

Title: Time Location Sampling and Respondent Driven Sampling: techniques implementation for monitoring concentrated HIV/AIDS epidemic in Mexico.

Authors: Gayet, Cecilia (1), Fernández-Cerdeño, Araceli (2)

(1) Facultad Latinoamericana de Ciencias Sociales México (FLACSO México)

(2) University of California, San Diego / San Diego State University, Joint Doctoral Program in Public Health (HB)

Background

Monitoring HIV/AIDS in countries with concentrated epidemics requires special methods to carry out sexual behavior surveys or estimate the prevalence of HIV. Accessing stigmatized population groups such as men who have sex with men (MSM), male and female sex workers (MSW and FSW) or mobile populations such as long-distance truck drivers (truckers) is a challenge, but their data are necessary to establish the progress of the epidemic and to offer prevention and HIV testing in places where these groups gather.

Time-Location Sampling (TLS)

Time-Location Sampling (also known as venue sampling) is a probabilistic method used to recruit members of a target population at specific times in set venues. The sampling framework consists of venue-day-time units (VDT) – also known as time-location units - which represent the potential universe of venues, days and times. For example, a VDT unit could be a defined period of four hours on a Monday in a specific venue. The fieldwork team identifies a range of time-location units to locate the members of the target population through interviews and key informants, service providers, and members of the target population. Then, the team visits the venues and prepares a list of VDT units which are considered potentially eligible on the basis of checking the number of people present. In addition, interviews are conducted with those in charge of the venue to ascertain affluence on certain days and at certain times. With this information, population size for each VDT unit, and the number eligible for each sample are estimated.

The sample is selected in stages. In the first stage of the sampling strategy, a simple or stratified sample of all the time-location units which appear in the sampling frame list (preferably with probability proportional to the total number of members of the population eligible for each time-location unit) is selected. In the second stage, the participants are systematically selected for each time-location unit selected randomly. Using Time-Location Sampling allows the surveying of informal venues, such as private houses, into the sampling frame, to reach the least visible members of the target population, or those who do not typically frequent public places (Semaan et al., 2002).

TLS has several advantages (Magnani et al., 2005). In an unbiased sample, all members of a population must have a known, non-zero probability of selection. TLS provides a sampling framework that allows calculation of the probability selection of each individual in the sample. Arbitrary convenience samples compromise validity of comparisons between survey rounds since change in the composition of the successive sample may stem from the difference in methods. TLS significantly reduces arbitrary selection of venues and individuals and provides a replicable method of sample selection. TLS is recommended when all population members can be reached at

certain sites at different times and where no comprehensive list (census) of the target population exists (Lemp et al., 1994; MacKellar et al., 1996).

Respondent-driven sampling

RDS is a modified form of snowball sampling which allows researchers to recruit highly-stigmatized groups who do not congregate in well-known places. RDS does not only provide a probabilistic method to reach the desired sample size, but also allows the research group to identify networks and the characteristics of those belonging to the networks (Heckathorn, 2002).

Objectives

- To discuss the advantages of and challenges for the field implementation of two sampling methods for sexual behavior surveillance surveys of hard-to-reach populations: Time location sampling (TLS) and Respondent driven sampling (RDS).
- To present lessons learned from surveys of four target populations: male sex workers (MSW), female sex workers (FSW), men who have sex with men (MSM) and long distance truck drivers (truckers).¹
- To reflect on the feasibility of the TLS and RDS methods for use with these hard to reach groups.
- To reflect on the TLS limitations to randomness imposed by these unique field work conditions and their implications in terms of sample biases.

Methods

This reflection is based on the experiences gathered during a behavioral surveillance and HIV prevalence (BSS+) survey among populations at high risk of HIV/AIDS carried out during 2005 and 2006. Fieldwork took place in four Mexican cities (Acapulco, Monterrey, Nezahualcoyotl and Tampico). We used TLS to obtain samples of three target populations: 603 female sex workers (FSW), 1,111 men who have sex with men (MSM) and 313 long distance truck drivers (truckers). Also, we used RDS to obtain one sample of 102 male sex workers. This presentation revisits the feasibility, advantages and limitations of these sampling techniques.

For TLS, we compare the list of venues (TLS units) initially identified, those from the randomly selected sample and those where actual data could be collected. We analyze rejection rates from the selected venues and potential individual participants to provide an analysis of quantifiable biases. We revisit other selection biases introduced by the interviewers.

For RDS, we analyze productivity of seeds and time-delay to achieve the desirable number of branches. We also reflect on the field conditions during data collection, such as safety of field-team members and confidentiality during the interview.

The data were analyzed with SPSSv15, with the exception of the survey carried out with the RDS method. In the case of RDS, we used RDSAT v5.4, a specifically designed software published by the Sociology Department of Cornell University, New York, USA.

¹ For results of these surveys see Gayet C, Magis C, Sacknoff D, Guli L (2007). Prácticas sexuales de las poblaciones vulnerables a la epidemia de VIH/SIDA en México. México DF: FLACSO México/CENSIDA, Colección Ángulos del SIDA.

For RDS analysis, a partition analysis was carried out, which allowed the analyst to define multiple sub-groups exclusive to a population (for example, participants who systematically used a condom with regular partners). For these sub-groups, the RDS computer program generates estimates of population size and confidence intervals. These population estimates and corresponding confidence intervals provide a method to characterize the wider community of MSW in Nezahualcoyotl. Through statistical processing with the RDS statistical package, simple percentages are obtained for the sample according to a specific characteristic (for example, the percentage of participants who speak an indigenous language among MSW) and the estimated population size, i.e. the broader network, not just the people interviewed.

Results

Expected sample sizes for the FSW, MSM and truck driver populations

In order to determine the sample size needed to detect a change of 15 percentage points, for different indicators for each target group, the initial value (P1) has been estimated at 50%. The design effect has been estimated at 2 because of the group design used to sample the target groups. The significance level has been established at 0.05. Calculations using the relevant formula result in a sample size of 267 for each target group. Given the lack of available information on these target groups' sexual behavior, a correction factor was not considered, except to compensate for refusal to participate. It is generally accepted that between 10 and 15% refuse to participate. To deal with these rejection rates, the expected sample size of 267 increased to 300 for each target group in each city.

Expected sample size for MSW

Male sex workers are considered to be a high-risk group, but assessing sample sizes is a real challenge given the lack of previous estimates concerning the size of this population. Due to the difficulties encountered in accessing this population, owing to the stigmatization and discrimination which they often experience, the RDS method was piloted in Nezahualcoyotl to assess its feasibility and effectiveness. An expected sample size of 300 respondents was established, with the assumption that more cases could be obtained by this method, especially taking into account that the survey would take place in the suburbs of Mexico City, the most populated city in Mexico.

Table 1. Expected sample size by population and city studied.

City (State)	MSM	FSW	MSW	Truck drivers	Total
Acapulco (Guerrero)	300	300			600
Monterrey (Nuevo León)	300	300		300	900
Mexico City (Estado de México)*	300		300*		600
Tampico (Tamaulipas)	300				300
Subtotal	1200	600	300	300	
TOTAL					2400

*RDS.

Achieved sample sizes can be observed in Table 2. Some cities reached expected sample sizes. Implementation and other field challenges prevented achieving the expected sample size in other cities. The following sections discuss the application of each method and the challenges encountered to reach the target populations.

Table 2. Achieved sample sizes by city and population.

City (State)	MSM	FSW	MSW	Truck drivers	Total
Acapulco (Guerrero)	301	285			586
Monterrey (Nuevo León)	315	318		313	946
Mexico City (Estado de México)*	225		102*		327
Tampico (Tamaulipas)	270				270
Subtotal	1111	603	102	313	
TOTAL					2129

*RDS method.

Time Location Sampling

In order to use the TLS method with MSM, FSW and truck drivers, it was necessary to first construct the sampling frame for each population in every city. Using structured interviews with key persons, lists of venues used by these populations were drawn up with geographical indications. The physical existence of these venues was verified, as well as their opening and closing times, and an estimate of the size of the target population present. The 'time-location' units were established, a random selection took place, and quotas were set for interviews at each location, taking into account the estimated population size in each venue. The conglomerates were represented by the venues identified for each population. For example, for male and female sex workers, these were formed by bars, strip clubs, meeting places in specific streets or sections of beach, etc. For long distance truck drivers, the venues were car parks, cafes and restaurants for truck drivers, transport companies, and resting places along motorways.

Once the location had been chosen on one day of the week, and for a specific time, the members of the target population to be interviewed were randomly selected if the number of potential participants exceeded that set for the quota. If the number of people from the target population was lower than that established in the quota on the first visit to the venue, the visit was repeated the following week, respecting the day of the week and the time established through random selection.

Refusal Percentage

Time-Location sampling (TLS) can be problematic for various reasons. Magnani et al. list the following as significant: the omission of unidentified venues which remain outside the sampling frame, the omission of those members of the population who do not visit these types of venues, and the refusal of many members of the subpopulation to test for HIV and answer a questionnaire in the venues (Magnani et al., 2005). MacKellar and colleagues (2007) mention additional shortcomings, such as staffing difficulties, lack of community support and the challenges of conducting ongoing formative research (in an adaptive sampling framework) and recruitment simultaneously (MacKellar, Gallagher and Finlayson, 2007).

Based on our field experience, we add the refusal of the owners or managers of some establishments who control access to some sites considered in the sampling frame, which led to a significant number of interviews being missed as contact with members of the target populations could not be established. This problem may be related to insufficient community support, but we believe that it is necessary to consider it separately, as it may introduce significant biases constraining the external validity of the sample. Previous research included only venues with owner's permission in the sampling frame (Muhib et al., 2001), but in our opinion, that sampling

frame does not show the actual refusal rate. An additional problem was a possible selection bias by recruiters.

The sampling frame of venues was designed using a list drawn up in consultation with key individuals related to the target population. When the locations were checked, it was found that some were outside the geographical area established for the study and others were not actually venues with high affluence of members of the target population at the time of the study. Among those venues which fulfilled the requirements, two types of refusal were experienced: firstly by the owners or managers of the establishment, which prevented any access to the venue; and within establishments where the survey was accepted, by members of the target population. The percentage of refusal can be seen in Table 3.

Table 3. Refusal percentages of venues and individuals by target population

	MSM	FSW	Truckers
Refusal of venues	(4 cities)	(2 cities)	(1 city)
Total number of venues reported by key informants	N=232	N=153	N=99
Total number of venues after verification	N=100	N=130	N=48
Total number of venues selected	n=75	n=130	n=47
Agreed to participate	68%	65%	62%
Declined to participate	24%	34%	34%
Agreed and then declined to participate	8%	1%	4%
Refusal of individuals within venues			
Average of all venues selected	35%	13%	23%

A higher proportion of MSM venues than FSW and trucker venues rejected participation in the survey. However, FSW venues had higher rates of lost interviews due to venues that prevented interviewers from doing their work. Therefore, rejection by venues likely caused a more significant bias in the FSW study than in other populations. In broad terms, the project didn't have access to sex work venues with middle and high income clients. Sex workers who are reached by telephone or escort services were excluded from the sample. In the MSM study, the biggest biases stemmed from venue owners' problems with public authorities or conflicts of interest between them and the local implementing organizations.

At the individual level, the rejection rate was higher in the MSM survey and lowest in the FSW survey. It is hard to get patrons who came to a venue to have fun, to participate in a half-hour interview and have an HIV test. In contrast, despite our emphasis on voluntary participation, in most closed FSW venues it was the owners or managers who seemed to decide for the workers.

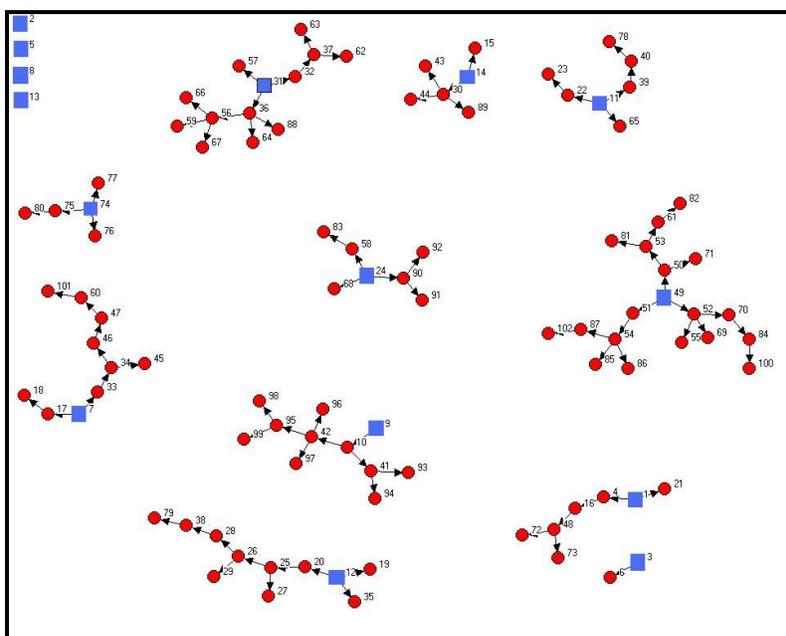
Interviewers' selection biases

It was difficult to identify MSM in open gathering venues, such as parks. Thus, in deference to those who did not want to be identified as MSM, subjects were approached according to subjective markers, which caused a selection bias against masculine looking MSM. Additionally, there was a noticeable age bias (older men were not approached by interviewers even if there was evidence that they sought to purchase sexual services from younger men).

Respondent driven sampling

The first step in RDS was to collaborate with NGOs working with target populations in order to recruit 15 initial respondents known as 'seeds'. The MSW 'seeds' in Nezahualcoyotl went to the venue where the study was being carried out, completed the questionnaire on behavior and took the HIV rapid test. After completing the test, the 'seeds' were given a small sum of money to cover transport costs and for their time (60 pesos, equivalent to 6 American dollars at the time of the survey). They were also given three coupons to recruit three MSW to participate in the study. The 'seeds' received an additional payment for each participant recruited and who completed the study (30 pesos (US\$3) for each referral). This process was repeated until the last possible day of the study (eight weeks), yielding a sample of 102 people. The limit of three coupons per participant was set so that a wide group of subjects had the opportunity to recruit participants, reducing homophily* between the study participants, and preventing the emergence of semi-professional recruiters and potential competition over recruitment (Heckathorn, 2002). [*Homophily is defined as the tendency to affiliate oneself with those who have similar features; for example, level of education, income, ethnic origin etc.]

Figure 1. MSW participant chains using 'seeds' in RDS method



RDS also provides information on recruitment patterns among the respondents. To evaluate whether the recruiters and those they recruit are similar, RDS calculates a measure of homophily according to a self affiliation bias. For example, if homophily is equal to 1, then all the contacts in the network belong to the same group (i.e. young people only recruit young people). Homophily is equal to zero when all the contacts in the network are formed randomly, and it is equal to minus 1 if all the contacts in the network are outside the group (i.e. young people only recruit older people). A score of -1 is also called heterophily. The ideal levels of homophily, which guarantee a random recruitment process, are between -0.3 and +0.3. In the case of MSW in Nezahualcoyotl, recruitment patterns and homophily were examined according to age. The indices were located between -0.3 and +0.3 for each age category, indicating a random selection of participants. It is

important to consider these factors as they enable us to have a better understanding of the demographic characteristics and risk factors of this hard-to-reach population.

Regarding the MSW case using the RDS method, small obstacles arose because of the location of the venue where potential respondents had to arrive. Once a week, a market took place outside the front door of the building, which made it difficult to see and which blocked the entrance. Moreover, both the venue and the interviewing team had been hired for a limited time, which may not have been long enough to achieve the expected sample size.

Other limitations encountered in the field included that four out of 15 seeds did not produce additional referrals and one seed produced only 1 referral. The referral chains generally had few waves (probably because of the limited time-scope of the study). We lacked an estimate of the total size of the population of MSW before the study. Therefore, it is impossible to determine the main reason why the targeted sample size could not be reached. We propose one or several of the following field limitations to be considered:

- a) the small sample may have been because the population was smaller than expected and a large part of the target population was reached;
- b) the weak social networks of the male sex workers (Abdul-Quader et al., 2006);
- c) the survey period was too limited to allow enough recruitment waves;
- d) the survey location was the inconvenient;
- e) there may have been safety concerns for participants who had to go to an unknown survey-site in a context where police harassment is frequent.

Conclusions

In time-location-sampling, field conditions make it difficult to adhere to random selection methods. Documented biases resulted from rejection by venue managers. Additional biases occurred in the selection of subjects who were approached by interviewers. Despite these limitations, TLS has important advantages: Once the sampling framework has been established, TLS provides a systematic procedure and precise instructions for field-staff on “where, when and how to select” participants to interview. This reduces arbitrary selection of participants by interviewers as long as adequate supervision and quality control procedures are implemented. As a result, samples are much more diversified and have fewer induced biases than if participants are selected by convenience sampling.

The main challenge in our application of RDS was not having reached the expected sample size. However, analysis shows a clearly random selection at least regarding age of participants. Lacking a previous estimate of population size, it must be noted that it cannot be determined if our low sample size was due to implementation challenges such as a short survey period, inaccessible location, safety considerations and fear of harassment by participants. Further research is needed to determine if the achieved sample actually covered a significant majority of the target population.

The innovative sampling methods described above provide systematic sampling procedures that yield better samples than convenience sampling. However, they have important limitations in the field, which need to be considered. To improve survey methods in the future, there remains a need to report practical difficulties of implementing complex sampling methods. The validity of

surveys for hard-to-reach populations would greatly benefit if these limitations were widely acknowledged and discussed.

References

- Abdul-Quader AS, Heckathorn DD, Sabin K, Saidel T (2006). Implementation and Analysis of Respondent Driven Sampling: Lessons Learned from the Field. *J Urban Health*. 2006 Nov;83(6 Suppl):i1-5.
- Gayet C, Magis C, Sacknoff D, Guli L (2007). *Prácticas sexuales de las poblaciones vulnerables a la epidemia de VIH/SIDA en México*. México DF: FLACSO México/CENSIDA, Colección Ángulos del SIDA.
- Lemp GF, Hirozawa AM, Givertz D, Nieri GN, Anderson L, Lindegren ML, Janssen RS, Katz M (1994). Seroprevalence of HIV and risk behaviors among young homosexual and bisexual men: the San Francisco/Berkeley Young Men's Survey. *JAMA*. 1994;272:449-454.
- MacKellar DA, Gallagher KM, Finlayson T, Sanchez T, Lansky A, Sullivan PS (2007). Behaviors of Men Who Have Sex with Men—A National Application of Venue-Based, Time-Space Sampling. *Public Health Rep*. 2007 Supplement 1;122:39-47.
- MacKellar D, Valleroy L, Karon J, Lemp G, Janssen R. The Young Men's Survey: methods for estimating HIV seroprevalence and risk factors among young men who have sex with men. *Public Health Rep*. 1996;111(suppl 1):138-144.
- Magnani R, Sabin K, Saidel T, Heckathorn D (2005). Review of sampling hard-to-reach and hidden populations for HIV surveillance, *AIDS* 2005, 19 (suppl 2): S67-72.
- Muhib FB, Lin LS, Stueve A, Miller RL, Ford WL, Johnson WD, Smith PJ; Community Intervention Trial for Youth Study Team (2001). A venue-based method for sampling hard-to-reach populations, *Public Health Rep*. 2001;116 Suppl 1:216-22.
- Semaan S, Lauby J, Liebman J (2002). Street and Network Sampling in Evaluation Studies of HIB Risk-reduction Interventions, *AIDS Rev* 2002: 4: 213-223.