

Estimating duration in partnership studies: issues, methods and examples

Bart Burington,¹ James P Hughes,² William L H Whittington,³ Brad Stoner,⁴ Geoff Garnett,⁵ Sevgi O Aral,⁶ King K Holmes³

¹Genentech, San Francisco, California, USA

²Department of Biostatistics, University of Washington, Seattle, Washington, USA

³Department of Medicine, University of Washington, Seattle, Washington, USA

⁴Department of Anthropology, Washington University, St Louis, Missouri, USA

⁵Department of Infectious Disease Epidemiology, Imperial College, London, UK

⁶Centers for Disease Control and Prevention, Atlanta, Georgia, USA

Correspondence to

Dr James P Hughes, Department of Biostatistics, University of Washington 357232, Seattle, WA 98115, USA; jphughes@u.washington.edu

Accepted 11 December 2009

ABSTRACT

Background and objectives Understanding the time course of sexual partnerships is important for understanding sexual behaviour, transmission risks for sexually transmitted infections (STI) and development of mathematical models of disease transmission.

Study design The authors describe issues and biases relating to censoring, truncation and sampling that arise when estimating partnership duration. Recommendations for study design and analysis methods are presented and illustrated using data from a sexual-behaviour survey that enrolled individuals from an adolescent-health clinic and two STD clinics. Survey participants were queried, for each of (up to) four partnerships in the last 3 months, about the month and year of first sex, the number of days since last sex and whether partnerships were limited to single encounters. Participants were followed every 4 months for up to 1 year.

Results After adjustment for censoring and truncation, the estimated median duration of sexual partnerships declined from 9 months (unadjusted) to 1.6 months (adjusted). Similarly, adjustment for censoring and truncation reduced the bias in relative risks for the effect of age in a Cox model. Other approaches, such as weighted estimation, also reduced bias in the estimated duration distribution.

Conclusion Methods are available for estimating partnership duration from censored and truncated samples. Ignoring censoring, truncation and other sampling issues results in biased estimates.

INTRODUCTION

Theoretical investigations of the transmission dynamics of sexually transmitted infections (STI) have demonstrated the importance of partnerships and partnership networks on disease spread. A number of studies have provided cross-sectional summaries of partnerships and/or networks, such as numbers of partners in a network,¹ mixing patterns² and concurrency,³ and these results been incorporated into STI epidemic models.^{4–6} Often, such models use only summary statistics of partnerships and treat partnerships as instantaneous entities so that their time course (ie, duration) is lost. Alternative models that simulate the formation, evolution and dissolution of partnerships suggest that partnership duration is a key determinant of partnership turnover rate, concurrency and STI risk.⁷ To date, however, there have been only a few studies that report empirical data on the time course of sexual partnerships.^{7–9} In this manuscript we review issues that complicate estimation of the distribution of the duration of sexual partnerships

from typical sexual-behaviour surveys, and we recommend approaches to the design and analysis of such surveys. To illustrate these methods we analyse data from the Seattle Sex Partner Survey,^{2 10} a sexual behaviour survey of individuals recruited through an adolescent health clinic and two STD clinics.

METHODS

Conceptually, our goal is to estimate the distribution of duration of all partnerships for all members of a defined population. In the absence of age and cohort effects, one could sample this 'population of partnerships' in a number of ways: select a random sample of individuals and obtain duration information on all or a random sample of their completed partnerships; assemble a cohort of individuals prior to sexual debut and follow them over time to obtain duration information for all partnerships during their life; or some combination of these. Additional care must be taken in defining the sample if age and/or cohort effects are expected, but estimation of the duration distribution from such data would be relatively straightforward. However, such sampling strategies are generally not practical or possible—individuals may not recall all partnerships, less recent partnerships may be subject to recall bias, follow-up periods are limited, etc. In contrast, the strategies used in typical sexual-behaviour surveys rely on non-random sampling of partnerships (eg, most recent partners, all partners in a time interval) and do not follow all partnerships to completion. Without careful analysis, these procedures can lead to biased estimates of partnership duration. In this section, we review the statistical issues and methods relevant to estimating partnership duration from typical sexual behaviour surveys using the Seattle Sex Partner Survey (SSPS) as an example.

Seattle sex partner survey

Briefly, heterosexual men and women were recruited from 1992 to 1995 into a study of sexual behaviour and STI risk ('source 1' in Aral *et al*²); 66% of those approached agreed to participate in the study. Participants were enrolled and followed every 4 months for 1 year. At enrolment and each follow-up visit, participants received a physical exam and STI testing, and a standardised questionnaire was administered. For a maximum of four heterosexual partners within the past 3 months (at entry) or since the last visit (follow-up visits), participants were asked to provide details on the month and year of first sex, the number of days since last sex and whether the partnership was limited to a single encounter. Other than this single encounter

indicator, participants were not asked whether partnerships were 'ongoing.' However, partnerships were tracked over follow-up visits (by initials or other unique identifiers provided by the participant) so that the visit date when each partnership ceased to be reported could be determined.

Statistical issues and methods

Definition of the start and end of a partnership

The concept of 'partnership duration' implies that partnerships have a well-defined beginning and end. While the start of a sexual partnership is conceptually well defined, ascertainment of the exact date based on retrospective interviews may be difficult. In the SSPS, participants were asked to specify only the month and year of the start of partnerships. Thus, for the purposes of these analyses, partnership start time was fixed at the middle of the month.

Defining the end of a partnership is more difficult, since the duration reported by the participant at the time of the interview is the 'current duration' and not necessarily the 'total duration' of the partnership. The latter is important for estimating partnership turnover rates and incorporating partnership time course into mathematical models of sexually transmitted infections. Ideally, participants should be queried about the status of their partnerships, and this information can be used to classify the partnership as ongoing or terminated (possibly with some misclassification). If such information is unavailable or deemed unreliable, then current duration can still be determined. However, it is impossible to estimate total partnership duration without further information and/or assumptions. For example, since participants in the SSPS were not asked about the status of their partnerships, we assumed that partnerships that were no longer reported at a given follow-up visit (effectively, 4 months with no reported sexual activity) had ended at the date of last sex reported at the previous visit. Partnerships that were described continuously through the end of follow-up were considered ongoing and were right-censored.

Partnership subsampling

Partnership surveys often ask participants to describe either all sexual partners within a fixed time period ('sampling window') or to describe a fixed number of most recent partners. In the SSPS, for example, participants were asked at entry to describe their most recent partners (up to four) within the past 3 months. Both subsampling schemes can lead to bias.

Fixed window subsampling tends to oversample longer partnerships, since longer partnerships are more likely to fall into the window, a problem known as length-biased sampling or left-truncation bias¹¹ (left truncation means that shorter partnerships are more likely to be excluded from the sample). This bias occurs even if all partners within the window are described. As a result, the standard product limit estimator of Kaplan and Meier¹² overestimates the median length (and all other quantiles) of sexual partnerships. As the length of the sampling window increases, the bias decreases. One solution is to consider only partnerships which began within the sampling window. However, this approach discards much of the data and limits the range of the duration distribution that can be estimated. A better approach is to use a modified Kaplan–Meier estimator that explicitly accounts for left truncation¹³ (see appendix). A similar modification can be made to Cox's proportional hazards model to evaluate factors that effect the duration distribution.

An alternative to fixed-window subsampling is to sample a fixed number of partnerships for each participant. Participants may be queried about their *k* (say) most recent partners or, less

commonly, about a random sample of *k* partnerships. Such sampling schemes may also bias duration estimates. For instance, suppose that within a hypothetical population of participants, 50% have one partner, and 50% have two partners. Thus, in this population, 33% of the partnerships are monogamous. However, if we sample only one partnership per participant, then 50% of the partnerships in the sample will be monogamous. If partnership duration is associated with the number of partners, then the duration estimates from such a sample will be biased. However, if the single partner from each participant is chosen randomly, the bias can be removed by weighting each sampled partnership according to the participant's total lifetime number of partnerships. For instance, if a participant reports five total lifetime partnerships, and information is available on two of those, then each of those two is given a weight of 5/2 in the analysis (an approach known as inverse probability weighting). If most recent (rather than random) partners are sampled, such a correction may not eliminate the bias, since the recent partners may not be representative of all partnerships (although more complex weighting schemes have recently been proposed¹⁴). Sampling weights can be incorporated into estimation of the distribution function or a Cox model using modern statistical packages.

CIs

Since each participant may contribute multiple partnerships, and the partnership durations from a single individual may be correlated, standard methods for computing CIs are suspect. To assess the variability of the estimated duration distribution, we recommend computing bootstrap-based CIs. The bootstrap procedure resamples participants and uses all partnerships associated with each resampled participant, thereby preserving correlation among the partners of the same participant.¹⁵ Similarly, robust variances¹⁶ may be used to provide valid CIs for the relative hazards in a Cox model.

Table 1 summarises the issues discussed above. Descriptive statistics, figures and algorithms were produced with the R package, version 2.4.1 (<http://www.r-project.org>).

RESULTS

SSPS participants (*n*=593) were followed for a median (IQR) of 197 days (0–370) and provided information on 1286 sexual partnerships. Participants were primarily female (55%), white (59%) and never married (84%) (table 2). Since some individuals were recruited from an adolescent health facility, a relatively high proportion of participants (26%) had not completed high school.

This dataset illustrates all of the complexities noted above: participants were not explicitly asked whether partnerships had ended, illustrating the problem of unknown censoring status, and participants described up to four most recent partners in the previous 3 months, illustrating the issues of left-truncation, partner subsampling and correlated data. We apply the methods discussed above to demonstrate the impact of these issues on estimation of the distribution of partnership duration.

Current duration versus total duration

Current duration is the length of time from the start of a partnership through the date of last sex. In contrast, total duration incorporates censoring information (ie, whether or not the partnership has ended). The lower line in figure 1 depicts the Kaplan–Meier (KM) estimate of the distribution of current partnership duration without adjustment for left truncation. The upper curve in figure 1 depicts the KM estimate of total duration, also without adjustment for left truncation. Not

Table 1 Summary of methodological issues

Scientific/study design issue	Statistical concern	Solution(s)
Current versus total duration	Censoring status	<ul style="list-style-type: none"> ▶ Current duration—ignore censoring ▶ Total duration—infer censoring from questionnaire or follow-up
Fixed window sampling	Left truncation (length-biased sampling)	<ul style="list-style-type: none"> ▶ Design study with longer sampling window ▶ Restrict analysis to partnerships initiated within the sampling window, or ▶ Use Kaplan–Meier estimate modified to account for left truncation ▶ Randomly sample partnerships within a long window
Fixed number of partners sampling	Sample of partnerships biased towards individuals with fewer partners; possible length-biased sample	<ul style="list-style-type: none"> ▶ Weight partnerships by total number of partners in lifetime ▶ Randomly sample partnerships within a long window
Multiple partnerships per individual	Correlated data	<ul style="list-style-type: none"> ▶ Use bootstrap or robust variance to compute CIs

surprisingly, the median and all other quantiles for current duration are uniformly shorter than the corresponding values for total duration (table 3), reflecting the effect of ignoring censoring.

Estimating duration with left-truncated partnerships

The estimate of total duration in figure 1 makes no adjustment for left truncation and is reproduced as the dashed line in figure 2A. The median duration from this curve is 9.0 months, but as explained previously, this figure is biased towards longer durations. Using the truncation-adjusted estimator of the duration distribution produces a median estimated duration of 1.6 months (solid line in figure 2A). Figure 2B shows that there are large differences between the duration distributions reported by participants recruited through the STD clinics versus the adolescent health clinic ($p < 0.001$). Interestingly, after stratifying by clinic, the differences between the duration distributions for

different age groups are relatively small (RR=0.99 (95% CI 0.73 to 1.32) and RR=1.1 (95% CI 0.80 to 1.54) for comparing 19–22-year-olds and >22-year-olds to those less than or equal to 18 years old, respectively), although our ability to look for such age differences is limited in the adolescent health facility. Failure to adjust for left truncation also affects the estimated relative risks in this example (RR=0.70 and RR=0.85 for the same comparisons when no adjustment is made for left truncation), although the degree and direction of RR bias are difficult to predict. When plotted on a log scale (not shown), the truncation-adjusted estimators in figure 2B are similar in shape to the second to fifth most recent partnership curves shown in figure 4 of Foxman *et al*⁹—approximately linear—suggesting an exponential distribution of partnership durations. Quantiles of the various duration distribution estimates are summarised in table 3.

Partnership sampling strategies

The effect of partnership sampling strategy (recent vs random) is illustrated in figure 3. For simplicity, only partnerships described at entry—when subsampling of partnerships is most relevant—are included. Figures 3A,B show, respectively, duration distribution estimates from most-recent and random

Table 2 Descriptive statistics for participants (N=593) from the Seattle Sex Partner Survey sample

	No	%
Gender		
Male	265	45
Female	328	55
Race*		
White	348	59
Black	176	30
Other	68	11
Age group		
less than 20	199	34
20–24	181	30
25 and older	213	36
Marital status		
Never married	502	84
Married	22	4
Divorced	52	9
Other	17	3
Education		
Less than high school or still in high school	154	26
High school	194	33
More than high school	245	41
No partners described over follow-up		
1	282	48
2	133	28
3	80	9
4	51	6
5 or more	47	9
Age	Median (25th–75th)	23 (19–28)

*One missing observation.

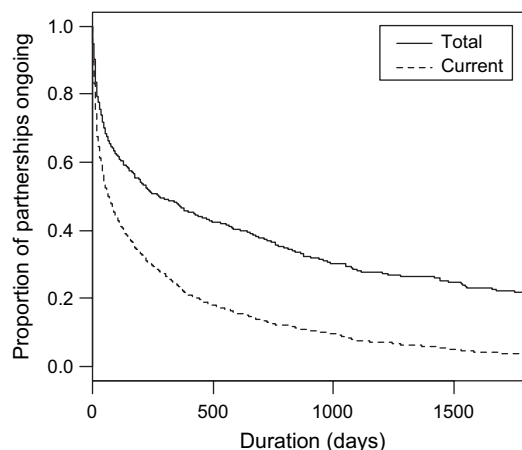


Figure 1 Current duration versus unadjusted total duration—all reported partnerships. Current duration is the time from the start of the partnership to the last reported date of sex. Total duration is similar except that partnerships are considered ongoing unless (i) they are listed as single encounters or (ii) there has been a 1 visit (~120 day) hiatus in sexual activity. Neither estimate is adjusted for left truncation.

1- and 2-partnerships subsamples compared with the full sample of partnerships (which includes up to the four most recent partnerships and may, therefore, suffer some bias as well). All the estimates are left-truncation-adjusted, and censoring information is incorporated. For both sampling schemes, the over-representation of partnerships from participants who report few total partners (within the previous 3 months) shifts the duration distribution towards longer durations (compared with the distribution from all partnerships). The estimates based on randomly selected partnerships (figure 3B) can be corrected with weights calculated from the total number of reported partnerships for each participant (figure 3C). Correcting the bias induced by most recent partner sampling is more difficult due to (potential) dependence between duration and number of reported partners. Nonetheless, figure 3C shows that weighting the most recent partner sample by number of partnerships does substantially reduce the bias in this example. There is no guarantee, however, that such weighting will always reduce the bias attributable to most recent partner sampling, since the most recent partners may not be representative of all partnerships.

DISCUSSION

Analyses of data on the duration of sexual partnerships collected in sexual behaviour surveys are complicated by a number of issues including censoring, truncation and sampling biases, and correlated data. Although some of these issues have been recognised previously,^{9, 14} no comprehensive analysis of their impact has been presented. We review these issues and illustrate their effect in the SSPS. We also suggest solutions that are appropriate for different study designs. A recent paper by Copas *et al*¹⁴ provides alternative approaches based on weighting and focusing on the most recent partner sampling.

We distinguish between the 'current duration' and 'total duration' of a partnership. Current duration may be more relevant to risk of STI, while total duration is more relevant to investigations of disease transmission and epidemic course, and is necessary to parameterise mathematical models of disease transmission. To estimate total duration, the end of a partnership must be unambiguously defined, and ongoing partnerships must be treated as right-censored. To determine if a partnership has ended, the researcher can explicitly ask whether the partnership is ongoing; alternatively, in longitudinal surveys, partnerships that have not been active for a given period may be assumed to have ended. In the SSPS, the end of a partnership was defined as a hiatus in sexual activity of approximately 4 months, although this definition was driven as much by the study design as by any underlying behavioural construct. However, once one distinguishes between current duration and total duration, other issues related to endpoint definition (eg, interval censoring) are less critical.

Table 3 Quantiles of the duration distribution of the Seattle Sex Partner Survey data by various methods (see text)

		0.1	0.25 (lower quartile)	0.5 (median)	0.75 (upper quartile)	0.9
Current duration*		0 days †	11.5 days	2.1 months	10.9 months	31.4 months
Total duration						
Unadjusted*	All	7.0 days	33.5 days	9.0 months	49.5 months	103 months
Truncation adjusted	All (95% CI) ‡	3.0 days (1.5 to 5.0)	12.0 days (10.5 to 14.0)	1.6 months (1.4 to 2.0)	6.8 months (5.7 to 7.9)	16.8 months (14.2 to 20.9)
	Adolescent health facility clinic (95% CI) ‡	5.0 days (3.5 to 10.5)	19.0 days (13.0 to 53.5)	5.3 months (3.1 to 7.3)	19.1 months (12.3 to 24.4)	28.5 months (24.3 to 50.4)
	STD clinic (95% CI) ‡	2.0 days (1.0 to 4.5)	11.0 days (9.5 to 12.5)	1.3 months (1.1 to 1.6)	4.9 months (3.8 to 6.3)	11.9 months (10.0 to 14.9)

*No adjustment for left truncation.

†Single encounter.

‡CIs are based on the 2.5th and 97.5th percentiles the bootstrap distribution based on 2000 bootstrap samples.

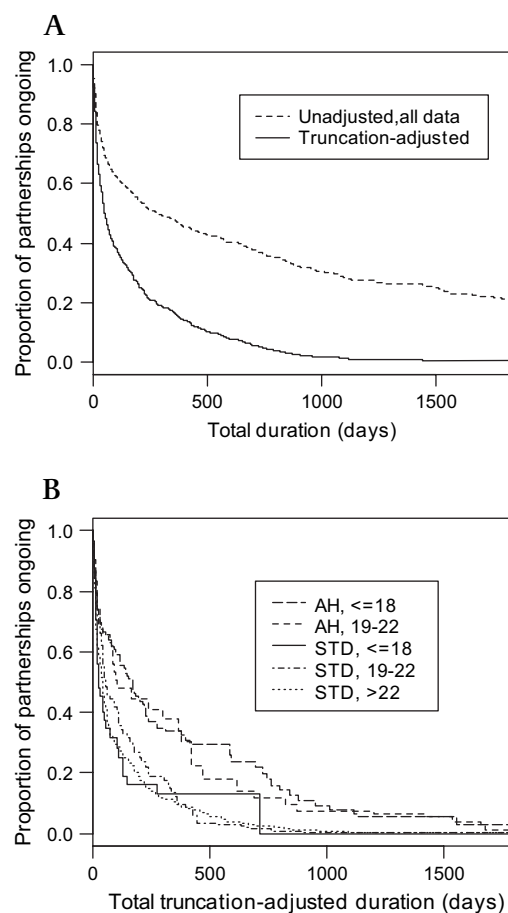


Figure 2 Truncation-adjusted total partnership duration—all reported partnerships. (A) Estimates of the Seattle Sex Partner Survey total duration distribution corresponding to an unadjusted Kaplan–Meier (KM) estimator for all partnerships (median duration=9.0 months) and a truncation-adjusted estimator for all partnerships (median=1.6 months). (B) Truncation-adjusted estimator of total partnership duration separated by clinic type and age group. AH, adolescent health facility; STD, STD clinic.

In addition to right censoring, the design of sexual-behaviour surveys often leads to left truncation and length-biased sampling. For instance, fixed window sampling preferentially includes longer partnerships and can lead to significant biases in the estimated duration distribution. This bias can be removed with appropriate statistical methods that incorporate information on the window length. Sampling schemes that select a fixed number of most recent partnerships can also bias estimates of partnership duration. In the SSPS data, the bias induced by most recent partner sampling appears to be less than that seen with fixed window sampling (and declines as more partners are sampled). However, bias due to most

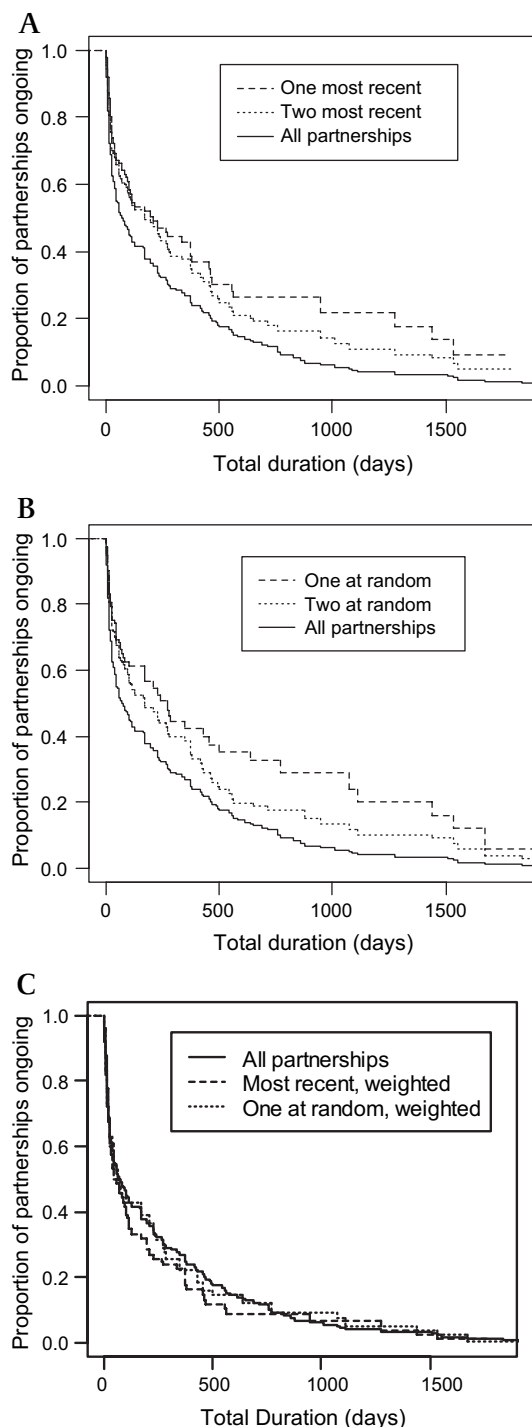


Figure 3 Partner sampling schemes. All panels include only participants described at entry, when partner sampling is most pronounced. (A) Duration estimates using only the 1 and 2 most-recent partnerships compared with the duration based on all partnerships. (B) Duration estimates when at most 1 and 2 partnerships are randomly sampled for each participant. (C) Effect of weighting by number of reported partners on the duration estimates based on one partnership selected at random or the most recent partnership.

recent partner sampling is less well defined, and statistical methods to adjust for it depend on assumptions about the representativeness of the sampled partners.

An alternative to adjusting for left truncation and sampling biases in the analysis is to design the survey to reduce or eliminate the biases by limiting the extent to which partnership

inclusion criteria are associated with the outcome of interest (ie, partnership duration). One possible strategy is to first query participants for a small number of key statistics about all of their partnerships within a relatively long window prior to the interview. Then, the interviewer could randomly sample a subset of these partnerships from which more extensive information is to be gathered. Any estimates based on the random subsample should use weights to correct the estimates for the total number of partners reported by each participant. Such a strategy may be subject to recall bias, however, and is likely to be most effective in individuals who have had a small to moderate number of partners (ie, so all can be listed and recalled). The trade-off between recall bias and length bias is a potential area for future research.

A key assumption of the statistical methods presented here is that the censoring and truncation processes are independent of duration. The independence assumption for censoring is commonly made in time-to-event analyses and may be reasonable in many settings. Independent truncation means that the sampling window does not depend on the time to the event of interest (eg, end of partnership). The assumption of independent truncation may be violated if there is a significant secular trend in duration (ie, partnerships are systematically getting longer or shorter over time). The sensitivity of duration distribution estimates to such violations can be explored through simulations tailored to the particular setting, but ultimately, such assumptions are not testable, and conclusions based on these methods should be appropriately circumspect.

The methods described here have been applied to a longitudinal sexual-behaviour survey. However, all of the methods described, with the exception of our approach to distinguishing completed from ongoing partnerships (which depends on the availability of follow-up data), are equally relevant and applicable in cross-sectional surveys.

We have not discussed in detail more common sampling issues, but these are important as well. For instance, the data analysed here are collected from participants coming to STI clinics and an adolescent health clinic. The duration distribution estimated from these data may be representative of similar clinic populations but is unlikely to be representative of the broader population outside such clinics¹⁷ For instance, the quantiles of the duration distribution summarised in table 3 from this clinic population are uniformly shorter than those seen in Foxman *et al*'s telephone survey of the Seattle general population.⁹

In summary, estimates of partnership duration based on surveys of sexual behaviour are subject to multiple sources of bias and variability. However, through careful design and analysis, the biases can be substantially reduced and the variability appropriately quantified.

Key messages

- ▶ Estimates of partnership duration depend on a careful definition of the end of a partnership to distinguish completed from ongoing (right-censored) partnerships.
- ▶ Typical sexual behaviour surveys produce biased estimates of duration; such biases can be reduced through a combination of design and analysis.
- ▶ CIs must adjust for correlated data when information on multiple partnerships is gathered from each participant.

Contributors BB and JPH developed the statistical methods and analysed the data. WLW, BS, SA, GG and KKH contributed to the design and WLW and BS supervised the data collection of the Seattle Sex Partner Survey. All authors contributed to the interpretation of results and write-up.

Funding The work was supported by National Institutes of Health grants AI31448 and AI29168.

Competing interests None.

Ethics approval Ethics approval was provided by the University of Washington.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Stoner BP**, Whittington WL, Hughes JP, *et al*. Comparative epidemiology of heterosexual gonococcal and chlamydial networks: Implications for transmission patterns. *Sex Transm Dis* 2000;**27**:215–23.
2. **Aral SO**, Hughes JP, Stoner BP, *et al*. Demographic and behavioral concordance between sex partners: Effects of demographic and partnership characteristics, sexual behavior and infection status. *Am J Public Health* 1999;**89**:825–33.
3. **Morris M**, Kretzschmar M. Concurrent partnerships and the spread of HIV. *AIDS* 1997;**11**:642–8.
4. **Garnett GP**, Anderson RM. Contact tracing and the estimation of sexual mixing patterns: the epidemiology of gonococcal infections. *Sex Transm Dis* 1993;**20**:181–91.
5. **Oxman GL**, Smolkowski K, Noell J. Mathematical modeling of epidemic syphilis transmission. *Sex Transm Dis* 1996;**23**:30–9.
6. **Renton AM**, Whitaker L, Riddlesdell M. Heterosexual HIV transmission and STD prevalence: predictions of a theoretical model. *Sex Transm Infect* 1998;**74**:339–44.
7. **Chen MI**, Ghani AC, Edmunds J. Mind the gap: The role of time between sex with two consecutive partners on the transmission dynamics of gonorrhoea. *Sex Transm Dis* 2008;**35**:435–44.
8. **Kraut-Becher JR**, Aral SO. Gap length: An important factor in sexually transmitted disease transmission. *Sex Transm Dis* 2003;**30**:221–5.
9. **Foxman B**, Newman M, Percha B, *et al*. Measures of sexual partnerships: Lengths, gaps, overlaps and sexually transmitted infections. *Sex Transm Dis* 2006;**33**:209–14.
10. **Aral SO**, Hughes J, Gorbach P, *et al*. The Seattle “Sexual Mixing,” “Sexual Networks,” and “Sexual Partnership Types” Studies. In: Morris M, ed. *Network epidemiology: a handbook for survey design and data collection* Oxford: Oxford University Press, 2004:139–73.
11. **Simon R**. Length-biased sampling in etiological studies. *Am J Epidemiol* 1980;**111**:444–52.
12. **Kaplan EL**, Meier P. Nonparametric estimator from incomplete observations. *J Am Stat Assoc* 1958;**53**:457–81.
13. **Tsai W**, Jewell NP, Wang M. A note on the product-limit estimator under right censoring and left truncation. *Biometrika* 1987;**74**:883–6.
14. **Copas AJ**, Mercer CH, Farewell VT, *et al*. Recent heterosexual partnerships and patterns of condom use. *Epidemiology* 2009;**20**:44–51.
15. **Davison AC**, Hinkley DV. *Bootstrap methods and their application* Cambridge: Cambridge University Press, 1997.
16. **Therneau TM**, Grambsch PM. *Modeling survival data: extending the Cox model* New York: Springer, 2000.
17. **Manhart LE**, Aral SO, Holmes KK, *et al*. Influence of study population on the identification of risk factors for sexually transmitted diseases using a case-control design. *Am J Epidemiol* 2004;**160**:393–402.

APPENDIX

The usual product limit estimator of the survival curve of the duration distribution may be written as

$$S(x) = \prod_{y_i \leq x} \left(1 - \frac{d(y_i)}{R(y_i)}\right)$$

where the product is over all unique event times (durations, y_i) less than or equal to x , $d(y_i)$ is the number of uncensored events (durations) exactly equal to y_i and $R(y_i) = N - \sum_j I(y_j < y_i)$ where N is the sample size and $I(\cdot)$ is the indicator

function (ie, $I(z) = 1$ if z is true and 0 otherwise). For truncated data, we use the same estimator but define $R(y_i)$ as $R(y_i) = \sum_j I(t_j < y_i) - \sum_j I(y_j < y_i)$ where t_j is the truncation time for the j th observation. In the present context, the truncation time is the time between the start of the partnership and the start of the sampling window ($t_j = 0$ for partnerships that began after the start of the sampling window).

Statistical programs that allow for a ‘counting process’¹⁶ style of input for survival data (eg, SAS, Stata, R) can also be used to estimate the duration distribution for left-truncated data by entering the truncation time (t_j in the above) as ‘time 1’ and the partnership duration (y_j in the above) as ‘time 2.’ Each of these packages supports this data format for both product limit estimates and Cox proportional hazards modelling. In addition, robust variances for the relative hazards in a proportional hazards model are supported.