Assessing respondent-driven sampling

Sharad Goel[a,1] and Matthew J. Salganik[b,1]

[a]Microeconomics and Social Systems, Yahoo! Research, 111 West 40th Street, New York, NY, 10018; and [b]Department of Sociology and Office of Population Research, Princeton University, Wallace Hall, Princeton, NJ 08544

Respondent-driven sampling (RDS) is a network-based technique for estimating traits in hard-to-reach populations, for example, the prevalence of HIV among drug injectors. In recent years RDS has been used in more than 120 studies in more than 20 countries and by leading public health organizations, including the Centers for Disease Control and Prevention in the United States. Despite the widespread use and growing popularity of RDS, there has been little empirical validation of the methodology. Here we investigate the performance of RDS by simulating sampling from 85 known, network populations. Across a variety of traits we find that RDS is substantially less accurate than generally acknowledged and that reported RDS confidence intervals are misleadingly narrow. Moreover, because we model a best-case scenario in which the theoretical RDS sampling assumptions hold exactly, it is unlikely that RDS performs any better in practice than in our simulations. Notably, the poor performance of RDS is driven not by the bias but by the high variance of estimates, a possibility that had been largely overlooked in the RDS literature. Given the consistency of our results across networks and our generous sampling conditions, we conclude that RDS as currently practiced may not be suitable for key aspects of public health surveillance where it is now extensively applied.

disease surveillance | snowball sampling | social networks

The development and evaluation of public health policies often require detailed information about so-called hard-to-reach or hidden populations. For example, HIV researchers are especially interested in monitoring risk behavior and disease prevalence among injection drug users, men who have sex with men, and commercial sex workers—the groups at highest risk for HIV in most countries. Unfortunately, however, these high-risk groups are not easily studied with standard sampling methods, including institutional sampling, targeted sampling, and time-location sampling (1).

Respondent-driven sampling (RDS) (2–4) facilitates examination of such hidden populations via a chain-referral procedure in which participants recruit one another, akin to snowball sampling. RDS is now widely used in the public health community and has been recently applied in more than 120 studies in more than 20 countries, involving a total of more than 32,000 participants (5). In particular, in helping to track the HIV epidemic, RDS is used by the Centers for Disease Control and Prevention (CDC) (6, 7) and by the United States President's Emergency Plan for AIDS Relief.

RDS is a method both for data collection and for statistical inference. To generate an RDS sample, one begins by selecting a small number of initial participants ("seeds") from the target population who are asked—and typically provided financial incentive—to recruit their contacts in the population (2). The sampling proceeds with current sample members recruiting the next wave of sample members, continuing until the desired sample size is reached. Participants are usually allowed to recruit up to three other contacts in order to ensure sampling continues even if some sample members do not recruit.

With the RDS sampling design, individuals with more contacts in the target population are more likely to be recruited (4). To adjust for this selection bias, respondents are weighted inversely proportional to their network degree or number of contacts.

Specifically, for any individual trait f (e.g., age), the RDS estimate $\hat{\mu}_f$ of the population mean of f is defined to be

$$\hat{\mu}_f = \frac{1}{\sum_{i=1}^{n} 1/\text{degree}(X_i)} \sum_{i=1}^{n} \frac{f(X_i)}{\text{degree}(X_i)},\qquad [1]$$

where $X_1, \ldots, X_n$ are the n participants in the study.[†] Typically, f is 0–1, for example, indicating infectivity of a specific disease, in which case $\hat{\mu}_f$ estimates prevalence of the trait in the target population (8, 10).

The accuracy of RDS estimates is affected by the structure of the underlying social network, the distribution of traits within the network, and the recruitment dynamics. In particular, RDS can perform poorly when traits cluster in cohesive subpopulations (10)—a phenomenon that may be especially acute in the case of infectious diseases (11). Gauging the combined effect of these factors has proven difficult, and previous attempts to assess the performance of RDS have been largely inconclusive.

The earliest evaluations of RDS come from simulation studies on synthetic networks (4, 8, 12). These studies, however, likely overestimated the accuracy of RDS by neglecting to design synthetic networks that adequately mimic the community structure of real social networks (10). Recently, Gile and Handcock (13) have evaluated the bias of RDS estimates on more realistic synthetic networks. Beyond simulation studies, there have been three approaches for assessing the quality of RDS. First, RDS has been carried out on a population with known characteristics: undergraduates at a large residential university (9, 14). Because only a single RDS sample was taken, however, it is difficult to ascertain sample-to-sample variability of estimates.[‡] Second, RDS estimates have been compared to estimates derived from alternative sampling methods (15–21) (Table S1). These comparisons have not yielded consistent patterns and are hindered by the fact that true population values remain unknown.[§] Finally, the

STATISTICS

stability of repeated cross-sectional RDS estimates has been examined, again yielding ambiguous results. For example, in studies of men who have sex with men (MSM) in Beijing, China in 2004, 2005, and 2006 (22), year-to-year RDS estimates for age and employment status were stable, whereas estimates of education and sexual orientation were suspiciously volatile.[¶]

In contrast to previous approaches, we evaluate the performance of RDS by simulating the sampling and estimation process on 85 real populations mapped in two previous studies. In all cases, both the network structure of the population and demographic traits for each individual are available. We are thus able to directly compare empirical RDS estimates to true population values and, in particular, to measure the variability of estimates.

Data and Methods
Our first source of data, Project 90, was a large, multiyear study that began in 1987 as a prospective examination of the influence of network structure on the propagation of infectious disease (23). As such, researchers attempted to construct a network census of high-risk heterosexuals in Colorado Springs, focusing particularly on sex workers and drug injectors and their sexual and drug partners (23–26). We restrict attention to the giant component of the network, comprising 4,430 individuals and 18,407 edges, representing social, sexual, and drug affiliation. Our second data source, the National Longitudinal Study of Adolescent Health (Add Health), mapped the friendship networks of 84 middle and high schools in the United States (27–29). The giant components for these school networks range in size from 25 to 2,539 students—with a median size of 753—and collectively include a total of 72,262 individuals and 258,688 edges.

Rather than attempting to model the complex social dynamics that play out during the RDS recruitment process, in our simulations we assume the same, idealized sampling conditions considered in the theoretical RDS literature (4, 8, 10, 30). Specifically, (*i*) initial sample members are chosen independently and proportional to network degree; (*ii*) relationships within the population are symmetric (i.e., if *A* is a contact of *B*, then *B* is also a contact of *A*); (*iii*) participants recruit uniformly at random from their contacts; (*iv*) those who are recruited always participate in the study; (*v*) individuals can be recruited into the sample more than once; (*vi*) the number of recruits per participant does not depend on individual traits; and (*vii*) respondents accurately report their social network degree. The remaining parameters of our simulations are modeled after common RDS study features (5). Starting from ten initial seeds, each participant recruits between 0 and 3 other individuals. The exact recruitment distribution mimics an RDS study of drug injectors in Tijuana and Ciudad Juarez, Mexico (31), in which 1/3 of participants recruited no one, 1/6 recruited one other participant, 1/6 recruited two other participants, and 1/3 recruited three other participants, the maximum allowed. The simulated recruitment procedure continues until a sample size of 500 is reached. The entire sampling process was repeated 10,000 times on each network to generate replicate estimates.

Results
From each simulated sample, the RDS estimator (Eq. **1**) was used to infer the population proportion of a given trait—for example, the proportion of drug dealers in the Project 90 network or the proportion of students on the soccer team in a particular Add Health school. Consistent with theoretical results (8, 10), we find RDS generates approximately unbiased estimates: Across all networks and traits, both the mean and median bias are less than 0.0005.

The variability of RDS estimates, however, is significantly larger than generally acknowledged. We quantify this variability in terms of design effect (32), which benchmarks the performance of RDS against that of simple random sampling (SRS). Specifically, the design effect d is $\mathrm{Var}(\hat{p}_{\mathrm{RDS}})/\mathrm{Var}(\hat{p}_{\mathrm{SRS}})$, where $\hat{p}_{\mathrm{RDS}}$ is the RDS estimate and $\hat{p}_{\mathrm{SRS}}$ is the estimate obtained from SRS. It follows that an RDS estimate with sample size n and design effect d has the same variance as a simple random sample of size n/d. A design effect of 10, for example, effectively reduces an RDS sample of nominal size 500 to an SRS sample of size 50.

Consistently large design effects are seen in both Project 90 and Add Health (Fig. 1). The 13 binary traits in Project 90 have design effects that range from 5.7 to 58.3, with a median design effect of 11.0. As a consequence, estimating the 17% unemployment rate in Project 90, which has a design effect of 10, with reasonable precision ($\pm 5\%$, 95% confidence) requires an RDS sample of approximately 2,300 people—a sample size 5 times larger than in nearly all previous RDS studies (5). We observe a similar phenomenon for each of the 46 binary traits in Add Health. The median design effect for traits ranges from 4.2 to 14.4, where for each trait the median is taken across the 84 Add Health schools. The overall median design effect for all traits in all schools is 5.9.

All traits in the Project 90 and the Add Health networks yield design effects larger than what is commonly assumed in the planning stages of RDS studies. A review of 91 studies found that more than half assumed a design effect less than 1.5, and all assumed a design effect less than 2.5 (5). Furthermore, a rule-of-thumb design effect of 2 had been suggested by Salganik (12). Given that we find typical design effects greater than 5, even
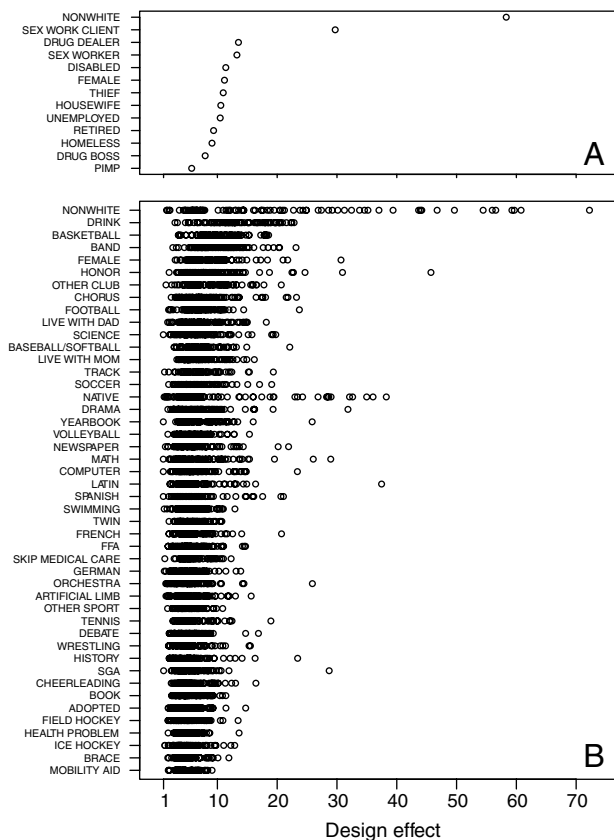

Fig. 1. Design effects for the 13 binary traits in Project 90 (*A*) and the 46 binary traits in Add Health (*B*); in Add Health, each circle indicates the design effect of a trait in one of 84 schools.

under optimistic sampling conditions, it is likely that many RDS studies do not have sufficiently large sample sizes to meet their stated study objectives. In particular, the CDC, in its assessment of RDS for use in the National HIV Behavioral Surveillance System, concludes that a sample of size 500 is adequate to estimate traits with 5% precision (95% confidence) (33), implicitly assuming a design effect of 1. Though valid if subjects are chosen by SRS, our results suggest an RDS sample of 500 may lead to errors of 10% or more.

Compounding these large design effects, the standard RDS confidence intervals developed by Salganik (12, 34) are often misleadingly narrow, masking the effects of inadequate sample sizes (Fig. 2). In Project 90, the coverage probability of nominal 95% confidence intervals ranges across traits from 42% to 65%, with median coverage of 52%. In other words, though 95% of such intervals should contain the true population proportion, in our simulations only about half of the intervals in fact do. We find similar results for the Add Health networks, where nominal 95% intervals have coverage probability across traits ranging from 44% to 72%, with median coverage of 62%. It thus seems likely that, in a substantial fraction of field studies, true population values fall outside the reported ranges.

To put our results in context, we compare the RDS estimator (Eq. 1)—which weights sample members inversely proportional to their network degree—to the unweighted mean of the RDS sample, an estimation method that mimics traditional snowball sampling and that is widely considered problematic for public health surveillance (35, 36). As shown in Fig. 3, though the sample mean generally has larger bias than the RDS estimator, the two estimators are comparable in terms of their standard error and their overall performance, as quantified by root-mean-squared error (RMSE). Specifically, in Add Health the

Fig. 3. Comparison of bias, standard error, and RMSE between the RDS estimator and the unweighted mean of the RDS sample in Project 90 (*A–C*) and Add Health (*D–F*). Each point corresponds to a given trait in a given network population.

median absolute bias, standard error, and RMSE of the sample mean are 0.005, 0.023, and 0.025, respectively, compared to 0.0004, 0.026, and 0.026, respectively, for the RDS estimator; and in Project 90 the analogous values for the sample mean are 0.022, 0.035, and 0.041, respectively, compared to 0.0004, 0.034, and 0.034, respectively, for RDS.

Helping to explain why RDS and the sample mean perform similarly, we note that weighted and unweighted means differ to the extent that weights and traits are correlated, a fact that is well-documented in the survey sampling literature (37). Thus the advantages of RDS are manifest primarily when network degree correlates with, for example, drug use, sexual behavior, or other traits of interest. At least in the datasets we examine, however, we see mostly weak correlations and thus generally find weighting to have minimal impact. In Add Health, where traits and degree have median absolute correlation of 0.05, RDS and the sample mean perform almost identically: The median difference in RMSE (RMSE of the sample mean minus RMSE of the RDS estimator) is −0.001 across all school-trait pairs. Ten of the thirteen traits in Project 90 are likewise weakly correlated with degree—median absolute correlation among these ten is 0.06—and RDS and the sample mean are correspondingly similar, with a median difference in RMSE of 0.005. The remaining traits—indicating, respectively, whether an individual is a drug dealer, a sex worker, or unemployed—are more highly correlated with degree (0.25, 0.30, and 0.31), and the relative value of RDS is more pronounced, with reductions in RMSE of 0.06, 0.09, and 0.13. Though, even in these three cases where RDS substantially

Fig. 2. Coverage rates of nominal 95% RDS confidence intervals in Project 90 (*A*) and Add Health (*B*). For each trait in Add Health, the mean coverage over all 84 schools is shown.

STATISTICS

outperforms the sample mean, RDS is still far from ideal, with design effects greater than 10 for all three traits.

In the above analysis, we considered the performance of RDS for 85 previously mapped real social networks. To check the robustness of our results, we simulate RDS on several variants of these network populations. We find that large design effects persist, suggesting that our results extend beyond the particular populations we study and hold more generally. First, a significant structural anomaly with the Project 90 network is the large number of so-called leaf nodes (i.e., individuals connected to exactly one other node), which comprise 18% of the giant component. These leaves often correspond to individuals who were identified in the study by another participant but who themselves were not directly interviewed. Repeating our analysis on the Project 90 subnetwork that excludes these leaves, we find no appreciable change in the performance of RDS, with a median design effect of 10.8 compared to 11.0 in the original network (Fig. S2). Second, of the 84 Add Health networks, 44 represent joint middle and high schools whereas the remaining 40 are exclusively high schools. In the joint middle and high schools there is substantial segregation between these two constituent subpopulations (29, 38, 39), and as such, one may worry that our results are driven by this single, atypical, network feature. We find, however, that the distribution of design effects in the joint schools is approximately the same as in those that are strictly high schools, with median design effects of 5.9 and 6.0, respectively (Fig. S3). Third, students in the Add Health study were asked to name up to five male and five female friends, and in our primary analysis we inferred a symmetric edge between students A and B if either individual named the other. Alternatively, one could consider only reciprocal nominations—inferring a symmetric tie between A and B only if A and B both nominate one another—that potentially correspond to stronger friendships (29).$^{\parallel}$ We find that the performance of RDS on these reciprocal-nomination networks is in fact considerably worse, with a median design effect of 18.9, compared to 5.9 for the one-sided-nomination networks (Fig. S4). For a final robustness check, we confirm that large design effects are not driven by obvious structural or demographic properties of the target populations. Specifically, across the 84 Add Health schools, we find that design effect is weakly correlated with both school size (0.07) and the true population proportion of the trait being estimated (0.27) (Fig. S5).**

Discussion
Past work has emphasized that RDS in theory generates approximately unbiased estimates (4, 8, 30)—and we indeed find this to be the case in our simulations. However, by neglecting to consider the variance, this result has been widely interpreted as indicating that RDS has low error. Explicitly examining the variance of RDS, we find that estimates are much less precise than previously believed. In particular, RDS as currently practiced may be poorly suited for important aspects of public health surveillance where it is now extensively applied. For example, to reliably detect a decline in unsafe injection practice from 40% to 30% with SRS requires approximately 350 people at each of the two time points.†† With a design effect of 5—a typical finding in our simulations—the required sample size jumps to 1750, substantially larger than nearly all RDS studies (5). Consequently, it seems that many existing studies do not have sufficient power to identify even quite large changes in behavior and almost certainly could not identify with statistical confidence small changes in disease prevalence.

Our findings are subject to two potential objections. First, the Project 90 and the Add Health networks are not perfect representations of hidden populations. In particular, networks of high school students are unlikely to be representative of networks of populations at high risk for HIV, and the Project 90 network, although it maps such a hidden population, probably suffers from significant missing data. We attempted to mitigate this shortcoming by analyzing two distinct datasets, each with different limitations, and by analyzing several thousand network-trait pairs (46 traits × 84 school networks from Add Health, and 13 traits in the Project 90 network), thus decreasing the chance that our findings are driven by anomalous features of any one trait or network. Furthermore, we considered several modified versions of these networks to check for robustness to structural perturbations. The qualitative consistency of our results suggests that the observed high variance of RDS estimates is the norm rather than the exception.

Second, in order to simulate recruitment, we followed the same idealized sampling design assumed in the theoretical RDS literature; it is thus possible that RDS could perform better under real-world sampling conditions than it does in our simulated environment. It seems much more likely, however, that actual RDS sampling dynamics only exacerbate the problems indicated by our results (13). Specifically, initial participants are generally a convenience sample and are almost certainly not chosen in the judicious, independent manner of our simulations (40). There is also evidence of nonrandom recruitment of peers and of differential participation and recruitment rates (40–44). In a study of MSM in Brazil (45), for example, participants were more likely to recruit those who they thought engaged in riskier behavior and who would therefore most benefit from HIV testing; this same study found that some individuals refused to participate for fear of disclosing their sexual orientation. Evidence of differential recruitment rates was seen in a study of jazz musicians in New York, with women on average recruiting more than 60% more participants than men (30). Furthermore, in practice, and in contrast to our simulations, participants are prohibited from entering an RDS study multiple times, a policy intended to deter fraudulent recruitment tactics, but one that also generally increases the bias of estimates (13). Finally, ascertaining an individual's network degree is a challenging problem (46–48)—particularly in the context of RDS (40)—and so self-reported network size represents a possibly significant source of nonsampling error absent from our simulations.

Conclusion
Simulating sampling across 85 real social networks, we find the variance of RDS is typically 5–10 times greater than that of SRS and, moreover, that standard RDS confidence intervals are misleadingly narrow. In light of our generous sampling assumptions, and the robustness of our results to network perturbations, it is likely that RDS will perform no better—and may perhaps perform considerably worse—in field studies. In particular, our results highlight the considerable obstacles facing RDS in applications such as disease surveillance. By clarifying the limitations of RDS, we hope to encourage its further development, systematic evaluation, and cautious application.

$^{\parallel}$This change reduces the median degree for students in a giant component from 7 for the original, one-sided-nomination networks to 3 for the reciprocal-nomination networks.

**For any trait, the design effect for estimating the proportion p of students having the trait is the same as for estimating the proportion $1-p$ not having the trait; for example, the design effect for estimating the proportion of white students is the same as the design effect for estimating the proportion of nonwhite students. As a result, we find that $\bar{p} = \min(p, 1-p)$ is slightly more predictive of design effect than the proportion p itself, with correlation 0.35 between design effect and $\bar{p}$, compared to 0.27 between design effect and p.

††The sample size estimate is based on a standard power calculation with $\alpha = 0.05$ and $\beta = 0.80$.

1. Magnani R, Sabin K, Saidel T, Heckathorn D (2005) Review of sampling hard-to-reach and hidden populations for HIV surveillance. *AIDS* 19:S67–S72.
2. Heckathorn DD (1997) Respondent-driven sampling: A new approach to the study of hidden populations. *Soc Probl* 44:174–199.
3. Heckathorn DD (2002) Respondent-driven sampling II: Deriving valid population estimates from chain-referral samples of hidden populations. *Soc Probl* 49:11–34.
4. Salganik MJ, Heckathorn DD (2004) Sampling and estimation in hidden populations using respondent-driven sampling. *Sociol Methodol* 34:193–239.
5. Malekinejad M, et al. (2008) Using respondent-driven sampling methodology for HIV biological and behavioral surveillance in international settings: A systematic review. *AIDS Behav* 12:105–130.
6. Lansky A, et al. (2007) Developing an HIV behavioral surveillance system for injecting drug users: The National HIV Behavioral Surveillance System. *Public Health Rep* 122:48–55.
7. Lansky A, Drake A (2009) HIV-associated behaviors among injecting-drug users—23 cities, United States, May 2005–February 2006. *Morb Mort Wkly Rep* 58:329–332.
8. Volz E, Heckathorn DD (2008) Probability-based estimation theory for respondent-driven sampling. *J Off Stat* 24:79–97.
9. Wejnert C (2009) An empirical test of respondent-driven sampling: Point estimates, variance, degree measures, and out-of-equilibrium data. *Sociol Methodol* 39:73–116.
10. Goel S, Salganik MJ (2009) Respondent-driven sampling as Markov chain Monte Carlo. *Stat Med* 28:2202–2229.
11. Friedman SR, et al. (2000) Network-related mechanisms may help explain long-term HIV-1 seroprevalence levels that remain high but do not approach population-group saturation. *Am J Epidemiol* 152:913–922.
12. Salganik MJ (2006) Variance estimation, design effects, and sample size calculations for respondent-driven sampling. *J Urban Health* 83:98–112.
13. Gile KJ, Handcock MS (2010) Respondent-driven sampling: An assessment of current methodology. *Sociol Methodol*, in press.
14. Wejnert C, Heckathorn DD (2008) Web-based network sampling: Efficiency and efficacy of respondent-driven sampling for on-line research. *Sociol Methods Res* 37:105–134.
15. Platt L, et al. (2006) Methods to recruit hard-to-reach groups: Comparing two chain referral sampling methods of recruiting injection drug users across nine studies in Russia and Estonia. *J Urban Health* 83:39–53.
16. Robinson WT, et al. (2006) Recruiting injection drug users: A three-site comparison of results and experiences with respondent-driven and targeted sampling procedures. *J Urban Health* 83:29–38.
17. Burt RD, Hagan H, Sabin K, Thiede H (2010) Evaluating respondent-driven sampling in a major metropolitan area: Comparing injection drug users in the 2005 Seattle area National HIV Behavioral Surveillance System with injectors in the RAVEN and Kiwi studies. *Ann Epidemiol* 20:159–167.
18. Abdul-Quader AS, et al. (2006) Effectiveness of respondent-driven sampling for recruiting drug users in New York City: Findings from a pilot study. *J Urban Health* 83:459–476.
19. Kendall C, et al. (2008) An empirical comparison of respondent-driven sampling, time location sampling, and snowball sampling for behavioral surveillance in men who have sex with men, Fortaleza, Brazil. *AIDS Behav* 12:97–104.
20. Johnston LG, Trummal A, Lohmus L, Ravalepik A (2009) Efficacy of convenience sampling through the internet versus respondent driven sampling among males who have sex with males in Tallin and Harju County, Estonia: Challenges reaching a hidden population. *AIDS Care* 21:1195–1202.
21. Ramirez-Valles J, Heckathorn DD, Vázquez R, Diaz RM, Campbell RT (2005) From networks to populations: The development and application of respondent-driven sampling among IDUs and Latino gay men. *AIDS Behav* 9:387–402.
22. Ma X, et al. (2007) Trends in prevalence of HIV, syphilis, hepatitis C, hepatitis B, and sexual risk behavior among men who have sex with men. Results of 3 consecutive respondent-driven sampling surveys in Beijing, 2004 through 2006. *JAIDS* 45:581–587.
23. Potterat JJ, et al. (2004) *Network Epidemiology: A Handbook for Survey Design and Data Collection*, ed M Morris (Oxford Univ Press, Oxford), pp 87–114.
24. Woodhouse DE, et al. (1994) Mapping a social network of heterosexuals at high risk for HIV infection. *AIDS* 8:1331–1336.
25. Klovdahl AS, et al. (1994) Social networks and infectious disease: The Colorado Springs study. *Soc Sci Med* 38:79–88.
26. Rothenberg RB, et al. (1995) *Social Networks, Drug Abuse, and HIV Transmission*, eds RH Needle, S Coyle, SG Genser, and RT Trotter (Natl Inst on Drug Abuse, Rockville, MD), Vol 151, pp 3–19.
27. Moody J (2001) Race, school integration, and friendship segregation in America. *Am J Sociol* 107:679–716.
28. Bearman PS, Moody J, Stovel K, Thalji L (2004) *Network Epidemiology: A Handbook for Survey Design and Data Collection*, ed M Morris (Oxford Univ Press, Oxford), pp 201–220.
29. Goodreau S, Kitts JA, Morris M (2009) Birds of a feather or friend of a friend? Using exponential random graph models to investigate adolescent friendship networks. *Demography* 46:103–125.
30. Heckathorn DD (2007) Extensions of respondent-driven sampling: Analyzing continuous variables and controlling for differential recruitment. *Sociol Methodol* 37:151–208.
31. Frost SDW, et al. (2006) Respondent-driven sampling of injection drug users in two US.-Mexico border cities: Recruitment dynamics and impact on estimate of HIV and Syphilis prevalence. *J Urban Health* 83:83–97.
32. Kish L (1965) *Survey Sampling* (Wiley, New York).
33. Gallagher KM, Sullivan PS, Lansky A, Onorato IM (2007) Behavioral surveillance among people at risk for HIV infection in the U.S.: The National HIV Behavioral Surveillance System. *Public Health Rep* 122(Suppl 1):32–38.
34. Volz E, Wejnert C, Degani I, Heckathorn DD (2007) *Respondent-Driven Sampling Analysis Tool (RDSAT)* (Cornell University, Ithaca, NY), Version 5.6.
35. Kalton G, Anderson DW (1986) Sampling rare populations. *J R Stat Soc A* 149:65–82.
36. Thompson SK, Collins LM (2002) Adaptive sampling in research on risk-related behaviors. *Drug Alcohol Depen* 68:S57–S67.
37. Kish L (1992) Weighting for unequal p_i. *J Off Stat* 8:183–200.
38. Moody J (2005) Add health network structure files. (Technical Report, Carolina Population Center).
39. Gonzalez M, Herrmann H, Kertesz J, Vicsek T (2007) Community structure and ethnic preferences in school friendship networks. *Physica A* 379:307–316.
40. 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 Behav* 12:131–141.
41. Johnston LG, et al. (2008) The effectiveness of respondent driven sampling for recruiting males who have sex with males in Dhaka, Bangladesh. *AIDS Behav* 12:294–304.
42. Scott G (2008) "They got their program, and I got mine": A cautionary tale concerning the ethical implications of using respondent-driven sampling to study injection drug users. *Int J Drug Policy* 19:42–51.
43. Poon AFY, et al. (2009) Parsing social network survey data from hidden populations using stochastic context-free grammars. *PLoS ONE* 4:e6777.
44. Abramovitz D, et al. (2009) Using respondent-driven sampling in a hidden population at risk of HIV infection: Who do HIV-positive recruiters recruit?. *Sex Transm Dis* 36:750–756.
45. de Mello M, et al. (2008) Assessment of risk factors for HIV infection among men who have sex with men in the metropolitan area of Campinas City, Brazil, using respondent-driven sampling. (Population Council), Technical report.
46. McCarty C, Killworth PD, Bernard HR, Johnsen EC, Shelley GA (2001) Comparing two methods for estimating network size. *Hum Organ* 60:28–39.
47. Zheng T, Salganik MJ, Gelman A (2006) How many people do you know in prison?: Using overdispersion in count data to estimate social structure in networks. *J Am Stat Assoc* 101:409–423.
48. McCormick T, Salganik MJ, Zheng T (2010) How many people do you know?: Efficiently estimating personal network size. *J Am Stat Assoc*, in press.

STATISTICS