Author's response to reviews

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

Authors:

John L Wylie (John.Wylie@gov.mb.ca)
Ann M Jolly (Ann_M_Jolly@phac-aspc.gc.ca)

Version: 3 Date: 16 May 2013

Author's response to reviews: see over
May 16, 2013

Dear Dr. Aldcroft,

Please find attached a revised manuscript entitled “Understanding recruitment: outcomes associated with alternate methods for seed selection in respondent driven sampling”, authored by John. L. Wylie and Ann M. Jolly.

We would like to thank the reviewers for their comments. We have addressed their concerns and have summarized our responses below on the following pages. If you have any further questions or require any clarification, please let us know. Thank you for considering this manuscript for publication in BMC Medical Research Methodology.

Sincerely yours,

[Signature]

John L. Wylie, PhD
University of Manitoba and Cadham Provincial Laboratory
Winnipeg, Manitoba
Canada
P: 204-945-7473
F: 204-786-4770
E: John.Wylie@gov.mb.ca
Reviewer's report 1  
Reviewer: Edward Fottrell  
Minor Essential Revisions

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.

We have added additional details on RDS as follows, line 13, page 1: “Briefly, the approach as originally described involves the selection of a small number of “seeds”; i.e. individuals who will be instructed to recruit others, with recruitment being restricted to some maximum number (typically 3 recruits maximum per person). Subsequently recruited individuals continue the process such that multiple waves of recruitment occur. Ultimately any bias associated with initial seed selection would be eliminated and the resultant sample could be used to produce reliable and valid population estimates via RDS software designed for that purpose”.

For sample proportions vs. RDS estimates we added some additional details in parentheses, line 14, page 2, as follows: “These researchers found that across 7 variables, the majority of RDS sample proportions (the observed proportions of the final RDS sample) were closer to the true population proportion than the RDS estimates (the estimated population proportions as generated by RDS software) and that many RDS confidence intervals did not contain the true population proportion.

If the reviewer has any suggestions for additional details needed or rewording of what we have inserted, we would welcome them.

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

Full stop has been added.

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

See my comments under point 11) below. Based on what has been raised by reviewer 3, several of the RDS analyses may not be necessary and did not add substantially to the empirical description of our results. This has allowed simplification by removing measures such as convergence and representativeness. Now under sample analyses, we mention only the estimated population proportion and homophily and define homophily there as follows, line 2, page 13:

“The analysis for table 3 used RDSAT version 5.6 [21] to generate the RDS measures of estimated population proportion and homophily. Homophily values in RDS can vary from -1.0 to 1.0. Values near 0 indicate random recruitment (e.g. a value of 0 for individuals with male gender would indicate that males were equally likely to recruit a male participant as a female
participant). Positive homophily values indicate a tendency to recruit others who share a given characteristic, while the opposite is true for negative values.”

4) Pajek should be identified as computer software.

A reference for Pajek has been added on page 12 (reference 20).

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.

The overall name of the study was Social Network Study III (SNS III) but no publications are yet available. Our first goal was analysis of the data as presented in this manuscript in case it informs our later analyses. As such we do not have a publication to cite yet that describes the overall study or other parts of it. However, we have now made note of the name of the overall study for future cross reference, line 18, page 6: Data collection took place in Winnipeg, Manitoba, Canada as part of a larger survey (Social Network Study III – SNS III) designed to better understand interactions between individuals at risk for STBBI.

For citation of the formative research mentioned by the reviewer, we do have a paper that we can cite there; reference 19 has now been added, line 11, page 9.

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.

This paragraph has been reworded to clarify, line 5, page 14. It now reads “Within arm 2, 118 individuals self presented to study staff and were designated as Arm 2 seeds. Of these, 108 agreed to attempt recruitment with 63 successfully recruiting. At close of data collection, arm 2 recruits numbered 264 resulting in a mean recruitment of 2.2 individuals per arm 2 seed. The largest recruitment chain consisted of 34 individuals. The mean number of recruits per chain within arm 2 was 4.2, with 6 chains containing 10 or more individuals. For these latter 6 chains, the number of waves of recruitment ranged from 4-6, with a mean of 5”.

It now keeps consistent the idea that all self-presenters were given the opportunity to recruit and designated as seeds so that they could be included in later analyses. This then further created breakdowns such as those that refused coupons, those that took coupons but ultimately decided to not follow through or could not recruit, followed by those that took coupons and did recruit. We prefer to keep the 10 non-coupon acceptors as part of our arm 2 seed group for the following reasons. First, given the number of overall non-recruiters, this small group of 10 may be people who simply knew right from the start they were unlikely to be able to recruit. And second, out of respect for their time and with no clear indication they were different than other non-recruiters, we prefer to have them included within the designation of seed.
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.

This has now been noted, line 13, page 12, as follows: “Of the 600 people in the study, 552 (92%) indicated their network consisted of 10 or fewer 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.

This has been noted, line 14, page 12, as 508/600 (84.7%).

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?

Analysis has now been clarified as using a Chi square analysis for the initial univariable step in table 1 to decide on possible variables for the multinomial logistic modelling of table 2, line 18, page 12, as follows: “The analysis summarized in table 1 used Chi square analysis to identify overall differences between the arm 1 recruits, the arm 2 seeds, and the arm 2 recruits. The analysis of table 2 used multinomial logistic regression to identify differences between the arm 1 recruits (used as the reference group) and the arm 2 seeds or arm 2 recruits. The 22 arm 1 seeds were not included, given their small number and purposeful selection. In the multinomial analysis, the effect of removing variables was assessed through the likelihood ratio test.”

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.

We did try following this suggestion in terms of presentation of table 2, but upon reviewing it, we realized this could potentially create another source of confusion as it would mean that the presentation of table 2 would have now deviated from the format used in the other tables where the arms/seeds were presented as columns. Our solution was to remove the continual notation of “Ref” from the table and only list the category for each variable that relates directly to the odds ratio. This doesn’t change the results and is an approach that we have used in other papers and hopefully better addresses the confusion. For consistency, we also removed the various “Ref” lines from table 5 so there is consistency across all tables.

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?
Initially we split the data into tables 3 and 4 even though they all relate to the same analysis. This was done simply to accommodate the amount of data, but the reviewer is correct in terms of this creating confusion, since different tables would typically contain a different set of analyses. However, rather than trying to better explain what tables 3 and 4 contain, reviewer 3 has raised an excellent point in terms of his concerns that the representativeness measure that we employed as published by Wang et al. may in fact be in error (an error in the sense of the original publication of the measure; see his comments below) and not typically used in RDS studies. Therefore, given that a) analyses such as representativeness and other aspects such as convergence could, in fact, be in doubt, b) the presence or absence of these measures doesn’t affect our empirical description of our results (i.e. they garner only a few lines in the manuscript or are not specifically mentioned at all, such as the Pe measure in our old table 4 and c) the need to clarify the contents of tables 3 and 4, our suggestion is to remove several of the more esoteric RDS measures that we included, specifically representativeness and convergence and some of the RDS output data that are used for the aforementioned measures. In particular this allows us to collapse the data of table 4 into table 3 and brings the data into the more common pattern of putting a given analysis into one table. This has now been done and table 3 contains all aspects of the RDS measures such as homophily values and the confidence intervals generated by the software, that are discussed in detail in the manuscript. We believe this also helps simplify the overall manuscript and its readability. If this is acceptable, this is the approach we would like to take in terms of clarifying tables 3 and 4.

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.

See my comments above, and reviewer three’s comments regarding representativeness and our thoughts that it should be removed altogether from the manuscript.

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?

After reading the paragraph in question again and thinking about what we were trying to convey, I suspect we could easily remove it and not alter any aspects of the analysis or its key discussion points. When we first conceived of the paragraph, we wondered whether readers might look at the results and pose questions such as - if a typical RDS analysis is done are there any clues that specific segments of the target population corresponding to specific behavioural attributes had been missed, given our insight in terms of having two distinct arms of data collection? Essentially the answer is “no” as we tried to explain in that paragraph. However, really the paragraph involves quite a bit of speculation and projection on our part in terms of trying to predict what readers may think. We are therefore suggesting that this entire paragraph be removed, which again helps to make the paper much more concise and to the point, without changing any of the more concrete observations made in the paper.

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.

We do agree that this specific analysis did seem counter-intuitive and ties in well with our comments above and reviewer three’s comments about the representativeness measure potentially being in error and not typically used in RDS papers since the time of its original publication. This further reinforces our suggestion that this measure be removed from the paper.

**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.

This may be the case but it’s somewhat difficult to say. Overall, all of the variables in table 2 could act as proxy measures of marginalization, depending on their accompanying behavioural patterns. In that scenario, as one examines which OR are significant, it could be argued that depending on the variable in question marginalization was highest amongst arm 2 seeds but not their recruits (education and income), both arm 2 seeds and recruits (sex work), arm 2 recruits but not the arm 2 seeds (solvent users) or arm 1 recruits over arm 2 seeds and recruits (MSM). But we do agree that the variables that could be argued as the most likely direct measure of marginalization such as having dropped out of school or being unemployed do appear to be associated with the arm 2 seeds (but not their recruits who do not differ significantly from arm 1 recruits for those two variables).

Overall, we do therefore agree with the reviewer’s suggestion and have inserted the following, line 11, page 17: “The individuals in Arm 2, in particular the arm 2 seeds, may represent the most marginalized members of the overall population from which we were sampling (for example, based on their lower education and income levels and greater likelihood of being solvent users – see table 2). This marginalization may be one of the underlying determinants that governed their apparent lesser likelihood of obtaining an RDS coupon from any of the individuals in Arm 1. This occurred despite their apparent social connection to the population (i.e. without any advertising they still became aware of the study and obtained sufficient study information to initiate contact with the study nurse). Our data does not reveal whether this would have been inadvertent or purposeful on the part of the individuals enrolled in Arm 1, but it does raise questions as to whether the most marginalized members of a target population may be the least likely to have the means to enter a typical RDS study. Marginalization and enrolment in studies of this kind is an area that deserves further research to ensure the most marginalized and vulnerable members of a population are not inadvertently being excluded from enrolment and hence essentially remaining unknown to study staff.”
Reviewer's report 2
Reviewer: Anneli Uuskula

Reviewer's report:
Major Compulsory Revisions

The paper would benefit from clearer and more detail methods presentation:
1) Without clear and justified study inclusion criteria it is very hard to capture and interpret the results/discussion. Currently, the information provided is limited and scattered throughout the text: requesting age to meet the minimum age requirement of 14 (page 8/line 7), people at risk for STBBI, including people/seeds based on their “their interconnectedness with other members of the target populations” (page 7/ line 9)?? It is really hard to understand to what population group the results can be generalized??

We have now moved all of the various recruitment details together such that they appear, line 12, page 8, as follows: Recruitment coupons were provided at the end of questionnaire administration. Given their familiarity with the types of questions asked, study participants were instructed to recruit other friends or family members who they believed practiced some of the risk behaviours they had been questioned about. Upon presentation, potential study participants were asked their age to meet the minimum age requirement of 14, with no other pre-screening occurring. This broad criteria was largely driven by the wide range of risk groups under investigation in the SNS III study, as one of the research aims of this larger study was to better understand bridging between different risk groups (analyses to be presented in future publications).

Essentially it is a very broad group consisting of many subgroups generally considered to be at higher risk of HIV such as IDU, sex workers, MSM, street-involved youth, etc. This is broader than most RDS studies, but in general, the literature around RDS now suggests that any generalization may be problematic as RDS, like other typically used sampling methods, may not create a representative sample, which is also backed up by our work.

2) It seems that comparison meeting the set aims should be conducted between Arm 1 and Arm 2 recruits (excluding seeds). It is not clear to this reviewer whether these 2 groups (Arm 1 and Arm 2 recruits) are actually comparable: Arm 1 with the mean recruitment of 8.9 per seed, and Arm 2 with mean recruitment 2.4 per seed. In the paper one can read that there was a preset sample size, and currently there is a doubt that recruitment to the Arm 2 was stopped ‘artificially’ at the targeted study sample size. Thou the recruitment chains in Arm 2 might be significantly shorter, and one can expect differences in sample / population proportions related to this fact only. Information on rules on recruitment cessation in the study and data that would enable assessing comparability of the Arm 1 and Arm 2 recruits should be presented;

Essentially because of the way we presented results it is possible to compare only the Arm 1 and Arm 2 recruits if a reader prefers, using the data in tables 1 and 2 and only looking at the numbers/results corresponding to the recruits in both arms. We felt it important to include the arm 2 seeds as well since they did essentially self-recruit and were not artificially selected by our study staff. They therefore represent a relatively large group that could typically enter a study depending on what type of sampling strategy is undertaken. And yes, the arm 2 seeds were
poorer at recruitment which may be linked back to the comments we made at the end of reviewer 1’s section regarding marginalization. But even if recruitment success does differ, it is still meaningful to better understand how these groups do compare to each other. Poor recruitment may be a characteristic of these individuals and a comparison between the groups simply helps to better understand how they differ. It is also important to note the same number of seeds in Arm 1 and Arm 2 (6 in each case) did generate relatively large recruitment chains exceeding 10 people and five or more waves of recruitment. It was clearly easier to find Arm 1 seeds who were able to do this (6 of 22) vs. Arm 2 (6 of 118), but this is important to know as it does reveal some information as to the number of seeds that could be necessary to generate recruitment chains. For example, it may imply that if a researcher wants to do an RDS study with very marginalized individuals, it may be necessary to select a large number of seeds, but this is an open question for future studies.

3) Study / data collection site – Was this a research center? A public health service (harm reduction, HIV testing, counseling) center known to the target group? Else? This information is needed for better contextualization of the results (i.e. whom the self presenting seeds represent – service clients?; more/less marginalized group of people – someone who know that the study is ongoing but to whom the study participants will not render the recruitment coupon … because they are not considered to be ‘like themselves’?);

We’ve clarified study logistics, line 1, page 7, as follows: “Questionnaire administration occurred over an 11 month period from January to December 2009. Interviewing and specimen collection was conducted by one research nurse. A variety of interview sites had been established by this nurse prior to study implementation. These interview sites were located within local clinics or resource centres geographically dispersed throughout the areas of Winnipeg where it was expected most participants would reside. Upon first phone contact with the nurse, a mutually agreed upon interview time and place was established and the nurse traveled to a given interview site at the appointed time. Similar approaches have been used by others to ensure RDS can be carried out in a cost-effective manner.”

It’s important to note that we were not catering to any specific clients that may have been attending some type of clinic or NGO. All recruitment was being conducted by community members in the community and we were only using various institutions in terms of trying to have a relatively large number of potential interview rooms scattered throughout the central part of Winnipeg where most study participants resided. Any study participant could choose any site that was most convenient to them. The study nurse always accommodated their wishes in terms of where they preferred to be interviewed.

4) The information presented on data collection is conflicting. At page 6/line 16 authors write “Recruitment, interviewing and specimen collection was conducted by one research nurse”; at page 8/line 17-18 one can read “A variety of interview sites from which participants could freely choose had been established prior to study implementation within local clinics or resource centres”;}
We’ve clarified this in the manuscript, as noted above. There is no conflict and what we wrote is correct. There was only one research nurse and she would go to whatever site a study participant chose to be interviewed. They simply set up an appointment time to meet and she went to that spot at a given time to meet with a given study participant.

5) In their justification for the selected recruitment methodology authors highlight the potential to study how bridge individuals link diverse groups of people (page 8/lines 9-10). However, none of this analysis is presented?

No, this data is not present. In rereading our manuscript we can understand that our original wording may be taken to imply that this data might be a topic of discussion later in this manuscript, but this was not our intention. This is just one of several future analyses that we intend with this dataset and will be the subject of future papers. It did not play directly into the empirical description of the RDS recruitment that we wanted to cover in this paper. We have now reworded this section, line 16, page 8, as follows: “This broad criteria was largely driven by the wide range of risk groups under investigation in the SNS III study, as one of the research aims of this larger study was to better understand bridging between different risk groups (analyses to be presented in future publications).”

6) In the Table 3 it is not clear whether the Ps calculations include the seeds or not?? (should not);

The reviewer is correct; it did not include the seeds. This is now clearer in table 3 as we’ve moved the proportion data directly into the same column as the recruitment numbers.

7) The information on homophily (pages 14/15) is better placed in the Methods section;

We have moved the homophily background data to the methods section and it appears at line 3, page 13, as follows: “The analysis for table 3 used RDSAT version 5.6 [21] to generate the RDS measures of estimated population proportion and homophily. Homophily values in RDS can vary from -1.0 to 1.0. Values near 0 indicate random recruitment (e.g. a value of 0 for individuals with male gender would indicate that males were equally likely to recruit a male participant as a female participant). Positive homophily values indicate a tendency to recruit others who share a given characteristic, while the opposite is true for negative values.”

8) Authors used a personal egocentric network measure (probably also in the RDS analysis). Has this measure been validated against ‘traditional RDS network measure’?

No, but we did not have any clear alternative given the complexity of the number of possible risk groups that we were allowing into the study. We have added the following sentences to the limitations section, line 4, page 20, and, as noted, many of the empirical comparisons we did would not involve the network size and would not be affected by validity issues associated with that measure.

“Our egocentric network measure that was used as an input for the RDS software differs somewhat from the typically much narrower type of risk behaviour network measure used in
most RDS studies. This was necessary given the broad range of risk groups that were a part of this study and could affect some RDS measures such as the estimated population proportions. However, the majority of results presented in this paper (i.e. tables 1, 2, 4 and 5) would not be affected by this network size data”
Reviewer's report 3

Reviewer: Richard D Burt

Reviewer's report:

Major compulsory revisions

1) It seems that recruitment chains were quite short in both the sample based on researcher-selected seeds and, in particular, the sample derived from self-selected seeds. The authors report that the first sample had 6 chains containing 10 or more individuals, the second 2 chains containing 10 or more individuals. It would be important to know the number of recruitment waves in both groups. If, as seems likely, the vast majority of study participants were recruited in the first few waves, then it would appear that RDS recruitment has not had an opportunity to adequately penetrate through the target population. This makes it difficult to compare the paper’s results with RDS studies which achieved a wider elaboration into their target population. This limitation does not seem amenable to correction during analysis but requires consideration in the discussion of the paper’s results.

The first clarification we have to make here is that we poorly worded a sentence which appears to have lead to the reviewer misreading the sentence. The reviewer states above the first sample had 6 chains containing 10 or more individuals, the second 2 chains containing 10 or more individuals (we acknowledge that our original sentence of “Six arm 2 recruitment chains contained 10 or more individuals” can be easily misread, as appears to have happened in this case). In reality both arms generated the same number of recruitment chains of 10 or more individuals; Arm 1 recruitment had 6 chains of 10 or more individuals, while Arm 2 also had 6 chains of 10 or more individuals. We have now reworded the sentences so they have exactly the same structure, line 1, page 14: “the mean number of recruits per RDS chains within arm 1 was 11.5, with 6 chains containing 10 or more individuals” and, line 8, page 14: “the mean number of recruits per chain within arm 2 was 4.2, with 6 chains containing 10 or more individuals”.

Having said that, we have also added the data the reviewer requests to these respective paragraphs and note that the number of waves in these chains ranges from 5-9 waves for Arm 1 and 4-6 waves for Arm 2. At line 9, page 20, we also inserted as a limitation “the number of waves of recruitment seen in some RDS studies exceeds the maximum number of waves we obtained (9 waves in one of the Arm 1 recruitment chains) and it is possible that eventually recruitment differentials of the type we observed would diminish if a sufficiently large number of waves can be completed. Future studies can be designed to address this question”. This therefore indicates that if we had continued to collect more waves of data some variables may have ceased to show differences. It would be an interesting future study to specifically see if more (and how many) waves of data are necessary to alleviate the recruitment differentials we observed.

2) The target population does not appear to be explicitly defined. I do not see a description of eligibility requirements in the manuscript and the collection of characteristics evaluated
(IDU, MSM, sex work, street involved youth) indicate a very broad study population. By casting such a wide net the problem of adequately penetrating the study population is compounded. Further, the capacity of the paper’s finding to be generalizable to RDS studies with more conventional target populations is called into question. Again, this problem would seem to be amenable only to ex post facto discussion.

As per the suggestions of reviewer 2, we have attempted to bring all relevant recruitment criteria together into one place, line 12, page 8, as follows: “Recruitment coupons were provided at the end of questionnaire administration. Given their familiarity with the types of questions asked, study participants were instructed to recruit other friends or family members who they believed practiced some of the risk behaviours they had been questioned about. Upon presentation, potential study participants were asked their age to meet the minimum age requirement of 14, with no other pre-screening occurring. This broad criteria was largely driven by the wide range of risk groups under investigation in the SNS III study, as one of the research aims of this larger study was to better understand bridging between different risk groups (analyses to be presented in future publications)”.

But yes, the reviewer is correct that our target catchment was very broad, although given the number of RDS studies that have now been done around the world it is conceivable that other groups have used or are planning on using similarly broad criteria. We do agree that this is worth noting in the limitations paragraph, which we have now done, as follows, line 13, page 20: “Our recruitment involved very broad risk groups whereas the majority of RDS studies typically have narrower recruitment criteria, and, as noted above, recruitment differentials may have eventually diminished in our sample. Overall, the criteria for enrolment and recruitment in published RDS studies do vary depending on the research question. Given this variation it would be important to understand what effect enrolment criteria has on the number of waves of recruitment that may be required in different scenarios”.

**Minor essential revision**

I have doubts that the t-test in the Wang paper allows the assertion that an RDS sample is representative. I would not presume the lack of a clearly significant difference offers strong proof that there is no difference; low power would bias the test towards representativeness. I corresponded with Wang some years ago and his response indicated that he believed the presentation was in error. Outside of publications from the Wright State group, I am unaware of the test being reported in RDS publications.

We think this is an important point to have raised and please see my reply for point 11 of reviewer 1. Briefly, we think representativeness and some other RDS measures can be removed from the manuscript without affecting our results, which also has the side benefit of simplifying the manuscript.

**Discretionary revisions**

I would not include the self-recruited seeds in the derivation of the univariate statistics of Table 1. One would expect self-selected seeds to differ from the other groups and including
them in the statistical test masks univariate differences between the two RDS-recruited groups.

The reason we thought it important to include these arm 2 seeds with respect to data descriptions and analysis is that they were self-selecting and we did not have any role in their contacting us for potential entry into the study. They therefore represent a group that could have easily entered some other study if we had used, for example, a different type of sampling strategy, such as time-location sampling or some other type of convenience sampling. Since they were also a large group in number, we think their absence from the paper would lead many readers to wonder how they would compare to the recruits in both arms. Further, especially for table 2, a reader can still compare only arm 1 and arm 2 recruits if they prefer, since the arm 1 recruits are the reference group in that table. For these reasons, we feel that it is of benefit to leave these arm 2 seeds in the manuscript and let their data contribute to the paper and our understanding of how recruitment proceeded. If necessary, we could however also consider increasing the columns in table 2 so that, for example, arm 1 recruits are compared individually to arm 2 seeds and then to arm 2 recruits.

I’m unfamiliar with exact logistic regression. Is there a reference?

Sorry, we had planned on including a reference, but neglected to actually do so. We have now added a reference to the paper (reference 22; Ammann 2004). Many software packages likely now include exact regression methods (the cited paper specifically mentioned LogXact 5, but, for example, Stata version 11.1 that we used also now includes it).