

Estimating Design Effect and Calculating Sample Size for Respondent-Driven Sampling Studies of Injection Drug Users in the United States

Cyprian Wejnert · Huong Pham · Nevin Krishna ·
Binh Le · Elizabeth DiNenno

Published online: 15 February 2012
© Springer Science+Business Media, LLC (outside the USA) 2012

Abstract Respondent-driven sampling (RDS) has become increasingly popular for sampling hidden populations, including injecting drug users (IDU). However, RDS data are unique and require specialized analysis techniques, many of which remain underdeveloped. RDS sample size estimation requires knowing design effect (DE), which can only be calculated post hoc. Few studies have analyzed RDS DE using real world empirical data. We analyze estimated DE from 43 samples of IDU collected using a standardized protocol. We find the previous recommendation that sample size be at least doubled, consistent with $DE = 2$, underestimates true DE and recommend researchers use $DE = 4$ as an alternate estimate when calculating sample size. A formula for calculating sample size for RDS studies among IDU is presented. Researchers faced with limited resources may wish to accept slightly higher standard errors to keep sample size requirements low. Our results highlight dangers of ignoring sampling design in analysis.

Keywords Respondent-driven sampling · Design effect · Sample size · Injecting drug users · HIV · Hidden populations

Disclaimer The findings and conclusions in this paper are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention

C. Wejnert (✉) · N. Krishna · B. Le · E. DiNenno
Division of HIV/AIDS Prevention, National Center for HIV/
AIDS, Viral Hepatitis, STD and TB Prevention, Centers for
Disease Control and Prevention, 1600 Clifton Road NE,
Mailstop E-46, Atlanta, GA 30333, USA
e-mail: cwejnert@cdc.gov

H. Pham
ICF International, Atlanta, GA, USA

Resumen El muestreo dirigido por los participantes (RDS, por sus siglas en inglés) es cada vez más utilizado para tomar muestras de poblaciones ocultas, como las de usuarios de drogas inyectables (UDI). Sin embargo, los datos del RDS son muy particulares y requieren de técnicas de análisis especializado, muchas de las cuales no se han desarrollado. Para estimar el tamaño de la muestra del RDS es necesario conocer los efectos del diseño (DE, por sus siglas en inglés), los cuales solo pueden ser calculados en forma posterior. Pocos estudios han analizado el DE del RDS utilizando datos empíricos reales. Nosotros analizamos el DE de 43 muestras de UDI recolectadas a través de un protocolo estandarizado. Determinamos que la recomendación anterior de que por lo menos se duplique el tamaño de la muestra, congruente con $DE = 2$, subestima DE verdaderos, por lo que recomendamos a los investigadores utilizar $DE = 4$ como una estimación alternativa para calcular el tamaño de la muestra. Se presenta una fórmula para calcular el tamaño de la muestra para estudios con RDS que incluyan UDI. Es posible que los investigadores que cuenten con pocos recursos tengan que aceptar errores estandarizados ligeramente superiores para mantener limitados los requisitos del tamaño de la muestra. Nuestros resultados destacan los peligros al ignorar el diseño del muestreo en el análisis.

Introduction

Although the number of human immunodeficiency virus (HIV) infections attributed to injection drug use decreased between 2006 and 2009 in the United States [1], persons who inject drugs remain at increased risk of HIV infection. A recent analysis of IDU in 23 cities in the United States

[2] found high proportions engaging in HIV risk behaviors. The National HIV/AIDS Strategy for the United States [3] identifies injecting drug users (IDU) as a priority population for HIV prevention efforts. However, the illicit, stigmatized nature of injection drug use makes surveillance and sampling of IDU for research difficult. IDU are a hidden population which cannot be accessed using standard sampling methodologies.

Respondent-driven sampling (RDS) is a peer-referral sampling and analysis method that provides a way to account for several sources of bias and calculate population estimates [4–6] now widely used to reach hidden populations including many at high risk for HIV [7]. IDU are an especially well-suited population for RDS methodology because they rely on social networks for much of their livelihood, including access to drugs, income, and safety. This reliance requires them to form cohesive social exchange networks, which are ideal for RDS. Furthermore, IDU are difficult for researchers to study due to the stigmatized, illegal nature of their activities and the danger posed to field researchers working in their communities [8]. RDS' peer-to-peer recruitment of respondents fosters trust and promotes participation while also reducing field staff's exposure to dangerous environments while conducting research [9].

RDS has become increasingly popular in studies of IDU in the U.S. [10, 11] and abroad [7]. While studies collecting RDS data are numerous, many RDS analytical techniques are still under development due to the complex nature of RDS data and relatively short time since its debut in 1997 [12]. One currently under-developed technique is a priori sample size calculation.

Calculating sample size requirements is a necessary preliminary step to proposing, planning, and implementing a successful study. However, in RDS, this calculation is complicated by the peer-driven nature of RDS data and their analysis [13]. RDS analysis is similar to analysis of stratified samples where sampling weights are applied during analysis to adjust for non-uniform sampling probability. However, while such samples are usually stratified by variables of interest, such as race or income, with preset selection probabilities determined by the researcher, RDS samples are stratified by the target population's underlying network structure which is unknown to researchers. For this reason, data on network structure are collected during sampling and used to calculate sampling weights post hoc [4]. Thus, sample size estimation, which requires knowing selection probabilities, cannot currently be directly calculated a priori.

Cornfield [14] proposed that sample size for complex surveys can be calculated by first calculating the sample size required for a simple random sample (SRS) and then adjusted by a measure of the complex sampling method's

efficiency compared to SRS, termed the design effect (DE) [15]:

$$n = \text{DE} \cdot \frac{P_a(1 - P_a)}{(\text{SE}(P_a))^2} \quad (1)$$

where n is the sample size; $\frac{P_a(1-P_a)}{(\text{SE}(P_a))^2}$ is the common formula for calculating required SRS sample size for some proportion P_a ; SE is the standard error; and DE is the design effect of the actual sampling method used.

Following Kish [15] we define DE as the ratio of the variance of the estimate observed with RDS, $\text{Var}^{\text{RDS}}(P_a)$, to the expected variance of the estimate had the sample been collected using simple random sampling (SRS), $\text{Var}^{\text{SRS}}(P_a)$ as follows:

$$\text{DE} = \frac{\text{Var}^{\text{RDS}}(P_a)}{\text{Var}^{\text{SRS}}(P_a)} \quad (2)$$

where $\text{Var}^{\text{SRS}}(P_a) = \frac{P_a(1-P_a)}{n}$. Consequently, DE compares the observed variance under a complex sampling method—in this case RDS—to that expected for the same estimate under an SRS of similar size [15]. Consequently, DE measures the increase in sample size required to achieve the same power as that of an SRS. For example, a sampling design with $\text{DE} = 3$ requires a sample three times as large as an SRS to achieve the same power [16]. Comparing sampling methods to SRS is mathematically convenient here because calculating power, and consequently estimating sample size, is straightforward for SRS. Consequently, knowing DE for a given sampling method provides a means for estimating sample size [14]. However, SRS and RDS are not directly comparable in an empirical setting. RDS was developed to reach hidden-populations which, by definition, cannot be sampled using SRS methods. A comparison of RDS to other hidden-population methodologies, such as time-location sampling, is beyond the scope of this paper; however we expect all such methods to have similar limitations in comparison to SRS [17].

The definition of DE assumes knowledge of variance associated with a given estimate. However, the only way to truly measure the variance would be to take repeated samples from the same population simultaneously. Given limited resources, researchers are left with two options for estimating the underlying variance: (1) estimate the variance based on a single sample of real world data or (2) create an estimate of the population and simulate repeated samples from that estimate. In both approaches, the researcher is forced to make mathematical assumptions regarding the behavior of participants and the probability of selection. For established sampling methods, such as SRS, this is not problematic because the probabilities of selection are well understood and a single, agreed upon variance calculation exists for real (not simulated) data. For

less established methods, such as RDS, the sampling probabilities are not yet well understood and multiple methods of calculating variance may exist with no agreed upon best approach. In such cases, the calculated DE represents an estimated design effect, denoted \widehat{DE} throughout this paper.

To date, few studies of RDS \widehat{DE} s have been conducted. One often-cited publication suggests RDS samples have a \widehat{DE} of at least two and recommends samples sizes should be at least double that required for a comparable SRS design [18]. However, An RDS study of sex workers in Brazil reported $\widehat{DE} = 2.63$ [19]. Another RDS study of undergraduate students found an average \widehat{DE} of 3.14 [13]. Furthermore, using simulated data, Goel and Salganik [20] find RDS \widehat{DE} s may reach above 20, suggesting RDS analysis may produce highly unstable samples. While empirical evidence suggests RDS estimates are too accurate to support such large \widehat{DE} s [21], further research is needed to determine if a generalized \widehat{DE} can be applied to RDS studies of certain populations or if the commonly applied recommendation of $\widehat{DE} = 2$ should be adjusted. If, for example, a review of many RDS studies found relatively consistent \widehat{DE} s across variables and samples, then this \widehat{DE} could be applied to calculate sample size in future RDS studies.

Methods and Data

Methods for NHBS among IDU (NHBS-IDU) are described in detail elsewhere [10] and briefly reported here. NHBS is a community-based survey that conducts interviews and HIV testing among 3 high risk populations: IDU, men who have sex with men, and heterosexuals at increased risk for HIV infection [22]. Data used in this paper were collected during the first two cycles of NHBS-IDU (NHBS-IDU1 from 2005 to 2006 and NHBS-IDU2 in 2009). NHBS-IDU is conducted by the Centers for Disease Control and Prevention (CDC) in collaboration with state and local health departments in over 20 of 96 large metropolitan statistical areas (those with population greater than 500,000) within the United States (termed “cities” throughout), where approximately 60% of the nation’s AIDS cases had been reported [23]. Health departments in each city received local IRB approval for study activities. CDC reviewed the protocol and determined CDC staff was not-engaged; therefore CDC IRB approval was not required.

Each city operated at least one interview field site that was chosen to be accessible to drug-use networks identified during formative research. Following standard RDS procedures [12, 24], each city began RDS with a limited

number of diverse initial recruits, or “seeds” ($n = 3\text{--}35$). Respondents were provided number-coded coupons with which to recruit other IDU they knew personally. The number of coupons given to each respondent varied by city and ranged from three to six. Within some cities, the number of coupons given varied throughout sampling to regulate the flow of individuals seeking interviews and to reduce the number of coupons in the community as the sample size was approached. Respondents were compensated both for their participation and for each eligible recruit who completed the survey.

NHBS-IDU procedures included eligibility screening, informed consent from participants, and an interviewer-administered survey. Eligibility for NHBS-IDU included being age 18 or older, a resident of the city, not having previously participated in the current NHBS data collection cycle, being able to complete the survey in English or Spanish, and having injected drugs within 12 months preceding the interview date as measured by self-report and either evidence of recent injection or adequate description of injection practices [10]. The survey measured characteristics of participants’ IDU networks, demographics, drug use and injection practices, sexual behaviors, HIV testing history, and use of HIV prevention services. Participants in NHBS-IDU2 were also offered an HIV test in conjunction with the survey.

NHBS-IDU1 was conducted from May 2005 through February 2006 in 23 cities. NHBS-IDU2 was conducted from June 2009 through December 2009 in 20 cities, 18 of which were included in NHBS-IDU1. Our results are based on two cycles of data collection from 18 cities and 1 cycle of data collection from seven cities, five from NHBS-IDU1 and two from NHBS-IDU2 (Fig. 1). For this analysis, RDS data from each city are treated as independent samples. Data from each city were analyzed separately. Results from city level analysis are pooled and presented here by cycle. In total 43 samples (23 samples from NHBS-IDU1 and 20 from NHBS-IDU2) were included. Data collection time varied across cities, due to differences in timing for human subjects approvals, logistics, and speed of sampling.

All cities implemented a single protocol during each cycle. Field operations across all cities were standardized and followed common RDS procedures [24], however, individual cities were provided flexibility, such as determining the number of coupons given or interview locations used, to meet local challenges. While not its primary purpose, the presence of multiple simultaneous samples in the NHBS-IDU research design provides a means for evaluating RDS methodology when used to study populations at increased risk for HIV.

To meet public health goals, NHBS focused on those cities with the largest burden of AIDS disease based on most recent available data at the time cities were being

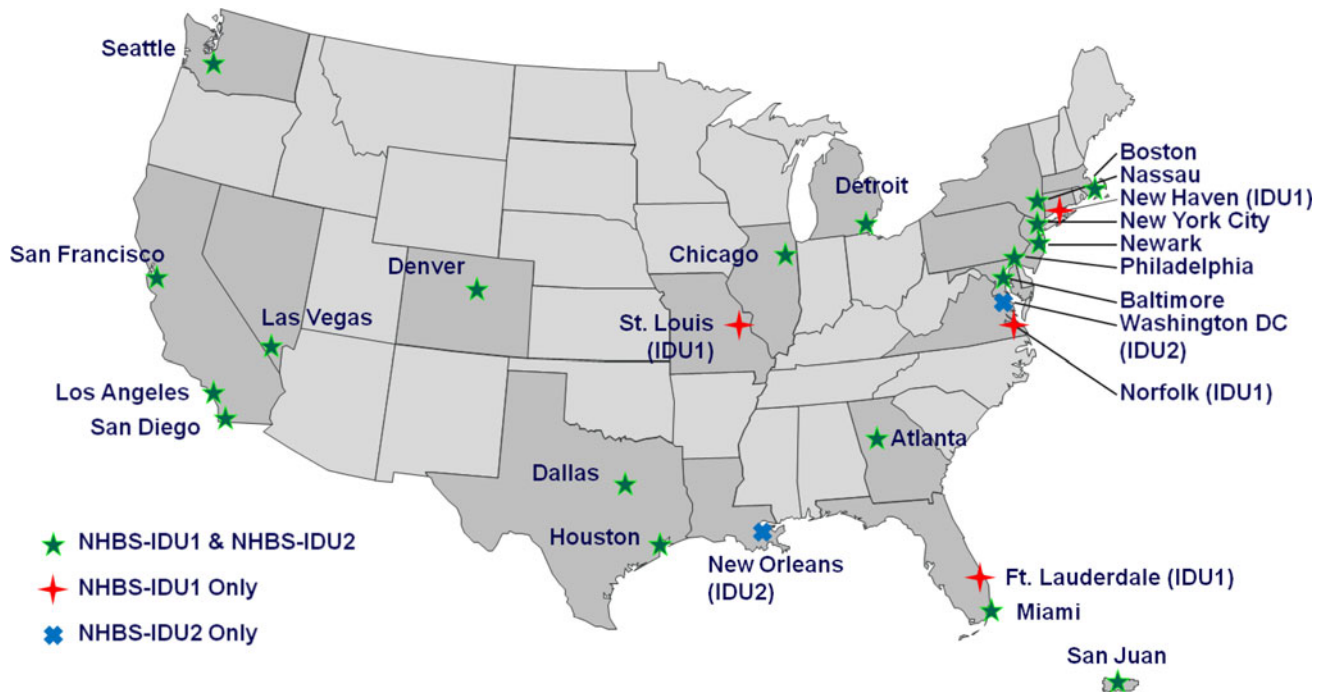


Fig. 1 Map of NHBS-IDU1 and NHBS-IDU2 sampling sites by participating cycle. Cycle for which data are available is shown in *parenthesis* next to city names. If no cycle is shown, data are available for both NHBS-IDU1 and NHBS-IDU2

chosen: NHBS-IDU1 cities were chosen based on AIDS data available from 2000; NHBS-IDU2 cities were chosen based on AIDS data available from 2004 [23]. As such, the cities are not necessarily representative of all U.S. cities or IDU populations. However, NHBS cities are chosen to ensure coverage of diverse geographic areas in the United States and likely represent typical U.S. cities in which RDS studies of IDU or other hard-to-reach populations would be conducted.

Measures

Efficiency of network-based samples, such as RDS, is correlated with homophily in the social network [4]. Homophily is the network principal that similar individuals are more likely to form social connections than dissimilar ones.

In networks where members are defined by a specific stigmatized activity—such as injection drug use—the highest homophily variables tend to be basic demographic characteristics [25]. Based on formative research, we identified three key NHBS-IDU variables likely to have high homophily and, consequently, the largest \widehat{DE} s: race/ethnicity, gender, and age. Race and Hispanic ethnicity were asked separately, then coded into one variable with mutually exclusive categories: white, black, Hispanic (regardless of race), and other (including American Indian or Alaska Natives, Asian, Native Hawaiian and Pacific

Islander, and multiracial). Gender was coded as male or female. Age was grouped into five categories: 18–24, 25–29, 30–39, 40–49, 50 years and over. In addition, we analyze \widehat{DE} for two variables related to HIV risk: sharing syringes and self-reported HIV status. Sharing syringes was defined as having shared any syringes or needles in the past 12 months. Self-reported HIV status was coded as HIV-positive or not (HIV-negative, indeterminate results or status, never received the result, never tested).

Data

As shown in Fig. 2, during the NHBS-IDU1 and NHBS-IDU2 a total of 26,705 persons were recruited to participate (13,519 in NHBS-IDU1; 13,186 in NHBS-IDU2), 524 of whom were seeds (384 in NHBS-IDU1; 140 in NHBS-IDU2). The target sample size for each city in each cycle was 500 IDU (range: 186–631 IDU).

In NHBS-IDU1, a total of 1,563 (12%) persons did not meet NHBS-IDU eligibility criteria and were excluded from analysis. An additional 46 persons had no recruitment information and their records were also excluded. Among the 11,910 eligible persons, we retained only recruitment data for 439 (3.2%) persons whose survey data were either lost during data transfer (334), who were not identified as male or female (67) or whose responses to survey questions were invalid (38). The purpose of this analysis procedure is

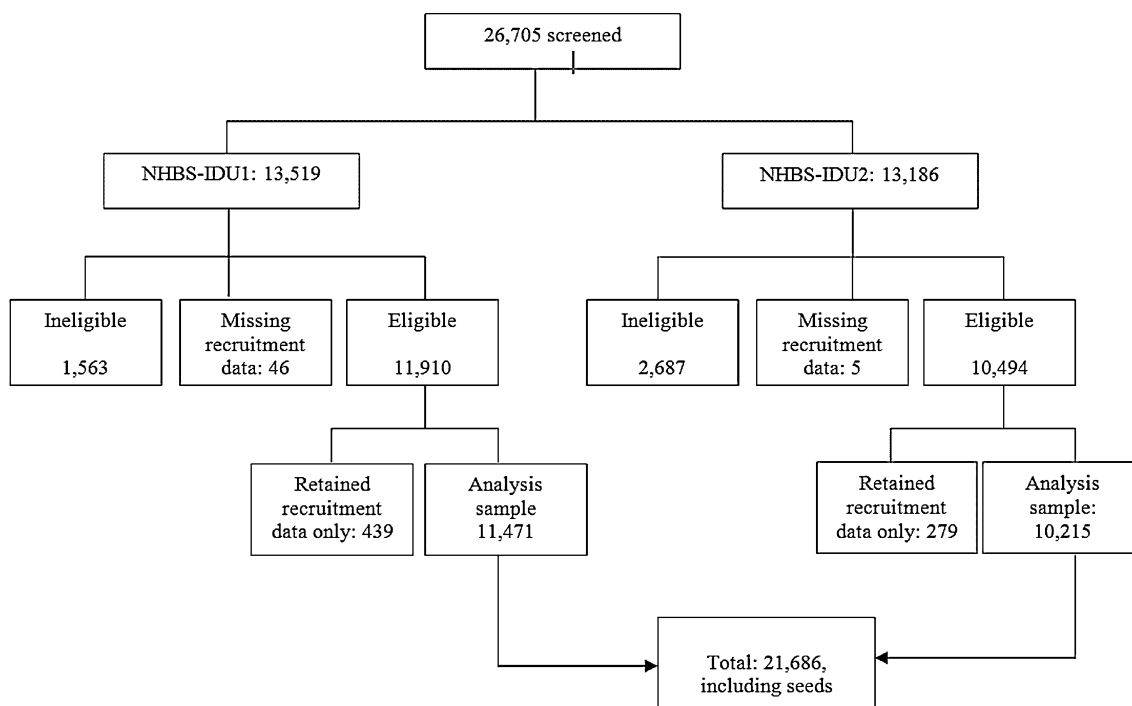


Fig. 2 NHBS-IDU1 and NHBS-IDU2 analysis data

to maintain recruitment chains. Their survey data were coded as missing thus excluded from analysis.

In NHBS-IDU2 a total of 2,692 (20.4%) persons were screened ineligible. These included 2,687 persons who did not meet NHBS-IDU eligibility criteria and 5 persons without recruitment information. Among the 10,494 persons included in the analysis, 279 persons (2.7%) were included with only recruitment data in order to maintain recruitment chains. These include 142 lost records, 55 persons who were not identified as male or female, 64 persons with incomplete survey, and 18 persons for other reasons (repeated participants, invalid survey response or invalid participation coupons).

The final analysis sample included 21,686 persons (11,471 for NHBS-IDU1 and 10,215 for NHBS-IDU2), including seeds. As this analysis does not present test results, participants with missing or indeterminate HIV test results were not excluded from the analysis. Raw sample proportions (unweighted), aggregated national estimates (weighted), and median homophily for all five analysis variables are presented in Table 1.

Analysis Techniques

As discussed above, measurement of \widehat{DE} requires a means of calculating variance. Several methods of estimating RDS variance have been presented [5, 18, 19, 26]. A detailed discussion of these estimates is beyond the scope of this paper. For this analysis, Salganik’s [18] bootstrap variance estimate procedure was used for two reasons. First, this paper revisits

Salganik’s [18] recommendation that $\widehat{DE} = 2$ should be used in calculation of RDS sample size. Our use of the same variance estimation provides a consistent comparison. Second, this is the variance estimate employed by RDS Analysis Tool (RDSAT). To date, RDSAT is the only RDS analysis software publically available. While multiple RDS variance estimators have been proposed, RDSAT remains the primary RDS analysis option for most researchers not involved in the development of new estimators.

RDS analysis was conducted using RDSAT 8.0.8 with $\alpha = 0.025$ and 10,000 resamples for bootstrapping to calculate estimates and estimate standard errors [27]. \widehat{DE} s were calculated as the ratio of RDS variance to variance expected under SRS, as defined above. \widehat{DE} s were calculated independently for each variable within each city. Observed \widehat{DE} s for each variable across all cities within a given cycle are presented in each box plot. A tall box plot represents large variation in \widehat{DE} across cities. Homophily was calculated in RDSAT using Heckathorn’s formula [4, 24]:

$$\begin{aligned}
 H_a &= \frac{S_{aa} - \widehat{P}_a}{1 - \widehat{P}_a} & \text{if } S_{aa} \geq \widehat{P}_a \\
 H_a &= \frac{S_{aa} - \widehat{P}_a}{\widehat{P}_a} & \text{if } S_{aa} < \widehat{P}_a
 \end{aligned}
 \tag{3}$$

where H_a is homophily of subgroup a, S_{aa} is the proportion of in-group recruitments of individuals in subgroup a, and \widehat{P}_a is the estimated proportions of a individuals in the population. As defined, homophily ranges from -1 to 1

Table 1 Demographic characteristics of participants: National HIV Behavioral Surveillance System—Injecting Drug Users, United States, 2005–2006 and 2009

	NHBS-IDU1, 2005–2006				NHBS-IDU2, 2009			
	Unweighted		Weighted		Unweighted		Weighted	
	<i>N</i>	Sample (%)	Proportion (95% CI)	Median homophily	<i>N</i>	Sample (%)	Proportion (95% CI)	Median homophily
Gender								
Male	8,158	71.1	68.0 (66.1–70)	0.27	7,389	72.3	70.9 (69–72.7)	0.19
Female	3,313	28.9	32.0 (30–33.9)	0.11	2,825	27.7	29.1 (27.3–31)	0.15
Race/ethnicity								
Hispanic	2,429	21.2	23.1 (21.1–25.2)	0.35	2,199	21.5	23.1 (21–25.2)	0.37
Black	5,630	49.1	47.7 (45.3–50.1)	0.47	4,756	46.6	40.5 (38.2–42.9)	0.57
White	2,921	25.5	24.8 (22.9–26.7)	0.37	2,786	27.3	31.8 (29.3–34.3)	0.37
Multiple/other	482	4.2	4.3 (3.5–5.1)	0.07	458	4.5	4.6 (3.8–5.4)	0.04
Age (years)								
18–24	443	3.9	4.5 (3.6–5.4)	0.14	349	3.4	4.4 (3.5–5.3)	0.24
25–29	793	6.9	7.2 (6.3–8.1)	0.09	669	6.5	7.1 (6–8.1)	0.12
30–39	2,527	22.0	21.4 (19.9–22.9)	0.07	1,838	18.0	19.9 (18.3–21.6)	0.12
40–49	4,314	37.6	39.1 (37.2–41)	0.06	3,176	31.1	32.1 (30.2–34)	0.08
≥50	3,394	29.6	27.8 (26–29.6)	0.18	4,183	40.9	36.5 (34.5–38.5)	0.23
Self-reported HIV status								
HIV positive	882	7.7	7.8 (6.6–9)	0.18	549	5.4	5.8 (4.9–6.7)	0.10
HIV negative ^a	10,535	91.8	92.9 (91–93.4)	0.25	9,596	93.9	94.2 (93.3–95.1)	0.21
Share syringes^b								
Yes	4,133	36.0	33.0 (31.2–34.7)	0.16	3,575	35.0	33.3 (31.4–35.1)	0.17
No	7,322	63.8	67.0 (65.3–68.8)	0.01	6,495	63.6	66.7 (64.9–68.6)	–0.01
Total	11,471				10,215			

^a Includes participants who reported negative or unknown status

^b Sharing syringes or needles in the 12 months prior to interview

and provides of an estimate of the proportion of in-group ties after accounting for group size.

We utilized NHBS-IDU's extensive data to determine whether and to what extent \widehat{DE} s vary in RDS studies of U.S. IDU and to test the current recommendation that a \widehat{DE} of at least two should be used when calculating sample size requirements for RDS studies. A finding that \widehat{DE} s remain consistent across variables and cities would suggest a single \widehat{DE} can be applied to studies of U.S. IDU using RDS to conduct a priori sample size estimation. If \widehat{DE} s vary within an identifiable range, then the upper bound of that range can be used as a conservative estimate in the calculation of sample size for future RDS studies of U.S. IDU.

Results

Figures 3, 4, 5, and 6 show box plots of \widehat{DE} for five analysis variables across two cycles of NHBS-IDU.

Figure 3 shows \widehat{DE} s for estimates by gender, self-reported HIV status, and syringe sharing behavior. In all cases, the 25th percentile occurs above $\widehat{DE} = 2$ and the 75th percentile occurs below $\widehat{DE} = 4$. The distribution of \widehat{DE} s differs between variables, but appears consistent across cycles.

Figure 4 shows \widehat{DE} s for age estimates by cycle. DE s for younger age categories are more variant than those for older age categories. Overall, Fig. 4 displays a similar pattern to Fig. 3. The majority of DE s fall above $\widehat{DE} = 2$ and below $\widehat{DE} = 4$.

Figure 5 shows DE s for race estimates by cycle. Again the majority of \widehat{DE} s fall above $\widehat{DE} = 2$, however, there is more variation in \widehat{DE} across racial categories and across NHBS cycles by race than by other variables.

The association between homophily and \widehat{DE} is shown in Fig. 6. For both NHBS-IDU1 and NHBS-IDU2 data, a non-linear model fits the data well, with approximately

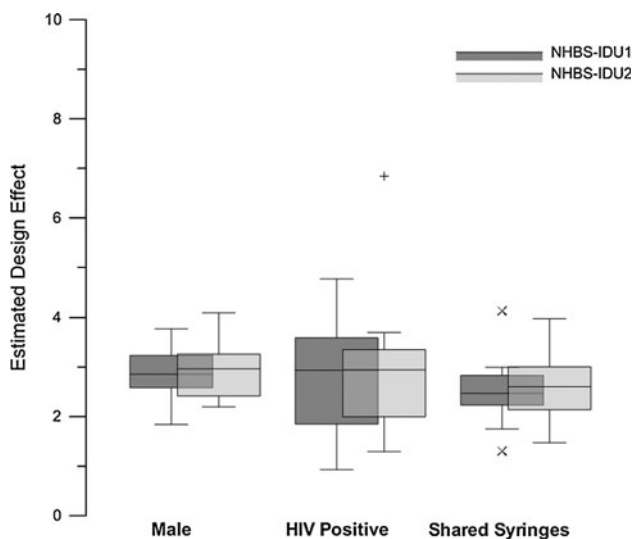


Fig. 3 \widehat{DE} by gender, self-reported HIV status, and syringe sharing behavior for two cycles of NHBS-IDU. \widehat{DE} s of dichotomous variables are equivalent across category (i.e. \widehat{DE} of males = \widehat{DE} of females)

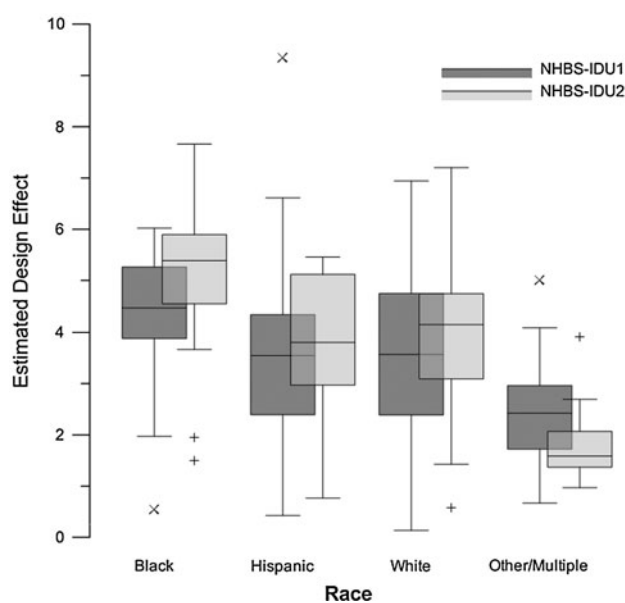


Fig. 5 \widehat{DE} of estimates of race of IDU in NHBS-IDU1 and NHBS-IDU2

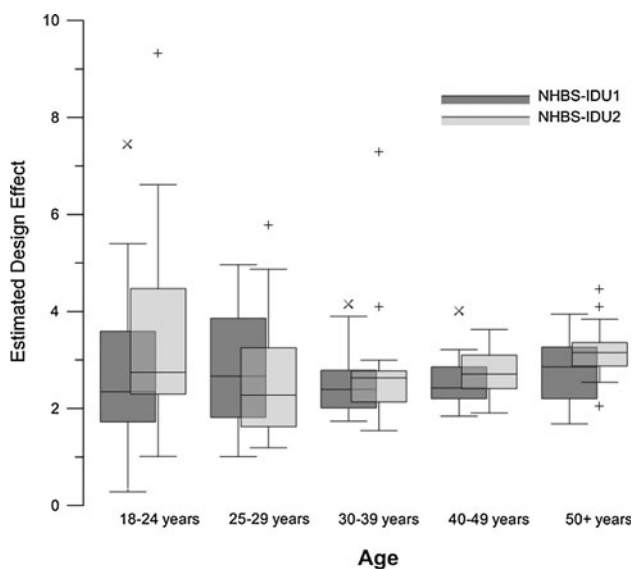


Fig. 4 \widehat{DE} of estimates for age of IDU in NHBS-IDU1 and NHBS-IDU2

50% of the variation in \widehat{DE} attributable to differences in homophily. The intercept in both models, 2.55 in IDU1 and 2.73 in IDU2, is above 2. Thus, even when there is no homophily, the expected \widehat{DE} is greater than the current recommendation. A linear fit was tested but ruled out when the residuals plots showed non-random clear patterns suggesting a non-linear fit. A non-linear association is consistent with Heckathorn [4] who hypothesizes a non-linear association between homophily and standard error in RDS studies.

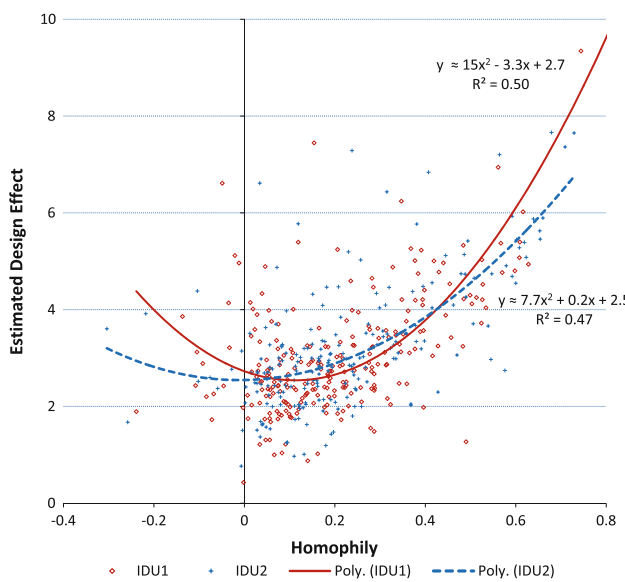


Fig. 6 Association between \widehat{DE} and homophily by gender, race, age, HIV positive status and sharing syringes among IDU in 43 RDS samples in NHBS-IDU1 and NHBS-IDU2. Poly (IDU1) and Poly (IDU2) are the non-linear best fit lines for NHBS-IDU1 and NHBS-IDU2, respectively. Linear fit lines were tested and ruled out when residual plots showed clear nonrandom patterns

Discussion

Our results show that while RDS \widehat{DE} s tend to vary by city and analysis variable, the majority of \widehat{DE} s fall between $\widehat{DE} = 2$ and $\widehat{DE} = 4$ with the exception of several race categories. As mentioned above, the NHBS populations

tended toward insularity by race. Consistent with other work [4], we found high \widehat{DE} s were associated with high homophily. High \widehat{DE} s for blacks, Hispanics, and whites are likely due to the high homophily we observed for these groups.

Our results support two conclusions. First, the original \widehat{DE} recommended by Salganik [18] for use in sample size calculations of RDS studies ($\widehat{DE} = 2$), is unlikely to provide adequate statistical power in RDS studies of IDU in the U.S. While some \widehat{DE} s at or below two were observed, the vast majority fell above $\widehat{DE} = 2$. Second, with the exception of estimates of race, \widehat{DE} s tended to fall below $\widehat{DE} = 4$. Coupled with our previous assessment that these data can be viewed as representative of typical RDS studies of IDU in the U.S., the results suggest that \widehat{DE} s for successful RDS studies focusing on this population will generally fall in the range of two to four. Consequently, we recommend $\widehat{DE} = 4$ as a more appropriate, realistic estimate of \widehat{DE} to use when calculating sample size requirements for RDS studies of IDU in the U.S. In multi-racial studies, formative research should be conducted to determine the level of racial homophily within the population. If race homophily is too high, separate studies may be necessary.

Calculating Sample Size

Based on our finding that $\widehat{DE} = 4$ is a more realistic estimate for the \widehat{DE} of RDS studies of U.S. IDU populations, we can now calculate sample size estimates for future research using Eq. 1. For example, if based on pre-existing knowledge we suspect approximately 30% of IDU engage in a high-risk behavior and we want to estimate this prevalence with a standard error no greater than 0.03, the required sample size is calculated as follows:

$$n = 4 \cdot \frac{(0.3)(1 - 0.3)}{(0.03)^2} = 933 \quad (4)$$

Thus, we would need a sample of 933 IDU in our study. Note that while the relationship between sample size and \widehat{DE} is linear, the relationship between sample size and standard error is exponential. Therefore, while achieving the same statistical power requires a sample size four times larger than SRS, an RDS study with sample size similar to SRS will reduce statistical power by less than four times. For example, if we make the above estimate with a desired maximum standard error of 0.04 instead of 0.03 the new sample size requirement is:

$$n = 4 \cdot \frac{(0.3)(1 - 0.3)}{(0.04)^2} = 525 \quad (5)$$

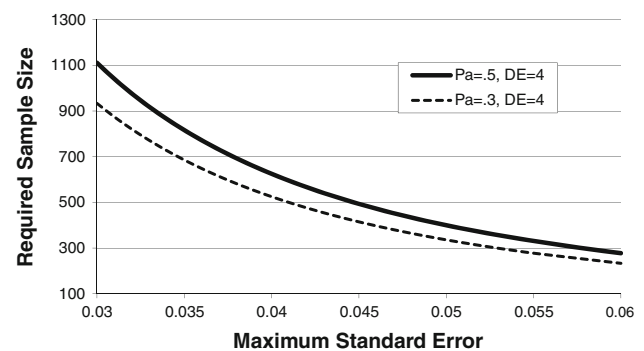


Fig. 7 Required sample size decreases sharply as the maximum allowable standard error of estimates is relaxed for samples with \widehat{DE} of four based on Eq. 1. P_a is the population proportion of individuals with characteristic ‘a’

By slightly reducing statistical power (i.e., increasing maximum standard error), we reduce the required sample size by about 50% to 525 IDU. Figure 7 shows the relationship between sample size and standard error of estimates for RDS studies with $\widehat{DE} = 4$ for $P_a = 0.3$ and $P_a = 0.5$. The relationship is exponential, so a reduction in standard error from 0.03 to 0.04 provides a greater reduction in absolute sample size than reducing standard error from 0.04 to 0.05. A population proportion estimate of 0.5 ($P_a = 0.5$) provides the most conservative estimates and should be used in the absence of outside information. Researchers faced with limited resources may be willing to accept higher standard errors to keep sample size requirements low. Additionally, because $\widehat{DE} = 4$ is a conservative estimate, studies planning for higher standard errors may find observed standard errors are lower than initially expected for many variables.

Conclusion

Our analyses suggest that a \widehat{DE} of four ($\widehat{DE} = 4$) is preferred for applying to calculations of sample size for future RDS studies of IDU in the U.S. This \widehat{DE} is higher than Salganik’s [18] estimate, but lower than some recent theoretical estimates suggested by Goel and Salganik [20]. The advantage of our recommendation is its empirical basis and practical emphasis.

Our results are likely generalizable to RDS studies of IDU populations in the U.S. Studies of non-IDU or populations outside the U.S. should apply our results with caution. Second, our data originate from larger urban populations. Studies of rural IDU may find different \widehat{DE} outcomes. Third, while the large number of cities included in NHBS covers a wide range of IDU populations, city selection favored those cities with the highest overall HIV

burden. Thus, results from studies conducted in cities with lower HIV burdens may differ from our own.

Beyond generalizability, our results have several limitations. First, our recommendation that $\widehat{DE} = 4$ should be used in estimating sample size requirements may underestimate \widehat{DE} with respect to race. The large \widehat{DE} s we observed are likely due to higher homophily by race than other variables, a common finding in U.S. populations. Fortunately, racial homophily is relatively easy to monitor during sampling and to address in formative research. If racial homophily is too high, stratified results can be reported by race. If racial homophily is excessive, such that almost no cross-race recruitment is observed, samples can be separated by race and analyzed separately. If external information on the relative size of the different racial groups is available, the results can be aggregated. Second, our analysis relies on confidence interval bounds generated using the RDSAT bootstrapping algorithm [18, 27]. The algorithm is not the only method of calculating RDS variances [5, 19, 26] and has been shown to underestimate variance under certain conditions [20, 21]. If this variance estimation procedure is not correct the true DE could be higher, possibly much higher, than \widehat{DE} . Similarly, as new, more efficient estimation procedures are developed a lower \widehat{DE} may become more appropriate for estimating sample size. Unfortunately, there is often a significant delay between when new methods are developed and when they are accessible for use by the scientific community. For practicality, we chose a variance estimate that is most accessible to researchers utilizing RDS today. Third, differences in implementation within city across time, such as the number of coupons given to each respondent, negate our ability to explore the effect of some implementation differences on \widehat{DE} . Fourth, our analysis focused on the relationship between \widehat{DE} and homophily, a trait level measure of clustering. Recent work suggests bottlenecks may be a more appropriate level of analysis than homophily [20]. Bottlenecks are a function of the entire network structure and how traits are distributed across that network structure. Unfortunately, our RDS data do not provide sufficient information to analyze the global network structure. Based on the literature, we expect the association between \widehat{DE} and bottlenecks to be stronger than the association between homophily and \widehat{DE} . Finally, this analysis analyzed data from 43 RDS samples implementing a standard NHBS protocol. While the NHBS protocol follows standard RDS procedures and allowed flexibility meet the unique conditions of each city, it is possible that different studies could yield different results. This is especially true of studies that use modified RDS procedures. Further research is needed to explore the effect of differences in implementation on \widehat{DE} .

Despite these limitations, we argue that these results provide an alternative to earlier recommendations for calculating RDS sample size in studies of U.S. IDU and serve as a guide to researchers planning future RDS studies. Previous research presenting weighted RDS estimates and confidence intervals is not impacted by our results, as confidence intervals from RDS analysis account for DE. However, our results further highlight the need for conducting RDS analysis on RDS data. Unweighted analyses of RDS data, which treat the sample as an SRS, not only risk presenting biased estimates, but also risk underestimating the variance of those estimates by as much a factor of four.

Public health researchers working with RDS data will benefit from our results by ensuring they have adequate power for identifying health outcomes such as HIV prevalence and/or risk behaviors. Data collections, including NHBS-IDU, must balance the need for precise estimates with the need to limit burden on the public and to ensure the best use of limited resources. The current target sample size for NHBS-IDU is 500 per city and 10,000 nationally. This sample size is adequate for national estimates, but may have limited power locally for some variables of interest. Given the exponential relationship between standard error and sample size, researchers may be willing to accept and plan for higher standard errors to keep sample size requirements low.

Acknowledgments This paper is based, in part, on contributions by National HIV Behavioral Surveillance System staff members, including J. Taussig, R. Gern, T. Hoyte, L. Salazar, B. Haddock, Atlanta, Georgia; C. Flynn, F. Sifakis, Baltimore, Maryland; D. Isenberg, M. Driscoll, E. Hurwitz, Boston, Massachusetts; N. Prachand, N. Benbow, Chicago, Illinois; S. Melville, R. Yeager, A. Sayegh, J. Dyer, A. Novoa, Dallas, Texas; M. Thrun, A. Al-Tayyib, R. Wilmoth, Denver, Colorado; E. Higgins, V. Griffin, E. Mokotoff, Detroit, Michigan; M. Wolverton, J. Risser, H. Rehman, Houston, Texas; T. Bingham, E. Sey, Los Angeles, California; M. LaLota, L. Metsch, D. Beck, D. Forrest, G. Cardenas, Miami, Florida; C. Nemeth, C.-A. Watson, Nassau-Suffolk, New York; W. T. Robinson, D. Gruber, New Orleans, Louisiana; C. Murrill, A. Neaigus, S. Jenness, H. Hagan, T. Wendel, New York, New York; H. Cross, B. Bolden, S. D'Errico, Newark, New Jersey; K. Brady, A. Kirkland, Philadelphia, Pennsylvania; V. Miguelino, A. Velasco, San Diego, California; H. Raymond, W. McFarland, San Francisco, California; S. M. De León, Y. Rolón-Colón, San Juan, Puerto Rico; M. Courogen, H. Thiede, N. Snyder, R. Burt, Seattle, Washington; M. Herbert, Y. Friedberg, D. Wrigley, J. Fisher, St. Louis, Missouri; and P. Cunningham, M. Sansone, T. West-Ojo, M. Magnus, I. Kuo, District of Columbia.

References

1. CDC. HIV Surveillance Report, 2009-2011, vol. 21. Available from www.cdc.gov/hiv/surveillance/resources/reports.
2. CDC. HIV-associated behaviors among injecting-drug users—23 Cities, United States, May 2005–February 2006. *Morb Mortal Wkly Rep.* 2009;58(13):329–32.

3. White House Office of National AIDS Policy. National HIV/AIDS Strategy for the United States. Washington: The White House Office of National AIDS Policy; 2010.
4. Heckathorn DD. Respondent-driven sampling. II. Deriving valid population estimates from chain-referral samples of hidden populations. *Soc Probl.* 2002;49(1):11–34.
5. Volz E, Heckathorn DD. Probability based estimation theory for respondent driven sampling. *J Off Stat.* 2008;24(1):79–97.
6. Salganik MJ, Heckathorn DD. Sampling and estimation in hidden populations using respondent-driven sampling. *Sociol Methodol.* 2004;34:193–239.
7. Malekinejad M, Johnston LG, Kendall C, Kerr LR, Rifkin MR, Rutherford GW. Using respondent-driven sampling methodology for HIV biological and behavioral surveillance in international settings: a systematic review. *AIDS Behav.* 2008;12(4 Suppl):S105–30.
8. Broadhead RS, Heckathorn DD. Aids-prevention outreach among injection-drug users—agency problems and new approaches. *Soc Probl.* 1994;41(3):473–95.
9. Platt L, Wall M, Rhodes T, Judd A, Hickman M, Johnston LG, et al. Methods to recruit hard-to-reach groups: comparing two chain referral sampling methods of recruiting injecting drug users across nine studies in Russia and Estonia. *J Urban Health.* 2006;83(6 Suppl):i39–53.
10. Lansky A, Abdul-Quader AS, Cribbin M, Hall T, Finlayson TJ, Garfein RS, et al. Developing an HIV behavioral surveillance system for injecting drug users: the National HIV Behavioral Surveillance System. *Public Health Rep.* 2007;122(Suppl 1):48–55.
11. Des Jarlais DC, Arasteh K, Perlis T, Hagan H, Abdul-Quader A, Heckathorn DD, et al. Convergence of HIV seroprevalence among injecting and non-injecting drug users in New York City. *AIDS.* 2007;21(2):231–5.
12. Heckathorn DD. Respondent-driven sampling: a new approach to the study of hidden populations. *Soc Probl.* 1997;44(2):174–99.
13. Wejnert C, Heckathorn DD. Web-based network sampling—efficiency and efficacy of respondent-driven sampling for online research. *Sociol Method Res.* 2008;37(1):105–34.
14. Cornfield J. The determination of sample size. *Am J Public Health Nations Health.* 1951;41(6):654–61.
15. Kish L. *Survey sampling.* New York: Wiley; 1965.
16. Lohr S. *Sampling: design and analysis.* Pacific Grove: Brooks/Cole Publishing Company; 1999.
17. Karon JM. The analysis of time-location sampling study data. In: *Proceeding of the joint statistical meeting, section on survey research methods.* Minneapolis, MN: American Statistical Association; 2005. p. 3180–6.
18. Salganik MJ. Variance estimation, design effects, and sample size calculations for respondent-driven sampling. *J Urban Health.* 2006;83(6 Suppl):i98–112.
19. Szwarcwald CL, de Souza Junior PR, Damacena GN, Junior AB, Kendall C. Analysis of data collected by RDS among sex workers in 10 Brazilian cities, 2009: estimation of the prevalence of HIV, variance, and design effect. *J Acquir Immune Defic Syndr.* 2011;57(Suppl 3):S129–35.
20. Goel S, Salganik MJ. Assessing respondent-driven sampling. *Proc Natl Acad Sci USA.* 2010;107(15):6743–7.
21. Wejnert C. An empirical test of respondent-driven sampling: point estimates, variance, degree measures, and out-of-equilibrium data. *Sociol Methodol.* 2009;39(1):73–116.
22. Lansky A, Brooks JT, Dinunno E, Heffelfinger J, Hall HI, Mermin J. Epidemiology of HIV in the United States. *J Acquir Immune Defic Syndr.* 2010;55(Suppl 2):S64–8.
23. Gallagher KM, Sullivan PS, Lansky A, Onorato IM. Behavioral surveillance among people at risk for HIV infection in the U.S.: the National HIV Behavioral Surveillance System. *Public Health Rep.* 2007;122(Suppl 1):32–8.
24. Wejnert C, Heckathorn D. Respondent-driven sampling: operational procedures, evolution of estimators, and topics for future research. In: Williams M, Vogt P, editors. *The SAGE handbook of innovation in social research methods.* London: SAGE Publications, Ltd.; 2011. p. 473–97.
25. McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: homophily in social networks. *Annu Rev Sociol.* 2001;27:415–44.
26. Gile KJ. Improved inference for respondent-driven sampling data with application to HIV prevalence estimation. *J Am Stat Assoc.* 2011;106(493):135–46.
27. Volz E, Wejnert C, Degani I, Heckathorn D. *Respondent-driven sampling analysis tool (RDSAT).* 8th ed. Ithaca: Cornell University; 2010.