ANNUAL REVIEWS

Annual Review of Statistics and Its Application

Methods for Inference from Respondent-Driven Sampling Data

Krista J. Gile,[1] Isabelle S. Beaudry,[2] Mark S. Handcock,[3] and Miles Q. Ott[4]

[1]Department of Mathematics and Statistics, University of Massachusetts, Amherst, Massachusetts 01003, USA; email: gile@math.umass.edu

[2]Department of Statistics, Pontificia Universidad Católica de Chile, 7820436 Santiago, Chile

[3]Department of Statistics, University of California, Los Angeles, California 90095, USA

[4]Statistical and Data Science Program, Smith College, Northampton, Massachusetts 10163, USA

Annu. Rev. Stat. Appl. 2018. 5:4.1–4.29

The *Annual Review of Statistics and Its Application* is online at statistics.annualreviews.org

https://doi.org/10.1146/annurev-statistics-031017-100704

Keywords

respondent-driven sampling, link-tracing, not missing at random, network sampling, social network, hard-to-reach population, snowball sampling

Abstract

Respondent-driven sampling is a commonly used method for sampling from hard-to-reach human populations connected by an underlying social network of relations. Beginning with a convenience sample, participants pass coupons to invite their contacts to join the sample. Although the method is often effective at attaining large and varied samples, its reliance on convenience samples, social network contacts, and participant decisions makes it subject to a large number of statistical concerns. This article reviews inferential methods available for data collected by respondent-driven sampling.

1. INTRODUCTION TO RESPONDENT-DRIVEN SAMPLING (RDS)

Respondent-driven sampling (RDS) is a network-based sampling method designed to sample hard-to-reach human populations connected by an underlying social network of relations. Introduced by sociologist Douglas Heckathorn (Heckathorn 1997), RDS begins with a small sample of seeds known to the researchers. Each participant then receives a small number of uniquely identified coupons to pass to their contacts in the target population, making them eligible for the study and enlarging the sample until the desired sample size is reached. This is a specialization of the widely used class of so-called snowball sampling network sampling techniques, often used as a last-resort convenience sampling approach when no conventional sampling frame is available (Handcock & Gile 2011). Although versions of snowball sampling (Goodman 1961) are subject to clear statistical principles, in practice snowball sampling typically involves beginning with a convenience sample, then soliciting and including all relevant contacts of each person sampled. This results in a great deal of dependence among the resulting samples.

RDS provides two major advantages over more conventional snowball sampling: First, because respondents receive only a small number of coupons, a fixed sample size will sample farther from the seeds than a typical snowball sample of equivalent size, reducing the dependence between samples, including between the seeds and the final sample. This makes it more reasonable to approximate the final sample as a probability sample and allows for more reasonable statistical inference.

Second, relying on the respondents to pass coupons to conduct recruitment reduces the confidentiality concerns associated with asking respondents to enumerate their contacts in a potentially stigmatized population. This innovation makes RDS an attractive approach for sampling from populations that are both hard-to-reach and stigmatized, including men who have sex with men (MSM), female sex workers (FSW), and people who inject drugs (PWID).

The prospect of a practically workable sampling method with plausibly principled inferential methods has made RDS especially popular for sampling populations that are hard to reach and carry a high burden of disease, especially MSM, FSW, and PWID. Since its inception, it has therefore been used in hundreds of studies of these populations in dozens of countries (White et al. 2015). These studies often have important public health implications and are instrumental in directing aid moneys aimed at addressing the global burden of diseases such as HIV/AIDS (Magnani et al. 2005, Malekinejad et al. 2008), thus creating a critical public health need for improved understanding of and performance of inferential procedures for RDS data.

In the remainder of this section, we introduce the standard RDS sampling procedure, some useful notation, and common theoretical approximations to the sampling process. We follow this with Section 2 on methods for the most common inferential goal of RDS: point estimation of population prevalence. We turn to additional inferential goals in Section 3. Section 4 discusses available software, and we conclude with a brief discussion in Section 5.

1.1. Sampling Process

RDS sampling is typically conducted using methods promoted by Salganik & Heckathorn (2004), a slight modification of the original procedure of Heckathorn (1997), and codified in protocols such as those of Johnston (2007) and White et al. (2015). An initial sample of typically 2–10 seeds is selected from among population members reachable to researchers. Seeds are chosen to be as heterogeneous as possible, hopefully representing each of the major social divisions researchers expect to see in the target population. They are also chosen to be opinion leaders in the community, such that their buy-in and support for the survey process is likely to encourage the broad

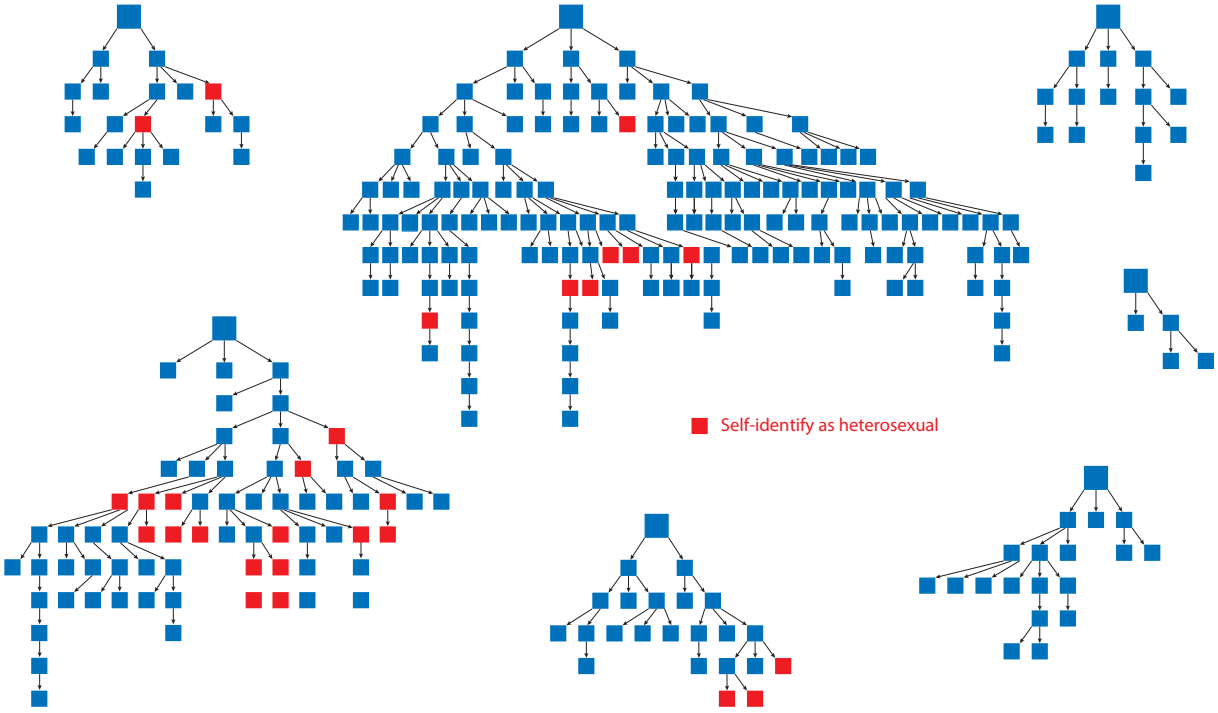

Figure 1

Recruitment trees plot among of men who have sex with men in Higuey, Dominican Republic. Adapted with permission from Gile et al. (2015).

participation of other population members. This buy-in is particularly important in populations where behaviors characterizing the target population are illegal or stigmatized, especially if the population has any experience with research procedures that put the population at risk.

The seeds are interviewed, then given typically three uniquely identified coupons to distribute to other population members. This number is chosen to strike a balance between the inferential desire to limit the branching of the sample and the practical necessity of guarding against the early termination of the sample trees.

Population members receiving coupons visit a study center where they are either interviewed directly or given an appointment for an interview. Subsequent respondents are also given three coupons, and in such a way the sample grows until it nears the desired sample size and coupons are tapered or stopped. **Figure 1** illustrates a recruitment structure among MSM in Higuey, Dominican Republic. As in most survey research, participants are compensated for their time completing the survey. In addition, participants are given recruitment incentives for each successful recruitment.

The survey must measure the number of target population contacts of each respondent. This is typically done by asking a series of questions that narrow the recruit's contacts to the precise definition of the target population, and typically to a recency of contact. Studies must also verify membership in the target population. Methods vary, but may include being able to explicitly describe procedures for preparing and injecting drugs or showing physical injection marks (for PWID). Researchers must also assure participants do not participate in the study more than once. In many settings, study staff are familiar enough with the target population to notice repeat

participation attempts. In other cases, repetition is prevented by collecting nonidentifiable but unique information about participants. These and other considerations of implementation are discussed by Johnston (2013a).

1.2. Notation

For most RDS studies, the object of inference is a population proportion of a nodal covariate, typically a binary nodal covariate such as HIV status. For population size N, we denote the values of this covariate as $\mathbf{z} = \{z_i\}, i \in 1 \ldots N$. Our object of inference is then $\mu = \frac{1}{N}\sum z_i$. To estimate this, we observe a sample of size n. Without loss of generality, we assume we observe units indexed $\{1, \ldots, n\}$.

We conduct the sample over a (typically) binary (typically) undirected network represented by $N \times N$ symmetric binary matrix $\mathbf{Y}$, where $Y_{ij} = 1$ indicates the presence of a relation between i and j. Also of central importance to the sampling process is the degree of each node, denote d_i, where $d_i = \sum_{j=1}^{N} Y_{ij}$ is the total number of contacts of node i.

We also consider some features of the network structure. We define differential activity as the ratio of mean degrees of the (typically) two groups under consideration. We define $DA = \frac{d_{(1)}}{d_{(0)}}$, where $d_{(k)}$ is the mean degree of nodes of class k. Also of interest is homophily, or the tendency for nodes to preferentially form network contacts with others similar to them. We are typically concerned with homophily on z. Several definitions of homophily have been used in RDS. The concepts discussed in this review are not dependent on a particular definition.

1.3. Approximations to the RDS Process

RDS is usually applied in survey settings where conventional methods are infeasible. Unlike in more ideal sampling contexts, in RDS, sampling typically depends on initial convenience samples, social network contacts, and participant decisions. As such, the RDS design is very complicated. In practice, it is not feasible to model all relevant features of the RDS sampling process directly. Therefore, all inferential methods rely on approximations. In this section, we review the major approaches to approximate to the sampling process. In later sections, we illustrate how these relate to inferential methods.

1.3.1. Random walk. The random walk (RW) approximation, and the corresponding first-order Markov process on the state space of the network nodes, has been used extensively to represent the RDS sampling process (Salganik & Heckathorn 2004, Volz & Heckathorn 2008, Goel & Salganik 2009, Lu 2013).

The RW process begins with the selection of a single initial sampled node. Peer recruitment is represented by allowing transitions only between adjacent nodes in the social network, assumed to be undirected and binary. Therefore, for W_k denoting the label of the node selected at step k, the probability of transitioning to the next state is given by $P(W_{k+1} = w_{k+1}|W_1, W_2, \ldots, W_k, Y = y) = P(W_{k+1} = w_{k+1}|W_k = w_k, Y = y) = y_{w_{k+1},w_k}/d_{w_k}$.

A key advantage of this approximation is the derivation of the sampling weights. Salganik & Heckathorn (2004) show that for a fully connected, undirected network, the described RW properties guarantee the existence of a unique stationary distribution with probabilities proportional to nodal degrees, $\pi_i \propto d_i$. Since then, various RDS estimators have been developed under the assumption that RDS sample weights can be approximated by the sampling weights of this RW stationary distribution. Note that an initial sample chosen with $\pi_i \propto d_i$ begins the walk at the

stationary distribution, otherwise the sample must converge before samples may be considered from this distribution.

The simplicity of the sampling weights makes this approximation particularly appealing for practitioners analyzing RDS data. However, this representation fails to capture various RDS features. The seed selection, for example, assumes only one seed, selected at random. This is typically unreasonable in practice. In addition, although RDS is a without-replacement process, the RW allows for nodes to be repeatedly sampled. The branching nature of RDS is also ignored in this representation. Finally, the RW presumes that participants recruit completely at random despite the growing evidence that RDS participants may engage in differential recruitment, as discussed in Section 2.1.7.

Malmros et al. (2016b) propose to approximate RDS with a variation of the standard RW that allows the process not only to transition to an adjacent node but also to visit a randomly selected node in the network. This process is referred to as the random walk with teleportation (RWWT). The main purpose is to replicate the multiple seeding feature of RDS as the jumps in the RW are interpreted as the starting points of the new recruitment trees. The jumps are assumed to occur at a constant probability that, they argue, is equivalent to a random seed selection regime.

Note that in addition to most RW theory treating a nonbranching RW, Rohe (2015) and Li & Rohe (2016) treat a branching with-replacement network sampling processes. Rohe (2015) uses conditions on the branching (based on referral rates) and network structure (based on the second eigenvalue of the Markov transition matrix), as well as the sample fraction to characterize the design effect, or amount of information in the sample. Li & Rohe (2016) provide a central limit theorem for RDS based on a branching with-replacement RW and give fairly mild conditions under which a standard inverse-probability weighted estimator (Volz & Heckathorn 2008, see Section 2.1.1) has an asymptotic normal distribution.

1.3.2. First order Markov process on categories. In this approximation, introduced by Heckathorn (1997), the sampling process is treated as a first order Markov process on the state space of the categories of nodes, according to $\mathbf{z}$. Under this approximation, if we consider a single-nonbranching process indexed by the sample order i, $P(Z_{i+1} = k|Z_1, z_2, \ldots Z_i) = P(Z_{i+1} = k|Z_i)$. In a study of HIV status in a population of PWID, for example, this would imply that every HIV-positive PWID has the same probability of linking to every other HIV-positive PWID.

Heckathorn's (1997) first suggested use of this approximation—to assure, under prohibitively strict sampling conditions, that the sample composition would converge to the population composition—was quickly replaced by the weighted alternative based on a Markov process on nodes, as in Salganik & Heckathorn (2004). The assumption still remains, however, in several areas of RDS inference.

First, although later RDS point estimators such as those introduced by Salganik & Heckathorn (2004) and Volz & Heckathorn (2008) involve weighting by nodal degrees, based on assuming a first order Markov process on nodes (as in the previous section), it is certainly unreasonable to assume a sample covering only a small portion of the target population could reach a stationary distribution on the state space of nodes. In arguing that the samples will approach a stationary distribution, it is therefore necessary to consider a stationary distribution on classes larger than individual nodes. Indeed, a popular diagnostic procedure (Heckathorn et al. 2002, Johnston et al. 2008, Montealegre et al. 2013) involves using the cross-group referral matrix to diagnose the number of waves needed to attain approximate stationarity. This approach conflates the two types of Markov processes. In truth, neither a first order Markov process on categories nor stationarity on nodes is realistic. In practice, some sort of approximate stationarity may be attainable at a scale between these two, as suggested by convergence plots as in Gile et al. (2015).

Separately, the first order Markov approximation remains central to the most widely used method of variance estimation, the bootstrap introduced by Salganik (2006), discussed in Section 2.2.

The assumption of a first order Markov structure on categories has been studied in RDS samples, especially by Neely (2010), Yamanis et al. (2013), and Verdery et al. (2015b), and found to be an inadequate representation of the RDS process. In practice, for example, HIV-positive PWID recruited by HIV-negative PWID are more likely to recruit HIV-negative PWID than are HIV-positive PWID recruited by HIV-positive PWID.

1.3.3. Successive sampling. The successive sampling approximation to the RDS sampling procedure, introduced by Gile (2011), has the impact of adjusting the RW approximation for finite population effects, which are important for large sample fractions. Gile (2011) argues that if the sampling process is a self-avoiding RW over the edges of a network generated from a configuration model (Molloy & Reed 1995), and sampling probabilities are marginalized over all networks generated from this configuration model, then the sampling process is equivalent to successive sampling, or probability proportional to size without replacement sampling. Under this process, all units have sizes, and units are selected sequentially with probability proportional to size from among the remaining unsampled units in the population. Gile (2011) argues that in RDS, the nodal degrees serve as the relevant unit sizes.

The model for the sample takes the following form:

$$P(\mathbf{W} = \mathbf{w}) = \prod_{i=1}^{n} \frac{\mathbf{d}_{w_i}}{\sum_{j=1}^{N} \mathbf{d}_j - \sum_{j=1}^{i-1} \mathbf{d}_{w_j}},$$

where W_i is the random index of the ith observed node, and w_i is its value. This results in stepwise sampling probabilities proportional to degree from among remaining unsampled nodes, and marginal sampling probabilities that are not available in closed form, but can be estimated to arbitrary precision.

1.3.4. Two-sided rational choice. Most of the approximations to the RDS sampling process assume that recruiters recruit uniformly at random from their network contacts, although it is likely that people recruit in ways that may depend on nodal or dyadic covariates. To model this preferential recruitment, McLaughlin (2016) developed a sequential two-sided rational-choice preferential recruitment model. In this model, each recruiter has utility for selecting each contact (or no one), and each contact has a utility for being recruited by each recruiter (or not participating). Participants behave to maximize their utility.

1.3.5. Epidemic model. At least two models have been used to approximate RDS sampling as an epidemic process. Malmros et al. (2016a) treat RDS as a Reed-Frost, wave-wise epidemic process, in the interest of evaluating the efficacy of the recruitment process in terms of the epidemic threshold R_0 for sustained spreading of the recruitment. Berchenko et al. (2017) treat RDS as a stochastic epidemic process, where the times of recruitments are dependent on the number of susceptible unsampled nodes, the number of active recruiters, and recruitment rates for each nodal class. Crawford et al. (2017) use a related approximation, in which sampling rates are assumed to depend on the number of ties to already-sampled population members.

1.3.6. Modeling the network. In addition to the explicit and general approximations to the RDS process introduced above, some authors have directly modeled the underlying network,

then explicitly or implicitly modeled the overlaid RDS process. We discuss these in more detail in Section 2.1.3.2.

2. ESTIMATION OF POPULATION PROPORTIONS

2.1. Point Estimation

The majority of inferential methods for RDS data aim to estimate population proportions of a binary or categorical nodal covariate. This may be partially due to specific reporting requirements such as the use of estimated proportions infected in the Estimation and Projection Program (EPP) (UNAIDS 2009), used by UNAIDS (the Joint United Nations Programme on HIV/AIDS) to estimate HIV rates in all countries with concentrated epidemics. It is also a logical first inferential question to address and may be less challenging than multivariate inference. Other common applications of prevalence estimators include estimating prevalence of risk behaviors, prevalence of service utilization, and the proportions of a target population from various demographic groups.

We assume the object of inference is the population proportion μ of binary covariate $\mathbf{z}$, as in Section 1.2. The initial paper on RDS (Heckathorn 1997) made strong enough assumptions that the sample proportion was assumed to be an unbiased estimate of the population proportion. These assumptions were quite heroic and were never widely believed. Nonetheless, the naive sample mean,

$$\hat{\mu}_{\text{mean}} = \frac{1}{n}\sum_{i=1}^{n} z_i, \qquad 1.$$

is often used as a benchmark for the performance of other estimators.

We first discuss two classical RDS approaches to inference for the population proportion, then discuss the key innovations according to the assumptions they relax. The choice of which estimator to use, when not guided by software availability (Section 4), should be based on which potential biases are deemed most critical in a particular study, perhaps guided by diagnostic procedures (Section 3.1). In addition, many newer papers compare the performance of proposed estimators to earlier ones. Systematic reviews of multiple estimators also exist, including those of Tomas & Gile (2011) and Verdery et al. (2015a).

2.1.1. Random walk approximations. The first widely used estimator for RDS data was introduced by Salganik & Heckathorn (2004), with a related popular refinement introduced by Volz & Heckathorn (2008). We refer to the estimators introduced in these papers as $\hat{\mu}_{\text{SH}}$ and $\hat{\mu}_{\text{VH}}$, respectively. Both of these estimators rely on the RW approximation to the RDS sampling process described in Section 1.3.1.

Both $\hat{\mu}_{\text{mean}}$ and $\hat{\mu}_{\text{VH}}$ can be considered estimators of the Hájek style:

$$\hat{\mu} = \frac{1}{\hat{N}}\sum_{i=1}^{n}\frac{z_i}{\pi_i} = \frac{\sum_{i=1}^{n}\frac{z_i}{\pi_i}}{\sum_{i=1}^{n}\frac{1}{\pi_i}}. \qquad 2.$$

When π_i is the inclusion probability of unit i, and $\pi_i > 0$ for all i, the numerator is an unbiased Horvitz-Thompson (Horvitz & Thompson 1952) estimator of the population total $\sum_{i=1}^{N} z_i$, and the denominator is an unbiased Horvitz-Thompson estimator of the population size N. The ratio is therefore approximately unbiased for the population mean of z_i for large n.

Sample mean $\hat{\mu}_{\text{mean}}$ is of this style, with $\pi_i = n/N$ for all i. $\hat{\mu}_{\text{VH}}$ is given by

$$\hat{\mu}_{\text{VH}} = \frac{1}{\hat{N}}\sum_{i=1}^{n}\frac{z_i}{d_i} = \frac{\sum_{i=1}^{n}\frac{z_i}{d_i}}{\sum_{i=1}^{n}\frac{1}{d_i}}. \qquad 3.$$

Table 1 Assumptions of the Volz-Heckathorn estimator

	Network structure assumptions	Sampling assumptions
Random walk Model	Network size large ($N \gg n$)	***With-replacement sampling (2.1.2)*** Single nonbranching chain
Remove seed dependence (2.1.3)	Homophily weak enough Bottlenecks limited Connected network (2.1.4)	Enough sample waves
Respondent behavior	***All contacts reciprocated (2.1.5)***	***Accurate measurement (2.1.6)*** ***Random recruitment (2.1.7)***

Other estimators are organized by the primary assumption they relax, according to the section numbers given in parentheses. Diagnostics for assumptions in are considered in Section 3.1. Versions of this table appeared in Gile & Handcock (2010) and Gile et al. (2015).

This is a Hájek-style estimator, substituting d_i for π_i. If sampling probabilities are proportional to degrees and $\pi_i = \alpha d_i$, then the constant of proportionality in the numerator and denominator cancels, and this estimator, too, is approximately unbiased for population mean μ. Aronow & Crawford (2015) give further technical conditions for consistency of this estimator.

The assumption that $\pi_i = \alpha d_i$ is justified by the RW approximation in Section 1.3.1. Note one additional subtlety here: Under the RW approximation, the π_i^*s, which are proportional to the d_is, are the draw-wise sampling probabilities, or the probabilities of drawing a given unit on a given draw, as opposed to list-wise sampling probabilities, which would reflect the overall probability of being included in the final sample (see Gile & Handcock 2010 for an extensive discussion of this distinction). Note that the numerator and denominator of Equation 2 are still unbiased for the population total and population size if π is replaced by π^*, but now each of these estimators is of the Hansen-Hurwitz (Hansen & Hurwitz 1943) form. The ratio in Equation 2, and therefore also in Equation 3, is thus also approximately unbiased for large n. Note that the challenge of considering draw-wise sampling probabilities in a truly without-replacement sampling setting is addressed by the estimators in Section 2.1.2 and also discussed by Gile & Handcock (2010).

Estimator $\hat{\mu}_{\mathrm{VH}}$ depends on a large number of assumptions on the sampling process and underlying network structure, as summarized in **Table 1**.

Most of these assumptions have been addressed in separate work, and nearly all of them have been addressed with the development of new estimators relaxing that assumption (typically by substituting other assumptions). We briefly discuss each assumption and its implications in the sections related to relaxing the assumption. Gile & Handcock (2010) give a more theoretical discussion of the implications of many of these assumptions for $\hat{\mu}_{\mathrm{VH}}$, and Lu et al. (2012) give a series of data-driven examples.

The estimator in Salganik & Heckathorn (2004), $\hat{\mu}_{\mathrm{SH}}$, relies on very similar assumptions to $\hat{\mu}_{\mathrm{VH}}$. Its form is quite different, and it is therefore differentially sensitive to violations of assumptions. Suppose two classes of nodes correspond to $z_i = 0$ and $z_i = 1$. Then let $t_{01} = t_{10} = \sum_{i \neq j} Y_{ij} z_i (1 - z_j)$, the total number of network ties between zeroes and ones. Let $d_{(k)} = \frac{\sum_{i=1}^{N} \sum_{j \neq i} Y_{ij} \mathbb{I}(z_i = k)}{\sum_{i=1}^{N} \mathbb{I}(z_i = k)}$, the mean degree of nodes of class k, where $\mathbb{I}(\cdot)$ is the indicator function on $\cdot$. Let $N_1 = N \cdot \mu$ and $N_0 = N \cdot (1 - \mu)$ represent the population sizes of the subgroups. Then let

$$c_{01} = \frac{t_{01}}{N_0 \cdot d_{(0)}} \quad \text{and} \quad c_{10} = \frac{t_{10}}{N_1 \cdot d_{(1)}},$$

representing the proportions of ties from each group that cross groups. Then algebraic manipulation gives

$$\mu = \frac{N_1}{N} = \frac{d_{(0)} \cdot c_{01}}{d_{(1)} \cdot c_{10} + d_{(0)} \cdot c_{01}}. \quad 4.$$

The estimator $\hat{\mu}_{\mathrm{SH}}$ plugs sample approximations of $d_{(0)}, d_{(1)}, c_{01}$, and c_{10} into Equation 4 to get estimator

$$\hat{\mu}_{\mathrm{SH}} = \frac{\hat{d}_{(0)} \cdot \hat{c}_{01}}{\hat{d}_{(1)} \cdot \hat{c}_{10} + \hat{d}_{(0)} \cdot \hat{c}_{01}}. \quad 5.$$

Under the RW model, the stationary distribution on the edges traversed is uniform. Therefore, the c_{kl} terms are estimated as the proportion of referrals from nodes of class k that go to class l. The $d_{(k)}$ terms are estimated using a Horvitz-Thompson or Hansen-Hurwitz approach, weighting by degrees.

The approach in Salganik & Heckathorn (2004) is therefore a hybrid design-based (estimation of $d_{(0)}, d_{(1)}, c_{01}$, and c_{10}) and method of moments (computation of $\hat{\mu}_{\mathrm{SH}}$) approach. Both Volz & Heckathorn (2008) and Beaudry et al. (2017) suggest that under some circumstances, the estimators $\hat{\mu}_{\mathrm{SH}}$ and $\hat{\mu}_{\mathrm{VH}}$ approach one another. Beaudry et al. (2017) suggests that the two are nearly equal when the numbers of cross-group recruitments in both directions are close to the same.

Because it does not directly weight the sampled units, $\hat{\mu}_{\mathrm{SH}}$ is less sensitive to the sample composition and therefore less sensitive to seed bias (Tomas & Gile 2011). It relies more heavily on the random referral assumption, as the estimated rates of cross-group referrals figure centrally. It also has particularly high variance when the sample proportion of one group is very small owing to the resulting imprecision of the estimated referral matrix.

2.1.2. Without-replacement sampling. When the sample fraction (n/N) is large, the with-replacement assumption of the RW approximation may be inadequate. Gile (2011) shows that earlier estimators $\hat{\mu}_{\mathrm{SH}}$ and $\hat{\mu}_{\mathrm{VH}}$ exhibit considerable bias due to large sample fractions, especially when there is differential activity (DA) far from 1. Gile (2011) shows that in large sample fractions, the mapping from degree to sampling probability deviates from the proportional relation in the RW approximation and approaches uniform as $n \to N$. If the degree distributions are similar for the groups, this may distort uncertainty estimation but does not induce substantial bias in $\hat{\mu}_{\mathrm{VH}}$ or $\hat{\mu}_{\mathrm{SH}}$. If DA is far from 1, however, the inaccuracy of the assumed proportional mapping of $\hat{\mu}_{\mathrm{SH}}$ and $\hat{\mu}_{\mathrm{VH}}$ differentially impacts the groups, inducing bias. Barash et al. (2016) evaluate a number of conditions, including without-replacement sampling, especially with more common smaller sample fractions. They conclude that with sample fractions of up to 20% of the population, and in some conditions even up to 40%, without-replacement sampling contributes very little bias to estimates.

In the presence of finite population effects, Gile (2011) suggests using sampling probabilities estimated under the successive sampling approximation in Section 1.3.3 in a Hájek-style estimator, as in Equation 2. A critical limitation of this estimator is that the computation of sampling probabilities requires knowledge of the population size. Without this knowledge, the method can also be used to conduct sensitivity analysis. When the working population size is reasonable (and especially not unreasonably low), the resulting estimator $\hat{\mu}_{\mathrm{SS}}$ consistently outperforms $\hat{\mu}_{\mathrm{VH}}$ and $\hat{\mu}_{\mathrm{SH}}$ for large sample fractions with DA far from 1, and it is nearly identical to $\hat{\mu}_{\mathrm{VH}}$ for small sample fractions and DA close to 1.

One advantage of $\hat{\mu}_{\mathrm{SH}}$ over $\hat{\mu}_{\mathrm{VH}}$ is its relative robustness to seed bias (Tomas & Gile 2011). It is natural to consider whether $\hat{\mu}_{\mathrm{SH}}$ can also be adapted to be robust to finite population effects. Just as the distortion of nodal sampling probabilities will have some impact on $\hat{\mu}_{\mathrm{SH}}$, Ott & Gile

(2016) show that finite population effects can also cause edge traversal probabilities to deviate from the uniform distribution implied by the RW. Ott et al. (M.Q. Ott, K.J. Gile, J.W. Hogan & M. Harrison, manuscript in preparation) therefore propose a new estimator, referred to as the weighted Salganik-Heckathorn ($\hat{\mu}_{\mathrm{WSH}}$), a modification of $\hat{\mu}_{\mathrm{SH}}$ that uses the successive-sampling based nodal weights of $\hat{\mu}_{\mathrm{SS}}$, along with novel estimates of c_{kl} based on edge-weights approximated using successive sampling. This estimator combines the robustness to finite population effects of $\hat{\mu}_{\mathrm{SS}}$ with robustness to seed bias of $\hat{\mu}_{\mathrm{SH}}$. Like $\hat{\mu}_{\mathrm{SS}}$, it requires knowledge of the population size for effective adjustment for finite population effects.

Relatedly, Fellows (I.E. Fellows, manuscript in preparation) introduced the homophily configuration graph (HCG) model extending the standard configuration graph model to incorporate homophily. This model maintains the strengths of the configuration model (e.g., allowing the distributions of nodal degrees to be fully general) while allowing the within- and between-group tie rates to be fully general. Fellows develops a new estimator, $\hat{\mu}_{\mathrm{HCG}}$, based on assuming the network follows a HCG. Like $\hat{\mu}_{\mathrm{WSH}}$, $\hat{\mu}_{\mathrm{HCG}}$ outperforms $\hat{\mu}_{\mathrm{SH}}$ and $\hat{\mu}_{\mathrm{VH}}$ when the sample fraction is large and is superior to $\hat{\mu}_{\mathrm{SS}}$ when there are biased seeds, retaining the advantages of each of these estimators without dramatically elevating variance.

2.1.3. Seed dependence. If the mechanism for choosing the initial sample is unknown, then the full sampling design for the final sample will be, too. This unknown sampling mechanism is unlikely to be ignorable, in the sense of Rubin (1976). As such, the unsampled members of the population will be not missing at random (NMAR) data (Rubin 1976). There are three broad approaches to NMAR data that are exploited by methods for RDS inference:

1. Assume the NMAR impact is negligible and ignore it (assumed by many methods),
2. Use estimators robust to the specific class of NMAR (Section 2.1.3.1), or
3. Jointly model the data and the sampling process (Section 2.1.3.2).

RDS data is likely NMAR for many reasons, including seed bias and variants of participation bias (Section 2.1.7), as well as the fact that sampling strictly depends on features of an unknown underlying network (see Lunagomez & Airoldi 2014 for a discussion of this). In this section, we focus on dependence on seeds.

The premise of treating RDS as a process at stationarity is that the sample will move enough steps or waves from the initial sample that the impact of the initial convenience sample on the final sample and resulting estimates is adequately attenuated. In practice, it is not reasonable to assume that the numbers of waves in RDS samples (nearly always <20) are adequate to reach a stationarity distribution on nodes in a sample of a sizeable social network. It is more reasonable to assume that the process might reach stationarity on classes of nodes (as in Section 1.3.2), although the first order Markov assumption on this space is known to be inadequate. Researchers typically begin RDS samples at a diverse set of seeds, which also helps speed convergence. Of course, beyond reaching stationarity, valid inference would require that the full set of units used in inference were sampled from a stationary distribution (see Gile & Handcock 2010 for a discussion of this). In addition, the notion of stationarity is already fraught in a without-replacement sample.

The impact of the seed selection regime is also less for deeper sample trees (or larger numbers of waves), and less when the underlying network is well approximated by random mixing. High homophily, especially on $\mathbf{z}$ or on characteristics closely related to $\mathbf{z}$, exacerbates the impact of seed selection (Gile & Handcock 2010).

An especially challenging and extreme form of homophily has been referred to as bottlenecks (Goel & Salganik 2009, Yamanis et al. 2013). Yamanis et al. (2013) and related papers give a poignant example of this: Sex workers in China are strongly segregated by a tier system governed

by the types of settings where clients are found and the background of the sex workers, as well as the riskiness of practices and HIV rates. Because RDS coupons are rarely passed between tiers, a sample tree starting in one tier is likely to stay in that tier, leading to seed dependence that is nearly impossible to remove, and suggesting that studies should be conducted separately within each tier.

In contrast, in the ideal case, the impact of the initial sample may be minimized by features of the network structure or sampling process, and most estimators will be robust to the composition of seeds. This is not generally true, as RDS is typically conducted in settings where the population is not well known to researchers, and seeds are typically selected by some convenience sample. We discuss two classes of methods with particular robustness to seed selection: those that do not weight the sample directly and those that implicitly or explicitly model the sampling process.

2.1.3.1. Methods that do not weight the sample. When the impact of seeds is strong, estimators that directly weight the sample, such as those of the Hájek form (Equation 2) ($\hat{\mu}_{\text{mean}}$, $\hat{\mu}_{\text{VH}}$, $\hat{\mu}_{\text{SS}}$), can have large biases, whereas estimators related to $\hat{\mu}_{\text{SH}}$, which do not directly weight the sample, will not be as directly impacted (Tomas & Gile 2011). For this reason, estimators such as $\hat{\mu}_{\text{SH}}$ (Salganik & Heckathorn 2004), $\hat{\mu}_{\text{WSH}}$ Ott et al. (M.Q. Ott, K.J. Gile, J.W. Hogan & M. Harrison, manuscript in preparation), and $\hat{\mu}_{\text{HCG}}$ (I.E. Fellows, unpublished manuscript) may be preferred.

2.1.3.2. Methods that model the network and sampling process. In RDS, modeling both the data and the sampling requires modeling the nodal attributes of interest and the network structure, as well as an overlaid sampling mechanism. There have been at least three approaches to this type of joint modeling.

Rather than explicitly or parametrically modeling all of these processes, Gile & Handcock (2015) use a model-assisted approach to model the network explicitly, model the sampling process implicitly, and estimate the population proportions of nodal covariates using weights estimated based on the network and sampling models. They argue that this method is able to adjust for seed bias as well as other measurable aberrations in the network or sampling process. This approach involves iteratively estimating nodal sampling probabilities, then fitting a network model with weighted sample information. Weights are approximated by simulating full social networks, then simulating RDS samples from those models. This approach has the advantage of allowing for finite population adjustments as well as adjustment for seed composition. The application in Gile & Handcock (2015) also treats the issue of differential recruitment effectiveness, illustrating that this approach may, in principle, be extended to other measurable features of the network or sampling process.

The approach of Fellows (2012b) is more ambitious in its modeling. This method leverages the exponential-family random network model (ERNM) model class (Fellows & Handcock 2012), which jointly models the network and covariate structures of network data. Using the partially observed data, this method therefore involves fitting a joint model for Y and Z, while also modeling the sampling process.

Lunagomez & Airoldi (2014) also propose an ambitious joint modeling strategy, including marginalizing over the large class of networks consistent with the sample and jointly modeling the network, covariates, and sampling process. They use a Markov random field to represent the network dependence in nodal covariates and a simple Erdős-Rényi (Erdős & Rényi 1959) network model. This modeling approach corresponds to modeling nodal covariates Z conditional on network y, rather than Y conditional on z as in Gile & Handcock (2015) or joint modeling as in Fellows (2012b). Note that this framework does make it difficult to include differential activity.

Owing to the computational complexity of these three methods, as well as the lack of available software, there are no available side-by-side comparisons of these three methods. The

model-assisted method of Gile & Handcock (2015) is available in the RDS Analyst software package (see Section 4) (Handcock et al. 2014a).

2.1.4. Connected network. Malmros et al. (2016b) also developed an estimator of the Hájek style. The estimated sampling probabilities in their proposed methodology are derived from the stationarity distribution of a RWWT, as discussed in Section 1.3.1. The RWWT is presumed to take place on the nodes of an undirected network of infinite size generated from a configuration network model (Molloy & Reed 1995). Contrary to the RW approximation, which presumes that the sampling process begins at a single seed and is largely selected from the stationary distribution, this estimator allows for the random selection of multiple seeds, modeled as random jumps. The random jumps also allow the process to visit different components of the network, should it be disconnected. This therefore lifts the common restriction that the network must be fully connected. The proposed estimator could be particularly vulnerable to seed selection bias and bias induced by network homophily because the random selection of seeds is an unlikely scenario and the transition probabilities of the RWWT assume jumps as well as edges selected at random. It is unclear how sensitive the method is to practical deviations from these assumptions.

2.1.5. Unreciprocated contacts. Most RDS estimators assume that relationships are reciprocated, or that the edges in the underlying social network are undirected. If person i recruits j into the sample, it is assumed that j might have also chosen to recruit i if possible. Thus, an individual's out-degree (number of reported ties) is necessarily equal to their in-degree (number who would count them as a tie). If networks are directed, sampling probabilities are related to in-degree, although only out-degree is practically observable.

In a study of male heads of household in 25 rural villages in Uganda, McCreesh et al. (2012) collected a sample through RDS and also had access to population level information. In this study, each participant was asked to list their contacts who fit the study criteria, though only 30% of those recruited were named as contacts by their recruiter. In a paper explicating the various assumptions inherent to RDS inference, Rudolph et al. (2013) cite the network literature and reciprocation rates varying between 36% and 43%.

If the underlying network is directed, then out-degree will not be equal to the in-degree, and estimation may suffer. Lu et al. (2012) simulated RDS on the largest single connected component of an online social network. Simulations of the sampling stage of RDS on this network were conducted using all ties and also using only the reciprocated ties. Although the inference using the $\hat{\mu}_{\mathrm{VH}}$ estimator on the undirected network resulted in low bias, the estimator produced biased estimates on data collected through the directed network simulation.

New RDS estimators have been introduced that aim to address the biases that are incurred owing to a violation of the directed network assumption. Lu et al. (2013) use prior information on the in-degree ratio for specific groups to alter existing RDS estimators. An in-degree ratio of 1 would indicate that for individuals in that group, the average in-degree is equal to the average out degree, which would correspond to the undirected network assumption. This estimator requires that the in-degree ratio is either known, or can be reasonably estimated. Malmros et al. (2016b) develop an estimator that requires no prior information on in-degree ratio but leverages the properties of RWs on directed networks and assumes that the proportion of unreciprocated edges may be estimated. To do this, they assume an Erdős-Rényi model for the underlying network.

2.1.6. Measurement error. Inference from RDS data is not immune to the impact of measurement error, which can plague any study involving measurement. Beaudry et al. (2017) address the problem of misclassification (categorical measurement error) on the outcome of interest, z.

Although some outcomes of interest, such as modern individual HIV tests, may have high sensitivity and specificity, reducing the impact of measurement error, other quantities are less accurately measured. Beaudry et al. (2017) use the example of self-reported HIV status as a proxy for HIV status. In addition to illustrating the potentially considerable impact of misclassification error on resulting estimates, they present two methods for adjustment of common RDS estimators for misclassification [the matrix method and the SIMEX-MC (simulation-extrapolation-misclassification) method (Kuchenhoff et al. 2006)] and describe the circumstances under which each method might be more suitable. These methods require external knowledge of the misclassification rates.

In addition to measurement error on measures of substantive interest, the critical role played by observed degrees in weighting RDS samples raises concerns about the impact of measurement error on degrees. There have been several studies of the impacts of inaccurate degree reports on RDS inference (Lu et al. 2012, Rudolph et al. 2013, Mills et al. 2014). They show that RDS point estimators are robust to many well-behaved types of measurement error on degrees but may exhibit considerable bias in cases such as when the misreporting pattern is associated with z, or in the case of heaping, when participants systematically round degrees to nearby salient numbers such as multiples of 5 or 10.

McLaughlin et al. (2015) address the related question of the relationship between reported degrees and network visibility. Primarily in the context of population size estimation based on the methods in Section 3.2.2, they argue that RDS inferential methods that assume degrees are proportional to sampling weights rely on the assumption that all reported edges are equally likely to lead to a referral. This is true in a RW on a theoretical graph with equal edge-weights. This is unlikely to be true in practice. Instead, McLaughlin et al. (2015) argue that of greater interest is the respondent's connectedness or visibility in the target population, reflecting their number of connections as well as whatever can be gleaned of the salience of those connections. They propose a method for computing respondent visibility scores that combine self-reported degrees with elements of the sampling process such as number of successful recruits. A critical advantage of this approach for the methods in Section 3.2.2 is that extremely high degrees tend to shrink toward the rest of the degree data, which dramatically diminishes the potential impact of high degree outliers late in the sample for the successive sampling population size estimator (SSPSE) (Section 3.2.2) estimates.

Although it is not strictly a method to deal with measurement error, Berchenko et al. (2017) provide a poststratification estimator using estimated sampling weights for binned sets of degrees that are not assumed to be proportional to degrees. Like the adjustments of McLaughlin et al. (2015), these are empirically estimated; however, they are estimated based on the estimated population proportions of each degree group based on a stochastic epidemic model. It seems that this estimator is designed to address the situation in which sampling probabilities are not related to degrees in a straightforward manner. It is unclear under which sets of conditions on network sampling processes this estimator might be expected to perform well.

2.1.7. Nonrandom recruitment. Most methods for RDS data assume idealized participation-related behaviors. When offered a coupon, individuals are presumed to always accept it and complete the survey (absence of nonresponse). Recruiters of all classes are expected to distribute coupons at the same rate (absence of differential recruitment effectiveness) and to do so completely at random among their contacts (absence of preferential recruitment or differential recruitment).

Although mathematically convenient, these assumptions are often unrealistic in practice. Growing empirical evidence suggests RDS surveys may be subject to substantial nonresponse (Yamanis et al. 2013), differential recruitment effectiveness (McCreesh et al. 2011), and

preferential recruitment (Iguchi et al. 2009, Liu et al. 2012, Yamanis et al. 2013, Merli et al. 2015). Because departures from these idealized behaviors may significantly impact the quality of the inference (Gile & Handcock 2010, Tomas & Gile 2011, Lu et al. 2012, McCreesh et al. 2012, Lu 2013, Rudolph et al. 2013, Shi et al. 2016), recent methods relax some of these assumptions. Critically, most of these methods may only be applied to estimate the prevalence of characteristics that are visible to other members of the target population.

McCreesh et al. (2013) propose adjusting the sampling probabilities (αd_i) to incorporate unsuccessful recruitment attempts reported on follow-up questionnaires. Extending $\hat{\mu}_{\mathrm{VH}}$, the sampling probability for an individual presenting with characteristic X is adjusted by the ratio of the observed proportion of individuals in the sample with these characteristics to the proportion of people with these characteristics who were offered a coupon. This adjusts the sampling probability for groups that disproportionately rejected recruitment attempts while preserving the weighting by network connectivity. As pointed out by the authors, this method is dependent on collecting representative information in follow-up questionnaires.

Shi et al. (2016) also leverage the information collected from follow-up questionnaires to extend $\hat{\mu}_{\mathrm{SH}}$. The estimated matrix of cross-group ties used for $\hat{\mu}_{\mathrm{SH}}$, originally derived from observed recruitments, is modified to take into account failed recruitment attempts. This approach shows robustness to some forms of nonresponse but retains considerable bias and variance in others, especially in the presence of homophily.

One important advance is suggested by Lu (2013), who proposes collecting information on the composition of respondents' contacts in the target population as part of the RDS survey. The estimator developed by Lu (2013), which we call $\hat{\mu}_{\mathrm{Lu}}$, is an extension of $\hat{\mu}_{\mathrm{SH}}$ in which the cross-tie matrix is estimated from the local network data instead of the observed recruitment patterns. The estimator has the advantage of being significantly more robust to nonrandom recruitment as well as reducing considerably the variance of the original estimator. In a study of FSW in China, Verdery et al. (2015a) verify that Lu's improved estimator provides substantial improvement over other methods. Verdery et al. (2015a) argue that this improvement is particularly critical in this setting because of the substantial biases induced by the tier-based social structure of FSW in China, where social networks, recruitment hierarchies, and disease risk are all heavily structured by tier.

Using the two-sided rational choice model descried in Section 1.3.4, McLaughlin (2016) creates models for the latent utilities of recruiters and recruits that can be estimated from RDS data. She uses Bayesian inference to infer the model parameters and develops new prevalence estimates based on the rational-choice preferential recruitment sampling model.

Beaudry (2017) extends the RW approximation in $\hat{\mu}_{\mathrm{Lu}}$ and $\hat{\mu}_{\mathrm{VH}}$, deriving stationary distributions of RWs impacted by three forms of preferential recruitment. The stationary distributions then serve as sampling weights to explicitly correct for preferential recruitment. Similarly to $\hat{\mu}_{\mathrm{Lu}}$, the proposed estimator requires the collection of ego-network data. However, this estimator has the additional advantage of being able to correct for preferential recruitment occurring on any observable discrete variables, not only the outcome variable. These adjusted estimators perform best when there is no homophily and retain some bias in cases of strong homophily. To address cases of stronger homophily, Beaudry (2017) also proposes a method involving fitting a network model, then treating sampling as a (preferential) successive sampling process. The population parameters are estimated under a Bayesian framework.

2.1.8. Sampling innovations. Some inferential methods for RDS rely on novel or specialized adaptations of the RDS sampling procedure. We discuss these here, focusing on specialized sampling methods that are more extensive than simply adding questions, and focusing on methods

designed to allow for improved inferential methods rather than improving the practicality of RDS sampling.

2.1.8.1. Anticluster RDS. Khabbazian et al. (2016) propose novel variants of the RDS sampling method which they refer to as anticluster RDS. They propose distributing coupons as usual to RDS respondents, but asking respondents to distribute them according to one of several restrictions:

- Recruits should not know each other.
- Recruits should know many people unknown to the recruiter.
- Recruits should not know the respondent's recruiter.

They demonstrate that these innovations reduce the dependence in the RDS sample trees, leading to decreased variance of samples and therefore of estimators. Using stochastic blockmodels, they demonstrate theoretical properties of these sampling procedures. A related method, using deliberate sampling decisions that force the sample away from regions already well covered, but based on full elicitation of the local social network, was introduced by Mouw & Verdery (2012). Although these methods could clearly speed the mixing of an RDS sample, it is not clear how these new sampling procedures may change the stationary distribution, and therefore the relative sampling weights, in general populations.

2.1.8.2. Privatized network sampling. Fellows (2012b) proposed privatized network sampling that collects more information on the network while preserving the privacy of the networked population. Each RDS respondent is asked for not just their number of contacts, but identifiers for each. Privatized network sampling uses an idea from cryptography called a hash function to create subject identifiers that are unique and verifiable, but would be difficult to use to identify a member of the population (Buchmann 2004). Hash functions might include initials, birthdays, or digits of a phone number, for example. Privatized network sampling dramatically improves the inference possible from the method of Fellows (2012b) using the ERNM model class.

2.1.8.3. Web-based RDS. This review focuses on statistical procedures for RDS data. Such procedures are theoretically agnostic to the medium of the surveys and recruitment. In principle, the process of collecting an RDS sample may be conducted over the Internet, through distribution of electronic coupons. Wejnert & Heckathorn (2008) were the first to do this, conducting a validation study of RDS methods with students on a college campus. Although the procedures used were mathematically similar to those of in-person RDS samples, they encountered several specialized practical advantages and challenges. Since then, several web-based RDS studies have been conducted, including those of Bengtsson et al. (2012) (and others by the same group) and Arayasirikul et al. (2016). Rather than detail these practical circumstances that might impact survey and inferential design choices, we refer readers to these works.

2.2. Estimation of Uncertainty for Population Proportions

Because of the complex dependence in the sampling process, along with the unequal and unknown sampling weights and unknown respondent behavior, estimation of uncertainty for RDS data is a very challenging problem. Indeed, in conditions where approximations such as those above give nearly unbiased point estimators, the variance may be wildly underestimated. For this reason, the approximations to the RDS process used for estimation of uncertainty are often different from the approximations used for point estimation, and they are typically implemented as bootstrap procedures.

The most widely used uncertainty estimator for RDS was introduced by Salganik (2006). It consists of a bootstrap procedure based on assuming a first order Markov dependence on categories, implemented as follows:

1. Sampled nodes are partitioned according to the **z** status of their recruiters. Let R_k represent the set of nodes recruited by those with $z_i = k$.
2. For $m \in 1 : M$ bootstrap replicates,
   (a) One initial sample, or seed, is selected uniformly at random from the whole sample.
   (b) While sample size $< n$:
      - Let z_i represent the class of the most recent sample.
      - Select the next sample uniformly at random with replacement from R_{z_i}.
   (c) Compute $\hat{\mu}_m$ as the estimator based on the mth resample.
3. An approximate level-α confidence interval for μ is given by the middle $1 - \alpha$ quantiles of the distribution of $\hat{\mu}_m$s (i.e., the percentile bootstrap method).

As the dominant method of uncertainty estimation for RDS for a long time, this method has been subjected to much use, evaluation, and criticism. Major concerns highlighted in the literature, especially by Goel & Salganik (2010) and Verdery et al. (2015b), include the following:

- The single nonbranching chain ignores the elevated dependence induced by the branching multi-tree structure of the true sampling, leading to underestimated variance.
- The first order Markov assumption on categories underestimates the dependence resulting from true network structures, especially in the presence of bottlenecks (i.e., parts of the network that are only weakly connected to each other).

Despite these limitations, the difficulties inherent in estimating the variance induced by the true RDS procedure have contributed to the dearth of available alternatives. Most of the revised inferential procedures in Section 2.1 propose corresponding estimators of uncertainty that simply modify the procedure of Salganik (2006) to account for the modification of the proposed point estimator. Exceptions include the following:

- The methods to adjust for without-replacement sampling use without-replacement bootstraps. To do so, they simulate finite population structures with homophily-type dependence from which to sample (Gile 2011; Ott, et al 2017).
- Some methods similar to that of Salganik (2006) replicate the branching structure (Yamanis et al. 2013).
- Yamanis et al. (2013) also suggest using information in a follow-up questionnaire to compare variants using transitions based on the composition of contacts versus the composition of invited contacts. This is related to the method of Lu (2013).
- Some methods fit a full Bayesian model and use posterior credible intervals rather than bootstraps (Lunagomez & Airoldi 2014).
- The methods that fit a full network model simulate bootstrap samples based on full networks (Gile & Handcock 2015, Fellows 2012b).
- Beaudry et al. (2017) include modifications that resample misclassification error.

Baraff et al. (2016) recently introduced a novel alternative RDS bootstrap. This method, called treeboot, involves bootstrap resampling from among the actual recruits of the active recruiter. A node with three recruits in the true sample appearing in the bootstrap resample, for example, would have three bootstrapped recruits, sampled with replacement from their three true recruits. The resulting branching resampling process runs until there are no more active recruiters. This

process replicates the dependence in the RDS process more faithfully than Salganik (2006). This is a promising contribution. As this is a new procedure, its performance is still being evaluated. Some evaluations suggest it may substantially overestimate uncertainty.

Spiller et al. (2017) provide a systematic review of the uncertainty estimators of Salganik (2006) and Gile (2011), paired with $\hat{\mu}_{\text{mean}}$, $\hat{\mu}_{\text{SH}}$, $\hat{\mu}_{\text{VH}}$, and $\hat{\mu}_{\text{SS}}$, on networks and RDS samples simulated to match 40 samples of PWID in 20 cities over two cycles from the National HIV Behavioral Surveillance (NHBS) studies (Lansky et al. 2007). Although their findings do support those of Goel & Salganik (2010) and Verdery et al. (2015b) in showing that the true variance is often underestimated, especially for highly structured network populations, they find a degree of underestimation far less than the nearly catastrophic levels suggested by Goel & Salganik (2010) and Verdery et al. (2015b). They argue that part of this difference is because their simulations are conducted without replacement. In highly structured populations, it is easier for a simulated sample to get stuck in a region or subgroup using with-replacement sampling, inducing higher variance than without-replacement sampling. They also argue that studentized bootstrap intervals tend to outperform the percentile intervals suggested by Salganik (2006).

3. OTHER INFERENTIAL GOALS

3.1. Diagnostics for Assumptions about the Sampling Process

Given the large number of assumptions necessary to treat an RDS sample as a probability sample, it is important to consider diagnostics for violations of these assumptions whenever possible. Many authors have contributed diagnostic methods.

Gile et al. (2015) present a battery of diagnostics for many of the assumptions in **Table 1** and apply them to example data from 12 studies in the Dominican Republic. They also include a number of helpful citations and background on each of the assumptions. Here, we summarize major approaches to evaluating each of the assumptions but refer readers to the original papers for details. Our discussion follows the structure of **Table 1**.

3.1.1. Without-replacement sampling. In truth, RDS samples are without-replacement. The question is whether this known approximation impacts inference. Gile et al. (2015) suggest using the differences between $\hat{\mu}_{\text{VH}}$ and $\hat{\mu}_{\text{SS}}$ for quantities of interest as a diagnostic. They also suggest that features of the sampling process, including failure to reach the intended sample size, failed recruitment attempts, and high proportions of contacts who already participated in the sample, may indicate large sample fractions. In their supplemental materials, they do not find evidence of systematically decreasing degrees over time in the sample, as would be suggested by the approximation in Section 1.3.3.

3.1.2. Seed dependence. RDS estimates that depend heavily on the initial sample, or seeds, are unlikely to give valid estimates of population proportions. Although earlier methods suggested comparing the estimated transition matrix on classes to the number of waves of the sample (Heckathorn et al. 2002, Johnston 2013b, Malekinejad et al. 2008, Montealegre et al. 2013), this is problematic for reasons discussed in Section 1.3.2, as well as by Gile et al. (2015).

Gile et al. (2015) present three types of diagnostic plots to evaluate the removal of seed dependence, as illustrated in **Figure 2**:

- Convergence plots (see **Figure 2**) plot the value of the estimator as a function of the number of samples observed, in order. Large changes or trends in this plot toward the end of the

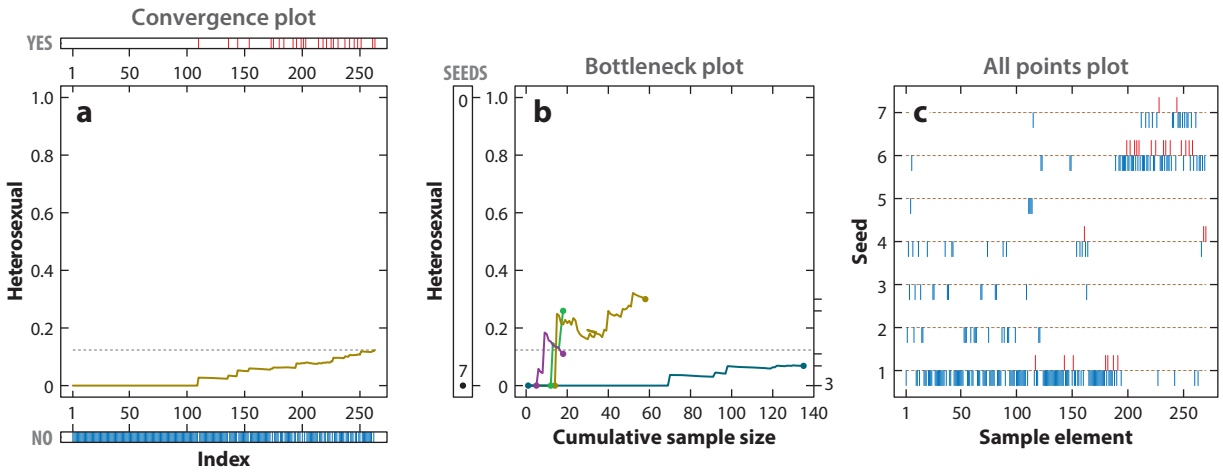

Figure 2

Three diagnostic plots for estimates of men who have sex with men in Higuey, Dominican Republic who self-identify as heterosexual. (*a*) The convergence plot, plotting the estimated proportion as the sample size increases, shows that the estimate has not stabilized, suggesting data collected late in the sample differ from earlier data. (*b*) The bottleneck plot, plotting separate estimates for each sample tree by cumulative sample size, shows that the chains explore different subgroups. (*c*) The all points plot, plotting the outcome for each sample by sample order and seed (or tree), shows that the self-identified heterosexuals (*red ticks*) were both late in the sample and from a small number of chains. Adapted with permission from Gile et al. (2015).

sample suggest the sample may not yet be well mixed within the target population. This is similar to an approach suggested by Bengtsson et al. (2012).

- Bottleneck plots plot the value of the estimator as a function of the number of samples observed in order with a separate trace for samples originating with each seed. Traces that do not seem to converge to a common range suggest strong divisions within the population that may impede the mixing of the sample.
- All points plots show the value of z for each sample as a function of its seed-based subtree and timing in the sample. These plots can be used to help explain aberrations in the previous two plots.

3.1.3. Reciprocation. Reciprocation of the relationships over which coupons are passed allows the self-reported degree (necessarily out-degree, or number of ties perceived by the respondent) to be used in place of the in-degree, or number of ties to the respondent perceived by others, which is more closely tied to recruitment rates. Additionally, if the network is directed (i.e., not all ties are reciprocated), the stationary distribution of the RW (see Section 1.3.1) is not proportional to degree. One level of assessment of reciprocation is given by asking respondents to report their relationships with recruiters (Heckathorn 2002, Lansky et al. 2012). Gile et al. (2015) also suggest asking respondents additional questions to directly assess the likelihood of reciprocal coupon-passing.

3.1.4. Measurement of degree. As discussed in Section 2.1.6, valid estimates depend on accurately reported degrees. Many methods have been proposed to assess degree measurement, primarily in terms of reliability, validity, and the sensitivity of estimates. Gile et al. (2015) measures reliability by comparing self-reported degrees on the standard survey to questions repeated when respondents returned to collect secondary incentives. They evaluate the validity of the time

interval of the degree question by comparing it to the time differences between recruiters and recruits in the sample, and they evaluate sensitivity by comparing estimates based on different versions of the degree question on initial and follow-up surveys. These issues are also evaluated empirically by several authors, including Wejnert & Heckathorn (2008), Wejnert (2009), McCreesh et al. (2012), and Yamanis et al. (2013).

3.1.5. Random recruitment. Estimators of the Hájek form (Equation 2) directly weight the resulting sample, meaning that unless they directly adjust weights for sampling aberrations, they are susceptible to biases due to deviations from the participation assumptions of the standard models. Three critical features of participation bias are differential recruitment effectiveness, preferential recruitment, and nonresponse. Note that many methods to measure recruitment bias are based on adding questions either to the original survey or to a follow-up administered when respondents return to collect secondary incentives.

Gile et al. (2015) measure three types of nonresponse, corresponding to the rates at which coupons were refused or not returned, as well as the overall nonresponse rate.

Differential recruitment effectiveness can be evaluated by comparing the mean numbers of recruits for members of different groups (Gile et al. 2015), excluding respondents at the end of the study whose recruitment opportunities were limited.

Preferential recruitment (sometimes called differential recruitment) can be assessed by asking about the composition of respondents' local social networks, then comparing them to the composition of recruits. Several authors consider this question and related approaches, some including proposed statistical tests (e.g., Heckathorn et al. 2002, Wang et al. 2005, Wejnert & Heckathorn 2008, Iguchi et al. 2009, Rudolph et al. 2011, Liu et al. 2012, Yamanis et al. 2013, Gile et al. 2015).

3.1.6. Feasibility analysis. Griffin et al. (M. Griffin, E. Erosheva, M. Handcock & K. Fredriksen-Goldsen, manuscript in preparation) present a method for assessing the feasibility of conducting an RDS study based on egocentric pilot study data, including self-reported local network composition and willingness to recruit others into a theoretical RDS study. They propose using these data to model the dependence in the underlying social network structure, then simulating RDS samples according to estimated sampling parameters. They suggest evaluating the resulting simulations in terms of whether they provide better estimates than any competing methods, whether they provide enough information to be worthwhile, and whether the uncertainty estimation is reasonably calibrated. Of course, such methods can only represent the measurable properties of the network and sample, so they provide an upper limit on the performance of an actual sample.

In related work from a different perspective, Malmros et al. (2016a) consider RDS recruitment as an epidemic process on a network. They evaluate the epidemic threshold, R_0, required to collect an effective sample size and conclude that the number of coupons used may often be too low.

3.2. Population Size Estimation

RDS is often conducted in settings where fundamental features of the population, including its size, are unknown. It is often of great interest to estimate the size of the population. In fact, the UNAIDS EPP (UNAIDS 2009), used to estimate national HIV cases for all countries with concentrated epidemics, requires both prevalence and size estimates for each key population. Because RDS is often used for these prevalence estimates, there is much interest in leveraging these data for population size estimates as well. There are several ways to do this, and we discuss the major ways here. We restrict our discussion to methods leveraging RDS data. Comparisons

of related methods, including but not limited to RDS, include those of Kendall et al. (2008), Paz-Bailey et al. (2013), and Tran et al. (2015).

3.2.1. Capture-recapture and multiplier methods. By far the most common use of RDS data for population size estimation uses a variant of the principle of capture-recapture. In traditional capture-recapture sampling, classically in wildlife populations, two or more independent samples are drawn from the target population. The overlap in the samples is used to estimate the size of the target population: Large overlap suggests a large portion of the population was sampled. Mathematically, with two independent captures, the estimated population size is given by

$$\hat{N} = \frac{n_1 n_2}{n_{12}},$$

where n_k is the size of the kth capture and n_{12} is the overlap. Related approaches can incorporate information from more than two samples or use a Bayesian framework to address nonhomogenous capture probabilities (Fienberg et al. 1999). Critically, the captures must be independent. If, for example, slow-swimming fish are more likely to be captured by all captures, the overlap between samples will be artificially inflated, and the population size estimate will have negative bias.

3.2.1.1. Service multipliers. The most straightforward use of multiple captures for RDS involves using service multipliers (Johnston et al. 2013). In this approach, the RDS sample serves as one capture. RDS respondents are asked whether they have used a variety of services available to the local target population. The set of unique population members served by each provider serves as an additional capture event, and this information can be combined to estimate the size of the target population. For example, an RDS proportion estimate that estimates 25% of PWID use a particular needle exchange program can be combined with knowledge that the exchange program has served 1,000 people to estimate a population size of 4,000 PWID. This type of estimate relies heavily on the assumption that service usage is independent of survey participation, as well as on the accuracy of both self-report and the lists maintained by the service provider, and on the correspondence of time intervals reported by respondents and service providers.

3.2.1.2. Unique object multiplier. Another approach is to use RDS data alongside another study-created capture or marking event using the distribution of a unique object. Unique objects are often key chains or other small personal items that recipients can keep with them at all times. The objects are distributed, typically by people close to the key population. RDS participants are then asked whether they have received the unique object. Some studies require participants to show the object and others require them describe it.

Because the number of objects distributed is known, this serves as a known-size reference population and can be used in a manner similar to the service multiplier. Again, a critical assumption is that object distribution is independent of the RDS. If the persons distributing the objects are also involved in selecting the seeds for the RDS study, for example, artifactual overlap may result. It is also essential to ensure that the unique objects are given to eligible population members and that the objects are not passed from person to person. A good example of this approach is in Paz-Bailey et al. (2011).

3.2.2. Successive sampling. In contrast to multiple capture methods, the method of Handcock et al. (2014b, 2015b) uses information in only the RDS sample. Relying heavily on the successive sampling approximation to the sampling process (Section 1.3.3) (Gile 2011), this estimator is based on the following logic: If units are sampled sequentially with probability proportional to

degree (d_i), then higher-degree nodes should tend to be sampled earlier in the sample. The trend of decreasing degree of observed nodes as the sample progresses contains information on the population size.

This method leverages the nonignorable structure of the sampling process by jointly modeling the sample and the population distribution of degrees. It is implemented in a Bayesian framework that allows for the natural incorporation of informative prior information, which is likely to be available in many RDS settings.

The greatest advantage of this approach is its low demand for novel data collection. RDS samples are often available in settings where population size estimates are needed. Unfortunately, there is often little information in the RDS sample about population size, leading to very large credible intervals. In addition, this method relies heavily on several assumptions. Strong clustering (violation of the random mixing assumption of the configuration model) can degrade performance, especially when combined with differential activity or different overall rates of ties by cluster. The method can also be sensitive to measurements of the nodal degrees and the validity of their use to measure connectedness in the target population. A respondent reporting 300 ties, for example, may be reporting weaker ties than a respondent reporting 10 ties, and may not be 30 times as connected to the target population. This method can be sensitive to high outlying degrees, especially later in the sample. It may also be sensitive to priors for the population size and other parameters.

One way to address this sensitivity to outlying degrees and degree misreporting is introduced by McLaughlin et al. (2015), who introduces a smoothing method to estimate effective unit sizes for RDS data based on smoothing outliers and using other information in the sample such as recruitment patterns. This method shows more robustness than the earlier method.

3.2.3. Methods based on recruitment timing. There are also two approaches that leverage information in the timing of recruitments to help infer population size. These have the advantage of using more diverse types of information in the sample but also rely on additional assumptions about these data.

3.2.3.1. The Crawford et al. (2017) method. Crawford et al. (2017) propose an alternate approach to estimating hidden population size based on RDS data alone. This method uses a different set of assumptions than the SSPSE method. First, it assumes an Erd˝os-R´enyi network model, in which every tie is equally likely. It then assumes that each population member is recruited at a rate dependent on their number of ties to already-recruited population members. Because the full network structure is not known, the authors use an algorithm to approximately marginalize over possible values of the unobserved network ties. Because it models recruitment rates, it also relies heavily on the exact timing of the interviews. This different set of assumptions makes it sensitive to different deviations from assumptions than the estimator in 3.2.2. Although the evaluation does not make the degree of sensitivity to these assumptions clear, this method will certainly be sensitive to violations in the assumption on the network model (i.e., nonrandom mixing), outliers and misreported degrees, and artifactual constraints on recruitment times (i.e., waiting times to be interviewed, study site closures, reduced coupons at the end of the study). Crawford et al. (2017) also illustrate sensitivity to assumed priors for the Bayesian model.

3.2.3.2. The Berchenko et al. (2017) method. Berchenko et al. (2017) also present a method for estimating population size dependent on the timing of recruitments. Rather than assuming a network model, this method assumes all recruiters are equally likely to recruit all nonrecruited "susceptible" population members. A critical difference from other methods (except that of McLaughlin et al.

2015) is that there is no explicit assumption of the relationship between degree and recruitment rate. Degrees are binned into classes (their application uses 10 classes), and each class is assigned an estimable sampling rate parameter. These rates are estimated based on the timing of recruitment events for nodes of each degree class and are used to infer the sizes of each group as well as the overall population size. This method is clearly dependent on a very different set of assumptions and is potentially sensitive to many different features than other estimators. These may include sensitivity to choices of degree-binning and sensitivity to logistical disruptions or limitations on the timing of recruitments. It is unclear how sensitive the method may be or how its performance might relate to earlier methods.

3.3. Regression and Multivariate Methods

RDS suffers from two particular challenges for multivariate modeling: unknown sampling weights and unknown dependence structure. However, investigators have strong interest in multivariate relationships. Some applied researchers with RDS data have proceeded as if the data were collected through simple random sampling, using methods including chi-squared tests, ANOVA, and linear and logistic regression (e.g., Štulhofer et al. 2008, Erausquin et al. 2014, Ramirez-Valles et al. 2013).

Another common approach has been to use sampling weights to control for unequal sampling probabilities, but to ignore possible correlation in observations, recognizing that this likely underestimates uncertainty and overstates significance (Johnston et al. 2010, Zhong et al. 2011, Silva-Santisteban et al. 2012).

Spiller (2009) suggests using mixed effects models to account for dependence. He introduces a several-step procedure including using ANOVA (analysis of variance) testing for dependence to assess appropriate layers of dependence, then building a model including possible random effects on features such as recruitment tree, geography, or shared recruiter, in either a weighted or unweighted fashion. He also suggests using sandwich estimators for variance estimation. These are certainly valid first steps in treating this challenging issue. Unfortunately, this brief master's thesis does not offer any technical considerations or evaluations of how these methods might perform under various RDS circumstances. Nonetheless, lacking more extensive guidance, many researchers have cited this thesis and used variants of these ideas.

Several researchers have used variants of the mixed effect modeling approach. Rhodes & McCoy (2015), for example, fit several models, including two types of hierarchal modeling and several missing data treatments. Such sensitivity analyses are a reasonable approach given the lack of a single superior strategy; however, it is challenging, as in the case of Rhodes & McCoy (2015), when the methods do not produce consistent findings.

Others, some citing Spiller (2009), have used generalized estimating equations (GEE) (Liang & Zeger 1986). This approach allows separate specification of a mean structure and a variance structure. Under many conditions, the model for the mean structure is consistent even if the covariance is misspecified. GEE is often used in settings where samples are expected to have correlated errors with known patterns. Perhaps owing to the limitations of currently available software, the most common use of GEE for RDS does not weight samples and uses a covariance structure based on clustering by RDS sample tree (see, for example, Abramovitz et al. 2009 and Pitpitan et al. 2016). This is a reasonable first step, however this correlation structure is clearly not ideal for RDS, as the covariances between all pairs of samples in the same tree are treated as equal.

We mention one further method for assessing multivariate association. D. Kim, K.J. Gile, H. Guarino & P. Mateu-Gelabert (manuscript in prep) introduce a nonparametric method for

testing for bivariate associations in RDS data. This method retains the full observed dependence structure for one of the variables by fixing the observed tree, then uses an estimated dependence structure (i.e., first order Markov) to simulate randomized replicates of the other variable to generate a null distribution for a test statistic. This method outperforms standard tests when the dependence in at least one variable can be represented by a simple estimable dependence structure.

3.4. Network Properties

In addition to the works described in Section 2.1.3.2 that infer network properties along with, or in the service of, estimating population means, a few authors have directly addressed the goal of inferring network properties from RDS data. Although many investigators are interested in the underlying network structures, the critical research questions are less well defined in this area, and the works are therefore less cohesive in their aims. There are severe limitations in the data. For example, Crawford et al. (2015) point out that homophily and preferential recruitment are typically confounded in RDS data and therefore are not separately identifiable.

Assuming no preferential recruitment, Wejnert (2010) discusses methods to estimate homophily from RDS data, as well as how to estimate mean degree by group, and triad closure, from information within the "recruiter–recruit–contacts of recruit" triad, the last based on asking respondents how many of their contacts know their recruiter. Triad closure, or clustering, is also treated by Verdery et al. (2016), who address estimation of the network clustering coefficient for RDS samples by applying algorithms designed for RWs on computer networks to simulated RDS samples. This requires asking respondents about connections between their recruiter and other contacts. They find some biases induced by common features of RDS, but the error rates in their study are not as extreme as one might expect given the necessary approximations.

Crawford (2016) uses the novel approach of modeling the recruitment process as a survival process over time, with sampling rates depending on the number of active recruiters in one's local social network. This results in an identifiable simple network model for the induced subgraph on the sample members.

3.5. Inference from Multiple RDS Studies

There are two settings where multiple RDS studies are often used together to make inference:

1. Multiple simultaneous samples in the same region (usually country) to get region-wide estimates
2. Multiple observations of the same population over time in the interest of inferring change

Both types of multi-RDS inference often occur in large ongoing surveillance studies. A critical example of both is the EPP used by UNAIDS to estimate HIV rates at the country level. This model combines prevalence estimates from multiple RDS studies of a key population in different geographical areas, along with population size estimates, to estimate national HIV burden (UNAIDS 2009). This goal can also be aided by modeling trends in HIV across time, as described by Brown et al. (2014).

Within the United States, the CDC uses the large ongoing NHBS system to monitor the status and change of populations at risk for HIV in 25 large US cities at three-year intervals (Gallagher et al. 2007). The goal here is not to combine estimates across cities, but to monitor the status and progress of the disease and related interventions across time. Thus, change over time is a major object of inference.

Note that in both of these applications, point and uncertainty estimates from RDS data are used and applied to the inferential goals, but the methods of extrapolation and testing for change do not consider the origin of the data used. Given the critical importance of the results of these methods, there is clearly call here for methodological advances.

4. SOFTWARE

The widespread use of RDS methods has depended strongly on the availability of user-friendly software. The original software, implementing a variety of methods introduced by Heckathorn and collaborators (i.e., Salganik & Heckathorn 2004, Salganik 2006, Heckathorn 2007) as well as a number of exploratory and diagnostic methods, is a stand-alone, freely available proprietary package called RDS Analysis Tool, or RDSAT (Volz et al. 2012). They maintain both newer and legacy versions of the software, and they provide a manual.

A newer alternative is the open-source `RDS` `R` package (Handcock et al. 2015a, R Core Team 2016) and the associated user-friendly front-end graphical user interface called RDS Analyst or RDS-A (Handcock et al. 2014a). It is freely available for download with a single installer file. This package implements several point estimators including those in Salganik & Heckathorn (2004), Volz & Heckathorn (2008), Gile (2011), Gile & Handcock (2015), and Fellows (I.E. Fellows, unpublished manuscript), as well as multiple estimators of uncertainty, population size estimates using the method of Handcock et al. (2014b), and diagnostics including those from Gile et al. (2015). The graphical user interface, built on top of the Deducer package (Fellows 2012a), allows for the analysis of RDS data from reading in the data through report generation. It also interfaces easily with the full capabilities of `R` and its many packages (R Core Team 2016). Johnston (2013a) provides a user-friendly manual.

Many other methods described here are available in stand-alone software or `R` packages.

5. DISCUSSION

RDS is a powerful method for attaining a sample of a hard-to-reach human population that is well-connected by a social network of relations. It is often more effective, economical, and ethical than other possible alternatives. However, because the method relies on respondents to act as recruiters, begins with a convenience sample, and is heavily dependent on a partially observed underlying social network, inference from the resulting data is challenging. These sampling patterns can be considered NMAR.

There are many methods for estimating population proportions or averages of nodal characteristics. None is strictly superior to the others, as many substitute one set of assumptions for another or require additional data collection. A critical area for future development is estimators of uncertainty. Most existing methods are known to systematically underestimate uncertainty under a broad set of conditions.

Methods are also under development to make inference about the violations of estimator assumptions, hidden population size, regression coefficients, and other network properties. There has been considerably less development in these areas. Despite great interest, there are still few methods for estimation outside a single sample, including extrapolation from many RDS studies to a broader population beyond the scope of any of them, and inference on change over time based on repeated RDS studies.

As with so much new methodological development, the deeper understanding and broader implementation of methodology is mediated by the availability of user-friendly software. For RDS users, many of whom are operating in low-resource situations without the opportunity to implement the methods themselves, this is especially important.

This review is far from a complete review of methodological advances for RDS. It focuses on statistical advances and insights directly related to inference from RDS data. There is also an extensive literature on other aspects of RDS, including those related to study design and implementation.

FUTURE ISSUES

1. Methods for multivariate (regression) modeling and testing are underdeveloped and in great demand.
2. Methods extrapolating to populations beyond the sampling frame are widely used, although they are often developed outside the RDS context.
3. RDS is often used in surveillance settings, where change from year to year is of interest. Methods for inference on change over time are needed.
4. Methods for uncertainty estimation still do a poor job of adequately representing many important sources of uncertainty.
5. Methods jointly modeling network and sampling features to adjust for NMAR patterns are promising but computationally expensive, and they may be aided by additional approximations and improved algorithms.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This material is partially based upon work supported by the National Science Foundation (NSF, SES-1230081, MMS-0851555, SES-1357619, IIS-1546259), including support from the National Agricultural Statistics Service, National Institute of Child Health and Human Development (NICHD, R21HD063000, R21HD075714, and R24-HD041022) and the Office of Naval Research (N00014-08-1-1015). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, the National Science Foundation or the Office of Naval Research. The authors thank Lisa G. Johnston for helpful comments.

LITERATURE CITED

Abramovitz D, Volz EM, Strathdee SA, Patterson TL, Vera A, Frost SD. 2009. Using respondent driven sampling in a hidden population at risk of HIV infection: Who do HIV-positive recruiters recruit? *Sex. Transm. Dis.* 36:750–56

Arayasirikul S, Chen YH, Jin H, Wilson E. 2016. A web 2.0 and epidemiology mash-up: using respondent-driven sampling in combination with social network site recruitment to reach young transwomen. *AIDS Behav.* 20:1265–74

Aronow PM, Crawford FW. 2015. Nonparametric identification for respondent-driven sampling. *Stat. Probab. Lett.* 106:100–2

Baraff AJ, McCormick TH, Raftery AE. 2016. Estimating uncertainty in respondent-driven sampling using a tree bootstrap method. *PNAS* 113:14668–73

Barash VD, Cameron CJ, Spiller MW, Heckathorn DD. 2016. Respondent-driven sampling—testing assumptions: sampling with replacement. *J. Off. Stat.* 32:29–73

Beaudry IS. 2017. *Inference from network data in hard-to-reach populations*. PhD Thesis, Univ. Mass., Amherst

Beaudry IS, Gile KJ, Mehta SH. 2017. Inference for respondent-driven sampling with misclassification. *Ann. Appl. Stat. 11:4:2111-2141*Bengtsson L, Lu X, Nguyen QC, Camitz M, Hoang NL, et al. 2012.

Implementation of web-based respondent-
driven sampling among men who have sex with men in Vietnam. *PLOS ONE* 7:e49417

Berchenko Y, Rosenblatt JD, Frost SDW. 2017. Modeling and analyzing respondent-driven sampling as a counting process. *Biometrics*. **http://dx.doi.org/10.1111/biom.12678**

Brown T, Bao L, Eaton JW, Hogan DR, Mahy M, et al. 2014. Improvements in prevalence trend fitting and incidence estimation in EPP 2013. *AIDS* 28:S415–25

Buchmann JA. 2004. Cryptographic hash functions. In *Introduction to Cryptography*, ed. JA Buchmann, pp. 235–48. New York: Springer. 2nd ed.

Crawford FW. 2016. The graphical structure of respondent-driven sampling. *Sociol. Methodol.* 46:187–211

Crawford FW, Aronow PM, Zeng L, Li J. 2015. Identification of homophily and preferential recruitment in respondent-driven sampling. arXiv:1511.05397 [stat.AP]

Crawford FW, Wu J, Heimer R. 2017. Hidden population size estimation from respondent-driven sampling: a network approach. *J. Am. Stat. Assoc.* **http://dx.doi.org/10.1080/01621459.2017.1285775**

Erausquin JT, Reed E, Blankenship KM. 2014. Change over time in police interactions and HIV risk behavior among female sex workers in Andhra Pradesh, India. *AIDS Behav*. 19:1108–115

Erdős P, Rényi A. 1959. On random graphs. *Publ. Math.* 6:290–97

Fellows IE. 2012a. Deducer: a data analysis GUI for R. *J. Stat. Softw.* 49:1–15

Fellows IE. 2012b. *Exponential family random network models*. PhD Thesis, Univ. Calif., Los Angeles

Fellows IE, Handcock MS. 2012. Exponential-family random network models. arXiv:1208.0121 [stat.ME]

Fienberg SE, Johnson MS, Junker BW. 1999. Classical multilevel and Bayesian approaches to population size estimation using multiple lists. *J. R. Stat. Soc. Ser. A* 162:383–405

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*. 122:32–38

Gile KJ. 2011. Improved inference for respondent-driven sampling data with application to HIV prevalence estimation. *J. Am. Stat. Assoc.* 106:493:135–46

Gile KJ, Handcock MS. 2010. Respondent-driven sampling: an assessment of current methodology. *Sociol. Methodol.* 40:285–327

Gile KJ, Handcock MS. 2015. Network model-assisted inference from respondent-driven sampling data. *J. R. Stat. Soc. Ser. A* 178:619–39

Gile KJ, Johnston LG, Salganik MJ. 2015. Diagnostics for respondent-driven sampling. *J. R. Stat. Soc. Ser. A* 178:241–69

Goel S, Salganik MJ. 2009. Respondent-driven sampling as Markov chain Monte Carlo. *Stat. Med.* 28:2202–29

Goel S, Salganik MJ. 2010. Assessing respondent-driven sampling. *PNAS* 107:6743–47

Goodman LA. 1961. Snowball sampling. *Ann. Math. Stat.* 32:148–70

Handcock MS, Fellows IE, Gile KJ. 2014a. RDS Analyst: software for the analysis of respondent-driven sampling data. *R package*, *version 0.42*. **http://wiki.stat.ucla.edu/hpmrg**

Handcock MS, Fellows IE, Gile KJ. 2015a. RDS: Respondent-driven sampling. Los Angeles, CA. *R package*, *version 0.7-2.* **http://wiki.stat.ucla.edu/hpmrg**

Handcock MS, Gile KJ. 2011. Comment: on the concept of snowball sampling. *Sociol. Methodol.* 41:367–71

Handcock MS, Gile KJ, Mar CM. 2014b. Estimating hidden population size using respondent-driven sampling data. *Electron. J. Stat.* 8:1491

Handcock MS, Gile KJ, Mar CM. 2015b. Estimating the size of populations at high risk for HIV using respondent-driven sampling data. *Biometrics* 71:258–66

Hansen MH, Hurwitz WN. 1943. On the theory of sampling from finite populations. *Ann. Math. Stat.* 14:333–62

Heckathorn DD. 1997. Respondent-driven sampling: a new approach to the study of hidden populations. *Soc. Probl.* 44:174–99

Heckathorn DD. 2002. Respondent-driven sampling II: deriving valid population estimates from chain-referral samples of hidden populations. *Soc. Probl.* 49:11–34

Heckathorn DD. 2007. Extensions of respondent-driven sampling: analyzing continuous variables and controlling for differential recruitment. *Sociol. Methodol.* 37:151–207

Heckathorn DD, Semaan S, Broadhead RS, Hughes JJ. 2002. Extensions of respondent-driven sampling: a new approach to the study of injection drug users aged 18–25. *AIDS Behav.* 6:55–67

Horvitz DG, Thompson DJ. 1952. A generalization of sampling without replacement from a finite universe. *J. Am. Stat. Assoc.* 47:663–85

Iguchi MY, Ober AJ, Berry SH, Fain T, Heckathorn DD, et al. 2009. Simultaneous recruitment of drug users and men who have sex with men in the United States and Russia using respondent-driven sampling: sampling methods and implications. *J. Urban Health* 86:5–31

Johnston LG. 2007. *Conducting respondent driven sampling studies in diverse settings: a manual for planning RDS studies*. Cent. Dis. Control Prev., Atlanta, GA

Johnston LG. 2013a. *Introduction to HIV/AIDS and sexually transmitted infection surveillance. Module 4. Introduction to respondent-driven sampling*. World Health Organ., Geneva. **http://www.lisagjohnston.com/respondent-driven-sampling/respondent-driven-sampling**

Johnston LG. 2013b. *Introduction to HIV/AIDS and sexually transmitted infection surveillance. Module 4, supplement: a guide to using RDS Analyst and NetDraw*. World Health Organ., Geneva. **http://www.lisagjohnston.com/respondent-driven-sampling/respondent-driven-sampling**

Johnston LG, Malekinejad M, Kendall C, Iuppa IM, Rutherford GW. 2008. Implementation challenges to using respondent-driven sampling methodology for HIV biological and behavioral surveillance: field experiences in international settings. *AIDS Behav.* 12:131–41

Johnston LG, O'Bra H, Chopra M, Mathews C, Townsend L, et al. 2010. The associations of voluntary counseling and testing acceptance and the perceived likelihood of being HIV-infected among men with multiple sex partners in a South African township. *AIDS Behav.* 14:922–31

Johnston LG, Prybylski D, Raymond HF, Mirzazadeh A, Manopaiboon C, McFarland W. 2013. Incorporating the service multiplier method in respondent-driven sampling surveys to estimate the size of hidden and hard-to-reach populations: case studies from around the world. *Sex. Transm. Dis.* 40:304–10

Kendall C, Kerr LRFS, Gondim RC, Werneck GL, Macena RHM, 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

Khabbazian M, Hanlon B, Russek Z, Rohe K. 2016. Novel sampling design for respondent-driven sampling. arXiv:1606.00387 [stat.ME]

Kuchenhoff H, Mwalili SM, Lesaffre E. 2006. A general method for dealing with misclassification in regression: the misclassification SIMEX. *Biometrics* 62:85–96

Lansky A, Abdul-Quader AS, Cribbin M, Hall T, Finlayson TJ, 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

Lansky A, Drake A, Wejnert C, Pham H, Cribbin M, Heckathorn DD. 2012. Assessing the assumptions of respondent-driven sampling in the national HIV behavioral surveillance system among injecting drug users. *Open AIDS J.* 6:77–82

Li X, Rohe K. 2016. Central limit theorems for network driven sampling. arXiv:1509.04704 [stat.ME]

Liang KY, Zeger SL. 1986. Longitudinal data analysis using generalized linear models. *Biometrika* 78:13–22

Liu H, Li J, Ha T, Li J. 2012. Assessment of random recruitment assumption in respondent-driven sampling in egocentric network data. *Soc. Netw.* 1:13

Lu X. 2013. Linked ego networks: improving estimate reliability and validity with respondent-driven sampling. *Soc. Netw.* 35:669–85

Lu X, Bengtsson L, Britton T, Camitz M, Kim BJ, et al. 2012. The sensitivity of respondent-driven sampling. *J. R. Stat. Soc. Ser. A* 175:191–216

Lu X, Malmros J, Liljeros F, Britton T. 2013. Respondent-driven sampling on directed networks. *Electron. J. Stat.* 7:292–322

Lunagomez S, Airoldi E. 2014. Bayesian inference from non-ignorable network sampling designs. arXiv:1401.4718 [stat.ME]

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

Malekinejad M, Johnston LG, Kendall C, Kerr LRFS, Rifkin MR, Rutherford GW. 2008. Using respondent-driven sampling methodology for HIV biological and behavioral surveillance in international settings: a systematic review. *AIDS Behav*. 12:105–30

Malmros J, Liljeros F, Britton T. 2016a. Respondent-driven sampling and an unusual epidemic. *J. Appl. Probab.* 53:518–30

Malmros J, Masuda N, Britton T. 2016b. Random walks on directed networks: inference and respondent-driven sampling. *J. Off. Stat.* 32:433–59

McCreesh N, Copas A, Seeley J, Johnston LG, Sonnenberg P, et al. 2013. Respondent driven sampling: determinants of recruitment and a method to improve point estimation. *PLOS ONE* 8:e78402

McCreesh N, Frost S, Seeley J, Katongole J, Tarsh MN, et al. 2012. Evaluation of respondent-driven sampling. *Epidemiology* 23:138–47

McCreesh N, Johnston LG, Copas A, Sonnenberg P, Seeley J, et al. 2011. Evaluation of the role of location and distance in recruitment in respondent-driven sampling. *Int. J. Health Geogr.* 10:1

McLaughlin KR. 2016. *Modeling preferential recruitment for respondent-driven sampling*. PhD Thesis, Univ. Calif., Los Angeles

McLaughlin KR, Handcock MS, Johnston LG. 2015. Inference for the visibility distribution for respondent-driven sampling. In *JSM Proceedings 2015*, *Statistical Computing Section*. Alexandria, VA: Am. Stat. Assoc.

Merli MG, Moody J, Smith J, Li J, Weir S, Chen X. 2015. Challenges to recruiting population representative samples of female sex workers in China using respondent driven sampling. *Soc. Sci. Med.* 125:79–93

Mills HL, Johnson S, Hickman M, Jones NS, Colijn C. 2014. Errors in reported degrees and respondent driven sampling: implications for bias. *Drug Alcohol Depend.* 142:120–26

Molloy MS, Reed BA. 1995. A critical point for random graphs with a given degree sequence. *Random Struct. Algorithms* 6:161–79

Montealegre JR, Johnston LG, Murrill C, Monterroso E. 2013. Respondent driven sampling for HIV biological and behavioral surveillance in Latin America and the Caribbean. *AIDS Behav*. 17:2313–40

Mouw T, Verdery AM. 2012. Network sampling with memory. *Sociol. Methodol.* 42:206–56

Neely WW. 2010. *Statistical theory for respondent driven sampling*. PhD Thesis, Univ. Wisc., Madison

Ott MQ, Gile KJ. 2016. Unequal edge inclusion probabilities in link-tracing network sampling with implications for respondent-driven sampling. *Electron. J. Stat.* 10:1109–32

Paz-Bailey G, Jacobson J, Guardado M, Hernandez F, Nieto A, et al. 2011. How many men who have sex with men and female sex workers live in El Salvador? Using respondent-driven sampling and capture–recapture to estimate population sizes. *Sex. Transm. Infect.* 87:279–82

Paz-Bailey G, Miller W, Shiraishi RW, Jacobson JO, Abimbola TO, Chen SY. 2013. Reaching men who have sex with men: a comparison of respondent-driven sampling and time-location sampling in Guatemala City. *AIDS Behav*. 17:3081–90

Pitpitan EV, Smith LR, Goodman-Meza D, Torres K, Semple SJ, et al. 2016. "Outness" as a moderator of the association between syndemic conditions and HIV risk-taking behavior among men who have sex with men in Tijuana, Mexico. *AIDS Behav*. 20:431–38

R Core Team. 2016. *R: a language and environment for statistical computing*. Vienna: R Found. Stat. Comput.

Ramirez-Valles J, Molina Y, Dirkes J. 2013. Stigma towards PLWHA: the role of internalized homosexual stigma in Latino gay/bisexual male and transgender communities. *AIDS Educ. Prev.* 25:179–89

Rhodes SD, McCoy TP. 2015. Condom use among immigrant Latino sexual minorities: multilevel analysis after respondent-driven sampling. *AIDS Educ. Prev.* 27:27–43

Rohe K. 2015. Network driven sampling; a critical threshold for design effects. arXiv:1505.05461 [math.ST]

Rubin DB. 1976. Inference and missing data. *Biometrika* 63:581–92

Rudolph AE, Crawford ND, Latkin C, Heimer R, Benjamin EO, et al. 2011. Subpopulations of illicit drug users reached by targeted street outreach and respondent-driven sampling strategies: implications for research and public health practice. *Ann. Epidemiol.* 21:280–89

Rudolph AE, Fuller CM, Latkin C. 2013. The importance of measuring and accounting for potential biases in respondent-driven samples. *AIDS Behav*. 17:2244–52

Ott, MQ, Gile KJ, Harrison MT, Johnston LG, and Hogan JW. 2017. Reduced Bias for respondent-driven sampling: accounting for non-uniform edge sampling probabilities in people pho inject drugs in Mauritius. arXiv:1712.09149 [stat.AP].

Salganik MJ. 2006. Variance estimation, design effects, and sample size calculations for respondent-driven sampling. *J. Urban Health* 83:98–112

Salganik MJ, Heckathorn DD. 2004. Sampling and estimation in hidden populations using respondent-driven sampling. *Sociol. Methodol.* 34:193–240

Shi Y, Cameron CJ, Heckathorn DD. 2016. Model-based and design-based inference. *Sociol. Methods Res.* **http://dx.doi.org/10.1177/0049124116672682**

Silva-Santisteban A, Raymond HF, Salazar X, Villayzan J, Leon S, et al. 2012. Understanding the HIV/AIDS epidemic in transgender women of Lima, Peru: results from a sero-epidemiologic study using respondent driven sampling. *AIDS Behav.* 16:872–81

Spiller MW. 2009. *Regression modeling of data collected using respondent-driven sampling*. Master's Thesis, Cornell Univ.

Spiller MW, Gile KJ, Handcock MS, Mar CM, Wejnert C. 2017. Evaluating variance estimators for respondent-driven sampling. *J. Surv. Stat. Methodol.* https://doi.org/10.1093/jssam/smx018

Štulhofer A, Ba´cak V, Bo˘zi˘cevi´c I, Begovac J. 2008. HIV-related sexual risk taking among HIV-negative men who have sex with men in Zagreb, Croatia. *AIDS* 12:505–12

Tomas A, Gile KJ. 2011. The effect of differential recruitment, non-response and non-recruitment on estimators for respondent-driven sampling. *Electron. J. Stat.* 5:899–934

Tran HV, Le LVN Johnston LG, Nadol P, Van Do A, et al. 2015. Sampling males who inject drugs in Haiphong, Vietnam: comparison of time-location and respondent-driven sampling methods. *J. Urban Health* 92:744–57

UNAIDS. 2009. *Estimating national adult prevalence of HIV-1 in concentrated epidemics*. Tech. Rep., UNAIDS Joint United Nations Programme on HIV/AIDS

Verdery AM, Fisher JC, Siripong N, Abdesselam K, Bauldry S. 2016. New survey questions and estimators for network clustering with respondent-driven sampling data. arXiv:1610.06683 [stat.ME]

Verdery AM, Merli MG, Moody J, Smith J, Fisher JC. 2015a. Respondent-driven sampling estimators under real and theoretical recruitment conditions of female sex workers in China. *Epidemiology* 26:661

Verdery AM, Mouw T, Bauldry S, Mucha PJ. 2015b. Network structure and biased variance estimation in respondent driven sampling. *PLOS ONE* 10:e0145296

Volz E, Heckathorn DD. 2008. Probability based estimation theory for respondent driven sampling. *J. Off. Stat.* 24:79

Volz E, Wejnert C, Cameron C, Barash V, Degani I, Heckathorn DD. 2012. Respondent-driven sampling analysis tool (RDSAT), version 7.1. *Statistical software*. **http://www.respondentdrivensampling.org/**

Wang J, Carlson RG, Falck RS, Siegal HA, Rahman A, Li L. 2005. Respondent-driven sampling to recruit MDMA users: a methodological assessment. *Drug Alcohol Depend.* 78:147–57

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

Wejnert C. 2010. Social network analysis with respondent-driven sampling data: a study of racial integration on campus. *Soc. Netw.* 32:112–24

Wejnert C, Heckathorn DD. 2008. Web-based network sampling: efficiency and efficacy of respondent-driven sampling for online research. *Sociol. Methods Res.* 37:105–34

White RG, Hakim AJ, Salganik MJ, Spiller MW, Johnston LG, et al. 2015. Strengthening the reporting of observational studies in epidemiology for respondent-driven sampling studies: STROBE-RDS statement. *J. Clin. Epidemiol.* 68:1463–71

Yamanis TJ, Merli MG, Neely WW, Tian FF, Moody J, et al. 2013. An empirical analysis of the impact of recruitment patterns on RDS estimates among a socially ordered population of female sex workers in China. *Sociol. Methods Res.* 42:392–425

Zhong F, Lin P, Xu H, Wang Y, Wang M, et al. 2011. Possible increase in HIV and syphilis prevalence among men who have sex with men in Guangzhou, China: results from a respondent-driven sampling survey. *AIDS Behav.* 15:1058–66