2024 March 13

Dr. William C. Miller
Editor-in-Chief
Sexually Transmitted Diseases

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 *Sexually Transmitted Diseases*

Dear Dr. Miller,

We are pleased to submit the attached manuscript entitled *Duration considerations for cross-sectional estimates of sexual behaviour: developing bias adjustments using Bayesian hierarchical models* for consideration as an *Original Article* in *Sexually Transmitted Diseases*.

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 measurement error, 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 parameters: 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 parameters explored. Since we use only aggregate data from this previously published study, there is no relevant guideline applicable to our study.

While we focus on these two key parameters, we expect that our approach and findings would be relevant to other estimates of sexual behaviour from cross-sectional surveys, and thus of interest to the broad readership of *Sexually Transmitted Diseases*.

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

Sincerly,

Jesse Knight, Siyi Wang, and Sharmistha Mishra

Title Duration considerations for cross-sectional estimates of sexual behaviour: developing bias adjustments using Bayesian hierarchical models

Authors Jesse Knight[1,2,*], Siyi Wang[1], and Sharmistha Mishra[1–5]

[1]MAP Centre for Urban Health Solutions, Unity Health Toronto
[2]Institute of Medical Science, University of Toronto
[3]Division of Infectious Diseases, Department of Medicine, University of Toronto
[4]Institute of Health Policy, Management and Evaluation & Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto
[5]Institute for Clinical Evaluative Sciences, Toronto, Ontario
[*]jesse.x.knight@protonmail.com

Date 2024 March 13

Journal Sexually Transmitted Diseases

Status submitted

Counts Words: 34 Summary, 248 Abstract, 2226 Body; 20 References, 4 Figures, 0 Tables

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 We used only published aggregate data. All code and selected results are available at: github.com/mishra-lab/duration-biases-sex-data

Ethics Approval Not applicable.

Competing Interests None declared.

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.

Summary Mean duration selling sex and partnership change rates are key parameters for mathematical models of STD transmission. We develop Bayesian hierarchical models to address several duration-related biases affecting parameter estimates from aggregate cross-sectional data.

Abstract **Background.** Two required parameters for mathematical models of sexually transmitted infections are the mean duration in epidemiological risk states (e.g., selling sex) and the mean rates of sexual partnership change. While much attention has been paid to sampling and reporting biases affecting these parameters, relatively little work has examined measurement error in quantifying dynamic sexual behaviour from cross-sectional data. **Methods.** We explore adjustments for several biases affecting aggregate estimates of duration in sex work and numbers of reported sexual partners from a published 2011 survey 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. **Results.** 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/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 useful when the survey recall period and partnership duration are similar in length. **Conclusions.** While we explore these bias adjustments using one 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, sexual behavior, sex work, sexually transmitted diseases

1 Introduction

Mathematical models of sexually transmitted infections require estimates of sexual behaviour for model inputs (parameters) [1]. In risk-stratified models, two important parameters are: the duration of time within epidemiological risk states/groups, and rates of sexual partnership change for each group [1–4]. For example, the mean duration of time selling sex can be used to define the modelled rate of "turnover" among sex workers [4], while, numbers of main, casual, and/or paying sexual partners per year can be used to define the modelled "force of infection" [3].

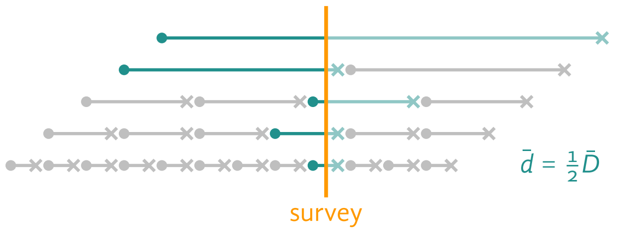
Ideally, these parameters could be informed by individual-level data from cohort studies. However, in many cases, only aggregate estimates from published cross-sectional studies are available. While much attention has been paid to sampling and reporting biases affecting sexual behaviour data [5, 6], relatively less work has examined measurement error in quantifying dynamic sexual behaviour from cross-sectional data [7, 8]. Our aim is therefore to described and address several types of duration-related biases when estimating: (1) mean duration selling sex, and (2) mean rates of partnership change, from aggregate cross-sectional estimates. We develop Bayesian hierarchical models to integrate multiple potential mechanisms of bias, and support inference of the unbiased parameters of interest. We use data from a 2011 female sex worker survey in Eswatini [9] to support parameterization of an HIV transmission model.

2 Methods

Our starting data are aggregate estimates of duration selling sex and numbers of sexual partners from [10]. In this study, 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) [11]. For each parameter — mean duration selling sex and mean partnership change rates — we describe different potential mechanisms of bias, and integrate these mechanisms stepwise into Bayesian hierarchical models. We then attempt to infer unbiased estimates of these parameters via Gibbs sampling [12]. Model implementation details are given in Appendix A.2. Figure 1 illustrates some helpful diagrams, while Figure 2 summarizes the final models.

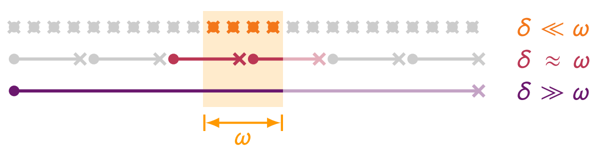
2.1 Duration Selling Sex

Data & Crude Estimate. The survey in [10] did not ask respondents about duration selling sex directly, only about current age and age of first selling sex. The difference between these ages could be used to estimate duration selling sex. Using this approach, the raw median duration was $\hat{d} = 4$ years. However, if durations are exponentially distributed — an implicit assumption in compartmental models [13] — then the mean $\bar{d}$ is related to the the median by $\bar{d} = \hat{d}/\log(2)$ due to skewness.

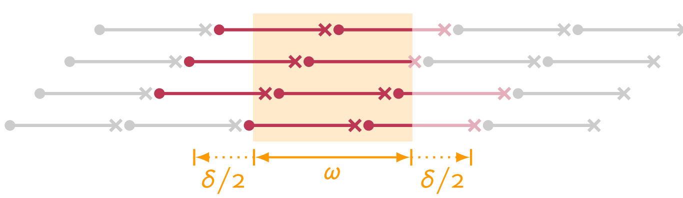


(a) Right censoring of reported durations selling sex in a steady state population

(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

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.

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

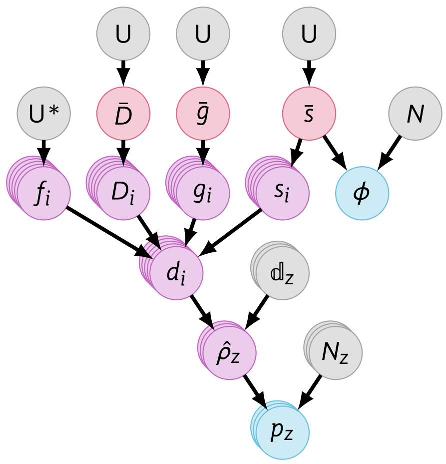
Sampling. The crude estimate above does not consider sampling bias. However, sampling bias was considered via RDS-adjustment in [10], yielding mean and 95% CI estimates of the proportions of respondents p_z who had sold sex starting $\mathbb{d}_z \in \{0\text{–}2, 3\text{–}5, 6\text{–}10, 11+\}$ years ago (Table A.1, "z" enumerates strata). Discarding the crude estimate above, we restart our estimation by defining a model to identify distributions of reported durations selling sex d_i which are consistent with these RDS-adjusted 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 — i.e., 100 simulated respondents.

Censoring. These reported durations d_i are effectively right censored because they only capture engagement in sex work up until the survey, and not additional sex work after the survey (Figure 1a) [8]. If we assume that the survey reaches respondents at a random time point during their total (eventual) duration selling sex D_i — i.e., censoring is uniform — 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 $\bar{d}/\bar{D} = \bar{f} = \frac{1}{2}$ [14].

Interruption. 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 [10] were not asked whether they ever temporarily stopped selling sex, but a later survey [15] 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 the rate $\bar{s}$ given ϕ.

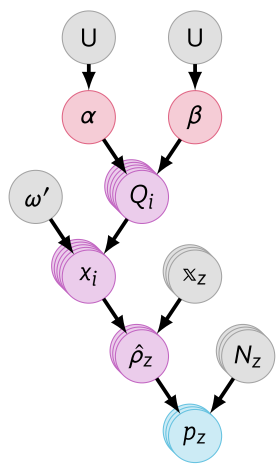
Next, we update the model for reported durations as $d_i = D_i\,(f_i + s_i\,(1 + g_i))$, where g_i is the relative duration of gaps between selling sex, with the following rationale. If $s_i = 0$, then $d_i = f_i D_i$ as before, reflecting the censored current period only. If $s_i > 0$, then d_i also includes s_i prior periods selling sex and the gaps between them (Figure 1b) — i.e., $s_i\,(D_i + g_i D_i) = D_i\, s_i\,(1 + g_i)$. 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 $g_i \sim \text{Exp}\,(1/\bar{g})$, 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.



$$p_z \sim \text{BAB}\,(N_z, \hat{\rho}_z) \tag{1a}$$
$$\hat{\rho}_z = \text{mean}\,(d_i \in \mathbb{d}_z) \tag{1b}$$
$$d_i = D_i\,(f_i + s_i\,(1 + g_i)) \tag{1c}$$
$$D_i \sim \text{Exp}\,(1/\bar{D}) \tag{1d}$$
$$f_i \sim \text{Unif}\,(0, 1) \tag{1e}$$
$$s_i \sim \text{Pois}\,(\bar{s}) \tag{1f}$$
$$g_i \sim \text{Exp}\,(1/\bar{g}) \tag{1g}$$
$$\phi \sim \text{BAB}\,(N, 1 - e^{-\bar{s}}\,) \tag{1h}$$

(a) Duration selling sex



$$p_z \sim \text{BAB}\,(N_z, \hat{\rho}_z) \tag{2a}$$
$$\hat{\rho}_z = \text{mean}\,(x_i \in \mathbb{x}_z) \tag{2b}$$
$$x_i \sim \text{Pois}\,(Q_i\,\omega') \tag{2c}$$
$$Q_i \sim \text{Gamma}\,(\alpha, \beta) \tag{2d}$$

(b) Rates of partnership change

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.

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

2.2 Rates of Partnership Change

Data & Assumptions. The survey [10] also asked respondents to report their numbers of sexual partners in a recall period of 30 days. Numbers were stratified by three types of partner: new paying clients, regular paying clients, and non-paying partners. However, the survey did not ask about partnership durations. We assume that only a small proportion of new clients go on to become regular clients; thus, we conceptualize "new" clients as effectively "one-off" clients.[1] 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.

Crude Estimates. Numbers of reported partners (x) in a given recall period (ω) 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):

$$Q \approx \frac{x}{\omega} \tag{3a}$$

or

$$K \approx x \tag{3b}$$

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 move beyond these crude estimates of Q and K, we develop the another hierarchical model as follows.

Sampling. As before, [10] 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 these proportion data for each partnership type.

Censoring. To account for reporting of tail-ends, complete, and ongoing partnerships within the recall period, we again assume that survey/recall period timing is effectively random. Then, if the start 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 end of the recall period. Thus, the recall period is effectively extended by half the partnership duration $\delta/2$ on each end, and δ overall [16], as illustrated in Figure 1d. We can

[1] The number of new clients per recall period could also be used to define a rate of partnership change [8], but we do not explore this approach here.

therefore define unbiased estimators of Q and K as:

$$Q = \frac{x}{\omega + \delta} \tag{4a}$$

$$K = \frac{x\delta}{\omega + \delta} = Q\delta \tag{4b}$$

To apply (4) 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 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 Approaches. To quantify the influence of using the crude vs adjusted estimators of Q and K, we fit the proposed model for each partnership type under three approaches: crude assuming *short* partnerships as in (3a) with $\omega' = \omega$; crude assuming *long* partnerships as in (3b) with $\omega' = \delta$; and our *adjusted* approach as in (4) with $\omega' = \omega + \delta$. To illustrate more general trends in the magnitude of potential biases, we further compared crude vs adjusted 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).

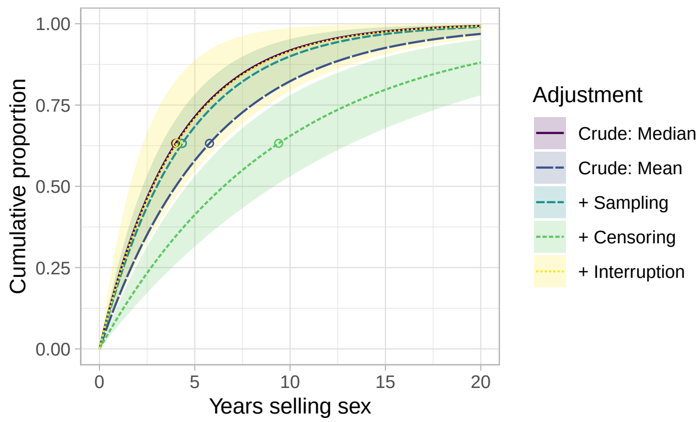
3 Results

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; 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 interruption adjustment yields the largest decrease.

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



Guide: lines: cumulative distribution under posterior mean, shaded ribbon: 95% CI, circles: posterior mean; "+": adding to existing adjustments.

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

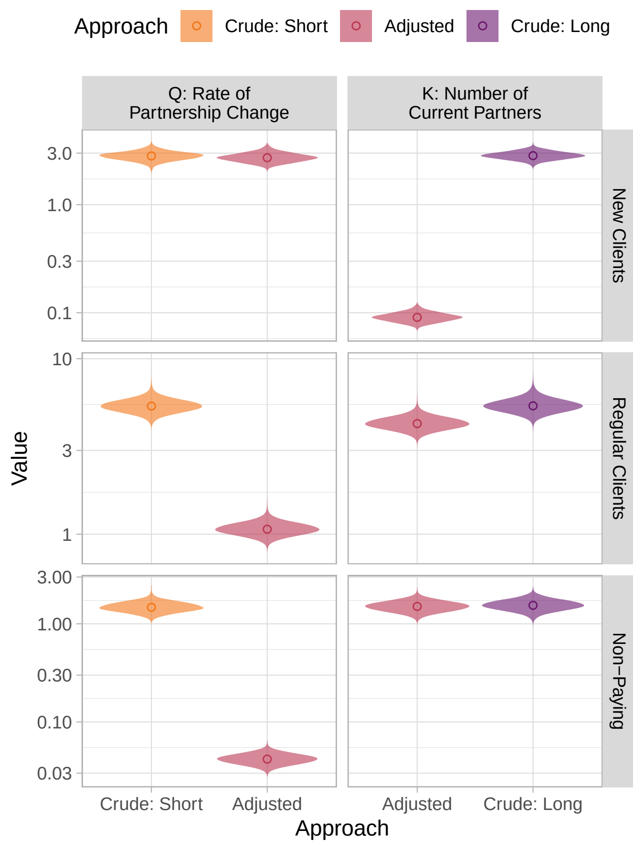
with long partnerships (e.g., non-paying partners) and K with short partnership (e.g., new clients). However, biases are significant for both Q and K with "medium-length" partnerships (e.g., regular clients), because the assumed partnership duration of $\delta = 4$ months is similar to the recall period of $\omega = 1$ month. Figure A.3 illustrates generalized trends in these biases.

4 Discussion

We sought to develop bias adjustments for estimating mean duration in epidemiological risk states (periods of risk) and mean 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 bias-generating processes. We showed that these adjustments can influence estimated parameter means by factors approaching 2, suggesting that unadjusted estimates of these parameters should be interpreted with care.

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 observation that: in many cross-sectional studies, reported exposure duration reflects only half of the total expected exposure duration.

Our work highlights key variables to collect in sexual behaviour surveys, such as numbers of sexual partners per recall period, sexual partnership durations, and number/durations of gaps within sexual partnerships and/or period of risk. Such survey questions could even be designed to support a pre-



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

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

specified hierarchical model, and use built-in redundancy to validate model assumptions, such as multiple recall periods.

Our work can 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 from additional data. Future work could also consider rounding error [17], recall bias [18], reporting bias [19], and the like [5]. Finally, our approach could be extended to support triangulation from multiple data sources [20], with potentially different mechanisms of bias for each source explicitly modelled.

References

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

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

[3] 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". *Journal of Acquired Immune Deficiency Syndromes* 68.S2 (2015), S74–82. https://doi.org/10.1097/QAI.0000000000000437.

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

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

[6] Lisa F. Langhaug, Lorraine Sherr, and Frances M. Cowan. "How to improve the validity of sexual behaviour reporting: Systematic review of questionnaire delivery modes in developing countries". *Tropical Medicine and International Health* 15.3 (2010), 362–381. https://doi.org/10.1111/j.1365-3156.2009.02464.x.

[7] Bart Burington et al. "Estimating duration in partnership studies: Issues, methods and examples". *Sexually Transmitted Infections* 86.2 (2010), 84–89. https://doi.org/10.1136/sti.2009.037960.

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

[9] 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". *Sexually Transmitted Diseases* 40.5 (2013), 406–412. https://doi.org/10.1097/OLQ.0b013e318283c16d.

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

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

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

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

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

[15] 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.

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

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

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

[19] 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". *AIDS and Behavior* 16.4 (2012), 1054–1062. https://doi.org/10.1007/s10461-011-0004-1.

[20] Debbie A Lawlor, Kate Tilling, and George Davey Smith. "Triangulation in aetiological epidemiology". *International Journal of Epidemiology* 45.6 (2016), 1866–1886. https://doi.org/10.1093/ije/dyw314.

Appendix

Title Duration considerations for cross-sectional estimates of sexual behaviour: developing bias adjustments using Bayesian hierarchical models

Authors Jesse Knight[1,2,*], Siyi Wang[1], and Sharmistha Mishra[1–5]

[1]MAP Centre for Urban Health Solutions, Unity Health Toronto
[2]Institute of Medical Science, University of Toronto
[3]Division of Infectious Diseases, Department of Medicine, University of Toronto
[4]Institute of Health Policy, Management and Evaluation & Division of Epidemiology, Dalla Lana School of Public Health, University of Toronto
[5]Institute for Clinical Evaluative Sciences, Toronto, Ontario
[*]jesse.x.knight@protonmail.com

Date 2024 March 13

Appendix A

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-biases-sex-data`. 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.

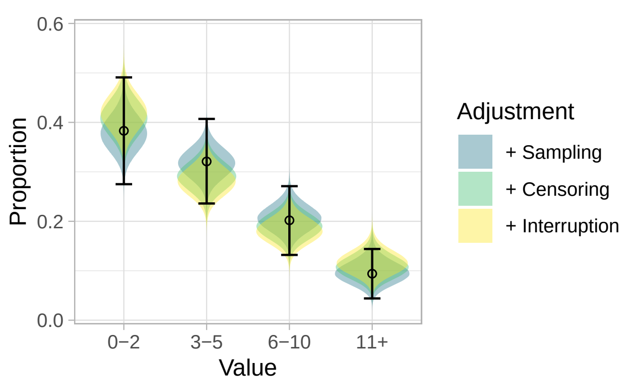


Figure A.1: Proportions of respondents reporting different durations selling sex: observed (points and ranges) vs inferred posterior (coloured regions) after 3 stages of adjustment

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)
+ Interruption	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.

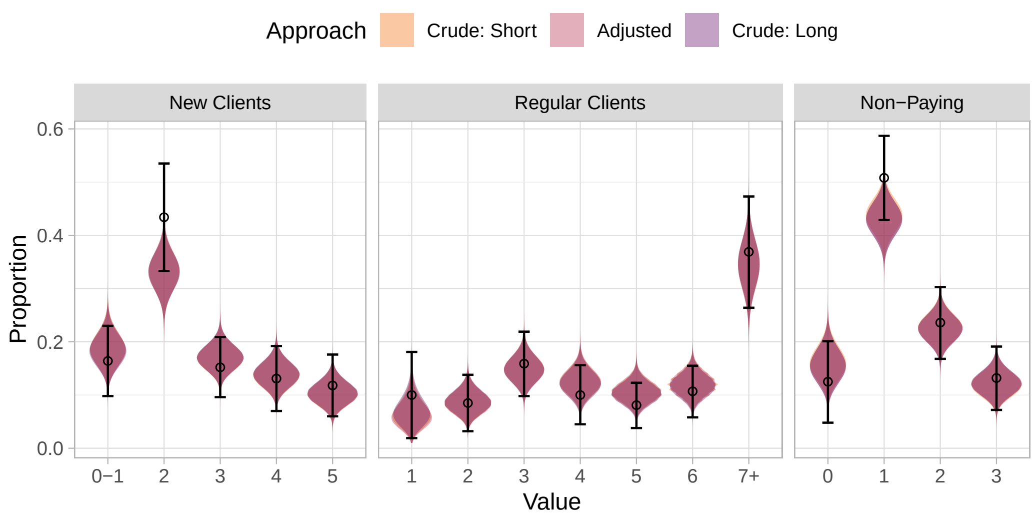


Figure A.2: Proportions of respondents reporting different numbers of partner in the past 30 days: observed (points and ranges) vs inferred posterior (coloured regions) under different partnership duration assumptions

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.

Generalized Trends. Figure A.3 illustrates generalized trends in estimated rates of partnership change and numbers of current partners under different partnership duration assumptions for different recall periods and partnership durations.

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

Partnership Type	Bias [b]	Rate Q [a] Mean	Rate Q [a] (95% CI)	Number K Mean	Number K (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 (3a); biased K assume long partnerships as in (3b).

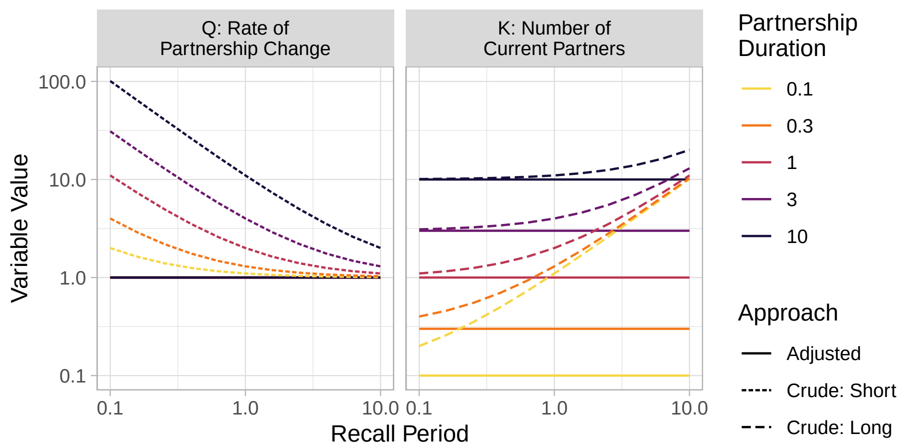


Figure A.3: Estimates of rates of partnership change and numbers of current partners under different partnership duration assumptions for different recall periods and partnership durations

Units are arbitrary.

References

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