August 16, 2023

Dr. Samuel Jenness
Editor
Epidemiology

Dr. Sharmistha Mishra
MAP Centre for Urban Health Solutions
St. Michael's Hospital, Unity Health Toronto
University of Toronto

Re: Submission of a manuscript to *Epidemiology*

Dear Dr. Jenness,

We are pleased to submit the attached manuscript entitled *Adjusting for hidden biases in sexual behaviour data: a mechanistic approach* for consideration as a *Validation* study in *Epidemiology*.

Quantitative estimates of sexual behaviour are required as inputs to mathematical models of sexually transmitted infections, and in other studies of sexually transmitted infection epidemiology. Such estimates are often derived from cross-sectional surveys. While previous work has explored established biases associated with survey-based estimates (e.g., recall bias, reporting bias), less attention has been paid to other "hidden" biases, such as estimate-estimand mismatch (e.g., numbers of partners in the past 30 days vs partnership change rate).

In this study, we explore precise interpretation of survey data to inform two key variables: durations in epidemiological risk states (e.g., selling sex) and rates of sexual partnership change (e.g., casual partners per year). We identify potential sources of bias, and develop Bayesian hierarchical models to reflect mechanistic assumptions about the bias-generating processes. Fitting these models to aggregate data from a previously published study of female sex workers in Eswatini, we show that failure to account for particular biases can substantially influence estimates of both variables explored.

While we focus on sexual behavior data, we expect that our approach and findings would be relevant to other estimates of intermittent risk exposure and event rates from cross-sectional surveys, and thus of interest to the broad readership of *Epidemiology*.

Thank you for your consideration and we look forward to hearing from you.

Sincerely,

Jesse Knight, Siyi Wang, and Sharmistha Mishra

Title Adjusting for hidden biases in sexual behaviour data: a mechanistic approach

Authors Jesse Knight[1,2], Siyi Wang[1], and Sharmistha Mishra[1,2]

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

Date August 16, 2023

Journal Epidemiology

Status draft

Funding The study was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC CGS-D); the Ontario Ministry of Colleges and Universities and the University of Toronto (QEII-GSST); the Canadian Institutes of Health Research Foundation Grant (FN-13455); the National Institute of Allergy and Infectious Diseases (R01AI170249).

Data & Code `github.com/mishra-lab/duration-bias`

Acknowledgements We thank: Linwei Wang, Korryn Bodner, Amrita Rao, Kate Rucinski, Le Bao, and Stefan Baral for helpful discussions on epidemiological implications of the work; Jarle Tufto and Michael Neely for guidance on mathematical modelling of censored durations and Poisson processes.

Abstract Two required inputs to mathematical models of sexually transmitted infections are the average duration in epidemiological risk states (e.g., selling sex) and the average rates of sexual partnership change. These variables are often only available as aggregate estimates from published cross-sectional studies, and may be subject to distributional, sampling, censoring, and measurement biases. We explore adjustments for these biases using aggregate estimates of duration in sex work and numbers of reported sexual partners from a published 2011 survery of female sex worker in Eswatini. We develop adjustments from first principles, and construct Bayesian hierarchical models to reflect our mechanistic assumptions about the bias-generating processes. We show that different mechanisms of bias for duration in sex work may "cancel out" by acting in opposite directions, but that failure to consider some mechanisms could over- or underestimate duration in sex work by factors approaching 2. We also show that conventional interpretations of sexual partner numbers are biased due to implicit assumptions about partnership duration, but that unbiased estimators of partnership change rate can be defined that explicitly incorporate a given partnership duration. We highlight how the unbiased estimator is most important when the survey recall period and partnership duration are similar in length. While we explore these bias adjustments using a particular dataset, and in the context of deriving inputs for mathematical modelling, we expect that our approach and insights would be applicable to other datasets and motivations for quantifying sexual behaviour data.

Keywords bias, uncertainty, Monte Carlo method, sexual behavior, sexual partners, sex work, sexually transmitted diseases

1 Introduction

Mathematical models of sexually transmitted infections require quantitative estimates of sexual behaviour for model inputs (parameters) [1]. In such models with risk heterogeneity — i.e., considering states that experience differential risks — two important parameters are: the duration of time within an epidemiological risk state/group (or period/season of risk) and the rate of sexual partnership change (often stratified by partnership type) [2, 3, 4, 5]. For example, the average duration of time engaged in sex work can be used to define the modelled rate of "turnover" among sex workers [4]. Similarly, the numbers of main, casual, transactional, and/or paying sexual partners per year can be used to define the modelled rate of infection incidence [6].

Data to inform these parameters largely come from cross-sectional studies, and are often only available as aggregate estimates (vs individual-level data). Such estimates may be subject to several "hidden" sources of bias which can easily go unnoticed, including distributional, sampling, censoring, and measurement biases. Our aim is therefore to explore bias adjuments for estimating: (1) duration in a risk state, and (2) rate of partnership change, from aggregate cross-sectional survey data. We explore and demonstrate these topics using data from a 2011 female sex worker survey in Eswatini [7], to support parameterization of a mathematical model of heterosexual HIV transmission.

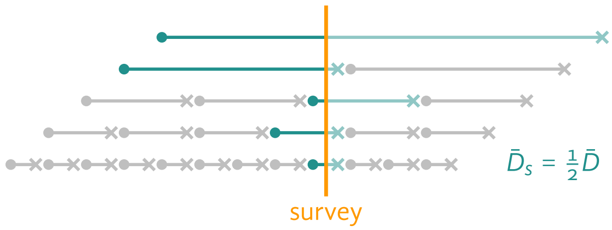
2 Methods

Data Source. Full details of the survey methodology are available in [8]. Briefly, 328 women aged 15+ who reported exchanging or selling sex for money, favors, or goods in the past 12 months were recruited via respondent-driven sampling (RDS) [9].

Approach. We conceptualize bias adjustments to the given data using Bayesian hierarchical models. Specifically, we define explicit distributions for the unbiased data and bias-generating mechanisms, and infer the parameters of these distributions based on the available data, using Gibbs sampling [10]. Implementation details are given in Appendix A.2. Figure 1 gives some supporting diagrams, while Figure 2 illustrates the complete models.

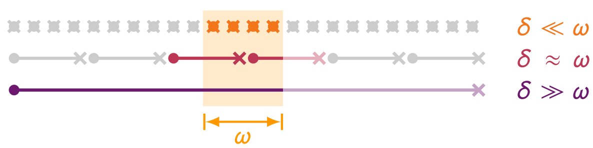
2.1 Duration Selling Sex

Crude Estimates. The survey [7] included 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 crude "duration selling sex". Using this approach, the crude median duration was $\tilde{d} = 4$ years. However, if durations are assumed to be exponentially distributed — a implicit assumption in compartmental models [11] — then the crude mean could be estimated from the crude median as $\bar{d} = \tilde{d}/\log(2)$ due to skewness. To move beyond crude estimates, next we develop the hierarchical model, considering the following potential biases.

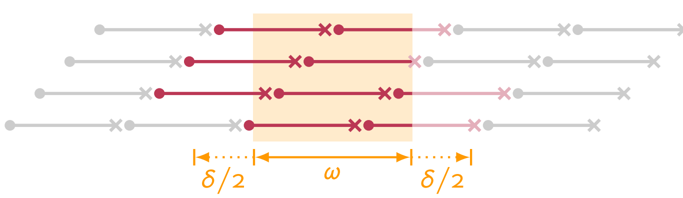


(a) Right censoring of reported durations selling sex in a steady state population

(b) Possible periods of selling sex for one respondent who stopped 0, 1, or 2 times



(c) Differences in partnership duration vs recall period



(d) Fully and partially observed partnerships during a given recall period

Figure 1: Diagrams of fully observed, censored, and unobserved periods selling sex or within ongoing sexual partnerships

Guide: • : start, × : end, yellow: survey/recall period, full colour: fully observed, faded colour: right censored, grey: unobserved, $\bar{d}$: mean duration at survey and $\bar{D}$: overall, s: number of times stopped selling sex, g: relative gap length vs D, ω: recall period, δ: partnership duration, x: number of reported partnerships.

Sampling. Sampling bias was considered via RDS-adjustment in [7], yielding mean and 95% CI estimates of the proportions of respondents p_z who had sold sex starting $\mathbb{d}_z \in \{0–2, 3–5, 6–10, 11+\}$ years ago (Table A.1, "z" enumerates strata). We start by defining a model to identify distributions of reported durations d_i which are consistent with these data. We model each proportion p_z as a random variable with a beta approximation of binomial (BAB) distribution (see Appendix A.3) with parameters N_z and ρ_z. We model each N_z as a fixed value, which we fit to the 95% CI of p_z as described in § A.3. We then model each ρ_z as the proportion of reported durations d_i within the interval $\mathbb{d}_z$. Since these proportions are difficult to define analytically, we estimate $\hat{\rho}_z = \text{mean}(d_i \in \mathbb{d}_z)$ from $N = 100$ samples.

Censoring. These reported durations d_i are effectively right censored because they only capture engagement in sex work up until the survey, and and not additional sex work after the survey (Figure 1a) [12]. If we assume that the survey reaches respondents at a random time point during their total (eventual) duration selling sex D_i, we can model this censoring via a random fraction $f_i \sim \text{Unif}(0, 1)$, such that $d_i = f_i D_i$; the expected means are then related by $\bar{d}/\bar{D} = \bar{f} = \frac{1}{2}$ [13].

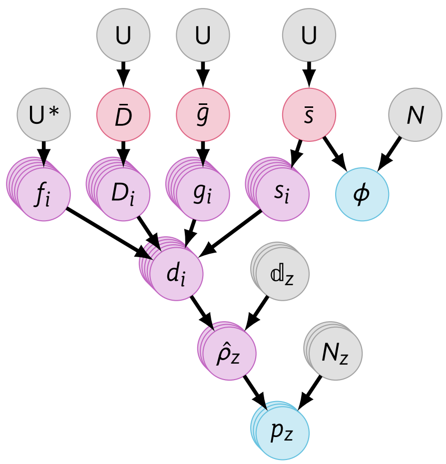
Measurement. Finally, respondents may not sell sex continuously. Reported durations d_i may therefore include multiple periods of selling sex with gaps in between, whereas we aim to model D_i as the durations of individual periods selling sex. Respondents in [7] were not asked whether they ever temporarily stopped selling sex, but a later survey [14] indicated that $\phi = 45\%$ had stopped at least once. We model the number of times a respondent may temporarily stop selling sex as a Poisson-distributed random variable s_i with mean $\bar{s}$. The expected value of ϕ given $\bar{s}$ is then $\text{P}(s > 0) = 1 - e^{-\bar{s}}$. Since $\phi = 45\%$ is an imperfect observation, we model ϕ as a random variable with a BAB distribution having parameters $N = 328$ and $\rho = 1 - e^{-\bar{s}}$, which allows inference on $\bar{s}$ given ϕ.

Next, we update the model for reported durations as $d_i = D_i\left(f_i + s_i\left(1 + g_i\right)\right)$, where g_i is the relative duration of gaps between selling sex, with the following rationale. If $s_i = 0$, then $d_i = f_i D_i$ as before, reflecting the censored current period only. If $s_i > 0$, then d_i also includes s_i prior periods selling sex and the gaps between them (Figure 1b) — i.e., $s_i\left(D_i + g_i D_i\right) = D_i\, s_i\left(1 + g_i\right)$. The major assumption we make here is that all successive periods are of equal length, and likewise for gaps between them. We must also assume a distribution for g_i, for which we choose $g_i \sim \text{Exp}\left(1/\bar{g}\right)$, arbitrarily.

Summary. Figure 2a summarizes the proposed model graphically. The primary parameter of interest is the mean duration selling sex (for a given period) $\bar{D}$, but we must also infer the mean number of times respondents stop selling sex $\bar{s}$, and the mean relative duration of gaps $\bar{g}$. We assume uninformative priors for these 3 parameters.

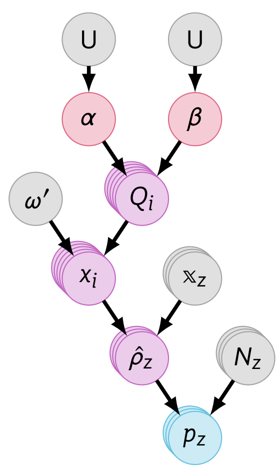
2.2 Rates of Partnership Change

Data & Assumptions. The survey [7] also asked respondents to report their numbers of sexual partners (x) in a recall period (ω) of 30 days. Numbers were stratified by three types of partner: new paying clients, regular paying clients, and non-paying partners. We assume that only a small



$$p_z \sim \text{BAB}\left(N_z, \hat{\rho}_z\right) \tag{1}$$
$$\hat{\rho}_z = \text{mean}\left(d_i \in \mathbb{d}_z\right) \tag{2}$$
$$d_i = D_i\left(f_i + s_i\left(1 + g_i\right)\right) \tag{3}$$
$$D_i \sim \text{Exp}\left(1/\bar{D}\right) \tag{4}$$
$$f_i \sim \text{Unif}\left(0, 1\right) \tag{5}$$
$$s_i \sim \text{Pois}\left(\bar{s}\right) \tag{6}$$
$$g_i \sim \text{Exp}\left(1/\bar{g}\right) \tag{7}$$
$$\phi \sim \text{BAB}\left(N, 1 - e^{-\bar{s}}\right) \tag{8}$$

(a) Duration selling sex



$$p_z \sim \text{BAB}\left(N_z, \hat{\rho}_z\right) \tag{9}$$
$$\hat{\rho}_z = \text{mean}\left(x_i \in \mathbb{x}_z\right) \tag{10}$$
$$x_i \sim \text{Pois}\left(Q_i\, \omega'\right) \tag{11}$$
$$Q_i \sim \text{Gamma}\left(\alpha, \beta\right) \tag{12}$$

(b) Rates of partnership change

Figure 2: Graphical and mathematical representations of the proposed Bayesian hierarchical models

Guide: gray: fixed variable/distribution; red: target; purple: intermediate; blue: observed. Variables: p_z: proportion of population; N_z: effective sample size; $\hat{\rho}_z$: empirically estimated p_z mean; $\mathbb{d}_z$: range of reported durations selling sex; d_i: reported duration at survey; D_i: total (eventual) duration; f_i: censoring fraction; s_i: number of times stopped selling sex; g_i: relative gap length; $\bar{D}$: true D mean; $\bar{s}$: true s mean; $\bar{g}$: true g mean; ϕ: proportion who stopped selling sex at least once; $\mathbb{x}$: range of reported partner numbers; x_i: reported partner numbers; Q_i: partnership change rate; ω': effective recall period; α, β: parameters of Q_i distribution. Distributions: U: uniform / uninformative; BAB: beta approximation of binomial distribution (see § A.3).

proportion of new clients go on to become regular clients; thus, we conceptualize "new" clients as effectively "one-off" clients.[1] Since no survey questions asked about partnership durations (δ), we further assume that these were: 1 day with new paying clients, 4 months with regular paying clients, and 3 years with non-paying partners. We now develop the hierarchical model to estimate the expected rate of partnership change for each type, considering the following potential biases.

Sampling. As before, [7] estimates RDS-adjusted proportions of respondents p_z (mean, 95% CI) reporting different numbers/ranges of partners $\mathbb{x}_z$ in the past 30 days (Table A.1). Thus, we take the same approach as in § 2.1 to identify distributions of reported partner numbers x_i which are consistent with the data for each partnership type.

Interpretation. Numbers of reported partners (x) have generally been interpreted in two ways — x/ω as the *rate* of partnership change (Q) or x as the *number* of current partners (K):

$$Q \approx \frac{x}{\omega} \tag{13a}$$

or

$$K \approx x \tag{13b}$$

Both interpretations are reasonable under certain conditions: If partnership duration is short and the recall period is long ($\delta \ll \omega$, e.g., 1 day vs 1 month), then reported partnerships mostly reflect *complete* partnerships, and thus $x/\omega \approx Q$. If partnership duration is long and the recall period is short ($\delta \gg \omega$, e.g., 1 year vs 1 month), then reported partnerships mostly reflect *ongoing* partnerships, and thus $x \approx K$. However, if partnership duration and recall period are similar in length ($\delta \approx \omega$, e.g., 1 month vs 1 month), then reported partnerships reflect a mixture of tail-ends, of complete, and of ongoing partnerships. Thus x/ω overestimates Q, but x also overestimates K. These three cases are illustrated in Figure 1c.

To adjust for this bias, we again assume that survey/recall period timing is effectively random. Then, if the *end* of the recall period would intersect an ongoing partnership, then a random fraction $f_i \sim \text{Unif}(0, 1)$ of the partnership duration δ would be outside the recall period. As before, the expected value $\bar{f} = \frac{1}{2}$. The same goes for the *start* of the recall period. Thus, the recall period is effectively extended by half the partnership duration $\delta/2$ on each end, and δ overall [15], as illustrated in Figure 1d. We can therefore define unbiased estimators of Q and K as:

$$Q = \frac{x}{\omega + \delta} \tag{14a}$$

$$K = \frac{x\delta}{\omega + \delta} = Q\delta \tag{14b}$$

To apply (14) in the hierarchical model, we sample the true rate of partnership change from an assumed distribution $Q_i \sim \text{Gamma}(\alpha, \beta)$, with unknown parameters α, β. Then, we model the

[1] The number of new clients per recall period could also be used to define a rate of partnership change [12], but we do not explore this approach here.

numbers of reported partners x_i given Q_i and $\omega' = (\omega + \delta)$ as: $x_i \sim \text{Poi}(Q\,\omega')$.

Summary. Figure 2b summarizes the proposed model graphically. The primary parameters of interest are α, β, which govern the distribution of rates of partnership change (for a given type) Q. We assume uninformative priors for these 2 parameters.

Comparing Assumptions. To quantify the influence of using the biased vs unbiased estimators of Q and K, we fit the proposed model for each partnership type under three assumptions: assuming *short* partnerships as in (13a) with $\omega' = \omega$; assuming *long* partnerships as in (13b) with $\omega' = \delta$; and *no* assumption on partnership duration as in (14) with $\omega' = \omega + \delta$. To illustrate more general trends in the magnitude of bias, we further compared biased vs unbiased estimates of Q and K across a range of different partnership durations $\delta \in [0.1, 10]$ and recall periods $\omega \in [0.1, 10]$, with fixed true rate $Q = 1$ (arbitrary units).

3 Results

3.1 Risk Group Duration

Figure A.1 illustrates the distributions of observed proportions p_z vs inferred proportions $\hat{p}_z$ of respondents reporting durations $d_i \in \mathbb{d}_z$ selling sex, following each stage of adjustment from § 2.1. Figure 3 illustrates the estimated cumulative distributions for years selling sex following each stage of adjustment, while Table A.2 provides the corresponding distribution means $\bar{D}$ and 95% CI. In this case, the final estimate of 4.06 (2.29, 6.34) is similar to the original median of 4, because each adjustment alternates betwen increasing and decreasing $\bar{D}$. The censoring adjustment yields the largest increase, while the measurement adjustment yields the largest decrease.

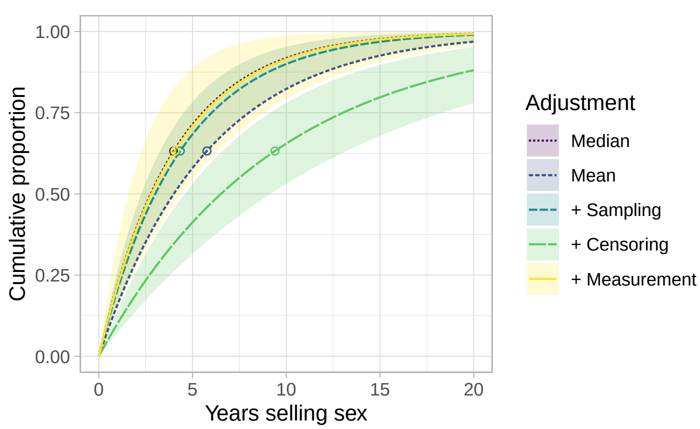


Figure 3: Estimated cumulative distribution for years selling sex following stages of adjustment

Guide: lines: cumulative distribution under posterior mean, shaded ribbon: 95% CI, circles: posterior mean.

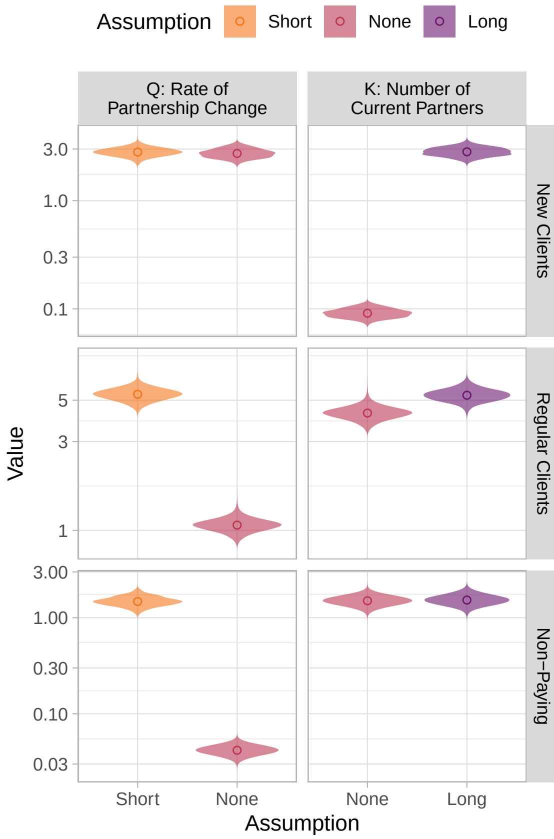


Figure 4: Estimates of rates of partnership change and numbers of current partners under different partnership duration assumptions for three partnership types reported by female sex workers

Guide: circles: posterior mean, shaded area: posterior distribution. Rates are per-month.

3.2 Rates of Partnership Change

Figure A.2 illustrates the distributions of observed proportions p_z vs inferred proportions $\hat{p}_z$ of women reporting $x_i \in \mathbb{x}_z$ partners in the past 30 days, under the three partnership duration assumptions. Figure 4 illustrates the inferred rates of partnership change (Q) and numbers of current partners (K) under each assumption, while Table A.3 provides the corresponding means and 95% CI. The biased estimates of Q and K appear equal because Q is defined as per-month. Biases are largest for Q with long partnerships (e.g., non-paying partners) and K with short partnership (e.g., new clients). However, biases are also large for both Q and K with "medium-length" partnerships (e.g., regular clients). Figure 5 illustrates generalized trends in these biases.

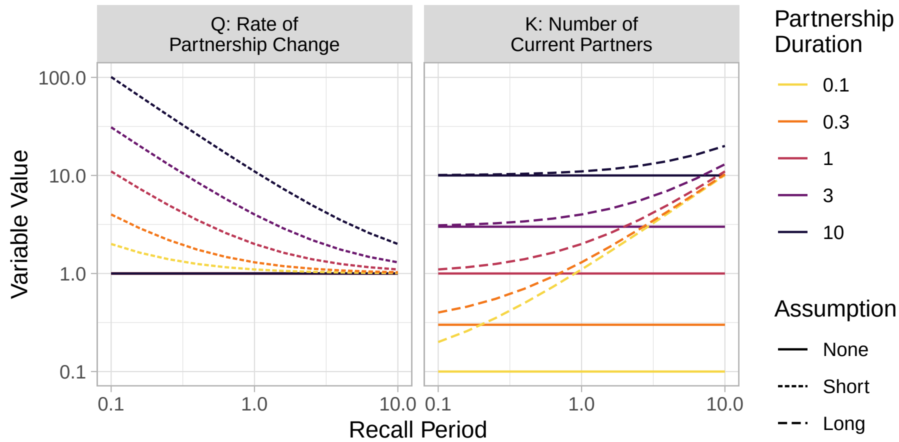


Figure 5: Estimates of rates of partnership change and numbers of current partners under different partnership duration assumptions for different recall periods and partnership durations

Units are arbitrary.

4 Discussion

We sought to develop bias adjustments for estimating the mean duration in epidemiological risk states (or periods of risk) and rates of sexual partnership change from aggregate cross-sectional data. We developed these adjustments using Bayesian hierarchical models to incorporate uncertainty in the available data and mechanistic assumptions about several "hidden" bias-generating processes. We showed that these adjustments can influence estimated variable means by factors approaching 2, suggesting that unadjusted estimates of these variables should be interpreted carefully.

We grounded our study in the analysis of aggregate sex work data to parameterize a mathematical model of HIV transmission. However, our approach should be broadly applicable to analysis of other intermittent risk exposures and event rates, including analysis of individual-level data for conventional statistical models. For example, periods of hazardous conditions may need to be quantified in an empiric study of workplace injury risk. Additionally, estimates of population-attributable fractions may be improved through our insight that: in some cross-sectional studies, reported exposure duration reflects only half of the total expected exposure duration.

Our work can also be built upon by considering further potential sources of bias and/or uncertainty. For example, we assumed a fixed duration for each sexual partnership type, but this duration could be modelled as another random variable whose distribution could also be inferred. Moreover, future work could consider rounding error [16], recall bias [17], reporting bias [18], and the like [19].

References

[1] G. P. Garnett. "An introduction to mathematical models in sexually transmitted disease epidemiology". In: *Sexually Transmitted Infections* 78.1 (Feb. 2002), pp. 7–12. https://doi.org/10.1136/sti.78.1.7.

[2] G. P. Garnett and R. M. Anderson. "Sexually transmitted diseases and sexual behavior: Insights from mathematical models". In: *Journal of Infectious Diseases* 174.S2 (Oct. 1996), S150–S161. https://doi.org/10.1093/infdis/174.Supplement˙2.S150.

[3] Hein Stigum, Per Magnus, and Leiv S. Bakketeig. "Effect of changing partnership formation rates on the spread of sexually transmitted diseases and human immunodeficiency virus". In: *American Journal of Epidemiology* 145.7 (Apr. 1997), pp. 644–652. https://10.0.4.69/oxfordjournals.aje.a009162.

[4] Charlotte Watts et al. "Remodelling core group theory: the role of sustaining populations in HIV transmission". In: *Sexually Transmitted Infections* 86.S3 (Dec. 2010), pp. iii85–iii92. https://doi.org/10.1136/sti.2010.044602.

[5] Jesse Knight et al. "Contribution of high risk groups' unmet needs may be underestimated in epidemic models without risk turnover: A mechanistic modelling analysis". In: *Infectious Disease Modelling* 5 (Jan. 2020), pp. 549–562. https://doi.org/10.1016/j.idm.2020.07.004.

[6] Marie-Claude Boily et al. "What really is a concentrated HIV epidemic and what does it mean for West and Central Africa? Insights from mathematical modeling". In: *Journal of Acquired Immune Deficiency Syndromes* 68.S2 (Mar. 2015), S74–82. https://doi.org/10.1097/QAI.0000000000000437.

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

[8] Eileen A. Yam et al. "Association between condom use and use of other contraceptive methods among female sex workers in swaziland: A relationship-level analysis of condom and contraceptive use". In: *Sexually Transmitted Diseases* 40.5 (May 2013), pp. 406–412. https://doi.org/10.1097/OLQ.0b013e318283c16d.

[9] Douglas D. Heckathorn. "Respondent-Driven Sampling: A new approach to the study of hidden populations". In: *Social Problems* 44.2 (May 1997), pp. 174–199. https://doi.org/10.2307/3096941.

[10] Stuart Geman and Donald Geman. "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-6.6 (Nov. 1984), pp. 721–741. https://doi.org/10.1109/TPAMI.1984.4767596.

[11] Roy M Anderson and Robert M May. *Infectious diseases of humans: dynamics and control*. Oxford University Press, 1991.

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

[13] Jarle Tufto. *Answer to: How to calculate average length of adherence to vegetarianism when we only have survey data about current vegetarians?* Cross Validated. Aug. 20, 2017. https://stats.stackexchange.com/a/298866/156508.

[14] 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.

[15] Michael Neely. *Answer to: Poisson-like distribution for partially or fully observed events with duration*. Mathematics Stack Exchange. July 7, 2023. https://math.stackexchange.com/a/4732425/375256.

[16] Harriet L. Mills et al. “Errors in reported degrees and respondent driven sampling: Implications for bias”. In: *Drug and Alcohol Dependence* 142 (Sept. 2014), pp. 120–126. https://www.sciencedirect.com/science/article/pii/S0376871614009375.

[17] Gita Ramjee, Amy E Weber, and Neetha S Morar. “Recording sexual behavior: Comparison of recall questionnaires with a coital diary”. In: *Sexually Transmitted Diseases* 26.7 (Aug. 1999), pp. 374–380. https://doi.org/10.1097/00007435-199908000-00002.

[18] Catherine M. Lowndes et al. “Polling booth surveys: A novel approach for reducing social desirability bias in HIV-related behavioural surveys in resource-poor settings”. In: *AIDS and Behavior* 16.4 (May 2012), pp. 1054–1062. https://doi.org/10.1007/s10461-011-0004-1.

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

Appendix

Title Adjusting for hidden biases in sexual behaviour data: a mechanistic approach

Authors Jesse Knight[1,2], Siyi Wang[1], and Sharmistha Mishra[1,2]

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

Date August 16, 2023

A Supplement

A.1 Source Data

Table A.1 gives the RDS-adjusted data from [1].

Table A.1: RDS-adjusted proportions for variables of interest

Variable	Stratum	Mean	(95% CI)
Years selling sex	0–2	38.3	(27.5, 49.1)
	3–5	32.1	(23.6, 40.7)
	6–10	20.2	(13.2, 27.1)
	11+	9.4	(04.4, 14.4)
New clients [a]	0–1	16.4	(09.8, 23.0)
	2	43.4	(33.3, 53.5)
	3	15.2	(09.6, 20.9)
	4	13.1	(07.0, 19.2)
	5	11.8	(06.0, 17.6)
Regular clients [a]	0–1	10.0	(01.9, 18.1)
	2	8.5	(03.2, 13.8)
	3	15.9	(09.8, 21.9)
	4	10.0	(04.5, 15.6)
	5	8.1	(03.8, 12.3)
	6	10.7	(05.8, 15.5)
	7+	36.9	(26.4, 47.3)
Non-paying partners [a]	0	12.5	(04.8, 20.1)
	1	50.8	(42.9, 58.7)
	2	23.6	(16.8, 30.3)
	3+	13.2	(07.2, 19.1)

[a] Number reported in the past 30 days. Data from [1].

A.2 Code

All analysis code is available online at: `github.com/mishra-lab/duration-bias`.
We fit the Bayesian hierarchical models using rjags: `cran.r-project.org/package=rjags`, with 1000 adaptive iterations and 100,000 sampling iterations.

A.3 Beta Approximation of the Binomial Distribution

The distributions of RDS-adjusted variables in [1] were reported as adjusted proportions (mean, 95% CI) for different stratifications of the variable value; e.g., 16.4 (9.8, 23.0) % of respondents reported 0–1 new clients in the past 30 days. For each proportion, we defined a beta approximation of the binomial (BAB) distribution:

$$
\begin{aligned}
\mathrm{P}(\rho) &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\,\rho^{\alpha-1}(1-\rho)^{\beta-1} \\
&\approx \binom{N}{n}\rho^{n}(1-\rho)^{N-n}
\end{aligned}
\tag{A.1}
$$

with $\alpha = N\rho$ and $\beta = N(1-\rho)$. We fixed ρ as the adjusted point estimate, and estimated N by minimizing the sum of squared differences between the 95% quantiles of (A.1) given N and the reported 95% CI for the adjusted proportion.

A.4 Risk Group Duration

Fitting to RDS-Adjusted Proportions. Figure A.1 illustrates the observed vs inferred proportions of respondents reporting different durations selling sex, following each stage of adjustment from § 2.1.

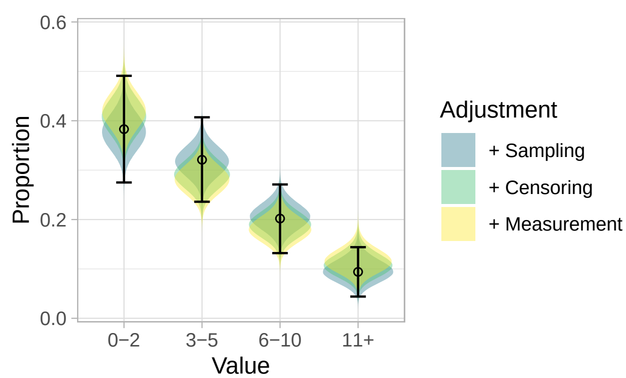


Figure A.1: Proportions of respondents reporting different durations selling sex: observed (points and ranges) vs inferred posterior (coloured regions) after 3 stages of adjustment

Numeric Summary. Table A.2 summarizes the estimated exponential distribution means (95% CI) for years selling sex following each stage of adjustment from § 2.1.

Table A.2: Estimated mean durations selling sex (years) following each stage of adjustment

Adjustment	Mean	(95% CI)
Median	4.00	—
Mean	5.77	—
+ Sampling	4.35	(3.27, 5.72)
+ Censoring	9.40	(6.60, 13.22)
+ Measurement	4.06	(2.29, 6.34)

A.5 Rate of Partnership Change

Fitting to RDS-Adjusted Proportions. Figure A.2 illustrates the observed vs inferred proportions of respondents reporting different numbers of partners in the past 30 days, under each partnership duration assumption from § 2.2.

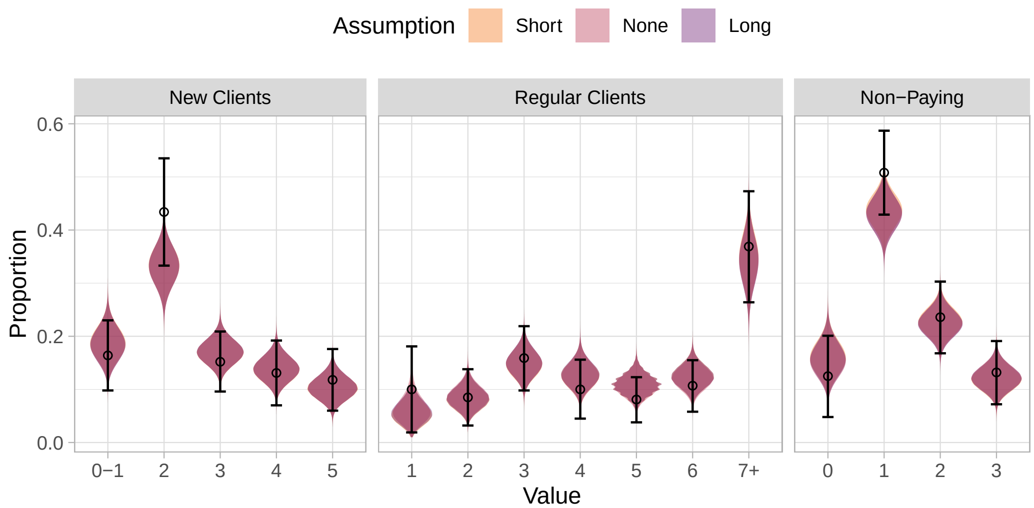


Figure A.2: Proportions of respondents reporting different numbers of partner in the past 30 days: observed (points and ranges) vs inferred posterior (coloured regions) under different partnership duration assumptions

Numeric Summary. Table A.3 summarizes the means (95% CI) for rates of partnership change and numbers of current partners estimated under each partnership duration assumption from § 2.2.

Table A.3: Biased vs unbiased estimates of rates of partnership change and numbers of current partners for three partnership types

		Rate Q [a]		Number K	
Partnership Type	Bias [b]	Mean	(95% CI)	Mean	(95% CI)
New Clients	Biased	2.82	(2.33, 3.35)	2.84	(2.36, 3.37)
	Unbiased	2.75	(2.29, 3.31)	0.09	(0.08, 0.11)
Regular Clients	Biased	5.38	(4.60, 6.20)	5.33	(4.57, 6.19)
	Unbiased	1.07	(0.90, 1.25)	4.28	(3.62, 5.02)
Non-Paying	Biased	1.49	(1.17, 1.86)	1.54	(1.20, 1.95)
	Unbiased	0.04	(0.03, 0.05)	1.51	(1.18, 1.88)

[a] Rates are per-month; [b] biased Q assume short partnerships as in (13a); biased K assume long partnerships as in (13b).

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. 2014), e115465. http://doi.org/10.1371/journal.pone.0115465.