Adjusting for duration biases in sexual behaviour data

Jesse Knight[1,2] and Sharmistha Mishra[1,2]

[1]Institute of Medical Science, University of Toronto
[2]MAP Centre for Urban Health Solutions, Unity Health Toronto

June 30, 2023

Abstract

TODO

Funding.

Aknowledgements. We wish to thank: Saulius Simcikas and Dr. Jarle Tufto for help identifying and discussing censoring biases.[1]

[1] stats.stackexchange.com/questions/298828

1 Introduction

Quantifying sexual behaviour is necessary to study the epidemiology of sexually transmitted infections (STI), including to inform inputs for STI transmission modelling [4]. Two important quantities are: the duration of time within a "risk group" such as female sex workers (FSW), and the rate of new partnership formation, possibly stratified by partnership type.

Our aims are to motivate and discuss bias adjuments for estimating:

1. duration in a risk group
2. rate of partnership change

from cross-sectional survey data, considering issues of sampling bias and censoring. We explore these topics using aggregate data from two female sex worker (FSW) surveys in Eswatini: [1] (2011, RDS sampling, $N = 328$) and [5] (2014, PLACE sampling, $N = 781$).

2 Methods

2.1 Risk Group Duration

The FSW survey data include questions about the current respondent's age, and the age of first selling sex. The difference between these ages could be used to define a "duration selling sex". Using this naive approach, the unadjusted durations D'_s among FSW in Eswatini were median [IQR] years: 4 [2, 7] in 2011 and 5 [3, 9] in 2014. However, such estimates have three sources of bias: sampling error, measurement error, and censoring.

Sampling error was considered via RDS-adjustment in the 2011 survey (only), yielding estimates of the proportions of FSW who had sold sex starting 0–2, 3–5, 6–10, and 10+ years ago. The adjusted proportions indicate fewer years selling sex *vs* the unadjusted proportions, which would be consistent with challenges in reaching women in the first year(s) of sex work [2]. Fitting an exponential distribution to the cumulative adjusted proportions (Appendix A.1)

(a) TODO (b) TODO

yielded an estimated distribution mean $\bar{D}_S$ of 4.2 (95% CI: 3.5, 5.3) years.

Regarding measurement error: FSW may not sell sex continuously. The 2014 survey (only) asked whether respondents ever stopped selling sex and 348/777 (45%) had stopped at least once. Among these FSW, the expected duration selling sex in the current period (*i.e.,* since re-starting most recently) must be less than half ($\rho < 1/2$) of the durations above, depending on the number and lengths of gaps in selling sex. Thus, a further adjustment to the mean duration could be defined as: $(0.45\,\rho + 0.55)\,\bar{D}_S$, with $\rho \sim \text{Unif}\,(0.2, 0.4)$ as an assumption.

Finally, these durations are also right censored because almost all respondents will continue selling sex after the survey [3]. If we assume that the survey reaches FSW at a random time point during their total (eventual) duration selling sex D, then the duration reported in the survey is effectively $D_S \sim \text{Unif}\,(0, D)$. Thus, the mean duration reported in the survey is $\bar{D}_S = \frac{1}{2}\bar{D}$, and we can define $f = \bar{D}/\bar{D}_S = 2$, to give the final adjusted estimate as: $\bar{D} = f\,(0.45\,\rho + 0.55)\,\bar{D}_S$. In case the RDS-adjustment did not actually account for delayed self-identification as FSW [2], we could use $f \sim \text{Unif}\,(1.5, 2)$.

Figure 1a illustrates the censoring issue in a steady-state population with 5 women selling sex at any given time. Another observation we can make from Figure 1a is that women who sell sex longer are more likely to be captured in the survey. That is, while the sampled durations are representative of women who *currently* sell sex, these durations are biased high *vs* the population of women who *ever* sell sex.

2.2 Sexual Partnership Duration

The FSW surveys asked respondents to report their numbers of unique sexual partners in the past 30 days, stratified by three types of partner: new paying clients, regular paying clients, and non-paying partners. For illustrative purposes, we assume that only a small proportion of new clients would go on to become regular clients; thus, we conceptualize "new" clients as effectively "one-off" clients. We further assume that partnership durations were: 1 day with new paying clients, 4 months with regular paying clients, and 3 years with non-paying partners (no survey questions asked about partnership durations) [**?**].

Using the published aggregate statistics in [1, 5], the reported numbers of partnerships of each type were, respectively: 1.77, 4.69, and 0.74 in 2011 (RDS-adjusted medians), and 5.04, 8.03, and 1.22 in 2014 (unadjusted means). Our aim is to use these reported numbers of partners (x) for the 30-day recall period (ω), with the assumed partnership durations (δ), to define expected rates of partnership change (Q) and numbers of concurrent partnerships (K).

We begin with some general observations about the relationships between the variables x, ω, δ, Q, K. If partnership duration is long and the recall period is short ($\delta \gg \omega$), the reported

partnerships mostly reflect *ongoing* partnerships, and thus $x \approx K$. If partnership duration is short and the recall period is long ($\delta \ll \omega$), the reported partnerships mostly reflect *complete* partnerships, and thus $x/\omega \approx Q$. However, if partnership duration and recall period are similar in length, the reported partnerships reflect a mixture of tail-ends, complete, and ongoing partnerships, and thus x overestimates K, but x/ω also overestimates Q.

Next, we make a similar assumption as in § 2.1: that survey timing is effectively random with respect to partnership duration. Then, if either end of the recall period would capture an ongoing partnership, the intersection point would be, on average, at the partnership mid-point. Thus, the recall period is effectively extended by half the partnership duration $\delta/2$ on each end, and δ overall, as illustrated in Figure 1b. As such, we can define Q and K as:

$$Q = \frac{x}{\omega + \delta} \tag{1}$$

$$K = \frac{x\delta}{\omega + \delta} = Q\delta \tag{2}$$

As an example, Figure 1b also illustrates a recall period of $\omega = 1$ year, for which $x = 9$ partnerships are reported, having durations of $\delta = 0.75$ years. Thus, we can compute $Q = 9/(1 + 0.75) = 5.14$ partnerships per year and $K = 5.14(0.75) = 3.86$ current partners; these ars slight underestimates of the true values $Q = 5.33, K = 4$, due to randomness in the exact "location" of the recall period.

3 Results

3.1 Risk Group Duration

Figure 2

3.2 Sexual Partnership Duration

4 Discussion

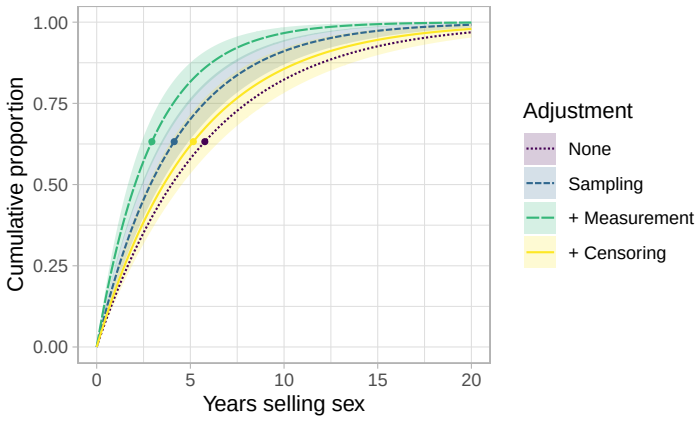

Figure 2: TODO

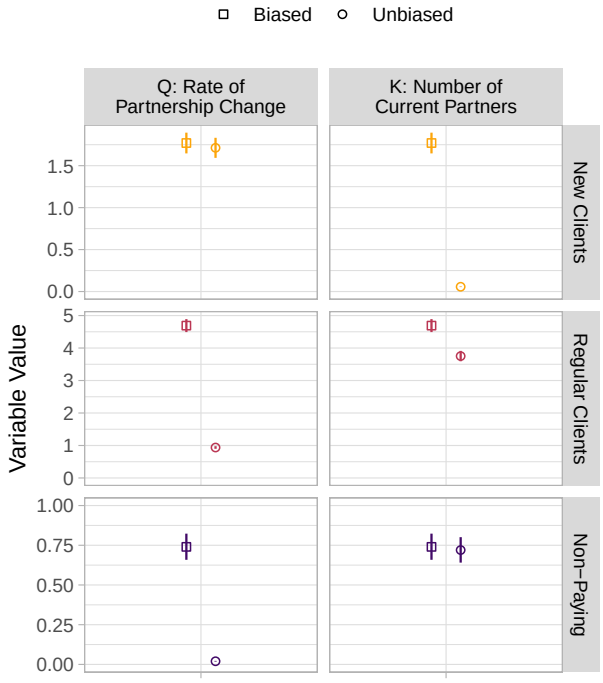

Figure 3: Biased vs unbiased estimates of: number of current partners and rate of partnership change, for three partnership types reported by female sex workers in [1]. Error bars show 95% CI from 10,000 simulated surveys with $N = 328$.

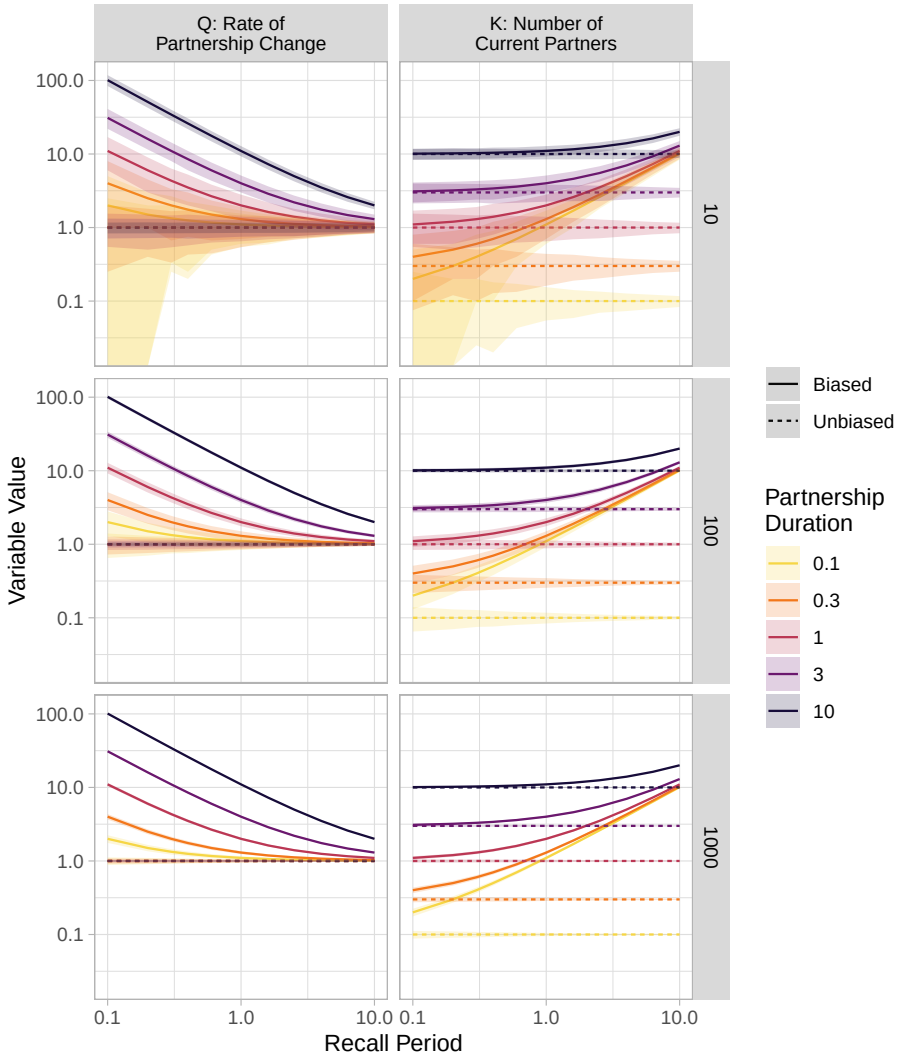

Figure 4: Biased vs unbiased estimates of: number of current partners and rate of partnership change, for different recall periods and partnership durations. Ribbons show 95% CI from 10,000 simulated surveys with N = 10, 100, 1000.

References

[1] Stefan Baral et al. "Reconceptualizing the HIV epidemiology and prevention needs of female sex workers (FSW) in Swaziland". In: *PLOS ONE* 9.12 (Dec. 22, 2014), e115465. URL: http://doi.org/10.1371/journal.pone.0115465.

[2] Eve Cheuk et al. "Transitions: Novel Study Methods to Understand Early HIV Risk Among Adolescent Girls and Young Women in Mombasa, Kenya, and Dnipro, Ukraine". In: *Frontiers in Reproductive Health* 2 (Sept. 10, 2020), p. 10. URL: https://doi.org/10.3389/frph.2020.00007.

[3] Erika Fazito et al. "Analysis of duration of risk behaviour for key populations: a literature review." In: *Sexually transmitted infections* 88.S2 (Dec. 1, 2012), pp. i24–i32. URL: https://doi.org/10.1136/sextrans-2012-050647.

[4] K. A. Fenton et al. "Measuring sexual behaviour: Methodological challenges in survey research". In: *Sexually Transmitted Infections* 77.2 (Apr. 1, 2001), pp. 84–92. URL: https://doi.org/10.1136/sti.77.2.84.

[5] PEPFAR. *Characterizing the HIV Prevention and Treatment Needs among Key Populations, including Men who Have Sex with Men and Female Sex Workers in Swaziland: From Evidence to Action*. Mbabane, Swaziland, 2015.

Table A.1: TODO

Adjustment	Mean	95% CI
None	5.77	—
Sampling	4.13	(3.43, 4.91)
+ Measurement	2.94	(2.39, 3.56)
+ Censoring	5.15	(3.95, 6.55)

Table A.2: TODO

		Rate Q		Number K	
Partnership Type	Bias	Mean	95% CI	Mean	95% CI
New Clients	Biased	1.77	(1.65, 1.89)	1.77	(1.65, 1.89)
	Unbiased	1.71	(1.60, 1.83)	0.06	(0.05, 0.06)
Regular Clients	Biased	4.69	(4.49, 4.89)	4.69	(4.49, 4.89)
	Unbiased	0.94	(0.90, 0.98)	3.75	(3.60, 3.91)
Non-Paying	Biased	0.74	(0.66, 0.82)	0.74	(0.66, 0.82)
	Unbiased	0.02	(0.02, 0.02)	0.72	(0.64, 0.80)

A Supplement

A.1 Risk Group Duration

A.2 Sexual Partnership Duration

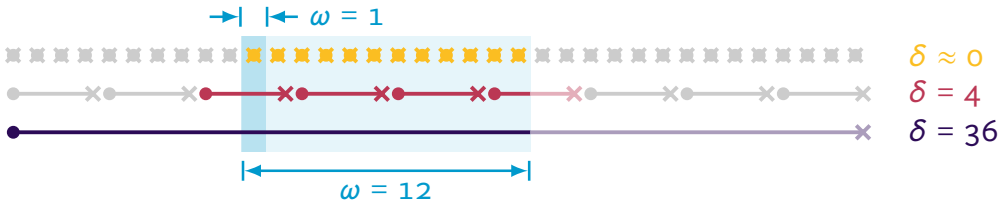

Figure A.1: TODO