Social Networks 33 (2011) 70–78

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

The game of contacts: Estimating the social visibility of groups☆

Matthew J. Salganik[a,*], Maeve B. Mello[b], Alexandre H. Abdo[c], Neilane Bertoni[b], Dimitri Fazito[d], Francisco I. Bastos[b,e]

[a] Department of Sociology and Office of Population Research, Princeton University, Princeton, NJ, USA
[b] Laboratory on Health Information, Oswaldo Cruz Foundation, Rio de Janeiro, Brazil
[c] Department of Physics, University of São Paulo, São Paulo, Brazil
[d] Department of Demography, Center of Development and Regional Planning, Federal University of Minas Gerais, Belo Horizonte, Brazil
[e] Fulbright/Capes Visiting Scholar, Brown University, Providence, RI, USA

ARTICLE INFO

Keywords:
Network scale-up method
Hidden populations
Network sampling
HIV/AIDS disease surveillance
Information flow

ABSTRACT

Estimating the sizes of hard-to-count populations is a challenging and important problem that occurs frequently in social science, public health, and public policy. This problem is particularly pressing in HIV/AIDS research because estimates of the sizes of the most at-risk populations—illicit drug users, men who have sex with men, and sex workers—are needed for designing, evaluating, and funding programs to curb the spread of the disease. A promising new approach in this area is the network scale-up method, which uses information about the personal networks of respondents to make population size estimates. However, if the target population has low social visibility, as is likely to be the case in HIV/AIDS research, scale-up estimates will be too low. In this paper we develop a game-like activity that we call *the game of contacts* in order to estimate the social visibility of groups, and report results from a study of heavy drug users in Curitiba, Brazil ($n = 294$). The game produced estimates of social visibility that were consistent with qualitative expectations but of surprising magnitude. Further, a number of checks suggest that the data are high-quality. While motivated by the specific problem of population size estimation, our method could be used by researchers more broadly and adds to long-standing efforts to combine the richness of social network analysis with the power and scale of sample surveys.

1. Introduction

Estimating the sizes of hard-to-count populations is important in social science, public health, and public policy. The problem of population size estimation is particularly pressing in HIV/AIDS research because reliable estimates of the sizes of the most-at risk populations—illicit drug users, sex workers, and men who have sex with men—are critical for designing, evaluating, and funding programs to curb the spread of the disease (UNAIDS, 2003). Unfortunately, estimating the sizes of these groups is extremely difficult. One promising approach, however, is the network scale-up method, which uses information about the personal networks of respondents to make population size estimates (Bernard et al., 1989, 1991, in press; Johnsen et al., 1995; Killworth et al., 1998a,b). But, because the method implicitly assumes that respondents know everything about everyone in their personal network, the network scale-up method can underestimate the sizes of groups where members keep their status "hidden," as is the case for the groups most at-risk for HIV/AIDS.

Therefore, in this paper we develop *the game of contacts*, a method to estimate the social visibility of groups defined by demographic characteristics, behaviors, or attitudes. We also report results from a study of heavy drug users in Curitiba, Brazil ($n = 294$). Because heavy drug users are a difficult population to survey and network data collection is often cognitively demanding and prone to error (Bernard et al., 1984; Marsden, 1990; Brewer, 2000), we formulated our data collection into a concrete, game-like activity involving playing cards, a game board,

☆ The authors thank Russ Bernard, Yea-Hung Chen, Dennis Feehan, Rob Lyerla, Mary Mahy, Chris McCarty, Tyler McCormick, and Tian Zheng for helpful discussions. We are grateful to all study participants and Curitiba's field work team, especially to the interviewers and the field coordinator Maria Lúcia Tozetto Vettorazzi. In addition, we thank Simone Tetu Moysés from PUC-Curitiba for her support and Francisca Lucena for providing the frequency of names from Brazilian Ministry of Health SIM database. This research was supported by The Joint United Nations Programme on HIV/AIDS (UNAIDS), the Brazilian Ministry of Health, the U.S. National Science Foundation (CNS-0905086), and the U.S. National Institutes of Health/NICHD (R01HD062366). The opinions expressed here represent the views of the authors and not the funding agencies.

* Corresponding author at: 145 Wallace Hall, Princeton, NJ 08544, USA. Tel.: +1 609 258 8867; fax: +1 609 258 180.
E-mail address: mjs3@princeton.edu (M.J. Salganik).

0378-8733/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.socnet.2010.10.006

and toy blocks. While introducing little extra cost or complexity, a number of data quality checks described in this paper suggest that our game lead to high-quality, consistent data. Further, the game of contacts is flexible enough that it could be generalized to collect data in very different contexts. Thus, while motivated by the problems of estimating hidden population size, we believe that our method could be used more broadly and is part of a long line of work that attempts to combine the richness of network analysis with the power and scale of sample surveys (Coleman, 1958; Barton, 1968; Granovetter, 1976; McCallister and Fischer, 1978; Burt, 1984; McCarty et al., 1997; Hogan et al., 2007).

This manuscript is organized as follows: Section 2 provides background on the network scale-up method. Section 3 describes our data collection procedure in Curitiba, Brazil. Section 4 presents the results. Because heavy drug users are a difficult population to survey, Section 5 presents a number of data quality and robustness checks. Finally, Section 6 concludes with a discussion of limitations, open questions, and directions for future research. Additional analysis is also available in the supporting online materials.

2. Background

The network scale-up method uses information about the personal networks of respondents to make population size estimates based on the assumption that personal networks are, on average, representative of the general population (Bernard et al., 1989, 1991, in press; Johnsen et al., 1995; Killworth et al., 1998a,b). Thus, information about the composition of many personal networks should allow for estimates of the composition of the general population. For example, if a respondent reports knowing 3 illicit drug users and knows 600 people overall, we can estimate that 3/600, or 0.5%, of the population are ilillicit drug users. This estimate can be improved by averaging over many respondents as described in Killworth et al. (1998b) and shown below.

The data needed for the scale-up method comes from interviews with a random sample of the general population. In addition to basic demographic questions, respondents in these surveys are asked how many people they "know" in the target population. In these studies, "know" is usually defined as follows: "you know them and they know you and you have been in contact with them in the last two years" (Bernard et al., in press). Respondents are then asked a battery of questions to estimate the number of people they know (i.e., their personal network size) (Killworth et al., 1998b; McCarty et al., 2001; McCormick et al., 2010). In all, the interviews take approximately 10 min and can be embedded within any nationally representative survey (McCarty et al., 2001).

To make population size estimates from this data, Killworth et al. (1998b) assumed a binomial model where the probability of a "success" (i.e., knowing someone in the target population) was $p = (N_t/N)$, where N_t is the size of the target population and N is the size of the general population. Thus, the likelihood function would be,

$$L(p|y_i, d_i) \propto \prod_i [(p)^{y_i} \cdot (1-p)^{(d_i - y_i)}] \quad (1)$$

where y_i is the number of people person i reports knowing in the target population and d_i is the number of people that person i knows.[1] Thus the maximum likelihood estimate for p is[2]

$$\hat{p} = \frac{\sum_i y_i}{\sum_i d_i} \quad (2)$$

Because the scale-up method uses data from a random sample of the general population it offers numerous logistical advantages over other methods for estimating the sizes of hidden populations. First, it is more amendable to standardization across time and across countries, a key feature for any method that is to be used for global public health surveillance. Second, it allows researchers to estimate the sizes of multiple target populations (e.g., illicit drug users, sex workers, and men who have sex with men) all in the same survey, whereas other methods such as capture-recapture would require three different data collections each with its own cost and distinct sources of error. A complete review of the advantages and disadvantages of the scale-up method is presented in Bernard et al. (in press).

One major limitation of the network scale-up method is that respondents do not know everything about the people in their personal networks. This problem is called "transmission error" in the scale-up literature because information about the characteristics of network members is not always "transmitted" to the survey respondents. This lack of information flow could be intentional—if members of stigmatized groups practice "information management" (Goffman, 1963) or "selective disclosure" (Kitts, 2003)—or unintentional—sometimes topics simply do not arise in conversation. Whether intentional or not, transmission error generally results in the number of people reported known in the target population (y_i in Eq. (2)) being too low, yielding underestimates of target population size. Estimating the magnitude of the transmission error and its effects on estimates, however, has proven difficult. For example, Shelley et al. (1995, 2006) conducted 3-h interviews with HIV-positive individuals which yielded rich insights into the social processes behind transmission error, but did not produce quantitative estimates that could be used to adjust population size estimates. Building on that work, Killworth et al. (2006) attempted an ambitious "alter-chasing" study that was ultimately unsuccessful due to logistical challenges. Most recently, Paniotto et al. (2009) collected data on the social stigma of groups as a proxy for transmission error, but did not measure transmission error directly or offer a pre-defined procedure for adjusting scale-up estimates.

Building on this earlier work, Salganik and Feehan (2009) offered a precise definition of transmission error and a procedure for adjusting population size estimates. They returned to the binomial model of Killworth et al. (1998b) (see Eq. (1)) and noted that a "success" actually involves two distinct steps. First, a respondent needs to be connected to a member of the target population, and second, the respondent needs to actually be aware that they are in fact in the target population. This second step is what researchers had meant by transmission error, but we find it easier to speak of *transmission rates*. The transmission rate can be defined as follows (using the standard network analysis terms *egos* and *alters*): consider the set of alters that are connected to the members of the target population, the transmission rate, τ, is the probability that a randomly chosen alter from this set will be aware that the ego they are connected to is in fact in the target population. More pre-

[1] For this paper we ignore the fact that the network size is estimated rather than known exactly. This is dealt with in some detail by Killworth et al. (1998b) and Salganik and Feehan (2009).

[2] Killworth et al. (1998b) multiplied $\hat{p}$ by the size of the population N (assumed to be known) to get an estimate of the size of the hidden population, N_t. In this paper, however, we choose to express results in terms of population prevalence which facilitates cross-national comparisons and is more typical in HIV/AIDS research (e.g., Cáceres et al., 2006).

cisely, let $a_{i,1}, a_{i,2}, \ldots a_{i,d_i}$ be the alters of person i where d_i is the personal network size of person i, and let T be set the set of people in the target population. The transmission rate is defined to be:

$$\tau = \frac{\sum_{i \in T} \sum_{j=1}^{d_i} I(a_{i,j})}{\sum_{i \in T} d_i} \tag{3}$$

where I is an indicator function that equals 1 if alter j is aware that ego i is in the target population and 0 otherwise. τ ranges from 0 for a completely invisible population to 1 for a completely visible population.

Therefore, Salganik and Feehan (2009) proposed the revised likelihood function:

$$L(p|y_i, d_i, \tau) \propto \prod_i [(p \cdot \tau)^{y_i} \cdot (1 - (p \cdot \tau))^{(d_i - y_i)}] \tag{4}$$

which yields the new estimator:

$$\hat{p} = \frac{\sum_i y_i}{\sum_i d_i} \cdot \frac{1}{\hat{\tau}} \tag{5}$$

The nature of the correction introduced, $1/\hat{\tau}$, makes it clear that the transmission rate can have a large impact on estimates.[3] For example, if $\hat{\tau} = 0.5$, a reasonable value, the current network scale-up estimator will under-estimate the size of the population by a factor of 2.

3. Data collection procedure

In this project we developed and implemented a method for estimating the transmission rate (τ) and various aspects of social visibility and network composition of specific groups. Although the network scale-up method requires a sample of the general population, our procedure requires a sample of the target population. Therefore, in order to use our resources most efficiently, we nested our data collection within a previously planned behavioral surveillance study of heavy drug users in Curitiba, Brazil, which was conducted by the Brazilian Ministry of Health and the Oswaldo Cruz Foundation as part of the 10-city Brazilian Behavioral Surveillance Study of heavy drug users (Bastos, 2009). For these studies, and our study, heavy drug users were defined to be individuals that have injected drugs at least once in the last 6 months and/or have used illicit drugs other than marijuana on at least 25 days during the last 6 months. Curitiba, a city of 1.8 million people, was a favorable location for the study because it has an existing official estimate for the size of its heavy drug user population. Our estimated transmission rate, therefore, could be combined with a future network scale-up study that we plan to conduct, and the resulting size estimates—both adjusted and unadjusted—could be compared to the existing official estimate.

Because it is not possible to select a simple random sample of heavy drug users, the behavioral surveillance study sampled heavy drug users via respondent-driven sampling (Heckathorn, 1997, 2002; Salganik and Heckathorn, 2004), a widely used sampling method for studying hidden or hard-to-reach populations (Malekinejad et al., 2008). Respondent-driven sampling data are collected through a chain referral process in which current sample members recruit future sample members. In Curitiba, the study began with an initial sample of five "seeds." After participating, the "seeds" were asked—and provided financial incentive—to recruit up to three other people in the target population. These new recruits were in turn encouraged to recruit others. The process continued in this way with current sample members recruiting the next wave of sample members until 303 people had been enrolled.[4]

Participants were interviewed at the PUC-Curitiba Psychiatric Hospital (*Hospital Nossa Senhora da Luz*), a hospital that is easily accessible by public transportation and has a tradition of providing services to drug users. They initially completed an audio computer-assisted interview (ACASI) that lasted approximately 1 h and collected information about life-history, drug use, and HIV-related knowledge, attitudes, and behavior. After this survey, participants took part in our nested data collection and voluntary counseling and testing for HIV/AIDS and Syphilis, with the sequence of these activities determined by the availability of staff and equipment. Before beginning the game of contacts, participants in our data collection were initially asked by our interviewers—interviewers that were not used in the larger study—several questions about their personal network size and participation in drug treatment programs. When describing our procedure in this paper, we follow standard network terminology and refer to the people we interview as *ego* and the people connected to them as *alters*.

To estimate the transmission rate, τ, we collected a sample of alters of members of the target population (see Eq. (3)) using a variation of the "first names" procedure described in McCarty et al. (1997). More concretely, the game of contacts began when the interviewer shuffled a deck of 24 playing cards. The respondent then drew a card from the pile and turned it over to reveal a name we had written on it (e.g., "Sebastião"). The respondent was then asked, "How many people do you know named Sebastião?"[5] Next the respondent picked up a token for each "Sebastião" that she knew and placed the tokens onto a board such that the location of each token indicated both whether the Sebastião under consideration did or did not use drugs and whether he did or did not know that the respondent used drugs (Fig. 1).[6] To reinforce the feel of a game, the board was brightly colored and respondents used toy blocks as tokens. After all the tokens were placed for a particular name, the interviewer recorded the location of the tokens, cleared the board, and the respondent drew another card from the deck. This process of drawing names and placing tokens continued until the respondent had been asked about 24 names (Table 1). Overall, the location of these tokens allows us to estimate various features of the network composition of the target population and the flow of information within it.

Two sources of concern are discussed briefly below and in greater detail later in the paper. First, the game of contacts relies

[3] The correction can also introduce additional uncertainty around the resulting estimates of $\hat{p}$ since the correction factor is itself estimated. A further discussion of variance estimation with the scale-up method is available in Salganik and Feehan (2009).

[4] The study was approved by the IRB of the School of Public Health of the Oswaldo Cruz Foundation (FWA00000389). Participants received approximately USD 10 as primary incentive and another USD 10 as secondary incentive for each participant recruited up to a maximum of three.

[5] For the game of contacts, the definition of "knowing" was: you know them by sight or by name, they know you by sight or by name, and you have gotten in touch in person, by phone, by mail or by e-mail at least once over the past two years. This definition matched the definition that, following standard practice (Bernard et al., in press), we will use in a network scale-up survey conducted in the general population. If the results from the game of contacts are going to be used to adjust results from a scale-up survey (see Eq. (5)), then the definition of "know" used in both must be the same.

[6] For the game of contacts, the definition of alter drug use was the same as the definition of drug use for enrollment into the study: individuals that have injected drugs at least once in the last 6 months and/or have used illicit drugs other than marijuana on at least 25 days during the last 6 months.

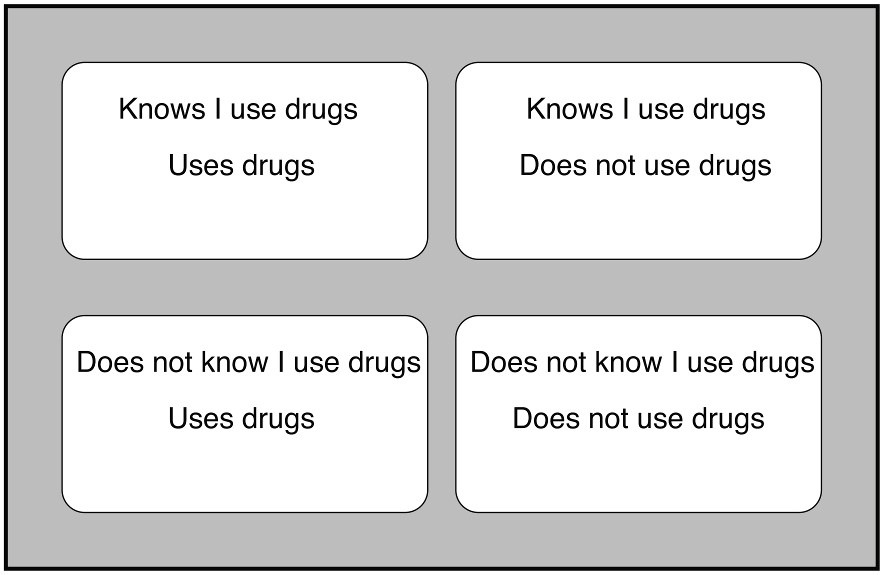

Fig. 1. Playing board for the game of contacts. Alters of participants were sampled by asking questions such as "How many people do you know named Sebastião?" For each alter known, the respondent would pick up a token and place it in the appropriate space on the board indicating whether the alter uses drugs and whether the alter knows that the ego uses drugs.

extensively on proxy reporting. In other words, consider the dyad between ego i and alter j, we collect i's proxy report about j's drug use as well as i's proxy report about j's proxy report about i's drug use. Ideally we would directly interview the alters that are sampled, but that is impractical given the difficulty of finding and interviewing the mentioned alters. The possible biases introduced by the use of proxy reports are discussed in Section 6. Second, the results from the game of contacts depend on the names that are used, and if the names are not selected properly, it could result in what McCarty et al. (1997) call *alter selection bias*. For example, if the names selected all belong to men, the set of alters learned about will only consist of men, or if the names selected tend to belong older individuals, the set of alters will be biased toward older people. Unfortunately, we could not locate information about the popularity or characteristics of people with specific first names living in Curitiba. Instead, we chose first names from mortality records from Curitiba using a procedure that attempted to minimize the possibility of alter selection bias.[7] Initially, we stratified first names into six groups: low-popularity female, middle-popularity female, high-popularity female, low-popularity male, middle-popularity male, and high-popularity male. Names were assigned to strata such that about 1/3 of all males had names in the low-popularity male strata; 1/3 of all males had names in the middle-popularity male strata; and 1/3 of all males had names in the high-popularity male strata. From within strata, we picked the names with the smallest variation in year-to-year popularity in an attempt to avoid names that were particularly trendy. If a name was commonly used as part of a compound name (e.g., Maria Helena) or was used by men and women (e.g., Alison), the name was discarded and another name was selected. This procedure resulted in the 24 first names presented in Table 1. The possible biases introduced by this procedure and set of first names are discussed and evaluated empirically in Section 5.5.

[7] The mortality data was provided from a larger national mortality database (the SIM database). To maintain privacy, we did not have access to the last names in this database.

Table 1
Names that were used in the game of contacts. Based on mortality data from the city of Curitiba, all names were stratified into six groups: low-popularity female, middle-popularity female, high-popularity female, low-popularity male, middle-popularity male, and high-popularity male. From within each strata, we selected four names that had the smallest year-to-year variation in popularity.

Female names			Male names		
Low	Middle	High	Low	Middle	High
Albertina	Alzira	Antônia	Acir	Benedito	Carlos
Doralice	Hilda	Cecília	Amadeu	Orlando	Pedro
Filomena	Iracema	Regina	Silvestre	Osvaldo	Mário
Wanda	Lídia	Rosa	Waldir	Sérgio	Sebastião

4. Results

The data set analyzed here includes 294 interviews conducted between July 28, 2009 and October 16, 2009 (80 days). In total, our game generated 4173 alters where the median number of alters per respondent was 10 (Fig. 2).[8] Table 2 presents a cross-tabulation of the data.

Although previous researchers have considered transmission error, again, we find it easier to express results in terms of *transmission rate*. In order to estimate transmission rate from our data, we must account for the fact that the sample was selected via respondent-driven sampling which means that respondents vary in their probability of selection (Salganik and Heckathorn, 2004; Volz and Heckathorn, 2008). Therefore, we estimate the transmission rate as

$$\hat{\tau} = \frac{\sum_i (w_i/\pi_i)}{\sum_i (x_i/\pi_i)} \qquad (6)$$

[8] There appears to be one outlier in the distribution. We checked the original survey form for this respondent and found that there was not a data entry error. Because we have no evidence of an error and because personal network size is known to be skewed (McCarty et al., 2001; Zheng et al., 2006; McCormick et al., 2010), we have opted to include this respondent in the analysis.

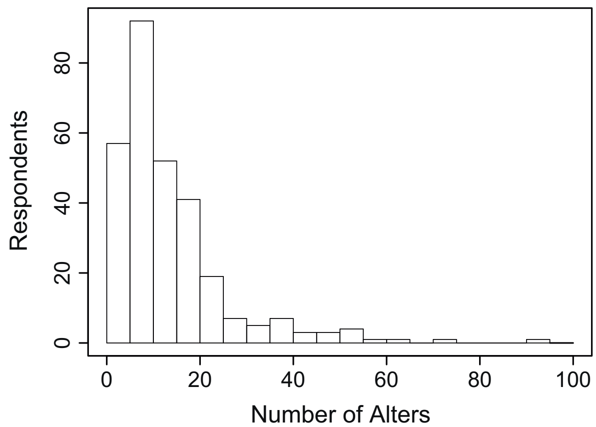

Fig. 2. Histogram of number of alters generated per respondent. The median number of alters generated by our procedure was 10 (mean = 14.2).

Table 2
Cross-classification of the reported characteristics of the 4173 alters generated by the game of contacts.

	Alter uses drugs	Alter does not use drugs
Alter aware that ego uses drugs	2082	1156
Alter not aware that ego uses drugs	225	710

where w_i is the number of alters that are aware that respondent i is in the target population, x_i is the total number of alters for respondent i, and π_i is the probability of selection for respondent i. Following standard practice with respondent-driven sampling, we assume that the probability of selection, π_i, is proportional to the number of people respondent i knows in the target population (Salganik and Heckathorn, 2004). This personal network size information was collected as part of the larger behavioral surveillance study, as described in greater detail in the supporting online materials.

Applying the estimator in Eq. (6) to our data, we estimate $\hat{\tau} = 0.76$. To place a confidence interval around this estimate, we preform a bootstrap procedure, described more fully in the supporting online materials, which produced an estimated 95% confidence interval of [0.72, 0.80]. In the supporting online materials, we show that this estimate, as well as all other weighted estimates presented in the paper, are qualitatively similar to unweighted estimates or estimates made using a different weighting approach (Figs. A4–A9).

In addition to the overall transmission rate, we can use the placement of the blocks on the game board to estimate other features of the social environment around heavy drug users in Curitiba. For example, Kitts (2003) predicted that stigmatized groups will tend to have both "selective exposure" (i.e., people who engage in stigmatized behavior are more likely to associate with others who engage in the same stigmatized behavior) and "selective disclosure" (i.e., people who engage in stigmatized behavior are more likely to share that information with other people who also engage in the same stigmatized behavior). We see evidence in support of both of these predictions in our data. An estimated 55% of the alters of heavy drug users are drug users (estimated 95% confidence interval [0.50, 0.60]). Since this is well above the proportion of heavy drug users in the population, this is strong evidence of selective exposure and is consistent with predictions based on theories of social homophily more generally (McPherson et al., 2001). We also see evidence of differential information transmission depending on the drug use behavior of the alter. When the alter is a drug user, the estimated transmission rate, $\widehat{\tau_U}$, is 0.88 (estimated 95% confidence interval [0.82, 0.92]); in other words, we estimate that about 90% of drug

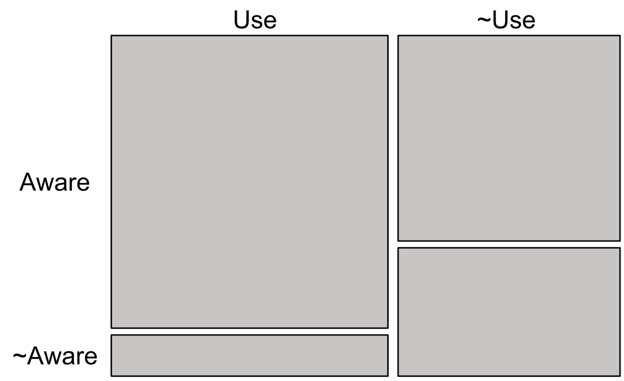

Fig. 3. Mosaic plot of estimates from the game of contacts revealing selective exposure and selective disclosure. The width of each block represents the estimated proportion of alters that use drug (Use) and do not use drugs (∼Use). The height of each block represents the proportion of alters in each category that are aware (Aware) and not aware (∼Aware) of ego's drug use.

using alters are aware that ego is a drug user. When the alter is a not a drug user, the estimated transmission rate is much lower, $\widehat{\tau_{\sim U}} = 0.62$ (estimated 95% confidence interval [0.56, 0.67]). The relationship between the two transmission rates is as expected: the social visibility to in-group members is higher than to out-group members. The evidence of selective exposure and selective disclosure is summarized in Fig. 3.

The match between our results and previous theoretical predictions provides some support for the validity of the game of contacts. However, while the qualitative pattern of these estimates might have been expected, their magnitudes were unknown. For example, the overall transmission rate was higher than we initially expected, but we think this is partially because a large proportion of alters of the target population were also drug users. Further, the target population was *heavy* drug users, a group for whom hiding their behavior may be particularly difficult. Other potential explanations for this finding are discussed in Section 6.

We can also further compare transmission rates for different types of egos and alters. For example, Fig. 4 plots the transmission rate of respondents who reported using, at least once in the past 12 months, each of nine drugs. Overall, the differences were quite small where the heavy drug users who used marijuana were estimated to be the most visible, $\hat{\tau} = 0.78$ (estimated 95% confidence interval [0.75, 0.82]), and the heavy drug users who used cocaine paste were estimated to be the least visible, $\hat{\tau} = 0.72$ (estimated 95% confidence interval [0.62, 0.84]). Further, we can harness the fact that 12 of our names lead to male alter and 12 lead to female alters to assess how transmission rate depends jointly on the char-

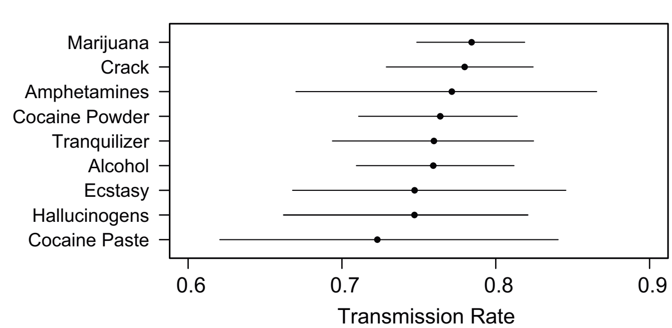

Fig. 4. Transmission rate for heavy drug users who used each drug at least once in the past 12 months. The figure shows a slight gradient of visibility for users of different drugs with heavy drug users who used marijuana estimated to be the most visible and heavy drug users who used cocaine paste estimated to be the least visible. Note that these categories are not mutually exclusive (e.g., some participants used both crack and marijuana).

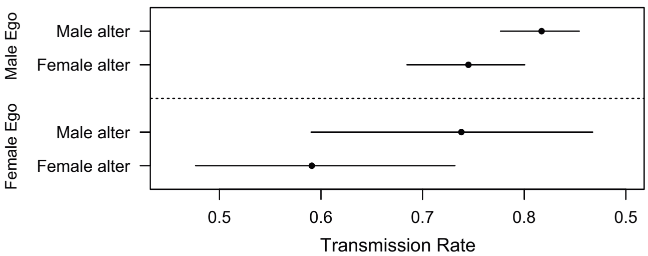

Fig. 5. Transmission rate by ego gender and alter gender. For both men and women, the transmission rate is higher to male alters.

acteristics of the ego and the alter. For example, Fig. 5 shows that independent of ego gender, the transmission rate tends to be higher to male alters than female alters. Results that further separate alters by both gender and drug user are presented in the supporting online material (Fig. A10). The ability to compare results for subsets of egos and alters shows the richness of what can be learned with the game of contacts beyond the overall transmission rate needed to adjust the scale-up estimates.

5. Data quality issues

Heavy drug users are a difficult population to survey and network data is notoriously challenging to collect (Marsden, 1990; Brewer, 2000; McCarty et al., 2007), so a number of data quality concerns naturally arise in a study such as this. Fortunately, many of these concerns can be addressed empirically using the data we collected, and we find that the data appear to be high-quality. These findings suggest that our “game,” while extremely cheap and simple, was able to turn the complex and boring process of network data collection into something interesting for both respondents and interviewers.

5.1. Question order effects

Our data collection had a median length of 11 min (25th percentile: 10 min; 75th percentile: 15 min) so there is a concern that responses to the early questions could be systematically different from those of the later questions due to learning, fatigue, or boredom. However, our procedure of shuffling the cards, and thus name order, allows us to detect question order effects, and we do not see strong evidence of them in our data. The number of alters generated per question does not seem to increase or decrease (Fig. 6a) or become more or less variable (Fig. 6b) as the survey progresses. The estimated transmission rate, however, does seem to show a slightly non-monotonic pattern as the survey progress (Fig. 6c), but since the magnitude of this pattern is small and it is hard to develop simple mechanisms that could explain it (unlike, for example, a linear increase or decrease), we do not think it represents serious cause for concern. If future researchers using the game of contacts observe this same non-monotic pattern, additional investigation would be called for.

5.2. Interviewer effects

While the study was underway we added a second interviewer in order to minimize wait times for participants, and by comparing results across interviewers we can measure interviewer effects for our procedure (Groves, 1989, Chapter 8). Previous research has found strong interviewer effects in network data collection in the general population (Van Tilburg, 1998; Marsden, 2003), and one might expect interviewer effects would be even larger in this context. However, we find surprisingly little difference between the data generated by the two interviewers. For this analysis we consider only the 222 interviews conducted after the second interviewer was added on August 26, 2009; during this time each interviewer completed 111 interviews.[9]

Interviewers A and B generated similar number of alters per interview (means of 12.5 and 14.9) and had similar variability in the number of alters per interview (standard deviations of 11.0 and 12.6); the complete histograms are plotted in the supporting online material (Fig. A11). The estimated transmission rates from each interviewer were also similar. From interviewer A we estimate $\hat{\tau} = 0.73$ (estimated 95% interval [0.62, 0.83]) and from interviewer B we estimate $\hat{\tau} = 0.76$ (estimated 95% interval [0.70, 0.81]). This difference is not statistically significant ($p = 0.78$). Results from both interviewers are presented visually in Fig. 7.

Two notes of caution are required. First, both interviewers had similar demographic characteristics—females in their 20’s—so it might be the case that interviewer effects would be detected with a more diverse pool of interviewers. Second, we did not explicitly randomize the assignment of respondents to interviewers. However, because our study procedure had interviewers alternate between participants, we think it is unlikely that there was a systematic difference between the type of participants encountered by the interviewers.

5.3. Nonresponse and missing data

Nonresponse—at both the level of the individual (unit nonresponse) and the level of the question (item nonresponse)—is a potentially serious threat to data quality (Groves et al., 2002). In our data, fortunately, we had little nonresponse. More specifically, 303 people completed the larger behavioral surveillance study and were thus eligible to participate. Of these 303, six completely refused to participate in our study and three participated in our initial data collection but refused to participate in the game of contacts. Of the 294 participants who began the game of contacts, 293 completed the process and one broke off the interview. Thus, we had very little unit nonresponse; 293 out of 303 eligible participants (97%) completed the entire data collection procedure. Further, of the 294 initiated game of contacts, there were a possible 28,224 responses (294 interviews $\times$ 24 names $\times$ 4 quadrants on the game board). Of these, only 71 responses are missing (0.3% of responses), and 36 of these missing responses are due to the one respondent who broke off the interview.

5.4. Relationship to external data

In addition to the internal consistency checks described above, we can also investigate how the data collected in the game of contacts compares to external sources of data. The first external source of data we compare to is the information on the popularity of the names that, as described previously, comes from a national mortality database. We find that the total number of alters known with each name is generally consistent with our pre-existing popularity categories. For example, based on the mortality data, Carlos was considered a high-popularity male name and Amadeu a low-popularity male name. This is consistent with the data from the game of contacts where respondents reported knowing a total of 584 alters named Carlos and only 67 alters named Amadeu. More generally, the median number of alters known for the high, middle, and low popularity males names were 367.5, 152, and 91.5. Similarly, the median number of alters known for the high, middle, and low popularity female names were 149, 87, 37.5, again consistent with expectations. A plot of the number of alters generated

[9] There is one interview during this period where the interviewer data is missing.

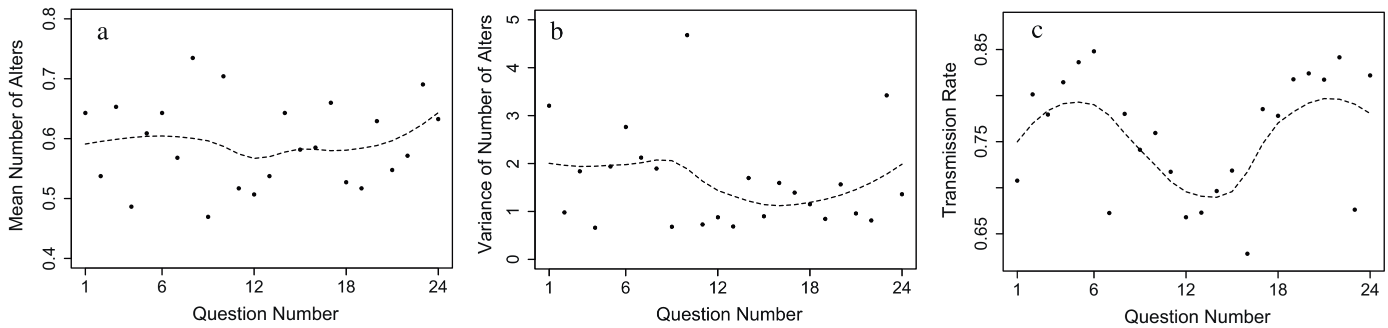

Fig. 6. Mean of number of alters generated (a); variance of number of alters generated (b); and transmission rate (c) by question order; dashed line is a loess curve. There does not seem to be a strong, consistent pattern as the survey processes.

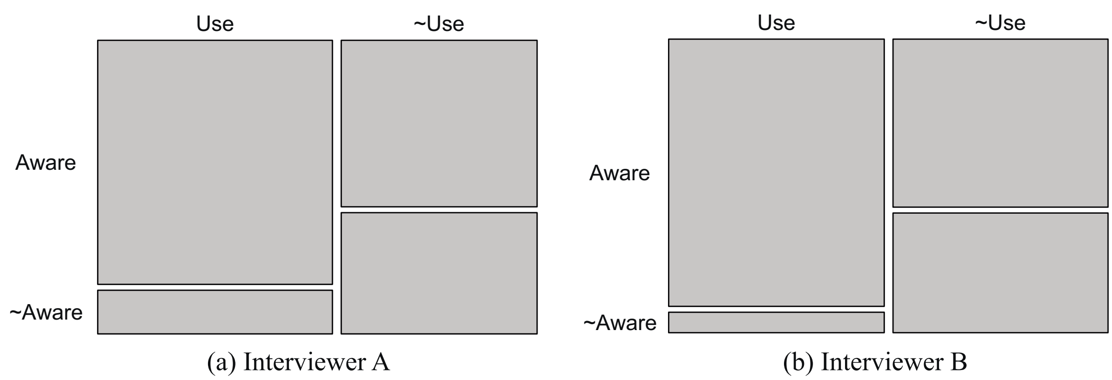

Fig. 7. Mosaic plot of estimates from interviewers A and B. Use and ∼Use refer to the drug use status of the alter and Aware and ∼Aware refer to the awareness that alter has about the behavior of ego. The estimates are similar from the two interviewers.

with each name is available in the supporting online materials (Fig. A12).

The second source of external data is the self-reported number of drug users known by the respondent, information that was collected during the larger behavioral surveillance study and was used to weight the respondent-driven sampling data. Following standard practice in respondent-driven sampling studies (Johnston et al., 2008), the interviewer in the larger study asked each respondent a series of four sequential network size questions in order to quickly approximate their personal network size in the target population (i.e., the number of heavy drug users each respondent knows). We found a positive correlation between these four self-reported measures of number of drug users known and the number of drug using alters generated by the game of contacts (Spearman's $\rho = 0.20, 0.18, 0.16, 0.22$; Pearson's $r = 0.14, 0.14, 0.14, 0.16$). Scatter plots of these relationships are presented in the supporting online material (Fig. A13). Overall, these two checks show that the data collected from the game of contacts is consistent with external sources of data, and thus provides some further support for the game of contacts.

5.5. Choice of names

One final concern is that our procedure for choosing the names resulted in a non-random set of alters, what McCarty et al. (1997) called alter selection bias (see also McCormick et al., 2010). For example, if we picked names that tended to belong to older people, alters involving those demographics would be over-represented in our sample, potentially biasing estimates. These concerns are particularly relevant because our names were selected from mortality records, a data source that is biased toward older people. To assess the possibility that the alters selected with one specific name might be different from the alters selected using the other names, we calculated the transmission rate for each name and found them to vary from 0.41 (Albertina) to 0.89 (Mário) (Fig. A16). Many of these estimates, however, have large uncertainty due to the small number of alters with low-popularity names. Therefore, to assess whether any name had a substantial impact on our estimates, we dropped each name from our sample in succession, thereby creating 24 distinct estimates each based on 23 names (e.g., the first estimate is based on all names except Pedro, the second estimate is based on all names except Carlos, etc). This procedure, akin to the jackknife procedure, produced replicate estimates of $\hat{\tau}$ ranging from 0.746 (when Carlos is dropped) to 0.766 (when Lidia is dropped). Thus, we see no evidence that any particular name we used had a substantial effect of on our estimates. The supporting online material includes mosaic plots of the alters with each name (Figs. A14 and A15) and the estimated transmission rate for each name (Fig. A16).

6. Discussion and limitations

Overall, the results from the game of contacts with heavy drug users in Curitiba, Brazil are encouraging. As predicted by Kitts (2003), we see evidence "selective exposure" and "selective disclosure." Further, the data seems to be high-quality: there is little missing data, almost no evidence of question order effects or interviewer effects, and the data collected during the game of contacts are consistent with external data.

However, there are numerous reasons for caution when interpreting data from the game of contacts. Our procedure measures target population members *perception* of network composition and social visibility, not ground-truth. While these perceptions may be more important than actual behavior in some cases (Marsden, 1990; Kitts, 2003), we speculate that there will be both systematic and random differences between the two. Random differences are less of a concern because our estimates are about heavy drug users in Curitiba as a group, not any particular individual, and, therefore, involve a large amount of averaging.[10] However, we suspect that there may be systematic error as well which could cause more serious problems. First, ego's reports about alter's drug use may be inflated because of the "false consensus effect" (Ross et al., 1977; Marks and Miller, 1987; Kitts, 2003). For example, in a network study of reports about contraceptive use in Kenya, White and Watkins (2000) found that respondents tended to project their behavior onto their network alters. Other evidence about the inaccuracy of reports about alters can be found in the literature on social networks (Laumann, 1969; Huckfeldt et al., 2000; Goel et al., 2010) and the literature on proxy reports in surveys (Moore, 1988; Blair et al., 1991). Second, ego's reports about alter's knowledge about ego may be inflated because of the "illusion of transparency" (Gilovich et al., 1998). However, the magnitudes of these cognitive biases in our context is unknown. We hope, therefore, that more research will be done on the accuracy, not just consistency, of social perceptions even given the tremendous logistical difficulties involved (see, for example, Goel et al., 2010). Systematic error could also be introduced by the choice of first names used. In this case, we do not see evidence that any particular name influenced our estimates very much, but this might not be the case in future studies. A final concern is that the game of contacts only measures one type of transmission error: false-negatives (i.e., cases where a person is connected to an alter that is a heavy drug user but is not aware the alter is a heavy drug user). However, false-positives are also possible; for example, a respondent in a network scale-up survey who is connected to an alter that uses marijuana once per month might report knowing a heavy drug user even though that alter is not a heavy drug user according to our study criteria. Thus, just as false-negatives can lead to population size estimates that are too low, false-positives can lead to estimates that are too high. Procedures for estimating the rate of false-positives and empirical work measuring the balance between false-positives and false-negatives are topics that we believe deserve further research.

Another set of problems could be caused by the fact that our sample was collected using respondent-driven sampling and our data collection was nested within the larger behavioral surveillance study. First, it is possible that the voluntary counseling and testing for HIV/AIDS and Syphilis or the 1-h long survey affected how respondents answered questions during the game of contacts. We could not assess this empirically with our data, but future researchers could detect this by randomizing the order of these different study components. Second, it is possible that not all of alters in the game of contacts were unique. For example, two respondents who report knowing one Sebastião could actually be referring to the same person. This is unlikely if respondents are selected at random from a large city, but possible if respondents know each other as happens in respondent-driven sampling studies. Thus, while it appears that we have information about 247 unique Sebastiãos, some of these could be duplicates adding additional uncertainty to our estimates. Third, our weighting scheme could introduce bias if the probability of selection is not proportional to self-reported network size, as could occur if personal network size is reported inaccurately (Wejnert, 2009) or due to violation of respondent-driven sampling modeling assumptions (Gile and Handcock, 2010). It is also possible that "bottlenecks" in the social network of heavy drug users in Curitiba prevented the respondent-driven sampling process from exploring all parts of the population of heavy drug users in Curitiba (Goel and Salganik, 2009, 2010). We have no evidence that this took place, but with the data that we have, we cannot rule out that possibility. Finally, it is possible that the most "hidden" heavy drug users refused to participate in the study for fear that their drug use would become public. This type of non-response would lead our estimates of the transmission rate to be too high. Again, we have no evidence that this actually occurred, but since the recruitment was done by participants and not researchers, we do not have good information about people who refused to participate. If the game of contacts is to be used in routine public health surveillance, then it will frequently be nested in larger studies that recruit participants with respondent-driven sampling. Therefore, we think that interactions between the game of contacts and the larger research context should be explored further.

Given these many possible sources bias and uncertainty, researchers may wonder if they should adjust their network scale-up estimate of population sizes using the estimated transmission rate. Ideally this decision would be made based on whether the adjusted estimate will have lower mean square error than the unadjusted estimate. Unfortunately, it is not known precisely when this will be the case. Therefore, we offer the following rough rule-of-thumb: if the transmission rate is estimated accurately or over-estimated, researchers should almost certainly adjust, but if the transmission rate is under-estimated, the decision is more complicated. In the case where the transmission rate is estimated accurately it will almost certainly make sense to use the adjusted estimate, even though it will have more variability than unadjusted estimate (in order to account for the fact that τ was estimated). If the estimated transmission rate is too high, it is still probably wise to use the adjusted estimate because, while not correct, the adjusted estimate will be closer to the true value than the unadjusted estimate. If the transmission rate is too low, the decision of whether or not to use adjusted estimates depends in some complex fashion on the true transmission rate, the estimated transmission rate, and the relative costs of under-estimating and over-estimating the size of the target population. A more precise specification of the conditions under which the adjusted population size estimate (Eq. (5)) will have lower mean square error than the unadjusted estimate (Eq. (2)) is a topic that could be explored more fully in future papers, especially as more data becomes available from other studies.

One somewhat novel aspect of our data collection was reformulating our survey as a concrete, game-like activity. We suspect that this resulted in higher-quality data, while only adding modestly to our costs. Focusing on participant enjoyment is common in online social research (e.g., von Ahn and Dabbish, 2008; Salganik and Watts, 2009; Goel et al., 2010), and we expect it will become more common in standard survey research. Therefore, we hope that future research will more directly explore whether enjoyable data collection leads to more accurate data.

We conclude by noting that the game of contacts can be used in other research contexts and can be generalized to collect other types of social network data. For example, the labels on game board (Fig. 1) could be adjusted to collect any two categorical pieces of information about the alter. Further, the game board could be turned into a series of concentric circles allowing for the collection of the strength of tie between ego and alter with virtually no additional effort (see, for example, Hogan et al., 2007). Finally, by recording the sequence that respondents place the tokens, not just their final configuration, it might be possible to learn additional information about how respondents organize social information in their memory (Brewer, 1995). Because the local environment around individuals and groups is so important, we hope that the game of contacts, and its generalizations, will be used in a wide range of social research.

[10] At the individual-level, a rough estimate for the standard error of an estimate from the game of contacts would be $p(1-p)/n$ where n is the relevant number of alters. An additional complication of analysis at the individual-level is that the accuracy of the estimate for each individual varies depending on the number of alters generated by the game.

Acknowledgment

All data analyzed in this paper are available from the data archive of the Office of Population Research at Princeton University.

Appendix A. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.socnet.2010.10.006.

References

Barton, A.H., 1968. Bringing society back in: survey research and macro-methodology. American Behavorial Scientist 12 (2), 1–9.

Bastos, F.I., 2009. Taxas de infecçao de HIV e sífilis e inventário de conhecimento, atitudes e práticas de risco relacionadas às infecçoes sexualmente transmissíveis entre usuários de drogas em 10 municípios brasileiros. Tech. Rep., Fundaçao Oswaldo Cruz. Brazilian Multicity Study Group on Drug Misuse.

Bernard, H.R., Hallett, T., Iovita, A., Johnsen, E.C., Lyerla, R., McCarty, C., Mahy, M., Salganik, M.J., Saliuk, T., Scutelniciuc, O., Shelley, G.A., Sirinirund, P., Stroup, D.F., Weir, S., in press. Counting hard-to-count populations: the network scale-up method for public health. Sexually Transmitted Infections, http://sti.bmj.com/content/86/Suppl_2/ii11.abstract.

Bernard, H.R., Johnsen, E.C., Killworth, P., Robinson, S., 1991. Estimating the size of an average personal network and of an event subpopulation: some empirical results. Social Science Research 20, 109–121.

Bernard, H.R., Johnsen, E.C., Killworth, P.D., Robinson, S., 1989. Estimating the size of an average personal network and of an event subpopulation. In: Kochen, M. (Ed.), The Small World. Ablex Publishing, pp. 159–175.

Bernard, H.R., Killworth, P., Kronenfeld, D., Sailer, L., 1984. The problem of informant accuracy: the validity of retrospective data. Annual Review of Anthropology 13, 495–517.

Blair, J., Menon, G., Bickart, B., 1991. Measurement effects in self vs. proxy response to survey questions: an information-processing perspective. In: Biemer, P., Groves, R.M., Lyberg, L.E., Mathiowetz, N.A., Sudman, S. (Eds.), Measurement Errors in Surveys. John Wiley & Sons, New York, pp. 145–166 (Chapter 9).

Brewer, D.D., 1995. The social structural basis of the organization of persons in memory. Human Nature 6 (4), 379–403.

Brewer, D.D., 2000. Forgetting in the recall-based elicitation of person and social networks. Social Networks 22, 29–43.

Burt, R.S., 1984. Network items and the general social survey. Social Networks 6, 293–339.

Cáceres, C., Konda, K., Pecheny, M., Chatterjee, A., Lyerla, R., 2006. Estimating the number of men who have sex with men in low and middle income countries. Sexually Transmitted Infections 82 (Suppl. 3), 3–9.

Coleman, J.S., 1958. Relational analysis: the study of social organizations with survey methods. Human Organization 17 (4), 28–36.

Gile, K., Handcock, M., 2010. Respondent-driven sampling: an assessment of current methodology. Sociological Methodology 40 (1), 285–327.

Gilovich, T., Savitsky, K., Medvec, V.H., 1998. The illusion of transparency: biased assessments of others' ability to read one's emotional states. Journal of Personality and Social Psychology 75 (2), 332–346.

Goel, S., Mason, W., Watts, D.J., 2010. Real and perceived attitude homophily in social networks. Journal of Personality and Social Psychology 99 (4), 611–621.

Goel, S., Salganik, M.J., 2009. Respondent-driven sampling as Markov chain Monte Carlo. Statistics in Medicine 28 (17), 2202–2229.

Goel, S., Salganik, M.J., 2010. Assessing respondent-driven sampling. Proceedings of the National Academy of Science, USA 107 (15), 6743–6747.

Goffman, E., 1963. Stigma: Notes on the Management of Spoiled Identity. Prentice-Hall, Englewood Cliffs, NJ.

Granovetter, M., 1976. Network sampling: some first steps. American Journal of Sociology 81 (6), 1287–1303.

Groves, R., 1989. Survey Errors and Survey Costs. John Wiley & Sons, Hoboken, NJ.

Groves, R., Dillman, D., Eltinge, J., Little, R., 2002. Survey Nonresponse. John Wiley & Sons, New York.

Heckathorn, D.D., 1997. Respondent-driven sampling: a new approach to the study of hidden populations. Social Problems 44 (2), 174–199.

Heckathorn, D.D., 2002. Respondent-driven sampling II: deriving valid population estimates from chain-referral samples of hidden populations. Social Problems 49 (1), 11–34.

Hogan, B., Carrasco, J., Wellman, B., 2007. Visualizing personal networks: working with participant-aided sociograms. Field Methods 19 (2), 116–144.

Huckfeldt, R., Sprague, J., Levine, J., 2000. The dynamics of collective deliberation in the 1996 election: campaign effects on accessibility, certainty, and accuracy. The American Political Science Review 94 (3), 641–651.

Johnsen, E., Bernard, R., Killworth, P., Shelley, G., McCarty, C., 1995. A social network approach to corroborating the number of AIDS/HIV+ victims in the US. Social Networks 17 (3–4), 167–187.

Johnston, L., Malekinejad, M., Kendall, C., Iuppa, I., Rutherford, G., 2008. Implementation challenges to using respondent-driven sampling methodology for HIV biological and behavioral surveillance: field experiences in international settings. AIDS and Behavior 12 (S1), 131–141.

Killworth, P.D., Johnsen, E.C., McCarty, C., Shelly, G.A., Bernard, H.R., 1998a. A social network approach to estimating seroprevalence in the United States. Social Networks 20, 23–50.

Killworth, P.D., McCarty, C., Bernard, H.R., Shelly, G.A., Johnsen, E.C., 1998b. Estimation of seroprevalence, rape, and homelessness in the U.S. using a social network approach. Evaluation Review 22, 289–308.

Killworth, P.D., McCarty, C., Johnsen, E.C., Bernard, H.R., Shelley, G.A., 2006. Investigating the variation of personal network size under unknown error conditions. Sociological Methods & Research 35 (1), 84–112.

Kitts, J., 2003. Egocentric bias or information management? Selective disclosure and the social roots of norm misperception. Social Psychology Quarterly 66 (3), 222–237.

Laumann, E.O., 1969. Friends of urban men: an assessment of accuracy in reporting their socioeconomic attributes, mutual choice, and attitude agreement. Sociometry 32 (1), 54–69.

Malekinejad, M., Johnston, L., Kendall, C., Kerr, L., Rifkin, M., Rutherford, G., 2008. Using respondent-driven sampling methodology for HIV biological and behavioral surveillance in international settings: a systematic review. AIDS and Behavoir 12 (Supp. 1), 105–130.

Marks, G., Miller, N., 1987. Ten years of research on the false-consensus effect: an empirical and theoretical review. Psychological Bulletin 102 (1), 72–90.

Marsden, P.V., 1990. Network data and measurement. Annual Review of Sociology 16, 435–463.

Marsden, P.V., 2003. Interviewer effects in measuring network size using a single name generator. Social Networks 25 (1), 1–16.

McCallister, L., Fischer, C.S., 1978. A procedure for surveying personal networks. Sociological Methods & Research 7 (2), 131–148, November.

McCarty, C., Bernard, H.R., Killworth, P.D., Shelley, G.A., Johnsen, E.C., 1997. Eliciting representative samples of personal networks. Social Networks 19 (4), 303–323, October.

McCarty, C., Killworth, P.D., Bernard, H.R., Johnsen, E., Shelley, G.A., 2001. Comparing two methods for estimating network size. Human Organization 60, 28–39.

McCarty, C., Killworth, P.D., Rennell, J., 2007. Impact of methods for reducing respondent burden on personal network structural measures. Social Networks 29 (2), 300–315.

McCormick, T., Salganik, M.J., Zheng, T., 2010. How many people do you know? Efficiently estimating personal network size. Journal of the American Statistical Association 105 (489), 59–70.

McPherson, M., Smith-Lovin, L., Cook, J.M., 2001. Birds of a feather: homophily in social networks. Annual Review of Sociology 27, 415–444.

Moore, J.C., 1988. Self/proxy response status and survey response quality, a review of the literature. Journal of Official Statistics 4 (2), 155–172.

Paniotto, V., Petrenko, T., Kupriyanov, V., Pakhok, O., 2009. Estimating the size of populations with high risk for HIV using the network scale-up method. Tech. Rep., Kiev Internation Institute of Sociology.

Ross, L., Greene, D., House, P., 1977. The "false consensus effect": an egocentric bias in social perception and attribution processes. Journal of Experimental Social Psychology 13 (3), 279–301.

Salganik, M.J., Feehan, D., 2009. Generalized network scale-up method for estimating the size of hard-to-count populations. Tech. Rep. Office of Population Research, Princeton University.

Salganik, M.J., Heckathorn, D.D., 2004. Sampling and estimation in hidden populations using respondent-driven sampling. Sociological Methodology 34, 193–239.

Salganik, M.J., Watts, D.J., 2009. Web-based experiments for the study of collective social dynamics in cultural markets. Topics in Cognitive Science 1 (3), 439–468.

Shelley, G.A., Bernard, H.R., Killworth, P., Johnsen, E., McCarty, C., 1995. Who knows your HIV status? What HIV+ patients and their network members know about each other. Social Networks 17 (3–4), 189–217, October.

Shelley, G.A., Killworth, P.D., Bernard, H.R., McCarty, C., Johnsen, E.C., Rice, R.E., 2006. Who knows your HIV status II? Information propagation within social networks of seropositive people. Human Organization 65 (4), 430–444.

UNAIDS, 2003. Estimating the Size of Popualtions at Risk for HIV. No. 03.36E. UNAIDS, Geneva.

Van Tilburg, T., 1998. Interviewer effects in the measurement of personal network size: a nonexperimental study. Sociological Methods & Research 26 (3), 300–328, February.

Volz, E., Heckathorn, D.D., 2008. Probability-based estimation theory for respondent-driven sampling. Journal of Official Statistics 24 (1), 79–97.

von Ahn, L., Dabbish, L., 2008. Designing games with a purpose. Communications of the ACM 51 (8), 58–67.

Wejnert, C., 2009. An emperical test of respondent-driven sampling: point estimates, variance, degree measures, and out-of-equalibrium data. Sociological Methodology 39 (1), 73–116.

White, K., Watkins, S.C., 2000. Accuracy, stability and reciprocity in informal conversational networks in rural Kenya. Social Networks 22, 337–355.

Zheng, T., Salganik, M.J., Gelman, A., 2006. How many people do you know in prison?: Using overdispersion in count data to estiamte social structure in networks. Journal of the American Statistical Association 101, 409–423.