Does Size Really Matter? A Sensitivity Analysis of Number of Seeds in a Respondent-Driven Sampling Study of Men who Have Sex with Men in Vancouver, BC

#898

Nathan John Lachowsky, PHD,[1,2] Justin T Sorge, RRT MPH,[1] Henry Fisher Raymond, DrPH, MPH,[3,4] Zishan Cui, MSc,[1] Paul Sereda, BA(Hons.),[1] Ashleigh Rich, MPH,[1] Eric A Roth, PhD,[5] Robert S Hogg, PhD,[1,6] David M Moore, MDCM, MHSc, FRCPC(C)[1,2]

1. British Columbia Centre for Excellence in HIV/AIDS, Vancouver, Canada.
2. Faculty of Medicine, University of British Columbia, Vancouver, Canada.
3. University of California San Francisco, San Francisco, United States.
4. San Francisco Department of Public Health, San Francisco, United States.
5. Department of Anthropology, University of Victoria, Victoria, Canada.
6. Faculty of Health Science, Simon Fraser University, Burnaby, Canada

Background

- Respondent-driven sampling (RDS) is an increasingly used peer chain-recruitment method to sample "hard-to-reach" populations[1]
- Implementation success of RDS varies; one potential negative factor being the number of "seeds" used to start recruitment chains.
- **We sought to examine the effect that differing restrictions of our sample based on seed and recruitment chain productivity had on various RDS-adjusted point estimates.**
- We hypothesized that point estimates would not be largely affected by the exclusion of unproductive seeds and short chains.

Methods

Study Protocol & Participants: Data were drawn from the baseline visit of the Momentum Health Study, a prospective bio-behavioural cohort of sexually-active MSM aged 16+ in Metro Vancouver recruited using RDS from 02/2012 - 02/2014. Participants completed a computer-assisted self-interview (CASI) on demographics, sexual behaviour, and substance use; and a nurse visit with a rapid point-of-care HIV test and specimen collection for HCV and syphilis serology.

Outcomes: Three main outcomes were selected for this analysis: 1) HIV serostatus (HIV-positive vs. HIV-negative), 2) risky sexual behaviour (any condomless anal intercourse with an HIV serodiscordant or unknown status partner in the past six months vs.not), and 3) any injection drug use in the past six months (vs.not).

Samples Used: To conduct this analysis, we used various sample cuts, starting with all participants and subsequently removing unproductive seeds (0 recruitment waves), chains of ≤ 1 recruitment waves, and chains of ≤ 2 recruitment waves.

Statistical Analysis: Using the RDS Analyst 0.52 and SAS 9.4, we calculated estimates for three different outcomes for each sample cut and three different RDS weighting procedures: RDS-I (SH), RDS-II (VH), and RDS-SS[1-4]. We also assessed seed dependence with bottleneck analyses and convergence plots.

Table 1. Recruitment chain characteristics of various samples

	Overall (n=719)	> 0 Wave (n=669)	> 1 Wave (n=594)	> 2 wave (n=537)
Seeds* n (% of sample)	119 (16.6%)	69 (10.3%)	37 (6.2%)	25 (4.7%)
Chain length: mean (median)	1.75 (1.00)	3.01 (2.00)	4.76 (3.00)	6.08 (5.00)
Chain length: range	0-16	1-16	2-16	3-16

Results

- 719 participants were recruited, which included 119 seeds and a maximum of 16 recruitment waves (mean recruitment chain length = 1.7).
- **Table 1 (lower left corner)** describes our study's wave-length characteristics
- The sample of >0 recruitment waves removed all unproductive seeds (n=50/119, 42.0%), resulting in 69 chains (mean length = 3.0).
- The sample of >1 recruitment waves removed 125 seeds or recruits (17.4% of overall sample), resulting in 37 chains (mean length = 4.8).
- The final sample of >2 recruitment waves removed a further 182 seeds or recruits (25.3% of overall sample), resulting in 25 chains (mean length = 6.1).
- **Table 2 (below)** describes the demographics of our study sample with crude and various RDS-adjusted estimates, including 95% confidence intervals (CIs)
- Our sample of 719 MSM contained 612 (85.1%) men that identified as gay. There was a reasonably even distribution of age among our sample. The majority of our sample consisted of respondents who identified their race/ethnicity as White (n=539, 75.0%). Most respondents reported earning less that $30,000/year (n=457, 63.6%).

Table 2. Momentum Health Study baseline sample demographics

	n	Crude %	RDS-I % (95% CI)	RDS-II % (95% CI)	RDS-SS % (95% CI)
Sexual Identity					
Gay	612	85.1	80.5 (75.8-85.2)	82.4 (77.8-87.0)	82.5 (77.9-87.1)
Other	107	14.9	19.5 (14.8-24.2)	17.6 (13.0-22.2)	17.6 (13.0-22.2)
Age (years)					
18-29	275	38.3	35.3 (31.5-39.2)	37.7 (31.3-44.1)	37.7 (31.3-44.1)
30-44	233	32.4	35.6 (31.2-40.0)	34.3 (27.9-40.6)	34.3 (27.9-40.7)
≥45	211	29.5	29.1 (25.4-32.8)	28.0 (22.5-33.6)	28.1 (22.5-33.6)
Race/Ethnicity					
White	539	75.0	66.7 (61.0-72.3)	67.4 (61.3-73.5)	67.5 (61.4-73.6)
Asian	72	10.0	12.3 (8.8-15.8)	12.1 (8.3-16.0)	12.1 (8.2-15.9)
Indigenous	50	7.0	10.5 (6.1-15.0)	10.0 (6.1-14.0)	10.0 (6.0-14.0)
Other	58	8.1	10.5 (7.0-14.0)	10.5 (6.2-14.7)	10.4 (6.2-14.7)
Income (CAD)					
<$30,000	457	63.6	74.0 (70.6-77.4)	72.9 (67.3-78.4)	72.8 (67.2-78.4)
$30,000-60,000	182	25.3	17.8 (15.2-20.3)	18.3 (13.4-23.1)	18.3 (13.4-23.2)
>$60,0000	80	11.1	8.2 (6.6-9.9)	8.9 (5.7-12.1)	8.9 (5.7-12.1)

Results

- **Figure 1 (below)** depicts convergence and bottleneck plots for each of the main study outcomes: HIV status, high-risk sex, injection drug use
- High-risk sex and injection drug use both converged on the population estimate, and the bottleneck plots suggest low homophily. However, HIV status converged late and the bottleneck plot suggests sample homophily.
- **Table 3 (right)** shows descriptive statistics for each of the study outcomes, for the four different sample cuts using three different RDS-adjustments
- For high risk sex and injection drug use, within each sample cut, crude, all RDS-adjusted estimates fell within each estimates' CIs. For HIV status, within each sample cut, crude estimates did not fall within the RDS-I (SH) CI, but did fall within RDS-II (VH) and RDS-SS CIs.
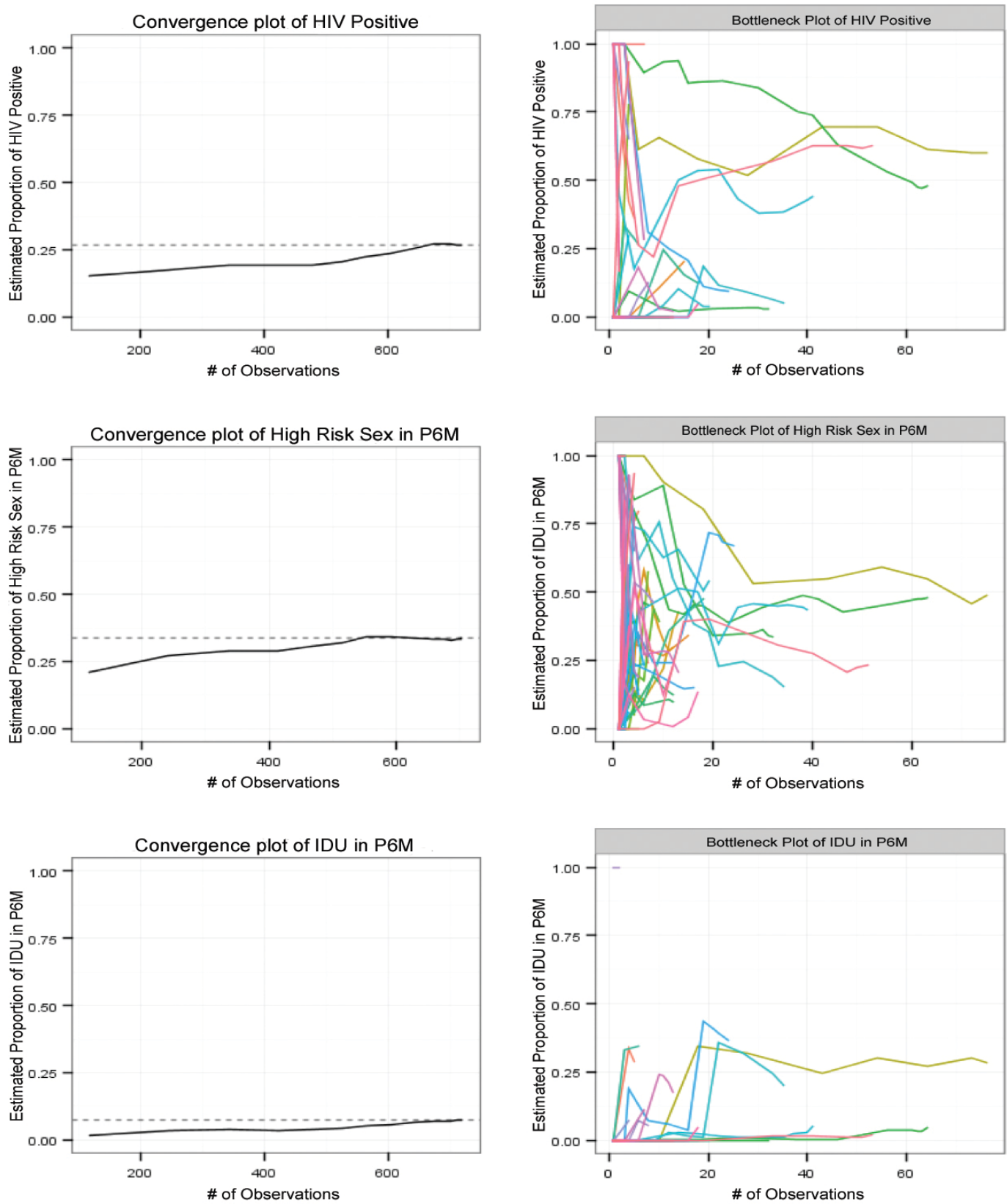

Figure 1. Convergence and bottleneck plots for three outcomes

Table 3. Three study outcomes using various sample cuts and RDS-weights

		n	Crude %	RDS I % (95% CI)	RDS II % (95% CI)	RDS SS % (95% CI)
HIV-positive serostatus (vs. HIV-negative)	Overall (n=719)	199	27.7	22.5 (19.4-25.6)	26.7 (20.7-32.7)	26.7 (20.7-32.7)
	>0 Wave (n=669)	189	28.3	22.9 (19.7-26.2)	27.7 (21.5-33.9)	27.7 (21.4-34.1)
	>1 Wave (n=594)	178	30.0	24.2 (20.6-27.7)	29.3 (22.6-36.0)	29.4 (22.6-36.1)
	>2 Wave (n=537)	161	30.0	23.8 (20.1-27.5)	29.1 (22.1-36.1)	29.1 (22.1-36.2)
Any high risk sex in past 6 months (vs. none)	Overall (n=719)	262	37.3	34.9 (31.1-38.8)	33.6 (27.6-39.6)	33.7 (27.7-39.7)
	>0 Wave (n=669)	251	38.4	35.5 (31.5-39.5)	35.2 (28.9-41.5)	35.3 (29.0-41.6)
	>1 Wave (n=594)	221	38.0	36.0 (31.7-40.3)	35.8 (29.3-42.4)	35.8 (29.2-42.4)
	>2 Wave (n=537)	203	38.7	37.0 (32.4-41.5)	36.4 (29.5-43.3)	36.5 (29.6-43.3)
Injected any drugs in past 6 months (vs. none)	Overall (n=719)	61	7.1	8.8 (5.5-12.1)	7.3 (3.9-10.7)	7.3 (3.9-10.7)
	>0 Wave (n=669)	51	7.6	9.1 (5.8-12.4)	8.1 (4.4-11.8)	8.1 (4.3-11.9)
	>1 Wave (n=594)	48	8.1	8.7 (5.7-11.7)	8.0 (4.7-11.3)	8.0 (4.7-11.3)
	>2 Wave (n=537)	43	8.0	8.7 (5.5-11.9)	8.3 (4.8-11.8)	8.3 (4.8-11.9)

Conclusions

- Nearly all RDS-weighted estimates were similar and fell within 95% CIs of each other
- Diagnostic plots were a useful method to assess for equilibrium and homophily within an RDS sample and a useful predictor of the validity of estimates
- Using diagnostic methods suggested by Gile, Johnston and Salganik,[5] for outcomes that have reached parameter stability within each sample cut, the crude proportions fell within 95% CIs of all RDS-weighted estimates
- **RDS studies, although potentially costly and time consuming (especially when many seeds are needed to reach desired sample size), are not negatively affected by large numbers of unproductive or lowly productive seeds when equilibrium has occurred**

References

(1) Heckathorn DD. Respondent-driven sampling: A new approach to the study of hidden populations. *Soc. Probl.* 1997;44(2):174-199
(2) Heckathorn DD. Respondent-Driven Sampling II: Deriving Valid Population Estimates from Chain-Referral Samples of Hidden Populations. *Soc. Probl.* 2002;49(1):11-34.
(3) Volz E, Heckathorn DD. Probability Based Estimation Theory for Respondent Driven Sampling. *J. Off. Stat.* 2008;24(1):79-97.
(4) Gile KJ. Improved Inference for Respondent-Driven Sampling Data with Application to HIV Prevalence Estimation. *Journal of Am. Stat. Assoc.* 2011;106(493):135-146.
(5) Gile K, Johnston LG, Salganik MJ. Diagnostics for respondent-driven sampling. *J. R.* 2014:241-269.

Acknowledgements

Presenter Contact Information:
Nathan J. Lachowsky
nlachowsky@cfenet.ubc.ca
P: 604-558-2017