Supporting online material

Evaluation of Respondent-Driven Sampling

McCreesh, N et al

Corresponding author: Dr Richard White, Department of Infectious Disease Epidemiology, Faculty of Epidemiology & Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT. Tel: + 44 (0) 20 7299 4626 Email: richard.white@lshtm.ac.uk

Supporting methods

Target population

The data used to define the target population were available from an ongoing general population cohort of 25 villages in rural Masaka, Uganda covering an area of approximately 38km^{2}[1,2] (main text Figure 1). Annually, households in the study villages are mapped and after obtaining consent, a total-population household census and an individual questionnaire are administered and blood taken for HIV-1 testing.

The study villages are in southwestern Uganda, not far from Lake Victoria. The vast majority of dwellings are distributed throughout the countryside rather than clustered in villages, that mainly represent administrative areas demarcated on maps rather than population centres. The study population are mostly subsistence farmers, whose staple diet consists of matooke (cooking bananas) with groundnuts. There are no tarmac roads and access may be difficult during the rains. People live in semi-permanent structures built from locally available materials. Levels of literacy are low and the main income-earning activities are growing bananas, coffee and beans, and trading produce including fish.[3]

The data used in this study to identify the target population (village residence and head of household status) were collated from ongoing general population cohort surveys on 25 villages in rural Masaka carried out during the 12 months immediately prior to the start of the respondent-driven sampling (February 2009 - Jan 2010). Household was defined by the general population cohort staff as a group of people who share food and other resources. Head of household status was self-defined by the members of the household. The characteristics of the target population were estimated for the start date of the respondent-driven sampling (8 March 2010). Data on the tribe, religion and date of birth were collated from any general population cohort survey. Household socioeconomic status was calculated

using principle components analysis from household ownership of 22 items recorded during an annual census (December 2008-October 2009) and categorised into quantiles based on the status of all households in the general population cohort villages. Data on the number of sexual partners in the preceding 12 months were collated from the most recent general population cohort survey round (carried out between December 2009 - October 2010), or if this was unavailable, from the previous survey round (December 2008 - October 2009). HIV testing algorithms and laboratory methods are reported elsewhere[4], briefly, HIV status was determined by two independent immunoassays (Wellcozyme HIV-1 recombinant VK 56/57 (Murex Biotech Ltd, Dartford,Kent, UK) and Recombigen HIV-1/2 (Trinity Biotech plc,Galway, Ireland)), confirmed by western blot (Cambridge Biotech HIV-1 Western blot, Calypte Biomedical Corporation, Rockville, MD, USA). Current infection status was imputed based on earlier positive results or later negative results.[1]

The target population consisted of 2402 men who were recorded as a male head of a household within the study villages between February 2009 and January 2010. Approximately equal proportions were aged under-30, 30-39, 40-49, and 50 or more years old (main text Table 1, 'Population proportion' column). Membership of the four main tribal groups ranged from 72% Muganda to 2% Mukiga. 60% were Catholic, 17% Protestant, and 23% Muslim. The proportion in each village ranged from 2% in village B to 9% in village Q. 42% reported one sexual partner in the preceding year and 6.3% were known to be HIV infected.

The respondent-driven sampling survey

People were eligible for the respondent-driven sampling if they were recorded as a male head of a household within the study villages between February 2009 and January 2010. Three interview sites were placed to minimise the maximum distance between the centre of any eligible village and the nearest interview site (4km) (main text Figure 1).

Seed selection

Ten seeds (number based on a typical number used in respondent-driven sampling studies[5]) were selected from the target population. Total-population and GPS data were available on the target population, but as data of this quality are typically unavailable to researchers using respondent-driven sampling these data were not used to select seeds. Instead it was assumed that during a typical respondent-driven sampling, pre-study mapping of the target population would yield limited information on the approximate geography, age and tribe distribution of the target population (Table S9, left), and this information was used to make a proposal for the variation in these characteristics that would be sought in the seeds. The criteria were that one seed would be from each of ten areas covering the study villages and that two seeds would be in each of five age and five tribe groups (Table S9, right). A list of candidate seeds was then drawn up in consultation with local community leaders by Medical Research Council employees with previous experience of working in these villages. For each of the ten geographic areas shown in Table S9 (right) one of three Medical Research Council employees identified, by convenience, five popular and well known male household heads who were willing to act as study seeds, and who said they were confident that they could recruit other male household heads for the study. The Medical Research Council employees were asked to select a range of male household heads within each area that approximately covered the desired range of ages and tribes. Thus in total a list of 50 male household heads was drawn up (five candidate seeds from each of the ten areas). Stata was then used to randomly select one seed from each of the ten areas. The characteristics of the set of ten candidate seeds were then compared to the criteria. This process was repeated (with replacement) until a set of seeds matching all criteria was identified. This first seed set identified in this way was used to initiate the study.

Seeds were given three coupons to recruit people into the study. All people receiving coupons were instructed that their potential recruits should attend for interviews within seven days, although potential recruits attending after this time were also interviewed. Potential

recruits arriving at the interview sites with valid coupons were assessed for eligibility using their existing general population cohort identity card or reported demographic information. If they were eligible for the study and gave consent, they were enrolled and given a first interview, and are defined as 'recruits' in this paper. In the first interview all recruits were asked to provide details of their relationship with their recruiter and of other male household heads they knew (their 'network'). All recruits were also asked if they wanted to recruit other people. If they accepted, the survey protocol specified that they would be offered three coupons to use to recruit up to three people. However, early in the survey, project staff could not cope with the rapidly increasing number of people who arrived for interviews each day (main text Figure 2a). Therefore, this protocol specification was modified so that the probability of each recruit being offered three coupons, was halved from 100% to 50% from the start of day nine (i.e. 50% were offered zero coupons). When the arrival rate had decreased later in the study (start of day 32), the probability of being offered three coupons was increased from 50% to 100% To close the study the probability of being offered coupons was reduced to 0% when the target sample (900) was about to be reached. Interviews, for those with coupons, continued for another seven days.

If recruits were offered and accepted coupons they were defined as 'recruiters' in this paper. Recruits received one primary incentive for completing the first interview. One 'incentive' was either soap, salt or school books to the value of ~$1US. Recruiters also received one secondary incentive for each person they successfully recruited. Receiving secondary incentives was conditional on also having completed a second interview, during which recruiters were asked to provide details of who they did or did not offer their coupons to, and who accepted or rejected coupons. All recruiters were instructed that they must give out all three coupons before returning to collect their secondary incentives.

The questionnaire was programmed in Access 2003 VBA[6] on Samsung Q1 UMPCs. The protocol ensured interviews could be carried out at any recruitment station by any

interviewer. This was achieved by downloading data from the ten fieldworkers UMPCs each evening; reconciling the data in London; uploading an identical copy of the reconciled database to each UMPC each morning; each potential recruit being instructed that they would not be interviewed until the day after they were given coupons; and each recruiter being instructed that they would not be given a second interview until the day after they (the recruiter) were given coupons to give out. As is typical in respondent-driven sampling studies, members of the target group were prevented from being recruited more than once.

We defined network size in five different ways. The first network size definition (NS1) was created to be comparable with other respondent-driven sampling studies.[7,8] Recruits were first asked the core question *"Baami bameka b'omanyi nga (i) mu myezi kkumi n'ebiri egiyise baali ba nannyinimu mu byalo bya MRC, (ii) era ng'obamanyi nabo bakumanyi, (iii) ng'obalabyeko mu week ewedde?* (*"How many men do you know who (i) were head of a household in the last 12 months in any of the Medical Research Council villages, and (ii) you know them and they know you, and (iii) you have seen them in the past week"*). We also re-asked the core question but asked the recruit to categorise based on residence (own village or not) (NS2) and then by residence and tribe (NS3). Each time the question was re-asked the recruit was reminded of their response to the previous question, but the recruit was not required to reconcile inconsistent responses. We based the final two network size definitions on data collected when the recruits were asked to recall the names and/or other demographic characteristics of each individual eligible member of their network (hereafter called 'individual-level network members'). These details were used by the interviewer to search the general population cohort database (containing details of all men known to the Medical Research Council irrespective of eligibility for the general population cohort or respondent-driven sampling) and attempt unique identification. If the man was positively identified as someone in the general population cohort database (hereafter called 'identified' individual-level network members), this was recorded, else the name/nickname and/or demographic data were recorded for later analysis. Using these data, network size was also

defined as the total number of individual-level network members (NS4), and as the total number of identified individual-level network members who were eligible for the study (NS5). By definition NS5 was a subset of NS4.

Statistical Methods

Pre-processing of the data was performed using Stata v11 (StataCorp, Texas).[9] Networks and trees were generated using scripts written in Stata and R v2.12.0(R Foundation, Vienna)[10] and visualized using GraphViz (AT&T Research, New Jersey).[11] Where possible, to maximise the comparability of our methods with those used in a typical RDS study, we analysed the dataset following current recommended statistical methods[12-14] employing RDSAT v6.0.1,[15] the custom written software package for the analysis of respondent-driven sampling studies.

Simple sample proportions and respondent-driven sampling estimates were calculated for two different sample sizes. The first was the 'Full' sample. The second was a 'Small' sample consisting of the first 250 recruits (including the 10 seeds) and was designed to be more typical of the sample sizes used in respondent-driven sampling studies (a recent systematic review of 123 respondent-driven sampling studies found a median sample size of 247 and a mean sample size of 273[5]).

Recruitment patterns, sample proportions, RDS-1 and RDS-2 estimates and 95% confidence intervals

Current respondent-driven sampling definitions and the statistical inference methods employed by RDSAT were used.[13,14,16-18] Sample proportions were calculated excluding seeds. Respondent-driven sampling 'transition probabilities' were calculated as the proportion of each sub-group's recruits who were in each subgroup e.g. proportion of all the recruits' of Catholics, who were Protestant.[14] Adjusted group network size was calculated by weighting individual network size by the inverse of the individual's network size, i.e. the

respondent-driven sampling 'multiplicity estimate' of group network size using RDSAT terminology.[16]

RDS-1 estimates were calculated using RDSAT by solving the set of simultaneous linear equations relating (using respondent-driven sampling theory) estimated network size, estimated proportions and transition probabilities, using the least squares algorithm.[14] 95% confidence intervals were generated using the modified bootstrap method employed by RDSAT that somewhat mimics the respondent-driven sampling recruitment method.[17] Using this method, for any characteristic, the sample is divided into groups based on which group recruited them e.g. recruited into 3 groups, those recruited by an HIV+, HIV- and HIV-unknown.[17] The seed is then chosen with uniform probability from the entire sample, eg an HIV+ seed. The next person is selected from the group that was recruited by people in the same group as the seed, eg in this example, by HIV+ people. If this new person was HIV- then the next person would be recruited from the group who were recruited by HIV- people, and so on. This continued until the bootstrap sample was the same size as the original sample, and the respondent-driven sampling estimator is applied to the bootstrap sample. For each bootstrap sample, RDS-1 estimates were calculated. The 2.5% and 97.5% percentiles of 20,000 bootstrap samples were used to construct 95% confidence intervals.

Root mean squared errors were calculated for the difference between the population proportions and the full and small sample proportions, and for the difference between the population proportions and the RDS-1 and RDS-2 estimates, for each variable and in total. As RDS-1 estimates could not be calculated for the variable village using the small sample, village was not included in the total root mean squared error for RDS-1 using the small sample. Therefore, the total root mean squared error for the small sample proportions was calculated twice: including village (to allow comparison with the total RDS-2 root mean squared error) and excluding village (to allow valid comparison with the total RDS-1 root mean squared error).

The RDS-2 point estimator weights individual-level data by the reciprocal of their reported network size, to adjust for expected over-recruitment of large-network size individuals.[18] RDS-2 point estimates were calculated, excluding seeds, using R. 95% CIs were estimated using the method described above, with RDS-2 estimates (instead of RDS-1 estimates) calculated for each bootstrap sample.

For comparison with the RDS-1 and RDS-2 estimates, we calculated recruitment probabilities for the target population, including seeds, using predictions from a logistic regression model[19] as weights. The outcome was recruitment into full sample for estimates using data from full sample, and outcome was recruitment into small sample for estimates using data from small sample. Variables were included if they were significant at the 95% confidence level.

Two methods were used to determine whether equilibrium had been reached. The first was based on methods employed by RDSAT.[13,14,16] This method simulates recruitment for a hypothetical sample, assuming that all of the seeds were homogeneous for a variable and using the sample recruitment probabilities to calculate the expected sample proportions in each wave. The numbers of waves required to reach equilibrium for each variable was calculated from this as the number of waves it takes for the proportions in each wave to change by less than 2% relative to the proportions in the wave before. This differed depending on the subgroup chosen for the initial seed and therefore the largest number of waves required was reported. Limitations of this method are that it does not take into account random variation in recruitment or the actual sample proportions by wave. The second method was to calculate recruitment weights as the ratio of the equilibrium proportions to the sample proportions (excluding seeds) for each group.[20] Equilibrium proportions are calculated by simulating recruitment using the sample recruitment probabilities. Recruitment weights that are far from one suggest that the sample has not reached equilibrium for that

group. Equilibrium was assumed to have been reached if the ratio was within the range 0.90 to 1.10.

The mixing pattern between population sub-groups was summarised using the respondent-driven sampling measure 'Homophily'. [14, Equation 19] Homophily (H) was defined to be equal to one if all the recruits of that group were within that group, equal to minus one if all the recruits of that group were outside that group, and equal to zero if the proportion of recruits of that group was equal to the RDS-1 estimate of that group. Our (arbitrary) cut off of for high or low within-group recruitment was $H \geq 0.1$ or $H \leq -0.1$, among groups of size >25. To test the respondent-driven sampling assumption that recruitment is random from the recruiters reported network, expected recruitment matrices were calculated for each variable using data collected in recruiters' first interview on identified individual-level network members who were a member of the target population. The data were weighted by the number of recruits of the recruiter (ie data from recruiters who recruited three recruits were given three times the weight of data from recruiters who only recruited one recruit). Age groups 0-19 and 20-29 were grouped and the category 'Other known/none/unknown' was excluded for religion due to zero values in the expected recruitment matrices. The expected recruitment matrices were compared with the actual recruitment matrices and a chi-squared test was used to test for evidence against random recruitment from the recruiters reported network.

We explored the robustness of our results to any bias in network size estimates caused by under-reporting by re-calculating RDS-1 and RDS-2 estimates for the full sample using network size data from subsets of the sample that were less likely to have been affected by this potential source of bias. These subsets were: 1) Men recruited during the first five weeks of the study (mean network size fell slightly between weeks 5 and 8), 2) Men interviewed at interview sites 1 and 3 only (qualitative data showed that staff at interview site 2 unofficially started requiring their respondents to give at least 10 contacts in response to perceived reductions in reported network size), and 3) Men who responded to the respondent-driven

sampling interview question "*How did your recruiter persuade you to come today?"* by saying that their recruiters had told them nothing about the study, or had told them only about the incentives. There were no recruits from the subgroups age '<20 years' and religion 'Other/none' who reported that they had been given no information about the interview by their recruiters. The mean network sizes for these subgroups were therefore calculated from the reported network sizes of all recruits for subset 3. Estimates were not calculated for the variable village due to the high proportion of villages with few or no recruits meeting the requirements for subsets 1, 2 and 3.

To test for the possibility that biases in the unadjusted and adjusted estimates for the variable socio-economic status were due to an association between socioeconomic status and age and biases in recruitment by age, unadjusted and adjusted estimates for socioeconomic status were calculated separately by age group. Age group 0-19 and 29-29 were grouped due to the small number of recruits aged 0-19. Combined estimates were produced by combining the estimates by age group, weighted according to the population proportions in each age group.

Spatial analysis

Geographic plots were performed in ArcGIS 9.2[21] and distances between villages were calculated using ArcMap as the minimum distance between the main village meeting points along well established paths and roads.

Simple random sample of non- respondent-driven sampling *-recruits*

To compare network size of the whole target population to the respondent-driven sampling recruits, 300 men in the target population who had not been recruited in the respondent-driven sampling study, were randomly selected to be interviewed using the first respondent-driven sampling questionnaire. The size of the eligible population was 1475 (ie 2402 – 927

(the number recruited by respondent-driven sampling). The T-test was used to test for differences between means.

A minimum estimate for the proportion of the target population that were in a single connected network was estimated by calculating the proportion of the target population who were given as a contact by at least one respondent-driven sampling recruit or by at least one member of the simple random sample who was given as a contact by a respondent-driven sampling recruit.

Qualitative survey

To help understand the quantitative study findings 54 members of the population in the study villages or Medical Research Council staff were selected for qualitative interview. The groups sampled, sample sizes, and sampling methods were 1) 10 respondent-driven sampling recruits were randomly selected from 917 eligible (excluding seeds), 2) 10 men who were reported by recruiters as having refused coupons and we knew had not enrolled in the respondent-driven sampling study (refusers) were randomly selected from 29 eligible, 3) 10 community members (men and women) who were not respondent-driven sampling recruits or refusers were randomly selected from 8695 eligible, 4) 10 key informants from the study population were selected purposively, 5) all 10 respondent-driven sampling interviewers were selected for interview, 6) 2 general population cohort census survey staff were randomly selected from 8 eligible, 7) 2 general population cohort medical survey staff were randomly selected from 17 eligible.

Ethical approval

The Science and Ethics Committee of the Uganda Virus Research Institute (GC/I27109108), the Uganda National Council for Science and Technology (SS2278) and the London School of Hygiene and Tropical Medicine Ethics Committee (5585) gave ethical approval for the study.

Supporting Results

Seed selection

All a-priori seed selection criteria (assuming limited knowledge) were met (Table S1). Two seeds were selected from each age and tribe group. The geographic distribution was slightly more uneven than expected when GPS data were used to examine the actual position of seed households (main paper Figure 1, seeds shown as black triangles).

Simple random sample survey

1475 (2402 - 927) men were eligible for the simple random sample. 55% (164/300) completed the interview. The reasons for non-interview are shown in supporting Table S10.

Qualitative survey

54 members of the population in the study villages or Medical Research Council staff were selected for qualitative interview. 53 were interviewed consisting of 10 out of 10 respondent-driven sampling recruits, 10 out of 10 men who were offered coupons but did not enrol in the respondent-driven sampling study (refusers), 10 out of 10 community members (men and women) who were not RDS recruits or refusers, 10 out of 10 key informants, 9 out of 10 respondent-driven sampling interviewers (refusal due to being too busy), 2 out of 2 general population cohort census survey staff, and 2 out of 2 general population cohort medical survey staff. During analysis four refusers were found to have been ineligible and their data were removed from the analysis leaving six valid interviews from this group. The final sample size was 49.

Recruitment pattern

A video illustrating recruitment in space and time is shown in ‘Video1.avi. There was very strong evidence against random recruitment from reported contacts by age ($p<0.001$) (Table S5). Compared to reported contacts, younger men were over-recruited. This is likely to be

due a bias against reporting young men to be household heads, rather than due to a genuine over-recruitment of younger men, as younger men were under-represented in the respondent-driven sampling sample. There was strong evidence that recruitment was not random by tribe ($p<0.001$), with a tendency for tribes that made up a smaller proportion of the eligible population to over-recruit from their own tribe by a larger amount (Mukiga by 300%, Murundi by 67%, Munyanrwanda/kole by 17%, and, in contrast Muganda under-recruited from their own tribe by 6%). There was good evidence against random recruitment by religion ($p=0.01$), due largely to an over-recruitment of Protestants by Muslims. There was strong evidence that recruiters did not recruit randomly by village ($p<0.001$) (Table S6). 11 out of 25 villages over-recruited from their own village. Recruiters in villages with a larger number of eligible villages within 3km tended to over-recruit less (correlation of -0.42, $p=0.04$, supporting Figure S5). Most recruits were recruited by recruiters who lived in the same village (70.6%). 24% were recruited by recruiters who lived in villages within 3km of their village. 5% were recruited by recruiters living in villages more than 3km from their village. A map and recruitment networks showing the recruitment pattern by village are shown in supporting Figure S6. A recruitment network showing whether they were offered and accepted coupons is shown in supporting Figure S7. There was very strong evidence against random recruitment by socioeconomic status ($p<0.001$) with men in the lowest two socioeconomic groups being over-recruited and men in the highest two groups (and men of unknown socioeconomic status) being under-recruited. The over/under-recruitment was greatest for men in the highest and lowest socioeconomic groups and for men of unknown socioeconomic status. There was very strong evidence against random recruitment by number of sexual partners ($p<0.001$), due largely to under-recruitment of people with unknown numbers of partners and over-recruitment of people with zero sexual partners. The over-recruitment may be due to over-recruitment of older men as a higher proportion of older men reported zero sexual partners compared to younger men (23% of 50+ year olds compared to 6% of <50 year olds). There was no evidence against random recruitment for HIV ($p=0.1$)

Comparison with target population data

The root mean squared error for the difference between the true population proportions and the respondent-driven sampling estimates was 6.9% for the RDS-1 estimates and 6.6% for the RDS-2 estimates for the full sample and 7.4% for the RDS-1 and RDS-2 estimates for the small sample (table S7). The root mean error was largest for the variable HIV status for both estimators and sample sizes. It was smallest for religion for the RDS-1 estimates using the full sample and tribe using the small sample, and for village for the RDS-2 estimates using both sample sizes.

Sensitivity to different network size definitions

The RDS-1 adjusted estimates were closer to the true population proportions than the sample proportions were for 36% (19 out of 52) categories for network size definition NS1, 33% (17 out of 52) for definition NS4 and 35% (18 out of 52) for definition NS5 for the full sample, and for 27% (7 out of 26) categories for definition NS1, 35% (9 out of 26) for definition NS4 and 39% (10 out of 26) for definition NS5 for the smaller sample. The RDS-2 adjusted estimates were closer for 33% (17 out of 52) categories for network size definitions NS1, NS4, and NS5 for the full sample, and for 35% (18 out of 52) categories for definition NS1 and for 31% (16 out of 52) for definitions NS4 and NS5 for the smaller sample (supporting Table S11).

Sensitivity of our results to potential bias in network size estimates

Mean network size rose slightly from 11.8 in week one to 13.8 in week five and subsequently fell slightly to 10.3 in week 8. There was very strong evidence for higher mean network size among men interviewed at interview sites 1 and 3 than site 2 (12.8 vs 11.0 $p<0.001$). There was very strong evidence that a higher proportion of recruits reported a network size of exactly 10 at interview site 2 than at sites 1 and 3 (28% vs 15%, $p<0.001$). There was weak evidence for a slightly higher mean network size among recruits whose recruiters told them

that *'there would be questions'* than among recruits whose recruiters had told them nothing about the study, or had told them only about the incentives (12.5 vs 11.7 p = 0.1).

The RDS-1 and RDS-2 estimates for the full sample generated using mean network sizes calculated from subsets of the samples were generally slightly worse than the estimates calculated using network size data from the whole sample (not shown). The exceptions to this were RDS-1 estimates calculated using network size data excluding site 2 and RDS-2 estimates calculated using network size data from the first 5 weeks of the study. In both cases, the RDS estimate was improved for just over half of the estimates (56%, 15/27) by using the subset network size data rather than the whole sample. However there was no evidence that this small improvement was significantly larger than 50% at the 95% confidence level (p=0.6). The other estimates were closer for 33% to 41% (9 to 11 out of 27) of subgroups. This may be due to chance (p=0.08-0.3), or it may be due to the fact that the average network sizes were calculated from fewer observations and were therefore more variable, making the average size of the RDS adjustments larger.

Socio-economic status by age group

40% (2 out of 5) sample proportions for socio-economic status were closer to the true population proportions after controlling for age group (Table S12). After controlling for age group, 100% (5 out of 5) RDS-1 estimates were closer to the true population proportions than the non-age-adjusted RDS-1 estimates were. 40% (2 out of 5) were closer than the age-adjusted sample proportions were. The under-representation of men in the highest socio-economic status group and over-representation of men in the lowest group in both the sample proportions and the RDS-1 estimates remained after adjusting for age group ([population proportion, age-adjusted sample proportion, age-adjusted RDS-1 estimate], highest socio-economic group [26%, 18%, 18%], lowest socio-economic group [21%, 26%, 30%]).

Comment on number of men who were reported to have accepted more than one coupon

Analysis of the data on identified individual-level network members collected from recruiters who had returned for the second interview, showed 92 men had accepted coupons from more than one recruiters (84 from two, seven from three, and one from four). As only 16 men were found to be ineligible due to previous recruitment the majority of these men did not attempt re-recruitment. It is likely that more people in the target population accepted coupons from more than one recruiter because only 66% of recruiters returned for a follow up interview and only 68% of the people in the target population who were given coupons by these recruiters were identified.

Equilibrium

Using the method employed by *RDSAT*, for both sample sizes the number of waves required to reach equilibrium was calculated as four for socio-economic status and five for religion and at least 500 for village when the full sample was used (supporting Table S3 and Table S4). The estimated number of waves differed between the full and small sample size for HIV (three for full and four for small), age group (four for full and three for small), tribe (five for full and seven for small) and number of sexual partners (three for full and four for small). The difference between the values obtained using the two different sample sizes shows one of the problems with this method. There were 16 waves of recruitment in the full sample and 6 waves in the smaller sample and therefore using this method suggests that equilibrium was reached for all variables except village for both sample sizes and possibly tribe for the small sample.

Using the second method, recruitment weights for the full sample ranged between 0.93 and 1.01 for tribe, 0.99 and 1.05 for religion, 1.00 and 1.01 for socioeconomic status, 0.94 and 1.02 for age group, 0.03 and 6.01 for village, 0.97 and 1.01 for HIV status, and 1.00 and 1.00 for number of sexual partners (supporting Table S3 and Table S4). For the smaller sample

they ranged between 0.62 and 1.08 for tribe, 0.99 and 1.05 for religion, 0.97 and 1.04 for socioeconomic status, 0.98 and 1.03 for age group, 0.00 and 13.13 for village, 0.93 and 1.02 for HIV status, and 0.97 and 1.02 for number of sexual partners. This suggests that equilibrium may not have been reached for tribe or village for either sample size.

Respondents all linked in single network

The recruitment networks from each seed were all linked to the same overall network and 73% of the eligible population were linked in a single network. This was likely to be an underestimate as network membership data were unavailable on many members of the target population and also because younger household heads tended not to be perceived as household heads by the target population (only 21% of eligible 0-19 years olds and 54% of eligible 20-29 years olds could be linked to the network compared to 79% of eligible 30+ year olds).

References

1. Shafer LA, Biraro S, Nakiyingi-Miiro J, Kamali A, Ssematimba D, Ouma J, Ojwiya A, Hughes P, Van der Paal L, Whitworth J, Opio A, Grosskurth H. HIV prevalence and incidence are no longer falling in southwest Uganda: evidence from a rural population cohort 1989-2005. *AIDS* 2008;**22**(13):1641-9.
2. Kamali A, Carpenter LM, Whitworth JA, Pool R, Ruberantwari A, Ojwiya A. Seven-year trends in HIV-1 infection rates, and changes in sexual behaviour, among adults in rural Uganda. *AIDS* 2000;**14**(4):427-34.
3. Nakibinge S, Maher D, Katende J, Kamali A, Grosskurth H, Seeley J. Community engagement in health research: two decades of experience from a research project on HIV in rural Uganda. *Trop Med Int Health* 2009;**14**(2):190-5.
4. Mbulaiteye SM, Mahe C, Whitworth JA, Ruberantwari A, Nakiyingi JS, Ojwiya A, Kamali A. Declining HIV-1 incidence and associated prevalence over 10 years in a rural population in south-west Uganda: a cohort study. *Lancet* 2002;**360**(9326):41-6.
5. Malekinejad M, Johnston L, Kendall C, Kerr L, Rifkin M, Rutherford G. Using respondent-driven sampling methodology for HIV biological and behavioral surveillance in international settings: a systematic review. *AIDS and Behavior* 2008;**Volume 12**(S1):105-130.
6. Microsoft Corporation. Microsoft Access 2003. 2003 ed. Washington, 2003.
7. McCarty C, Killworth PD, Bernard HR, Johnsen EC, Shelley GA. Comparing two methods for estimating network size. *Human Organization* 2001;**60**(1):28-39.
8. McCormick T, Salganik M, Zheng T. How many people do you know?: Efficiently estimating personal network size. *Journal of the American Statistical Association* 2010;**105**(489):59-70.
9. StataCorp. Stata Statistical Software: Release 11.0. 9 ed. College Station, Texas: Stata Press, 2010.
10. R Development Core Team. R language and environment for statistical computing and graphics Vienna, Austria: R Foundation for Statistical Computing, http://www.R-project.org., 2010.
11. Gansner ER, North SC. An open graph visualization system and its applications to software engineering. *Softw. Pract. Exper* 1999;**S1**:1-5.
12. Salganik MJ, Heckathorn DD. Sampling and Estimation in Hidden Populations Using Respondent-Driven Sampling. *Sociological Methodology* 2004;**34**(1):193-240.
13. Heckathorn DD. Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations. *Social Problems* 1997;**44**(2):174-199.
14. Heckathorn DD. Respondent-Driven Sampling II: Deriving Valid Population Estimates from Chain-Referral Samples of Hidden Populations. *Social Problems* 2002;**49**(1):11-34.
15. Volz E, Wejnert C, Deganii I, Heckathorn D. Respondent-Driven Sampling Analysis Tool (RDSAT). 6.0.1 ed. Ithaca, NY: Cornell University, 2007.
16. Heckathorn DD. EXTENSIONS OF RESPONDENT-DRIVEN SAMPLING: ANALYZING CONTINUOUS VARIABLES AND CONTROLLING FOR DIFFERENTIAL RECRUITMENT. *Sociological Methodology* 2007;**37**(1):151-207.
17. Salganik MJ. Variance estimation, design effects, and sample size calculations for respondent-driven sampling. *J Urban Health* 2006;**83**(6 Suppl):i98-112.
18. Volz E, Heckathorn D. Probability Based Estimation Theory for Respondent Driven Sampling. *Journal of Official Statistics* 2008;**24**(1):79-97.
19. Kirkwood BR, Sterne JAC. *Essential medical statistics* Wiley-Blackwell, 2003.
20. Frost SD, Brouwer KC, Firestone Cruz MA, Ramos R, Ramos ME, Lozada RM, Magis-Rodriguez C, Strathdee SA. Respondent-driven sampling of injection drug users in two U.S.-Mexico border cities: recruitment dynamics and impact on estimates of HIV and syphilis prevalence. *J Urban Health* 2006;**83**(6 Suppl):i83-97.
21. Environmental Systems Research Institute. ArcGIS. Version 9.2. Redlands, CA.

Table S1 Characteristics and recruitment patterns of the ten seeds. HIV status and sexual activity omitted for confidentiality

								Full sample (n=927, incl seeds)			Small sample (n=250 incl seeds)		
Seed	Age	Tribe	Religion	Village	Socio-economic status	Reported network size	Number of recruits	Number of waves	Total number of recruits	Percent recruited	Number of waves	Total number of recruits	Percent recruited
1	27	Munyanrwanda/kole	Catholic	C	Highest	7	2	13	157	17.1	6	30	12.5
2	27	Other known	Catholic	A	Lower	3	2	3	8	0.9	3	4	1.7
3	34	Other known	Catholic	F	Higher	5	3	9	32	3.5	2	6	2.5
4	33	Muganda	Protestant	H	Higher	6	2	16	177	19.3	6	65	27.1
5	42	Mukiga	Catholic	I	Highest	20	3	11	241	26.3	6	48	20.0
6	40	Murundi	Catholic	M	Lowest	20	3	16	129	14.1	6	32	13.3
7	74	Murundi	Muslim	O	Lower	17	3	6	22	2.4	3	10	4.2
8	19	Munyanrwanda/kole	Protestant	Q	Higher	9	2	8	53	5.8	5	27	11.3
9	63	Mukiga	Protestant	T	Unknown	10	3	11	57	6.2	3	14	5.8
10	18	Muganda	Catholic	V	Lowest	31	1	9	41	4.5	2	4	1.7

Table S2 The correlation coefficients among RDS recruits between the five measures of network used in the study (including seeds). $p<0.0001$ in all cases

	1	*2*	*3*	*4*
5	0.754	0.778	0.880	0.904
4	0.821	0.850	0.958	
3	0.840	0.886		
2	0.963			

Table S3 Recruitment matrices and other characteristics of the RDS sample for age, tribe, religion, socioeconomic status, sexual activity and HIV status, for the full and small sample. Table shows sample proportions, equilibrium proportions, recruitment weights, unadjusted and adjusted network sizes, homophily and wave at which equilibrium was estimated to have been reached using the RDSAT method. Recruitment weights (indicating 'equilibrium' had been reached using the Frost Method [20]) are shown in bold if they lie between 0.90 and 1.10. Sample size for age group is 238 rather than 240 because the two seeds in age group 0-19 were excluded to allow estimates to be calculated.

		Full RDS sample (n=927 including seeds)						Small RDS sample (n=250 including seeds)					
		Age group of recruits						Age group of recruits					
Age group of recruiters		**0-19**	**20-29**	**30-39**	**40-49**	**50+**	**Total**	**0-19**	**20-29**	**30-39**	**40-49**	**50+**	**Total**
0-19		0 (0.000)	0 (0.000)	1 (0.250)	1 (0.250)	2 (0.500)	4	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.333)	2 (0.667)	3
20-29		0 (0.000)	12 (0.167)	22 (0.306)	15 (0.208)	23 (0.319)	72	0 (0.000)	1 (0.100)	2 (0.200)	3 (0.300)	4 (0.400)	10
30-39		0 (0.000)	39 (0.163)	67 (0.280)	55 (0.230)	78 (0.326)	239	0 (0.000)	11 (0.159)	22 (0.319)	16 (0.232)	20 (0.290)	69
40-49		1 (0.004)	26 (0.104)	65 (0.261)	68 (0.273)	89 (0.357)	249	0 (0.000)	6 (0.075)	21 (0.263)	24 (0.300)	29 (0.363)	80
50+		3 (0.008)	45 (0.127)	74 (0.210)	81 (0.229)	150 (0.425)	353	0 (0.000)	7 (0.090)	19 (0.244)	15 (0.192)	37 (0.474)	78
Total (Sample proportion)		4 (0.004)	122 (0.133)	229 (0.250)	220 (0.240)	342 (0.373)	917	0 (0.000)	25 (0.105)	64 (0.269)	59 (0.248)	92 (0.387)	238
Equilibrium proportion		0.004	0.136	0.253	0.237	0.370	1.000	-	0.106	0.263	0.240	0.391	1.000
Recruitment weight		**0.939**	**1.021**	**1.013**	**0.989**	**0.992**		-	**1.001**	**0.975**	**0.980**	**1.030**	
Mean network size	**: Unadjusted**	9.750	11.926	12.459	13.050	11.307		-	9.240	11.469	11.441	8.674	
	: Adjusted	9.385	10.210	10.165	10.295	9.328		-	9.240	11.469	11.441	8.674	
Homophily		-1.000	0.039	0.050	0.062	0.051		-	-0.042	0.090	0.088	0.107	
Equilibrium at wave number		4						3					

		Tribe of recruits						Tribe of recruits					
Tribe of recruiters		**Muganda**	**M'r/kole**	**Mukiga**	**Murundi**	**Other**	**Total**	**Muganda**	**M'r/kole**	**Mukiga**	**Murundi**	**Other**	**Total**
Muganda		443 (0.752)	93 (0.158)	5 (0.008)	25 (0.042)	23 (0.039)	589	112 (0.824)	12 (0.088)	0 (0.000)	7 (0.051)	5 (0.037)	136
M'rwanda/kole		96 (0.475)	72 (0.356)	9 (0.045)	17 (0.084)	8 (0.040)	202	21 (0.375)	16 (0.286)	8 (0.143)	8 (0.143)	3 (0.054)	56
Mukiga		13 (0.542)	4 (0.167)	3 (0.125)	2 (0.083)	2 (0.083)	24	7 (0.583)	2 (0.167)	0 (0.000)	1 (0.083)	2 (0.167)	12
Murundi		31 (0.525)	16 (0.271)	0 (0.000)	11 (0.186)	1 (0.017)	59	9 (0.429)	7 (0.333)	0 (0.000)	5 (0.238)	0 (0.000)	21
Other known/ unknown		29 (0.674)	8 (0.186)	2 (0.047)	1 (0.023)	3 (0.070)	43	8 (0.533)	3 (0.200)	1 (0.067)	1 (0.067)	2 (0.133)	15
Total (Sample proportion)		612 (0.667)	193 (0.210)	19 (0.021)	56 (0.061)	37 (0.040)	917	157 (0.654)	40 (0.167)	9 (0.038)	22 (0.092)	12 (0.050)	240
Equilibrium proportion		0.674	0.207	0.019	0.060	0.040	1.000	0.709	0.143	0.023	0.081	0.043	1.000
Recruitment weight		**1.010**	**0.984**	**0.927**	**0.978**	**0.989**		**1.084**	0.858	0.622	0.884	0.868	
Mean network size	**: Unadjusted**	12.364	11.440	11.316	10.911	13.081		10.261	9.325	10.444	9.818	12.000	
	: Adjusted	9.968	9.785	9.263	8.716	11.080		7.170	7.345	7.607	6.675	9.527	
Homophily		0.265	0.173	0.111	0.130	0.036		0.384	0.157	-1.000	0.166	0.100	
Equilibrium at wave number		5						7					

		Religion of recruits					Religion of recruits				
Religion of recruiters		**Catholic**	**Protestant**	**Muslim**	**Other**	**Total**	**Catholic**	**Protestant**	**Muslim**	**Other**	**Total**
Catholic		427 (0.723)	81 (0.137)	82 (0.139)	1 (0.002)	591	144 (0.787)	17 (0.093)	21 (0.115)	1 (0.005)	183
Protestant		70 (0.534)	41 (0.313)	20 (0.153)	0 (0.000)	131	14 (0.667)	3 (0.143)	3 (0.143)	1 (0.048)	21
Muslim		73 (0.382)	34 (0.178)	83 (0.435)	1 (0.005)	191	16 (0.471)	4 (0.118)	14 (0.412)	0 (0.000)	34
Other known/ none/ unknown		3 (0.750)	1 (0.250)	0 (0.000)	0 (0.000)	4	2 (1.000)	0 (0.000)	0 (0.000)	0 (0.000)	2
Total (Sample proportion)		573 (0.625)	157 (0.171)	185 (0.202)	2 (0.002)	917	176 (0.733)	24 (0.100)	38 (0.158)	2 (0.008)	240
Equilibrium proportion		0.620	0.177	0.200	0.003	1.000	0.724	0.101	0.166	0.009	1.000
Recruitment weight		**0.994**	**1.032**	**0.991**	**1.053**		**0.987**	**1.012**	**1.047**	**1.053**	
Mean network size	**: Unadjusted**	11.853	12.248	12.757	7.333		9.761	10.708	11.789	7.500	
	: Adjusted	9.547	11.021	10.072	7.304		7.073	8.585	7.403	7.467	
Homophily		0.230	0.191	0.287	-1.000		0.180	0.062	0.292	-1.000	
Equilibrium at wave number		5					5				

		Socioeconomic status of recruits						Socioeconomic status of recruits					
Socioeconomic status of recruiters		**Highest**	**Higher**	**Lower**	**Lowest**	**Unknown**	**Total**	**Highest**	**Higher**	**Lower**	**Lowest**	**Unknown**	**Total**
Highest		47 (0.281)	46 (0.275)	38 (0.228)	26 (0.156)	10 (0.060)	167	16 (0.400)	9 (0.225)	4 (0.100)	8 (0.200)	3 (0.075)	40
Higher		41 (0.178)	62 (0.270)	70 (0.304)	54 (0.235)	3 (0.013)	230	12 (0.179)	20 (0.299)	20 (0.299)	15 (0.224)	0 (0.000)	67
Lower		39 (0.160)	57 (0.234)	66 (0.270)	74 (0.303)	8 (0.033)	244	9 (0.143)	18 (0.286)	15 (0.238)	17 (0.270)	4 (0.063)	63
Lowest		28 (0.116)	52 (0.215)	67 (0.277)	83 (0.343)	12 (0.050)	242	7 (0.127)	14 (0.255)	14 (0.255)	17 (0.309)	3 (0.055)	55
Unknown		9 (0.265)	5 (0.147)	11 (0.324)	7 (0.206)	2 (0.059)	34	4 (0.267)	1 (0.067)	4 (0.267)	4 (0.267)	2 (0.133)	15
Total (Sample proportion)		164 (0.179)	222 (0.242)	252 (0.275)	244 (0.266)	35 (0.038)	917	48 (0.200)	62 (0.258)	57 (0.238)	61 (0.254)	12 (0.050)	240
Equilibrium proportion		0.178	0.241	0.275	0.267	0.038	1.000	0.208	0.257	0.231	0.253	0.051	1.000
Recruitment weight		**0.996**	**0.997**	**1.000**	**1.005**	**1.004**		**1.040**	**0.996**	**0.971**	**0.996**	**1.017**	
Mean network size	**: Unadjusted**	12.518	12.572	11.770	11.525	13.229		10.208	11.806	8.649	9.443	12.250	
	: Adjusted	10.359	10.048	10.088	9.048	11.433		7.715	7.580	7.061	6.484	10.003	
Homophily		0.137	0.050	0.006	0.058	0.026		0.259	0.072	-0.036	0.040	0.097	
Equilibrium at wave number		4						4					

		Number of sexual partners in the past year of recruits						Number of sexual partners in the past year of recruits					
Number of sexual partners in the past year of		**0**	**1**	**2-3**	**4+**	**Unknown**	**Total**	**0**	**1**	**2-3**	**4+**	**Unknown**	**Total**
0		24 (0.200)	64 (0.533)	19 (0.158)	3 (0.025)	10 (0.083)	120	4 (0.118)	20 (0.588)	5 (0.147)	0 (0.000)	5 (0.147)	34
1		81 (0.149)	319 (0.589)	73 (0.135)	18 (0.033)	51 (0.094)	542	18 (0.142)	74 (0.583)	20 (0.157)	4 (0.031)	11 (0.087)	127
2-3		19 (0.140)	78 (0.574)	21 (0.154)	4 (0.029)	14 (0.103)	136	7 (0.184)	19 (0.500)	7 (0.184)	2 (0.053)	3 (0.079)	38
4+		5 (0.147)	19 (0.559)	3 (0.088)	2 (0.059)	5 (0.147)	34	1 (0.100)	7 (0.700)	2 (0.200)	0 (0.000)	0 (0.000)	10
Unknown		7 (0.082)	49 (0.576)	12 (0.141)	5 (0.059)	12 (0.141)	85	2 (0.065)	14 (0.452)	5 (0.161)	2 (0.065)	8 (0.258)	31
Total (Sample proportion)		136 (0.148)	529 (0.577)	128 (0.140)	32 (0.035)	92 (0.100)	917	32 (0.133)	134 (0.558)	39 (0.163)	8 (0.033)	27 (0.113)	240
Equilibrium proportion		0.149	0.576	0.140	0.035	0.100	1.000	0.136	0.560	0.162	0.033	0.109	1.000
Recruitment weight		**1.003**	**0.998**	**1.003**	**1.001**	**1.000**		**1.016**	**1.002**	**0.998**	**0.997**	**0.973**	
Mean network size	**: Unadjusted**	11.037	12.098	13.133	11.469	12.348		8.719	9.709	13.000	10.375	9.926	
	: Adjusted	9.113	9.914	10.803	8.727	10.070		6.946	7.094	8.382	6.676	7.233	
Homophily		0.036	0.038	0.033	0.021	0.052		0.169	0.022	0.043	-1.000	0.168	
Equilibrium at wave number		3						4					

		HIV status of recruits				HIV status of recruits			
HIV status of recruiters		**Positive**	**Negative**	**Unknown**	**Total**	**Positive**	**Negative**	**Unknown**	**Total**
Positive		13 (0.148)	70 (0.795)	5 (0.057)	88	6 (0.231)	19 (0.731)	1 (0.038)	26
Negative		58 (0.078)	608 (0.816)	79 (0.106)	745	12 (0.066)	150 (0.820)	21 (0.115)	183
Unknown		1 (0.012)	71 (0.845)	12 (0.143)	84	0 (0.000)	26 (0.839)	5 (0.161)	31
Total (Sample proportion)		72 (0.079)	749 (0.817)	96 (0.105)	917	18 (0.075)	195 (0.813)	27 (0.113)	240
Equilibrium proportion		0.076	0.818	0.106	1.000	0.070	0.816	0.115	1.000
Recruitment weight		**1.001**	**0.970**	**1.014**		**1.004**	**0.927**	**1.020**	
Mean network size	**: Unadjusted**	13.111	12.011	11.927		10.556	10.092	10.370	
	: Adjusted	10.466	9.834	9.708		7.017	7.210	7.783	
Homophily		0.078	-0.005	0.043		0.168	-0.001	0.064	
Equilibrium at wave number		3				4			

Table S4 Recruitment matrices for the characteristics village; equilibrium distributions, recruitment weights, network sizes and homophily of each group. '-' = could not be calculated.

Full RDS sample (n=927 including seeds)

	Village of recruits																								
Village of recruiters	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Total**
A	9 (0.600)	1 (0.067)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.067)	0 (0.000)	1 (0.067)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.067)	0 (0.000)	0 (0.000)	1 (0.067)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.067)	15
B	2 (0.133)	10 (0.667)	2 (0.133)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.067)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	15
C	0 (0.000)	5 (0.066)	57 (0.750)	6 (0.079)	0 (0.000)	1 (0.013)	5 (0.066)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	2 (0.026)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	76
D	0 (0.000)	0 (0.000)	4 (0.085)	27 (0.574)	0 (0.000)	1 (0.021)	3 (0.064)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	4 (0.085)	8 (0.170)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	47
E	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	22 (0.917)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.042)	1 (0.042)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	24
F	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	5 (0.455)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.091)	0 (0.000)	2 (0.182)	2 (0.182)	0 (0.000)	1 (0.091)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	11
G	0 (0.000)	0 (0.000)	1 (0.040)	6 (0.240)	0 (0.000)	1 (0.040)	16 (0.640)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.040)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	25
H	2 (0.143)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	7 (0.500)	0 (0.000)	5 (0.357)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	14
I	1 (0.025)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.025)	0 (0.000)	29 (0.725)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.025)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	3 (0.075)	4 (0.100)	1 (0.025)	40
J	2 (0.061)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	2 (0.061)	0 (0.000)	20 (0.606)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	9 (0.273)	33
K	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	31 (0.912)	1 (0.029)	0 (0.000)	0 (0.000)	2 (0.059)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	34
L	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	2 (0.045)	28 (0.636)	3 (0.068)	6 (0.136)	4 (0.091)	0 (0.000)	0 (0.000)	1 (0.023)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	44
M	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	8 (0.229)	24 (0.686)	3 (0.086)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	35
N	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.067)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	2 (0.133)	12 (0.800)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	15
O	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	6 (0.079)	5 (0.066)	1 (0.013)	11 (0.145)	47 (0.618)	0 (0.000)	3 (0.039)	2 (0.026)	0 (0.000)	1 (0.013)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	76
P	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.063)	0 (0.000)	2 (0.125)	1 (0.063)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	10 (0.625)	2 (0.125)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	16
Q	0 (0.000)	0 (0.000)	0 (0.000)	2 (0.049)	2 (0.049)	3 (0.073)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.024)	1 (0.024)	3 (0.073)	26 (0.634)	0 (0.000)	3 (0.073)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	41
R	2 (0.042)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	2 (0.042)	0 (0.000)	0 (0.000)	0 (0.000)	2 (0.042)	0 (0.000)	0 (0.000)	41 (0.854)	1 (0.021)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	48
S	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	2 (0.067)	23 (0.767)	5 (0.167)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	30
T	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.017)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.017)	1 (0.017)	7 (0.119)	49 (0.831)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	59
U	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.021)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	2 (0.043)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.021)	36 (0.766)	2 (0.043)	0 (0.000)	1 (0.021)	4 (0.085)	47
V	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	36 (0.923)	3 (0.077)	0 (0.000)	0 (0.000)	39
W	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	18 (0.900)	2 (0.100)	0 (0.000)	20
X	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	9 (0.214)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	2 (0.048)	6 (0.143)	25 (0.595)	0 (0.000)	42
Y	1 (0.014)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.014)	0 (0.000)	0 (0.000)	1 (0.014)	5 (0.070)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	23 (0.324)	1 (0.014)	0 (0.000)	0 (0.000)	39 (0.549)	71
Total (Sample proportion)	19 (0.021)	16 (0.017)	64 (0.070)	43 (0.047)	25 (0.027)	15 (0.016)	26 (0.028)	9 (0.010)	41 (0.045)	31 (0.034)	41 (0.045)	43 (0.047)	30 (0.033)	35 (0.038)	57 (0.062)	21 (0.023)	43 (0.047)	50 (0.055)	35 (0.038)	56 (0.061)	60 (0.065)	41 (0.045)	30 (0.033)	32 (0.035)	54 (0.059)	917
Equilibrium proportion	0.007	0.002	0.005	0.008	0.164	0.007	0.005	0.000	0.010	0.002	0.126	0.088	0.102	0.190	0.065	0.010	0.020	0.049	0.033	0.038	0.008	0.012	0.033	0.011	0.005	
Recruitment weight	0.343	0.133	0.065	0.161	6.012	0.426	0.162	0.028	0.227	0.067	2.824	1.870	3.133	4.987	**1.043**	0.431	0.423	0.896	0.859	0.627	0.127	0.275	**1.004**	0.317	0.077	
Mean network size : Unadjusted	11.421	13.688	13.594	13.000	12.280	13.000	11.308	11.000	11.317	9.129	13.146	10.465	11.333	13.343	11.825	12.286	13.767	14.020	12.086	12.286	11.783	12.073	13.233	11.219	9.037	
: Adjusted	8.126	10.413	11.550	11.496	10.918	12.33	9.203	7.796	10.424	7.432	12.081	8.325	9.172	9.288	9.177	11.256	11.268	11.991	9.379	10.048	10.131	10.269	11.693	10.347	7.059	
Homophily	0.587	0.661	0.743	0.566	0.910	0.448	0.635	0.498	0.712	0.600	0.906	0.627	0.681	0.794	0.608	0.616	0.621	0.845	0.739	0.801	0.750	0.920	0.894	0.583	0.535	
Equilibrium at wave number	>500																									

Small RDS sample (n=250 including seeds)

	Village of recruits																								
Village of recruiters	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Total**
A	5 (0.556)	1 (0.111)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.111)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.111)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.111)	9
B	0 (0.000)	2 (0.667)	1 (0.333)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	3
C	0 (0.000)	2 (0.074)	23 (0.852)	1 (0.037)	0 (0.000)	0 (0.000)	1 (0.037)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	27
D	0 (0.000)	0 (0.000)	1 (0.500)	1 (0.500)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	2
E	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	2 (0.400)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.200)	0 (0.000)	1 (0.200)	1 (0.200)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	5
G	0 (0.000)	0 (0.000)	0 (0.000)	2 (0.143)	0 (0.000)	1 (0.071)	10 (0.714)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.071)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	14
H	2 (0.200)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	4 (0.400)	0 (0.000)	4 (0.400)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	10
I	1 (0.125)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.125)	0 (0.000)	2 (0.250)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.125)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.125)	1 (0.125)	1 (0.125)	8
J	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.050)	0 (0.000)	10 (0.500)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	9 (0.450)	20
K	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	11 (0.733)	1 (0.067)	0 (0.000)	2 (0.133)	0 (0.000)	0 (0.000)	1 (0.067)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	15
M	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	4 (0.182)	16 (0.727)	2 (0.091)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	22
N	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	3 (1.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	3
O	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.053)	0 (0.000)	2 (0.105)	16 (0.842)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	19
P	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	2 (1.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	2
Q	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.100)	1 (0.100)	1 (0.100)	4 (0.400)	0 (0.000)	3 (0.300)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	10
R	2 (1.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	2
S	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (0.077)	12 (0.923)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	13
T	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	2 (0.154)	11 (0.846)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	13
U	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1 (1.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	1
V	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	4 (1.000)	0 (0.000)	0 (0.000)	0 (0.000)	4
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Y	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	4 (0.105)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	0 (0.000)	9 (0.237)	0 (0.000)	0 (0.000)	0 (0.000)	25 (0.658)	38
Total (Sample proportion)	10 (0.042)	5 (0.021)	25 (0.104)	4 (0.017)	0 (0.000)	3 (0.013)	12 (0.050)	5 (0.021)	2 (0.008)	18 (0.075)	0 (0.000)	17 (0.071)	17 (0.071)	9 (0.038)	19 (0.079)	5 (0.021)	6 (0.025)	3 (0.013)	17 (0.071)	11 (0.046)	10 (0.042)	4 (0.017)	1 (0.004)	1 (0.004)	36 (0.150)	240
Equilibrium proportion	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.334	0.000	0.274	0.000	0.000	0.000	0.000	0.390	0.000	0.000	0.000	0.000	
Recruitment weight	0.000	0.018	0.015	0.011	-	0.002	0.004	0.000	0.000	0.000	-	0.000	0.000	8.902	0.000	13.131	0.000	0.000	0.001	0.000	9.363	0.000	0.000	0.000	0.000	
Mean network size : Unadjusted	11.600	9.400	12.680	12.250	-	13.000	10.917	11.000	15.000	7.389	-	7.765	9.882	16.667	8.737	10.800	12.667	10.000	12.882	9.727	7.900	5.750	19.000	6.000	8.111	
: Adjusted	7.083	6.853	10.376	8.981	-	12.389	7.825	6.442	13.333	6.022	-	6.279	7.818	6.653	7.137	9.367	9.474	6.096	9.420	6.945	5.917	5.393	19.000	6.000	6.210	
Homophily	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Equilibrium at wave number	-																									

Table S5 Observed and expected recruitment matrices. Expected recruitment matrices were calculated from data on identified individual-level network members. P-values are calculated using a chi-squared test and indicate the strength of evidence against random recruitment. Category 'Other known/none/unknown' was excluded for religion due to zero values in the expected recruitment matrices.

Observed recruitment

Recruiter	Recruit: 0-29	30-39	40-49	50+
0-29	12	23	16	25
30-39	39	67	55	78
40-49	27	65	68	89
50+	48	74	81	150

Recruiter	Recruit: Muganda	Murwanda/kole	Mukiga	Murundi	Other
Muganda	443	93	5	25	23
Murwanda/kole	96	72	9	17	8
Mukiga	13	4	3	2	2
Murundi	31	16	0	11	1
Other	29	8	2	1	3

Recruiter	Recruit: Catholic	Protestant	Muslim
Catholic	427	81	82
Protestant	69	41	20
Muslim	73	34	83

Recruiter	Recruit: Highest	Higher	Lower	Lowest	Unknown
Highest	47	46	38	26	10
Higher	41	62	70	54	3
Lower	39	57	66	74	8
Lowest	28	52	67	83	12
Unknown	9	5	11	7	2

Recruiter	Recruit: 0	1	2-3	4+	Unknown
0	24	64	19	3	10
1	81	319	73	18	51
2-3	19	78	21	4	14
4+	5	19	3	2	5
Unknown	7	49	12	5	12

Recruiter	Recruit: Positive	Negative	Unknown
Positive	13	66	9
Negative	51	548	110
Unknown	5	92	23

Expected recruitment

Recruiter	Recruit: 0-29	30-39	40-49	50+	
0-29	8.6	22.6	20.1	24.7	
30-39	19.7	71.5	67.2	80.6	**p<0.001**
40-49	13.4	58.6	77.2	99.8	
50+	17.6	64.3	92.1	179.0	

Recruiter	Recruit: Muganda	Murwanda/kole	Mukiga	Murundi	Other	
Muganda	473.6	68.5	6.8	26.4	13.7	
Murwanda/kole	112.6	61.5	4.8	17.8	5.4	**p<0.001**
Mukiga	13.7	7.2	0.8	1.3	1.1	
Murundi	35.0	13.4	1.2	6.6	2.7	
Other	29.2	6.6	0.8	4.9	1.5	

Recruiter	Recruit: Catholic	Protestant	Muslim	
Catholic	443.9	71.2	74.9	**p=0.01**
Protestant	77.3	35.1	17.7	
Muslim	77.3	21.7	91.0	

Recruiter	Recruit: Highest	Higher	Lower	Lowest	Unknown	
Highest	63.2	48.3	36.4	14.3	4.8	
Higher	68.3	70.7	51.0	35.9	4.1	**p<0.001**
Lower	59.5	72.0	67.1	38.8	6.6	
Lowest	58.6	68.5	64.4	44.4	6.0	
Unknown	8.7	8.0	11.0	5.6	0.7	

Recruiter	Recruit: 0	1	2-3	4+	Unknown	
0	17.6	59.7	13.9	3.2	25.6	
1	52.7	291.9	86.4	24.4	86.6	**p<0.001**
2-3	9.6	73.0	24.4	4.5	24.4	
4+	2.2	17.6	5.5	1.8	6.9	
Unknown	6.2	44.9	14.1	2.9	16.9	

Recruiter	Recruit: Positive	Negative	Unknown	
Positive	8.4	68.9	10.7	**p=0.1**
Negative	51.8	536.6	120.5	
Unknown	8.6	83.2	28.3	

Table S6. Observed and expected recruitment from own vs. other village. Expected recruitment was calculated from data on identified individual-level network members. P-values are calculated using a chi-squared test and indicate the strength of evidence against random recruitment

	Observed		Expected	
Village	**Own**	**Other**	**Own**	**Other**
A	9	6	12.2	2.8
B	10	5	6.6	8.4
C	57	19	62.5	13.5
D	27	20	32.8	14.2
E	22	2	17.6	6.4
F	5	6	6.1	4.9
G	16	9	14.2	10.8
H	7	7	8.6	5.4
I	29	11	32.0	8.0
J	20	13	23.4	9.6
K	31	3	25.5	8.5
L	28	16	31.2	12.8
M	24	11	23.4	11.6
N	12	3	6.7	8.3
O	47	29	49.5	26.5
P	10	6	11.9	4.1
Q	26	15	24.7	16.3
R	41	7	32.8	15.2
S	23	7	17.4	12.6
T	49	10	50.7	8.3
U	36	11	40.3	6.7
V	36	3	35.2	3.8
W	18	2	16.4	3.6
X	25	17	29.0	13.0
Y	39	32	44.5	26.5

$p<0.001$

Table S7 Root mean squared error for the difference between the true population proportions and the sample proportions and RDS estimates

'-' indicated that the RDS estimates could not be calculated

	Full sample			Small sample		
	Sample	RDS1	RDS2	Sample	RDS1	RDS2
Age group (years)	4.99%	5.60%	5.77%	6.16%	6.99%	6.90%
Tribe	2.20%	2.79%	2.63%	2.99%	2.52%	3.17%
Religion	1.81%	2.51%	2.88%	8.35%	8.72%	9.51%
Socio-economic status	4.72%	6.00%	5.54%	3.17%	4.35%	4.51%
Village	1.75%	3.26%	1.95%	3.96%	-	4.26%
Number of sex partners in the last year	12.10%	12.32%	12.15%	11.27%	11.73%	11.46%
HIV status	18.40%	18.54%	18.42%	17.89%	18.58%	18.42%
Total	6.35%	6.87%	6.56%	7.00% (8.88% excluding village)	- 7.40%	7.44%

Table S8 Target population proportions, full and small sample proportions, and regression-weight adjusted estimates with 95% confidence intervals (CIs). Regression-weight adjusted point estimates are shown in bold if they are closer to the target population proportions than the unadjusted sample proportions. CIs are shown in bold if they include the population proportion. '-' = could not be calculated. Full sample regression model included all variables shown except religion. Small sample model regression model included all variables except religion, village, and socioeconomic status. Village was excluded from the small sample regression model because no-one was recruited from two villages in the small sample and therefore everyone in those villages would have been excluded from the regression model if it had been included.

		Population proportion	Full sample			Small sample		
			Sample	Regression weight adjusted	95% CI	Sample	Regression weight adjusted	95% CI
Age group (years)	0-19	0.020	0.004	**0.019**	**0.004-0.044**	0.000	-	-
	20-29	0.202	0.133	**0.173**	**0.141-0.211**	0.104	**0.208**	-
	30-39	0.275	0.250	**0.287**	**0.248-0.343**	0.267	**0.249**	-
	40-49	0.207	0.240	**0.236**	**0.191-0.274**	0.246	**0.247**	-
	50+	0.297	0.373	**0.284**	**0.243-0.320**	0.383	**0.296**	-
Tribe	Muganda	0.718	0.667	**0.689**	**0.642-0.744**	0.654	**0.731**	**0.675-0.850**
	M'rwanda/kole	0.185	0.210	**0.198**	**0.156-0.243**	0.167	**0.166**	**0.086-0.231**
	Mukiga	0.018	0.021	**0.024**	**0.008-0.042**	0.038	**0.021**	**0.002-0.025**
	Murundi	0.048	0.061	**0.046**	**0.029-0.062**	0.092	**0.039**	**0.016-0.055**
	Other known/ unknown	0.031	0.040	**0.043**	**0.027-0.063**	0.053	**0.043**	-
Religion	Catholic	0.600	0.624	**0.595**	**0.534-0.643**	0.733	**0.602**	**0.491-0.697**
	Protestant	0.171	0.171	**0.176**	**0.147-0.230**	0.100	**0.152**	**0.092-0.238**
	Muslim	0.228	0.202	**0.216**	**0.167-0.258**	0.158	**0.238**	**0.143-0.34**
	Other known/ none/ unknown	0.002	0.003	**0.014**	-	0.008	**0.009**	-
Socio-economic status	Highest	0.257	0.179	**0.276**	**0.219-0.328**	0.200	**0.249**	**0.164-0.351**
	Higher	0.249	0.243	**0.231**	**0.200-0.277**	0.260	**0.262**	**0.190-0.360**
	Lower	0.229	0.274	**0.227**	**0.192-0.261**	0.236	**0.217**	**0.142-0.278**
	Lowest	0.214	0.265	**0.202**	**0.172-0.236**	0.252	**0.234**	**0.168-0.303**
	Unknown	0.052	0.039	**0.064**	**0.035-0.096**	0.052	**0.038**	**0.015-0.062**
Village	A	0.033	0.032	**0.038**	-	0.042	**0.036**	-
	B	0.017	0.017	**0.013**	-	0.021	**0.011**	-
	C	0.042	0.028	**0.042**	-	0.104	**0.116**	-
	D	0.033	0.019	**0.038**	-	0.017	**0.012**	-
	E	0.027	0.072	**0.024**	-	0.000	-	-
	F	0.067	0.013	**0.061**	-	0.013	**0.010**	-
	G	0.025	0.012	**0.022**	-	0.050	**0.044**	-
	H	0.031	0.004	**0.014**	-	0.021	**0.017**	-
	I	0.060	0.047	**0.067**	-	0.008	**0.010**	-
	J	0.028	0.014	**0.029**	-	0.075	**0.101**	-
	K	0.031	0.060	**0.032**	-	0.000	-	-
	L	0.040	0.026	**0.041**	-	0.071	**0.079**	-
	M	0.026	0.016	**0.023**	-	0.071	**0.049**	-
	N	0.033	0.030	**0.036**	-	0.038	**0.058**	-
	O	0.049	0.026	**0.048**	-	0.079	**0.072**	-
	P	0.034	0.024	**0.029**	-	0.021	**0.014**	-
	Q	0.086	0.034	**0.070**	-	0.025	**0.026**	-
	R	0.038	0.061	**0.042**	-	0.013	**0.018**	-
	S	0.038	0.107	**0.053**	-	0.071	**0.061**	-
	T	0.050	0.147	**0.049**	-	0.046	**0.027**	-
	U	0.050	0.064	**0.050**	-	0.042	**0.047**	-
	V	0.039	0.034	**0.035**	-	0.017	**0.014**	-
	W	0.040	0.054	**0.047**	-	0.004	**0.003**	-
	X	0.043	0.030	**0.055**	-	0.004	**0.005**	-
	Y	0.041	0.030	**0.044**	-	0.150	**0.171**	-
Number of sexual partners in last year	0	0.113	0.148	**0.119**	**0.099-0.150**	0.133	**0.107**	**0.072-0.151**
	1	0.419	0.577	**0.459**	**0.410-0.504**	0.558	**0.458**	**0.370-0.546**
	2-3	0.114	0.140	**0.132**	**0.105-0.161**	0.163	**0.133**	**0.084-0.179**
	4+	0.037	0.035	**0.035**	**0.022-0.050**	0.033	**0.037**	**0.015-0.090**
	Unknown	0.316	0.100	**0.256**	0.198-0.310	0.113	**0.265**	-
HIV status	Positive	0.063	0.079	**0.070**	**0.051-0.094**	0.075	**0.060**	**0.026-0.103**
	Negative	0.600	0.817	**0.683**	0.629-0.735	0.813	**0.670**	**0.581-0.774**
	Unknown	0.337	0.105	**0.247**	0.193-0.304	0.113	**0.270**	**0.163-0.365**

	Closer to pop. prop	Within CI	Closer to pop. prop	Within CI
Number of comparisons	52	26	49	19
Number met criteria	45	23	29	19
% met criteria	87%	88%	59%	100%

Table S9 Assumed (limited) prior information on target population (male household heads) (left) and a-priori desired characteristics of the ten seeds (right). Village names removed for confidentiality.

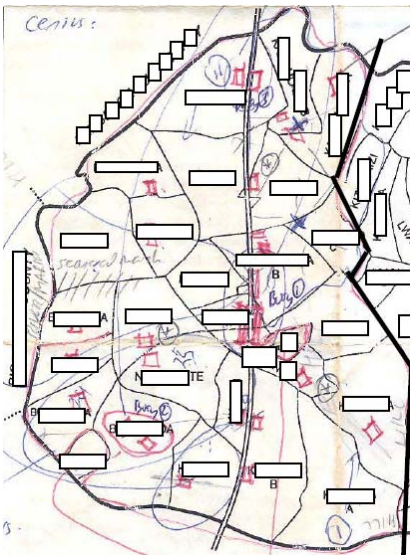
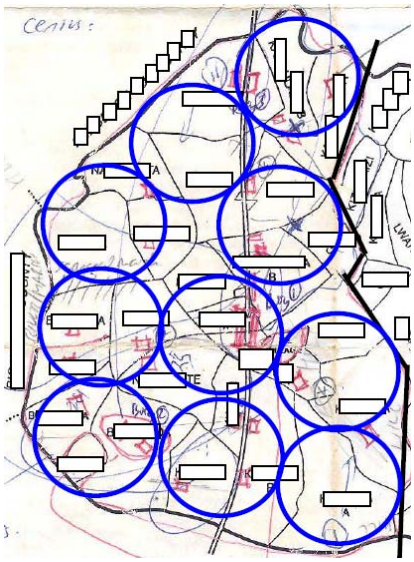
	Assumed (limited) prior knowledge	A-priori desired characteristics of seeds	
Geographic distribution	Map used by Medical Research Council mapper 	One seed from within each of the following ten areas 	
Age	Information from a Medical Research Council staff member working with study villagers: "*Most male household heads aged about 25 to 50 years. Min about 18 years. Max 70+ years.*"	10-19 yrs	2
		20-29 yrs	2
		30-39 yrs	2
		40-49 yrs	2
		50+ yrs	2
Tribe	Information from a Medical Research Council staff member working with study villagers: *"Most common tribe is Muganda followed by 'Munyanrwanda/kole'. There are also Mukiga, Murindi and other tribes in the area"*	Muganda	2
		Munyanrwanda/kole	2
		Mukiga	2
		Murundi	2
		Other known tribe	2

Table S10 Reasons for non-interview in simple random sample survey

Away	59	43.4%
Refused	26	19.1%
Couldn't find	20	14.7%
Died	8	5.9%
Health	4	2.9%
Other	19	14.0%
	136	100.0%

Table S11 Percentage of categories for which the RDS adjustments improve the estimates of the population proportions using different measures of network size. '-', could not be calculated.

Variable	RDS-1 Full sample			Small sample			RDS-2 Full sample			Small sample		
	NS1	NS4	NS5	NS1	NS4	NS5	NS1	NS4	NS5	NS1	NS4	NS5
Age group	40.0 (2/5)	60.0 (3/5)	40.0 (2/5)	0.0 (0/4)	40.0 (2/4)	20.0 (1/4)	40.0 (2/5)	40.0 (2/5)	40.0 (2/5)	20.0 (1/4)	40.0 (2/4)	40.0 (2/4)
Tribe	20.0 (1/5)	20.0 (1/5)	20.0 (1/5)	60 (3/5)	60 (3/5)	80 (4/5)	0.0 (0/0)	0.0 (0/0)	0.0 (0/0)	40.0 (2/5)	40.0 (2/5)	40.0 (2/5)
Religion	25.0 (1/4)	25.0 (1/4)	0.0 (0/4)	50.0 (2/4)	50.0 (2/4)	50.0 (2/4)	25.0 (1/4)	25.0 (1/4)	25.0 (1/4)	25.0 (1/4)	25.0 (1/4)	25.0 (1/4)
SES	20.0 (1/5)	20.0 (1/5)	20.0 (1/5)	20.0 (1/5)	0.0 (0/5)	20.0 (1/5)	20.0 (1/5)	20.0 (1/5)	20.0 (1/5)	20.0 (1/5)	20.0 (1/5)	20.0 (1/5)
Village	40.0 (10/25)	36.0 (9/25)	40.0 (10/25)	-	-	-	36.0 (9/25)	36.0 (9/25)	36.0 (9/25)	40.0 (10/23)	32.0 (8/23)	32.0 (8/23)
HIV status	33.3 (1/3)	66.7 (2/3)	33.3 (1/3)	0.0 (0/3)	33.3 (1/3)	33.3 (1/3)	66.7 (2/3)	66.7 (2/3)	66.7 (2/3)	0.0 (0/3)	0.0 (0/3)	0.0 (0/3)
Sexual partners	60.0 (3/5)	0.0 (0/5)	60.0 (3/5)	20.0 (1/5)	20.0 (1/5)	20.0 (1/5)	40.0 (2/5)	40.0 (2/5)	40.0 (2/5)	60.0 (3/5)	40.0 (2/5)	40.0 (2/5)
Overall	36.5 (19/52)	32.7 (17/52)	34.6 (18/52)	26.9 (7/26)	34.5 (9/26)	38.5 (10/26)	32.7 (17/52)	32.7 (17/52)	32.7 (17/52)	34.6 (18/52)	30.8 (16/52)	30.8 (16/52)

Table S12 Socioeconomic status results controlling for age. RDS-1 estimates are shown in bold if they are closer to the population proportions than the sample proportions.

Age group (years)	SES	Population proportions	Sample proportions	RDS-1 estimates
0-29	Highest	0.218	0.159	**0.178**
	Higher	0.237	0.246	0.215
	Lower	0.237	0.286	0.336
	Lowest	0.207	0.214	0.224
	Unknown	0.102	0.095	0.047
30-39	Highest	0.289	0.197	**0.198**
	Higher	0.250	0.227	**0.266**
	Lower	0.244	0.323	**0.290**
	Lowest	0.165	0.218	0.234
	Unknown	0.052	0.035	0.012
40-49	Highest	0.292	0.214	**0.254**
	Higher	0.282	0.268	0.219
	Lower	0.212	0.255	**0.237**
	Lowest	0.183	0.236	0.263
	Unknown	0.030	0.027	0.026
50+	Highest	0.231	0.152	0.109
	Higher	0.232	0.234	0.223
	Lower	0.221	0.251	**0.201**
	Lowest	0.286	0.336	0.448
	Unknown	0.029	0.026	0.019
Combined	Highest	0.257	0.178	**0.179**
	Higher	0.249	0.242	0.232
	Lower	0.229	0.279	**0.263**
	Lowest	0.214	0.256	0.302
	Unknown	0.052	0.044	0.025

Figure S1 Summary of reported network size of RDS recruits (excluding seeds)

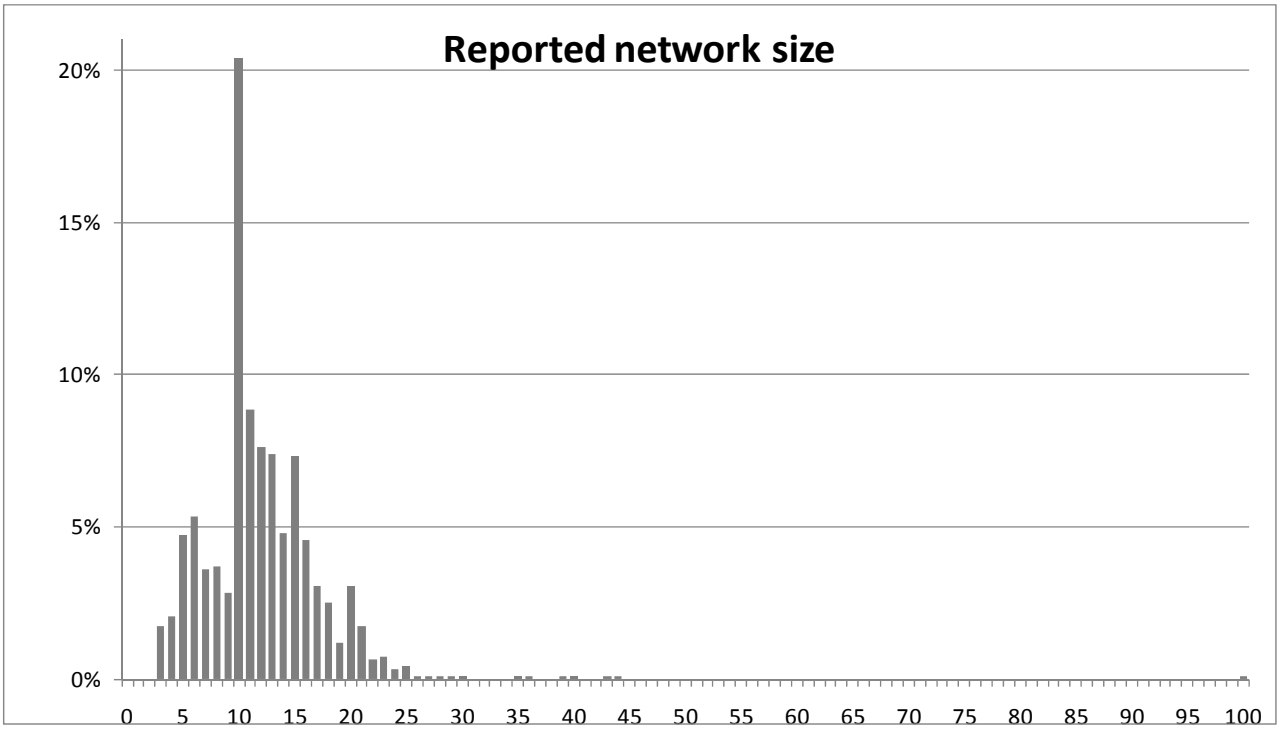


Figure S2 The distribution of network size, by definition (including seeds)

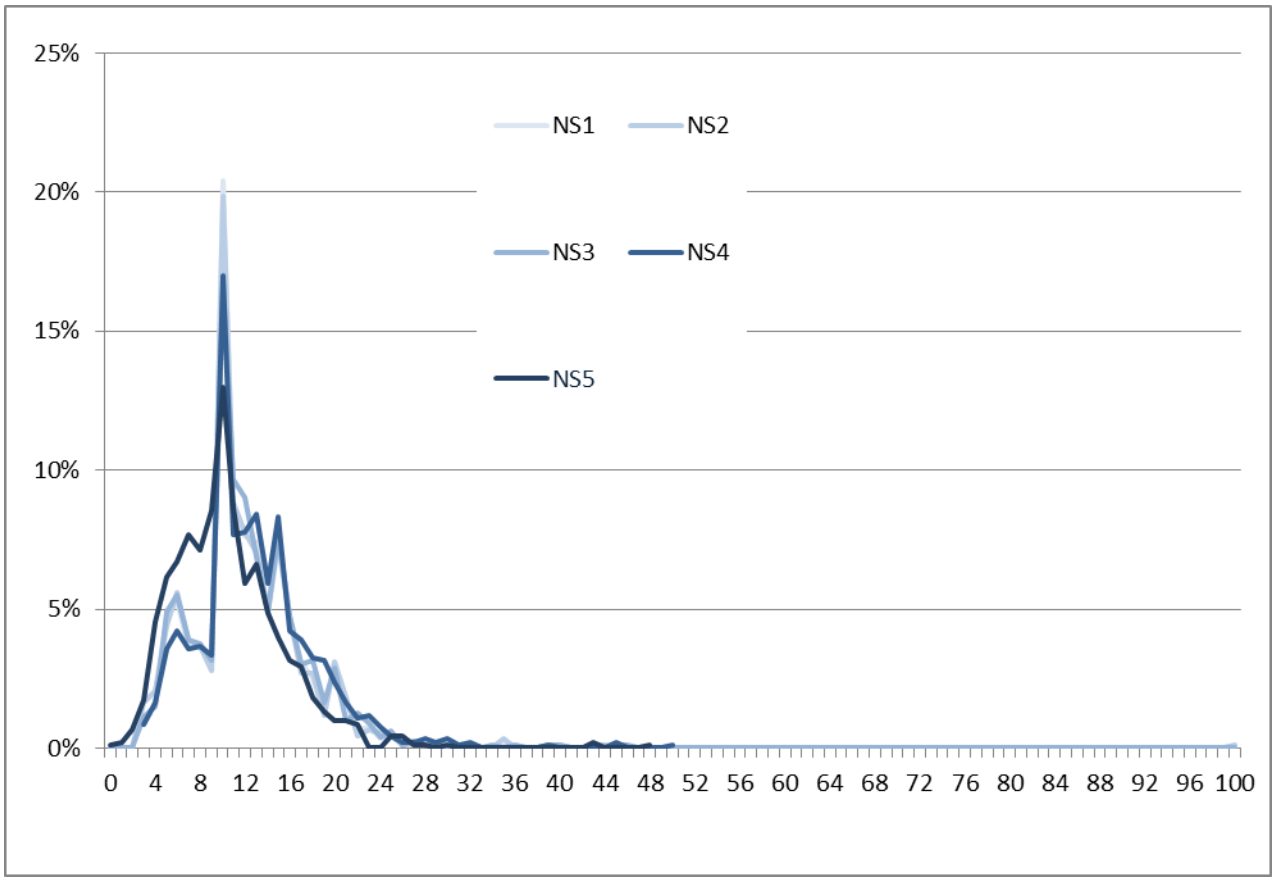


Figure S3 The distribution of network size among the target population. Men recruited into the RDS study are shown in black Network size definiton used was NS1. Recruits had a mean network size of 12.1 (based on 917 observations) and non-recruits 7.4 (162). The estimated mean network size in the whole target population was 9.2.

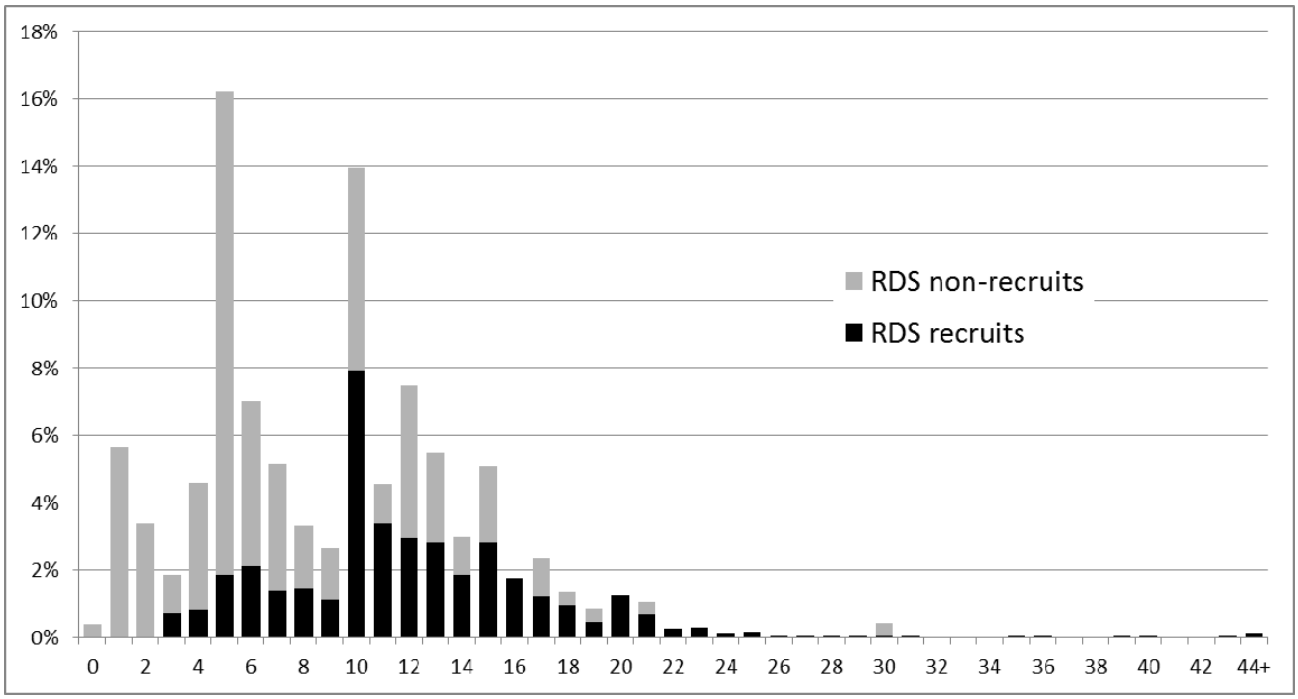


Figure S4 The number of times members of the target population were identified as contacts by other recruits

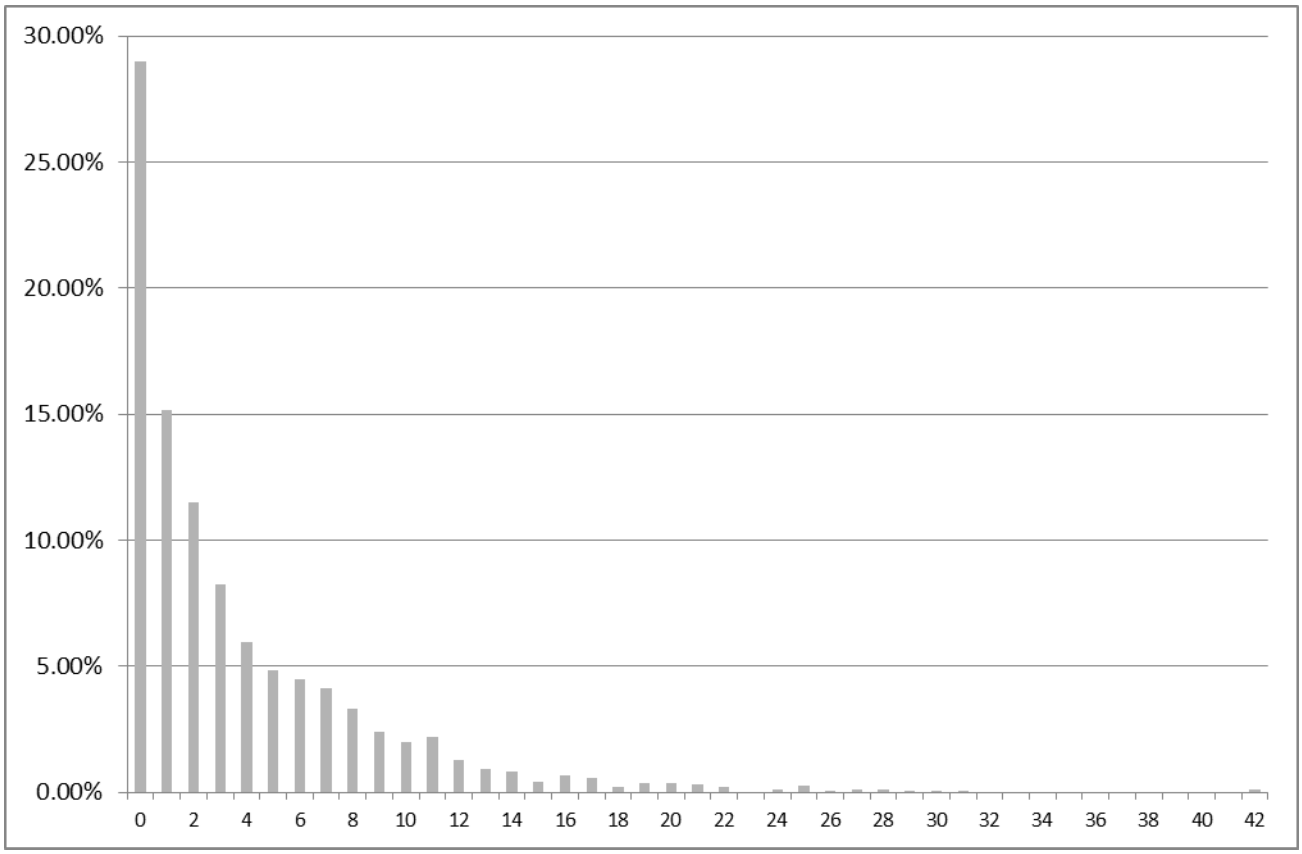


Figure S5 Proportion recruits over-recruited from their own village, by number of villages within 3km of a village. Network size definition NS5.

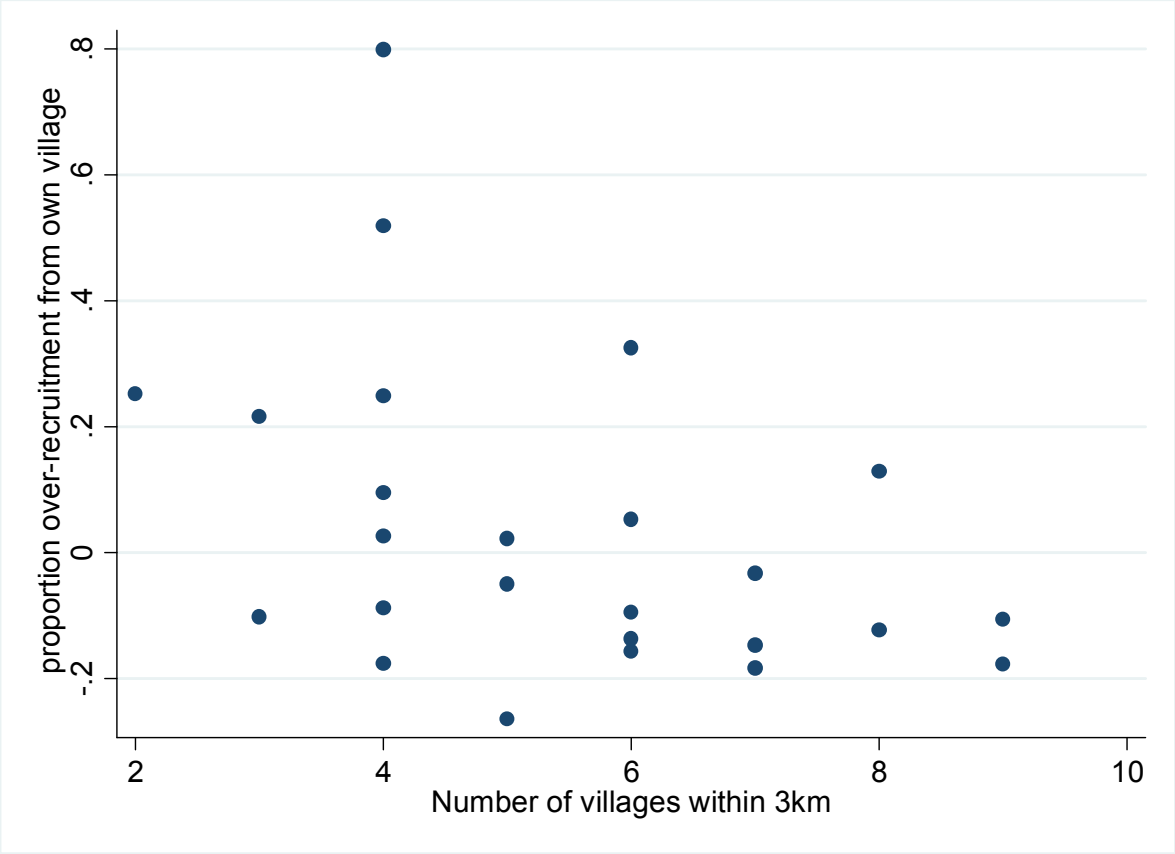


Figure S6. **Pattern of recruitment, by village and HIV status. Map (left):** symbols show the location of recruits' houses and colours indicate the recruiters' villages. Circles indicate that the recruit and recruiter were from the same village and triangles indicate that they were from different villages. **Recruitment networks (right):** The colour of the symbol indicates the recruit's village and the shape their HIV status (triangle=HIV positive, circle=HIV negative, square=HIV status unknown/not shown for seeds).

Figure S6

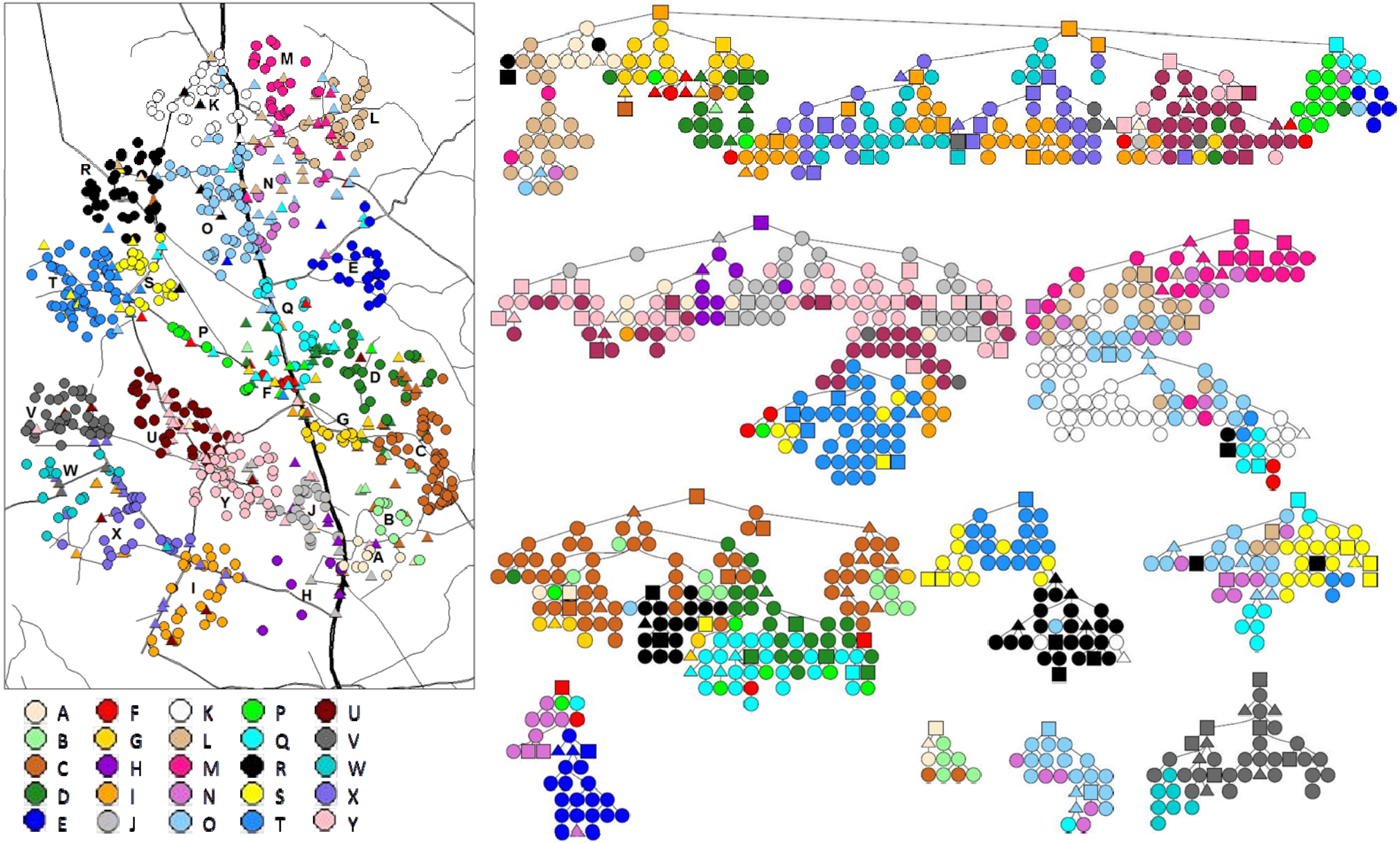

M
K
L
R
N
O
E
T
S
Q
P
D
F
V
G
U
C
W
Y
J
B
X
A
I
H
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y

Figure S7 Recruitment networks, by seed. Seeds are shown at the top of each recruitment network. Symbol area is proportional to network size. Symbol shading indicates week of recruitment (darkest = earliest). Symbol shape indicates whether the recruit was not offered coupons (square), was offered coupons but did not accept them (triangles), or was offered and accepted coupons (circles).