VERSITA

Journal of Official Statistics, Vol. 30, No. 2, 2014, pp. 291–310, http://dx.doi.org/10.2478/jos-2014-0018

Recruiting an Internet Panel Using Respondent-Driven Sampling

Matthias Schonlau[1], *Beverly Weidmer*[2], *and Arie Kapteyn*[3]

Respondent-driven sampling (RDS) is a network sampling technique typically employed for hard-to-reach populations when traditional sampling approaches are not feasible (e.g., homeless) or do not work well (e.g., people with HIV). In RDS, seed respondents recruit additional respondents from their network of friends. The recruiting process repeats iteratively, thereby forming long referral chains.

RDS is typically implemented face to face in individual cities. In contrast, we conducted Internet-based RDS in the American Life Panel (ALP), a web survey panel, targeting the general US population. We found that when friends are selected at random, as RDS methodology requires, recruiting chains die out. When self-selecting friends, self-selected friends tend to be older than randomly selected friends but share the same demographic characteristics otherwise.

Using randomized experiments, we also found that respondents list more friends when the respondent's number of friends is preloaded from an earlier question. The results suggest that with careful selection of parameters, RDS can be used to select population-wide Internet panels and we discuss a number of elements that are critical for success.

Key words: Web survey; RDS.

1. Introduction

Respondent-driven sampling (RDS) is a chain referral sampling technique typically conducted face to face with hard-to-reach populations in individual locations such as cities. Implementing RDS on the web can be advantageous. First, web implementation is much less expensive than a face-to-face approach because the interviewer labor costs and costs associated with setting up a field operation can be avoided. Second, RDS requires the assumption that respondents will recruit from among their friends at random. This is more easily enforceable on the web, where randomization from a list of friends is easy. Third, web implementation of RDS allows recruitment across a much wider geographic area as it is not tied to one geographic location as typically happens with face-to-face implementation of RDS. This is also true for geographic spread inside a large city, where friends who live or work close to the field station may be more likely to enroll.

[1] University of Waterloo, Statistics and Act.Sci.200 University Ave, Waterloo, Ontario N2L3G1, Canada. Email: schonlau@uwaterloo.ca
[2] RAND Corporation, Survey Research Group, Los Angeles, California, USA. Email: weidmer@rand.org
[3] University of Southern California and RAND Corporation. Email: kapteyn@usc.edu
Acknowledgments: Support for this research comes from grant R01AG20717 from the National Institute on Aging to RAND (Arie Kapteyn, P.I.). We thank the ALP team for all their help throughout. The project would not have been possible without their enthusiasm.

RDS studies conducted face to face have to be conducted in one location or city at a time. Fourth, in light of declining response rates to phone and mail surveys, a successful nationwide implementation of a network sample would provide a useful alternative. Fifth, populations who are routinely on the web may be more likely to participate. This may include at-risk college students, but also Internet sex workers. Finally, once the software for a web implementation is in place, it is easy to conduct additional RDS studies.

While implementing RDS on the web has several advantages, it also has different challenges. The web environment is more anonymous than face-to-face encounters, which may be useful for some hard-to-reach populations. However, recruiters may find it harder to motivate their friends to enroll and it may affect trust. Some populations do not have access to the Internet. For example, it may not be possible to reliably reach the homeless in this way. This limitation may decrease in importance over time as more and more people gain access to the Internet. Finally, operationalizing RDS through a web survey where new recruits enroll throughout the study and in turn become recruiters is not trivial.

Very few attempts have been made to conduct RDS on the Internet and many studies have either required multiple attempts or were unsuccessful. Because of the small number of studies involved, there is no conclusive evidence for any factor that may explain implementation challenges. We nonetheless find it useful to list factors we believe contributed to implementation challenges or failure. First and perhaps most importantly, requiring respondents to provide information such as email addresses of their friends appears to be a bad idea. We also note that "respondent-driven" implies that the respondents should contact their friends; not interviewers or an automated computer program. An Internet-based RDS study in Cambridge, England, about cars and the environment failed (RAND Cambridge, personal communication) because respondents were unwilling to contact friends.

Second, not providing an incentive or providing too small an incentive seems to also affect recruitment rates. The aforementioned Cambridge study did not have sufficient funds to provide incentives and failed. An attempt to recruit parents of students studying at Tilburg University in the Netherlands into a panel survey failed to generate sufficient response (personal communication, Department of Leisure Studies). Students were asked to contact and enroll their parents and were offered a small monetary incentive for doing so. Only 120 persons out of the 4,000 invited joined the panel (a 3% enrollment rate). A study of students at Wayne State University (Detroit) that invited feedback from seed respondents made it clear respondents wanted to earn more money (Bauermeister et al. 2012), among other things. In response, investigators increased the total number of referrals allowing respondents to earn money for the first five referrals (5∗$10 for referrals plus $20 for filling out the survey). The study succeeded in recruiting 3,426 respondents.

Incentives can be delivered in a variety of forms. A study of men who had sex with men (MSM) in Vietnam (Bengtsson et al. 2012) used credit on SIM chips or a donation to an organization the target population cared about as well a lottery draw for an iPad. The aforementioned study at Wayne State University (Bauermeister et al. 2012) provided VISA e-gift cards for filling out the questionnaire. The cards were reloaded when friends were referred successfully. A study of Muslim students, also at Wayne State University, used unspecified gift certificates (Arfken et al. 2013). In the first published Internet RDS

study (Wejnert and Heckathorn 2008) at Cornell University, respondents (or friends) had to pick up incentives in person.

Third, there is at present no evidence that the survey topic plays an important role in recruiting success. Consider two contrasting examples: The RAND Corporation studied the opinions of gays and lesbians in the military about the "Don't Ask, Don't Tell" policy (Berry et al. 2010). The RDS study was a resounding policy success and won the "policy impact award" from the American Association for Public Opinion Research (AAPOR) in 2011. However, the study failed as an RDS study because referral chains were not long enough. The role of gays and lesbians in the U.S. military is a topic that is presumably very important to those in the military who are gay and lesbian, and if the topic were important for recruiting it should have been more successful. In terms of speed, the most successful Internet RDS study (Wejnert and Heckathorn 2008) reached the intended sample size of 150 respondents (plus nine additional seeds) in only 72 hours. Participants were "invited to participate in a research study to empirically validate Respondent Driven Sampling (RDS) as an analytical tool for the study of social structure" (Wejnert and Heckathorn 2008, online Appendix A). This does not appear to be a 'sexy' topic for university students that would have contributed to the success of the study.

Fourth, adjusting the number of seeds after a study begins is an important tool to avoid recruiting chains dying out. The Vietnam study initially started with 15 seeds and then increased to 20 two weeks later (Bengtsson et al. 2012). However, increasing the number of seeds does not always work – the RAND study (Berry et al. 2010) increased the number of seeds from five to 189 and still came up short. Except for this study, the number of seeds in Internet RDS studies tends to be much smaller with numbers ranging from nine (Wejnert and Heckathorn 2008) to 22 (Bauermeister et al. 2012).

Fifth, adjusting the number of referrals allowed is a useful tool to avoid recruiting chains dying out. As mentioned above, one study (Bauermeister et al. 2012) increased the number of referrals while paying only for the first five successful referrals. This study also allowed respondents to copy the referral codes into text messages and social media (Facebook). The study eventually reached 3,448 respondents. Overall, the number of referrals in Internet RDS studies has ranged from three to five.

The purpose of our study was to explore the feasibility of recruiting respondents into a web panel using Internet-based RDS. Specifically, our goal was to recruit respondents into the American Life Panel (ALP) (https://mmicdata.rand.org/alp/), a probability-based Internet panel. It was hoped that the availability of a network sample would make the ALP more attractive to researchers with such needs. Already existing respondents in the ALP panel were **not** recruited over the Internet. Potential respondents without Internet access receive a laptop and broadband Internet access for free.

An overview of the sequence of experiments and recruitment efforts is given in Figure 1. In our initial pilot run we found that a large number of respondents would only list a single respondent, presumably to avoid follow-up questions. We designed Experiment 1 to find out which survey design would lead to a more successful elicitation of friends (Section 2). Next, we found respondents were listing friends but those friends would not contact us to enroll in the study. In response we designed a second experiment (Section 3), varying incentive levels and how friends were selected. Based on results from Experiment 2 we started an RDS sample in the American Life Panel (ALP) (Section 4). Section 4 also

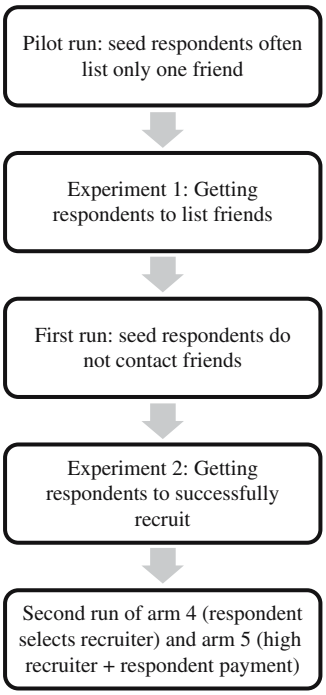

Fig. 1. Sequence of experiments and recruitment efforts.

compares the demographic composition of the recruited sample for self-selected friends and randomized friends. Section 5 concludes with a discussion. We first provide an overview of respondent-driven sampling.

1.1 Respondent-Driven Sampling

RDS is a chain referral sampling technique (Heckathorn 1997, 2002, 2007). A small number of seed respondents recruit additional respondents from their network of friends. The recruiting process repeats iteratively, thereby forming long referral chains. Suppose we want to estimate the percentage of males, p_1, and females, p_2, in a population. Of interest is the two-by-two gender transition matrix (male/female recruiting male/female) between the recruiter and recruit. Assuming a first-order Markov process (recruit gender depends on recruiter's gender, but not on earlier recruiters), a sample equilibrium is reached if referral chains are sufficiently long. The sample equilibrium is not the population equilibrium because well-connected people are overrepresented in the sample. For example, if women have more friends than men, the sample equilibrium would have more women than men. The assumption of reciprocity (explained below) yields an equation: The number of possible edges (links between two persons) with a male recruiter recruiting a female is the same as the number of possible edges with female recruiters recruiting a male: $n_1 D_1 S_{12} = n_2 D_2 S_{21}$ where n_i is the number of respondents in group i, D_i is the average group degree (e.g., average network size among females) and S_{ij} is the estimated transition probability between categories i and j. Dividing by the total sample size turns frequencies into proportions: $p_1 D_1 S_{12} = p_2 D_2 S_{21}$.

The average group degree of group i, D_i, is estimated from individual degrees using the "multiplicity" formula (Salganik and Heckathorn 2004) $D_i = n_i/[\sum_j(1/d_{ij})]$ where n_i is the number of respondents in group i and d_{ij} is an estimate of individual degree (number of friends). Unlike the arithmetic mean, this formula takes into account that respondents with a greater network are overrepresented. Because the formula relies on the inverse of self-reported degrees, this estimate is robust against large positive outliers in individual degrees.

In the example above, the equation depends on two unknown proportions, p_1 and p_2 (the proportion of females and males in the population), which must sum to one: $1 = p_1 + p_2$. Two equations with two unknowns can be solved and yield $p_1 = S_{21}D_2/(S_{21}D_2 + S_{12}D_1)$ and $p_2 = 1 - p_1$. When there are more than two categories, the estimates result from solving an over-determined system of equations. In general, RDS only allows the estimation of proportions, not absolute frequencies or totals. However, frequencies can be computed from the proportions if the population total is known from elsewhere or if the population total is estimated by capture/recapture methods (Berchenko and Frost 2011; Heckathorn et al. 2002).

RDS requires the following assumptions: *Assumption 1. Reciprocity*. If respondent A recruited respondent B, then in principle B could have recruited A also. In practice, this assumption is tested by verifying that the recruiter is part of the recruit's social network. *Assumption 2. Networked population*. Respondents are all linked to a single component in the network (i.e., there are no isolated pockets of people without friends). *Assumption 3. Sampling is with replacement*. This assumption never holds because the same respondent is not sampled twice. In practice, this assumption is innocuous unless the sample represents a large fraction of the population. *Assumption 4. Network size*. Respondents can accurately report their degree (personal network size). Consistent under- or overestimation of network size among all respondents cancels out and is unproblematic (Wejnert 2009, sec. "Degree Estimation"). Moreover, estimates may be robust for different assessments of network size (Wejnert 2009). *Assumption 5. Random Recruitments.* Respondents recruit from their network at random. This assumption is the most controversial by far. In some arms of our experiments we have done the randomization ourselves by choosing from a list of first names or initials provided by the respondent, thus avoiding this problem.

In practice, recruitment is facilitated through a dual incentive system which includes payments for both the respondent and each referral who agrees to participate. While RDS has not yet been used to recruit a national sample, it emerges as a natural choice when social network analyses are of interest (Wejnert 2010). Respondent-driven sampling is implemented in a stand-alone package (www.respondentdrivensampling.org), in Stata (Schonlau and Liebau 2012), and a package is in preparation for R.

2. EXPERIMENT 1: Eliciting the Number of Friends

One assumption of RDS is the random recruitment of friends. Given a list of friends (first names or nicknames suffice), respondents are asked to contact specific friends selected at random by the computer software. During initial trials we found that many respondents tended to only list a single friend. The purpose of Experiment 1 was to investigate how to

ask about respondents' friends such that respondents list their friends in greater numbers. Respondents were not told why they were supposed to list the friends to avoid having respondents selectively list only friends they wanted to contact. In each case *we first ask for the number of friends* (Numerical question given in Appendix). On the following screen we then ask the same question but ask the respondent to list first names or initials. (The friends question is reproduced in Appendix) The experiment had five experimental arms:

Experimental Arm 1 ("1 row"): We asked the respondent to list one person at a time (Figure 2). On the same screen, we asked whether the respondent wanted to list an additional person. If this question went unanswered, we prompted for an answer on the next screen: "You did not answer the previous question(s). Your answers are important to us. Please return to the previous question and answer it to the best of your ability."

Experimental Arm 2 ("10 rows"): We asked the respondent to list ten persons at a time (Figure 3). If respondents listed ten persons they were asked whether they would like to list additional persons. If respondents listed less than ten persons they proceeded to the next question (without prompt).

Experimental Arm 3 ("prompt w preloaded #"): As in Experimental Arm 2, respondents were asked to list ten persons at a time. However, if fewer people were listed than indicated in the preceding numerical question, we prompted for additional people: "You answered earlier you had [preloaded number] close friends and family members, but you listed a smaller number. This question is very important to us. If possible please go back and add more people." To avoid a problem if respondents gave an unreasonably high numerical value (e.g., 100), respondents were not prompted if they listed at least ten

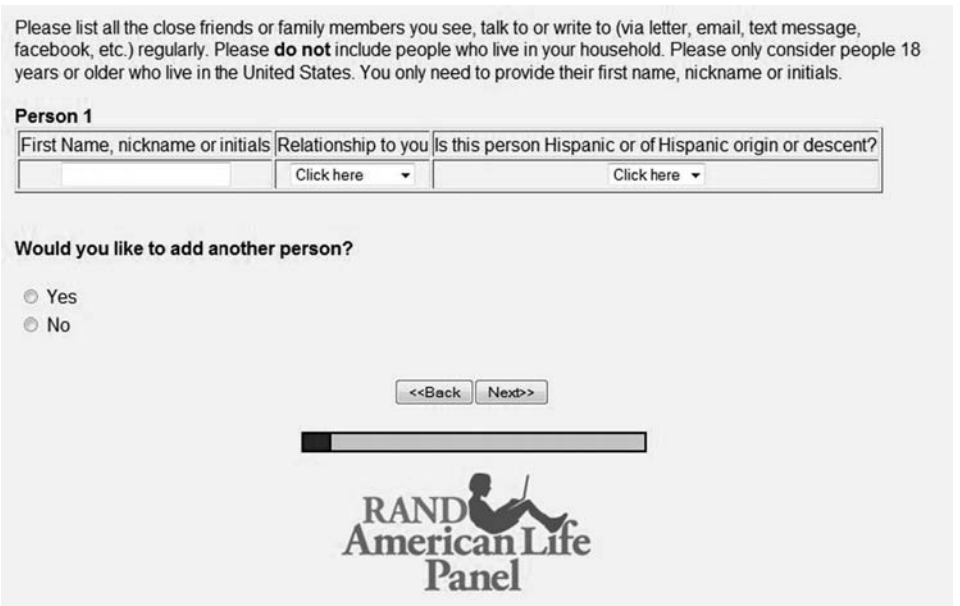

Fig. 2. Screenshot of Experimental Arm 1. The dropdown menu for relationship had the categories: child, parent, other relative, work friend, school friend, family friend, acquaintance, other.

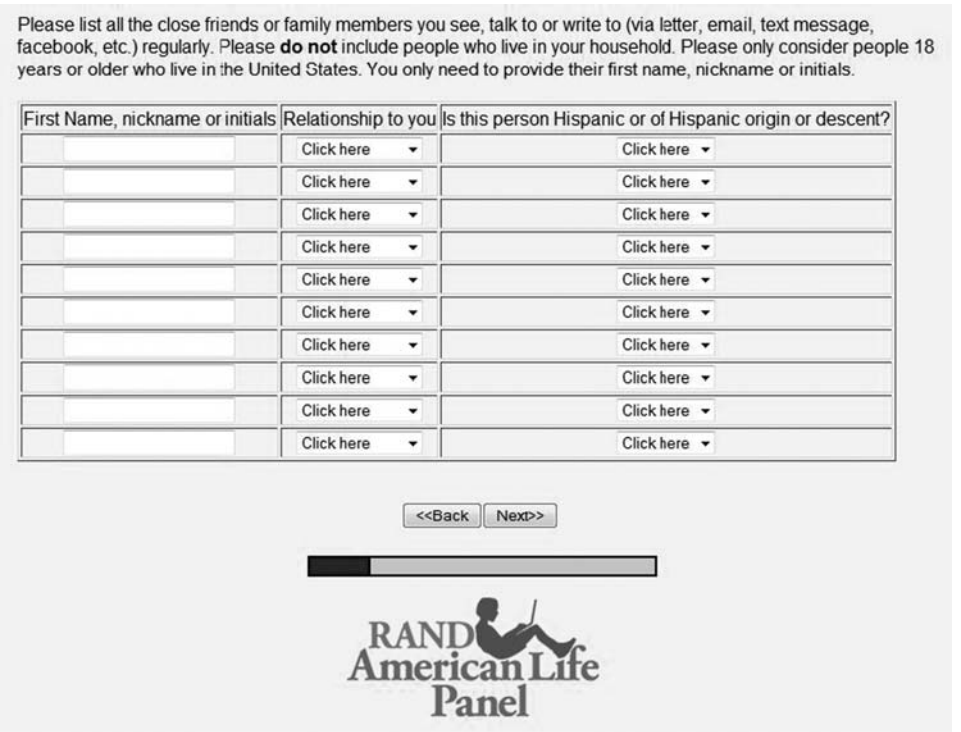

Fig. 3. Screenshot of Experimental Arm 2 (identical to Experiment Arm 3). The dropdown menu for relationship had the categories: child, parent, other relative, work friend, school friend, family friend, acquaintance, other.

friends. Respondents were allowed to list more people than indicated in the numerical question.

Experimental Arm 4 ("ask preloaded #"): Respondents were specifically asked for the number of friends given in the numerical question: "Please list these [preloaded number] close friends or family members. You only need to provide their first name, nickname or initials" (Figure 4). As before, if there were more than ten friends, we listed ten on each screen until the total number was exhausted. For example, if there were twelve friends, the respondents saw a screen with ten rows and a second screen with two rows. If the respondent failed to list ten friends, the second screen was not shown. There was no additional prompt if respondents listed fewer friends than indicated by the numerical question.

Experimental Arm 5 ("single column"): We first asked respondents to list first names only (Figure 5). The number of friends was not preloaded. In a second question, we asked respondents to list information about their friends (Figure 6). If respondents did not list anybody, they were not prompted to remind them of their earlier numerical answer. If respondents listed ten friends, they were shown an additional screen and asked to list additional friends as before.

Respondents received a $5 incentive for responding to this survey. To avoid overly complicated programming, we did not allow listing more than 50 people.

Fig. 4. Screenshot of Experimental Arm 4 with a listed number of rows given from numerical question. The dropdown menu for relationship had the categories: child, parent, other relative, work friend, school friend, family friend, acquaintance, other.

Fig. 5. Screenshot of Experimental Arm 5. First names are listed first, and then additional information is prompted.

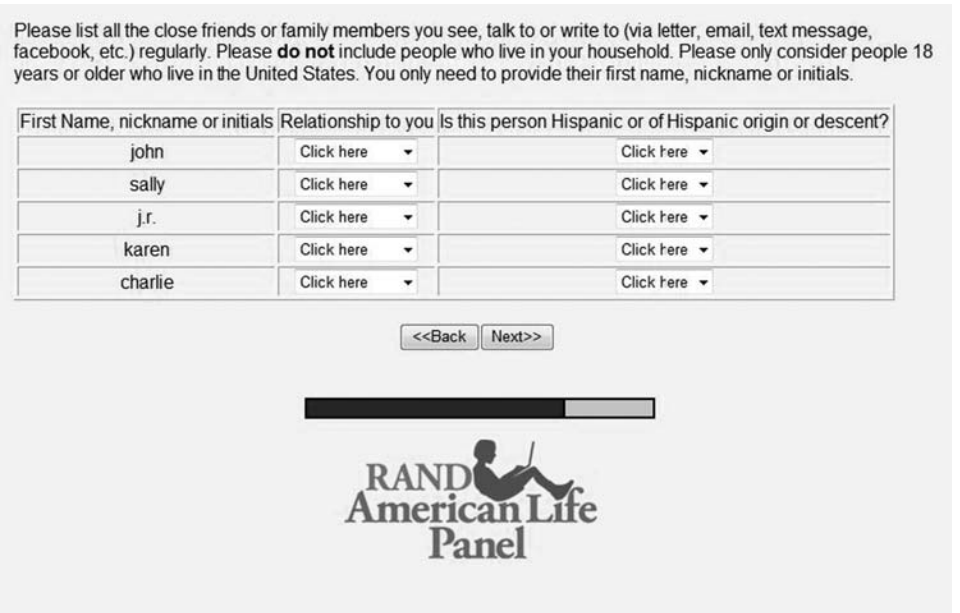

Fig. 6. Screenshot of Experimental Arm 5 soliciting additional information. The dropdown menu for relationship had the categories: child, parent, other relative, work friend, school friend, family friend, acquaintance, other.

Results: 473 respondents were invited to participate. Respondents were randomized one at a time to an experimental arm. The response rate was 86%. Respondents were randomized on the fly while they took the survey. The number of completed surveys was equal to 63 (earlier pilot experiment), 70 (Arm 1), 65 (Arm 2), 76 (Arm 3), 68 (Arm 4), 80 (Arm 5).

Box plots of the number of friends listed by experimental arm are shown in Figure 7. Listing the preloaded number of close friends and family and prompting with the preloaded number elicited the largest number of friends listed (Arms 3 and 4). Listing only one column (first names; Arm 5) instead of three columns (first names, relationship, Hispanic; Arm 2) does not affect the number of respondents listed. Asking to list one person at a time as opposed to ten persons at a time (Arm 1) does not work well at all.

3. EXPERIMENT 2: Getting Respondents to Contact Their Friends

3.1 Motivation

Experiment 1 explored how to get respondents to list more friends. Using the most successful design from Experiment 1 (asking for the preloaded number of friends; Experimental Arm 4), we started an RDS sample with five Hispanic seed respondents looking to recruit Hispanic friends. Four of five invited Hispanic respondents completed the survey. One respondent listed only one Hispanic friend; the others had at least four Hispanic friends each. We asked respondents to invite all of the listed Hispanic friends, so between them, respondents were supposed to contact 13 friends. However, in a follow-up survey the respondents indicated they tried to contact only two out of

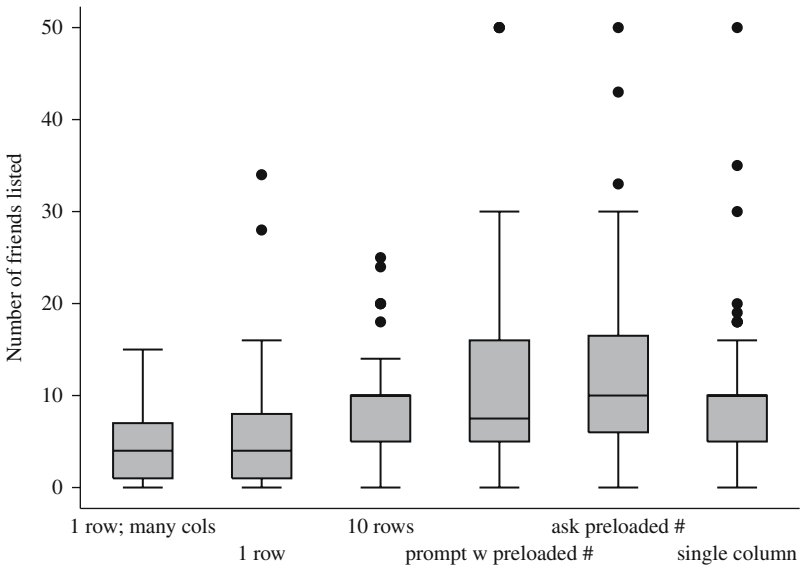

Fig. 7. Box plots of the number of close friends and family listed. (Note: "1 row; many cols" refers to an earlier pilot experiment; the five experimental arms are shown in order after that.)

13 respondents. Reasons for not contacting some respondents include "I forgot", "I did not feel comfortable asking them", and "He/she probably would not have participated". The two respondents contacting one friend each thought both friends were unlikely to participate because the friend "Thinks it will be too complicated". The referral codes were transferred by email (1) and phone (1). Overall, all four respondents thought passing on the referral code was very easy.

One of the two respondents contacted joined the panel; this person was not asked to refer additional friends because we felt we needed to conduct another experiment to test a different approach in the hope of finding a way to obtain a higher yield for the friend referrals. In conclusion, the four Hispanics who responded to the survey did not contact the friends they had listed.

3.2 Experiment 2

The purpose of the second experiment was to learn how to make it easier and more attractive for respondents to contact their friends. For all experimental arms, we gave respondents the option of receiving one prewritten email for each friend or one prewritten letter for each friend to be sent through the U.S. post office (Figure 8). Each email contained an explanation of the ALP, the referral code and a link to the webpage, as well as the respondent's name in the subject line. By the end of Experiment 1, the ALP had recruited a Hispanic subsample using address-based sampling. This obviated the need to specifically recruit respondents of Hispanic ethnicity. In the second experiment we therefore no longer restricted new recruits to those of Hispanic ethnicity.

The experimental arms for Experiment 2 are displayed in Table 1. As before, we paid respondents $5 for filling out the short referral survey (an earlier informal test revealed that

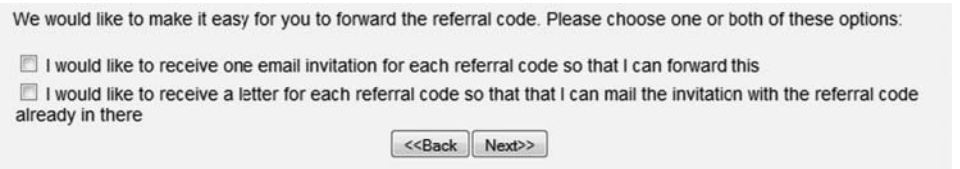

Fig. 8. Option to receive one prewritten email (or postal mail letter) per referral in Experiment 2.

respondents were equally likely to respond with a $5 payment as with a $10 payment). We experimentally varied the amount paid for each successful respondent ($15 vs. $30), a sign-up bonus to the friend ($20 vs. none), and whether the friends were selected at random by the computer or by the respondent, as is customary in RDS. Respondents were asked to nominate a maximum of four friends. With the $30 incentive per successful referral, the respondent could earn a total of $120 if all four friends filled out the first survey. The friends could earn the sign-up bonus as well as the regular survey payments ($20 for each 30 minute survey). In all arms, ten friends at a time could be entered on each page.

Results: The response rates (86%-92%) and number of completes (44-50) are reported in Table 1. Respondents listed on average 13.5 friends (median 10, 1st quartile 6, 3rd quartile 20). The number of friends listed did not vary significantly by experimental arm (based on a Poisson regression on indicator variables for arms with robust standard errors). A histogram of the number of friends listed is shown in Figure 9. Respondents could list up to ten friends. Correspondingly, there is some heaping on the values 10, 20 and 30. The number of friends that could be listed was limited to 50 for programming reasons.

The average number of recruits differed substantially by arm (see bottom rows of Table 1). When respondents self-selected friends (Arm 4) and when both larger recruiter incentives and sign-up bonuses were provided (Arm 5), the number of recruits was significantly larger than in the control group. We conducted a logistic regression of the indicator variable "$>=1$ referral" (vs. "0 referrals") on four indicator variables for the five experimental arms. Coefficients for Arm 4 ($p = 0.02$) and Arm 5 ($p = 0.03$) differ significantly from the control (Arm 1).

For Arms 4 and 5, the ratio of friends recruited to the number of respondents recruiting is close to 1. A ratio greater than one would imply an increasing number of recruits from wave to wave and the referral chain would not die out. Using both larger recruiter payments and recruit sign-up bonuses (Arm 5) works much better than either one of these on its own (Arms 2 and 3). A recruit sign-up bonus by itself has little effect relative to the control group.

4. Continued RDS Waves With Arms 4 and 5

We continued two separate RDS recruiting efforts corresponding to Arms 4 and 5 from Experiment 2. Monthly surveys were conducted on the second Wednesday of every month from March to October 2012. Any recruit who had responded in the previous month would be invited to recruit their friends. If the recruitment process took longer than a month for any one person, that person would simply be invited in the following month.

Table 1. Experiment 2 experimental arms, response rates, and friends recruited.

Experimental Arm	1	2	3	4	5
Name	control	high recruiter payment	respondent incentive	self-selected friends	arm 2 + arm 3
Payment to recruiter for filling out survey	5	5	5	5	5
Payment to recruiter for each successful respondent	15	30	15	15	30
Sign-up bonus payment for new respondent	0	0	20	0	20
Selection of recruits	computer	computer	computer	respondents	computer
Number invited	57	56	53	49	50
Number of respondents	49	50	46	45	44
Response rate	86%	89%	87%	92%	88%
Number of recruits	22	34	23	43	41
Ratio recruits per recruiter	0.45	0.68	0.50	0.96	0.93

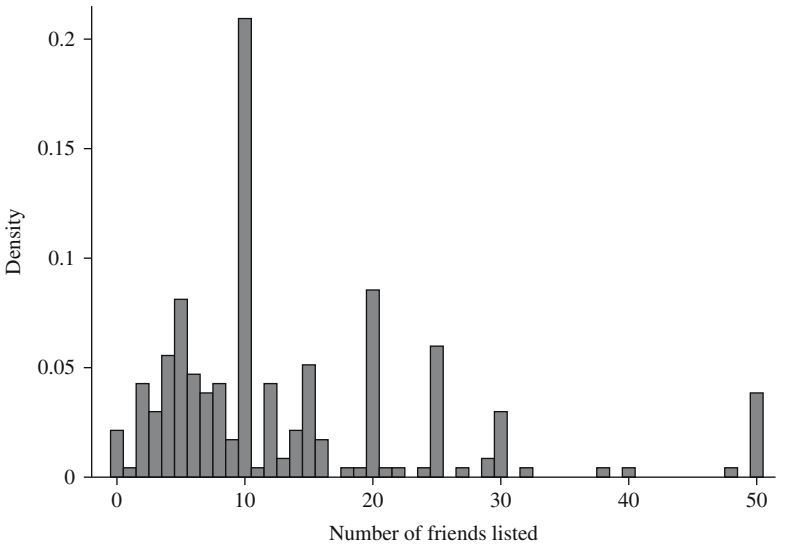

Fig. 9. Experiment 2: Histogram of the number of friends listed. Heaping effects for multiples of 10 and a ceiling effect (50) are visible.

Arm 4 (recruiter selects respondent) was only added after a two-month delay after it appeared that the enrolment in Arm 5 was slow.

Table 2 shows recruitment by wave (also referred to as depth) for both arms. Recruitment in Arm 5 decreased after Wave 1 and effectively died out in Wave 5. The recruiting strategy including incentives did not change after Wave 1. A number of respondents tried to circumvent the random recruiting assignment and gave the coupon code to self-selected friends. To avoid contamination of the ongoing experiment they were not allowed into the panel.

The response to Arm 4 was much stronger than that for Arm 5. The total number of recruits (excluding seeds) was more than twice as large. Recruiting was very slow, often taking longer than a month. The smaller number of recruits at Depth 5 does not reflect a

Table 2. Enrolment by Wave. Wave 0 refers to seed respondents; Wave 1 to friends of seed respondents; Wave 2 to friends of friends, etc.

Wave	Arm = 4	Arm = 5
0	45	44
1	46	47
2	50	19
3	63	18
4	48	10
5	16	1
Total	268	139
Total Recruits	223	95

Note: Wave 5 is not complete; only faster-recruiting chains reached this wave.

decline in enrollment but rather represents the fastest group of respondents. Some of the remaining respondents did not have the chance to reach Wave 5.

Demographics by Experimental Arm 4 (self-selected friends) vs. Arm 5 (friends selected at random) are shown in Table 3. There are no statistically significant differences (based on χ^2 tests) between self-selected and randomly selected friends (recruits) with respect to gender, education, race/ethnicity, family income, or marital status. However, self-selected friends are on average eight years older than randomly selected friends (mean age = 45 vs. mean age = 37, $p = 0.0001$ based on a t-test). There are also regional differences: Self-selected friends are more likely to live in the Midwest and less likely to live in the South or Northeast ($p = 0.01$; based on a χ^2 test).

There is also a gender imbalance; roughly 70% of recruits were female. Recruits are somewhat more educated than the general public; in particular, very few recruits have less than a high school degree. Recruits are predominantly non-Hispanic whites (88%), even though there were Spanish versions of all surveys. Most of the remainder are (non-Hispanic) African Americans. Recruits have a wide range of family incomes. Thirty-nine percent of families have a household income of less than $40,000; two thirds have a household income of less than $75,000. Ninety percent of those under the age of 65 are working. Recruits are geographically spread across 40 U.S. states. All of the recruits had Internet access and we therefore did not have to provide Internet access in the form of a laptop and broadband for any of the recruits.

For completeness, Table 3 provides the demographic characteristics of seed respondents by experimental arm. Comparisons of demographic characteristics of seeds and respondents must proceed with caution. When the equilibrium is reached, the demographic distribution of seed respondents and recruits are theoretically independent of each other. Large shifts between seed and respondent distributions only show the recruiting chain does not get stuck in any one category. Here we find that compared to the seed respondents, recruits are more often in the 18–29 age group (almost all ALP panel members are 18 years or older by design; though the occasional 17-year old is not rejected), they are more often "never married", they are less often in the highest income category (particularly in Arm 4), and more often live in the southern part of the U.S.

5. Discussion

This is the first RDS study to attempt recruiting respondents throughout the U.S. rather than in individual U.S. cities. In fact, the only other RDS study we are aware of that recruited at a national level is the Vietnam study (Bengtsson et al. 2012). It was initially unclear whether respondents would spread across the United States or would remain in a confined region or state. With respondents in 40 different states, the overall geographic spread is good.

Respondents resist random selection. While incentives are important, the respondents' overriding desire was to choose whom they recruit. Respondents might have preferred self-selection to increase the probability of getting their own incentive, to channel money to specific friends and family, or because they were more comfortable contacting certain friends and family members. Self-selected respondents tend to be older than randomized

Table 3. Demographic composition of seed respondents and recruits by experimental arm.

		Arm 4 (self-selected friends)			Arm 5 (high payment + respondent incentive)		
		Seeds	Recruits		Seeds	Recruits	
		%	Freq.	%	%	Freq.	%
Gender	Male	40	55	29	37	26	36
	Female	60	133	71	63	46	64
Age category	<18	0	0	0	0	2	3
	18-29	0	40	21	14	25	34
	30-39	24	48	26	14	23	32
	40-49	22	22	12	19	5	7
	50-59	16	23	12	26	13	18
	60-69	20	32	17	21	3	4
	70 +	18	23	12	7	2	3
Education	Less than high school	4	2	1	0	4	5
	High school	11	33	18	19	12	16
	Up to Bachelor	56	128	68	56	45	62
	More than Bachelor	29	25	13	26	12	16
Race / Ethn.	Non-Hispanic White	87	162	86	79	66	90
	Non-Hispanic African American	7	20	11	7	5	7
	Non-Hispanic Other	0	2	1	5	1	1
	Hispanic	7	4	2	9	1	1
Family income	<10,000	2	9	5	5	7	10
	10,000-19,999	7	17	9	5	7	10
	20,000-29,999	11	19	10	12	8	11
	30,000-39,999	16	27	14	9	10	14
	40,000-49,999	2	12	6	14	7	10

Table 3. Continued

		Arm 4 (self-selected friends)			Arm 5 (high payment + respondent incentive)		
		Seeds	Recruits		Seeds	Recruits	
		%	Freq.	%	%	Freq.	%
	50,000-59,999	4	23	12	5	6	8
	60,000-74,999	13	26	14	19	6	8
	>75,000	44	54	29	33	22	30
Region	Northeast	22	10	5	16	10	14
	Midwest	27	63	34	28	12	16
	South	31	76	40	35	36	49
	West	20	39	21	21	15	21
Marital status	Married	64	119	63	65	40	55
	Separated/widowed/divorced	29	28	15	19	11	15
	Never married	7	41	22	16	22	30

respondents, but except for increased recruiting in the Midwest, there are no discernible differences with respect to other demographic characteristics.

The slow speed of recruiting remains a challenge. A study in a single college finished recruiting in a single weekend (Wejnert and Heckathorn 2008). Web recruiting for the drug and alcohol study at a single university (Bauermeister et al. 2012) concluded after 2.5 months. Web recruiting for the study about men who have sex with men in Vietnam (Bengtsson et al. 2012) took about two months. Our recruiting effort may have been slower for several reasons: 1) We had a broader target population that was more difficult to incentivize. 2) We only mailed invitations out once a month, thereby possibly signaling a lack of urgency. The fastest Internet RDS study was automated, requiring intervention only to end recruiting and incentive payment (Wejnert and Heckathorn 2008). 3) There was no focus on a specific topic. Respondents would become ALP panel members and would be asked to participate in a variety of surveys. The ongoing recruiting effort was only one of several surveys they were asked to participate in.

It is unclear what caused the gender imbalance (Table 3). We have observed a similar gender imbalance in international face-to-face surveys (www.itcproject.org) where women are thought to be more likely at home. Slight gender imbalances were also reported for most countries in the European Community Household panel (Behr et al. 2005) even after Wave 1 of the panel.

Once enrolled, the response rate of RDS panel members is similar to that of regular panel members recruited in the same time period. (Long-standing panel members tend to have higher response rates.) Response rates of RDS panel members ranged from 79–86% in three large surveys conducted between January and April 2013. Response rates among regular panel members recruited since July 1, 2011 for these surveys ranged from 80% to 84%.

It appears that this approach reaches low-income populations, those not working, and the elderly. It works less well for reaching racial/ethnic minorities other than African Americans, those with less than high school education or those without access to the Internet. This replicates the demographics among the seed respondents. An alternative interpretation therefore might be that the approach thus far has been unable to reach hard-to-reach populations not already represented among the seed respondents. Any Internet implementation of RDS will require careful pilot testing and experimentation. Importantly, we discovered respondents' preloaded number of friends helped to generate a longer list of friends. Other studies also had difficulties in "calibrating" Internet-based RDS (Bauermeister et al. 2012); others go unreported in the literature because they failed. While our self-selected friends chain (Arm 4) would have continued past Wave 5, the number of recruits was certainly not rising exponentially.

Our study has a number of limitations. The first may have to do with our name generator. First introduced by Laumann (Laumann 1966), name generators have become an active area of research (Marin and Hampton 2007). Our name-generating question asked for "friends", whereas many studies ask for both "friends and acquaintances". The term "friend" alone for name generating has been shown to be interpreted differently by different socioeconomic strata (Burt 1983). Therefore our name generating question may have introduced some bias, and consequently the computer-generated random recruitment was also conducted on the potentially biased list of names. Most studies work with specific

subpopulations (e.g., "Men who have sex with men"). Because we were targeting a general population we consciously decided not to include acquaintances because we felt the number of acquaintances might have been too large. Either way, the difficulties in recruiting were not due to a lack of friends listed.

Second, impersonation and duplication of respondents, while unlikely, cannot be ruled out. Duplication was probably less likely than impersonation as respondents had to provide contact information and an address in order to receive their respondent payment checks and duplicate addresses would have been discovered.

A number of issues may have affected respondents' decision to cooperate. Recruiting respondents to an Internet panel is harder than recruiting respondents to a single survey, and the study must be seen in this context. Words such as "referral code" added to the complexity of a questionnaire and also increased respondents' burden. Further, asking respondents to list their friends added another step to the recruiting process. While this was needed to choose friends at random, each step increases respondents' burden and this may have contributed to the outcome.

RDS recruiting into an Internet panel also needs to be seen through the opportunity that the ALP affords. The ALP is a well-established open-access Internet panel. Open access implies that surveys cover a wide variety of topics. Survey length varies but is typically of the order of 15–30 minutes. Survey frequency also varies; 1–2 surveys a month is typical. Respondents are surely motivated by the payments structure (US $20 for every 30 minutes of interview time). Relatively well-paid long-term panel members should contribute to successful recruitment in at least the first wave. The drop-off observed after the first wave for arm 5 (recruiter and recruit payment) might be explained by whether the recruiter is a long term panel member. Wave 0 (seed) recruiters were long-term panel members, whereas Wave 1 recruiters were not.

Finally, encouraged by an anonymous referee, we provide recommendations for the implementation of Internet RDS studies. First, experiment as much as possible. Likely fine tuning is required. Second, automate the recruiting process as much as possible; this will help to get back to potential recruits as soon as possible. Third, select seeds with great care to encourage recruiting. Fourth, allow self-selection of friends. While this is less than desirable, it is a strategy that can be used for pragmatic reasons in order to avoid the recruiting chains dying out. Fifth, for a general population, choose a high incentive/payment. Sixth, if necessary consider letting respondents invite more friends. One study (Bauermeister et al. 2012) increased the number of friends to ten. To control costs this study paid only for the first five friends who respond. Alternatively, one can argue paying for the last five friends who respond is preferable because it motivates the recruiter to get all invited friends to participate – but whether this works in practice requires empirical study. Again, this is a pragmatic suggestion where necessary and not our first choice.

Appendix

1) Numerical Question

"How many close friends or family members would you say you have? By close, we mean friends or family members you talk to or write to (via letter, email, text message,

Facebook, etc.) regularly. Please do not include people who live in your household. Please only consider people 18 years or older who live in the United States."

2) Friends question

"Please list all the close friends or family members you see, talk to or write to (via letter, email, text message, Facebook, etc.) regularly. Please **do not** include people who live in your household. Please only consider people 18 years or older who live in the United States. You only need to provide their first name, nickname or initials."

6. References

Arfken, C.L., Ahmed, S., and Abu-Ras, W. (2013). Respondent-Driven Sampling of Muslim Undergraduate U.S. College Students and Alcohol Use: Pilot study. Social Psychiatry and Psychiatric Epidemiology, 48, 945–953. DOI: http://www.dx.doi.org/10.1007/s00127-012-0588-4.

Bauermeister, J.A., Zimmerman, M.A., Johns, M.M., Glowacki, P., Stoddard, S., and Volz, E. (2012). Innovative Recruitment Using Online Networks: Lessons Learned from an Online Study of Alcohol and Other Drug Use Utilizing a Web-Based, Respondent-Driven Sampling (webrds) Strategy. Journal of Studies on Alcohol and Drugs, 73, 834–838.

Behr, A., Bellgardt, E., and Rendtel, U. (2005). Extent and Determinants of Panel Attrition in the European Community Household Panel. European Sociological Review, 21, 489–512. DOI: http://www.dx.doi.org/10.1093/esr/jci037.

Bengtsson, L., Lu, X., Nguyen, Q.C., Camitz, M., Hoang, N.L., Liljeros, F. and Thorron, A. (2012). Implementation of Web-Based Respondent-Driven Sampling Among Men Who Have Sex With Men in Vietnam. arXiv preprint arXiv:1206.1739. DOI: http://www.dx.doi.org/10.1371/journal.pone.0049417.

Berchenko, Y. and Frost, S.D. (2011). Capture-Recapture Methods and Respondent-Driven Sampling: Their Potential and Limitations. Sexually Transmitted Infections, 87, 267–268. DOI: http://www.dx.doi.org/10.1136/sti.2011.049171.

Berry, S., Bogdon, T., Brown, R., Corey, C., Hickey, S., Hill, L., and Schell, T. (2010). Survey of Gay Service Members. In Sexual Orientation and U.S. Military Personnel Policy, B.D. Rostker, S. Hosek, and J.D. Winkler (eds). Santa Monica, CA: RAND Corporation.

Burt, R.S. (1983). Distinguishing Relational Contents. Applied Network Analysis: A Methodological Introduction, Ronald S. Burt, Michael J. Minor, and Richard D. Alba (eds). Thousand Oaks: Sage Publications, 35–74.

Heckathorn, D. (1997). Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations. Social Problems, 44, 174–199. DOI: http://www.dx.doi.org/10.1525/sp.1997.44.2.03x0221m.

Heckathorn, D. (2002). Respondent-Driven Sampling ii: Deriving Valid Population Estimates from Chain-Referral Samples of Hidden Populations. Social Problems, 49, 11–34, DOI: http://www.dx.doi.org/10.1525/sp.2002.49.1.11.

Heckathorn, D. (2007). Extensions of Respondent-Driven Sampling: Analyzing Continuous Variables and Controlling for Differential Recruitment. Sociological

Methodology, 37, 151–207, DOI: http://www.dx.doi.org/10.1111/j.1467-9531.2007.00188.x.

Heckathorn, D., Semaan, S., Broadhead, R., and Hughes, J. (2002). Extensions of Respondent-Driven Sampling: A New Approach to the Study of Injection Drug Users Aged 18–25. AIDS and Behavior, 6, 55–67, DOI: http://www.dx.doi.org/10.1023/A:1014528612685.

Laumann, E.O. (1966). Prestige and Association in an Urban Community: An Analysis of an Urban Stratification System. New York: Bobbs-Merrill.

Marin, A. and Hampton, K.N. (2007). Simplifying the Personal Network Name Generator Alternatives to Traditional Multiple and Single Name Generators. Field Methods, 19, 163–193, DOI: http://www.dx.doi.org/10.1177/1525822X06298588.

Salganik, M. and Heckathorn, D. (2004). Sampling and Estimation in Hidden Populations Using Respondent-Driven Sampling. Sociological Methodology, 34, 193, DOI: http://www.dx.doi.org/10.1111/j.0081-1750.2004.00152.x.

Schonlau, M. and Liebau, E. (2012). Respondent-Driven Sampling. Stata Journal, 12, 72–93.

Wejnert, C. (2009). An Empirical Test of Respondent-Driven Sampling: Point Estimates, Variance, Degree Measures, and Out-of-Equilibrium Data. Sociological Methodology, 39, 73–116, DOI: http://www.dx.doi.org/10.1111/j.1467-9531.2009.01216.x.

Wejnert, C. (2010). Social Network Analysis With Respondent-Driven Sampling Data: A Study of Racial Integration on Campus. Social Networks, 32, 112–124, DOI: http://www.dx.doi.org/10.1016/j.socnet.2009.09.002.

Wejnert, C. and Heckathorn, D. (2008). Web-Based Network Sampling: Efficiency and Efficacy of Respondent-Driven Sampling for Online Research. Sociological Methods & Research, 37, 105–134, DOI: http://www.dx.doi.org/10.1177/0049124108318333.

Received February 2013
Revised October 2013
Accepted November 2013