Published in final edited form as:

Netw Sci (Camb Univ Press). 2014 ; 2(2): 298–301. doi:10.1017/nws.2014.16.

Stickiness of respondent-driven sampling recruitment chains

Jacob C. Fisher and

Department of Sociology, Duke University, Box 90088, Durham, NC 27708-0088, USA (jcf26@duke.edu)

M. Giovanna Merli

Sanford School of Public Policy & Duke Global Health Institute, Duke Population Research Institute, Duke University, Box 90312, Durham, NC 27708, USA

Department of Sociology, Duke University, Durham, NC 27708, USA

Can respondent-driven sampling (RDS) recruitment chains get stuck among socially stratified subgroups?

Respondent-driven sampling (RDS) is an increasingly popular chain-referral sampling method. Although it has proved effective at generating samples of hard to reach populations —meaning populations for which sampling frames are not available because they are hidden or socially stigmatized like sex workers or injecting drug users—quickly and cost-effectively, the ease of collecting the sample comes with a cost: bias or inefficiency in the estimates of population parameters (Gile & Handcock, 2010; Goel & Salganik, 2010). One way that RDS can produce inefficient estimates is if one or more of the recruitment chains gets stuck among members of a cohesive subpopulation, preventing the RDS sampling process from exploring other areas of the network. If that happens, members of the population subgroup recruit one another repeatedly, leading to an increase in sample size without increasing the diversity of the sample. This type of stickiness is particularly likely when hidden populations are stratified, and the stratified groups are organized into venues that provide opportunities to recruit other members of the same stratum. Female sex workers (FSW) in China, who are stratified into tiers of sex work that are correlated with marital status, age, and risk behaviors, are a prime example (Merli et al., 2014; Yamanis et al., 2013). Chinese FSW recruit clients from venues such as karaoke bars, massage parlors, or street corners. At larger venues, sex workers who participate in an RDS study might recruit other members of the same venue into the study at a higher rate than expected, leading to inefficient estimates. In short, the chain could get stuck in a venue.

To determine whether chains were getting stuck among members of stratified subpopulations, we created a recruitment chain diagram, which shows the pattern of recruitment in the RDS sample, using data from a study of 583 FSW recruited into an RDS sample in Liuzhou, China. The data used for this diagram are described in Weir et al. (2012), who used them to compare estimates of characteristics of FSW recruited by two concurrent studies: an RDS study and a venue-based study. Vertices represent FSW who participated in the survey, and edges indicate that one FSW invited another to participate in the survey. The diagram starts at the top with the initial members of the sample, or seeds, and proceeds from top to bottom. Vertices are sized according to the size of the venue where

the sex worker primarily solicited her clients. We assume that FSW in larger venues have more opportunity to recruit others from the same venue, and therefore we would expect that recruitment chains are more likely to get stuck in larger venues. Edges are colored according to whether both the recruiter and the recruit solicited clients from the same venue. Vertices are colored according to the tier of sex work. High tier venues include places like karaoke bars or dance halls, middle tier venues include places like massage parlors, and low tier venues include outdoor places like streets or parks. Non-venue based sex work means that sex workers solicit clients by telephone or text message.

Visual examination of the figure begins with consideration of the vertices. Through inspection of the vertices, two relationships become apparent. First, size of venue is related to venue tier. Almost all of the large venues employ sex workers classified as high tier; most middle and low tier venues are small, employing fewer than 20 people. Second, there is a significant amount of autocorrelation among the chains. Most of the high tier sex workers were recruited by other high tier sex workers, and two of the chains are almost entirely composed of FSW who solicit clients in large, high tier venues. Subsequent examination of the edges reveals that many of the recruitments between members of the high tier, large venue clusters of FSW are, in fact, from the same venue. In those cases—particularly in the 3rd chain and the 4th chain from the left—the final members of the sample do not appear to be independent of the characteristics of the seeds, as early work on RDS would have suggested (Heckathorn, 1997, 2002). This illustration suggests that as new estimators for RDS (e.g., Gile, 2011) begin to account for more complicated dependencies in the recruitment chains, the estimators should incorporate information about where respondents meet others. Places like venues where sex workers operate can provide opportunities for sex workers with similar characteristics to interact, and therefore can provide opportunities for RDS chains to get stuck among subpopulations, leading to inefficient samples.

Acknowledgements

This research was supported in part by an NIA training grant to the Center for Population Health and Aging at Duke University (T32 AG000139), and by NICHD grant 1R01HD068523 to Duke University (Merli, PI). The authors thank Ashton Verdery, Jim Moody, and two anonymous reviewers for their helpful comments.

References

Gile KJ. Improved inference for respondent-driven sampling data with application to HIV prevalence estimation. Journal of the American Statistical Association. 2011; 106(493):135–146.

Gile KJ, Handcock MS. Respondent-driven sampling: An assessment of current methodology. Sociological Methodology. 2010; 40(1):285–327. [PubMed: 22969167]

Goel S, Salganik MJ. Assessing respondent-driven sampling. Proceedings of the National Academy of Sciences. 2010; 107(15):6743–6747.

Heckathorn DD. Respondent-driven sampling: A new approach to the study of hidden populations. Social Problems. 1997; 44(2):174–199.

Heckathorn DD. Respondent-driven sampling II: Deriving valid population estimates from chain-referral samples of hidden populations. Social Problems. 2002; 49(1):11–34.

Merli, MG.; Moody, J.; Smith, J.; Li, J.; Weir, S.; Chen, X. Social Science and Medicine. 2014. Challenges to recruiting population representative samples of female sex workers in China using respondent driven sampling. Advance online publication.

Weir SS, Merli MG, Li J, Gandhi AD, Neely WW, Edwards JK, Chen X-S. A comparison of respondent-driven and venue-based sampling of female sex workers in Liuzhou, China. Sexually Transmitted Infections. 2012; 88(2):i95–i101. [PubMed: 23172350]

Yamanis TJ, Merli MG, Neely WW, Tian FF, Moody J, Tu X, Gao E. An empirical analysis of the impact of recruitment patterns on RDS estimates among a socially ordered population of female sex workers in China. Sociological Methods and Research. 2013; 42(3):392–425.

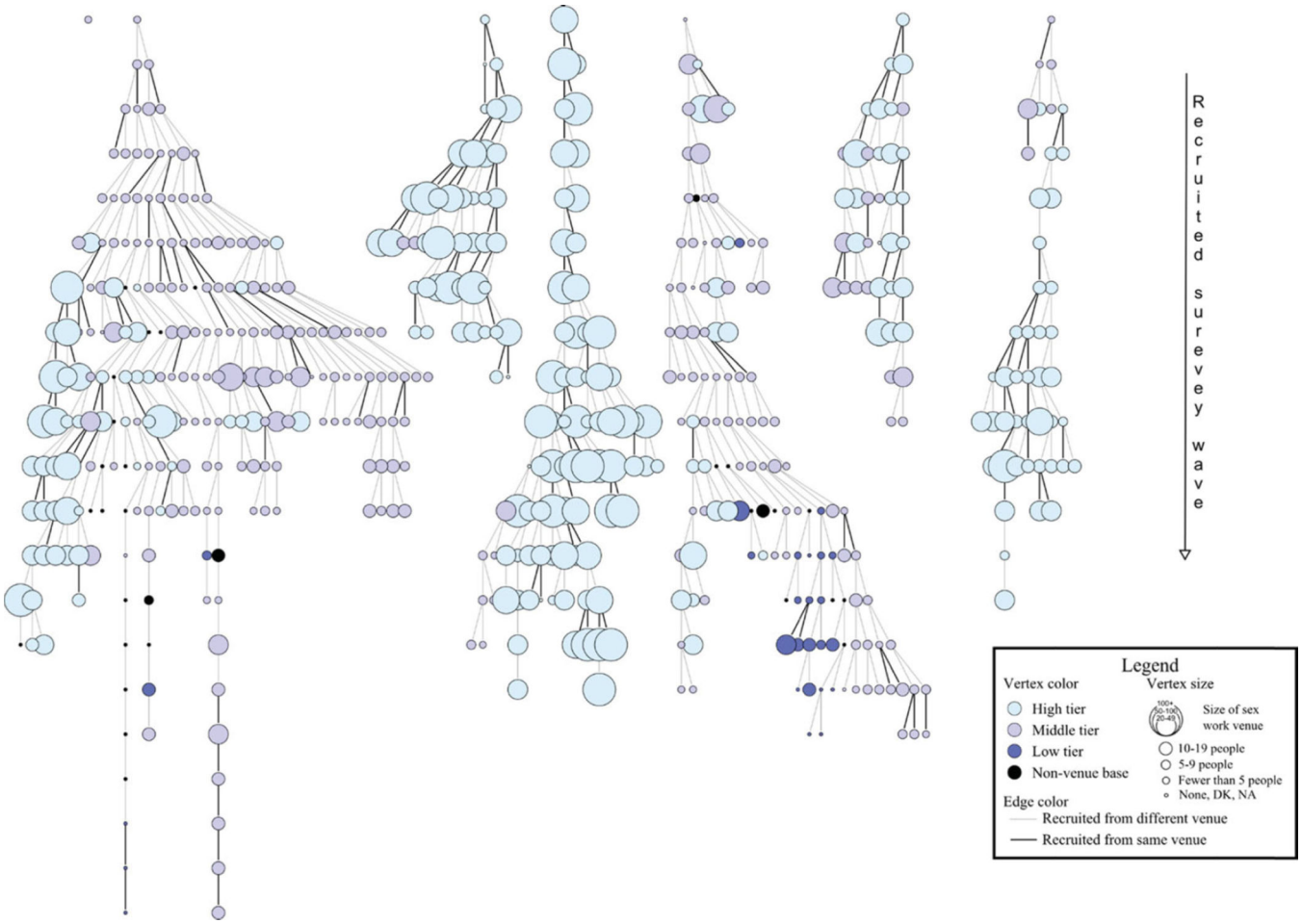

Fig. 1.

Recruitment chains from a respondent-driven sample of FSW collected in Liuzhou, China. (Color online)