Respondent driven sampling—where we are and where should we be going?

Richard G White,1 Amy Lansky,2 Sharad Goel,3 David Wilson,4 Wolfgang Hladik,5 Avi Hakim,5 Simon DW Frost6

Respondent Driven Sampling (RDS) is a novel variant of link tracing sampling that has primarily been used to estimate the characteristics of hard-to-reach groups, such as the HIV prevalence of drug users.1 ‘Seeds’ are selected by convenience from a population of interest (target population) and given coupons. Seeds then use these coupons to recruit other people, who themselves become recruiters. Recruits are given compensation, usually money, for taking part in the survey and also an incentive for recruiting others. This process continues in recruitment ‘waves’ until the survey is stopped. Estimation methods are then applied to account for the biased recruitment, for example, the presumed over-recruitment of people with more acquaintances, in an attempt to generate estimates for the underlying population.

RDS has quickly become popular and relied on by major public health organisations, including the US Centers for Disease Control and Prevention and Family Health International, chiefly because it is often found to be an efficient method of recruitment in hard-to-reach groups, but also because of the availability of custom written software incorporating inference methods that are designed to generate estimates that are representative of the wider population of interest, despite the biased sampling.

As demonstrated by RDS’s popularity,1 there was a clear need for new methods of data collection on hard-to-reach groups. However, RDS has not been without its critics. Its reliance on the target population for recruitment introduced ethical and sampling concerns.2–5 If RDS estimates are overly biased or the variance is unacceptably high, then RDS will be little more than another method of convenience sampling. If these errors can be minimised however, then RDS has the potential to become a very useful survey methodology.

In this editorial we highlight that ‘RDS’ includes both data collection and statistical inference methods, discuss the limitations of current RDS inference methods for generating representative estimates, highlight other applications of RDS for which it might not be a panacea, but could still be the best method to collect data on many hard-to-reach groups. However, it is critical that the concerns about statistical inference be addressed, and the current benefits and limitations of RDS be better communicated to the broader public health community.

Another concern is that currently RDS studies are not being adequately reported.1 Ultimately his reduces
the utility of published data and hinders assessment of study quality. Development of specific RDS reporting guidelines will assist in the interpretation of estimates and findings from RDS studies and in the evaluation of the RDS method itself. To facilitate this process, we have drafted a RDS study reporting checklist (Summary in table 1; Full version in web appendix table W1 or on Equator Network website) that we have adapted from the STROBE guidelines for cross-sectional studies.\(^9\)\(^–\)\(^10\)\(^\text{[42]}\) after receiving feedback from RDS experts contacted via the RDS list server\(^\text{[43]}\) and personal contacts. We invite further comments on the full draft checklist (web appendix table W1), either directly to the corresponding author, a Rapid Response on the STI website, or via the Equator Network website.\(^\text{[44]}\) These comments will feed directly into a planned guidelines-setting meeting during which the contents of this checklist and the accompanying guidelines will be discussed and hopefully a consensus reached, followed by formal publication of the resulting guidelines.

There are many current priority areas for RDS research. There needs to be a clearer distinction between the methods used for RDS sampling and the methods used for statistical inference. There is a need for more systematic reporting of RDS studies. There is a critical need for more robust empirical evaluations to measure RDS sampling errors (bias, and variance if possible) in a range of different populations as context is likely to be important. Current efforts to develop new inference methods should be intensified,\(^9\)\(^–\)\(^10\) and more focus should be given to designing diagnostics to identify when problems are occurring during RDS recruitment. If problems are detected during data collection, steps could be taken to alleviate these problems immediately by correcting the problem or collecting additional data that may allow correction during the estimation stage. As new inference methods are developed, it would be preferable if they were released as open source programmes in commonly used statistics packages (R, STATA, SAS etc) so they can be more easily evaluated and compared to existing methods.

RDS has undoubtedly generated a wealth of new data on populations that have historically been difficult to access, and RDS is here to stay. The challenge now is to improve the methods (both sampling and inference) and ensure studies are reported well enough so that we can make the most of these data to improve public health.

Acknowledgements This editorial was largely based on presentations and discussions at the symposium organised by RGW “Respondent Driven Sampling: Where we are and where should we be going?” at the 19th Biennial International Society for STD Research conference in Quebec, Canada in July 2011. We would like to thank many individuals for comments on the draft reporting checklist, in particular Amy Drake and Lisa Johnston.

Additional supplementary files are published online only. To view these please visit the journal online (http://dx.doi.org/10.1136/sextrans-2012-050703).

Contributors RGW wrote the first draft of the editorial and STROBE-RDS check list. All other authors revised the editorial and checklist critically for important intellectual content and helped redraft. RGW wrote the final draft. All authors agreed with the final version of the manuscript.

Funding RGW is funded by a Medical Research Council (UK) Methodology Research Fellowship (G0802414), the Consortium to Respond Effectively to the AIDS/TB Epidemic (19/90/01), and the Bill and Melinda Gates Foundation (21675). SDWF is funded by the National
Institute of Nursing Research (grant NR10961), the National Institute on Drug Abuse (grant DA24998), and by a Royal Society Wolfson Research Merit Award. The funders had no involvement in the design, collection, analysis or interpretation of the data, in writing the report or in the decision to submit.

Competing interests None.

Provenance and peer review Non commissioned, internally peer reviewed.

This paper is freely available online under the BMJ Journals unlocked scheme, see http://bjo.bmj.com/site/about/unlocked.xhtml

Accepted 8 August 2012

doi:10.1136/sextrans-2012-050703

REFERENCES


