Tran Supplemental Appendix

Table of Contents

INVERSE PROBABILITY OF EXPOSURE AND ATTRITION WEIGHTS 2

QUANTITATIVE BIAS ANALYSIS FOR OUTCOME NONDIFFERENTIAL MISCLASSIFICATION 3

SUPPLEMENTARY FIGURES & TABLES 4

Supplemental Figure 1. Directed acyclic graph for the association of pre-exposure prophylaxis uptake and bacterial sexually transmitted infections (STI) over follow-up in the Diary arm of the iCruise Study, Ontario, Canada, 2017-2018. 4
Supplemental Figure 2. Distribution of nonstabilized and stabilized weights for exposure by pre-exposure prophylaxis uptake over follow-up. The box for each group shows the median (middle horizontal bar) as well as the first and third quartiles (border horizontal bars). Vertical lines extend to observations that are 1.5 × interquartile range. 5
Supplemental Figure 3. Distribution of unstabilized and stabilized weights (exposure weights × lost to follow-up weights) summed across all follow-up visits by pre-exposure prophylaxis uptake. The box for each group shows the median (middle horizontal bar) as well as the first and third quartiles (border horizontal bars). Vertical lines extend to observations that are 1.5 × interquartile range. 6
Supplemental Figure 4. Lagged association of PrEP uptake and incidence rate ratio (95% confidence interval) of any bacterial sexually transmitted infections, chlamydia, gonorrhea, and syphilis 7

Supplemental Table 1. Characteristics of gay, bisexual, and other men who have sex with men by study completion in the Diary arm of the iCruise study, Ontario, Canada, 2017-2018 8
Supplemental Table 2. Characteristics of the study population for the LA LGBT Center and iCruise study 10
Supplemental Table 3. Summary of parameters from external literature for quantitative bias analysis of nondifferential outcome misclassification 11

INVERSE PROBABILITY OF EXPOSURE AND ATTRITION WEIGHTS

Inverse probability of exposure weights (IPEW) was used to create a weighted sample in which exposure assignment is unconfounded by measured covariates, emulating a randomized controlled trial.[1–3] The IPEW are the inverse of the probability of being exposed (or unexposed) at each study visit. Using logistic regression, we fitted nonstabilized and stabilized IPEW for PrEP uptake as follows:

$$w_{ij} = \prod_{j=0}^{int(t)} \frac{1}{P(A_{ij} = a_{ij} | \bar{A}_{i(j-1)} = \bar{a}_{i(j-1)}, \bar{L}_{ij} = \bar{l}_{ij}}$$

and

$$sw_{ij} = \prod_{j=0}^{int(t)} \frac{P(A_{ij} = a_{ij} | \bar{A}_{i(j-1)} = \bar{a}_{i(j-1)}, \boldsymbol{V_{i0}} = \boldsymbol{v_{i0}})}{P(A_{ij} = a_{ij} | \bar{A}_{i(j-1)} = \bar{a}_{i(j-1)}, \bar{L}_{ij} = \bar{l}_{ij}}$$

where w_{ij} and sw_{ij} are the nonstabilized and stabilized weight for person i at time j, respectively; $int(i)$ = largest integer $\leq t$; A_{ij} = PrEP uptake for person i at time j; a_{ij} = observed PrEP uptake for person i at time j; $\bar{A}_{i(j-1)}$ = PrEP use history for person i through time $j-1$; $\bar{a}_{i(j-1)}$ = observed PrEP uptake history for person i through time $j=1$; $\bar{L}_{ij}$ = time-invariant and time-varying covariate history for person i through time j; $\bar{l}_{ij}$ = observed time-invariant and time-varying covariate history for person i through time j; $\boldsymbol{V_{i0}}$ = time-invariant covariates at baseline; and $\boldsymbol{v_{i0}}$ = observed time-invariant at baseline for person i. The final stabilized IPEW used in the analyses was created by multiplying the stabilized weights across all time

Inverse probability of attrition weights was used to create to account for selection bias due to differential lost to follow-up. Based on potential differences between the iCruise participants who completed the study compared to those who were lost to follow-up, I constructed attrition weights using logistic regression in which the probability of attrition at the end of follow-up was conditional on age, sexual orientation, education, alcohol use, baseline STI diagnosis, and baseline PrEP uptake, where:

$$w_c = \frac{1}{P(C = 0 | A, V)}$$

We defined attrition as $C = 1$ if a participant was lost to follow-up at the end of the survey (i.e., missed four more consecutive diary surveys) and $C = 0$ otherwise. Further details on the characteristics of men by study completion are presented below in **Supplemental Table 1**.

QUANTITATIVE BIAS ANALYSIS FOR OUTCOME NONDIFFERENTIAL MISCLASSIFICATION

In secondary analysis of data collected from online studies, there are potential biases that distort the accuracy of the exposure-outcome relation. For the analysis, we conducted a quantitative bias analysis to understand the magnitude and direction in which nondifferential outcome misclassification may have impacted the findings. These methods were adapted based on the work by Goldstein et al. and readers who are interested in learning about the connection between theory and applied methodology should refer to this paper.[4]

We assumed nondifferential misclassification, in that, self-reported bacterial STI incidence, Y^*, was likely to underestimate the true incidence, Y, given that these infections are often asymptomatic[92] and is independent of the exposure, PrEP uptake (A). We also assumed that the misclassified outcome, Y^*, was related to the true outcome, Y, based on the sensitivity (SN), specificity (SP), negative predictive value (NPV), and positive predictive value (PPV).

To obtain these accuracy measures and their complements (false omission rate [FOR] and false discovery rate [FDR]), we calculated them using an external validation study,[5] assuming that estimates from the validation study are transportable to the iCruise study. A comparison between the validation cohort and iCruise participants are presented in **Supplemental Table 2**.

Based on data from the Los Angeles LGBT Center, we estimated the accuracy parameters for any bacterial STI and each specific microorganism as well as the 95% confidence intervals (**Supplemental Table 3**).

We then calculated β_0 and β_1 which represents the intercept and slope of the relation between the misclassified (Y^*) and true outcome (Y):

$$Logit(Y) = \beta_0 + \beta_1(Y^*)$$

To estimate β_0, we used the following equation: $\ln\left(\frac{1-NPV}{NPV}\right)$. This equation was derived from:

$$NPV = 1 - \frac{\exp(\beta_0)}{1 + \exp(\beta)}$$

To estimate β_1, we used the following equation: $\ln\left(\frac{PPV}{1-PPV}\right) - \beta_0$. This equation was derived from:

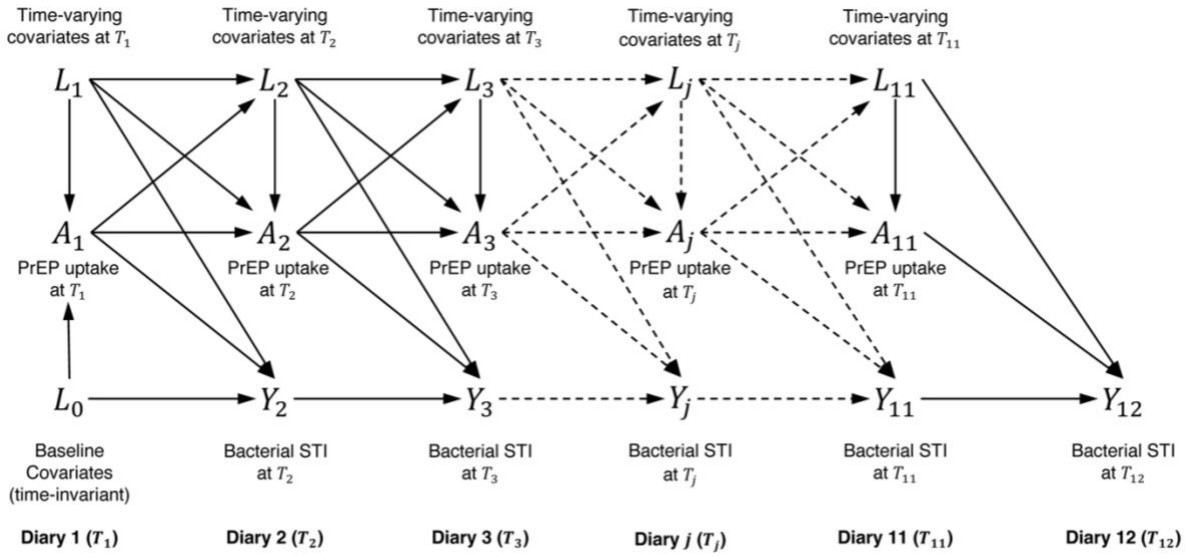
$$PPV = \frac{\exp(\beta_0 + \beta_1)}{1 + \exp(\beta_0 + \beta_1)}$$

Finally, we simulated the potential values of Y over 1,000 replications to reflect what the true incidence would have been if there was no nondifferential misclassification. Precision is estimated using bootstrapping and is expressed as 95% simulation intervals.

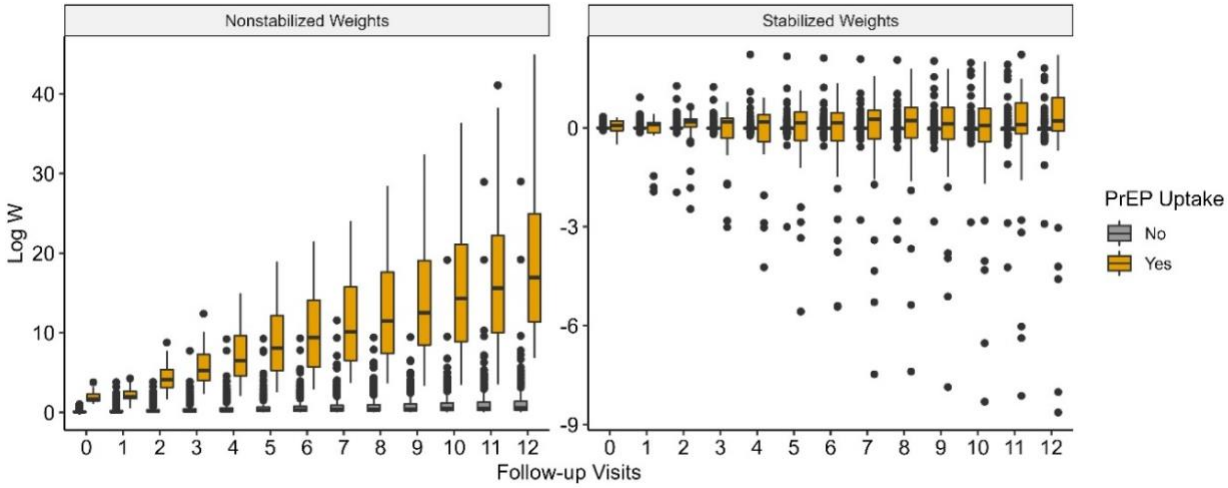
SUPPLEMENTAL FIGURES & TABLES

Supplemental Figure 1. Directed acyclic graph for the association of pre-exposure prophylaxis uptake and bacterial sexually transmitted infections (STI) over follow-up in the Diary arm of the iCruise Study, Ontario, Canada, 2017-2018.

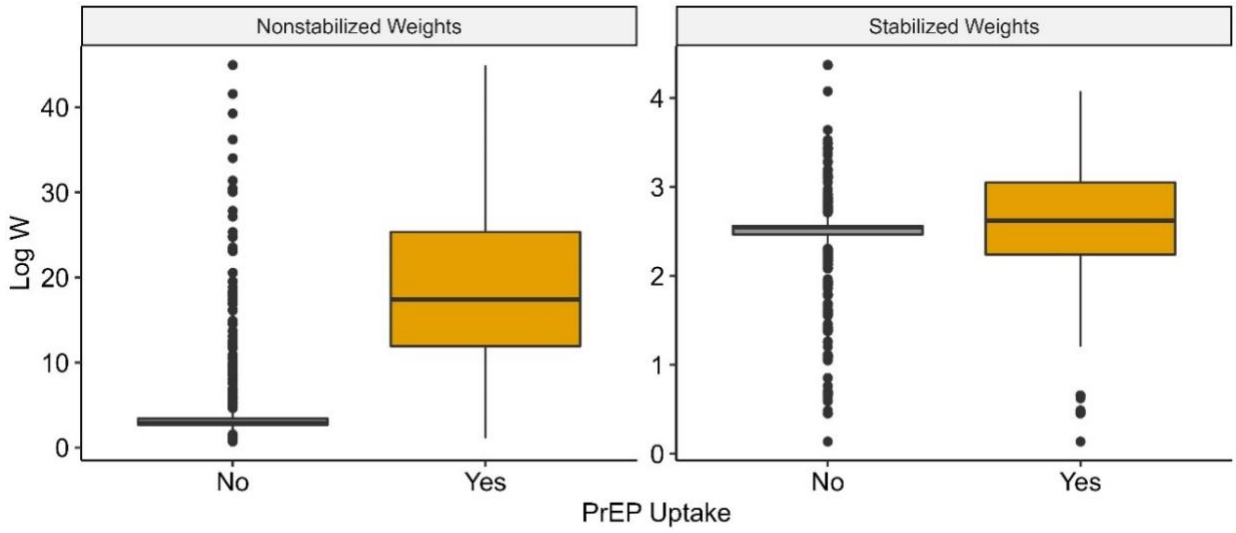
L_0 represents the baseline covariates (age, employment, income, access to primary care provider, alcohol and drug use, and lifetime bacterial STI diagnosis). L_t represents the time-varying covariates (number of anal sex partners, condomless anal sex, bacterial STI screening). A_0 and A_t represent PrEP uptake at baseline and over follow-up. Y represent cumulative incidence of bacterial STI over 12 weeks of follow-up.



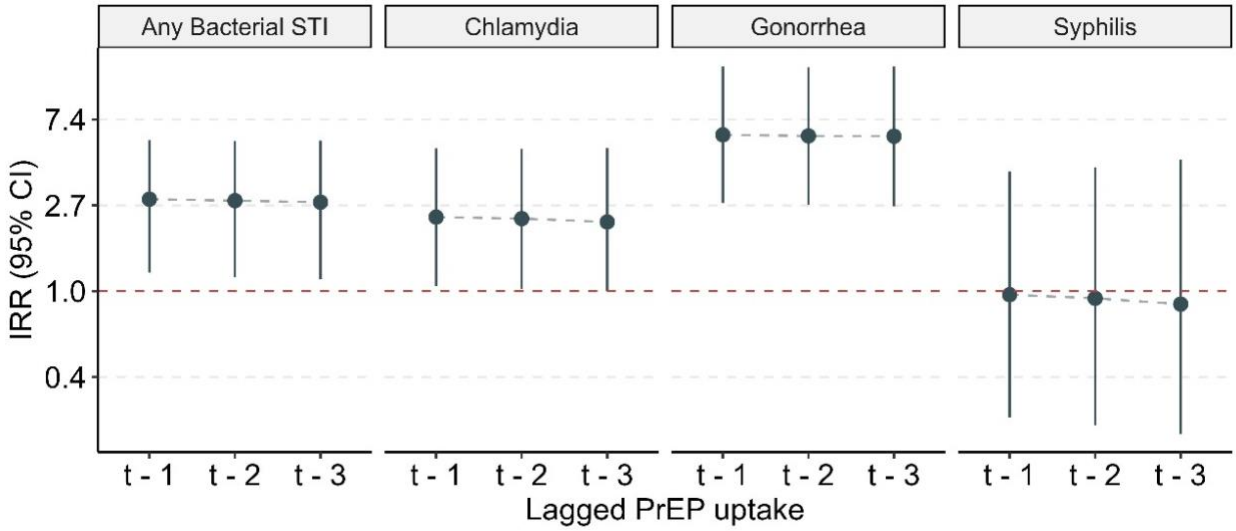
Supplemental Figure 2. Distribution of nonstabilized and stabilized weights for exposure by pre-exposure prophylaxis uptake over follow-up. The box for each group shows the median (middle horizontal bar) as well as the first and third quartiles (border horizontal bars). Vertical lines extend to observations that are 1.5 × interquartile range.



Supplemental Figure 3. Distribution of unstabilized and stabilized weights (exposure weights × lost to follow-up weights) summed across all follow-up visits by pre-exposure prophylaxis uptake. The box for each group shows the median (middle horizontal bar) as well as the first and third quartiles (border horizontal bars). Vertical lines extend to observations that are 1.5 × interquartile range.



Supplemental Figure 4. Lagged association of PrEP uptake and incidence rate ratio (95% confidence interval) of any bacterial sexually transmitted infections, chlamydia, gonorrhea, and syphilis.



Supplemental Table 1. Characteristics of gay, bisexual, and other men who have sex with men by study completion in the Diary arm of the iCruise study, Ontario, Canada, 2017-2018

	Completed study (*n* = 462)		Lost to follow up (*n* = 73)		*P* value
Characteristic	No.	%	No.	%	
Mean age, years (SD)	34.5 (13.0)		30.6 (11.2)		0.007
Racioethnic identity (*n* = 532)					0.23
African/African Caribbean/Black	17	3.7	2	2.7	
East/Southeast/South Asian	72	15.6	11	15.1	
Indigenous	24	5.2	6	8.2	
Latino	34	7.4	9	12.3	
Middle Eastern/North African	11	2.4	2	2.7	
Multiracial	29	6.3	2	2.7	
White	273	59.1	40	54.8	
Sexual orientation					0.04
Bisexual	79	17.1	23	31.5	
Gay	330	71.4	39	53.4	
Queer	35	7.6	7	9.6	
Other[a]	18	3.9	4	5.5	
Education					0.03
High school or less	46	10.0	12	16.4	
College/trade/technical	145	31.4	14	19.2	
University/Post-graduate	271	58.7	47	64.4	
Employment (*n* = 527)					0.38
Employed (Full-time/Part-time)	327	71.7	51	71.8	
Student (Full-time/Part-time)	47	10.3	9	12.7	
ODSP/OW/EI	41	9.0	5	7.0	
Unemployed	41	9.0	6	8.5	
Income (*n* = 505)					0.25
$0 – $19,999	128	29.2	19	23.9	
$20,000 – $39,999	114	26.0	26	31.3	
$40,000 – $59,999	101	23.1	13	16.4	
$60,000 – $79,999	48	11.0	13	14.9	
$80,000 or more	47	10.7	10	13.4	
Alcohol use, 3 months	385	83.3	67	91.8	0.03
Drug use, 3 months[b]	134	29.0	26	35.6	0.31
Number of anal sex partners, 3 months					
0	290	62.8	44	60.3	0.64
1 – 2	72	15.6	12	16.4	
3 – 5	53	11.5	5	6.8	
6 – 9	27	5.8	6	8.2	
10 or more	20	4.3	6	8.2	
Relationship types, 3 months (n = 533)					0.22
No partners	252	54.5	33	45.2	
Main partners exclusively	17	3.7	1	1.4	
Casual or one-time partners	139	30.1	21	28.8	
Main and casual or one-time partners	54	11.7	18	24.7	
Sexual positions, 3 months (*n* = 532)					0.29
No partners	293	63.4	44	60.3	
Insertive	42	9.1	6	8.2	

Receptive	51	11.0	7	9.6	
Versatile	76	16.5	16	21.9	
Condomless anal sex, 3 months	79	17.1	18	24.7	0.36
STI testing, 3 months	170	36.8	23	31.5	0.36
Lifetime STI diagnosis	210	45.5	32	43.8	0.71
Baseline PrEP uptake	45	9.7	8	11.0	0.87

Abbreviations: SD = standard deviation; IQR = interquartile range; ODSP = Ontario Disability Support Program; OW = Ontario Works; EI = employment insurance; STI = sexually transmitted infections

[a] Included men who identified as questioning, asexual, pansexual, and two-spirit

[b] Used the following drugs 2 hours before sex or during sex: cocaine, crystal methamphetamine

Supplemental Table 2. Characteristics of the study population for the LA LGBT Center and iCruise study

	LA LGBT Center (n = 10,529)	Ontario iCruise Study (n = 535)
	n (%)	n (%)
Age Group		
<25	1,567 (15)	128 (24)
25-29	2,717 (26)	147 (27)
30-39	3,527 (33)	122 (23)
≥40	2,718 (26)	138 (26)
Race/Ethnicity		
Black	833 (8)	18 (3)
Hispanic	3,125 (30)	45 (8)
White	5,180 (49)	314 (59)
Other	1,382 (13)	158 (29)
Education level		
High school or below	1,337 (13)	58 (11)
Some college	2,663 (25)	158 (29)
College or higher	5,421 (51)	319 (59)
History of bacterial STI diagnosis		
Chlamydia	2,275 (22)	112 (21)
Gonorrhea	2,669 (25)	115 (22)
Syphilis	289 (3)	47 (9)

Abbreviations: LA = Los Angeles

Supplemental Table 3. Summary of parameters from external literature for quantitative bias analysis of nondifferential outcome misclassification

	Estimate (95% confidence interval)			
Parameter	Any bacterial STI	Chlamydia	Gonorrhea	Syphilis
SN	0.778 (0.758, 0.797)	0.732 (0.699, 0.763)	0.809 (0.782, 0.834)	0.864 (0.770, 0.930)
SP	0.949 (0.946, 0.952)	0.945 (0.939, 0.951)	0.924 (0.917, 0.931)	0.966 (0.962, 0.970)
PPV	0.557 (0.541, 0.572)	0.619 (0.592, 0.644)	0.629 (0.607, 0.651)	0.182 (0.163, 0.204)
NPV	0.981 (0.979, 0.983)	0.967 (0.963, 0.970)	0.968 (0.964, 0.972)	0.999 (0.999, 0.999)

Abbreviations: SN = sensitivity; SP = specificity; PPV = positive predictive value; NPV = negative predicted value.

REFERENCES

1. Robins JM, Hernán MÁ, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11(5):550-560. doi:10.1097/00001648-200009000-00011

2. Cole SR, Hernán MA. Constructing Inverse Probability Weights for Marginal Structural Models. *Am J Epidemiol*. 2008;168(6):656-664. doi:10.1093/AJE/KWN164

3. Hernán MA, Brumback BA, Robins JM. Estimating the causal effect of zidovudine on CD4 count with a marginal structural model for repeated measures. *Stat Med*. 2002;21(12):1689-1709. doi:10.1002/sim.1144

4. Goldstein ND, Kahal D, Testa K, et al. Data Quality in Electronic Health Record Research: An Approach for Validation and Quantitative Bias Analysis for Imperfectly Ascertained Health Outcomes Via Diagnostic Codes. *Harv Data Sci Rev*. 2022;4(2). doi:10.1162/99608F92.CBE67E91

5. Cunningham NJ, Beymer MR, Javanbakht M, Shover CL, Bolan RK. Concordance between self-reported STI history and biomedical results among men who have sex with men in Los Angeles, California. *Sex Transm Infect*. 2017;93(7):514-519. doi:10.1136/sextrans-2016-052933