

IMPLEMENTATION, POLICY AND COMMUNITY ENGAGEMENT BRIEF REPORT

Respondent-driven sampling to recruit Latinos in a midwest micropolitan area: Lessons learned and recommendations for translational work

Paul A. Gilbert^{1,2,3*}, Heidi Haines^{1,2}, Barbara Baquero^{1,2,3} and Edith A. Parker^{1,2,3}

¹ Department of Community and Behavioral Health, University of Iowa College of Public Health, Iowa City, IA, USA

² Prevention Research Center, University of Iowa College of Public Health, Iowa City, IA, USA

³ Institute for Clinical and Translational Science, University of Iowa, Iowa City, IA, USA

Journal of Clinical and Translational Science (2018), 2, pp. 245–248 doi:10.1017/cts.2018.322

Latinos constitute a hard-to-reach minority population in Iowa. We used respondent-driven sampling (RDS) to supplement random digit dialing to recruit Latinos for a community physical activity intervention. RDS yielded a 59% increase in Latino participation in just 2 months, with few demographic differences between RDS and random digit dialing groups. RDS may increase recruitment of underrepresented populations and strengthen community engagement; however, it is not a quick fix for underperforming recruitment methods.

Received 3 April 2018; Revised 11 July 2018; Accepted 17 July 2018

Key words: Respondent-driven sampling, community engagement, Hispanic, intervention, physical activity.

Background

Respondent-driven sampling (RDS) is a methodology that was developed to overcome the challenges of sampling “hidden” and hard-to-reach populations [1, 2]. An extension of chain referral methods, RDS leverages social networks and takes into account network properties to generate approximately representative samples of populations for whom no sampling frame exists. Originally used in studies of injection drug users, RDS has been successfully applied in studies with other populations for whom the stigma associated with group membership makes recruitment difficult, such as sex workers, immigrants, and sexual or gender minorities [3–7]. Indeed, over the past 2 decades RDS has been widely adopted in epidemiological studies; however, it has been used less often in applied public health work, such as intervention trials. Among notable examples, however, RDS formed the

basis of a social network intervention to promote uptake of pre-exposure prophylaxis for HIV prevention among young Black men who have sex with men [8] and to recruit parents of adolescents into a family-based substance use prevention program [9].

Recognizing its potential utility to recruit underrepresented minority participants, The University of Iowa Prevention Research Center collaborated with the Integrating Special Populations core of the Iowa Institute for Clinical and Translational Studies to incorporate RDS into *Active Ottumwa*, a CDC-funded community-level intervention trial to promote physical activity in a micropolitan community in southeast Iowa (U48DP005021). The project has been described in detail elsewhere [10]; in brief, *Active Ottumwa* is a 5-year community-based physical activity intervention that uses lay health advisors to inform residents about physical activity, provide social and behavioral support, and advocate for policy and environmental changes. The evaluation assessments of *Active Ottumwa* take place at the individual, community, and policy levels. One of these evaluations is a longitudinal cohort study with a sample of community residents to measure individual changes in physical activity. Latinos constitute ~11% of Ottumwa residents [11] but are considered a hard-to-reach population in Iowa, due to their relatively recent migration to the state and because their social networks remain largely unknown to social services and public health providers. Furthermore, cultural differences and government policies often compel Latinos to isolate themselves from the larger

* Address for correspondence: P. A. Gilbert, PhD, ScM, Department of Community and Behavioral Health, University of Iowa College of Public Health, 145N. Riverside Drive, N414 CPHB, Iowa City, IA 52242, USA.

(Email: paul-gilbert@uiowa.edu)

community. Therefore, to ensure adequate statistical power for comparisons between Latino and non-Latino participants in our cohort study, we used RDS methods to increase the sample of Latinos. This paper briefly describes our experience and reports lessons learned that may inform other intervention trials.

Methods

In its initial year, *Active Ottumwa* used a random digit dialing (RDD) telephone survey to recruit a representative cohort of town residents to complete a baseline behavioral survey. Subsequently, Latinos who were recruited via RDD were asked to serve as RDS “seeds,” thereby initiating the RDS recruitment process. Those who agreed to serve as RDS seeds received an explanation of the study’s eligibility criteria and an overview of RDS methods. We told seeds that they could invite up to 3 people in their social network to participate in the survey and gave them 3 recruitment coupons with unique identifier numbers to distribute to each person whom they invited. Individuals receiving these coupons (“referrals”) then contacted the study office if they were interested in participating. In turn, referrals who were eligible and completed a baseline survey were then given 3 coupons of their own to distribute to members of their social network, thereby continuing the RDS recruitment process. Dual incentives are a hallmark of RDS methods. In addition to participants being offered a \$25 gift card for their own participation, seeds received an additional \$5 gift card for each referral who participated in the survey; however, seeds had no knowledge of their referrals’ actual participation in the survey unless the referral provided this information to them. To ensure confidentiality, referrals were not asked to provide the name of the person who referred them to the study, only to present the referral coupon that they had received. We obtained Institutional Review Board (IRB) approval to add RDS recruitment in May 2016 and implemented it in May and June 2016. We were only able to devote 2 months to active RDS recruitment due to the deadline to complete Year 1 baseline recruitment and begin follow-ups.

We collected and managed data using the Research Electronic Data Capture (REDCap) application hosted at the University of Iowa Institute for Clinical and Translational Sciences [12] and tracked recruitment chains using unique identifier numbers in an Excel database, which was kept separate from other study data. RDS data require special handling in analyses and cannot be treated as a simple random sample for statistical tests. Analytic methods have been described in detail elsewhere [13–16]. As this brief report focused only on describing the recruitment process and the resulting sample rather than making inferences about the larger population, such adjustments were not necessary. We calculated descriptive statistics and compared Latinos recruited via RDD and RDS using SAS/STAT software v9.4 (SAS Institute, Cary, NC). Due to the small sample size, we used exact statistical test (e.g., Fisher’s test). We also estimated the proportion of participants in each group who were retained for follow-up surveys. Process notes by the study team provided additional data for lessons learned.

Results

The *Active Ottumwa* cohort evaluation was designed to have a target sample size of 174. Based on a priori power estimates for comparisons between Latinos and non-Latinos on physical activity minutes (the main study outcome), our goal was to include at least 50 Latino participants. Near the end of Year 1 baseline recruitment we noted that RDD recruitment had yielded only 22 Latino participants, which fell short of our target and prompted our adoption of RDS as a supplemental recruitment strategy.

Of the 22 Latinos recruited via RDD, half ($n = 11$) agreed to serve as RDS seeds. Among those who did not serve as seeds, 4 individuals

stated that they did not know anyone who would want to participate and refused, 2 expressed initial interest but failed to keep study appointments 3 times or more, at which point we stopped contacting them. We were unable to contact the remaining 5 RDD Latino participants to invite them to serve as RDS seeds.

Of those who agreed to serve as RDS seeds, 6 participants produced no referrals. In contrast, 5 RDS seeds produced on average 2.6 referrals each, yielding 13 additional Latino participants. Fig. 1 shows recruitment chains. To gain further insights about seeds, we contrasted demographic characteristics of productive versus nonproductive seeds (online Supplementary Appendix A). Although the very small number precluded statistical tests of difference, nonproductive seeds had an older average age, a greater proportion was in the lowest income stratum, and all reported good/fair/poor health. Thus, productive seeds may have leveraged their better health, higher socioeconomic position, and younger age to successfully recruit other Latinos.

Overall, there were many similarities in demographic characteristics between RDD and RDS participants, such as no significant differences in average age or distributions of gender, educational attainment, income, marital status, and self-rated health (Table 1). Among differences, the majority (64%) of RDD participants owned their apartment or house while a comparable majority (62%) of RDS participants rented their apartment or house. In addition, the majority (77%) of RDD participants had health insurance but the majority (69%) of RDS participants did not. It appeared that RDD participants had lived in Ottumwa longer on average than RDS participants; however, the difference was marginally significant. We noted that our tests may have been underpowered to detect differences between RDD and RDS participants; for example, with a larger sample we might have seen a significant difference in the distribution of marital status. Approximately equivalent proportions of RDD and RDS participants were retained for 12-month follow-up (68 vs. 62%, respectively; $p = 0.69$). At this writing, 24-month follow-up is underway.

Reviewing process data, we identified 2 main challenges to recruitment via RDS: (1) logistical challenges in recruiting RDS seeds; (2) and limited study personnel resources. First, reaching Latino RDD participants was challenging as many of them worked multiple jobs and/or different shifts. On average, study staff called participants 6 times in order to ask if they would serve as RDS seeds. Once the agreement was obtained, we then had to schedule a new appointment at the *Active Ottumwa* office to reconsent seeds (a stipulation of our IRB as the study procedures had changed), explain the RDS process, and provide the recruitment coupons. This was a time-consuming process. We often had to reschedule these appointments as well as encountering frequent no-shows. Accordingly, we ceased efforts to enroll seeds after 3 missed appointments, which resulted in the exclusion of 2 potential seeds. Second, *Active Ottumwa* had only 2 part-time bilingual employees at the time of RDS recruitment. The limited Spanish-speaking staff meant that making the multiple recruitment calls and rescheduling appointments was especially

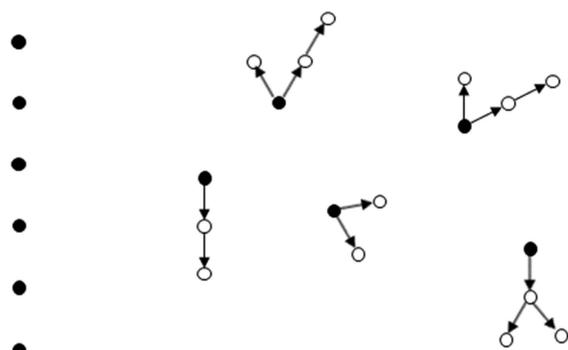


Fig. 1. Respondent-driven recruitment chains. Solid circle indicates seed.

Table 1. Latino participants' demographic characteristics by sampling method

| | RDD (n = 22) | RDS (n = 13) |
|------------------------------------|--------------|----------------|
| Age, years (mean) | 37 | 34 |
| Gender (%) | | |
| Female | 64 | 77 |
| Male | 36 | 23 |
| Educational attainment (%) | | |
| Primary | 32 | 31 |
| Secondary | 41 | 54 |
| Post-secondary | 27 | 15 |
| Income (%) | | |
| < \$15,000 | 25 | 33 |
| \$15,000–\$29,999 | 40 | 17 |
| \$30,000–\$49,999 | 10 | 17 |
| ≤ \$50,000 | 25 | 33 |
| Marital status (%) | | |
| Married/cohabitating | 68 | 46 |
| Divorced/separated | 9 | 15 |
| Never married | 23 | 38 |
| Housing tenure (%) | | |
| Own apartment/house | 64 | 15* |
| Rent apartment/house | 27 | 62 |
| Do not own or rent apartment/house | 9 | 23 |
| Years living in Ottumwa (mean) | 13 | 8 [†] |
| Health insurance (%) | | |
| Any | 77 | 31* |
| None | 23 | 69 |
| Self-rated health (%) | | |
| Excellent/very good | 27 | 23 |
| Good/fair/poor | 73 | 77 |

RDD, random digit dialing; RDS, respondent-driven sampling.

* $p < 0.05$; [†] $p < 0.10$.

challenging. In effect, participants' scheduling challenges were compounded by limited availability of study personnel.

Discussion

We used RDS methods as an adjunct to RDD to recruit Latino participants—an ethnic minority population that is considered hard-to-reach in Iowa—for a community-level physical activity intervention trial. RDS methods were moderately successful, yielding a 59% increase in Latino participation in just 2 months of active recruitment; however, despite combined RDD and RDS methods we failed to reach the target sample size of Latinos. Nevertheless, the process of implementing RDS recruitment provides several lessons that may inform future translational research activities, particularly those related to patient and community engagement.

First, the high proportion of nonproductive seeds is partially responsible for our failure to recruit the target number of Latinos. It is common practice to recruit additional seeds when faced with low referrals or nonproductive seeds [17]. That was not possible in our case as we had exhausted potential seeds recruited via RDD. Furthermore, we did not initially assess characteristics that may have hindered recruitment, such as poor health and low socioeconomic status. We encourage future work to select RDS seeds based on capacity to engage in recruitment activities as well as social network connections. Second, the short period devoted to RDS recruitment is also partially responsible for our failure to recruit the target number of Latinos. Indeed, we note that productive seeds made a good number of referrals on average. If we had extended the recruitment period, RDS chains might have continued and

yielded the target sample size. However, that was not possible due to the need to complete baseline surveys and begin the intervention to comply with the overall study timeline. In addition, the study was originally designed to use RDD recruitment only. We incorporated RDS as an adjunct strategy late in Year 1 when we realized that the original recruitment plan would not yield the desired sample size. Although flexibility in research is good, it is likely necessary to plan earlier to maximize the utility of RDS methods, particularly through pilot tests of its feasibility and planning for contingencies. Finally, we detected demographic differences between RDD recruits and RDS recruits on housing tenure, health insurance coverage, and years living in Ottumwa. Combining these sub-groups will inflate variance for those variables, which would introduce bias toward the null hypothesis (i.e., less likelihood of finding an association with the study outcome). While we recognize this possibility, the small number of RDS recruits ensures only a minimal effect on inferences.

We note that there has been considerable attention to improving RDS analytic methods [14, 18]; however, there has been less attention to strengthening the implementation process, particularly in applied research. Thus, we report our experiences as a means of sharing lessons learned. Despite our limited success, we strongly believe that RDS methods have a role to play in translational research, particularly in attempts to integrate special populations which have previously not been included in sufficient numbers in translational research. For example, we think it is important to build trusted community relationships and maintain open channels of communication as a precondition of patient and community engagement. In fact, RDS supports several tenants of community-engaged research, such as engaging participants in the research process, allowing for a representative sample of an otherwise “hidden population,” and necessitating that researchers understand the population's target patterns and characteristics [19, 20].

In sum, we support the use of RDS as a potential method to both increase recruitment of underrepresented populations in research and to address a key challenge in patient and community engagement. However, our experience suggests that RDS is not a quick fix for other underperforming recruitment methods and that studies that use RDS must plan carefully and ensure that sufficient staffing and resources are allotted to this endeavor.

Acknowledgments

Research reported in this publication was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number U54TR001356 and by the Centers for Disease Control and Prevention under Cooperative Agreement U48DP005021. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, the Centers for Disease Control and Prevention, or the Department of Health and Human Services.

Disclosure

The authors have no conflicts of interest to declare.

Supplementary material

To view supplementary material for this article, please visit <https://doi.org/10.1017/cts.2018.322>

References

1. Heckathorn DD. Respondent-driven sampling: a new approach to the study of hidden populations. *Social Problems* 1997; 44: 174–199.

2. **Ramirez-Valles J, et al.** From networks to populations: the development and application of respondent-driven sampling among IDUs and Latino gay men. *AIDS and Behavior* 2005; **9**: 387–402.
3. **Montealegre JR, et al.** Effectiveness of respondent driven sampling to recruit undocumented Central American immigrant women in Houston, Texas for an HIV behavioral survey. *AIDS and Behavior* 2013; **17**: 719–727.
4. **Bauer GR, et al.** Intervenable factors associated with suicide risk in transgender persons: a respondent driven sampling study in Ontario, Canada. *BMC Public Health* 2015; **15**: 525.
5. **Carballo-Diequez A, et al.** Use of respondent driven sampling (RDS) generates a very diverse sample of men who have sex with men (MSM) in Buenos Aires, Argentina. *PLoS One* 2011; **6**: e27447.
6. **Musyoki H, et al.** Prevalence of HIV, sexually transmitted infections, and risk behaviours among female sex workers in Nairobi, Kenya: results of a respondent driven sampling study. *AIDS and Behavior* 2015; **19**(Suppl. 1): S46–S58.
7. **Manopaiboon C, et al.** Unexpectedly high HIV prevalence among female sex workers in Bangkok, Thailand in a respondent-driven sampling survey. *International Journal of STD and AIDS* 2013; **24**: 34–38.
8. **Young LE, et al.** PrEP Chicago: a randomized controlled peer change agent intervention to promote the adoption of pre-exposure prophylaxis for HIV prevention among young Black men who have sex with men. *Clinical Trials (London, England)* 2018; **15**: 44–52.
9. **Oesterle S, et al.** Using Facebook to recruit parents to participate in a family program to prevent teen drug use. *Prevention Science: The Official Journal of the Society for Prevention Research* 2017; **19**: 559–569.
10. **Baquero B, et al.** Active Ottumwa: adapting evidence-based recommendations to promote physical activity in a micropolitan new destination community. *International Journal of Environmental Research and Public Health* 2018; **15**: e917.
11. **Census Bureau/American Fact Finder.** *Profile of General Population and Housing Characteristics* [Internet], 2010 [cited Feb 28, 2018]. (<https://factfinder.census.gov>)
12. **Harris PA, et al.** Research electronic data capture (REDCap) – a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics* 2009; **42**: 377–381.
13. **Heckathorn DD.** Respondent-driven sampling II: deriving valid population estimates from chain-referral samples of hidden populations. *Social Problems* 2002; **49**: 11–34.
14. **Salganik MJ.** Variance estimation, design effects, and sample size calculations for respondent-driven sampling. *Journal of Urban Health* 2006; **83**: 198–1112.
15. **McCreesh N, et al.** Respondent driven sampling: determinants of recruitment and a method to improve point estimation. *PLoS One* 2013; **8**: e78402.
16. **Salganik MJ, Heckathorn DD.** Sampling and estimation in hidden populations using respondent-driven sampling. *Sociological Methodology* 2004; **34**: 193–239.
17. **Rhodes SD, et al.** Prevalence estimates of health risk behaviors of immigrant Latino men who have sex with men. *Journal of Rural Health* 2012; **28**: 73–83.
18. **Gile KJ, Handcock MS.** Respondent-driven sampling: an assessment of current methodology. *Sociological Methodology* 2010; **40**: 285–327.
19. **Minkler M, Wallerstein N.** *Community Based Participatory Research for Health*. San Francisco, CA: Jossey-Bass, 2003.
20. **Israel BA, et al.** *Methods in Community-Based Participatory Research for Health*. San Francisco, CA: Jossey-Bass, 2005.