

Analysis of Data Collected by RDS Among Sex Workers in 10 Brazilian Cities, 2009: Estimation of the Prevalence of HIV, Variance, and Design Effect

Célia Landmann Szwarcwald, PhD,* Paulo Roberto Borges de Souza Júnior, PhD,*
Giseli Nogueira Damacena, MS,* Aristides Barbosa Junior, PhD,† and Carl Kendall, PhD‡

Background: Respondent-driven sampling (RDS) is a chain-referral method that is being widely used to recruit most at-risk populations. Because the method is respondent driven, observations are dependent. However, few publications have focused on methodological challenges in the analysis of data collected by RDS.

Methods: In this article, we propose a method for estimating the variance of the HIV prevalence rate, based on the Markov transition probabilities and within recruitment cluster variation. The method was applied to a female commercial sex workers study carried out in 10 Brazilian cities in 2008. Both the inverse of network size and the size of the city were considered in the estimation of overall sampling weights. The study included a behavior questionnaire and rapid tests for HIV and syphilis.

Results: About 2523 interviews were conducted successfully, excluding the seeds. Results show a positive homophily between recruits for those HIV+; HIV- recruiters selected HIV+ recruits 4% of the time; HIV+ recruiters selected other HIV+ recruits 19.6% of the time, about 5 times higher. The prevalence rate was estimated at 4.8% (95% confidence interval: 3.4 to 6.1), and a design effect of 2.63.

Conclusions: Using statistical methods for complex sample designs, it was possible to estimate HIV prevalence, standard error, and the design effect analytically. Additionally, the proposed analysis lends itself to logistic regression, permitting multivariate models. The stratification in cities has proved suitable for reducing the effect of design and can be adopted in other RDS studies, provided the weights of the strata are known.

Key Words: Brazil, complex sample, design effect, FCSW, HIV prevalence, RDS

(*J Acquir Immune Defic Syndr* 2011;57:S129–S135)

From the *Institute of Communication and Information Science and Technology in Health, Oswaldo Cruz Foundation; †Department of STD, AIDS and Viral Hepatitis; and ‡Department of Community Health Sciences, Center for Global Health Equity, Tulane University.

This work was carried out by the Oswaldo Cruz Foundation with technical and financial support of the Ministry of Health/Secretariat of Health Surveillance/Department of STD, AIDS and Viral Hepatitis through the Project of International Technical Cooperation AD/BRA/03/H34 between the Brazilian Government and the United Nations Office on Drugs and Crime—UNODC.

The authors have no conflicts of interest to disclose.

Correspondence to: Célia Landmann Szwarcwald, PhD, Fundação Oswaldo Cruz, ICICT/LIS, Sala 225, Av. Brasil, 4365—Manguinhos, 21040-360, Rio de Janeiro/RJ (e-mail: celials@icict.fiocruz.br).

Copyright © 2011 by Lippincott Williams & Wilkins

INTRODUCTION

Studies have shown the importance of interventions in populations at highest risk of HIV.¹ Depending on the patterns of sexual relationships among population subgroups, small changes in the rate of contact between those at low risk with those at high risk may change the pattern of spread of HIV/AIDS in the general population.² However, these high-risk groups are often small in number and are often hard to reach populations, especially if they practice illicit or stigmatized behaviors.³

The difficulties in monitoring the sexual risk and HIV prevalence in higher risk groups has led to the development of specific sampling methods to collect information in hard to reach populations,^{4,5} including time-space sampling⁶ and respondent-driven sampling (RDS).⁷

The time-space sampling method combines traditional techniques of ethnographic mapping to build a list of the primary units of selection, under the assumption that the groups at highest risk of HIV tend to gather in specific locations.⁸ In RDS, the data are collected through a chain-link recruitment process in which participants recruit future participants of the same population group, forming a network of recruits.⁹ Unlike snowball sampling methods, seeds can only recruit a limited number of people, generally no more than 3, and actual recruitment links are validated by a unique coupon provided by the recruiter to each recruitee. Given certain theoretical assumptions, it is possible to calculate the probabilities of selection and thus classify RDS as a probability sampling method.¹⁰

Since its development, RDS method has been used in many countries worldwide in studies with population subgroups at greatest risk of HIV.^{11–16} The US Centers of Disease Control uses RDS for HIV biological and behavioral sampling among populations of intravenous drug users. A recent review of studies conducted outside the United States between the years 2003 and 2007, reported 123 studies that used RDS, 59 in Europe, 40 in Asia and the Pacific, 14 in Latin America, 7 in Africa, and 3 in Oceania.¹⁷ Surveillance in these groups have also been necessitated by UN General Assembly Special Session on HIV/AIDS requirements for reporting national-level indicator estimates in most-at-risk groups such as sex workers, injection drug users, and men who have sex with men. Brazil has experienced 2 multisite rounds of surveillance with RDS; and in 2008 and 2009, multicenter studies were conducted in 3 population groups at higher risk

for HIV—men who have sex with men, female sex workers (FSW) and drug users.

Although specifications for organizing and collecting RDS data are well developed, questions remain about analysis. The recruitment chain generates dependence between observations, making it difficult to estimate the parameters of interest, the variance, and the design effect.^{18,19}

So far, 2 estimators to calculate statistical inferences with data collected by RDS have been developed. Under the assumption that the recruitment process follows a Markov process, the first estimation method is based on sample estimates of the transition probabilities and the average size of the personal network in each subgroup.¹⁰ Using a different approach, based on the probability of an individual being recruited for the research, a second method of estimation has been proposed, which has the advantage of permitting the analytical estimation of the variance without having to resort to simulation procedures.²⁰

Regarding the estimation of the average values and dispersion measures, the 2 proposed estimators^{10,20} were compared in a study in 2 samples of university students. Although the estimators computed by both methods were very close, the estimation of the variance was problematic for both methods.²¹

For estimation of the variance, Salganick¹⁸ proposes the use of bootstrap simulation methods. However, slight changes in initial assumptions produce large differences in the design effect.^{19,22} In the second estimation method,²⁰ the dependence between observations is considered only in the estimation of variance under the assumption that the correlation between observations decreases with the distance between them in the chain of recruitment.

Although univariate statistics may be sufficient for surveillance in single populations, researchers have been frustrated by the methodological challenges to multivariate analysis with RDS data, which has been little studied. The majority of RDS uses traditional procedures of multivariate logistic regression,^{14,23–27} treating observations as independent.

This article proposes a method for estimating HIV prevalence and its variance, taking into account unequal selection probabilities and the dependence between observations, produced by the pattern of recruitment and by the intraclass correlation among participants invited by the same recruiter. The proposed analysis lends itself to logistic regression, permitting multivariate models. This method was applied to the study of FSWs conducted in 10 cities in 2009. Finally, because we treat the sample as a stratified sample, we argue that the method permits multiple samples to be treated as a single sample, a novel argument for RDS.

METHODS

The Survey Among FSWs

Cities were chosen by Brazilian Department of STD, AIDS and Viral Hepatitis both by geographical criteria (at least 1 in each macroregion) and by its importance in the AIDS epidemic in the country. The sample size was set at 2500,

calculated to estimate an HIV prevalence of 6% with a confidence interval (CI) of 95%, 2-tailed error of 1.3%, and design effect of 2. We tried to distribute the sample in proportion to the female population 18–59 years in the city, but set a minimum sample size of 100 women in each city.

The project was approved by the Ethics Committee of the Oswaldo Cruz Foundation, and endorsed by the committee on National Research Ethics.

Women were eligible to participate in the study if they met the following inclusion criteria: being at least 18 years of age, working as a sex worker in one of the municipalities of the study, have had at least 1 sexual intercourse in exchange for money in the past 4 months; and present a valid coupon to participate.

Fieldwork was conducted in health centers located in the 10 municipalities. For each site, 5–10 initial participants—called seeds—were chosen purposively. Each seed received 3 invitations to give to other sex workers known to them. The recruits of the seeds in the survey were the first wave of the study. After participating in the interview, they received 3 coupons to distribute. This process was repeated until the sample size was achieved in each location. RDS requires a system of primary and secondary incentives. The primary incentive in this study was a gift (beauty products), payment of lunch, and transportation. The secondary incentive was a payment of R\$10 for each recruited person who participated in the study. Sites for data collection and level of incentives were selected following formative research by study PIs in each city.

The questionnaire was presented as an audio computer-assisted self-interview. The questionnaire included modules on the following: information and sociodemographic characteristics related to professional activity, knowledge about HIV transmission, sexual behavior, previous HIV test—in life and in the last year, signs of a sexually transmitted infection, use of alcohol and illicit drugs, access to prevention activities and health services, and discrimination and violence.

Tests for HIV and syphilis were conducted using rapid tests (capillary blood collection), according to the protocols recommended by the Brazilian Department of STD, AIDS and Viral Hepatitis. All participants received pretest and posttest counseling. Participants who tested positive received additional posttest counseling, both to address the psychological impact and to encourage partner notification, and they were referred to public health services for follow-up.

Data Analysis

The rationale for data analysis was to use statistical methods appropriate for data collected using this complex sampling design. This analysis attempted to take into account the dependence among observations, resulting from the recruitment chains and the unequal probabilities of selection, resulting from the different sizes of networks of each participant.

Weighting Of Data

The original authors of RDS propose an weighting based on the inverse probability of selection proportional to the size of the network of each participant.¹⁰ In this study, the question

used to measure the network size of each participant and the resulting weighting is: “How many sex workers who work here in town do you know personally?”

In addition, as the research was conducted in 10 municipalities, the sample was weighted by the relative population size of women 18–59 years of age in each site, assuming the same proportion of women sex workers in all sites, and considering each municipality as a stratum. Mathematically, the sample weights were calculated by:

$$w_{ij} = (1/\delta_{ij}) / \sum_i (1/\delta_{ij}) \cdot m_j \cdot n$$

where:

i represents participant in city j ($j = 1, \dots, 10$).

δ_{ij} = network size of participant i in city j ,

$$m_j = \frac{M_j}{\sum_j M_j}$$

M_j = female population of 18–59 years in city j .

n = sample size.

Estimation of HIV Prevalence, Variance, and Design Effect

The method proposed here takes into consideration both the chain-link effects and the unequal selection probabilities to estimate the prevalence rate, the standard error, and CI, and the design effect.

The tendency of a participant to recruit peers with similar characteristics is called homophily.²⁸ To take into account this bias in recruitment pattern and possible over-representation of individuals with certain characteristics in the study population, recruitment is modeled as a Markov process and estimates of, for example prevalence, can be generated from the equilibrium equation.^{9,10}

Take P as the prevalence rate of HIV, the parameter to be estimated. Then P can be written as a function of the conditional probabilities of recruitment $P_{1,0}$ (probability that a negative participant recruits a positive one) and $P_{0,1}$ (probability of a positive participant recruiting a negative one), known as the transition probabilities of Markov states:

$$P = \frac{P_{1,0}}{P_{0,1} + P_{1,0}}$$

In turn, the conditional probabilities are estimated using appropriate sample weights because the participants in the sample have unequal probabilities of selection. Then, the sample weight, S_{ab} is the sum of sample weights corresponding to participants with result a for HIV testing who were recruited by participants with result b , where a and b are equal to 1 when the result is positive and 0 when negative.

Let $S_0 = S_{00} + S_{01}$; $S_1 = S_{10} + S_{11}$; $S_0 = S_{00} + S_{10}$;

$S_1 = S_{01} + S_{11}$; $S = S_0 + S_1$

Then the transition probabilities are estimated by:

$P_{1,0} = S_{10}/S_0$ and $P_{0,1} = S_{01}/S_1$

To calculate the variance of P , we can write P as a function of the logarithm of $\frac{P_{0,1}}{P_{1,0}}$. Let $r = \ln\left(\frac{P_{0,1}}{P_{1,0}}\right)$

Then:

$$P(r) = \frac{1}{1 + e^r}$$

Using the delta method,²⁹ the variance of $P(r)$ is estimated by:

$\text{var}(P(r)) = [P'(r)]^2 \cdot \text{var}(r)$, where, $P'(r)$ is the derivative of $P(r)$.

Then:

$\text{var}(P) = P^2 \cdot q^2 \cdot \text{var}(x)$, where $q=1-p$, and the variance of x is estimated by:

$$\text{var}(x) = \text{var}(\ln P_{0,1} - \ln P_{1,0})$$

$$= \frac{1}{(P_{0,1})^2} \cdot \text{var}(P_{0,1}) + \frac{1}{(P_{1,0})^2} \cdot \text{var}(P_{1,0})$$

Logistic Regression Analysis

Another way to estimate the transition probabilities is to use a logistic regression model. Under the assumption that recruitment in the RDS method follows a Markov process, in which relations of recruitment are determined by the direct recruiter, not the recruiter’s recruiter or individual members of earlier waves. In this context, we consider the following regression model:

$$\text{Logit}(P) = \alpha + \beta \cdot x_i$$

Where,

i represents participant ($i = 1, \dots, n$).

$x_i = 1$, if the recruiter of participant i is HIV+;

$x_i = 0$ otherwise.

In the regression model proposed above, the effect of the status of the participant who is recruited is incorporated into the model as a fixed effect. The influence of intraclass correlation, given the similarity between the participants recruited by the same person, should be incorporated into the model as a random effect. Therefore, the estimation of the regression model must be performed by specific statistical software that takes into account the complex sampling design.

Mathematically, it can be shown that the transition probabilities can be expressed as a function of the parameter estimators α and β of the logistic regression model. Moreover, the model allows the test of dependence between the serological status of the recruited and recruiter: if the odds ratio (OR) is not statistically significant, we can infer that there is no effect of homophily.

The equivalence of the 2 procedures in estimating the prevalence permits modeling of the multivariate case as follows:

$$\text{Logit}(P) = \alpha + \beta x_i + \sum_k \gamma_k z_{ik}$$

where,

i represents participant ($i = 1, \dots, n$).

k represents variable ($k = 1, \dots, K$), K = number of model variables,

$x_i = 1$ if the participant recruiter i is HIV+,
 $x_i = 0$ otherwise.
 z_{ik} = value of variable z_k of participant i .

RESULTS

About 2523 interviews were conducted successfully, excluding the seeds distributed by city as follows: Manaus (199), Recife (237), Salvador (260), Campo Grande (147), Brasilia (308), Belo Horizonte (289) Santos (191), Texas (601), Curitiba (201), Itajaí (90).

Figure 1 presents the distribution of the sample by price charged per sex act. The dispersion of the distribution shows that the study reached different kinds of sex workers, both in the lower tail distribution (4% with price less than or equal to R\$10.00) but also in the higher (5.7% with the program price greater than or equal to R\$ 200.00).

Table 1 presents the pattern of recruitment by sex work venue. Despite the effect of homophily, diversity is achieved in the sample, because only about 50% of sex workers are recruited from the same location.

The Figure 2 shows the patterns of recruitment according to serological status in Rio de Janeiro. The tendency of HIV+ participants to recruit other HIV+ participants shows the need to incorporate the dependency of observations in data analysis.

Table 2 presents the number of participants with a positive HIV result according to the test results of the corresponding recruiter, after weighting the data for the whole sample in 10 cities. The refusal rate for HIV testing was 0.7% (18 FSWs). Results show a positive homophily between recruits for those HIV+: HIV- recruiters selected HIV+ recruits 4% of the time; HIV+ recruiters selected other HIV+ recruits 19.6% of the time, about 5 times higher.

As described in the Materials and Methods section, the data shown in Table 2 are the basis for calculating the rate of HIV prevalence, standard error, CI, and design effect. The prevalence rate was estimated at 4.8% (95% CI: 3.4 to 6.1),

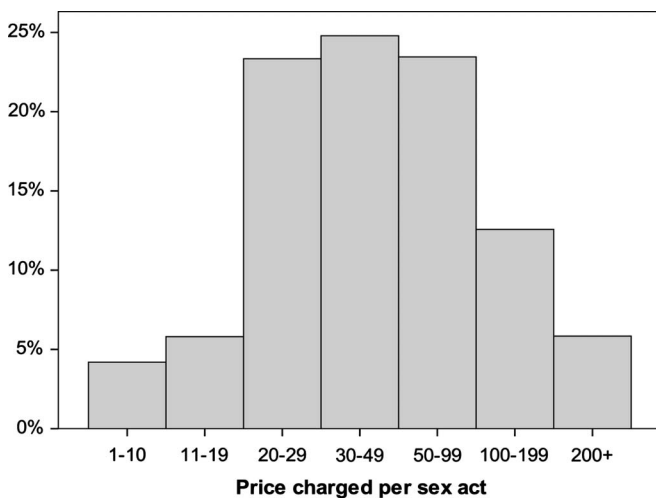


FIGURE 1. Distribution (%) of sex workers by price charged for sex act.

TABLE 1. Distribution of Sex Workers by Venue

Venue of Sex Workers	Venue of Recruiters									
	Club		Hotel		Street		Other		Total	
	n	%	n	%	n	%	n	%	n	%
Club	583	54.5	166	32.5	186	21.5	15	20.0	951	37.7
Hotel	151	14.1	135	26.3	215	24.9	28	36.4	529	20.9
Street	313	29.3	189	36.9	452	52.3	28	36.8	983	39.0
Other	22	2.1	22	4.2	11	1.3	5	6.8	61	2.4

and a design effect of 2.63. By way of comparison, the results of the logistic regression model are presented in Table 3, estimated using the R statistical software. The OR was 5.8 ($P < 0.0001$), indicating homophily for HIV seropositivity. The equivalence of results is demonstrated at the bottom of Table 3, where the conditional probabilities and their standard errors using the parameters estimates of the regression model are estimated.

DISCUSSION

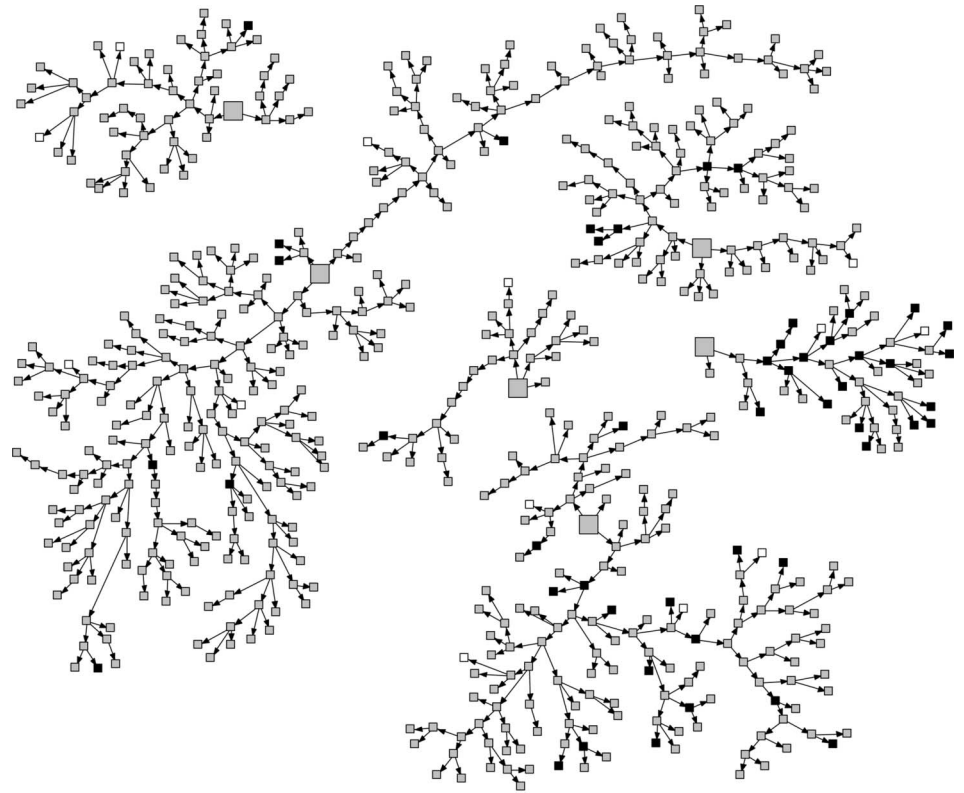
Studies conducted in groups at highest risk for sexually transmitted infections using conventional sampling strategies are generally problematic.³⁰ A review of the literature on estimating the rate of HIV prevalence among sex workers in the period 2000–2008, showed that among the 75 articles found, 84% used convenience samples and 47% of articles had insufficient sample size.³¹

In this context, RDS is a methodological advance, in the sense of providing a rationale for considering these samples as probabilistic, in the sense of carefully controlling recruitment and other procedures, and in the sense of utilizing the community members own knowledge to recruit “hidden” members, that is, participants not normally enrolled in surveillance.^{12,32}

A recent systematic review of 128 international published accounts of studies utilizing RDS for behavioral surveillance in populations at highest risk of HIV found a small proportion of studies with operational difficulties and the mean ratio between the achieved sample size and planned was 0.94.³³ In this project, it was possible to recruit more than 2500 sex workers in a period of just 4 months in 10 Brazilian cities. Only 2 municipalities did not achieve targeted size: Salvador because of a delay in starting the research; and in Itajaí, due to heavy rains and flooding in the city, the field work was interrupted for a month. Still, the ratio between planned and achieved sample in these cities were, respectively, 0.87 and 0.90.

Besides incentives, rapid testing for HIV and syphilis at the time of interview may have encouraged participation. In fact, the refusal rate for HIV testing was very small. Despite the fact that the incentives may have been more attractive to the sex workers of lower socioeconomic status, the results presented here show that the sample included sex workers of diverse socioeconomic status.

With respect to the analysis of data collected by RDS, there are still many unresolved issues. Although the point



Legend

■ HIV - (rapid test)

■ HIV + (rapid test)

□ Declined to take test

* Large icons represent seeds

FIGURE 2. Pattern of recruitment of commercial sex workers according to HIV test result in Rio de Janeiro. *Large icons represent seeds.

estimators can be argued to be asymptotically unbiased, researchers have questioned the plausibility of theoretical assumptions to real data.³⁴ Other issues concern the measurement of the size of individual networks. Studies have shown that changes in the questions about the size of the network change, the average estimates, and variability of parameters of interest.^{35,36}

Regarding the estimation of the average values and dispersion measures, the 2 proposed estimators^{10,20} were compared in a study in 2 samples of university students. Although the estimators computed by both methods were very close, the estimation of the variance was problematic for both methods.²¹

For estimation of the variance, Salganick¹⁸ proposes the use of bootstrap simulation methods. However, slight changes in initial assumptions produce large differences in the design

TABLE 2. Recruitree HIV Test Results According to HIV Test Results of the Corresponding Recruiter

Results of Recruited	Results of Recruiter			
	HIV-		HIV+	
HIV-	n	2238	n	115
	P _{0,0}	96.0%	P _{0,1}	80.4%
	Variance	0.093	Variance	0.288
HIV+	n	94	n	28
	P _{1,0}	4.0%	P _{1,1}	19.6%
	Variance	0.093	Variance	0.288
Total	2332		143	

TABLE 3. Logistic Regression of Outcome Variable Recruitree HIV Test Result, Controlling for the HIV Status of Recruiter

Parameters	Estimators	Standard Error	Exponential	P
α	-3.171	0.105	0.042	—
β	1.761	0.235	5.821	<0.001
Conditional Probabilities	Estimators	Variance		
P _{1,0}	0.0403	0.093		
P _{0,1}	0.8037	0.288		

effect.^{19,22} In the second estimation method,²⁰ the dependence between observations is considered only in the estimation of variance under the assumption that the correlation between observations decreases with the distance between them in the chain of recruitment.

In this article, we propose estimators of the prevalence and standard error, using statistical procedures suitable for analysis of data collected in complex sampling designs. We considered the effect of homophily, the intracluster correlation of participants recruited by the same person and the unequal selection probabilities. At equilibrium, our estimator is theoretically equal to the 2 other proposed estimators.^{10,20} However, the prevalence estimates and standard error proposed here are based on the estimation of transition probabilities from one state to another and can be calculated analytically without the need to consider equilibrium in Markov processes. Moreover, the procedure is equivalent to modeling using logistic regression, where the likelihood of HIV infection depends on serological status of the direct recruiter.

Similar to our proposal, the OR was used previously to test the significance of the effect of homophily.³⁷ Frost et al³⁸ have proposed the use of cross-tables between recruiters and recruits and log-linear models to model the transition probability. Similar to the approach used by Volzl and Heckathorn,²⁰ using the dependence of observations only in the estimation of the variance, Strathdee et al³⁹ proposed the use of a logistic regression model using fixed effects to represent covariates and random effects to express the possible correlation between recruiter and recruit. However, in none of these studies was the procedure used for the estimation of the variance in HIV prevalence and the design effect.

In summary, the findings of this study show that the RDS is a viable methodology for the study of FSWs in Brazil. In the analysis of HIV prevalence, the effect of homophily was significant, showing the need to consider the dependence of observations. The design effect was large, approaching 3. These findings and others suggest that design effect for RDS may be so large that RDS restricted to a single geographic locale in relatively small target populations—as originally proposed by the authors of the method—may not achieve adequate sample size or, if they can, be financially justifiable for routine surveillance. It is worth noting, however, that the design effect is calculated as the ratio between the estimate of the variance determined by the sampling plan and the estimate of variance obtained by a simple random sample of the same size,⁴⁰ which would not be a viable sampling method in studies of sex workers. The ability to calculate analytically the effect of design, as proposed in this article, allows the calculation of sample size for future studies that will use the same complex sampling design, facilitating the research project. The stratification in cities or districts has proved suitable for reducing the effect of design and can be adopted in other works, provided that the weights of the strata are known.

REFERENCES

- Williams JR, Foss AM, Vickerman P, et al. What is the achievable effectiveness of the India AIDS Initiative intervention among female sex workers under target coverage? Model projections from southern India. *Sex Transm Infect.* 2006;82:372–380.
- Boily MC, Lowndes C, Alary M. The impact of HIV epidemic phases on the effectiveness of core group interventions: insights from mathematical models. *Sex Transm Infect.* 2002;78(suppl 1):i78–i90.
- Penrod J, Preston DB, Cain RE, et al. A discussion of chain referral as a method of sampling hard-to-reach populations. *J Transcult Nurs.* 2003;14:100–107.
- Magnani R, Sabin K, Saidel T, et al. Review of sampling hard-to-reach and hidden populations for HIV surveillance. *AIDS.* 2005;19(suppl 2):S67–S72.
- Semaan S, Lauby J, Liebman J. Street and network sampling in evaluation studies of HIV risk-reduction interventions. *AIDS.* 2002;4:213–223.
- Muhib FB, Lin LS, Stueve A, et al. A venue-based method for sampling hard-to-reach populations. *Public Health Rep.* 2001;116(suppl 1):216–222.
- Heckathorn DD. Respondent-driven sampling: a new approach to the study of hidden populations. *Soc Probl.* 1997;44:174–199.
- MacKellar DA, Gallagher KM, Finlayson T, et al. Surveillance of HIV risk and prevention behaviors of men who have sex with men—a national application of venue-based, time-space sampling. *Public Health Rep.* 2007;122(suppl 1):39–47.
- Heckathorn DD. Respondent-driven sampling II: deriving valid population estimates from chain-referral samples of hidden populations. *Soc Probl.* 2002;49:11–34.
- Salganick MJ, Heckathorn DD. Sampling and estimation in hidden populations using respondent-driven sampling. *Social Methodol.* 2004;34:193–240.
- Jarlais DCD, Arasteh K, Semaan S, et al. HIV among injection drug users: current epidemiology, biologic markers, respondent-driven sampling, and supervised-injection facilities. *Curr Opin HIV AIDS.* 2009;4:308–313.
- Johnston L, Sabin K, Hien MT, et al. Assessment of respondent driven sampling for recruiting female sex workers in two Vietnamese cities: reaching the unseen sex worker. *J Urban Health.* 2006;83:16–28.
- Liu H, Liu H, Cai Y, et al. Money boys, HIV risks, and the associations between norms and safer sex: a respondent-driven sampling study in Shenzhen, China. *AIDS Behav.* 2008;13:652–662.
- Pollini RA, Alvelais J, Gallardo M, et al. The harm inside: injection during incarceration among male injection drug users in Tijuana, Mexico. *Drug Alcohol Depend.* 2009;103:52–58.
- Reisner S, Mimiaga MJ, Case P, et al. Predictors of identifying as a barebacker among high-risk new England HIV seronegative men who have sex with men. *J Urban Health.* 2008;86:250–262.
- Wattana W, van Griensven F, Rhucharoenpornpanich O, et al. Respondent-driven sampling to assess characteristics and estimate the number of injection drug users in Bangkok, Thailand. *Drug Alcohol Depend.* 2007;90:228–233.
- Malekinejad M, Johnston LG, Kendall C, et al. Using respondent-driven sampling methodology for HIV biological and behavioral surveillance in international settings: a systematic review. *AIDS Behav.* 2008;12:S105–S130.
- Salganick M. Variance estimation, design effects, and sample size calculations for respondent-driven sampling. *J Urban Health.* 2006;83:i98–i112.
- Goel S, Salganick MJ. Respondent-driven sampling as Markov chain Monte Carlo. *Stat Med.* 2009;28:2202–2229.
- Volzl E, Heckathorn DD. Probability based estimation theory for respondent driven sampling. *J Off Stat.* 2008;24:79–97.
- Wejnert C. An empirical test of respondent-driven sampling: point estimates, variance, degree measures, and out-of-equilibrium data. *Sociol Methodol.* 2009;39:73–116.
- Goel S, Salganick MJ. Assessing respondent-driven sampling. *Proc Natl Acad Sci U S A.* 2010;107:6743–6747.
- Bobashev GV, Zule WA, Osilla KC, et al. Transactional sex among men and women in the south at high risk for HIV and other STIs. *J Urban Health.* 2009;86(suppl 1):32–47.
- Shahmanesh M, Wyal S, Copas A, et al. A Study comparing sexually transmitted infections and HIV among Ex-Red-Light district and non-Red-Light district sex workers after the demolition of Baina Red-Light district. *J Acquir Immune Defic Syndr.* 2009;52:253–257.
- Lauby JL, Millett GA, LaPollo AB, et al. Sexual risk behaviors of HIV-positive, HIV-negative, and serostatus-unknown Black men who have sex with men and women. *Arch Sex Behav.* 2008;37:708–719.

26. Wheeler DP, Lauby JL, Liu KL, et al. A comparative analysis of sexual risk characteristics of Black men who have sex with men or with men and women. *Arch Sex Behav*. 2008;37:697–707.
27. Platt L, Wall M, Rhodes T, 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(suppl 6):39–53.
28. Heckathorn DD, Semaan S, Broadhead RS, et al. Extensions of respondent-driven sampling: a new approach to the study of injection drug users aged 18–25. *AIDS Behav*. 2002;6:55–67.
29. Bishop YMM, Fienberg SE, Holland PW. *Discrete Multivariate Analysis*. Cambridge, MA: MIT Press; 1975.
30. Mills S, Sidel T, Magnani R, et al. Surveillance and modelling of HIV, STI, and risk behaviours in concentrated HIV epidemics. *Sex Transm Infect*. 2004;80(suppl II):ii57–ii62.
31. Pascom AR, Szwarcwald CL, Barbosa Júnior A. Sampling studies to estimate the HIV prevalence rate in female commercial sex workers. *Braz J Infect Dis*. 2010;14:385–397.
32. Kendall C, Kerr LR, Gondim RC, et al. An empirical comparison of respondent-driven sampling, time location sampling, and snowball sampling for behavioral surveillance in men who have sex with men, Fortaleza, Brazil. *AIDS Behav*. 2008;12(suppl 4):S97–S104.
33. Johnston LG, Malekinejad M, Kendall C, et al. Implementation challenges to using respondent-driven sampling methodology for HIV biological and behavioral surveillance: field experiences in international settings. *AIDS Behav*. 2008;12(suppl 4):S131–S141.
34. Heimer R. Critical issues and further questions about respondent-driven sampling: comment on Ramirez-Valles, et al. (2005). *AIDS Behav*. 2005;9:403–408.
35. Ma X, Zhang Q, He X, et al. Trends in prevalence of HIV, syphilis, hepatitis C, hepatitis B, and sexual risk behavior among men who have sex with men. Results of 3 consecutive respondent-driven sampling surveys in Beijing, 2004 through 2006. *J Acquir Immune Defic Syndr*. 2007;45:581–587.
36. Barbosa Júnior A, Pascom AR, Szwarcwald CL, et al. Transfer of sampling methods for studies on most-at-risk populations (MARPs) in Brazil. *Cad Saude Publica*. 2011;27(suppl 1):S36–S44.
37. Deiss RG, Brouwer KC, Loza O, et al. High-risk sexual and drug using behaviors among male injection drug users who have sex with men in 2 Mexico-US border cities. *Sex Transm Dis*. 2008;35:243–249.
38. Frost SD, Brouwer KC, Cruz MF, et al. 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. *J Urban Health*. 2006;83(suppl 1):83–97.
39. Strathdee SA, Lozada R, Pollini RA, et al. Individual, social, and environmental influences associated with HIV infection among injection drug users in Tijuana, México. *J Acquir Immune Defic Syndr*. 2008;47:369–376.
40. Cochran WG. *Sampling Techniques*. 3rd Edition. New York, NY: John Wiley & Sons; 1977.