REVISTA ESPAÑOLA DE SALUD PÚBLICA · RESP

Issue
98
12/4/2024

SPECIAL COLLABORATION

Received: 5/2/2024
Accepted: 11/4/2024
Published: 12/4/2024

e202412062

e1-e6

Potenciales sesgos en estudios epidemiológicos basados en el método respondent-driven sampling: comparación según su aplicación presencial y en línea

Potential biases in epidemiological studies using respondent-driven sampling method: a comparison between its face-to-face and online application

AUTHORS

Pedro Ferrer Rosende	**(1,2,3)**	[ORCID: 0000-0002-3791-4696]
Laura Esteve Matalí	**(1,2)**	[ORCID: 0000-0001-6561-0131]
Valeria Stuardo Ávila	**(4)**	[ORCID: 0000-0001-9861-0285]
Mauricio Fuentes Alburquenque	**(5)**	[ORCID: 0000-0001-7040-0040]
Albert Navarro Giné	**(1,2,6)**	[ORCID: 0000-0001-7153-4673]

AFFILIATIONS

(1) Research group on psychosocial risks, work organization and health (POWAH), Autonomous University of Barcelona (UAB). Cerdanyola del Vallès. Spain.
(2) Biostatistics Unit, Faculty of Medicine, Autonomous University of Barcelona (UAB). Cerdanyola del Vallès. Spain.
(3) PhD Program in Biomedical Research Methodology and Public Health, Autonomous University of Barcelona (UAB). Cerdanyola del Vallès. Spain.
(4) Institute of Public Health, Andrés Bello University. Santiago. Chile.
(5) Biostatistics Program, School of Public Health, Faculty of Medicine, University of Chile. Santiago. Chile.
(6) Institute of Labor Studies (IET), Autonomous University of Barcelona (UAB). Cerdanyola del Vallès. Spain.

AUTHORSHIP CONTRIBUTIONS

All authors conceived the study. P Ferrer-Rosende wrote the original manuscript. All authors contributed to editing successive drafts of the manuscript, reviewed them, provided elements for improvement, and accepted the final version.

Authors declare that there is no conflict of interests

FUNDING

P Ferrer-Rosende acknowledges the support from ANID through CHILE SCHOLARSHIPS for doctoral studies in the foreigner.

CORRESPONDENCE

Laura Esteve Matalí
Avinguda Can Domènec, s/n.
Campus UAB. CP 08193.
Cerdanyola del Vallès. España.
laura.esteve@uab.cat

SUGGESTED CITATION

Ferrer Rosende P, Esteve Matalí L, Stuardo Ávila V, Fuentes Alburquenque M, Navarro Giné A. Potential biases in epidemiological studies using respondent-driven sampling method: a comparison between its face-to-face and online application. Rev Esp Salud Pública. 2024; 98: December 4th e202412062.

https://ojs.sanidad.gob.es/index.php/resp

ABSTRACT

In epidemiological and social studies on populations without a sampling frame or that are hard to reach, respondent-driven sampling (RDS), under certain assumptions, has the potential to produce asymptotically unbiased and efficient population estimates for these populations. While RDS is typically conducted face-to-face, the online version (WebRDS) has gained attention due to its potential advantages, although its disadvantages have also raised concern.

The objective of this study was to contrast these two formats at the level of application and potential biases, where the online version offers improved speed and lower costs but raises concerns about potential biases due to a lack of face-to-face instructions regarding definition of the social network size and peer-recruiting process. Both formats can generate unbiased estimates, however, it's crucial to carefully consider potential sources of bias to meet the necessary assumptions. Therefore, it is important to continue researching the most appropriate analytical approaches to address the specific biases of each modality.

KEYWORDS // Surveys and questionnaires; Social studies; Respondent-driven sampling; Bias; Internet.

RESUMEN

En estudios epidemiológicos sobre poblaciones sin un marco muestral o de difícil acceso, el muestreo dirigido por participantes (RDS, *respondent-driven sampling*, por sus siglas en inglés), bajo ciertos supuestos, tiene el potencial de producir estimaciones poblacionales eficientes y asintóticamente insesgadas. Si bien el RDS se realiza generalmente cara a cara, la versión en línea (WebRDS, *web-based respondent-driven sampling*) ha llamado la atención debido a sus posibles ventajas, aunque sus desventajas también han generado preocupación.

El objetivo de este estudio fue contrastar estos dos formatos en términos de aplicación y posibles sesgos, donde la versión en línea ofrece una mayor velocidad y menores costes pero plantea preocupaciones sobre posibles sesgos debido a la falta de instrucciones cara a cara con respecto a la definición del tamaño de la red social y el proceso de reclutamiento de pares. Ambos formatos pueden generar estimaciones insesgadas; sin embargo, es crucial considerar cuidadosamente las posibles fuentes de sesgo para cumplir con los supuestos necesarios. Por ello, es importante seguir investigando los enfoques de análisis más adecuados para abordar los sesgos específicos de cada forma de aplicación.

PALABRAS CLAVE // Encuestas y cuestionarios; Estudios sociales; Muestreo dirigido por participantes; Sesgo; Internet.

BACKGROUND

IN EPIDEMIOLOGICAL AND SOCIAL STUDIES, classical sampling and estimation techniques require a known sampling frame. However, for many populations of interest, this is not feasible due to the small size of the target population, the sensitive nature of the population, the difficulty of distinguishing members from the general population, or the absence of a mechanism to build the sampling frame. These populations are referred to as hidden or hard to reach populations and are challenging to study using traditional methods (1).

To overcome these challenges, several alternative methods have been developed. In the last decades respondent-driven sampling (RDS), a chain-referral method based on network theory, has gained visibility for its potential to generate asymptotically unbiased population estimates if certain assumptions are met (1), and has been widely used in the HIV field, in the study of sex workers, people who inject drugs and men who have sex with men, among other hard-to-reach populations (2).

While RDS is typically conducted face-to-face, the online version (WebRDS) has gained attention due to its potential advantages, although its disadvantages are also rising concern in the methodological debate.

This study contributes to the methodological debate by comparing these two formats at the level of application and potential biases.

RESPONDENT-DRIVEN SAMPLING (RDS)

RESPONDENT-DRIVEN SAMPLING (RDS) IS A chain-referral sampling method where participants are selected from the social network of the target population and not from a sampling frame. To achieve this, the researcher team selects a few initial participants known as *seeds*, who are asked to answer the survey and then recruit a limited number of people from their social network (usually through a coupon system). In addition to the survey, these people are required to answer some questions about the size of their social network. It is crucial that this information is accurately specified, as the precision of the estimates relies on the assumption that the social network size is well-defined. The chain-referral process continues for as many *waves'* are required until the desired sample size is reached, the participant characteristics have stabilized, or the chains go extinct. By tracking recruitment and collecting network information, RDS can generate unbiased population estimates under certain assumptions (3). The process of RDS can be broken down into three key stages (4): formative research, data collection, and data analysis. Each stage presents its own unique challenges and are critical to obtaining accurate population estimates.

Formative research is a crucial part to avoid failures on the implementation of RDS (5). In this phase researchers may address the target population network characteristics, evaluate the acceptability of RDS as a viable sampling method, seed selection and survey logistics such as incentives and coupon design.

The data collection process must be carefully planned and executed by constantly monitoring recruitment chains and estimates of selected traits as waves progress to see if equilibrium is reached (stabilization of some main estimates) while also controlling for homophily (i.e., tendency to have social ties with people of similar characteristics).

For data analysis, there are several proposed estimators for mean and variance and each one has its own assumptions to generate unbiased estimators along with its own limitations, for a complete review go to Abdesselam *et al.* (2)

The main assumptions for the sampling process are (4): (1) relationships between individuals are reciprocal, (2) individuals form a

single network (i.e., each individual can be reached by another individual through social ties), (3) sampling occurs with replacement, (4) respondents can accurately report the size of their personal network (or degree) and (5) recruits are randomly selected within the recruiter's social network.

While RDS has demonstrated potential, it also has practical limitations, including the speed at which data can be gathered, which is contingent upon staff availability and the length of the survey, as well as participants' willingness to recruit others in a period of time [6]. RDS also has time and location-related barriers as it is limited to the area in which the staff operates and where participants can move.

WEB-BASED RESPONDENT-DRIVEN SAMPLING (WEBRDS)

FOR THE REASONS MENTIONED ABOVE, the online version of RDS (WebRDS) is gaining attention over the last decade. WebRDS is being used for a variety of purposes, including generating population estimates, studying social network patterns of infectious diseases, and delivering targeted interventions to specific populations such as university students, men who have sex with men, smokers, and people with precarious employment [7].

Seed selection can be done both online and in person, while email or social media have been used for peer recruitment. However, lately, the use of the phone over computers has encouraged the application through phone-based platforms [4].

WebRDS, like other web-based methods, may face difficulties such as bias from unequal internet access, multiple responses for the same participant, uncertain credibility of online research, and lack of face-to-face interaction [7].

RDS VS WEBRDS

WHEN COMPARING FACE-TO-FACE RDS AND the online version, we can distinguish between issues related to implementation and other sources of bias. The implementation aspect can be broken down into four key elements: instructions, timing, location, and the application of the method. Other potential sources of bias includes adherence, multiple responses, and misreporting of network size.

Instructions. In the face-to-face version it is more likely that the participant understands the sampling method and follows the instructions correctly. On the other hand, online instructions must be very clear to minimize the risk of misunderstandings that could affect the sampling process and estimates. Face-to-face instructions can only be given to seeds.

Time. Conducting one-on-one interviews and then distributing coupons to each recruit is time consuming. The online version allows for reaching a larger number of participants in a shorter timeframe, as multiple individuals can respond simultaneously, and coupons can be delivered promptly.

Location. Face-to-face RDS is constrained by the geographic reach of the recruitment locations, necessitating careful selection of these places during field planning. In contrast, the online version can extend its coverage to a broader area, contingent upon the extent to which the assumption that the population belongs to the same network holds true. However, there is a potential risk of individuals outside the study scope responding if there isn't sufficient control in place.

Application method. The face-to-face version may face limitations based on the participants' ability and willingness to move. Nevertheless, individuals encountering challenges such as old age or illiteracy can receive assis-

tance in responding to the questionnaire. On the other hand, the effectiveness of the online version is dependent on access to the Internet and proficiency in using computers or smartphones.

Adherence (non-response). The time or cost involved for participants to attend the survey in person can impact adherence. With the online version, being able to answer the survey anywhere or at any time can reduce non-response rates.

Multiple answers. With adequate mechanisms to control and prevent participants from answering more than once, there is almost no chance of multiple answer in face-to-face application. For WebRDS there are higher chances of multiple responses, especially when monetary incentives are involved. Special controls are necessary to prevent this **(4)**.

Inaccurate report of network size (degree). The risk of inaccurate reporting can be mitigated through meticulous choices in the number and content of degree-related questions, as suggested by Wright **(8)**. This risk can be further reduced with clear face-to-face instructions during the survey. In the absence of face-to-face instructions, the potential for inaccurate reporting is higher. Following guidelines provided by Gille *et al.* **(9)** when selecting the number and content of degree-related questions is recommended to minimize such risks. Emphasizing the importance of these questions in the initial survey instructions is also crucial.

CONCLUSION

RDS HAS GAINED INTEREST AMONG EPIDEmiologists and social sciences researchers due to its potential to produce unbiased estimators in hard-to-reach populations, however caution must be exercised as some assumptions of the method are unrealistic (such as sampling with replacement).

The main advantages of WebRDS can be observed in the application of the recruitment process where speed and coverage are points clearly in favor of this version. However, the lack of in-person instruction may introduce bias, particularly in the reporting of network size and comprehension of the peer-recruitment process.

On the other hand, WebRDS is more susceptible to multiple answers for the same participant and is restricted to population with internet access, but it may attract more participants that are reluctant to go to an in-person interview.

To minimize the risk of low-quality studies and poor estimators, guidelines with diagnostics for evaluating RDS assumptions have been published **(8)**. Additionally, an adapted version of the *STROBE* guide for observational studies specifically for RDS has been released to improve the reporting **(10)**.

WebRDS has emerged as a cost-effective alternative to implement RDS, which under certain conditions as extensive formative research, good diffusion plan, good seed selection, and a population that have internet access and the motivation to take part of the process, it is possible to obtain a sample with recruitment performance like that of face-to-face RDS **(11)**. However, several challenges remain that must be considered, especially the lack of face-to- face instruction that can cause misreporting of network size and errors in the peer- recruitment process. If these challenges can be overcome, WebRDS emerges as a very good alternative that considerably reduce the effort and costs required to apply this sampling method.

ACKNOWLEDGMENTS

This work has been carried out within the framework of the doctoral program in Biomedical Research Methodology and Public Health at the Autonomous University of Barcelona.

REFERENCES

1. Faugier J, Sargeant M. *Sampling hard to reach populations.* J Adv Nurs [Internet]. 1997;26(4):790-797. Disponible en: http://doi.org/10.1046/j.1365-2648.1997.00371.x

2. Heckathorn DD. *Respondent-driven sampling: A new approach to the study of hidden populations.* Soc Probl [Internet]. 1997; 44 (2): 174-199. Disponible en: http://doi.org/10.2307/3096941

3. Goodman LA. *Snowball Sampling*. Ann Math Stat [Internet]. 1961; 32 (1): 148-170. Disponible en: http://doi.org/10.1214/aoms/1177705148

4. Deaux E, Callaghan JW. *Key informant versus self-report estimates of health-risk behavior.* Eval Rev [Internet]. 1985;9(3):365-368. Disponible en: http://doi.org/10.1177/0193841x8500900308

5. Abdesselam K, Verdery A, Pelude L, Dhami P, Momoli F, Jolly AM. *The development of respondent-driven sampling (RDS) inference: A systematic review of the population mean and variance estimates.* Drug Alcohol Depend [Internet]. 2020; 206 (107702): 107702. Disponible en: http://doi.org/10.1016/j.drugalcdep.2019.107702

6. Salganik MJ, Heckathorn DD. *Sampling and estimation in hidden populations using respondent-driven sampling.* Sociol Methodol [Internet]. 2004; 34 (1): 193-240. Disponible en: http://doi.org/10.1111/j.0081-1750.2004.00152.x

7. Sosenko FL, Bramley G. *Smartphone-based Respondent Driven Sampling (RDS): A methodological advance in surveying small or "hard-to-reach" populations.* PLoS One [Internet]. 2022; 17 (7): e0270673. Disponible en: http://doi.org/10.1371/journal.pone.0270673

8. Johnston LG, Whitehead S, Simic-Lawson M, Kendall C. *Formative research to optimize respondent-driven sampling surveys among hard-to-reach populations in HIV behavioral and biological surveillance: lessons learned from four case studies.* AIDS Care [Internet]. 2010; 22 (6): 784-792. Disponible en: http://doi.org/10.1080/09540120903373557

9. Lena Hipp E *et al. How to implement respondent-driven sampling in practice: Insights from surveying 24-hour migrant home care workers* [Internet]. Survey Methods: Insights from the Field (SMIF); 2019. Disponible en: http://doi.org/10.13094/SMIF-2019-00009

10. Ferrer-Rosende P, Feijoo-Cid M, Fernández-Cano MI, Salas-Nicás S, Stuardo-Ávila V, Navarro-Giné A. *Implementation of web-based respondent driven sampling in epidemiological studies.* BMC Med Res Methodol [Internet]. 2023; 23 (1): 217. Disponible en: http://doi.org/10.1186/s12874-023-02042-z

11. Wejnert C, Heckathorn DD. *Web-based network sampling: Efficiency and efficacy of respondent-driven sampling for online research.* Sociol Methods Res [Internet]. 2008;37(1): 105-134. Disponible en: http://doi.org/10.1177/0049124108318333

12. Helms YB, Hamdiui N, Kretzschmar MEE, Rocha LEC, van Steenbergen JE, Bengtsson L *et al. Applications and recruitment performance of web-based respondent-driven sampling: Scoping review.* J Med Internet Res [Internet]. 2021; 23 (1): e17564. Disponible en: http://doi.org/10.2196/17564

13. Wright KB. *Researching internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services.* J Comput Mediat Commun [Internet]. 2006; 10 (3). Disponible en: http://doi.org/10.1111/j.1083-6101.2005.tb00259.x

14. Gile KJ, Johnston LG, Salganik MJ. *Diagnostics for respondent-driven sampling.* J R Stat Soc Ser A Stat Soc [Internet]. 2015; 178 (1): 241-269. Disponible en: http://doi.org/10.1111/rssa.12059

15. Yauck M, Moodie EE, Apelian H, Fourmigue A, Grace D, Hart T *et al. General regression methods for respondent-driven sampling data.* Stat Methods Med Res [Internet]. 2021; 30 (9): 2105-2118. Disponible en: http://doi.org/10.1177/09622802211032713

16. White RG, Hakim AJ, Salganik MJ, Spiller MW, Johnston LG, Kerr L *et al. Strengthening the Reporting of Observational Studies in Epidemiology for respondent-driven sam-*

pling studies: "STROBE-RDS" statement. J Clin Epidemiol [Internet]. 2015; 68 (12): 1463-1471. Disponible en: http://doi.org/10.1016/j.jclinepi.2015.04.002

17. Hildebrand J, Burns S, Zhao Y, Lobo R, Howat P, Allsop S *et al. Potential and challenges in collecting social and behavioral data on adolescent alcohol norms: Comparing respondent-driven sampling and Web-based respondent-driven sampling.* J Med Internet Res [Internet]. 2015; 17 (12): e285. Disponible en: http://doi.org/10.2196/jmir.4762