Web appendix

Respondent Driven Sampling – where we are and where should we be going?

RG White et al

Dr Richard G White, Department of Infectious Disease Epidemiology, Faculty of Epidemiology & Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT. richard.white@lshtm.ac.uk

Web Table w1. Full version of proposed STROBE-RDS study reporting checklist. Guideline development proceeding according to *Moher et al* 2010(w12). This checklist adapted from STROBE guidelines(w9) checklist for cross-sectional studies(w10). Checklist scope limited to RDS reports that seek to generate representative estimates for populations or risk factor studies, as these are currently the most contentious and potentially most policy relevant uses of RDS. Specific journals will have their own, additional, reporting requirements. * = If risk factor study, give information separately for exposed and unexposed groups. [S] = unaltered from original STROBE checklist for cross-sectional studies.

	Item	Recommendation
Title and abstract	1	a) Indicate the study's design (Respondent-Driven Sampling) in the title or abstract b) Provide in the abstract an informative and balanced summary of what was done and what was found[S]
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported[S]
Objectives	3	State specific objectives, including any pre-specified hypotheses[S]
Methods		
Study design	4	a) Present key elements of study design early in the paper[S] b) State why RDS considered the most appropriate sampling method
Setting	5	a) Describe the setting, location(s), and relevant dates, including periods of recruitment and data collection. If a risk factor study, also describe exposures b) Describe formative research methods & findings used to inform RDS study design

	Item	Recommendation
Participants	6	a) Give the eligibility criteria, number, sources and methods of seed selection b) State if additional seeds were required, and if so, when and how recruited and started c) State if there was any variation in study design during data collection (eg changing numbers of coupons per recruit, or stopping chains) d) Give the eligibility criteria for subsequent recruits if it differs from seeds e) Give number, types (eg mobile/static) & location of recruitment venue(s) f) Consider reporting information on coupons(s). g) Report wording of network size question(s) h) Describe how participants were trained/ instructed to recruit others, including maximum number of recruitments, any maximum time referral permitted, and any efforts to encourage random sampling within recruits' network i) Refer to ethical review clearance documents j) Consider reporting recruitment challenges (eg selling of coupons, imposters, duplicate recruits)
Variables	7	a) Clearly define all outcomes, and if applicable exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable [S] b) State if and how recruiter-recruit relationship was tracked c) Consider reporting additional social network data, if available d) Consider reporting information on composition of personal networks, if available
Data sources/ measurement	8*	a) For each variable of interest, give sources of data (eg instrument) and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group[S] b) Describe incentives for participation and recruitment c) Describe methods to assess eligibility and reduce repeat enrollment (eg coupon

Item		Recommendation
		manager software, biometrics, detection of commercial exchange of coupons) d) QA/C checks (eg were returned coupons actually distributed & redeemed only once?)
Bias	9	Describe any efforts to address potential sources of bias[S]
Study size	10	Explain how the sample size was arrived at[S]
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why[S]
Statistical methods	12	a) Describe all statistical methods, including name and description of the analytical methods (ie point and interval estimators) used to take account of RDS sampling strategy. If appropriate, report software package used with version number and settings values b) Report any criteria used to support statements on whether estimator conditions or assumptions were met eg 'RDS equilibrium reached' c) State if seeds included in each analysis d) If applicable, describe methods used to control for confounding e) Describe any methods used to examine subgroups and interactions[S] f) Explain how missing data and small numbers were addressed g) Describe any sensitivity analyses[S]
Results		
Participants	13*	a) Report numbers of individuals at each stage of study, ie final number of seeds, number examined for eligibility, number confirmed eligible, number included in study, number returned for incentive collection and (if applicable) re-interview, and number analysed. Consider use of a flow diagram to summarise this. b) Give reasons for non-participation at each stage, including if data collected reported reason for coupon rejection

	Item	Recommendation
		c) Report number of coupons distributed and returned d) Report number of recruits by seed and number of RDS recruitment waves. Consider showing graph of entire recruitment network, whilst ensuring anonymity maintained. e) Consider reporting numbers potentially eligible if population size estimates made
Descriptive data	14*	a) Give characteristics of study participants (eg demographic, clinical, social) if appropriate, information on exposures and potential confounders[S] b) Indicate number of participants with missing data for each variable of interest[S]
Outcome data	15*	Report numbers of outcome events or summary measures[S]
Main results	16	a) Report unadjusted estimates and their stated precision (eg, 95% confidence interval) b) If applicable, report adjusted estimates and their stated precision (eg, 95% CI) c) If adjusted estimates presented, report enough information so that the reason for the magnitude of the adjustment is clear (eg network sizes and homophily by group) d) If appropriate, make clear which confounders were adjusted for and why included e) Report category boundaries when continuous variables were categorized[S] f) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period[S]
Other analyses	17	Report other analyses done—eg a) analyses of subgroups and interactions b) sensitivity analyses eg different RDS estimators, different definitions of network size
Discussion		
Key results	18	Summarise key results with reference to study objectives[S]
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Consider the limitations of cross-sectional studies, the RDS sampling

	Item	Recommendation
		method and, if used the RDS method(s) of inference. Include comment on how representative the unadjusted sample is thought to be. Indicate how participants compare to population description developed during formative assessment and other sources of information. Discuss both direction and magnitude of any potential bias
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence[S]
Generalisability	21	Discuss the generalisability (external validity) of the study results[S]
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based[S]
Data sharing	23	State whether access provided to data and survey, and if so, how to access

Web appendix references

w1 Semaan, S., et al., Ethical and regulatory considerations in HIV prevention studies employing respondent-driven sampling. Int J Drug Policy, 2009. 20(1): p. 14-27.

w2 Heimer, R., Critical issues and further questions about respondent-driven sampling: comment on Ramirez-Valles, et al. (2005). AIDS Behav, 2005. 9(4): p. 403-8; discussion 409-13.

w3 Johnston, L., Introduction to Respondent Driven Sampling. Participant Manual, 2008, Tulane Univeristy/CDC/GAP/UTAP: New Orleans.

w4 Heckathorn, D.D., Extensions of Respondent-Driven Sampling: Analyzing Continuous Variables and Controlling for Differential Recruitment. Sociological Methodology, 2007. 37(1): p. 151-207.

w5 Pollini, R.A., et al., Syringe possession arrests are associated with receptive syringe sharing in two Mexico–US border cities. Addiction, 2008. 103(1): p. 101-108.

w6 Wejnert, C., Social Network Analysis with Respondent-Driven Sampling Data: A Study of Racial Integration on Campus. Soc Networks, 2010. 32(2): p. 112-124.

w7 Paz-Bailey, G., et al., How many men who have sex with men and female sex workers live in El Salvador? Using respondent-driven sampling and capture-recapture to estimate population sizes. Sex Transm Infect, 2011. 87(4): p. 279-82.

w8 Broadhead, R.S., et al., Harnessing peer networks as an instrument for AIDS prevention: results from a peer-driven intervention. Public Health Reports, 1998. 113(Suppl 1): p. 42.

w9 von Elm, E., et al., The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. PLoS Med, 2007. 4(10): p. e296.

w10 STROBE. STROBE checklist for cross-sectional studies. 2007 [cited 2011 12 Dec 2011]; Available from: http://www.strobe-statement.org/index.php?id=available-checklists.

w11 RDS list server. RDS list server. Available from: RESPDRIVENSAMPLING@Princeton.EDU.

w12 Moher, D., et al., Guidance for Developers of Health Research Reporting Guidelines. PLoS Med, 2010. 7(2): p. e1000217.