integrated composite likelihood approach for incorporating real-world evidence in single-arm clinical studies.* Journal of Biopharmaceutical Statistics, 2020. **30**(3): p. 495-507.
25. Fang, Y., et al., *Key considerations in the design of real-world studies.* Contemporary Clinical Trials, 2020. **96**: p. 106091.
26. McCulloch, C.E., *observational studies, time-dependent confounding, and marginal structural models.* Arthritis & Rheumatology, 2015. **67**(3): p. 609-611.
27. Degtiar, I. and S. Rose, *A review of generalizability and transportability.* Annual Review of Statistics and Its Application, 2023. **10**: p. 501-524.
28. Thorlund, K., et al., *Synthetic and external controls in clinical trials–a primer for researchers.* Clinical Epidemiology, 2020: p. 457-467.
29. Levy, M.E., et al., *A longitudinal analysis of treatment optimism and HIV acquisition and transmission risk behaviors among black men who have sex with men in HPTN 061.* AIDS and Behavior, 2017. **21**: p. 2958-2972.
30. Whitfield, D.L., et al., *Risk compensation in HIV PrEP adherence among Black men who have sex with men in HPTN 073 study.* AIDS Care, 2021. **33**(5): p. 633-638.
31. Hernán, M. and J. Robins, *Causal Inference: What If?* 2020, Boca Raton: Chapman & Hall/CRC.
32. Sato, T. and Y. Matsuyama, *Marginal structural models as a tool for standardization.* Epidemiology, 2003: p. 680-686.
33. Hernán, M.A. and J.M. Robins, *Using big data to emulate a target trial when a randomized trial is not available.* American Journal of Epidemiology, 2016. **183**(8): p. 758-764.
34. Platt, R.W., J.A.C. Delaney, and S. Suissa, *The positivity assumption and marginal structural models: the example of warfarin use and risk of bleeding.* European Journal of Epidemiology, 2012. **27**: p. 77-83.
35. Austin, P.C., *An introduction to propensity score methods for reducing the effects of confounding in observational studies.* Multivariate Behavioral Research, 2011. **46**(3): p. 399-424.
36. Hernán, M.Á., B. Brumback, and J.M. Robins, *Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men.* Epidemiology, 2000: p. 561-570.
37. Li, F., K.L. Morgan, and A.M. Zaslavsky, *Balancing covariates via propensity score weighting.* Journal of the American Statistical Association, 2018. **113**(521): p. 390-400.

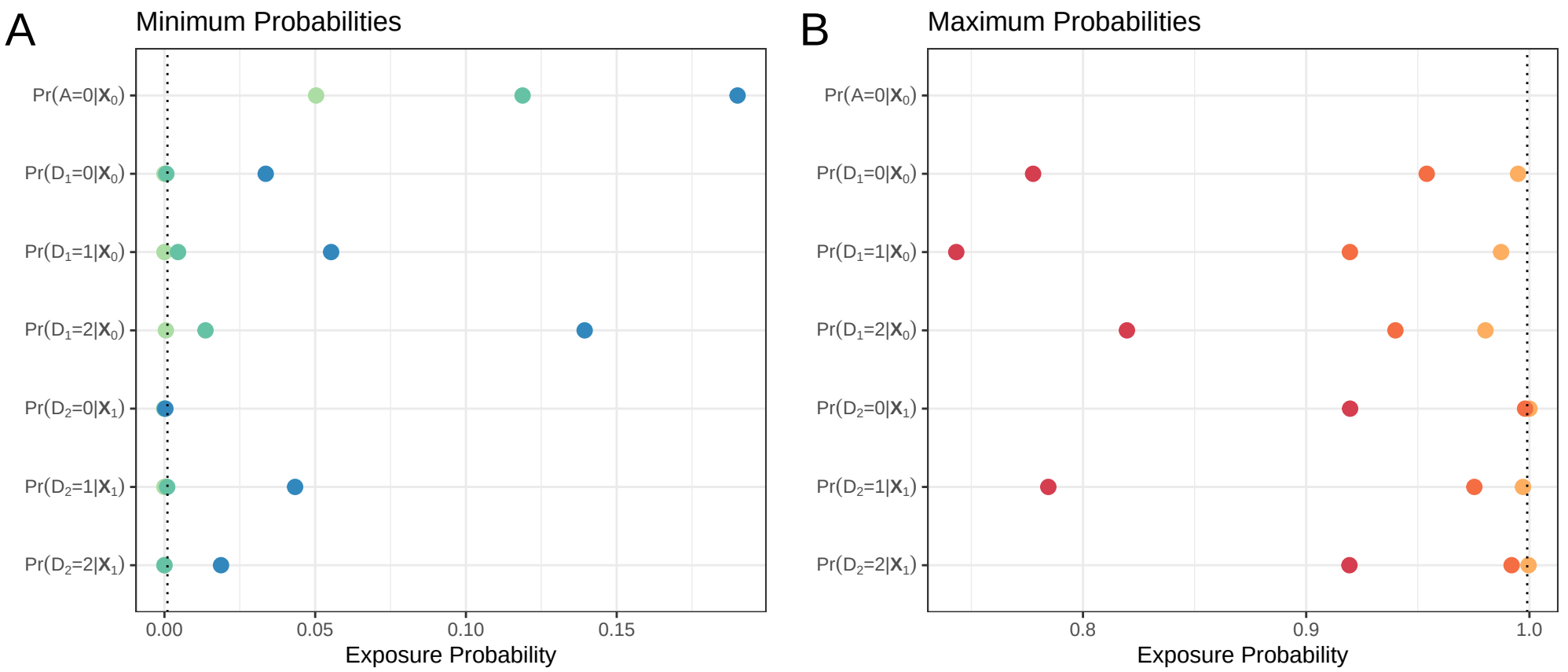
38. Imbens, G.W., *The role of the propensity score in estimating dose-response functions.* Biometrika, 2000. **87**(3): p. 706-710.
39. Westreich, D., et al., *Imputation approaches for potential outcomes in causal inference.* International Journal of Epidemiology, 2015. **44**(5): p. 1731-1737.
40. Daniel, R.M., et al., *Methods for dealing with time – dependent confounding.* Statistics in Medicine, 2013. **32**(9): p. 1584-1618.
41. Hernán, M.A. and J.M. Robins, *Estimating causal effects from epidemiological data.* Journal of Epidemiology & Community Health, 2006. **60**(7): p. 578-586.
42. Greifer, N. and E.A. Stuart, *Choosing the estimand when matching or weighting in observational studies.* 2021: arXiv preprint arXiv:2106.10577.
43. Hernán, M.A., B.A. Brumback, and J.M. Robins, *Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures.* Statistics in Medicine, 2002. **21**(12): p. 1689-1709.
44. Naimi, A.I., S.R. Cole, and E.H. Kennedy, *An introduction to g methods.* International Journal of Epidemiology, 2017. **46**(2): p. 756-762.
45. Xiao, Y., M. Abrahamowicz, and E.E. Moodie, *Accuracy of conventional and marginal structural Cox model estimators: a simulation study.* The International Journal of Biostatistics, 2010. **6**(2).
46. Yang, S., et al., *Application of marginal structural models in pharmacoepidemiologic studies: a systematic review.* Pharmacoepidemiology and Drug Safety, 2014. **23**(6): p. 560-571.
47. Hernán, M.A., *Beyond exchangeability: the other conditions for causal inference in medical research.* Statistical Methods in Medical Research 2012. **21**(1): p. 3-5.
48. Bacak, V. and E.H. Kennedy, *Marginal structural models: An application to incarceration and marriage during young adulthood.* Journal of Marriage and Family, 2015. **77**(1): p. 112-125.
49. Imai, K. and M. Ratkovic, *Robust estimation of inverse probability weights for marginal structural models.* Journal of the American Statistical Association, 2015. **110**(511): p. 1013-1023.
50. Imai, K. and M. Ratkovic, *Covariate balancing propensity score.* Journal of the Royal Statistical Society Series B: Statistical Methodology, 2014. **76**(1): p. 243-263.
51. Fong C, R.M., Imai K *CBPS: Covariate Balancing Propensity Score*. 2022. p. R package.
52. Team, R.C., *R: A language and environment for statistical computing* 2023, R Foundation for Statistical Computing: Vienna, Austria.
53. Ko, H., J.W. Hogan, and K.H. Mayer, *Estimating causal treatment effects from longitudinal HIV natural history studies using marginal structural models.* Biometrics, 2003. **59**(1): p. 152-162.
54. Richardson, T.S. and J.M. Robins, *Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality.* Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper, 2013. **128**(30): p. 2013.
55. Ertefaie, A. and D.A. Stephens, *Comparing approaches to causal inference for longitudinal data: inverse probability weighting versus propensity scores.* The International Journal of Biostatistics, 2010. **6**(2).
56. Li, F. and F. Li, *Propensity score weighting for causal inference with multiple treatments.* The Annals of Applied Statistics, 2019. **13**(4): p. 2389-2415.
57. Chiba, Y., *Sensitivity analysis of unmeasured confounding for the causal risk ratio by applying marginal structural models.* Communications in Statistics - Theory and Methods, 2009. **39**(1): p. 65-76.

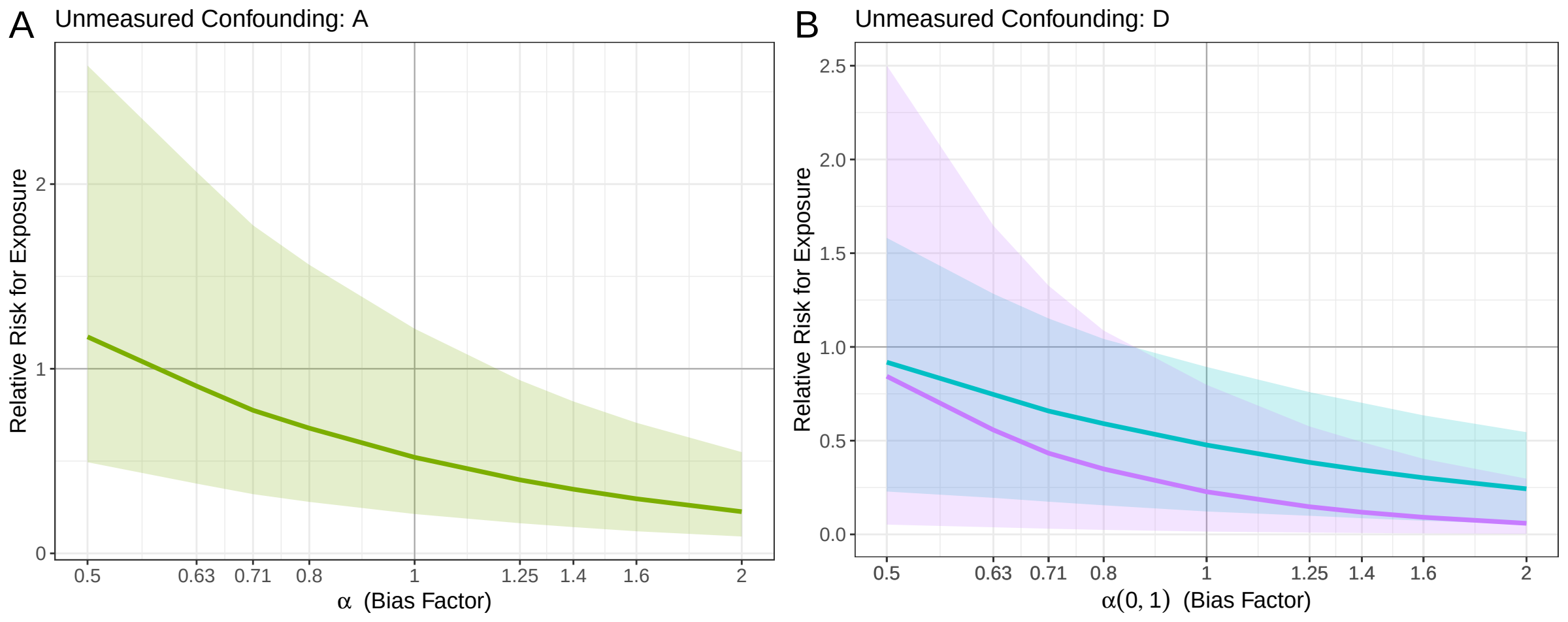
58. Ngwa, J.S., et al., *A comparison of time dependent Cox regression, pooled logistic regression and cross sectional pooling with simulations and an application to the Framingham Heart Study.* BMC Medical Research Methodology, 2016. **16**: p. 1-12.
59. Hernán, M.A., *A definition of causal effect for epidemiological research.* Journal of Epidemiology & Community Health, 2004. **58**(4): p. 265-271.
60. Cerdá, M., et al., *The relationship between neighborhood poverty and alcohol use: estimation by marginal structural models.* Epidemiology, 2010. **21**(4): p. 482.
61. Shinozaki, T. and E. Suzuki, *Understanding marginal structural models for time-varying exposures: pitfalls and tips.* Journal of Epidemiology, 2020. **30**(9): p. 377-389.
62. Howe, C.J., et al., *Estimating the effects of multiple time-varying exposures using joint marginal structural models: alcohol consumption, injection drug use, and HIV acquisition.* Epidemiology, 2012. **23**(4): p. 574.
63. Fewell, Z., et al., *Controlling for time-dependent confounding using marginal structural models.* The Stata Journal, 2004. **4**(4): p. 402-420.
64. Austin, P.C., P. Grootendorst, and G.M. Anderson, *A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study.* Statistics in Medicine, 2007. **26**(4): p. 734-753.
65. Brookhart, M.A., et al., *Variable selection for propensity score models.* American Journal of Epidemiology, 2006. **163**(12): p. 1149-1156.
66. Shiba, K. and T. Kawahara, *Using propensity scores for causal inference: pitfalls and tips.* Journal of Epidemiology, 2021. **31**(8): p. 457-463.
67. Crowson, C.S., et al., *The Basics of Propensity Scoring and Marginal Structural Models* 2013: Mayo Clinic.
68. Garofalo, R., et al., *Incidence of HIV infection and sexually transmitted infections and related risk factors among very young men who have sex with men.* JAIDS Journal of Acquired Immune Deficiency Syndromes, 2016. **72**(1): p. 79.
69. Stekhoven, D.J. and P. Bühlmann, *MissForest—non-parametric missing value imputation for mixed-type data.* Bioinformatics, 2012. **28**(1): p. 112-118.
70. Mayer, M., *missRanger: Fast Imputation of Missing Values*. 2023. p. R package.
71. Leyrat, C., et al., *Propensity score analysis with partially observed covariates: how should multiple imputation be used?* Statistical Methods in Medical Research, 2019. **28**(1): p. 3-19.
72. Mattei, A., *Estimating and using propensity score in presence of missing background data: an application to assess the impact of childbearing on wellbeing.* Statistical Methods and Applications, 2009. **18**: p. 257-273.
73. Cole, S.R., et al., *Effect of highly active antiretroviral therapy on time to acquired immunodeficiency syndrome or death using marginal structural models.* American Journal of Epidemiology, 2003. **158**(7): p. 687-694.
74. Reifeis, S.A. and M.G. Hudgens, *On variance of the treatment effect in the treated when estimated by inverse probability weighting.* American Journal of Epidemiology, 2022. **191**(6): p. 1092-1097.
75. Xiao, Y., E.E. Moodie, and M. Abrahamowicz, *Comparison of approaches to weight truncation for marginal structural Cox models.* Epidemiologic Methods, 2013. **2**(1): p. 1-20.
76. Schomaker, M. and C. Heumann, *Bootstrap inference when using multiple imputation.* Statistics in Medicine, 2018. **37**(14): p. 2252-2266.
77. Austin, P.C. and E.A. Stuart, *Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate*

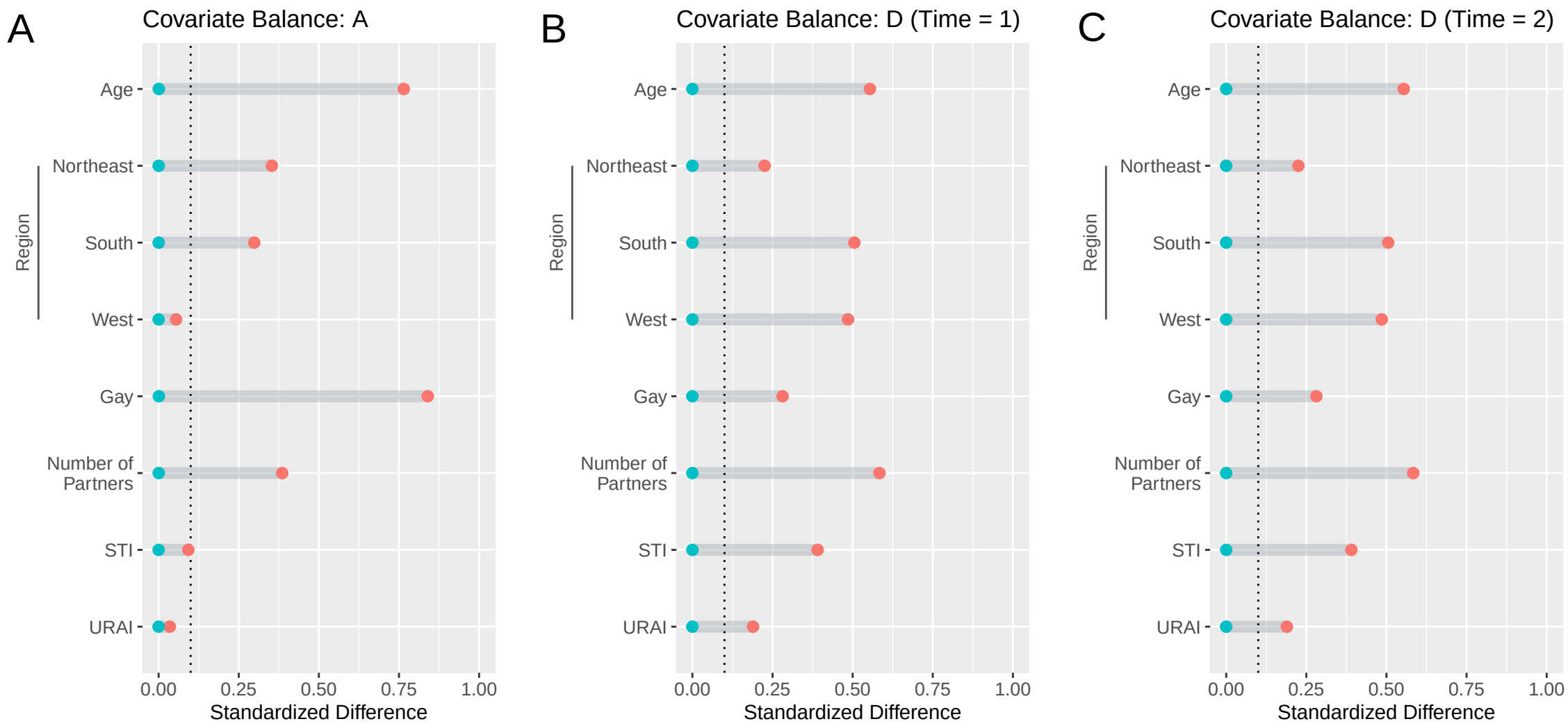
causal treatment effects in observational studies. Statistics in Medicine, 2015. **34**(28): p. 3661-3679.
78. Fong, C. and K. Imai, *Covariate balancing propensity score for general treatment regimes*, in *Princeton Manuscript*. 2014.
79. McCaffrey, D.F., et al., *A tutorial on propensity score estimation for multiple treatments using generalized boosted models.* Statistics in Medicine, 2013. **32**(19): p. 3388-3414.
80. Greifer, N., *cobalt: Covariate Balance Tables and Plots*. 2023. p. R package version.
81. Naimi, A.I., et al., *A comparison of methods to estimate the hazard ratio under conditions of time-varying confounding and nonpositivity.* Epidemiology (2011. **22**(5): p. 718.
82. Westreich, D. and S.R. Cole, *Invited commentary: positivity in practice.* American Journal of Epidemiology, 2010. **171**(6): p. 674-677.
83. Barber, J.S., S.A. Murphy, and N. Verbitsky, *Adjusting for time-varying confounding in survival analysis.* Sociological Methodology, 2004. **34**(1): p. 163-192.
84. Brumback, B.A., et al., *Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures.* Statistics in Medicine, 2004. **23**(5): p. 749-767.
85. Klungsøyr, O., et al., *Sensitivity analysis for unmeasured confounding in a marginal structural Cox proportional hazards model.* Lifetime Data Analysis, 2009. **15**: p. 278-294.
86. Molina, J.-M., et al., *On-demand preexposure prophylaxis in men at high risk for HIV-1 infection.* New England Journal of Medicine, 2015. **373**(23): p. 2237-2246.
87. Hernán, M.A. and T.J. VanderWeele, *Compound treatments and transportability of causal inference.* Epidemiology, 2011. **22**(3): p. 368.
88. VanderWeele, T.J. and M.J. Knol, *A tutorial on interaction.* Epidemiologic Methods, 2014. **3**(1): p. 33-72.
89. Słoczyński, T., *Interpreting OLS estimands when treatment effects are heterogeneous: Smaller groups get larger weights.* Review of Economics and Statistics, 2022. **104**(3): p. 501-509.
90. Goedel, W.C., et al., *A pilot study of a patient navigation intervention to improve HIV pre-exposure prophylaxis persistence among black/African American men who have sex with men.* JAIDS Journal of Acquired Immune Deficiency Syndromes, 2022: p. 10.1097.
91. Hernán, M.A., B. Brumback, and J.M. Robins, *Marginal structural models to estimate the joint causal effect of nonrandomized treatments.* Journal of the American Statistical Association, 2001. **96**(454): p. 440-448.
92. Burcu, M., et al., *Real‐world evidence to support regulatory decision‐making for medicines: considerations for external control arms.* Pharmacoepidemiology and Drug Safety, 2020. **29**(10): p. 1228-1235.
93. Westreich, D., et al., *Transportability of trial results using inverse odds of sampling weights.* American Journal of Epidemiology, 2017. **186**(8): p. 1010-1014.
94. Dimitrov, D.T., B.R. Mâsse, and D. Donnell, *PrEP adherence patterns strongly impact individual HIV risk and observed efficacy in randomized clinical trials.* JAIDS Journal of Acquired Immune Deficiency Syndromes, 2016. **72**(4): p. 444.
95. Landovitz, R.J., et al., *Cabotegravir for HIV prevention in cisgender men and transgender women.* New England Journal of Medicine, 2021. **385**(7): p. 595-608.
96. Olsen, R.B., et al., *External validity in policy evaluations that choose sites purposively.* Journal of Policy Analysis and Management, 2013. **32**(1): p. 107-121.

97. Haneuse, S., et al., *Adjustment for selection bias in observational studies with application to the analysis of autopsy data.* Neuroepidemiology, 2009. **32**(3): p. 229-239.
98. Choi, J., O.M. Dekkers, and S. le Cessie, *A comparison of different methods to handle missing data in the context of propensity score analysis.* European Journal of Epidemiology, 2019. **34**: p. 23-36.
99. Luehring-Jones, P., et al., *Pre-exposure prophylaxis (PrEP) use is associated with health risk behaviors among moderate-and heavy-drinking MSM.* AIDS Education and Prevention, 2019. **31**(5): p. 452-462.


A
X_0
D_1
Y_1
X_1
D_2
Y_2


A
Minimum Probabilities
Pr(A=0|X0)
Pr(D1=0|X0)
Pr(D1=1|X0)
Pr(D1=2|X0)
Pr(D2=0|X1)
Pr(D2=1|X1)
Pr(D2=2|X1)
0.00
0.05
0.10
0.15
Exposure Probability
B
Maximum Probabilities
Pr(A=0|X0)
Pr(D1=0|X0)
Pr(D1=1|X0)
Pr(D1=2|X0)
Pr(D2=0|X1)
Pr(D2=1|X1)
Pr(D2=2|X1)
0.8
0.9
1.0
Exposure Probability


A
Unmeasured Confounding: A
Relative Risk for Exposure
0
1
2
0.5
0.63
0.71
0.8
1
1.25
1.4
1.6
2
α (Bias Factor)
B
Unmeasured Confounding: D
Relative Risk for Exposure
0.0
0.5
1.0
1.5
2.0
2.5
0.5
0.63
0.71
0.8
1
1.25
1.4
1.6
2
α(0, 1) (Bias Factor)


A
Covariate Balance: A
B
Covariate Balance: D (Time = 1)
C
Covariate Balance: D (Time = 2)
Region
Age
Northeast
South
West
Gay
Number of Partners
STI
URAI
0.00
0.25
0.50
0.75
1.00
Standardized Difference