Reviewer's report

Title: Understanding recruitment: outcomes associated with alternate methods for seed selection in respondent driven sampling

Version: 2 Date: 21 February 2013

Reviewer: Edward Fottrell

Reviewer's report:

The question posed by the authors is well defined and essentially addresses the effect of two varying methods of seed selection in respondent driven sampling (RDS) on derived estimates of population sociodemographic characteristics and risk behaviours associated with HIV. The application of simultaneous yet separate RDS studies in the same population at the same time is original. The work is clearly described within the context of evaluations of RDS and appropriately recognises and cites related studies, which further helps to identify the originality and added-value of the current study. The experimental design and data capture methods of the study are sound and well described. As acknowledged by the authors, their study adds to a growing body of evidence rather than provides a definitive endpoint. The manuscript is well-written with only minor revisions of language necessary to clarify some points (detailed below).

Minor Essential Revisions

My comments largely relate to the clarity of presentation of methods and results and to the accessibility of this work to a more general health sciences readership, who may not have extensive knowledge on RDS methods or approaches to interpreting RDS derived data or evaluating RES approaches.

1) Although a summary of the purpose of RDS and its widespread application is provided, some more detail on the RDS method itself and a summary of the principles of this method (i.e. what it is) would be appropriate in the Background section. Similarly, I suggest that the authors clarify the distinction between RDS sample proportions and RDS estimates between lines 6-9 on page 4, as this may not be clear to general readers.

2) A full-stop needed at the end of page 6.

3) Outcome measures need to be more clearly defined in the methods section under sample analysis, especially ‘equilibrium sample proportion’ and ‘convergence’.

4) Pajek should be identified as computer software.

5) Citation of the larger study in which the current evaluation was based is needed in the first sentence of the Methods section. Similarly, citation of the formative research study described on line 4 page 9 should be provided.
6) The designation of arm 2 seeds needs further clarification. Line 15 page 13 states that all those presenting to study staff were designated as seeds. However, only 108 had the potential to actually act as seeds and these 108 are used as the denominator in calculation of mean recruitment per seed, suggesting that in fact only those who accepted coupons were designated as seeds.

7) To determine the suitability of the cut-off used in recording network sizes, it would be useful to know more precisely what proportion of participants did in fact report personal networks of greater than 10 people.

8) The proportion of participants that consented to serum testing (and thus contribute to the analyses relating to HIV) should be specified in the Results section.

9) The description of logistic regression modelling and the results in table 1 need more details to enable interpretation. What comparisons do the significant differences relate to and what statistic do the p-values relate to?

10) The presentation of results in Table 2 is confusing. The reference group for all analyses is arm 1 recruits, yet each parameter is presented with a reference category, implying that comparisons are made against that reference category within each parameter. Presentation of arms/seeds as rows and outcome parameters as columns might facilitate clearer presentation and interpretation of results.

11) The description of results of RDS measures needs to be clearer and more precise. For example, what different results are shown in tables 3 and 4 and what are the most important findings here?

12) I suggest the description of representativeness in the results section of the text follows the description of homophily because the latter is presented in table 3 and the former is presented in table 4.

13) Meaning of first paragraph on page 19 is difficult to follow (particularly the first two sentences). Can the authors consider how to rewrite this so that the meaning is absolutely clear?

14) In terms of representativeness, the results suggest that both RDS methods achieve representativeness. Yet this seems counter-intuitive given the observed differences in the samples between the two methods and the different parameter estimates derived from the two methods, as well as the conclusion that different methods of seed selection can produce different RDS results. Whilst the statistical tests used might indicate representativeness, some more detailed interpretation and discussion of this would be useful.

Discretionary Revisions

1) One possible interpretation of the results in tables 1 and 2 might suggest that arm 2 seeds actually recruited even more marginalised people (i.e. in terms of education, income, solvent use and sex work). Some comment on this possible
interpretation, its appropriateness, possible explanations for it and possible implications would be a useful addition to the manuscript.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Acceptable

**Statistical review:** Yes, but I do not feel adequately qualified to assess the statistics.