

Total Survey Error and Respondent Driven Sampling: Focus on Nonresponse and Measurement Errors in the Recruitment Process and the Network Size Reports and Implications for Inferences

Sunghee Lee¹, Tuba Suzer-Gurtekin¹, James Wagner¹, and Richard Valliant¹

This study attempted to integrate key assumptions in Respondent-Driven Sampling (RDS) into the Total Survey Error (TSE) perspectives and examine TSE as a new framework for a systematic assessment of RDS errors. Using two publicly available data sets on HIV-at-risk persons, nonresponse error in the RDS recruitment process and measurement error in network size reports were examined. On nonresponse, the ascertained partial nonresponse rate was high, and a substantial proportion of recruitment chains died early. Moreover, nonresponse occurred systematically: recruiters with lower income and higher health risks generated more recruits; and peers of closer relationships were more likely to accept recruitment coupons. This suggests a lack of randomness in the recruitment process, also shown through sizable intra-chain correlation. Self-reported network sizes suggested measurement error, given their wide dispersion and unreasonable reports. This measurement error has further implications for the current RDS estimators, which use network sizes as an adjustment factor on the assumption of a positive relationship between network sizes and selection probabilities in recruitment. The adjustment resulted in nontrivial unequal weighting effects and changed estimates in directions that were difficult to explain and, at times, illogical. Moreover, recruiters' network size played no role in actual recruitment. TSE may serve as a tool for evaluating errors in RDS, which further informs study design decisions and inference approaches.

Key words: Sampling hard-to-reach populations; chain referral; network-based sampling; measurement error; nonresponse error.

1. Introduction

This article attempts to provide a framework for evaluating Respondent-Driven Sampling (RDS) by integrating its key assumptions into the Total Survey Error (TSE), (Groves 1989) as suggested by Lee (2009). RDS, introduced by Heckathorn (1997, 2002), has gained tremendous popularity due to rising demands for data on rare, hidden and/or elusive populations, for example, sexual minorities (for example, Ramirez-Valles et al. 2005), injection drug users (for example, Burt et al. 2010), racial and ethnic minorities (for example, Dombrowski et al. 2013) and recent immigrants (for example, Montealegre et al. 2013). Not only in the scientific communities is RDS popular, but also in government statistical systems. RDS is practiced by the Centers of Disease Control and Prevention in

¹ Institute for Social Research, University of Michigan, 426 Thompson St., Ann Arbor, MI 48104, U.S.A. Emails: sungheel@umich.edu, tsuzer@umich.edu, jameswag@umich.edu, and rvalliant@umd.edu.

Acknowledgment: This research was supported by the National Science Foundation [grant number SES-1461470].

the United States (Lansky et al. 2007; Centers for Disease Control and Prevention (CDC) 2009, 2013; Lin et al. 2013).

While there is an attempt to improve analytic aspects of RDS (for example, Salganik and Heckathorn 2004; Volz and Heckathorn 2008; Gile 2011), design aspects aligned with the realities of data collection remain largely unexamined. The TSE framework allows a systematic examination of errors, which, in turn, further informs assessing design and analytic aspects and refining them to reduce overall error. This study examines TSE as a new framework for a systematic assessment of RDS errors by using two publicly available data sets on HIV-at-risk persons. Section 2 provides an overview of RDS by comparing its theoretical development and current practice and then turns to a set of assumptions in RDS and their relevance to TSE. Data sources and methods used in this study are introduced in Section 3. Sections 4 and 5 report results from the analysis. We offer a summary of this study and open questions in Section 6.

2. Respondent Driven Sampling

2.1. Overview of RDS

While rare in the general population, some population subgroups are interlinked. For instance, use of injection drugs often involves others who also inject drugs, and this connectedness directly forms informal social networks among Injection Drug Users (IDUs). Although rare and hidden from the outsiders, IDUs may be easily located within these networks. RDS attempts to locate these social networks and exploit them to generate samples.

RDS roughly follows these steps in practice: First, researchers recruit a few members of the target group typically through some type of convenience sampling and collect data from them. While data collection ends at this point in traditional sample surveys, these respondents in RDS are asked to recruit their peers in their social networks. Recruited peers become respondents as well as recruiters for further recruitment. Data collection and recruitment proceed in “waves”, as seen in Figure 1, until the cumulative sample size reaches the target sample size or some other criteria set in respective studies (for example, available resources, timeline). Respondents in RDS are not only the source of data, but also recruiters for participants in the immediately subsequent wave (hence, respondent driven sampling). For this reason, those recruited initially by researchers are called *seeds*. As noted in Figure 1, recruitment chains are formed from each seed. Under a set of assumptions examined in Subsection 2.2., these chains are regarded as Markov chains, necessitating the chain length to be reasonably long. This process then leads RDS to stationary probabilities (or equilibrium), where the characteristics of the cumulative sample become independent of seeds’ characteristics. This is also a point at which the sample is assumed to become unbiased (Heckathorn 1997, 2002).

A distinctive feature of RDS recruitment is the usage of recruitment coupons. In practice, a predetermined number of coupons are given to recruiters, who then distribute the coupons to their peers. These recruits need to redeem the coupons in order to participate in the study. Once they participate, they are given coupons to distribute to their peers. (Naturally, seeds participate in the study without coupons, and, hence, seeds only distribute coupons.) With serial numbers on the coupons, the link between recruiters and

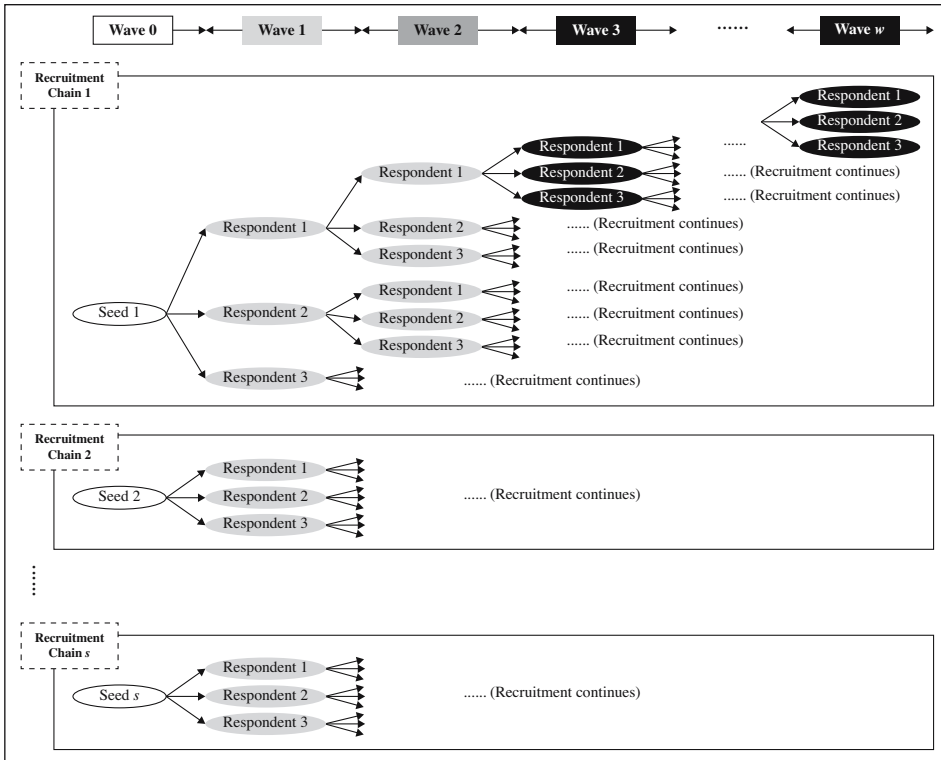


Fig. 1. Respondent-driven sampling recruitment process in theory.

their recruits can be traced. As recruitment is done through coupons, RDS does not require collecting any personal information about respondents’ peers. Coupons also play an important role in incentivizing participation, which requires redeeming coupons, and recruitment efforts, which is reflected in the number of redeemed coupons, equating to the number of recruits. Coupons not only decrease the data collection costs but also eliminate concerns with privacy, which led some to consider RDS innovative (for example, Baker et al. 2013) and to advocate RDS consistent with the “voluntariness” spirit of the research participants, while criticizing probability sampling as being intrusive (for example, Constantine 2010). On the other hand, others have raised concerns about bias in RDS due to the incentivized nature of recruitment and the potential for unwarranted influence or coercion in the recruitment process (for example, Phillips 2010; Simon and Mosavel 2010).

One piece of critical information in RDS is the number of peers in respondents’ networks, termed as *degree*, as it is a key element of RDS estimators (for example, Salganik and Heckathorn 2004; Volz and Heckathorn 2008; Gile 2011). For instance, the RDS-II estimator (Volz and Heckathorn 2008), takes the form of the Horvitz-Thompson or Hájek estimator as follows:

$$\hat{y} = \sum_{i \in S} (y_i d_i^{-1}) / \sum_{i \in S} (d_i^{-1}), \tag{1}$$

where y_i is a variable of interest measured on person i in the sample S and d_i is the network size of person i . Essentially, d_i^{-1} is used as an adjustment factor on the assumption that

persons with larger networks have higher chances of being sampled. This leads d_i^{-1} to be called a “weight” in the RDS literature. For instance, a weight of one will be given if a respondent has one peer and a weight of 0.01 if a respondent has 100 peers. By using this weight, estimation considers the characteristics of respondents with larger networks at a lower level than those with smaller networks. It should be noted that d_i^{-1} is different than weights in probability sampling, which are used to estimate population totals as incorporating population-level information. Weights in RDS simply rearrange the sample distribution by network sizes without incorporating population-level information and, hence, are irrelevant for estimating population totals. In fact, estimating population totals using RDS data requires intensive computing work (Handcock 2012).

2.2. Assumptions in Respondent Driven Sampling

Theoretical developments of RDS are based on a set of assumptions (Heimer 2005; Gile and Handcock 2010). Although essential for the claimed unbiasedness, these assumptions are strong, unrealistic and often difficult to verify, and violations are ignored in the inference. We discuss six assumptions. Note that the last two (Assumptions E and F) are not explicitly discussed in the RDS literature, but are critical for using existing RDS estimators.

A. Network Structure: RDS assumes that there is only one network that covers the entire population of interest. That is, everyone in the population can be traced from any starting point. This assumption requires a dense network with a single component. If the population includes multiple non-linked or loosely-linked networks, this assumption is violated. It was shown that estimates were sensitive to such a violation (Lu et al. 2012).

B. Equilibrium Condition: Let vector Q be the successive indices of units sampled by the random walk process and Q_k be the index of units sampled at the k^{th} wave. This follows the Markov process with a transition matrix,

$$P(Q_{k+1} = j | Q_k = i) = \begin{cases} 1/d_i, & \text{if } d_{ij} = 1 \\ 0, & \text{if } d_{ij} = 0 \end{cases}, \quad (2)$$

where d_{ij} is a link function in the sociomatrix of relations between unit i and unit j , defined as

$$d_{ij} = \begin{cases} 1, & \text{if there is a link between unit } i \text{ and unit } j \\ 0, & \text{if there is no link between unit } i \text{ and unit } j \end{cases}, \quad (3)$$

and $d_i = \sum_{j \neq i} d_{ij}$ in (1). Equilibrium assumption is that, as recruitment waves continue, the characteristics of recruits become independent of the seeds' characteristics. In other words, selection of seeds is not critical in the overall inference (Heckathorn 1997). This is directly related to the memorylessness of Markov chains, where the future and past states are independent given the present state. This allows us to rewrite (2) as $P(Q_{k+1} = j | Q_k = i) = P(Q_{k+1} = j | Q_k = i, Q_{k-1} = i - 1, \dots, Q_1 = 1)$. Further, the equilibrium state in RDS is assumed to be approached at a geometric rate.

C. Random Recruitment: RDS lets respondents control the recruitment process (Frost et al. 2006). The assumption here is that recruitment is done at random, implying that

recruiters do not use systematic criteria for selecting their recruits. The transition matrix in (2) is achieved only when any given unit j ($j = 1, \dots, d_i$) within the network of unit i selected at the k th wave has the equal probability to be selected into the $(k + 1)^{\text{th}}$ wave. Lu et al. (2012) showed that systematic recruitment results in a large bias and variance.

D. Equal Homophily: Homophily is the tendency of individuals of similar characteristics to associate with one another. RDS assumes that homophily rate in the recruitment is equal across subgroups: the tendency of a member of group G to recruit other members of group G is the same as a group H member recruiting group G members.

E. Complete Response: Unlike traditional surveys, where nonresponse occurs at the time of an interview attempt, nonresponse in RDS occurs in four stages:

- (1) whether respondents take coupons,
- (2) whether those who take coupons actually distribute them to their peers,
- (3) whether their peers accept coupons, and
- (4) whether the peers who accept coupons actually participate in the study. Overall, these stages can be expressed with a vector of 0/1 response indicators, $\mathbf{r} = (r_1, r_2, r_3, r_4)$.

Note that nonresponse on the last two stages implies not only their own nonresponse, but also nonresponse by their peers who may, otherwise, be open to accepting coupons and participating in data collection. Current RDS practice assumes a 100 percent response rate for all stages (i.e., $r_1 = r_2 = r_3 = r_4 = 1$). This compound nature of nonresponse has been recognized only very recently (for example, Lee et al. 2012; Gile et al. 2015).

This assumption affects error properties through nonresponse bias and overall sampling productivity in RDS. On the error properties, with the multiple stages of nonresponse introduced above, RDS is subject to a larger scope for nonresponse bias than traditional probability sampling. Under complete response, RDS sample sizes over waves should grow exponentially. However, in the presence of nonresponse at any of the four stages, this exponential growth becomes unlikely. This further leads to slow sample size growth or smaller sample sizes than expected, and short recruitment chains, breaking the Markov process, which is required for Assumptions A, B, and C. Hence, slow sample size growth, small sample sizes, or short recruitment chains may serve as evidence for equilibrium being not realized. It should be noted that, as nonresponse affects RDS sampling productivity, the number of seeds and the chain length are unascertainable during design stages.

F. Accurate and Complete Network Size Measures: In practice, RDS estimators use \tilde{d}_i , a self-reported network size, instead of d_i . For example, the RDS-II estimator in (1) becomes $\tilde{y} = \sum_{i \in S} (y_i \tilde{d}_i^{-1}) / \sum_{i \in S} (\tilde{d}_i^{-1})$. This implicitly assumes that \tilde{d}_i either is error free (i.e., $\tilde{d}_i = d_i$) or has a fixed error rate across i (for example, $\tilde{d}_i = p d_i$, where p is a constant, $0 < p \leq 1$). Notably, \tilde{d}_i is self-reported, subject to measurement error. The social network literature clearly indicates that obtaining an accurate network size is challenging because the scope and the nature of the networks are not standardized and that, even when the networks are defined narrowly, it is still found to be difficult (Laumann et al. 1983; Marsden 1990). This makes the error-free assumption or the fixed error rate assumption an unlikely scenario. Additionally, \tilde{d}_i is subject to item nonresponse as well as zero network size reports, posing additional difficulties in using \tilde{d}_i^{-1} as a weight variable.

2.3. Total Survey Error and Respondent Driven Sampling

Because RDS is a sampling method, one may conclude that RDS is subject only to sampling error. However, the chain referral in RDS affects all error types under the TSE framework. These errors are related to all assumptions in Subsection 2.2. In an attempt to frame these errors, we discuss each component of TSE in relation to the RDS assumptions below.

A. Coverage Error: While RDS does not use frames directly, obviously, networks with multiple components or with loose connections result in coverage error. Moreover, people's perceived social network structure determines coverage, making the network structure assumption relevant. Recruiters' understanding of the target population (for example, jazz musicians in Heckathorn and Jeffri 2001) is critical. This equates to the boundary specification, which has long been acknowledged as a problem in the social network literature (Laumann et al. 1983).

B. Sampling Error: Sampling error results from using a sample for inference rather than the entire population for inferences. With probability sampling, estimates are (approximately) unbiased in expectation, and the sampling variance is the sole source of sampling error. However, both sampling bias and variance come into play for nonprobability samples, including RDS. Assumptions about random recruitment and equal homophily directly influence the sampling bias through the unmet equilibrium assumption. If recruitment is done systematically, then the assumptions are violated and sampling bias is likely.

C. Nonresponse Error: A violation of the complete response assumption in RDS is the same as nonresponse error in TSE. While previous research recognizes this as an uncontrollable aspect of recruitment (for example, Gile and Handcock 2010), it is addressed as a sampling issue rather than a nonresponse issue. Understanding nonresponse in RDS is a complex task. Even calculating response rates is difficult, if not impossible. This is because the denominator required for calculating response rates includes all eligible peers in a participant's network to whom recruitment is attempted. Of course, this is not the same as the number of distributed coupons, because participants may attempt to recruit without involving coupons. If coupons are not involved, the number of unsuccessful recruitment attempts is unknown.

Moreover, of the four nonresponse stages (r) in Subsection 2.2.E, the current RDS practice captures information about Stages 1 and 4 only (whether participants take coupons and whether coupons are redeemed by their peers), providing partial information about nonresponse. Without monitoring the entire recruitment process, the magnitude and the effect of nonresponse cannot be ascertained.

Despite nonresponse compounded of multiple stages, there is little effort to understand nonresponse in RDS. For probability sample surveys, covariates of nonresponse have been studied extensively (for example, Groves and Couper 1998) and are incorporated through post-survey adjustments. Nonresponse follow-up studies have been recommended for RDS and implemented (for example, Gile et al. 2015), but as discussed in Section 6, the design is yet to be established to generate useful data.

D. Measurement Error: While participants being recruiters is a unique feature of RDS, they are also a unique source of measurement error, which also affects other types of error. First, their social network structure (for example, density, single vs. multiple components) and their understanding of the target population definition (that is, the boundary specification

problems in Laumann et al. 1983) both affect noncoverage error. Second, criteria recruiters use for selecting their recruits determine the selection mechanism, influencing sampling error. More importantly, \tilde{d}_i is subject to measurement error, potentially affecting overall inferences. Unreasonable network size reports have also been noted (for example, Wejnert and Heckathorn 2008; Schonlau 2014) with evidence that the report of network size is sensitive to question wording (McCreesh et al. 2012; Schonlau 2014). Measurement error in \tilde{d}_i has implication for the bias of \tilde{y} with an unclear direction. Also, \tilde{d}_i influences the variance of \tilde{y} : the larger the variation of \tilde{d}_i , the larger the variance of \tilde{y} .

In summary, what sets RDS apart from traditional probability or adaptive sampling is who has the control over sample selection (Frost et al. 2006). As sample selection is controlled by participants, not by researchers in RDS, statistical inferences are challenging (Frost et al. 2006), requiring a set of strong assumptions (Heimer 2005; Gile and Handcock 2010). Some studies provide cautionary remarks about RDS in terms of bias (for example, Martin et al. 2003; Wejnert and Heckathorn 2008; McCreesh et al. 2012) and variance (for example, Goel and Salganik 2010; Verdery and Mouw 2012; Verdery et al. 2015) and call for its evaluation on empirical data (for example, Heimer 2005; Burt et al. 2010; Simon and Mosavel 2010; Lu et al. 2012, Salganik 2012; Gile et al. 2015). Still, others use RDS data without considering or acknowledging potential limitations. For instance, Lee and colleagues (2011) asserted for replacing probability sampling with RDS entirely, even for general population studies.

It is important to note two clear differences between RDS and the network sampling by Sirken (1972, 1975, 1997). First, Sirken's network sampling uses well-specified networks, such as direct family members and biological siblings, whereas RDS uses loosely defined networks, such as acquaintances and friends. Second, in Sirken's network sampling, respondents provide the information about their peers that researchers use for drawing a sample. On the other hand, in RDS, participants sample on their own. Therefore, who controls the sampling process is completely different, although the word "network" may appear to suggest similarities.

Reflecting the recency of its introduction, the realities of data collection using RDS remain to be scrutinized. The scarcity of publicly available RDS data is a further impediment to methodological assessments (Salganik 2012). See Appendix Table A1 for a list of publicly available RDS data sets. This study uses two publicly available RDS data sets with recruitment information on similar topics (for example, HIV), in similar locales (for example, Chicago), and examines the realities of RDS data using the TSE framework on two specific errors: nonresponse error arising in the recruitment process and measurement error in the network size reports. We focus on these two errors, because the current practice of RDS does not provide adequate data for assessing remaining errors.

3. Data and Methods

3.1. Data

3.1.1. Overview

We use data sets from two RDS studies available from the Inter-University Consortium for Political and Social Research (ICPSR): the Sexual Acquisition and Transmission of HIV

Cooperative Agreement Program (SATHCAP) and the Latino MSM Community Involvement (LMSM). SATHCAP targeted those at high risk of HIV/AIDS (for example, IDUs, men who have sex with men) and their sexual partners in four cities: Los Angeles (LA), California; Chicago, Illinois; Raleigh-Durham, North Carolina; and St. Petersburg, Russia and was conducted in two phases using independent samples (Compton et al. 2009; Iguchi et al. 2009). SATHCAP data from ICPSR included the three US cities between November 2006 and August 2008 (Iguchi et al. 2010). LMSM was conducted in San Francisco (SF), California and Chicago, Illinois through 2003 and 2004, targeting Latino gay or bisexual men and transgenders (Ramirez-Valles 2013).

The reasons for using these data sets are three-fold. First, to the best of our knowledge, SATHCAP and LMSM are the only publicly available RDS data sources with coupon distribution information, which is necessary to trace the link between recruiters and recruits and to ascertain the recruitment process, including how many coupons were given to each recruiter and how many were redeemed by his/her recruits. Second, using two independent RDS studies on similar topics allows us to examine whether the errors and their impact replicate across studies. Third, as these studies include roughly consistent study sites, the effect of geography, that may, otherwise, confound the results, can be minimized. For geographical consistency, this study included LA and Chicago from SATHCAP and SF and Chicago from LMSM, resulting in the sample of 3,584 for SATHCAP (845 for LA and 2,739 for Chicago) and 643 for LMSM (323 for SF and 320 for Chicago). However, it should be noted that information about these studies is limited to what is publicly available. Information about, for example, decisions around incentives and sample sizes could not be verified.

3.1.2. Nonresponse Follow-Up Study

In addition to the main data collection, SATHCAP conducted a follow-up study at the time of their return visit to study sites to obtain recruitment incentives. It included questions ascertaining the number and characteristics of peers who had accepted (“accepters”) and refused (“refusers”) coupons from participants: for example, “How many people accepted study coupons from you?”, “How many people refused to accept study coupons from you?”, and “Of the [reported number] people who accepted study coupons from you, how many are friends of yours?” With this data set, accepters and refusers can be compared on various characteristics. The follow-up study participation rate was 45.2% ($n = 382$) for LA and 56.1% ($n = 1,537$) for Chicago.

3.1.3. Measurement of Network Size

The network size in SATHCAP was measured by combining information from the following three questions: 1) “How many people do you know personally (that is, you know their name, you know who they are, and they know you, and you have seen them in the last six months) who use heroin, methamphetamines, and/or powder or crack cocaine or who inject some other drug?”; 2) “How many people do you know personally (that is, you know their name, you know who they are and they know you and you have seen them in the last six months) who are men who have sex with men?”; and 3) “How many of the men who have sex with men that you know use heroin, methamphetamines, crack and/or powder cocaine or inject some other drug?” The network size in LMSM was based on the

answer to the question “how many Latino gay, bisexual and transgenders over 18 years old in San Francisco/Chicago do you know?”

3.2. Analysis Procedure

3.2.1. Nonresponse Error

Nonresponse error was first examined using coupon distribution data. By linking recruiters and their recruits, we assessed nonresponse at Stages 1 and 4 discussed in Subsection 2.2.E. We then connected nonresponse with the sample size growth and the recruitment chain length. We also examined the potential correlates of nonresponse. Specifically, we considered recruiters' age, race/ethnicity, nativity, education, income, living arrangement, HIV status, substance use, sexual behavior, incarceration, and network size and their relationship with the number of successful recruits in Poisson regression to reflect the distribution of the dependent variable. Recruitment chains are partly affected by nonresponse and could be considered as clusters. Hence, we also examined Intra-Chain Correlation (ICC) to assess homogeneity within chain.

Further, using the SATHCAP follow-up study, we examined the Stage 3 nonresponse pattern. Specifically, we compared those who accepted versus refused coupons. Usefulness of the follow-up study was assessed with respect to its own nonresponse and measurement issues.

3.2.2. Measurement Error

Measurement error of the reported network size was first examined through basic data checks. In addition to the standard weight (\tilde{d}_i^{-1}), we used smoothed weights by top and bottom coding the network size at its 10th and 90th percentile, adopting the idea of weight trimming routinely performed in survey sampling to minimize the effect of extreme weights (Potter 1988; Little et al. 1997). Although large weights are often discussed for trimming (for example, Elliott 2009), extreme weights include both small and large weights as they both increase variability in estimates (see Valliant et al. 2013, p. 388). In fact, some consider both small and large weights for trimming (for example, Cole and Hernan 2008; Izrael et al. 2009). As shown in Table 6, the 10th and 90th percentiles equated to a network size of 2 and 50 for LA and 3 and 50 Chicago in SATHCAP and 3 and 75 for SF and 3 and 40 for Chicago in LSM. Weights cannot be ascertained for cases with missing or zero network sizes. However, these occurred infrequently and imputed weights on these cases made no difference in analytic results. Hence, these cases were not assigned with weights. We compared the Unequal Weighting Effect (UWE, Kish 1992) between standard and smoothed weights and examined the relationship between respondent characteristics and their weights.

We then examined the effect of weighting on estimation by comparing unweighted estimates and estimates weighted by standard and smoothed weights. We considered both univariate and bivariate statistics. For univariate statistics, we examined proportions of various sociodemographics, health status, and risk behaviors. For bivariate statistics, the associations between HIV status and characteristics known to be related to HIV (for example, substance use, sexual behavior), as well as characteristics known to be unrelated

to HIV (for example, network size) were examined through simple logistic regression that modelled HIV status on these characteristics one by one.

There are very few verified estimators and software options for RDS. Some software (for example, RDS Analyst introduced shortly) requires entire coupon distribution information and accommodates standard weights only. This study used:

- (1) an unweighted naïve estimator that incorrectly assumed simple random sampling,
- (2) the RDS-II estimator (Volz and Heckathorn 2008) with standard weights, and
- (3) the RDS-II estimator with smoothed weights for univariate statistics.

Their standard errors were calculated using the bootstrap method in Salganik and Heckathorn (2004) and Salganik (2006). Note that, although Volz and Heckathorn (2008) introduced a variance estimator for the RDS-II (equation 17 in their article), it requires information about all network members in the population and, hence, cannot be used for sample data. Unweighted proportions and their standard errors were computed in SAS. An R package RDS by Handcock and his colleagues (2014) was used for the RDS-II. We also used RDS Analyst, a software by Hard-to-Reach Population Methods Research Group (<http://www.hpmrg.org/>) for the RDS-II with standard weights. The results when using the weights from RDS Analyst were virtually the same as estimates using standard weights and, hence, not presented in this article.

For bivariate statistics, there are no known or suggested estimators for logistic regression model parameters in the literature. Given this, we used proc surveylogistic in SAS with and without weights, focusing on the estimated coefficients and their significance.

4. Nonresponse Error

4.1. Recruitment Process Through Coupon Distribution and Redemption

Within each study, the recruitment started with similar numbers of seeds across cities. For SATHCAP, there were 117 seeds for LA and 132 for Chicago. LA seeds were recruited using passive recruitment (for example, flyers and advertisements) while Chicago added active recruitment (for example, study staff approaching potentially eligible community members) (Iguchi et al. 2009).

From these seeds, a total sample size of 845 was generated over 19 waves in LA and 2,739 in Chicago over 45 waves. Seeds in LMSM were recruited actively using prespecified sociodemographics criteria: country of origin, main language spoken, HIV status, gender, and sexual orientation (that is, gay, bisexual, transgender) (Ramirez-Valles et al. 2005). In LMSM, 17 seeds generated a total sample size of 323 over twelve waves in SF and 13 seeds generated 320 over nine waves in Chicago. Rows A through C in Table 1 summarize the recruitment process.

When examining the recruitment process at the recruiter level, there were 842 (row D) potential recruiters who could have taken and distributed coupons, for example, in SATHCAP LA. Among them, 769 (row E) actually took coupons and 410 (row L) generated actual recruits. This actual recruitment rate (row M) was 48.7% for LA and 55.8% for Chicago in SATHCAP and 50.0% for SF and 49.0% for Chicago in LMSM.

Up to six coupons were given to all respondents, except for those in the last wave in SATHCAP and three coupons in LMSM. However, not all potential recruiters took coupons for distribution. Rows D and E in Table 1 compare the number of potential recruiters (that is, who could have taken coupons) versus actual recruiters (that is, who took coupons). This equates to nonresponse Stage 1. The rate of actual recruiter (row F) ranged from 89.1% for LMSM SF to 96.9% for SATHCAP Chicago.

A total of 3,140, 8,245, 854, and 917 coupons were distributed in SATHCAP LA, SATHCAP Chicago, LMSM SF and LMSM Chicago (row G), equating to an average of 3.73, 3.02, 2.67, and 2.96 coupons taken by potential recruiters (row H). As expected, not all coupons were redeemed. The number of redeemed coupon per potential recruiter ranged from 0.87 for SATHCAP LA to 0.99 LMSM Chicago (row K), meaning that less than one recruit was generated per potential recruiter. Recall that each potential recruiter could have generated up to six additional recruits in the immediately subsequent wave in SATHAP and up to three in LMSM.

Table 1. Summary of recruitment process by city, the Sexual Acquisition and Transmission of HIV Cooperative Agreement Program (SATHCAP) and the Latino MSM Community Involvement (LMSM).

| | SATHCAP | | LMSM | |
|---|---------|---------|-------|---------|
| | LA | Chicago | SF | Chicago |
| Overall recruitment results | | | | |
| A. No. of seeds | 117 | 132 | 17 | 13 |
| B. No. of total data collection waves | 19 | 45 | 12 | 9 |
| C. Total sample size (i.e., all respondents, including seeds) | 845 | 2,739 | 323 | 320 |
| D. No. of potential recruiters (= C – no. of last wave respondents) | 842 | 2,735 | 320 | 310 |
| Coupon distribution | | | | |
| E. No. of actual recruiters (i.e., those who took coupons) | 769 | 2,650 | 285 | 299 |
| F. Rate of actual recruiters (= E/D) | 91.3% | 96.9% | 89.1% | 96.5% |
| G. No. of coupons taken by potential recruiters | 3,140 | 8,245 | 854 | 917 |
| H. Average no. of coupons taken by potential recruiters (= G/D) | 3.73 | 3.02 | 2.67 | 2.96 |
| Coupon redemption | | | | |
| I. No. of recruits (i.e., redeemed coupons) | 728 | 2,607 | 306 | 307 |
| J. Coupon redemption rate (= I/G) | 23.2% | 31.6% | 35.8% | 33.5% |
| K. Average no. of recruits generated by potential recruiters (= I/D) | 0.87 | 0.95 | 0.96 | 0.99 |
| L. No. of recruiters generating recruits (i.e., those whose coupons were redeemed by peers) | 410 | 1,526 | 160 | 152 |
| M. Actual recruitment rate (= L/D) | 48.7% | 55.8% | 50.0% | 49.0% |

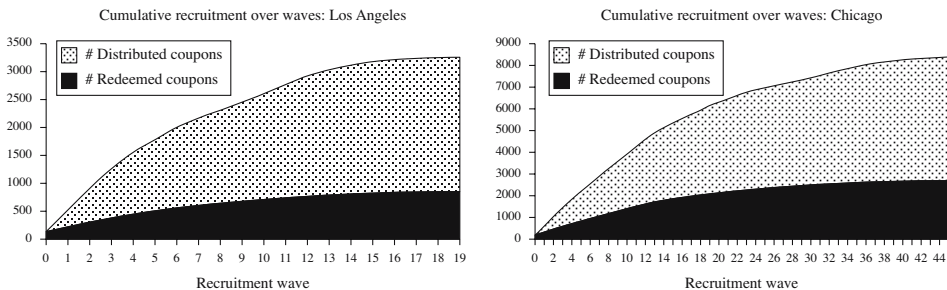
4.2. Coupon Redemption Rates, Sample Sizes, and Recruitment Chain Length

As there is no viable way of measuring response rates for RDS, we used coupon redemption rates (row J of Table 1) as a proxy. Although neither complete nor perfect, this was the only measure that reflected Stage 4 nonresponse. Because there are three other stages in RDS nonresponse, coupon redemption rates reported here indicate an upper bound for the true response rates. This rate ranged from 23.2% in SATHCAP LA to 35.8% in LMSM SF.

If all potential recruiters took coupons and their peers accepted and redeemed coupons, the cumulative sample size over recruitment waves would grow exponentially and all recruitment chains would reach the same length. However, with low coupon redemption rates and a small number of recruits per potential recruiter, cumulative sample sizes in Figure 2 grew in a quadratic, not the assumed exponential, pattern and approached a stationary phase rather rapidly. This was true across cities and studies.

At the recruitment chain level, nonresponse occurred differently, resulting in differential lengths as summarized in Table 2. On average, after seeds, chains lasted for as short as 1.56 waves in SATHCAP LA and as long as 4.38 waves in LMSM Chicago. Chains lasted longer in LMSM than in SATHCAP and in Chicago than in California cities. The distribution of chain lengths of SATHCAP was highly skewed, with the medians far

A. SATHCAP



B. LMSM

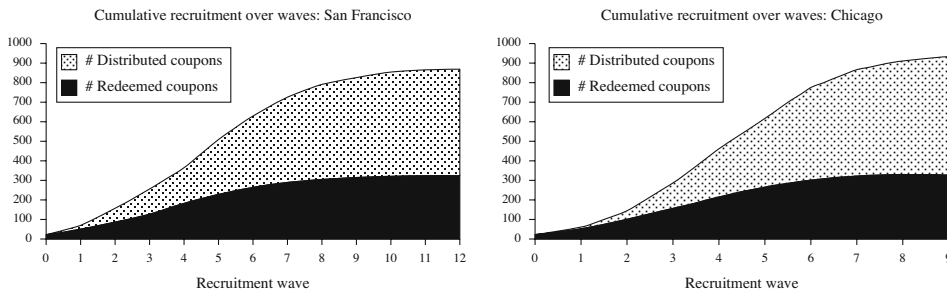


Fig. 2. Cumulative sample sizes (number of distributed coupons and redeemed coupons) by city, the Sexual Acquisition and Transmission of HIV Cooperative Agreement Program (SATHCAP) and the Latino MSM community involvement (LMSM). Note. The number of redeemed coupon equals to the sample size; Wave 0 is consisted of seeds only who did not have to redeem coupon to participate. Hence, the coupon redemption rates are inapplicable.

Table 2. Distribution of recruitment chain lengths, the Sexual Acquisition and Transmission of HIV Cooperative Agreement Program (SATHCAP) and the Latino MSM community involvement (LMSM).

| | SATHCAP | | LMSM | |
|---|---------|---------|------|---------|
| | LA | Chicago | SF | Chicago |
| No. of recruitment chains (i.e., no. of seeds) | 117 | 132 | 17 | 13 |
| Chain length | | | | |
| Average | 1.56 | 3.39 | 3.76 | 4.38 |
| Standard deviation | 2.33 | 5.52 | 2.25 | 1.93 |
| Maximum | 18 | 44 | 11 | 8 |
| 90th percentile | 4 | 7 | 9 | 8 |
| 75th percentile | 2 | 4 | 5 | 7 |
| Median | 0 | 1 | 4 | 5 |
| 25th percentile | 0 | 0 | 2 | 2 |
| 10th percentile | 0 | 0 | 0 | 0 |
| Minimum | 0 | 0 | 0 | 0 |
| Mode | 0 | 0 | 0 | 0 |
| % of chains died after seed (i.e., chain length = 0) | 58.1 | 32.6 | 23.5 | 15.4 |

below the means. In fact, 58.1% of chains in SATHCAP LA died immediately after seeds without generating recruits (that is, chain length = 0), meaning no chance for incorporating respondent-driven participant selection into the sample. This rate was 32.6% for SATHCAP Chicago 23.5% and 15.4% for LMSM SF and Chicago. The length varied widely across chains; for example, chains in SATHCAP Chicago lasted anywhere from 0 to 44 waves after seeds. The smallest variation was observed for LMSM Chicago with a range of 0 to 8. While this small variation in chain lengths for LMSM Chicago indicated that individual chains made similar contributions to the overall data, the relatively short maximum chain length suggested an issue for the memorylessness of the Markov chain.

4.3. Association Between Recruiter's Characteristics and the Number of Recruits

In order to further understand the recruitment process, we examined whether characteristics of potential recruiters were associated with the number of recruits they generated in Table 3. In SATHCAP LA, younger recruiters, those with lower income (less than USD 500 a month) and men who had sex with men generated more recruits than their counterparts, while in Chicago, it was IDUs and those who had ever been incarcerated that generated more recruits. In LMSM, foreign-borns in SF and those with lower income (less than USD 15,000 a year) in Chicago generated more recruits. It was notable that recruiter's network size had virtually no effect on recruitment across studies and cities. This contradicts the view by Johnston and Sabin (2010) that seeds with large and dense networks generate more recruits. Rather, socioeconomics (for example, income) and risk behaviors (for example, MSMs) of the recruiters made a difference in recruitment. Note that these characteristics were significantly related to HIV status, one of the key outcomes in these studies (results not shown), further suggesting nonresponse bias.

Table 3. Poisson regression of number of recruits on recruiters' characteristics, the Sexual Acquisition and Transmission of HIV Cooperative Agreement Program (SATHCAP) and the Latino MSM community involvement (LMSM).

| Recruiter characteristics | SATHCAP | | | LMSM | | |
|---------------------------|----------|-----------|---------|---------|---------|---------|
| | LA | Chicago | Chicago | SF | Chicago | Chicago |
| | Est. | Est. | Est. | Est. | Est. | Est. |
| Intercept | -0.551** | -0.289** | -0.183 | -0.323# | -0.183 | -0.183 |
| Age | -0.230* | -0.023 | -0.202 | 0.049 | -0.202 | -0.202 |
| Race/nativity | 0.143 | -0.051 | -0.235 | -0.531* | -0.235 | -0.235 |
| Education | 0.104 | -0.022 | -0.169 | 0.141 | -0.169 | -0.169 |
| Income | 0.290** | 0.066 | 0.423** | 0.026 | 0.423** | 0.423** |
| Living arrangement | 0.024 | 0.003 | -0.066 | 0.092 | -0.066 | -0.066 |
| HIV+ | -0.086 | -0.052 | 0.142 | 0.210 | 0.142 | 0.142 |
| Substance use | 0.080 | 0.165**** | 0.301 | 0.190 | 0.301 | 0.301 |
| Sexual behavior | 0.222* | 0.081 | -0.281 | 0.110 | -0.281 | -0.281 |
| Incarcerated | 0.040 | 0.219**** | - | - | - | - |
| Network size | 0.001 | 0.001# | 0.002 | 0.002 | 0.002 | 0.002 |

#Significant at $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4.4. Recruitment Chain Homogeneity

ICCs are reported in [Table 4](#). Recall that ICCs are evidence of correlated responses among respondents from the same recruitment chain, which further is not consistent with the equilibrium assumption. Overall, ICCs were sizable, indicating homogeneity within chain and heterogeneity between chains. Within-chain homogeneity was larger for SATHCAP than LMSM. ICC was notably large for race in SATHCAP Chicago at 0.619 and for HIV status in SATHCAP LA at 0.490, indicating that 61.9% and 49.0% of the overall variance in these variables were due to between-chain variance.

4.5. Nonresponse Follow-Up Study

In the SATHCAP follow-up study, recruiters were asked the number of coupon accepters and refusers. On average, follow-up respondents in LA reported 3.10 peers accepting and 1.60 refusing coupons; in Chicago 2.17 accepting and 1.06 refusing. If to examine any incidence of coupon being refused or accepted reported in the follow-up study, 46.2% of follow-up respondents in LA and 31.1% in Chicago reported any of their coupons being refused by the peers, while over 97% of respondents reported any of their coupons being accepted. The fact that coupon refusal was reported implies that nonresponse did arise at this stage and the true response rates were lower than the coupon redemption rates in Subsection 4.2.

Recruiters who reported any coupons being accepted or refused were asked about the characteristics of accepters and refusers separately. Their characteristics are listed in [Table 5](#). In both cities, the proportions of friends and sex partners were significantly higher among accepters than among refusers. For example, 87.6% of the coupon accepters were friends of recruiters, while 78.59% of the refusers were so in Chicago, a significant difference at $p < 0.001$. Coupon accepters in Chicago were less likely to be homeless and more likely to be IDUs, compared to coupon refusers (both at $p < 0.05$).

The follow-up study itself was subject to own nonresponse and measurement errors. As noted previously, about 53.5% of the potential recruiters participated in the follow-up. Given that this was conducted at the time of recruitment incentive payment, it is not surprising that follow-up study respondents had distributed more coupons than nonrespondents (4.40 vs. 3.15 for LA and 3.32 vs. 2.61 for Chicago) and were associated with a larger number of recruits (1.75 vs. 0.13 for LA and 1.66 vs. 0.04 for Chicago), all significant at $p < 0.001$, results similar to [Gile et al. \(2015\)](#). With logistic regression, we examined whether recruiters' characteristics beyond the number of coupons they took affected their follow-up study participation. The results in [Appendix Table A2](#) suggested that those with lower income or with HIV were more likely to participate in the follow-up study than their counterparts in LA, while it was Black recruiters who were more likely to participate in Chicago.

Additionally, the number of accepted coupons reported by the recruiters in the follow-up study matched neither the number of coupons they took nor the number of coupons redeemed by their peers in the coupon distribution data. While recruiters in the follow-up study reported 3.10 and 2.17 coupon accepters in LA and Chicago, respectively, their coupon distribution data showed that they took 4.40 and 3.32 coupons in LA and Chicago and that 1.75 and 1.66 coupons were redeemed in LA and Chicago.

Table 4. Intra – chain correlation on respondent characteristics, the Sexual Acquisition and Transmission of HIV Cooperative Agreement Program (SATHCAP) and the Latino MSM community involvement (LMSM).

| Respondent characteristics | SATHCAP | | | | LMSM | | | |
|----------------------------|----------------------------|-------|---------|----------------------------|-------|-------|---------|--|
| | LA | | Chicago | | LA | | Chicago | |
| | Est. | Est. | Est. | Est. | Est. | Est. | Est. | |
| Age | > 45 vs. ≤ 45 yrs | 0.102 | 0.117 | > 35 vs. ≤ 35 yrs | 0.117 | 0.133 | | |
| Race/nativity | Black vs. No | 0.176 | 0.619 | US Born vs. No | 0.026 | 0.236 | | |
| Education | ≤ High school vs. > HS | 0.108 | 0.017 | ≤ High school vs. > HS | 0.101 | 0.103 | | |
| Income | < \$500/mo vs. ≥ \$500/mo | 0.124 | 0.000 | < \$15K/yr vs. ≥ \$15K/yr | 0.114 | 0.099 | | |
| Living arrangement | Homeless vs. No | 0.171 | 0.064 | Live alone vs. No | 0.000 | 0.017 | | |
| HIV+ | Yes vs. No | 0.490 | 0.192 | Yes vs. No | 0.110 | 0.152 | | |
| Substance use | Injection drug ever vs. No | 0.158 | 0.199 | Substance use 6 mos vs. No | 0.141 | 0.079 | | |
| Sexual behavior | MSM vs. No | 0.114 | 0.182 | Bi/Transgender vs. No | 0.077 | 0.060 | | |
| Incarcerated | Ever vs. No | 0.134 | 0.001 | | – | – | | |

Table 5. Comparison of characteristics between coupon accepters and refusers, the Sexual Acquisition and Transmission of HIV Cooperative Agreement Program (SATHCAP).

| Characteristics | LA | | | | Chicago | | | |
|----------------------------|------------------|------------|-----------------|-------------------------|------------------|------------|-----------------|---------------------------|
| | Coupon accepters | | Coupon refusers | | Coupon accepters | | Coupon refusers | |
| | n | Mean (SE) | n | Mean (SE) | n | Mean (SE) | n | Mean (SE) |
| Friend (%) | 351 | 76.5 (1.8) | 163 | 69.3 (3.2) [#] | 1,433 | 87.6 (0.7) | 457 | 78.6 (1.6) ^{***} |
| Sex partner (%) | 349 | 36.1 (2.1) | 162 | 29.5 (3.1) [#] | 1,428 | 48.5 (1.1) | 453 | 39.8 (2.0) ^{***} |
| Known for 6+ months (%) | 349 | 70.0 (2.0) | 164 | 66.8 (3.3) | 1,430 | 89.1 (0.7) | 456 | 86.6 (1.4) |
| See daily (%) | 348 | 60.1 (2.1) | 164 | 57.7 (3.3) | 1,429 | 74.4 (1.0) | 455 | 72.0 (1.7) |
| Live in the same city (%) | 351 | 76.1 (2.1) | 164 | 74.4 (3.2) | 1,429 | 96.5 (0.4) | 452 | 95.5 (0.9) |
| Male (%) | 351 | 75.4 (1.6) | 137 | 74.6 (2.9) | 1,430 | 61.5 (1.0) | 382 | 61.2 (2.0) |
| Black (%) | 350 | 30.2 (2.0) | 164 | 26.3 (2.9) | 1,431 | 48.9 (1.3) | 456 | 49.1 (2.2) |
| Homeless (%) | 346 | 36.1 (2.3) | 160 | 35.0 (3.4) | 1,425 | 18.7 (0.9) | 447 | 23.8 (1.8) [*] |
| Injected drug together (%) | 345 | 50.3 (2.4) | 162 | 51.4 (3.5) | 1,430 | 81.4 (0.9) | 455 | 76.9 (1.7) [*] |

[#] Significantly different from coupon accepters at $p < 0.1$, ^{*} $p < 0.05$, ^{**} $p < 0.01$, ^{***} $p < 0.001$

5. Measurement Error

5.1. Reported Network Size

We examined the distribution of reported network sizes in [Table 6](#). First, in a small number of cases, networks sizes were not reported. On average, respondents reported their network sizes being in the neighborhood of 20: in SATHCAP 17.5 for LA and 21.1 for Chicago; and in LMSM 23.9 for SF and 36.7 for Chicago. The network size showed a wide variation, as small as zero and as large as 2,100; however, the median was modest at 7, 10, 10, and 11 for across cities and studies, resulting in large positive skewness. In fact, 90% of the respondents reported network sizes smaller than 50 for both cities in SATHCAP and 40 and 75 for LMSM SF and Chicago.

By default in the RDS recruitment and by the reciprocal nature of social networks, non-seed respondents should report at least one network member. This is because their recruiters considered them as a network member, and so should they. However, 43 non-seeds (5.9%) in SATHCAP LA reported zero network size, a problem reported by [McCreesh et al. \(2012\)](#). However, this zero network size reported by non-seeds occurred infrequently for SATHCAP Chicago and both cities in LMSM at 19 (0.7%), 3 (1.0%), and 2 (0.7%).

5.2. Reported Network Sizes for Weights

The maximum network size respondents reported was 400 and 591 for SATHCAP LA and Chicago and 1,000 and 2,100 for LMSM SF and Chicago. While not impossible, it is difficult to imagine a LMSM respondent in Chicago knowing exactly 2,100 Latino gay, bisexual, and transgenders over 18 years old in Chicago. This observation had an implication for inference because the inverse of network size was used as weights, which

Table 6. Distribution of reported network size, the Sexual Acquisition and Transmission of HIV Cooperative Agreement Program (SATHCAP) and the Latino MSM community involvement (LMSM).

| | SATHCAP | | LMSM | |
|--|---------|---------|-------|---------|
| | LA | Chicago | SF | Chicago |
| <i>n</i> | 845 | 2,739 | 323 | 320 |
| No. of cases with missing network size | 12 | 12 | 0 | 0 |
| Reported network size | | | | |
| Average | 17.5 | 21.1 | 23.9 | 36.7 |
| Standard deviation | 33.6 | 36.4 | 69.2 | 128.2 |
| Maximum | 400 | 591 | 1,000 | 2,100 |
| 90th percentile | 50 | 50 | 40 | 75 |
| 75th percentile | 16 | 25 | 20 | 30 |
| Median | 7 | 10 | 10 | 11 |
| 25th percentile | 3 | 6 | 5 | 5 |
| 10th percentile | 2 | 3 | 3 | 3 |
| Minimum | 0 | 0 | 0 | 0 |
| Mode | 2 | 10 | 2 | 10 |
| No. of non-seeds with 0 network size | 43 | 19 | 3 | 2 |

ranged from 0.0005 ($= 1/2,100$) to 1.0000 ($= 1/1$) in LSM Chicago. Weight dispersion resulted in UWEs of 2.19 and 2.29 for SATHCAP LA and Chicago and 2.21 and 2.57 for LSM SF and Chicago. Weight smoothing reduced UWEs substantially to 1.73, 1.62, 1.61, and 1.76 for respective study and city. This is not surprising given that the weights in LSM Chicago, for example, varied from 0.0005 to 1.000 without smoothing but in a smaller range, from 0.0133 ($= 1/75$) to 0.3333 ($= 1/3$), with smoothing. In Appendix Table A3, we examined respondents' characteristics associated with both weights. Overall, weights were related to certain respondent characteristics, and this relationship persisted regardless of weight smoothing.

5.3. Effects of Weights in Estimation

We used three types of estimation approaches: 1) unweighted; 2) weighted with standard weights; and 3) weighted with smoothed weights. The focus of this section is on whether the weights affected estimates and their variabilities or significance. Table 7A includes estimated proportions for various sociodemographic and health risk variables and their standard errors. Table 7B includes estimates of coefficients in simple logistic regression of HIV status and their p -values.

In Table 7A, weights changed univariate statistics. Estimates affected the most by weights were characteristics that were significantly related to weights in Appendix Table A3. For instance, incarceration, a significant covariate of weights in SATHCAP LA, changed from 66.4% (unweighted) to 61.6% (with standard weights), and to 61.1% (with smoothed weights). For LSM, HIV status was a significant covariate of weights in Chicago, and the HIV+ rate decreased from 21.5% (unweighted) to 17.1% (with smoothed weights), and to 14.3% (with standard weights). Not surprisingly, with weights, standard errors increased by a factor of 1.5 to 2.

Weights affected logistic regression coefficients in Table 7B both substantively and statistically. Risk factors known to be highly related to HIV status showed mixed results depending on the estimation approaches. For example, injection drug use was estimated to be significant in SATHCAP Chicago regardless of weights. However, in SATHCAP LA, it was not significant when unweighted or with smoothed weights but marginally significant with standard weights in the direction opposite of what one would expect: injection drug use was negatively related to HIV+. In LSM, regardless of the estimation approaches, substance use was not a significant predictor in SF, but was a significant predictor in Chicago without weights or smoothed weights. MSM in SATHCAP was a significant predictor regardless of weights in both cities. Whether someone had STD in LSM was a significant predictor of HIV+ for Chicago consistently across approaches, but was not so in SF when applying standard weights. Significance of network size in SATHCAP varied depending on the approaches. In LSM, network size was a consistently significant predictor in Chicago, but insignificant in SF.

6. Discussion

Our study showed 1) that there existed nonresponse and measurement errors pertinent to the assumptions and practices of RDS; 2) that these errors had implications for inferences; and 3) that this was observed commonly in two independent RDS studies.

Table 7. Comparison of unweighted and weighted estimates using standard and smoothed weights, the Sexual Acquisition and Transmission of HIV Cooperative Agreement Program (SATHCAP) and the Latino MSM community involvement (LMMSM).

A. Univariate statistics: proportions.

| Characteristics | SATHCAP: LA | | | | | | SATHCAP: Chicago | | | | | |
|-------------------|----------------------|--------|------------------------------|--------|------------------------------|--------|------------------------|--------|--------------------------------|--------|--------------------------------|--------|
| | Unweighted (n = 845) | | Weighted, standard (n = 802) | | Weighted, smoothed (n = 845) | | Unweighted (n = 2,739) | | Weighted, standard (n = 2,720) | | Weighted, smoothed (n = 2,739) | |
| | % | SE (%) | % | SE (%) | % | SE (%) | % | SE (%) | % | SE (%) | % | SE (%) |
| Age <45 yrs old | 54.3 | 1.7 | 51.7 | 3.1 | 50.8 | 2.8 | 47.1 | 1.0 | 48.2 | 1.7 | 47.1 | 1.5 |
| Race: Black | 48.4 | 1.7 | 46.0 | 3.5 | 47.4 | 3.5 | 81.5 | 0.7 | 79.4 | 1.7 | 79.8 | 1.5 |
| ≤ High school | 59.5 | 1.7 | 64.9 | 3.1 | 63.0 | 2.6 | 72.6 | 0.9 | 75.7 | 1.4 | 75.4 | 1.1 |
| Homeless | 56.2 | 1.7 | 52.5 | 3.3 | 53.8 | 2.9 | 38.6 | 0.9 | 36.1 | 1.4 | 36.1 | 1.3 |
| Income <\$500/ mo | 61.9 | 1.7 | 62.8 | 3.0 | 63.1 | 2.6 | 70.3 | 0.9 | 71.8 | 1.4 | 71.2 | 1.1 |
| HIV+ | 30.1 | 1.6 | 29.3 | 3.7 | 28.1 | 2.8 | 9.0 | 0.5 | 8.0 | 0.8 | 8.8 | 0.8 |
| MSM | 59.9 | 1.7 | 56.3 | 3.2 | 55.2 | 2.9 | 21.0 | 0.8 | 19.2 | 1.4 | 19.8 | 1.2 |
| Inject drug ever | 47.7 | 1.7 | 42.0 | 3.4 | 42.3 | 2.7 | 41.3 | 0.9 | 39.9 | 1.7 | 40.8 | 1.4 |
| Ever prison | 66.4 | 1.6 | 61.6 | 2.4 | 61.1 | 2.1 | 75.3 | 0.8 | 72.6 | 1.2 | 72.7 | 1.0 |
| | | | | | | | | | | | | |
| Characteristics | LMMS: SF | | | | | | LMMS: Chicago | | | | | |
| | Unweighted (n = 323) | | Weighted, standard (n = 320) | | Weighted, smoothed (n = 323) | | Unweighted (n = 320) | | Weighted, standard (n = 318) | | Weighted, smoothed (n = 320) | |
| | % | SE (%) | % | SE (%) | % | SE (%) | % | SE (%) | % | SE (%) | % | SE (%) |
| Age <35 yrs old | 41.2 | 2.7 | 42.4 | 4.9 | 40.7 | 4.7 | 60.0 | 2.7 | 57.6 | 5.1 | 58.0 | 5.0 |
| US Born | 14.2 | 1.9 | 19.7 | 2.7 | 16.8 | 2.5 | 30.9 | 2.6 | 30.5 | 5.0 | 32.0 | 4.6 |
| ≤ High school | 47.1 | 2.8 | 48.9 | 4.4 | 49.7 | 4.0 | 52.8 | 2.8 | 54.4 | 5.1 | 54.9 | 4.5 |
| Income <\$15K/yr | 62.2 | 2.7 | 67.6 | 4.1 | 65.7 | 3.5 | 50.0 | 2.8 | 52.3 | 4.6 | 51.1 | 4.2 |
| Live alone | 22.6 | 2.3 | 19.1 | 3.3 | 20.3 | 2.9 | 26.3 | 2.5 | 23.3 | 4.1 | 23.8 | 3.8 |
| HIV+ | 38.0 | 2.8 | 37.3 | 4.9 | 37.0 | 4.2 | 21.5 | 2.5 | 14.3 | 6.2 | 17.1 | 4.6 |
| Bi/Transgender | 33.5 | 2.6 | 40.5 | 4.4 | 39.6 | 4.3 | 27.2 | 2.5 | 31.4 | 4.6 | 32.7 | 4.0 |
| STD | 9.9 | 1.7 | 7.8 | 1.7 | 8.7 | 1.8 | 13.8 | 1.9 | 14.2 | 3.7 | 13.3 | 3.0 |
| Substance 6 mos | 42.7 | 2.8 | 40.8 | 3.9 | 40.8 | 3.7 | 52.8 | 2.8 | 56.0 | 4.8 | 54.9 | 4.0 |

B. Bivariate statistics: coefficients of simple logistic regression (dependent variable: HIV+).

| Independent variable | SATHCAP: LA | | | | | | SATHCAP: Chicago | | | | | |
|------------------------|-------------|---------------|--------------------|---------------|--------------------|---------------|------------------|---------------|--------------------|---------------|--------------------|---------------|
| | Unweighted | | Weighted, standard | | Weighted, smoothed | | Unweighted | | Weighted, standard | | Weighted, smoothed | |
| | Est. | <i>p-val.</i> | Est. | <i>p-val.</i> | Est. | <i>p-val.</i> | Est. | <i>p-val.</i> | Est. | <i>p-val.</i> | Est. | <i>p-val.</i> |
| Inject drug vs. No | 0.065 | 0.668 | -0.385 | 0.084 | -0.261 | 0.187 | 0.416 | 0.002 | 0.503 | 0.007 | 0.393 | 0.021 |
| MSM vs. No | 1.426 | <.001 | 1.143 | <.001 | 1.436 | <.001 | 1.442 | <.001 | 1.692 | <.001 | 1.611 | <.001 |
| Network size | 0.004 | 0.081 | 0.005 | 0.404 | 0.006 | 0.037 | 0.001 | 0.436 | 0.007 | 0.031 | 0.002 | 0.191 |
| | LMMS: SF | | | | | | LMMS: Chicago | | | | | |
| | Unweighted | | Weighted, standard | | Weighted, smoothed | | Unweighted | | Weighted, standard | | Weighted, smoothed | |
| Independent variable | Est. | <i>p-val.</i> | Est. | <i>p-val.</i> | Est. | <i>p-val.</i> | Est. | <i>p-val.</i> | Est. | <i>p-val.</i> | Est. | <i>p-val.</i> |
| Substance 6 mos vs. No | 0.225 | 0.350 | 0.083 | 0.817 | 0.187 | 0.539 | 0.790 | 0.011 | 0.624 | 0.199 | 0.741 | 0.088 |
| STD vs. No | 0.651 | 0.009 | 0.607 | 0.144 | 0.967 | 0.003 | 1.432 | <.0001 | 1.547 | 0.006 | 1.717 | 0.001 |
| Network size | -0.002 | 0.280 | 0.001 | 0.847 | -0.001 | 0.546 | 0.006 | 0.015 | 0.017 | 0.023 | 0.007 | 0.032 |

6.1. Summary

Nonresponse in the recruitment process impacted not only the sample size growth but also the recruitment chain length. The assumed exponential growth was far from the reality, and a substantial proportion of chains died immediately after seeds. Moreover, coupon distribution data as well as follow-up data suggested that nonresponse did not occur at random. First, closeness of the relationship between participants and their peers influenced peers' coupon acceptance. Proportions of friends and sex partners were significantly larger among coupon accepters than among refusers by about ten percent points. Second, participants with certain characteristics, most notably lower income, generated more recruits than the counterpart. This systematic nonresponse, nonrandom recruitment pattern and unequal chain length, when combined with large ICCs, further suggest that the Markov chain is not achieved in the practice of RDS.

Self-reported network sizes showed a wide variation with some unrealistic extreme values, strong evidence for measurement error. This measurement error is of concern on its own, of course. In RDS, this is also of concern for inference: as the inverse of network sizes is used as weights, the accuracy of the report matters. In particular, their variability means variability in weights, which, in turn, decreases efficiency of estimates shown through UWE that ranged around two in our analysis. Weights changed estimated prevalence in directions that were not entirely explainable. It is true that whether a person reports 2,100 or 2,150 for the network size has a little effect on the weight assigned to this person, with both resulting in a weight of 0.0005. However, whether a person reports 1 versus 50 does have an effect on the weight, with the weight being 1 versus 0.02. Moreover, in principle, while the current RDS estimators of prevalence attempt to take a form of model-based estimation (Valliant 2013), the information used in the estimation is subject to measurement error, making the estimators inadequate to account for such an error. These may hamper inferences in an unknown direction.

One may argue that the purpose of RDS is to study relationships between variables, not to estimate prevalence. Our analysis of simple models that regressed HIV status on various characteristics with different applications of weights (for example, unweighted, standard weights, smoothed weights) showed unexplainable patterns. For instance, in SATHCAP, injection drug use was a significant and positive predictor of HIV status regardless of weights in Chicago; but in LA, it was not a significant predictor without weights and with smoothed weights, and was a marginally significant and *negative* predictor with standard weights. While one may suspect that applying weights, particularly standard weights, would decrease the significance of covariates due to increased variability of their estimates, this was not always the case. Overall, weights did affect the inferences about bivariate relationships, but in a yet unexplainable and, in some cases, unreasonable way.

Moreover, participants' network sizes played no role in their recruitment success (Table 3), providing no support for the theoretical rationale of using them as weights in RDS to account for unequal selection probabilities. The significant relationships between the weights and respondent sociodemographic characteristics (for example, younger age), as well as outcome variables (for example, HIV+) in Appendix Table A3 make it very difficult to understand what these weights are adjusting for.

While errors examined in this article are important for understanding the key assumptions in RDS that further affect sampling productivity and inference, they are considered in neither the data collection nor the inferences. It is true that there is no practical and clear solution for sampling rare, hidden, and elusive populations. However, with obvious violations of these assumptions shown in this study, it is questionable how long the lack of practical solutions for sampling rare populations can be used as a justification of practicing RDS without improving design features that may minimize the effects of these breakdowns or accounting for them.

Undoubtedly, this study is limited in a number of ways. First, it addressed only two of four components of TSE, because existing RDS data do not provide information about remaining errors. Even with the two errors examined in this study, the breadth of examination was bounded by data availability. While the implications of nonresponse patterns and the effect of potential measurement error in network sizes on inferences were consistent between the two data sources, they may be specific only to these two studies. It should, however, be noted that the majority of methodological studies of RDS rely on a single data source or data that are not publicly available (for example, [Wejnert and Heckathorn, 2008](#); [McCreesh et al. 2012](#); [Gile et al. 2015](#)), making replication difficult, if not impossible. Rather than using the findings from this article against RDS, it would be productive to take them to develop a new framework for evaluating and improving RDS data collection practices and inferences.

6.2. Open Questions

With RDS, until a clear guidance is developed for assessing errors in RDS and for improving inferences, we may run the risk of mischaracterizing the hidden, rare and elusive populations, unintentionally negatively impacting these groups. In this section, we pose questions about diagnostics and estimations that may be considered in improving RDS.

On diagnostics, a recent study by [Gile and her colleagues \(2015\)](#) provides a set of approaches for examining RDS assumptions. As one of the first focusing on diagnostics, their study is innovative. However, their approaches rely heavily on follow-up interviews, which our study found not free from own nonresponse and measurement errors. For example, participants who did not recruit their peers were less likely to participate in the follow-up study than the counterpart. Questions used in their follow-up study were difficult to answer. For example, a question, “How many people did you try to give a coupon but they had already participated in the study?” was used to study failed recruitment attempts. This question assumes that participants are familiar with the recruitment status of their peers and/or are able to recall the number of own recruitment attempts. Other questions in their study include, “How many people do you know who have used illegal drugs in the past three months?”, “If we were to give you as many coupons as you wanted, how many of these drug users do you think you could give a coupon to by this time tomorrow?” and “What is the principal reason why these persons did not accept a coupon?” Undoubtedly, these questions are difficult as respondents simply may not have information for them (for example, peers’ illegal drug use). Data from such questions are not free from measurement error and may not provide meaningful information for understanding the recruitment process.

While follow-up interviews are a logical and attractive option for studying nonresponse error, their design cannot be taken lightly with respect to the types of questions and the timing of the follow-up interview. It would be advantageous to consider questions that provide meaningful data for investigating nonresponse, yet with little room for measurement error. For the timing, it would be ideal to pick a time that is reasonably long after the main interview so that all recruitment efforts can be captured, yet reasonably short so that recall does not become overly demanding. While the follow-up study in [Gile et al. \(2015\)](#) was conducted within one week after the main interview, our analysis of SATHCAP and LMSM showed that the average time gap between participants' main interview and their peers' interview was 14 to 46 days. Timing of the follow-up study should be informed either by the time gap observed in the field or by recruitment protocol designs (for example, assigning expiration dates to the coupons and conducting follow-up studies shortly after the expiration).

It would be ideal to account for these errors in inference. Despite the systematic nature of nonresponse examined in this article, there were no variables that explained nonresponse commonly across cities and studies. Hence, more organized efforts should be made to understand this mechanism. While the idea of accounting for unequal selection probabilities through weighting by network sizes, their measurement error needs to be addressed. One may consider using estimated network sizes through appropriate models, such as variants of the Fay-Herriot model ([Fay and Herriot 1979](#)) or those used by [Beaumont \(2008\)](#). Additionally, for increasing accuracy in network size measurements, the scale-up method for estimating network structures through specific questions ([McCarthy et al. 2001](#); [Zheng et al. 2006](#)) may serve as a reasonable approach.

Appendix

Table A1. Publicly available respondent driven sampling data sets through Interuniversity Consortium for Political and Social Research (ICPSR)[§] as of January, 2016.

| Study name (ICPSR study number) | Year of data collection | Year of data release | Coupon information |
|---|--|----------------------|--------------------|
| 1. Information on Artists (ICPSR 35585) | 1989, 1997, 2004, 2006–2007, 2009–2010, 2001 | 2015 | Not available |
| 2. Study of Jazz Artists [United States] (ICPSR 35593) | | 2015 | Not available |
| 3. Latino MSM Community Involvement: HIV Protective Effects (ICPSR 34385) | 2003–2004 | 2013 | Available |
| 4. Sexual Acquisition and Transmission of HIV Cooperative Agreement Program (SATHCAP) [United States] (ICPSR 29181) | 2005–2006, 2006–2008 | 2010 | Available |
| 5. The Commercial Sexual Exploitation of Children in New York City, 1982–2007 (ICPSR 34657) | 2006–2007 | 2015 | Available |
| 6. Dynamics of Retail Methamphetamine Markets in New York City (ICPSR 29821) | 2007–2009 | 2014 | Not available |
| 7. Health Consequences of Long-Term Injection Heroin Use Among Aging Mexican American Men in Houston, Texas (ICPSR 34896) | 2008–2011 | 2014 | Not available |
| 8. Social Justice Sexuality Project: 2010 National Survey, including Puerto Rico (ICPSR 34363) | 2010 | 2013 | Not available |

[§] Interuniversity Consortium for Political and Social Research (ICPSR) is a major data archive for social science research (<https://www.icpsr.umich.edu/icpsrweb/landing.jsp>). To our best knowledge, there are no other publicly available data using RDS located outside of ICPSR.

Table A2. Logistic regression of follow-up study participation on respondent characteristics, the Sexual Acquisition and Transmission of HIV Cooperative Agreement Program (SATHCAP).

| Respondent characteristics | | LA | Chicago |
|----------------------------|----------------------------|-----------------------|-----------------------|
| | | Est. | Est. |
| Intercept | | -2.133 ^{***} | -1.883 ^{***} |
| Age | > 45 vs. ≤ 45 yrs | -0.010 | -0.069 |
| Race/nativity | Black vs. No | 0.190 | 0.365 [*] |
| Education | ≤ High school vs. > HS | -0.143 | 0.054 |
| Income | < \$500/mo vs. ≥ \$500/mo | 0.384 [#] | 0.120 |
| Living arrangement | Homeless vs. No | 0.078 | 0.023 |
| HIV+ | Yes vs. No | 0.735 ^{***} | 0.231 |
| Substance use | Injection drug ever vs. No | -0.024 | 0.133 |
| Sexual behavior | MSM vs. No | -0.140 | -0.024 |
| Incarcerated | Ever vs. No | -0.068 | -0.006 |
| Network size | | 0.004 | 0.001 |
| No. coupons | | 0.385 ^{***} | 0.555 ^{***} |

[#] Significant at $p < 0.1$, ^{*} $p < 0.05$, ^{**} $p < 0.01$, ^{***} $p < 0.001$

Table A3. Linear regression of log transformed standard and smoothed weights on respondent characteristics, the Sexual Acquisition and Transmission of HIV Cooperative Agreement Program (SATHCAP) and the Latino MSM Community Involvement (LMSM).
A. Standard and smoothed weights in SATHCAP.

| Respondent Characteristics | Dependent variable: standard weight | | Dependent variable: smoothed weight | |
|----------------------------|--|-----------|--|-----------|
| | LA | Chicago | LA | Chicago |
| | Est. | Est. | Est. | Est. |
| Intercept | -1.605*** | -1.995*** | -1.633*** | -2.045*** |
| Age | -0.411*** | -0.069 | -0.365*** | -0.072# |
| Race/nativity | | | | |
| > 45 vs. ≤45 yrs | | | | |
| Black vs. No | -0.220* | -0.322*** | -0.201* | -0.279*** |
| Education | | | | |
| ≤High school vs. > HS | 0.246* | 0.137** | 0.210* | 0.133** |
| Income | | | | |
| <\$500/mo vs. ≥\$500/mo | 0.251* | 0.059 | 0.228* | 0.054 |
| Living arrangement | | | | |
| Homeless vs. No | -0.164 | -0.144** | -0.120 | -0.126** |
| HIV+ | | | | |
| Yes vs. No | 0.071 | 0.014 | 0.081 | 0.046 |
| Substance use | | | | |
| Injection drug ever vs. No | -0.228* | -0.068 | -0.200* | -0.054 |
| Sexual behavior | | | | |
| MSM vs. No | -0.271* | -0.006 | -0.263** | -0.004 |
| Incarcerated | | | | |
| Ever vs. No | -0.426*** | -0.201*** | -0.371*** | -0.176*** |
| No. distributed coupons | 0.039 | -0.028 | 0.029 | -0.026 |

Significant at $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

B. Standard and smoothed weights in LMSM.

| Respondent characteristics | Dependent variable: standard weight | | Dependent variable: smoothed weight | |
|----------------------------|--|-----------------------|--|-----------------------|
| | SF | Chicago | SF | Chicago |
| | Est. | Est. | Est. | Est. |
| Intercept | -2.768 ^{***} | -1.720 ^{***} | -2.706 ^{***} | -1.704 ^{***} |
| Age | -0.051 | -0.347 [*] | -0.072 | -0.331 [*] |
| Race/nativity | 0.290 | 0.128 | 0.201 | 0.067 |
| Education | 0.095 | 0.292 [#] | 0.061 | 0.254 [#] |
| Income | 0.203 | -0.051 | 0.177 | -0.036 |
| Living arrangement | -0.213 | -0.120 | -0.114 | -0.081 |
| HIV+ | 0.057 | -0.768 ^{***} | 0.017 | -0.628 ^{***} |
| Substance use | -0.031 | -0.134 | 0.016 | -0.138 |
| Sexual behavior | 0.157 | 0.165 | 0.139 | 0.142 |
| No. distributed coupons | 0.044 | -0.248 | 0.051 | -0.247 [#] |

[#]Significant at $p < 0.1$, ^{*} $p < 0.05$, ^{**} $p < 0.01$, ^{***} $p < 0.001$

7. References

- Baker, R., J.M. Brick, N.A. Bates, M. Battaglia, M.P. Couper, J.A. Dever, K.J. Gile, and R. Tourangeau. 2013. "Summary Report of the AAPOR Task Force on Non-Probability Sampling." *Journal of Survey Statistics and Methodology* 1(2): 90–143. Doi: <https://doi.org/10.1093/jssam/smt008>.
- Beaumont, J.-F. 2008. "A New Approach to Weighting and Inference in Sample Surveys." *Biometrika* 95(3): 539–553. Doi: <https://doi.org/10.1093/biomet/asn028>.
- Burt, R.D., H. Hagan, K. Sabin, and H. Thiede. 2010. "Evaluating Respondent-Driven Sampling in a Major Metropolitan Area: Comparing Injection Drug Users in the 2005 Seattle Area National HIV Behavioral Surveillance System Survey with Participants in the RACEN and Kiwi Studies." *Annals of Epidemiology* 20(2): 159–167. Doi: <https://doi.org/10.1016/j.annepidem.2009.10.002>.
- Centers for Disease Control and Prevention (CDC). 2009. HIV-Associated Behaviors Among Injecting-Drug Users—23 Cities, United States, May 2005–February 2006. *Morbidity and Mortality Weekly Report*, 58, 329–332. Available at: <http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5813a1.htm> (accessed September 2015).
- Centers for Disease Control and Prevention (CDC). 2013. *National HIV Behavioral Surveillance System Round 4: Model Surveillance Protocol*. Available at: http://www.cdc.gov/hiv/pdf/NHBS_Round4ModelSurveillanceProtocol.pdf (accessed September 2015).
- Cole, S.R. and M.A. Hernan. 2008. "Constructing Inverse Probability Weights for Marginal Structural Models." *American Journal of Epidemiology* 168: 656–664. Doi: <https://doi.org/10.1093/aje/kwn164>.
- Compton, W., J. Normand, and E. Lambert. 2009. "Sexual Acquisition and Transmission of HIV Cooperative Agreement Program (SATHCAP)." *Journal of Urban Health* 86(1): 1–4. Doi: <https://doi.org/10.1007/s11524-009-9373-4>.
- Constantine, M. 2010. "Disentangling Methodologies: The Ethics of Traditional Sampling Methodologies, Community-Based Participatory Research, and Respondent-Drive Sampling." *American Journal of Bioethics* 10(3): 22–24. Doi: <https://doi.org/10.1080/15265160903585628>.
- Dombrowski, R., B. Khan, J. Moses, E. Channell, and E. Misshula. 2013. "Assessing Respondent Driven Sampling for Network Studies in Ethnographic Contexts." *Advances in Anthropology* 3(1): 1–9. Doi: <https://doi.org/10.4236/aa.2013.31001>.
- Elliott, M.R. 2009. "Model Averaging Methods for Weight Trimming in Generalized Linear Regression Models." *Journal of Official Statistics* 25(1): 1–21. Doi: <https://doi.org/10.1.1.552.9050>.
- Fay, R.E. and R.A. Herriot. 1979. "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data." *Journal of American Statistical Association* 74: 269–277. Doi: <https://doi.org/10.2307/2286322>.
- Frost, S.D.W., K.C. Brouwer, M.A.F. Cruz, R. Ramos, M.E. Ramos, R.M. Lozada, C. Magis-Rodriguez, and S.A. Strathdee. 2006. "Respondent-Driven Sampling of Injection Drug Users in Two U.S.-Mexico Border Cities: Recruitment Dynamics and Impact on Estimates of HIV and Syphilis Prevalence." *Journal of Urban Health* 83(1): 83–97. Doi: <https://doi.org/10.1007/s11524-006-9104-z>.

- Gile, K.J. 2011. "Improved Inference for Respondent-Driven Sampling Data with Application to HIV Prevalence Estimation." *Journal of American Statistical Association* 106(493): 135–146. Doi: <https://doi.org/10.1198/jasa.2011.ap09475>.
- Gile, K.J. and M.S. Handcock. 2010. "Respondent-Driven Sampling: An Assessment of Current Methodology." *Sociological Methodology* 40(1): 286–327. Doi: <https://doi.org/10.1111/j.1467-9531.2010.01223.x>.
- Gile, K.J., L.G. Johnston, and M.J. Salganik. 2015. "Diagnostics for Respondent-Driven Sampling." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 178(1): 241–269. Doi: <https://doi.org/10.1111/rssa.12059>.
- Goel, S. and M.J. Salganik. 2010. "Assessing Respondent-Driven Sampling." *Proceedings of the National Academy of Sciences of the United States of America* 107(15): 6743–6747. Doi: <https://doi.org/10.1073/pnas.1000261107>. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2872407/> (accessed September 2015).
- Groves, R.M. 1989. *Survey Errors and Survey Costs*. New York: Wiley.
- Groves, R.M. and M.P. Couper. 1998. *Nonresponse in Household Surveys*. New York: Wiley.
- Handcock, M.S. 2012. *Estimating the Size of Hard-to-Reach Populations Using Respondent-Driven Sampling Data*. Paper for the International Conference on Methods for Surveying and Enumerating Hard-to-Reach Populations, October 31–November 3, New Orleans, LA.
- Handcock, M.S., K.J. Gile, I.E. Fellows, and W.W. Neeley. 2014. *Package 'RDS.'* Available at: <http://cran.r-project.org/web/packages/RDS/RDS.pdf> (accessed September 2015).
- Heckathorn, D.D. 1997. "Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations." *Society for the Study of Social Problems* 44(2): 174–199. Doi: <https://doi.org/10.2307/3096941>.
- Heckathorn, D.D. 2002. "Respondent-Driven Sampling II: Deriving Valid Population Estimates from Chain-Referral Samples of Hidden Populations." *Social Problems* 49(1): 11–34. Doi: <https://doi.org/10.1525/sp.2002.49.1.11>.
- Heckathorn, D.D. and J. Jeffri. 2001. "Finding the Beat: Using Respondent-Driven Sampling to Study Jazz Musicians." *Poetics* 28(4): 307–329. Doi: [https://doi.org/10.1016/S0304-422X\(01\)80006-1](https://doi.org/10.1016/S0304-422X(01)80006-1).
- Heimer, R. 2005. "Critical Issues and Further Questions About Respondent-Driven Sampling: Comment on Ramierz-Valles et al. (2005)." *AIDS and Behavior* 9(4): 403–408. Doi: <https://doi.org/10.1007/s10461-005-9030-1>.
- Iguchi, M.Y., S.H. Berry, A.J. Ober, T. Fain, D.D. Heckathorn, P.M. Gorbach, R. Heimer, A. Kozlov, L.J. Ouellet, S. Shoptaw, and W. Zule. 2010. *Sexual Acquisition and Transmission of HIV Cooperative Agreement Program (SATHCAP) 2006–2008 [United States]*. ICPSR29181-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research. Doi: <https://doi.org/10.3886/ICPSR29181>.
- Iguchi, M.Y., A.J. Ober, S.H. Berry, T. Fain, D.D. Heckathorn, P.M. Gorbach, R. Heimer, A. Kozlov, L.J. Ouellet, S. Shoptaw, and W.A. Zule. 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." *Journal of Urban Health* 88(1): 5–31. Doi: <https://doi.org/10.1007/s11524-009-9365-4>.

- Izrael, D., M. Battaglia, and M. Frankel. 2009. "Extreme Survey Weight Adjustment as a Component of Sample Balancing (a.k.a. Raking)." *Proceedings from the 2009 SAS Global Forum*. Cary, NC: SAS Institute. Available at: <http://abttassociates.com/AbtAssociates/files/c1/c1bc376c-1931-4721-b71c-cb823a0fe809.pdf> (accessed January 2017).
- Johnston, L.G. and K. Sabin. 2010. "Sampling Hard-to-Reach Populations with Respondent Driven Sampling." *Methodological Innovations Online* 5(2): 38–48. Doi: <https://doi.org/10.4256/mio.2010.0017>.
- Kish, L. 1992. "Weighting for unequal Pi." *Journal of Official Statistics* 8(2): 183–200.
- Lansky, A., A. Abdul-Quader, M. Cribbin, T. Hall, T.J. Finlayson, R.S. Garfein, L.S. Lin, and P.S. Sullivan. 2007. "Developing an HIV Behavioral Surveillance System for Injecting Drug Users: The National HIV Behavioral Surveillance System." *Public Health Reports* 122: 48–55. Doi: <https://doi.org/10.1177/00333549071220S108>.
- Laumann, E.O., P.V. Marsden, and D. Prensky. 1983. "The Boundary Specification Problem in Network Analysis." In *Applied Network Analysis. A Methodological Introduction*, edited by R.S. Burt and M.J. Minor. 18–34. Beverly Hills, CA: Sage.
- Lee, S. 2009. "Understanding Respondent Driven Sampling from a Total Survey Error Perspective." *Survey Practice*. Available at: <http://www.surveypractice.org/index.php/SurveyPractice/article/view/187/html> (accessed September 2015).
- Lee, R., J. Ranaldi, M. Cummings, J.N. Crucetti, H. Stratton, and L.-A. McNutt. 2011. "Given the Increasing Bias in Random Digit Dial Sampling, Could Respondent-Driven Sampling be a Practical Alternative?" *Annals of Epidemiology* 21(4): 272–279. Doi: <https://doi.org/10.1016/j.annepidem.2010.11.018>.
- Lee, S., Z.T. Suzer-Gurtekin, J. Wagner, and R. Valliant. 2012. *Exploring Error Properties of Respondent Driven Sampling*. Paper presented at the Joint Statistical Meeting, July 28–August 2, San Diego, CA.
- Lin, L., T. Finlayson, R. Iachan, M.C.B. Mendoza, and C. Wejnert. 2013. "Sampling Designs for Populations at High Risk for HIV." Paper presented at the Joint Statistical Meeting, August 3–August 8, Montréal, Canada.
- Little, R.J.A., S. Lewitzky, S. Heeringa, J. Lepkowski, and R.C. Kessler. 1997. "Assessment of Weighting Methodology of the National Comorbidity Survey." *American Journal of Epidemiology* 146(5): 439–449. Doi: <https://doi.org/10.1093/oxfordjournals.aje.a009297>.
- Lu, X., L. Bengtsson, T. Britton, M. Camitz, B.J. Kim, A. Thorson, and F. Liljeros. 2012. "The Sensitivity of Respondent-Driven Sampling." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 175(1): 1–26. Doi: <https://doi.org/10.1111/j.1467-985X.2011.00711.x>.
- Marsden, P.V. 1990. "Network Data and Measurement." *Annual Review Sociology* 16: 435–463. Doi: <https://doi.org/10.1146/annurev.so.16.080190.002251>.
- Martin, J.L., J. Wiley, and D. Osmond. 2003. "Social Networks and Unobserved Heterogeneity in Risk for AIDS." *Population Research and Policy Review* 22(1): 65–90. Doi: <https://doi.org/10.1023/A:1023509211339>.
- McCarty, C., P.D. Killworth, H.R. Bernard, E.C. Johnsen, and G.A. Shelley. 2001. "Comparing Two Methods for Estimating Network Size." *Human Organization* 60: 28–39. Doi: <https://doi.org/10.17730/humo.60.1.efx5t9gjtgmga73y>.

- McCreesh, N., S.D. Frost, J. Seeley, J. Katongole, M.N. Tarsh, R. Ndunguse, F. Jichi, N.L. Lunel, D. Maher, L.G. Johnston, P. Sonnenberg, A.J. Copas, R.J. Hayes, and R.G. White. 2012. "Evaluation of Respondent-Driven Sampling." *Epidemiology* 23(1): 138–147. Doi: <https://doi.org/10.1097/EDE.0b013e31823ac17c>.
- Montealegre, J.R., J.M. Risser, B.J. Selwyn, S.A. McCurdy, and K. Sabin. 2013. "Effectiveness of Respondent Driven Sampling to Recruit Undocumented Central American Immigrant Women in Houston, Texas for an HIV Behavioral Survey." *AIDS and Behavior* 17(2): 719–727. Doi: <https://doi.org/10.1007/s10461-012-0306-y>.
- Phillips, T. 2010. "Protecting the Subject: PDR and the Potential for Compromised Consent." *American Journal of Bioethics* 10(3): 14–15. Doi: <https://doi.org/10.1080/15265160903585602>.
- Potter, F. 1988. "Survey of Procedures to Control Extreme Sampling Weights." *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 453–458. Available at: http://www.websm.org/uploadi/editor/1368363852Potter_1988_Survey_of_procedures_to_control_extreme_sampling_weights.pdf (accessed January 2017).
- Ramirez-Valles, J. 2013. *Latino MSM Community Involvement: HIV Protective Effects*. ICPSR34385-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research. Doi: <https://doi.org/10.3886/ICPSR34385.v1>.
- Ramirez-Valles, J., D.D. Heckathorn, R. Vázquez, R.M. Diaz, and R.T. Campbell. 2005. "From Networks to Populations: the Development and Application of Respondent-Driven Sampling Among IDUs and Latino Gay Men." *AIDS and Behavior* 9(4): 387–402. Doi: <https://doi.org/10.1007/s10461-005-9012-3>.
- Salganik, M.J. 2006. "Variance Estimation, Design Effects and Sample Size Calculations for Respondent Driven Sampling." *Journal of Urban Health* 83(7): 98–112. Doi: <https://doi.org/10.1007/s11524-006-9106-x>.
- Salganik, M. 2012. "Commentary: Respondent-Driven Sampling in the Real World." *Epidemiology* 23(1): 148–150. Doi: <https://doi.org/10.1097/EDE.0b013e31823b6979>.
- Salganik, M.J. and D.D. Heckathorn. 2004. "Sampling and Estimation in Hidden Populations Using Respondent-Driven Sampling." *Sociological Methodology* 34: 193–239. Doi: <https://doi.org/10.1111/j.0081-1750.2004.00152.x>.
- Schonlau, M. 2014. "Recruiting an Internet Panel Using Respondent Driven Sampling." *Journal of Official Statistics* 30(2): 291–310. Doi: <https://doi.org/10.2478/jos-2014-0018>.
- Simon, C. and M. Mosavel. 2010. "Community Members as Recruiters of Human Subjects: Ethical Considerations." *American Journal of Bioethics* 10(3): 3–11. Doi: <https://doi.org/10.1080/15265160903585578>.
- Sirken, M.G. 1972. "Stratified Sample Surveys with Multiplicity." *Journal of American Statistical Association* 67: 224–227. Doi: <https://doi.org/10.1080/01621459.1972.10481236>.
- Sirken, M.G. 1975. "Network Surveys of Rare and Sensitive Conditions." *Advances in Health Survey Research Methods, NCHSR Research Proceedings* 31. Hyattsville, MD: National Center Health Statistics.
- Sirken, M.G. 1997. "Network Sampling." In *Encyclopedia of Biostatistics*, edited by P. Armitage and T. Colton, 2977–2986. Hoboken, NJ: Wiley & Sons.

- Valliant, R. 2013. "Comment." *Journal of Survey Statistics and Methodology* 1(2): 105–111. Doi: <https://doi.org/10.1093/jssam/smt010>.
- Valliant, R., J.A. Dever, and F. Kreuter. 2013. *Practical Tools for Designing and Weighting Survey Samples*. New York: Springer.
- Verdery, A.M. and T.D. Mouw. 2012. *Estimated Sampling Variance in Respondent Driven Sampling Data: Mathematical Derivations, Simulated Tests on Empirical Data, and Evidence from Other Forms of Chain-Referral Data Collection*. Paper for the International Conference on Methods for Surveying and Enumerating Hard-to-Reach Populations, October 31–November 3, New Orleans, LA.
- Verdery, A.M., T.D. Mouw, S. Bauldry, and P.J. Mucha. 2015. "Network Structure and Biased Variance Estimation in Respondent Driven Sampling." *PLOS ONE* 10(12): e0145296. Doi: <http://dx.doi.org/10.1371/journal.pone.0145296>.
- Volz, E. and D.D. Heckathorn. 2008. "Probability Based Estimation Theory for Respondent Driven Sampling." *Journal of Official Statistics* 24(1): 79–97.
- Wejnert, C. and D.D. Heckathorn. 2008. "Web-Based Network Sampling: Efficiency and Efficacy of Respondent-Driven Sampling for Online Research." *Sociological Methods and Research* 37: 105–134. Doi: <https://doi.org/10.1177/0049124108318333>.
- Zheng, T., M.J. Salganik, and A. Gelman. 2006. "How Many People Do You Know in Prison? Using Overdispersion in Count Data to Estimate Social Structure in Networks" *Journal of American Statistical Association* 101(474): 409–423. Doi: <https://doi.org/10.1198/01621450500000116>.

Received January 2016

Revised February 2017

Accepted March 2017