

**COLABORACIÓN
 ESPECIAL**

Recibida: 2/5/2024
 Aceptada: 4/11/2024
 Publicada: 4/12/2024
 e202412062es
 e1-e7

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

Los autores declaran que no existe ningún conflicto de intereses

FINANCIACIÓN

P Ferrer-Rosende reconoce el apoyo de ANID a través de BECAS CHILE para estudios de doctorado en el extranjero.

CORRESPONDENCIA

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

CITA SUGERIDA

Ferrer Rosende P, Esteve Matalí L, Stuardo Ávila V, Fuentes Alburquenque M, Navarro Giné A. Potenciales sesgos en estudios epidemiológicos basados en el método *respondent-driven sampling*: comparación según su aplicación presencial y en línea. Rev Esp Salud Pública. 2024; 98: 4 de diciembre e202412062es.

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

AUTORES

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]

CONTRIBUCIONES DE AUTORÍA

Todos los autores concibieron el estudio. P Ferrer-Rosende redactó el manuscrito original. Todos los autores contribuyeron a editar los borradores sucesivos del manuscrito, los revisaron, proporcionaron elementos para su mejora y aceptaron la versión final.

FILIACIONES

- (1) Grupo de investigación en riesgos psicosociales, organización del trabajo y salud (POWAH), *Universitat Autònoma de Barcelona (UAB)*, Cerdanyola del Vallès, España.
- (2) *Unitat de Bioestadística, Facultat de Medicina, Universitat Autònoma de Barcelona (UAB)*, Cerdanyola del Vallès, España.
- (3) Programa de doctorado en Metodología de la Investigación Biomédica y Salud Pública, *Universitat Autònoma de Barcelona (UAB)*, Cerdanyola del Vallès, España.
- (4) Instituto de Salud Pública, Universidad Andrés Bello, Santiago, Chile.
- (5) Programa de Bioestadística, Escuela de Salud Pública, Facultad de Medicina, Universidad de Chile, Santiago, Chile.
- (6) Instituto de Estudios del Trabajo (IET), *Universitat Autònoma de Barcelona (UAB)*, Cerdanyola del Vallès, España.

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.

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.

INTRODUCCIÓN

EN ESTUDIOS EPIDEMIOLÓGICOS, LAS TÉCNICAS clásicas de muestreo y estimación requieren un marco muestral conocido, es decir, un conjunto accesible y completo de todos los elementos de la población objetivo desde el cual se pueda seleccionar una muestra. Sin embargo, para muchas poblaciones de interés, esto no es factible debido al tamaño reducido de la población objetivo, la naturaleza sensible de la población, la dificultad para distinguir a sus miembros o la ausencia de un mecanismo para generar un marco de muestreo. Estas poblaciones se conocen como poblaciones ocultas o de difícil acceso (*hard-to-reach*) y su estudio es desafiante utilizando métodos tradicionales (1-4).

Para superar estos desafíos, a lo largo de los años se han desarrollado diversos métodos alternativos como el muestreo en bola de nieve (3) o el muestreo de informantes clave (4) pero, a pesar de ser los métodos dominantes para el estudio de las poblaciones de difícil acceso, se han descrito importantes limitaciones con respecto a la inferencia que se realiza a partir de estos medios de muestreo (2).

En respuesta a estas limitaciones, en las últimas décadas ha ganado visibilidad el muestreo dirigido por participantes (RDS, *respondent-driven sampling*, por sus siglas en inglés), un método de referencia en cadena basado en teoría de redes que puede producir muestras independientes de los sujetos iniciales y tiene el potencial para generar estimaciones poblacionales asintóticamente insesgadas, siendo ampliamente utilizado en el campo del VIH, en el estudio de trabajadores/as sexuales, de personas que se inyectan drogas y de hombres que tienen sexo con hombres, entre otras poblaciones de difícil acceso (2,5).

Si bien el RDS se realiza típicamente cara a cara, la versión en línea (WebRDS) ha llamado la atención debido a sus ventajas, aun-

que sus limitaciones también generan preocupación. Este estudio tuvo como objetivo contrastar estos dos formatos en términos de sus potenciales sesgos.

RESPONDENT-DRIVEN SAMPLING (RDS)

EL MUESTREO DIRIGIDO POR PARTICIPANTES (RDS) es un método de muestreo de referencia en cadena donde los/las participantes son seleccionados/as a partir de la red social de la población objetivo y no de un marco muestral. Para lograr esto, el equipo de investigación selecciona algunos participantes iniciales conocidos como *semillas*, a quienes se les pide que respondan la encuesta y luego recluten a un número limitado de personas (generalmente a través de un sistema de cupones). Además de la encuesta, a estas personas se les pide que respondan algunas preguntas sobre el tamaño de su red social. Es crucial que esta información esté correctamente especificada, ya que la precisión de las estimaciones depende de la suposición de que el tamaño de la red social (o *grado*) está bien definido. El proceso de referencia en cadena continúa durante tantas *olas* como sean necesarias hasta que se alcance el tamaño de muestra deseado, las características de los/las participantes se hayan estabilizado o las cadenas se extingan. Al rastrear el reclutamiento y recopilar información de la red, el RDS puede generar estimaciones poblacionales insesgadas bajo ciertos supuestos, donde se incluyen los siguientes (6):

- 1) Las relaciones entre los individuos son recíprocas.
- 2) Los individuos forman una sola red (es decir, cada individuo puede ser alcanzado por otro individuo a través de lazos sociales).
- 3) El muestreo ocurre con reemplazo.
- 4) Los encuestados pueden reportar con precisión el tamaño de su red personal.

Potenciales sesgos en estudios epidemiológicos basados en el método *respondent-driven sampling*: comparación según su aplicación presencial y en línea
PEDRO FERRER ROSENDE et al.

5) Los reclutados son seleccionados aleatoriamente dentro de la red social del reclutador.

Por otro lado, el proceso de RDS se puede dividir en tres etapas clave (7):

- 1) Investigación formativa (*formative research*).
- 2) Recolección de datos.
- 3) Análisis.

Cada etapa presenta sus propios desafíos y son fundamentales para obtener estimaciones precisas de la población.

La investigación formativa es una parte crucial para evitar fallos en la implementación del RDS (8). En esta fase, el equipo investigador debe abordar las características de red de la población objetivo, evaluar la aceptabilidad del RDS como un método de muestreo viable, la selección de las semillas y la logística de la encuesta, como son los incentivos y el diseño de los cupones.

El proceso de recolección de datos debe ser cuidadosamente planificado y ejecutado, monitoreando constantemente las cadenas de reclutamiento y las estimaciones de variables seleccionadas a medida que avanzan las *olas* para ver si se alcanza el equilibrio (estabilización de algunos estimadores principales), al mismo tiempo que se controla la homofilia (es decir, la tendencia a tener lazos sociales con personas de características similares). Para una descripción detallada del proceso de recolección de datos en versión presencial, ver Hipp *et al.* (9) Para versión web, consultar Ferrer-Rosende *et al.* (10)

Para el análisis de datos, existen varias propuestas de estimadores de la media y la varianza, cada uno con sus propios supuestos y limitaciones. La discusión sobre los diferentes estimadores propuestos y sus sesgos queda fuera del alcance de este trabajo. Para una revisión completa, consultar Abdesselam *et al.* (5)

Si bien el RDS ha demostrado potencial, también tiene limitaciones prácticas, incluida la velocidad a la que se pueden recopilar datos, que depende de la disponibilidad del personal y de la longitud de la encuesta, así como de la disposición de los/las participantes para reclutar a otros en un período de tiempo (11). El RDS también tiene barreras relacionadas con el tiempo y la ubicación, ya que está limitado al área en la que opera el personal y donde los/las participantes pueden moverse.

WEB-BASED RESPONDENT-DRIVEN SAMPLING (WEBRDS)



POR LAS RAZONES MENCIONADAS ANTERIORMENTE, la versión en línea del RDS (WebRDS) ha ido ganando atención en la última década. El WebRDS se está utilizando para una variedad de propósitos, incluida la generación de estimaciones poblacionales, el estudio de patrones en redes sociales de enfermedades infecciosas y la entrega de intervenciones dirigidas a poblaciones específicas tales como estudiantes universitarios, hombres que tienen sexo con hombres, fumadores y personas con precariedad laboral (12).

La selección de semillas puede realizarse tanto en línea como en persona, mientras que el reclutamiento entre pares se realiza mediante correo electrónico o redes sociales. Sin embargo, últimamente el uso del teléfono frente a los ordenadores ha fomentado la aplicación a través de plataformas basadas en teléfono (7).

El WebRDS, al igual que otros métodos basados en la web, puede enfrentar dificultades tales como sesgos derivados del acceso desigual a Internet, respuestas múltiples del mismo participante o la falta de interacción cara a cara (12).

RDS FRENTE A WEBRDS



AL COMPARAR EL RDS CARA A CARA Y LA versión en línea, podemos distinguir entre ▶

Potenciales sesgos en estudios epidemiológicos basados en el método *respondent-driven sampling*: comparación según su aplicación presencial y en línea
PEDRO FERRER ROSENDE *et al.*

problemas relacionados con la implementación y con otras fuentes de sesgo. El aspecto de implementación se puede dividir en cuatro elementos clave: instrucciones; tiempo; ubicación; aplicación del método. Otras posibles fuentes de sesgo incluyen la adherencia, las respuestas múltiples y el reporte inexacto del tamaño de la red.

SESGOS RELACIONADOS CON LA IMPLEMENTACIÓN

Instrucciones. En la versión cara a cara es más probable que el participante comprenda el método de muestreo y siga las instrucciones correctamente. Por otro lado, las instrucciones para la versión en línea deben ser muy claras para minimizar el riesgo de malentendidos que podrían afectar el proceso de muestreo y las estimaciones. En la versión en línea, las instrucciones cara a cara solo pueden ser dadas a las semillas.

Tiempo. Llevar a cabo encuestas individuales y luego distribuir cupones a cada nuevo participante reclutado lleva tiempo. La versión en línea permite llegar a un mayor número de participantes en un período de tiempo más corto, ya que múltiples individuos pueden responder simultáneamente, y los cupones pueden ser entregados de inmediato.

Ubicación. El RDS cara a cara está limitado por el alcance geográfico de los lugares de reclutamiento, lo que hace necesario seleccionar cuidadosamente estos lugares durante la planificación del trabajo de campo. En contraste, la versión en línea puede extender su cobertura a un área más amplia, dependiendo de hasta qué punto se considere que se cumple el supuesto de que la población pertenece a la misma red. Sin embargo, también existe un riesgo de que individuos fuera del alcance del estudio respondan si no hay suficiente control.

Aplicación del método. La versión cara a cara puede enfrentar limitaciones basadas

en la capacidad y disposición de los/las participantes para desplazarse. Sin embargo, las personas que enfrentan desafíos como la vejez o el analfabetismo pueden recibir ayuda para responder al cuestionario. Por otro lado, la efectividad de la versión en línea depende del acceso a Internet y la competencia en el uso de ordenadores o teléfonos inteligentes.

OTRAS FUENTES DE SESGO

Adherencia (no respuesta). El tiempo o el coste invertido en que los/las participantes asistan a la encuesta presencial puede afectar a la adherencia. La versión en línea, al permitir que la encuesta se responda desde cualquier lugar o en cualquier momento, puede reducir la tasa de no respuesta.

Respuestas múltiples. Con mecanismos adecuados para controlar y prevenir que los/las participantes respondan más de una vez, prácticamente no hay posibilidad de respuestas múltiples en la aplicación cara a cara. Para WebRDS hay mayores probabilidades de respuestas múltiples, especialmente cuando están involucrados incentivos monetarios. Se requieren controles especiales para prevenir este fenómeno (7).

Reporte inexacto del tamaño de red (grado). El riesgo de reportes inexactos puede mitigarse mediante la selección meticulosa del número y contenido de las preguntas relacionadas con el grado, como sugiere Wright (13). Este riesgo puede reducirse aún más con instrucciones claras cara a cara durante la encuesta. En ausencia de instrucciones cara a cara, el potencial de reportes inexactos es mayor. Se recomienda seguir las pautas proporcionadas por Gille *et al.* (14) al seleccionar el número y contenido de las preguntas relacionadas con el grado para minimizar tales riesgos. Es crucial enfatizar la importancia de estas preguntas en las instrucciones iniciales de la encuesta.

Potenciales sesgos en estudios epidemiológicos basados en el método *respondent-driven sampling*: comparación según su aplicación presencial y en línea
PEDRO FERRER ROSENDE *et al.*

LIMITACIONES GENERALES

EL MÉTODO RDS, AUNQUE HA DEMOSTRADO ser una herramienta útil para estudiar poblaciones de difícil acceso, presenta limitaciones independientemente de la forma de aplicación que son importantes de considerar. En primer lugar, el RDS depende de las redes sociales dentro de la población. Si estas redes no están lo suficientemente interconectadas, algunos subgrupos podrían verse infrarrepresentados o simplemente no representados, generando sesgo en los resultados (2). Otro desafío importante es la homofilia, es decir, la tendencia de las personas a reclutar a otros con características similares, lo que puede limitar la variabilidad dentro de la muestra y acentuar ciertos sesgos (15). Por otro lado, el método asume que los/las participantes pueden estimar con precisión el tamaño de sus redes sociales, lo cual rara vez es el caso, afectando negativamente los cálculos de probabilidad necesarios para estimar correctamente las características de la población (2). Estas limitaciones subrayan la importancia de una evaluación rigurosa al momento de determinar la idoneidad del RDS como método de muestreo en cada estudio. Es fundamental considerar cuidadosamente tanto la forma en que se implementará el RDS como su adecuación a las características específicas de la población objetivo y los objetivos del estudio, a fin de garantizar la validez y representatividad de los datos obtenidos.

DISCUSIÓN

EL MUESTREO RDS HA DESPERTADO INTERÉS en el campo de la epidemiología y la investigación en ciencias sociales debido a su potencial para producir estimadores insesgados en poblaciones de difícil acceso. Sin embargo, se debe tener precaución, ya que algunos supuestos del método, independientemente de la forma de aplicación, pueden ser poco realistas.

Las principales ventajas del WebRDS se pueden observar en la aplicación del pro-

ceso de reclutamiento, donde la velocidad y la cobertura son puntos a favor de esta versión. No obstante, la falta de instrucciones en persona puede introducir sesgos, especialmente en el reporte del tamaño de la red y la comprensión del proceso de reclutamiento entre pares.

Por otro lado, el WebRDS es más susceptible a respuestas múltiples por parte del mismo participante y está restringido a poblaciones con acceso a Internet, pero a la vez puede atraer a más participantes que son reticentes a participar en una entrevista presencial.

Para reducir el riesgo de estudios de baja calidad y estimaciones inadecuadas se han publicado recomendaciones con herramientas de diagnóstico para evaluar los supuestos del RDS (14). Además, se ha lanzado una versión adaptada de la guía STROBE para estudios observacionales específicamente para RDS con el fin de mejorar la presentación de reportes (STROBE-RDS) (16).

Finalmente, considerando que los sesgos pueden variar según el modo de aplicación del RDS, resulta fundamental continuar investigando para identificar los enfoques de análisis de datos más adecuados en cada contexto. Si bien se han logrado importantes avances en las técnicas analíticas del método RDS en términos generales, falta información sobre cómo abordar de manera efectiva los sesgos específicos asociados a la modalidad web del RDS en comparación con su versión presencial.

CONCLUSIÓN

WEBRDS HA SURGIDO COMO UNA ALTERNATIVA coste-efectiva para implementar RDS. Bajo ciertas condiciones, que incluyen una extensa investigación formativa, un buen plan de difusión, una buena selección de semillas y una población que tenga tanto acceso a Internet como la motivación para participar en el proceso, es posible obtener una muestra con un rendimiento de reclutamiento simi-

Potenciales sesgos en estudios epidemiológicos basados en el método respondent-driven sampling: comparación según su aplicación presencial y en línea
PEDRO FERRER ROSENDE et al.

lar al del RDS cara a cara (10,17). Sin embargo, persisten varios desafíos que deben ser considerados, especialmente la falta de instrucciones cara a cara, que puede causar un reporte incorrecto del tamaño de la red y errores en el proceso de reclutamiento entre pares. Si estos desafíos pueden ser superados, WebRDS emerge como una muy buena alternativa que reduce considerablemente el esfuerzo y los costes requeridos para aplicar este método de muestreo en estudios epidemiológicos. (2)

AGRADECIMIENTOS

Este trabajo ha sido realizado en el marco del programa de doctorado en Metodología de la Investigación Biomédica y Salud Pública de la Universitat Autònoma de Barcelona.

BIBLIOGRAFÍA

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, Dhimi P, Momi 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 Steenberg 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 Com-*

mun [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 sampling 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>