Original Article

An Extensive and Comprehensive Review on RDS: Techniques, Applications and Problems

Mihir Bhatta[1], Agniva Majumdar[2], Piyali Ghosh[3], Debjit Chakraborty[4], Shanta Dutta[5]

How to cite this article:

Mihir Bhatta, Agniva Majumdar, Piyali Ghosh, *et al*. An Extensive and Comprehensive Review on RDS: Techniques, Applications and Problems. Ind J Biol 2024; 11(1):7-13.

Abstract

Respondent Driven Sampling (RDS) is becoming a widely used method to sample hard-to-reach populations, especially in Public Health and Social Science research. The purpose of this systematic review is to present an overall view of RDS methodology, its applications in different areas as well as the difficulties faced when implementing it. A search of electronic databases was conducted systematically to find relevant studies for synthesis that would provide insight into the strengths, limitations and future directions of RDS.

Keywords: Respondent-Driven Sampling; RDS Methodology; Hidden Populations; Epidemiology; Social Science; Sampling Bias; Network Analysis.

INTRODUCTION

Respondent-Driven Sampling (RDS) refers to a sampling strategy which can be used in cases where hidden or hard-to-reach populations are involved.[1] Its use has grown from sociology to several other fields such as public health, epidemiology and social science research because it provides population characteristics that are not

Author's Affiliation: [1-3]Project Scientist-B, [2]Scientist-C, Department of VRDL Lab, [4]Scientist-E, Department of Epidemiology, [5]Scientist-G and Director, ICMR-National Institute of Cholera and Enteric Diseases, Beliaghata, Kolkata 700010, West Bengal, India.
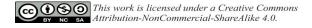
Corresponding Author: Mihir Bhatta, Project Scientist-B, ICMR-National Institute of Cholera and Enteric Diseases, Beliaghata, Kolkata 700010, West Bengal, India.

E-mail: mihirbhatta@gmail.com

biased. Present review will aim at reviewing the methodology of RDS critically, its applications in research and challenges accompanying its implementation around the world. Respondent-Driven Sampling (RDS) emerged in the late 1990s as a novel sampling methodology designed to overcome the challenges of sampling hidden or hard-to-reach populations. Developed by Douglas D. Heckathorn, RDS was introduced as a systematic approach to studying populations for which traditional sampling methods were ineffective or impractical.[2] The history of RDS is characterized by its evolution from conceptualization to widespread adoption across various disciplines.

MATERIALS AND METHODS

A systematic search was conducted across academic databases including PubMed, Web of Science, and Google Scholar. Keywords such as "respondent driven sampling", "RDS methodology", and "RDS applications" were used to identify

relevant published articles. Articles were screened based on inclusion criteria focusing on RDS methodology, applications, and challenges.[3] These published articles were then regorusly reviewed to find out the key findings.

Key Findings: Methodology of Respondent Driven Sampling (RDS)

Respondent Driven Sampling (RDS) is a sampling method designed to study hidden or hard-to-reach populations by leveraging social networks within these populations. RDS involves a systematic process of participant recruitment, whereby initial participants (seeds) recruit their peers, who in turn recruit additional participants, creating a chain referral sampling approach.[4] This methodology aims to overcome the limitations of traditional sampling methods and provide unbiased estimates of population characteristics. Below, we outline the key steps and principles of RDS methodology:

Selection of Seeds: RDS begins with the selection of a small number of initial participants, known as seeds. Seeds should be well-connected within the target population and possess diverse characteristics to ensure the representativeness of the sample.[5] The selection of seeds is critical, as they initiate the recruitment process and influence the composition of the sample.

Recruitment of Participants: Seeds recruit their peers from within their social networks to participate in the study. Each recruited participant, known as a "wave 1" participant, is given a limited number of coupons or referral cards to distribute to their peers.[6] Participants are incentivized for successful recruitment, typically through monetary compensation or non-monetary incentives.

Snowball Effect: As recruitment progresses, participants recruit their peers, who in turn recruit additional participants, creating a snowball effect. This chain-referral process continues until the desired sample size is reached, resulting in a network based sample.[7]

Monitoring and Tracking Recruitments: Throughout the recruitment process, researchers monitor and track the flow of participants through the social network. Data on the number of coupons distributed, redeemed, and remaining are collected to assess the progress of recruitment and adjust sampling efforts as needed.[8]

Dual Incentive System: RDS employs a dual incentive system to motivate participation and ensure the integrity of the sampling process. Participants receive incentives for both their own participation and for successfully referring their peers to the study.[9]

Weighting Adjustments: To account for biases inherent in the sampling process, weighting adjustments are applied to the collected data. Weighting adjustments correct for differences in individuals' network sizes and recruitment probabilities, ensuring that estimates are representative of the target population.

Statistical Analysis: Statistical methods specific to RDS are employed to analyze the collected data and derive population estimates. Commonly used estimators include the RDS-II estimator, which incorporates weighting adjustments, and bootstrapping techniques to assess the uncertainty of estimates.[10]

Convergence and Sample Adequacy: Convergence is assessed to determine when recruitment has reached equilibrium and the sample is adequately representative of the target population. Adequacy of the sample is evaluated based on criteria such as network size, recruitment diversity, and convergence of key characteristics.[11]

Ethical Considerations: Ethical considerations are paramount in RDS studies, particularly concerning participant confidentiality, informed consent, and protection of privacy. Researchers must adhere to ethical guidelines and obtain approval from institutional review boards to ensure the welfare and rights of participants.[12]

Applications of Respondent-Driven Sampling (RDS)

Respondent-Driven Sampling (RDS) has diverse applications across public health, social science, and market research domains. By providing a systematic approach to sampling hidden populations, RDS enables researchers to generate valuable insights into health disparities, social determinants of health, consumer behavior, and political attitudes. RDS studies contribute to evidence based interventions, policies, and programs aimed at addressing the needs of marginalized and underserved populations. By leveraging social networks within these populations, RDS offers a unique approach to data collection that overcomes many of the limitations of traditional sampling methods.[13] The key applications of RDS in research are as follows:

Public Health Research:

HIV/AIDS and STI Surveillance: RDS is extensively used in epidemiological studies to estimate the prevalence of HIV/AIDS, sexually transmitted infections (STIs), and other infectious

Mihir Bhatta, Agniva Majumdar, Piyali Ghosh, *et al*. An Extensive and Comprehensive Review on
RDS: Techniques, Applications and Problems

9

diseases among high-risk populations such as men who have sex with men (MSM), people who inject drugs (PWID), and sex workers. These estimates are crucial for monitoring disease trends, assessing the effectiveness of prevention programs, and allocating resources for targeted interventions.[14]

Substance Use and Harm Reduction: RDS is employed to study substance use patterns, prevalence of drug use disorders, and access to harm reduction services among marginalized populations, including PWID and individuals experiencing homelessness. These studies inform the development of harm reduction strategies, needle exchange programs, and substance use treatment services.[15]

Maternal and Child Health: RDS is utilized to study maternal and child health outcomes among vulnerable populations, such as pregnant women living with HIV, undocumented immigrants, and refugees. These studies assess access to prenatal care, maternal health services, and pediatric healthcare interventions, contributing to efforts to reduce disparities in maternal and child health outcomes.[12]

Social Science Research:

Migration and Mobility: RDS is employed to study migration patterns, social networks, and health outcomes among migrant and mobile populations, including undocumented immigrants, refugees, and temporary workers. These studies explore the social determinants of health, access to healthcare services, and experiences of discrimination and marginalization among migrant populations.[16]

Sexual and Gender Minority Health: RDS is utilized to study the health needs, experiences, and disparities among sexual and gender minority populations, including LGBTQ+ individuals and gender nonconforming individuals. These studies examine access to healthcare, prevalence of mental health disorders, and experiences of stigma and discrimination, informing policies and programs to promote LGBTQ+ health equity.[17]

Criminal Justice and Incarceration: RDS is employed to study the health and social needs of individuals involved in the criminal justice system, including current and former inmates, justice involved youth, and individuals on probation or parole. These studies assess access to healthcare services, prevalence of substance use disorders, and barriers to reintegration into society post release.[18]

Market Research

Consumer Behavior and Market Segmentation: RDS is utilized in market research to study consumer behavior, preferences, and purchasing patterns within hard-to-reach populations, such as niche markets or subcultures. These studies provide insights into consumer motivations, brand loyalty, and product preferences, guiding marketing strategies and product development efforts.[19]

Opinion Polling and Political Campaigns: RDS can be employed in opinion polling and political campaigns to survey hard-to-reach populations, such as minority voters, young adults, and low-income communities. These studies assess public opinion, political attitudes, and voting behavior, informing campaign strategies and policy advocacy efforts.[20]

Challenges in Respondent-Driven Sampling (RDS) Implementation

Respondent-Driven Sampling (RDS) offers a promising approach to sampling hidden or hard-to-reach populations, but its implementation presents several challenges that researchers must navigate. These challenges can impact the validity, reliability, and generalizability of study findings.[16] The key challenges in RDS implementation are depicted as:

Seed Selection Bias

Challenge: The selection of initial participants (seeds) can introduce bias if seeds do not adequately represent the diversity of the target population. Biased seed selection may lead to underrepresentation or overrepresentation of certain subgroups in the sample.[21]

Mitigation: Researchers should employ strategies to ensure diverse seed selection, such as purposive sampling based on key demographic or network characteristics. Additionally, researchers may consider recruiting seeds through community-based organizations or key informants to enhance representativeness.[22]

Recruitment Heterogeneity

Challenge: Variability in participants' social networks, recruitment patterns, and willingness to participate can impact the efficiency and representativeness of RDS. Certain subgroups within the population may be more connected or active in recruitment, leading to sampling biases.

Mitigation: Researchers should monitor recruitment progress and assess recruitment diversity to identify potential biases. Strategies to enhance recruitment heterogeneity may include increasing the number of recruitment waves, implementing targeted recruitment efforts, and offering incentives for participation and recruitment.[23]

Network Assumptions

Challenge: RDS relies on the assumption of a connected social network where individuals know each other and can effectively recruit one another. However, in some contexts, social networks may be fragmented, decentralized, or difficult to access, challenging the feasibility and validity of RDS.[24]

Mitigation: Researchers should conduct formative research to assess the structure and connectivity of social networks within the target population. Adaptations to the RDS methodology, such as increasing the number of seeds or implementing chain-referral reminders, may be necessary to overcome network limitations.[25]

Sample Size and Convergence:

Challenge: Achieving adequate sample size and convergence in RDS studies can be challenging, particularly in populations with small network sizes or low recruitment efficiency. Failure to reach convergence may compromise the validity and generalizability of study findings.

Mitigation: Researchers should carefully consider sample size calculations and recruitment targets based on population size, network connectivity, and expected design effects. Monitoring recruitment progress and implementing strategies to enhance recruitment efficiency, such as increasing incentives or expanding recruitment chains, can help facilitate convergence.[26]

Statistical Complexity:

Challenge: Correcting for biases and estimating population parameters through RDS can be statistically complex, requiring specialized weighting adjustments and estimation techniques. Inadequate understanding or application of RDS specific statistical methods may lead to biased or unreliable estimates.

Mitigation: Researchers should seek expert guidance or consultation from statisticians with experience in RDS methodology. Utilizing software programs designed for RDS analysis and conducting sensitivity analyses to assess the robustness of estimates can help ensure the validity

and reliability of study findings.[27]

Ethical Considerations

Challenge: RDS raises ethical concerns related to participant confidentiality, privacy, and informed consent. Participants may be reluctant to disclose sensitive information or refer their peers due to fear of stigma, discrimination, or legal repercussions.

Mitigation: Researchers should prioritize participant confidentiality and privacy by implementing strict data security measures, anonymizing data collection instruments, and obtaining informed consent from participants. Building trust and rapport with the target population through community engagement and partnership with local organizations can also facilitate ethical recruitment and data collection practices.[25]

DISCUSSION

Respondent-Driven Sampling (RDS) has proven to be a valuable tool for sampling hidden or hard-to-reach populations and generating population estimates. However, like any research methodology, RDS has its limitations. While RDS aims to achieve a form of random sampling through peer recruitment, it does not guarantee a truly random sample. The initial selection of seeds and the recruitment process may introduce biases, particularly if certain subgroups within the population are over or under represented. RDS relies heavily on the social networks of participants for recruitment. If the social networks are fragmented or if there are subgroups within the population that are disconnected from the main network, RDS may fail to reach certain segments of the population. RDS assumes that every individual in the population has a nonzero probability of being connected to every other individual. However, in reality, social networks may be fragmented, leading to limitations in the reach and representativeness of the sample.[28] RDS assumes that the recruitment process reaches equilibrium, meaning that recruitment patterns stabilize, and the composition of the sample reflects the underlying population distribution. Achieving equilibrium can be challenging, particularly in populations with high levels of mobility or turnover.

The findings from RDS studies may not be generalizable to the broader population due to the non-random nature of sampling and the reliance on

social networks for recruitment. While RDS provides valuable insights into hidden populations, caution should be exercised when extrapolating findings to other contexts. The effectiveness of RDS relies on participants' willingness to recruit their peers, which may be influenced by the incentive structure. If incentives are not perceived as sufficient or if there are concerns about confidentiality or safety, recruitment efforts may be compromised. RDS raises ethical considerations related to participant confidentiality, informed consent, and privacy. The use of peer recruitment may inadvertently disclose sensitive information about participants' social networks or behaviors, raising concerns about privacy and confidentiality. Analyzing data from RDS studies requires specialized statistical methods to account for the complex sampling design and biases inherent in the recruitment process. Researchers must carefully consider weighting adjustments and other statistical techniques to ensure the validity and reliability of estimates.[27]

Despite these limitations, RDS remains a valuable methodology for studying hidden populations and generating population estimates when traditional sampling methods are impractical or ineffective. By understanding and addressing these limitations, researchers can maximize the validity and utility of RDS studies in informing public health interventions, policy development, and social science research. Future research should focus on refining RDS methodology, exploring alternative sampling approaches, and addressing emerging issues in sampling hard-to-reach populations.[29]

CONCLUSION

Respondent-Driven Sampling (RDS) offers a valuable approach for studying hidden populations and estimating population parameters in public health, social science, and market research. The future of Respondent-Driven Sampling (RDS) holds exciting possibilities, with ongoing advancements and innovations aimed at addressing current limitations and expanding the methodology's applicability. Future advancements in RDS methodology will focus on addressing existing limitations and improving the validity and reliability of estimates. This includes refining techniques for seed selection, optimizing recruitment strategies, and developing innovative statistical methods to account for biases and uncertainties inherent in the sampling process.[26] The integration of digital technologies, such as mobile apps and social media platforms, presents

opportunities to enhance the efficiency and reach of RDS. Digital platforms can streamline participant recruitment, facilitate data collection, and enable real time monitoring of recruitment dynamics, thereby improving the timeliness and accuracy of RDS studies. Adaptive sampling designs, will allow dynamic adjustments to sampling procedures based on ongoing data collection, offer potential improvements to RDS. By incorporating feedback mechanisms and adaptive strategies, researchers can optimize recruitment efforts, achieve equilibrium more efficiently, and enhance the representativeness of the sample.[30]

Multiplex sampling involves simultaneously sampling multiple populations within a network or community, allowing for the study of interconnected subgroups and their interactions. Future advancements in multiplex sampling techniques will enable researchers to capture the complexity of social networks more comprehensively and explore the dynamics of diverse populations within a single study. Moreover, the integration of RDS with network analysis methods will offer opportunities to gain deeper insights into the structure and dynamics of social networks. By combining RDS data with network modeling techniques, researchers can elucidate patterns of social connectivity, identify influential nodes or clusters, and assess the impact of network structure on health outcomes and behaviors.[24] Future efforts might be focused on fostering global collaboration and standardization in RDS research, enabling cross-country comparisons and harmonization of data collection protocols. By establishing best practices, guidelines, and quality assurance mechanisms, researchers can enhance the reliability and validity of RDS estimates and facilitate evidence synthesis across diverse settings and populations.[31]

As RDS continues to evolve, it will be essential to prioritize ethical considerations and meaningful participant engagement. Future advancements will emphasize the importance of respecting participants' autonomy, ensuring confidentiality and privacy, and fostering partnerships with communities to co-create research protocols and interventions that are culturally sensitive and responsive to community needs. While RDS has been instrumental in advancing knowledge in various domains, challenges such as bias, network assumptions, and[32] ethical concerns warrant continued attention. By addressing these challenges and advancing methodological innovations, RDS can continue to serve as a robust tool for sampling hard-to-reach populations and informing

evidence based interventions and policies.[33]

REFERENCES

1. Johnston LG. Introduction to respondent-driven sampling. Geneva: World Health Organization; 2013 http://applications.emro.who.int/dsaf/ EMRPUB_2013_EN_1539.pdf. Accessed 15 Jun 2015.

2. Heckathorn DD. Respondent-driven sampling: a new approach to the study of hidden populations. SocProbl. 1997;44 (2):174–99.

3. Salganik MJ, Heckathorn DD. Sampling and estimation in hidden populations using respondent driven sampling. Sociological methodology. 2004 Dec; 34(1):193-240.

4. Volz E, Heckathorn DD. Probability based estimation theory for respondent driven sampling. J Off Stat. 2008;24 (1):79–97.

5. Simic, M., Johnston, L.G., Platt, L. *et al*. Exploring Barriers to 'Respondent Driven Sampling' in Sex Worker and Drug-Injecting Sex Worker Populations in Eastern Europe. J Urban Health 83, 6–15 (2006).

6. Johnston LG, Corceal S. Unexpectedly high injection drug use, HIV and hepatitis C prevalence among female sex workers in the Republic of Mauritius. AIDS Behav. 2013;17(2):574–84.

7. Salganik MJ, Heckathorn DD. Sampling and estimation in hidden populations using respondent driven sampling. Sociological methodology. 2004 Dec;34(1):193-240.

8. Hladik W, Barker J, Ssenkusu JM, Opio A, Tappero JW, Hakim A, *et al*. HIV infection among men who have sex with men in Kampala, Uganda—a respondent driven sampling survey. PLoS One. 2012;7(5):1–9.

9. McCreesh, N., Copas, A., Seeley, J., Johnston, L.G., Sonnenberg, P., Hayes, R.J., Frost, S.D. and White, R.G., 2013. Respondent driven sampling: determinants of recruitment and a method to improve point estimation. PLoS One, 8(10), p.e78402.

10. Gile, K.J. and Handcock, M.S., 2010. Respondent driven sampling: an assessment of current methodology. Sociological methodology, 40(1), pp.285-327.

11. Heckathorn, D.D., 2007. Extensions of respondent driven sampling: analyzing continuous variables and controlling for differential recruitment. Sociological methodology, 37(1), pp.151-207.

12. Fisher, J.C. and Merli, M.G., 2014. Stickiness of respondent-driven sampling recruitment chains. Network Science, 2(2), pp.298-301.

13. Wejnert, C. and Heckathorn, D., 2011. Respondent-driven sampling: operational procedures, evolution of estimators, and topics for future research. The SAGE handbook of innovation in social research methods.

14. Shi, Y., Cameron, C.J. and Heckathorn, D.D., 2019. Model-based and design-based inference: reducing bias due to differential recruitment in respondent-driven sampling. Sociological Methods & Research, 48(1), pp.3-33.

15. Lee, S., Ong, A.R. and Elliott, M., 2020. Exploring mechanisms of recruitment and recruitment cooperation in respondent driven sampling. Journal of official statistics, 36(2), pp.339-360.

16. Wejnert, C., 2009. An empirical test of respondent driven sampling: point estimates, variance, degree measures, and out of equilibrium data. Sociological methodology, 39(1), pp.73-116.

17. Schonlau, M., Weidmer, B. and Kapteyn, A., 2014. Recruiting an Internet panel using respondent-driven sampling. Journal of Official Statistics, 30(2), pp.291-310.

18. Lu, X., 2013. Respondent-driven sampling: theory, limitations & improvements. KarolinskaInstitutet (Sweden).

19. Verdery, A.M., Mouw, T., Bauldry, S. and Mucha, P.J., 2015. Network structure and biased variance estimation in respondent driven sampling. PloS one, 10(12), p.e0145296.

20. Bhatta, M., Majumdar, A., Banerjee, S., Ghosh, P., Biswas, S. and Dutta, S., 2023. Accumulation of Biological and Behavioral Data of Female Sex Workers Using Respondent-Driven Sampling: Protocol for a Systematic Review. JMIR Research Protocols, 12(1), p.e43722.

21. Wylie, J.L. and Jolly, A.M., 2013. Understanding recruitment: outcomes associated with alternate methods for seed selection in respondent driven sampling. BMC Medical Research Methodology, 13, pp.1-11.

22. Truong, H.H.M., Grasso, M., Chen, Y.H., Kellogg, T.A., Robertson, T., Curotto, A., Steward, W.T. and McFarland, W., 2013. Balancing theory and practice in respondent-driven sampling: a case study of innovations developed to overcome recruitment challenges. PloS one, 8(8), p.e70344.

23. Tomas, A. and Gile, K.J., 2011. The effect of differential recruitment, non-response and non-recruitment on estimators for respondent-driven sampling.

24. Lee, S., 2009. Understanding respondent driven sampling from a total survey error perspective. Survey Practice, 2(6).

25. Reisner, S.L., Mimiaga, M.J., Johnson, C.V., Bland, S., Case, P., Safren, S.A. and Mayer, K.H., 2010. What makes a respondent-driven sampling "seed" productive? Example of finding at-risk Massachusetts men who have sex with men. Journal of Urban Health, 87, pp.467-479.

26. Avery, L., Macpherson, A., Flicker, S. and Rotondi, M., 2021. A review of reported network degree and

London: SAGE Publications, Ltd, pp.473-97.

Mihir Bhatta, Agniva Majumdar, Piyali Ghosh, *et al*. An Extensive and Comprehensive Review on
RDS: Techniques, Applications and Problems

13

recruitment characteristics in respondent driven sampling implications for applied researchers and methodologists. Plos one, 16(4), p.e0249074.

27. Spiller, M., 2009. Regression modelling of data collected using respondentdriven sampling.

28. Berchenko, Y., Rosenblatt, J.D. and Frost, S.D., 2017. Modeling and analyzing respondent driven sampling as a counting process. Biometrics, 73(4), pp.1189-1198.

29. Nesterko, S. and Blitzstein, J., 2015. Bias–variance and breadth–depth tradeoffs in respondent-driven sampling. Journal of Statistical Computation and Simulation, 85(1), pp.89-102.

30. Handcock, M.S., Gile, K.J. and Mar, C.M., 2014. Estimating hidden population size using respondent-driven sampling data. Electronic journal of statistics, 8(1), p.1491.

31. Murrill, C.S., Bingham, T., Lauby, J., Liu, K.L., Wheeler, D., Carballo-Diéguez, A., Marks, G. and Millett, G.A., 2016. Respondent-driven sampling in a multi-site study of Black and Latino men who have sex with men. Journal of the National Medical Association, 108(1), pp.69-76.

32. Kuhns, L.M., Kwon, S., Ryan, D.T., Garofalo, R., Phillips, G. and Mustanski, B.S., 2015. Evaluation of respondent-driven sampling in a study of urban young men who have sex with men. Journal of Urban Health, 92, pp.151-167.

33. Kogan, S.M., Wejnert, C., Chen, Y.F., Brody, G.H. and Slater, L.M., 2011. Respondent-driven sampling with hard-to-reach emerging adults: An introduction and case study with rural African Americans. Journal of Adolescent Research, 26(1), pp.30-60.

———◆◆◆❍◆◆◆———