



Published in final edited form as:

*J Epidemiol Community Health*. 2014 August ; 68(8): 792–798. doi:10.1136/jech-2014-203935.

## Spatial, temporal and relational patterns in respondent-driven sampling: evidence from a social network study of rural drug users

April M Young<sup>1,2</sup>, Abby E Rudolph<sup>3</sup>, Deane Quillen<sup>4</sup>, and Jennifer R Havens<sup>2</sup>

<sup>1</sup>Department of Epidemiology, University of Kentucky, Lexington, Kentucky, USA

<sup>2</sup>Department of Behavioral Science, Center on Drug and Alcohol Research, University of Kentucky, Lexington, Kentucky, USA

<sup>3</sup>The Calverton Center, Pacific Institute for Research and Evaluation, Calverton, Maryland, USA

<sup>4</sup>Division of General Pediatrics, Vanderbilt University Medical Center, Nashville, Tennessee, USA

### Abstract

**Background**—Respondent-driven sampling (RDS) has become a common tool for recruiting high-risk populations for HIV research. However, few studies have explored the influence of geospatial proximity and relationship-level characteristics on RDS recruitment, particularly among high-risk individuals residing in rural areas of the US.

**Methods**—In a social network study of 503 drug users in rural Central Appalachia, interviewer-administered questionnaires were used to collect relationship-level data (eg, duration of relationship, frequency of communication, kinship, social/financial support, trust, drug use and sex) and residential location. Demographic and drug-use similarity were also evaluated. Residential data were geocoded and road distance (km) between participants and (1) their network members and (2) the study site were computed. Seasonal patterns were assessed using node-level analysis, and dyadic analyses were conducted using generalised linear mixed models. Adjusted ORs (AORs) and 95% CIs are reported.

---

Copyright Article author (or their employer) 2014. Produced by BMJ Publishing Group Ltd under licence.

Correspondence to: Dr April M Young, Center on Drug and Alcohol Research, University of Kentucky, College of Medicine, 333 Waller Avenue, Lexington, Kentucky 40504, USA; april.young@uky.edu.

#### Competing interests

None.

#### Ethics approval

University of Kentucky Institutional Review Board.

#### Provenance and peer review

Not commissioned; externally peer reviewed.

#### Contributors

AMY conceptualised and conducted the data analysis and drafted the manuscript. AER provided substantial contributions to the creation of the geospatial measures, conceptualisation of the analysis and manuscript drafts. DQ assisted with the conceptualisation of the overall aim of the study, provided support in the primary collection and management of the geospatial data, and provided feedback on the manuscript drafts. JRH oversaw the study, assisted with the conceptualisation of the analytic approach and gave substantive feedback on the manuscript draft. All authors approved the final manuscript.

**Results**—Differences in distance to the study office by season and order of study entry were not observed ( $F=1.49$ ,  $p=0.209$  and  $\beta=0.074$ ,  $p=0.050$ , respectively). Participants with transportation lived significantly further from the interview site than their counterparts ( $p<0.001$ ). Dyadic analyses revealed no association between RDS recruitment likelihood and geographic proximity. However, kinship (AOR 1.62; CI 1.02 to 2.58) and frequency of communication (AOR 1.63; CI 1.25 to 2.13) were significantly associated with RDS recruitment.

**Conclusions**—In this sample, recruitment from one's network was likely non-random, contradicting a core RDS assumption. These data underscore the importance of formative research to elucidate potential recruitment preferences and of quantifying recruitment preferences for use in analysis.

## INTRODUCTION

Surveillance is crucial for understanding factors that influence the spread of HIV and hepatitis C (HCV). However, due to issues of stigma, illegality and the absence of adequate sampling frames for populations at high risk for infection (eg, people who inject drugs (PWID), men who have sex with men (MSM) and commercial sex workers),<sup>1</sup> sampling from these populations can be difficult. Respondent-driven sampling (RDS) was developed to address this challenge.<sup>23</sup> RDS is a network-based sampling technique whereby purposively sampled initial participants or 'seeds' are identified and provided with a limited number of referral coupons to recruit their peers. The recruited peers are in turn asked to recruit their peers, and so on until the desired sample size is reached. Incentives are provided for peer recruitment. As an analytic strategy, RDS incorporates weights to account for sampling biases related to individuals' network sizes.<sup>23</sup>

RDS has become a common tool for recruiting high-risk populations for HIV surveillance.<sup>4</sup> The US National HIV Behavioral Surveillance System uses RDS to recruit PWID, MSM and high-risk heterosexual individuals in 25 metropolitan statistical areas.<sup>56</sup> From 2003 to 2007, RDS was used in over 120 HIV biological and behavioural studies in 28 countries involving over 32 000 high-risk individuals,<sup>7</sup> and by January 2013, RDS had been used by researchers in over 80 countries.<sup>8</sup>

Despite its widespread use and utility for recruitment, as an analytic tool, some have questioned its ability to produce representative samples<sup>910</sup> and the accuracy of the resulting variance estimates.<sup>91112</sup> In order for RDS to generate asymptotically unbiased prevalence estimates, a number of assumptions must be met. One of the most frequently challenged assumptions is that individuals *randomly* select recruits from their personal networks. Studies comparing demographic characteristics and risk behaviours of RDS peer recruits with those of egocentric network members reported by RDS participants typically reported differences, suggesting that peer recruitment may not be random.<sup>12–15</sup> Two studies demonstrated the presence of non-random recruitment in respondent-driven samples,<sup>1416</sup> and some suggest that recruitment probability is likely driven by factors other than network size.<sup>17</sup> Yet, few studies have used sociometric network data to evaluate the influence of demographic and behavioural similarity *and* relationship characteristics on RDS recruitment.

Heckathorn *et al*<sup>1</sup> reported that RDS could be used to recruit a diverse sample from a wide geographic area, citing their success in recruiting nearly 200 PWID from a city that was 15 miles away from the interview site. They also noted that racial and ethnic disparities in transportation access could cause sampling bias and subsequently moved the field site. The authors describe this as a *geographic filtering effect*, which occurs when there is differential transportation access across subgroups. The authors suggested that RDS samples are most representative when respondents are local and the interview site is easily accessible.<sup>1</sup>

Although some studies have examined the influence of different aspects of geography on peer recruitment,<sup>18–24</sup> few have specifically examined the role of geospatial proximity in RDS recruitment within peer networks. One study found that a geographically concentrated set of seeds could successfully recruit PWID from a broader geographic area.<sup>20</sup> Other studies have demonstrated more geographic heterogeneity in recruitment success.<sup>18,19</sup> Studies have suggested that RDS recruiters tend to refer individuals residing nearer,<sup>21–23</sup> and that spatial dispersal of networks may be greater than that of recruitment chains.<sup>24</sup> With the exception of one study conducted in Uganda,<sup>22</sup> the aforementioned studies have been conducted in urban settings.<sup>18–21,23,24</sup> Little is known about spatial patterns in RDS recruitment of high-risk populations in rural settings. Although some have suggested that geographic proximity may influence RDS recruitment of rural drug users,<sup>25</sup> no study to the authors' knowledge has empirically examined this proposition nor explored the impact of relationship-level characteristics in this population. The purpose of this analysis was to examine spatiotemporal and relationship-based characteristics associated with RDS recruitment of rural, high-risk drug users.

## METHODS

### Sample

The data for this analysis were collected during the baseline assessment of the longitudinal Social Networks among Appalachian People (SNAP) study (described in detail elsewhere<sup>26</sup>). The purpose of SNAP was to examine the prevalence of HIV, HCV and HSV-2 among illicit drug users in a rural, mountainous community located in Central Appalachia. Interviewer-administered questionnaires were conducted at a study office located in the local town centre. A total of 506 participants were interviewed, but three were excluded for reporting discrepant substance use information on their screening assessment and interview, leaving a final sample size of 503. All participants provided written informed consent. The protocol was approved by the University Institutional Review Board and a Certificate of Confidentiality was obtained. Participants were compensated \$50.

Participants were recruited using RDS. Seeds (n=108, including one participant who was later excluded from the study) were identified through outreach workers, community informants and flyers. Spatial distribution was not considered in seed selection. Upon completion of the interview, seeds received three coupons to recruit others. Due to administrative error, four participants recruited four individuals. Participants received \$10 for each coupon redeemed by an eligible peer recruit. From November 2008 to August 2010, recruitment proceeded through a maximum of 14 waves. Over 50% (n=57) of the seeds were

non-generative. Figure 1 displays the RDS referral chains (hereafter referred to collectively as the *RDS network*), and figure 2 displays recruitment by wave.

### Spatial data

Residential data were geocoded using ArcGIS (V.10) and ESRI's North American Street Map data.<sup>27</sup> Three participants who did not provide complete addresses and an additional 12 addresses that could not be geocoded were not analysed (final N=491). Participants whose addresses could not be geocoded (n=12) were not significantly different from those who were geocoded in terms of age, race, gender or months of education. Figure 1 displays the location of non-geocoded participants in the RDS network.

Data on streets, interstate highways and major roads was obtained from the North America Detailed Streets Layer Package (updated July 2013).<sup>27</sup> A road network for Kentucky and bordering states was built using Network Analyst in ArcMap V.10.1 and used to compute shortest distance (km) between each participant's residence and the study office and other participants' residences.

### Network data collection and analysis

The network data collection methods are described elsewhere.<sup>28</sup> Briefly, on a name-generating questionnaire, participants (ie, *egos*) gave the first name and last initial, gender and age of up to eight individuals (ie, *alters*) whom they had received and/or provided social support, used drugs (excluding alcohol and marijuana) and had sex with during the past 6 months (total of 24 alters possible). To determine whether the named alter was a participant, the names and demographic details were cross-referenced with those of other participants. In cases where the confirmation was questionable, the community-based interviewers were queried for their knowledge of existing relationships. These techniques are similar to those used in similar research.<sup>29</sup> Each confirmed relationship between two geocoded participants is hereafter referred to as a *network dyad*.

**Social and behavioural similarity**—Four dyadic measures of demographic similarity between participants were examined: gender (1=same gender, 0=different gender), race/ethnicity (1=same race/ethnicity, 0=different race/ethnicity), absolute difference in age (years) and absolute difference in education (years). Individual transportation access was defined as having a valid driver's license and access to an automobile. In dyadic analysis, this was operationalised with '1' indicating that one or both partners had transportation access, and '0' that neither member of the dyad had transportation access. Similarity in drug use was examined through binary measures in which a '1' indicated that both partners reported past 6-month use of the same substance; this was computed separately for stimulants (crack or cocaine), sedatives (barbiturates or other prescription sedatives), methamphetamine, heroin and illicit use of methadone, oxycodone, oxycontin and hydrocodone.

**Relationship characteristics**—Three binary measures indicating whether or not participants had recently (past 6 months) used drugs together, had sex and exchanged social support were also evaluated. For each named alter, respondents were asked the duration of

their relationship (months), frequency of communication (6-point Likert scale, with increasing values representing more frequent communication) and kinship (including both nuclear and extended family relations) (binary). Respondents were also asked how much they trusted each other (10-point scale), whether they shared drugs (binary) and whether they had received financial support from him/her in the past 6 months (binary). To construct a database in which each observation represented one network dyad, dyadic measures were symmetrised. For example, a report of kinship by either partner was sufficient to define the relationship as involving kinship (ie, even if the report of kinship was not reciprocated). For continuous dyadic measures, the average of the two values reported by partners was used for the value.

### Statistical analyses

To account for potential autocorrelation, direct estimation of the sampling distributions was needed to compute inferential statistics from the network data.<sup>30</sup> Each analysis was conducted within UCINET,<sup>31</sup> and statistical significance was determined based on permutation testing accounting for autocorrelation.<sup>30</sup> ANOVA followed by post hoc t tests (both of which determine significance based on permutation testing) were used to evaluate the association between timing of study entry (season) and distance to the study office and to recruits. t tests were used to examine differences in age and education between geocoded and non-geocoded participants and those with/without transportation access. For categorical covariates,  $\chi^2$  tests were conducted using SPSS (V.20), as permutation-based  $\chi^2$  testing is not currently available in UCINET.<sup>31</sup> Finally, quadratic assignment procedures correlation<sup>32</sup> was used to examine the association between matrices representing participants' spatial proximity and difference in order of study enrolment.

Dyadic analyses were conducted to determine the correlates to the likelihood of RDS recruitment. Given potential autocorrelation among dyads involving the same ego, generalised linear mixed models with a random effect for ego were used. Models were estimated using PROC GLIMMIX<sup>33</sup> (SAS V.9.3) and empirical (sandwich) estimators.

**Outcome variable**—The database contained 663 social, drug and sex network ties; 192 involved RDS recruitment. The outcome variable was RDS recruitment (binary, 1/0) in network ties. RDS recruitment was regressed on each of the dyadic measures described above (example data structure shown in table 1). Covariates reaching  $p < 0.05$  in bivariate analyses were entered into multivariate analyses to examine their independent association with RDS recruitment.

## RESULTS

The demographic and behavioural profile of participants is described in detail elsewhere.<sup>26</sup> Briefly, the majority were white (94%), male (57%), not currently married (74%), had graduated from high school (57%) and were employed (58%). The median age was 31 years (range 18–65). Most participants (61.8%) had transportation access.

### Spatial proximity of participants and study office

The median distance between participants was 11.7 km (range 0–220 km). Geographic distance between participants was positively associated with the difference in participants' order of study entry ( $r=0.312$ ,  $p<0.001$ ). The study office was centrally located, only 0.6 km (straight-line distance) from the median centre of the spatial distribution of participants. The median distance between the office and participants was 6.0 km (IQR 1.0–12.6; maximum 149.4) for all participants and 5.7 km (IQR 0.9–12.9) for seeds. Overall differences in distance to the office by season of study entry ( $F=1.49$ ,  $p=0.209$ ) and order of study entry ( $\beta=0.074$ ,  $p=0.050$ ) were not observed. The number of participants enrolled, the distance between each participant and the office, and the average distance between each recruiter and his/her recruits are displayed by recruitment wave in figure 2. Figure 3 displays the distance between each participant and the office by week of study entry. Of note, the median *and* maximum distances are displayed to show the high degree of variability in distance within each recruitment week.

### Transportation access

Participants with transportation access ( $n=180$ ) lived further from the interview site (mean=12.4 km, SD 15.7) than those without transportation access (mean=7.9 km, SD 10.3,  $p<0.001$ ), and had more years of education (mean=11.9 vs 10.7, respectively;  $p<0.001$ ). Other demographic differences between those with and without transportation were not observed. Among those who referred participants ( $n=227$ ), those with transportation access ( $n=80$ ) recruited participants residing further from them than those without (average=12.9 vs 9.2 km;  $p=0.055$ ; maximum=17.5 vs 11 km;  $p=0.006$ ).

### Seasonality of spatial distance in RDS recruitment

Overall, the average distance between recruiters and their recruits did not vary by season ( $F=2.03$ ,  $p=0.105$ ); the lack of an association was observed among those with and without transportation ( $F=1.59$ ,  $p=0.191$  and  $F=1.47$ ,  $p=0.218$ , respectively) and when controlling for RDS wave ( $\beta=-0.01$ ,  $p=0.509$ ).

### Correlates to RDS recruitment

The RDS recruitment network is shown in figure 1, and correlates to recruitment are shown in table 2. Bivariate analyses revealed no association between likelihood of RDS recruitment and geographic proximity. When stratified by transportation access, the negative association between recruitment and distance neared significance in those partnerships where both partners lacked transportation access (OR 0.97, 95% CI 0.94 to 1.00,  $p=0.083$ ) but not in those in which at least one had transportation (OR 1.00, 95% CI 0.98 to 1.01,  $p=0.498$ ).

Recruitment of opposite-sex peers was more likely than recruitment of same-sex peers; a phenomenon likely driven by the significant association between sexual relationships and RDS recruitment in this predominantly heterosexual sample. Compared with non-RDS dyads, RDS dyads were also more likely to be kin and were characterised by more frequent communication, increased trust, longer duration, exchange of financial support and drug sharing. In multivariate analysis, kinship (AOR 1.62, 95% CI 1.02 to 2.58) and frequency of



communication (AOR 1.63, 95% CI 1.25 to 2.13) remained associated with RDS recruitment.

## DISCUSSION

In this study of rural drug users, spatial proximity of participants' residences was positively associated with similarity in their time of entry into the study, indicating that participants who resided nearer to each other were more likely to enter the study at similar times. Contrary to previous research,<sup>21-23</sup> spatial proximity was not significantly associated with the likelihood of RDS recruitment. When stratified by transportation access, the effect sizes for the association between distance and RDS recruitment remained small; however, the association between distance and RDS recruitment was slightly stronger among those without transportation than those with transportation. This finding may indicate that distance is more likely to play a role in recruitment when both members of a dyad lack transportation. Though transportation access was significantly associated with recruiters' distance from their recruits, the bias referred to by Heckathorn as the *geographic filtering effect*<sup>1</sup> may not be present in this sample. Participants with transportation access lived significantly further from the interview site, but there was not substantial evidence to suggest that there was differential transportation access across demographic subgroups. Although these data seem to indicate that oversampling due to differences in transportation access was unlikely, uncertainty remains given that absence of data on transportation access and demographic characteristics of non-participants. Future studies examining spatially dispersed populations may benefit from purposely selecting a geographically diverse sample of seeds. This strategy would be particularly important if geographical variations in the epidemiological outcomes of interest are suspected and in areas where public transportation is unavailable.

In this study, we identified significant differences in the likelihood of RDS recruitment by relationship-based characteristics. For example, participants were significantly more likely to recruit kin and those with whom they more frequently communicated. Among relationships involving RDS recruitment (n=192), 42% were between family members and one-third were between individuals who communicated daily. Comparatively, kinship and daily communication characterised only 26% and 12% of non-recruitment dyads, respectively. These data contrast previous findings that reported that a majority of RDS recruitment ties were between friends or acquaintances<sup>15253435</sup> and few were between kin.<sup>13152534</sup>

The preferential recruitment of kin in this region may be explained by close interpersonal bonds that characterise many Appalachian families<sup>36</sup>; in fact, some consider the family the "central unit of rural Appalachian social organisation" (<sup>37</sup> p 24). Previous research suggests that the influence of kinship structures be considered during healthcare delivery among Appalachians<sup>3839</sup>; the present study extends this recommendation by underscoring the importance of considering familialism in the design of epidemiological studies among high-risk groups in the region. This finding is especially important given evidence of HIV and HCV risk behavior<sup>40</sup> and dense risk network structures<sup>28</sup> among drug users in Appalachian Kentucky.

This study significantly contributes to the methodological literature about RDS, but limitations warrant mention. First, data were only available on those enrolled in the study, precluding our ability to draw inferences about individuals who were not referred or did not enrol. For example, because participants were compensated for redeemed referrals, it is possible that recruiters preferentially referred those who they perceived to be *able* to participate (eg, residing near or having transportation to the study office). Future studies would benefit from a follow-up survey about logistics and preferences surrounding recruitment.<sup>17</sup> Second, use of travel distance may not have fully captured all factors affecting proximity, routes and travel time (ie, speed limits, construction, etc.). Road distances were calculated based on the shortest possible distance, but participants may have used alternate routes. Of note, the road distance measure was significantly associated with self-reported residential proximity (reported on a 9-point Likert scale). The use of residences as reference points for distance computation is also limited in that it does not account for interaction in other locations, including places where participants socialise and/or use/purchase drugs. Finally, while this study is the first to our knowledge to explore spatial and interpersonal influences on RDS recruitment preferences in a rural, high-risk population, the generalisability may be limited and further research is needed to examine whether similar trends occur among urban populations.

Despite limitations, this study has important implications for the planning, implementation and analysis of RDS studies. We identified recruitment preferences that suggest that recruitment from one's network may be non-random. Instead, individuals appeared to selectively recruit kin and individuals with whom they interacted more frequently. While the current analysis did not aim to estimate population prevalence of disease, the recruitment preferences revealed in this analysis could be used to adjust estimates or to weight individuals by probability of being sampled. Future studies using RDS should conduct formative research to elucidate potential recruitment preferences and should take a similar approach to quantifying preferences (both in terms of interpersonal characteristics and spatial dynamics) for use in analysis.

## Acknowledgments

### Funding

This research was supported by the National Institute on Drug Abuse Grant R01DA024598 (to JRH). AER was funded by the National Institute on Drug Abuse (NIH/NIDA: K01DA033879).

## References

1. Heckathorn DD, Semaan S, Broadhead R, et al. Extensions of respondent-driven sampling: a new approach to the study of injection drug users aged 18–25. *AIDS Behav.* 2002; 6:55–67.
2. Heckathorn D. Respondent-driven sampling: a new approach to the study of hidden populations. *Soc Probl.* 1997; 44:174–99.
3. Heckathorn D. Respondent-driven sampling II: deriving valid population estimates from chain-referral samples of hidden populations. *Soc Probl.* 2002; 49:11–34.
4. Magnani R, Sabin K, Saidel T, et al. Review of sampling hard-to-reach and hidden populations for HIV surveillance. *AIDS.* 2005; 19(Suppl 2):S67–72. [PubMed: 15930843]



5. Gallagher K, Sullivan P, Lansky A, et al. Behavioral surveillance among people at risk for HIV infection in the U.S. : the national HIV behavioral surveillance system. *Public Health Rep.* 2007; 122(Suppl 1):32–8. [PubMed: 17354525]
6. Lansky A, Abdul-Quader AS, Cribbin M, et al. Developing an HIV behavioral surveillance system for injecting drug users: the national HIV behavioral surveillance system. *Public Health Rep.* 2007; 122(Suppl 1):48. [PubMed: 17354527]
7. Malekinejad M, Johnston LG, Kendall C, et al. Using respondent-driven sampling methodology for HIV biological and behavioral surveillance in international settings: a systematic review. *AIDS Behav.* 2008; 12(Suppl 1):S105–S30. [PubMed: 18561018]
8. Lu, X. Respondent-driven sampling: theory, limitations & improvements. Karolinska Institutet; 2013.
9. Goel S, Salganik MJ. Assessing respondent-driven sampling. *Proc Natl Acad Sci USA.* 2010; 107:6743–47. [PubMed: 20351258]
10. Wylie JL, Jolly AM. Understanding recruitment: outcomes associated with alternate methods for seed selection in respondent driven sampling. *BMC Med Res Methodol.* 2013; 13:93. [PubMed: 23865487]
11. Salganik MJ. Variance estimation, design effects, and sample size calculations for respondent-driven sampling. *J Urban Health.* 2006; 83(Suppl 6):i98–112. [PubMed: 16937083]
12. Wejnert C. An empirical test of respondent-driven sampling: point estimates, variance, degree measures, and out-of-equilibrium data. *Sociol Methodol.* 2009; 39:73–116. [PubMed: 20161130]
13. Wang J, Carlson RG, Falck RS, et al. Respondent-driven sampling to recruit MDMA users: a methodological assessment. *Drug Alcohol Depend.* 2005; 78:147–57. [PubMed: 15845318]
14. Liu H, Li J, Ha T. Assessment of random recruitment assumption in respondent-driven sampling in egocentric network data. *Soc Netw.* 2012; 1:13–21. [PubMed: 23641317]
15. Rudolph AE, Crawford ND, Latkin C, et al. Subpopulations of illicit drug users reached by targeted street outreach and respondent-driven sampling strategies: implications for research and public health practice. *Ann Epidemiol.* 2011; 21:280–89. [PubMed: 21376275]
16. McCreesh N, Frost SDW, Seeley J, et al. Evaluation of respondent-driven sampling. *Epidemiology.* 2012; 23:138. [PubMed: 22157309]
17. Rudolph AE, Fuller CM, Latkin C. The importance of measuring and accounting for potential biases in respondent-driven samples. *AIDS Behav.* 2013; 17:2244–52. [PubMed: 23515641]
18. Kral A, Malekinejad M, Vaudrey J, et al. Comparing respondent-driven sampling and targeted sampling methods of recruiting injection drug users in San Francisco. *J Urban Health.* 2010; 87:839–50. [PubMed: 20582573]
19. Toledo L, Codeço CT, Bertoni N, et al. Putting respondent-driven sampling on the map: insights from Rio de Janeiro, Brazil. *J Acquir Immune Defic Syndr.* 2011; 57 (Suppl 3):S136–S43. [PubMed: 21857309]
20. Abdul-Quader AS, Heckathorn DD, McKnight C, et al. Effectiveness of respondent-driven sampling for recruiting drug users in New York City: findings from a pilot study. *J Urban Health.* 2006; 83:459–76. [PubMed: 16739048]
21. Burt RD, Hagan H, Sabin K, et al. Evaluating respondent-driven sampling in a major metropolitan area: Comparing injection drug users in the 2005 Seattle area national HIV behavioral surveillance system survey with participants in the RAVEN and Kiwi studies. *Ann Epidemiol.* 2010; 20:159–67. [PubMed: 20123167]
22. McCreesh N, Johnston LG, Copas A, et al. Evaluation of the role of location and distance in recruitment in respondent-driven sampling. *Int J Health Geogr.* 2011; 10:56–56. [PubMed: 22008416]
23. Qiu P, Yang Y, Ma X, et al. Respondent-driven sampling to recruit in-country migrant workers in China: a methodological assessment. *Scand J Public Health.* 2012; 40:92–101. [PubMed: 21948979]
24. Jenness SM, Neaigus A, Wendel T, et al. Spatial recruitment bias in respondent-driven sampling: implications for HIV prevalence estimation in urban heterosexuals. *AIDS Behav.* Published Online First: 13 Oct 2013. 10.1007/s10461-013-0640-8

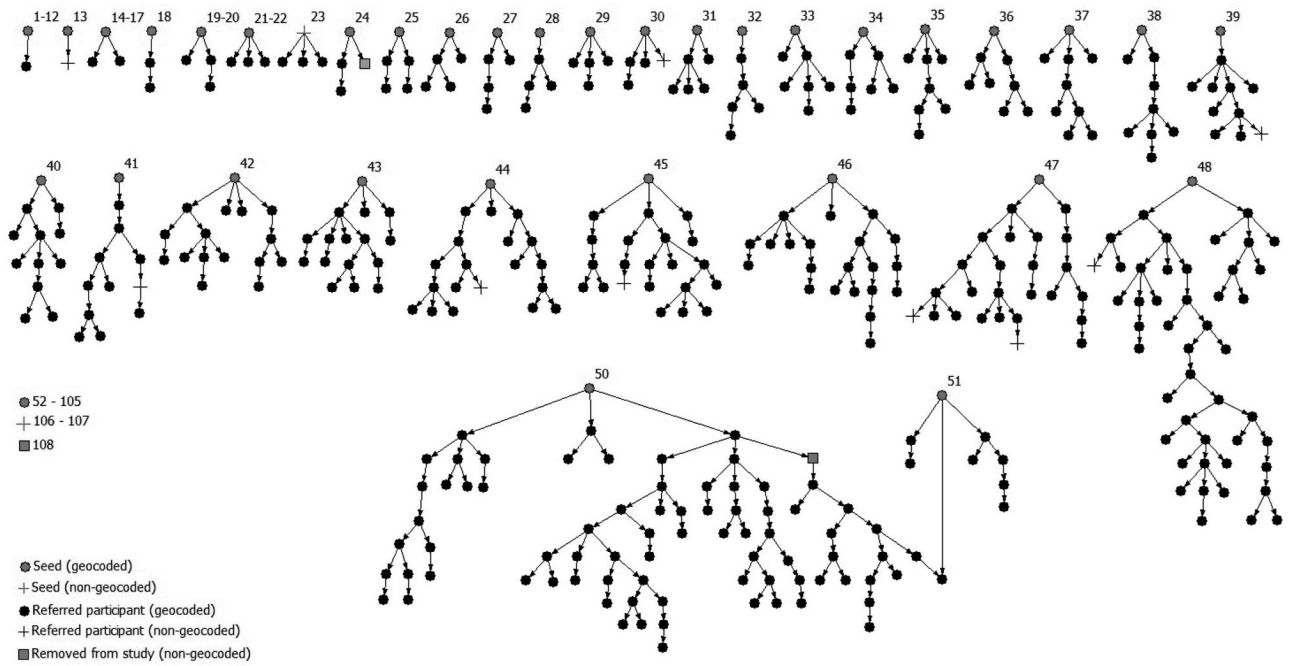
25. Wang J, Falck RS, Li L, et al. Respondent-driven sampling in the recruitment of illicit stimulant drug users in a rural setting: Findings and technical issues. *Addict Behav.* 2007; 32:924–37. [PubMed: 16901654]
26. Jonas A, Young AM, Oser CB, et al. OxyContin as currency: OxyContin use and increased social capital among rural Appalachian drug users. *Soc Sci Med.* 2012; 74:1602–9. [PubMed: 22465379]
27. Environmental Systems Research Institute (ESRI). North America Detailed Streets. Secondary North America Detailed Streets. Jul 23.2013 2007 <http://www.arcgis.com/home/item.html?id=f38b87cc295541fb88513d1ed7cec9fd>.
28. Young AM, Jonas AB, Mullins UL, et al. Network structure and the risk for HIV transmission among rural drug users. *AIDS Behav.* 2012; 17:2341–51. [PubMed: 23184464]
29. Friedman SR, Neaigus A, Jose B, et al. Sociometric risk networks and risk for HIV infection. *Am J Public Health.* 1997; 87:1289–96. [PubMed: 9279263]
30. Hanneman, R.; Riddle, M. Introduction to social network methods. Riverside, CA: University of California, Riverside; 2005.
31. Borgatti, S.; Everett, M.; Freeman, L. Ucinet for windows: software for social network analysis. Harvard, MA: Analytic Technologies; 2002.
32. Krackhardt D. Predicting with networks: Nonparametric multiple-regression analysis of dyadic data. *Social Netwks.* 1987; 10:359–81.
33. SAS Institute. The GLIMMIX procedure. SAS/STAT 93 user's guide. Cary, NC: SAS Institute; 2011.
34. Ramirez-Valles J, Heckathorn DD, Vazquez R, et al. From networks to populations: the development and application of respondent-driven sampling among IDUs and Latino gay men. *AIDS Behav.* 2005; 9:387–402. [PubMed: 16235135]
35. Frost SD, Brouwer KC, Firestone Cruz MA, et al. Respondent-driven sampling of injection drug users in two U.S-Mexico border cities: recruitment dynamics and impact on estimates of HIV and syphilis prevalence. *J Urban Health.* 2006; 83(6 Suppl):i83–97. [PubMed: 17072761]
36. Hicks, GL.; Ergood, B.; Kuhre, BE. Kinship roles and sex roles. In: Ergood, B.; Kuhre, BE., editors. *Appalachia: social context past and present.* 5. Dubuque, IA: Kendall/Hunt Publishing Company; 1976. p. 210
37. Keefe, S. Appalachian family ties. In: Keefe, S., editor. *Appalachian mental health.* Lexington, KY: University of Kentucky Press; 1988.
38. Hansen MM, Resick LK. Health beliefs, health care, and rural Appalachian subcultures from an ethnographic perspective. *Fam Community Health.* 1990; 13:1–10.
39. Keller SM, Helton LR. Culturally competent approaches for counseling urban appalachian clients: an exploratory case study. *J Soc Serv Res.* 2010; 36:142–50.
40. Havens J, Lofwall M, Frost S, et al. Individual and network factors associated with prevalent hepatitis C infection among rural appalachian injection drug users. *Am J Public Health.* 2013; 103:44–52.

**What is already known on this subject**

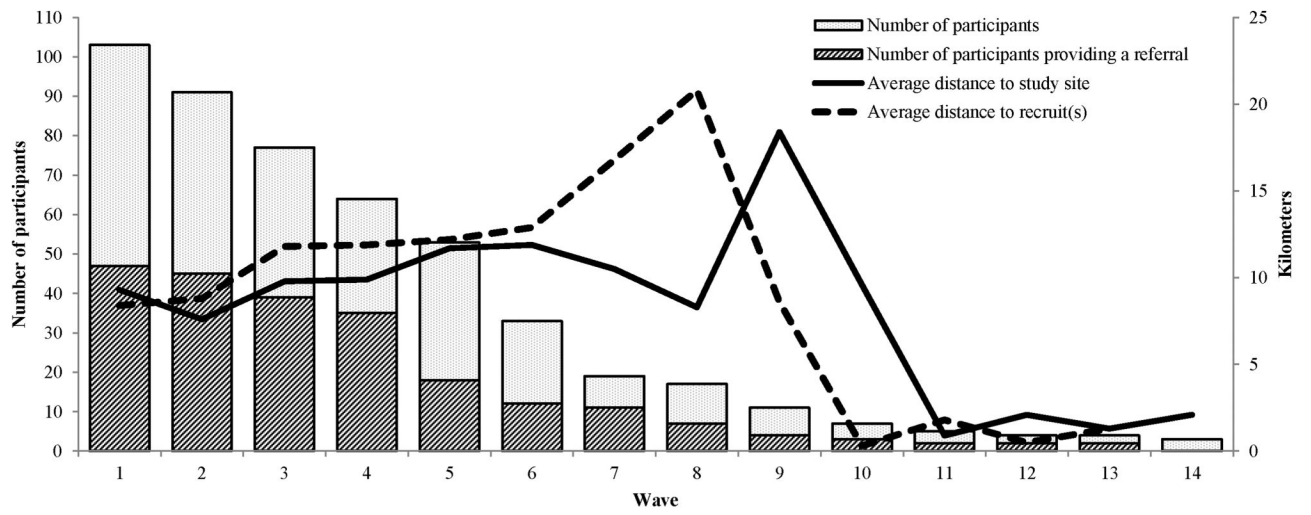
Respondent-driven sampling (RDS) is a modified version of chain-referral sampling designed to recruit 'hidden' populations such as people who use drugs and as an analytic tool used to generate unbiased population estimates. RDS has a demonstrated ability to generate large samples of key populations for HIV research worldwide. However, the accuracy of the RDS estimator depends on several assumptions that have been increasingly scrutinised, particularly the proposition that RDS participants *randomly* recruit peers from their social networks. Yet, few studies to date have directly explored the influence of relationship-level characteristics, including geospatial proximity, on RDS recruitment preferences.

**What this study adds**

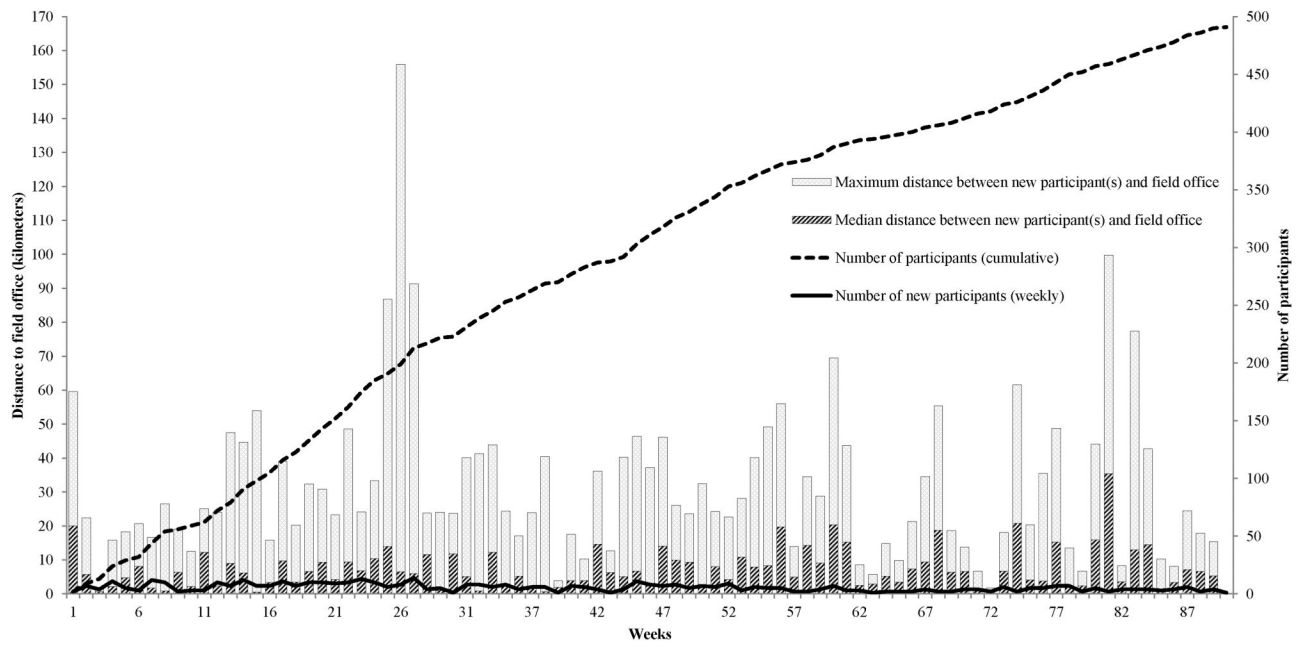
To our knowledge, this study was the first to explore the influence of relationship-level characteristics and geospatial proximity on the RDS recruitment preferences of high-risk individuals residing in a rural area of the USA. In this sample of drug users, spatial proximity was not associated with RDS recruitment. Contrary to research in other settings, kinship was significantly associated with recruitment, controlling for frequency of communication, relationship duration, trust and other measures of similarity. These findings highlight the value of considering the sociological context when planning RDS recruitment and highlight a need to quantify recruitment preferences for use in analyses.



**Figure 1.** Referral chains in respondent-driven sampling of 506 rural drug users.



**Figure 2.**  
Recruitment characteristics by wave of respondent-driven sampling.



**Figure 3.** Cumulative and weekly recruitment trends by participants' distance to study office site.



**Table 1**

Example of data structure for analysis

Ego	Alter	RDS recruitment	Distance (km)	Covariate 1	Covariate 2	...	
1	A	1	5	1	8		
1	B	1	22	0	4		
1	C	0	3	0	3		
2	D	0	33	1	6		
3	E	0	4	0	5		
3	C	1	3	1	2		
3	B	0	Excluded from analysis				

In this example, ego 1 named three alters as network members, but only referred alters A and B into the study. The third alter (alter C) entered the study through referral from ego 3. Ego 2 was a non-generative seed, but *did* name a network member (alter D); therefore, ego 2 remains in the analysis because he/she is part of a network dyad. Ego 3 named alter B, but the dyad was not included in the analysis because alter B was not eligible for recruitment (ie, alter B had already been recruited by ego 1).

RDS, respondent-driven sampling.

**Table 2**

Bivariate and multivariate analyses of respondent-driven sampling referral

Dyadic characteristic	OR (95% CI)	p Value	AOR (95% CI)	p Value
Distance between residences (km)	0.99 (0.98 to 1.01)	0.289		
Transportation access	1.26 (0.89 to 1.78)	0.190		
Demographic similarity				
Gender	0.70 (0.49 to 1.00)	0.047*	0.88 (0.54 to 1.41)	0.584
Age (absolute difference, years)	0.98 (0.95 to 1.01)	0.153		
Race	1.35 (0.73 to 2.51)	0.337		
Education (absolute difference, months)	0.94 (0.85 to 1.05)	0.275		
Similarity in drug use				
Heroin	1.07 (0.68 to 1.69)	0.760		
Crack/cocaine	0.65 (0.33 to 1.28)	0.212		
Sedatives/barbiturates	0.66 (0.36 to 1.21)	0.178		
Methamphetamine	0.98 (0.67 to 1.43)	0.901		
Illicit methadone	0.97 (0.54 to 1.72)	0.909		
Oxycodone	1.45 (0.72 to 2.89)	0.299		
Oxycontin	1.14 (0.63 to 2.07)	0.668		
Hydrocodone	0.64 (0.23 to 1.76)	0.386		
Relationship characteristics				
Exchange financial support	2.96 (1.99 to 4.40)	<0.001*	1.35 (0.75 to 2.43)	0.311
Kinship	2.07 (1.45 to 2.95)	<0.001*	1.62 (1.02 to 2.58)	0.042*
Frequency of communication	1.65 (1.48 to 1.85)	<0.001*	1.63 (1.25 to 2.13)	<0.001*
Duration of relationship (months)	1.03 (1.01 to 1.06)	0.002*	1.00 (1.00 to 1.00)	0.444
Trust	1.24 (1.17 to 1.33)	<0.001*	0.94 (0.84 to 1.06)	0.315
Share drugs	1.81 (1.53 to 2.14)	<0.001*	1.16 (0.90 to 1.49)	0.266
Network relationships				
Drug	2.86 (1.00 to 8.19)	0.051		
Sex	1.60 (1.10 to 2.34)	0.014*	0.68 (0.37 to 1.27)	0.223
Social support	1.16 (0.77 to 1.74)	0.470		

\* p&lt;0.05.

AOR, adjusted OR.