Respondent Driven Sampling with Online Recruitment and Adaptive Follow-ups

Presenter:
Deirdre Middleton, MPH ICF International
Co-authors:
Christian Evans, MA ICF International
Naomi Freedner, MPH ICF International
Ronaldo Iachan, Ph.D ICF International
Karen Trocki, Ph.D, Alcohol Research Group

Proceedings of the 2015 Federal Committee on Statistical Methodology (FCSM) Research Conference

Introduction

The National Alcohol Survey (NAS) has been conducted since 1979 by the Alcohol Research Group (ARG) to assess alcohol use and associated risk behaviors. Though these national surveys have limited ability to produce estimates in small subgroups, it has often been a launching pad for further study. In this case the data suggested there may be differences in alcohol use and other health risk behaviors in sexual minority women (SMW)—those who self-identify as lesbian or bisexual (Drabble et al., 2013; Trocki et al 2009). Both national and local surveys have identified that sexual minority women are suffering from substantial health disparities, including significantly higher rates of hazardous drinking. Despite the growing body of research that has identified sexual minority women as being at higher risk for alcohol problems, a limited number of studies have sought to understand why this disparity exists. Sexual minority women may be disproportionately impacted by risk factors that are known to correlate with alcohol consumption among women, including underemployment, job discrimination, stressors related to multiple roles and family conflict, and traumatic experiences such as early childhood sexual abuse (Drabble et al., 2016). Based on this background, the research aim of this study is to identify and compare mediators and moderators of substance use compared to heterosexual populations.

Methodological challenges to studying these problems include sexual minority women having a low population prevalence of 2-5% of women (Cochran, et. al, 2007; Hughes, McCabe et. al, 2010), which means there is not a sufficient sample size for in depth analysis in nationally representative surveys. Additionally, there is no convenient frame to sample sexual minority women from. In response to this public health research need and the methodological limitations, this study adapted a respondent driven sampling (RDS) methodology in order to assess alcohol use and covariates in a national sample of sexual minority women.

RDS is a methodology designed to sample "hard-to-reach" groups of individuals who are socially connected. In some cases, these "hidden" populations are socially stigmatized, and the surveys may involve sensitive topics.

Respondent driven sampling methods were introduced by Heckathorn (1997) as an effective approach to sample and survey hidden populations. Heckathorn (2002) developed weights based on the RDS network sampling structure to derive valid population estimates; Salganik (2006) and Volz and Heckathorn (2008) suggested ways of computing variance estimates. Wejnert and Heckathorn (2008) explored the use of web-based networks for online studies. This approach was then applied in several studies (Bauermeister et al, 2012, Arfken et al, 2012) of limited scope. Also limited in geography was the recent study of sexual minority women in Chicago (Martin et al, 2015). Examples of populations studied via RDS include intravenous drug users, sex workers, and gay men. Our study extends the approach to a national sample based on a nationally representative probability sample used to generate the seeds for the RDS sample.

During the formative research phase of an RDS study, researchers identify and recruit seeds: people who are part of the hard-to-reach population and in many cases well-respected by that population and eligible for the study. Traditionally, seeds participate in the survey and are given coupons to recruit additional participants from their social network. They hand out coupons to potential respondents and receive incentives for successful recruits. Respondents who complete interviews are then eligible to become recruiters themselves. Chains of respondents develop based on how successfully seeds and recruiters find willing respondents. Each participant is asked about the size of her social network (how many people they know and could potentially give coupons to). During analysis, the relative network size of participants is then taken into account. In this way, RDS combines aspects of snowball (convenience) sampling and probability sampling. RDS is most appropriate in cases where there is no sampling frame, members can identify each other, and the population is socially connected.

RDS has generally been conducted using face-to-face interviews and physical paper couponsfor recruitment. In many cases, residence within a particular geographic area is part of the eligibility criteria. This introduces limitations. Social networks need to be geographically clustered, which limits the methodology's usefulness for rural or geographically dispersed populations. This also means it is not effective for forming nationally representative samples. Additionally, with paper coupons, recruiters must share their coupons in face-to-face interactions. This limits with whom a recruiter can share her coupons and may increase the time it takes for recruitment to occur. For example, a recruit may have to wait to recruit people until she sees them at a weekly event rather than recruiting immediately after completing the survey.

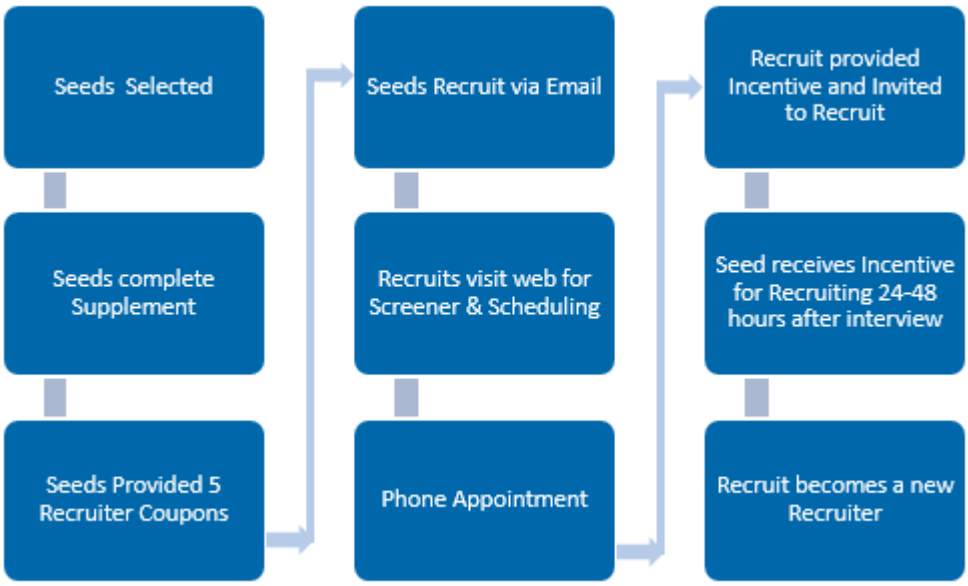
SMW Pilot Study Methodology

RDS typically uses a convenience sample or well-known members of the population of interest as seeds. In contrast, our pilot uses "seeds" from the NAS. Since the NAS uses random digit dialing to recruit participants, these seeds can be considered randomly selected.

The 2013 NAS yielded a total of 73 women who met the eligibility criteria for the SMW study. They were at least 18 years old, identified as lesbian or bisexual and indicated during participation in the NAS that they were willing to be contacted for additional research. From the 73 possible seeds, 15 were randomly selected for the first wave of recruitment into the SMW pilot study.

Seeds were recruited via phone and invited to participate in the survey. Our phone protocol was a total of 15 attempts until we received either a refusal, agreement to participate or confirmation of a wrong number.

Those who agreed to participate and completed the survey where invited to be recruiters and provided with 5 coupons via email. Seeds were provided an incentive (an Amazon gift card) for completing the survey and for each eligible recruit who completed the survey. The coupons provided via email contained links to a web screener. Recruits who complete the web screener and are eligible may schedule a time they are available by phone to complete the survey. If they are unavailable or do not answer at the designated time, the same protocol is used to attempt to recruit them and complete the survey as what is described for the seeds above. Recruits are also given the opportunity to recruit up to 5 eligible friends or family members. Seeds complete a shorter phone survey (approximately 15 minutes) which will be combined with their responses from the NAS for analysis. Recruits complete a longer more comprehensive survey (approximately 1 hour) which includes questions specific to this study and questions from the NAS.



The study aim is to collect survey results from 1,000 sexual minority women. The data will be compared to survey results of 600 heterosexual women. To obtain the heterosexual comparison sample, self-identified heterosexual women were randomly sampled from the 2013 NAS. They are contacted by phone and are asked to complete a 15-minute CATI supplemental survey.

Discussion

Using phone and web-based RDS allows us to pursue a nationally representative sample of a hard to reach population. We anticipate this will allow rapid recruitment of socially connected individuals that may not meet regularly or live in the same area.

A methodological challenge we faced in conducting this survey was determining the appropriate number of seeds needed to pilot the study. Starting with diverse seeds increases the chance of recruiting a representative sample, and larger numbers of seeds make diversity more likely. However, in RDS, short chains may exhibit greater seed bias (Gile et al., 2010), which is similar to clustering in other sampling designs. Longer chains exhibit less seed bias. Seed bias is the tendency for participants to be similar to the seed from their recruitment chain. The closer a person is on the recruitment chain to the seed, the greater the seed bias. Seed bias diminishes as recruitment chains lengthen and the chain's characteristics then begin to mimic a probability sample. Minimizing seed bias was an important consideration in deciding on the number of seeds.

Using a small number of seeds limits the ability to start with a diverse sample and can negatively impact recruitment goals. Nonresponding seeds or non-productive chains from seeds will impact ability to reach the desired sample size within the data collection timeline to a greater extent when there are fewer total seeds. However, a smaller number of seeds results in fewer recruitment chains used to reach the target population of 1,000 which means there is less seed bias.

Having a larger number of seeds allows greater diversity in the initial seed sample, and nonresponding or non-productive chains will have minimal impact. However, a larger number of recruitment chains may each be too short individually to adequately address the inherent seed bias. An additional consideration in this study was the limited pool of seeds to select from (73 total). Based on this, 15 seeds were selected with the assumption that approximately 5 would participate in the survey and recruitment.

Analysis of results will be conducted using *The Respondent Driven Sampling Analysis Tool* **(**RDSAT) in order to estimate population proportions of user defined groups, average personal network sizes, and measures of statistical significance for population estimates. Comparisons to the heterosexual sample will be conducted using standard analytic software with weights in the RDS branch created using RDSAT.

References

Bauermeister, J.A., Zimmerman, M.A., Johns, M.M., Glowacki, P., Stoddard, S. and Volz, E. (2012). Innovative recruitment using online networks: lessons learned from an online study of alcohol and other drug use utilizing a web-based respondent-driven sampling (webRDS) strategy. J. of Studies on Alcohol and Drugs, 834-838.

Cochran SD, Mays VM, Ortega AN, Alegria M, Takeuchi D. Mental health and substance abuse disorders among Latino and Asian American lesbian, gay, and bisexual adults. Journal of Consulting and Clinical Psychology. 2007;75(5):785–794.

Drabble L., Trocki K.F., Hughes T.L., Korcha R.A., & Lown E. A. (2013). Sexual orientation differences in the relationship between victimization and hazardous drinking among women in the National Alcohol Survey. Psychology of Addictive Behaviors, 27 (3) 639-48

Drabble, L., Trocki, K.F., Salcedo, B., Walker, P.C., Korcha, R.A. (2016). Conducting qualitative interviews by telephone: lessons learned from a study of alcohol use among sexual minority and heterosexual women. Qualitative Social Work, 15 (1) 118-133

Trocki, K.F., Drabble L.A., Midanik L.T. (2009). Tobacco, marijuana, and sensation seeking: comparisons across gay, lesbian, bisexual, and heterosexual groups. Psychology of Addictive Behaviors, 23 (4) 620-631

Gile KJ, Handcock MS. Respondent-driven sampling: an assessment of current methodology. Sociol Methodol. 2010;40(1):285–327.

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

Heckathorn, DD (1997) Respondent-driven sampling II: deriving valid population estimates from chain referral samples of hidden populations. Social Problems, 44(2), 174-199.

Hughes TL, McCabe SE, Wilsnack SC, West BT, Boyd CJ. Victimization and substance use disorders in a national sample of heterosexual and sexual minority women and men: Sexual identity, victimization and SUDs. Addiction. 2010;105(12):2130–2140.
Martin, K., Johnson, T.P., Hughes, T.L. (2015) Using respondent driven sampling to recruit sexual minority women. Survey Practice, 8(1), 1-10.

Salganik MJ (2006) Variance estimation, design effects, and sample size calculations for respondent-driven sampling. *Journal of Urban Health-Bulletin of the New York Academy of Medicine*, 83, 98-112.

Volz E, and Heckathorn DD (2008). Probability Based Estimation Theory for Respondent Driven Sampling. *Journal of Official Statistics*, 24, 79-97.

Wejnert, C. and Heckathorn, DD (2008). Web-based network sampling: efficiency and efficacy of respondent-driven sampling for online research. Sociological Methods and Research, 37 (1), 105-134.