Transfer of sampling methods for studies on most-at-risk populations (MARPs) in Brazil

Transferência de métodos de amostragem para estudos em populações sob maior risco à infecção pelo HIV no Brasil

Abstract

The objective of this paper was to describe the process of transferring two methods for sampling most-at-risk populations: respondent-driven sampling (RDS) and time-space sampling (TSS). The article describes steps in the process, the methods used in the 10 pilot studies, and lessons learned. The process was conducted in six steps, from a state-of-the-art seminar to a workshop on writing articles with the results of the pilot studies. The principal investigators reported difficulties in the fieldwork and data analysis, independently of the pilot sampling method. One of the most important results of the transfer process is that Brazil now has more than 100 researchers able to sample MARPs using RDS or TSS. The process also enabled the construction of baselines for MARPs, thus providing a broader understanding of the dynamics of HIV infection in the country and the use of evidence to plan the national response to the epidemic in these groups.

HIV; Vulnerable Populations; Sampling Studies

Introduction

Twenty-seven years since identification of the first AIDS case in Brazil, the Brazilian epidemic is still concentrated, with a disproportional share among most-at-risk population groups for HIV infection. According to Wilson & Halperin (p. 423) of the World Bank, the epidemic is considered to be concentrated “if transmission occurs in defined vulnerable populations, typically sex workers, men who have sex with men, and injecting drug users, and their sexual partners, and if protecting them would protect wider society”. An analysis of the dynamics in the AIDS epidemic in Brazil shows much higher rates in some groups, thereby contradicting more pessimistic forecasts of a generalized HIV/AIDS epidemic in the medium term. To illustrate the epidemic’s unequal distribution in Brazil, the AIDS incidence rate in 2004 among male injecting drug users (IDUs) was approximately 15 times that of heterosexual males in general. A similar contrast is observed when comparing men who have sex with men (MSM) and heterosexual men.

In the early years of the epidemic in Brazil, the main epidemiological surveillance activity was monitoring AIDS cases. Later, there was an evident need to monitor the principal risk factors related to sexually transmitted infection, including the knowledge, practices, and sexual behaviors of the Brazilian general population and among the subgroups with the most AIDS cases.

As for monitoring risk practices in most-at-risk groups for HIV infection, due to the inherent methodological difficulties in sampling hard-to-reach populations, until recently the surveys conducted in Brazil for these populations used convenience sampling. In addition, the studies were local, without a national scope, and their usefulness for monitoring activities was thus quite limited. Examples include the public opinion surveys conducted by the National STD/AIDS Program in 2001 and 2002 with male homosexuals over 16 years of age in gay gathering places (results unpublished). Other examples include the multicenter studies among IDUs participating in harm reduction projects, conducted in some Brazilian cities – the AjUDE-Brazil Project.

Currently, obtaining representative samples of hard-to-reach population subgroups still poses one of the main challenges for HIV surveillance. Traditional sampling methods are inadequate for generating representative samples, since in order to estimate parameters with the necessary robustness it is necessary to select very large samples, which is hindered by operational difficulties and costs. In addition, some of the population groups at greatest risk for HIV are involved in illegal activities and often remain hidden from society.

The development and expansion of the use of specific and probabilistic methods like time-space sampling (TSS) and respondent driven-sampling (RDS) brought new possibilities and stimulus for researchers interested in studies on most-at-risk groups for HIV infection. In recent years, the international scenario has witnessed an increase in the number of studies on hard-to-reach groups using probabilistic sampling methods.

The TSS method combines ethnographic techniques with venue sampling, and one of its premises is that hard-to-reach populations tend to gather in specific places, for example nightclubs, bars, saunas, parks, and certain stretches of streets. In each gathering place for the target population identified during the ethnographic research, the individuals are numbered in time periods (generally lasting four hours), on varying days, to construct a combined list of potential places, days, and times to be sampled. The time-space units (place, day, and hour of the day) where at least eight individuals from the target population were numbered constitute a list of units in the first selection stage. In the second selection stage, the individuals that frequent the primary sampled units and that are apparently eligible are numbered consecutively and approached to verify the study inclusion criteria. Eligible individuals are then invited to participate in the study.

The RDS sampling procedure is similar to that used in the so-called snowball techniques, since it uses referral chains for sampling, i.e., the recruitment process is done by the members of the study population themselves rather than by the researchers. The difference between the two methods is that the recruitment process in RDS was improved to enable calculation of the selection probabilities, and it can thus be classified among the probabilistic sampling methods.

Deployment of RDS begins with choosing individuals from the target population, called “seeds”, to participate in the research. Each seed receives a set number of coupons to give to their friends and acquaintances in the target population. Participants who come to the study site with a valid coupon and who meet the other inclusion criteria are considered eligible and constitute the first “wave” in the study. These new recruits then receive new coupons to invite friends and acquaintances from the same population group to participate in the study. This process is repeated until reaching the initially planned sample size. Damacena et al. describe in detail the underlying assumptions in RDS and the experience with implementing this sampling process among female sex workers in Brazil.

In addition to the increasing use of TSS and RDS, other events had positive effects on the adoption of surveillance activities by national HIV/AIDS programs in most-at-risk populations. Such developments include the use of specific indicators for most-at-risk populations for HIV in countries with a concentrated epidemic, in monitoring HIV/AIDS commitments by UNGASS (the United Nations General Assembly Special Session). This process encouraged countries to seek ways of responding to this demand. Meanwhile, recognition of the difficulty in developing an effective preventive vaccine against HIV, at least in the medium term, also encouraged the quantitative and qualitative expansion of prevention activities targeting most-at-risk groups.

Recognizing the importance of monitoring indicators of knowledge and behaviors related to HIV infection, as well as for prevalence rates for HIV and other sexually transmitted infections (STI) among subgroups at greatest risk of HIV, Brazil’s Department of STD, AIDS, and Viral Hepatitis, in partnership with the Global AIDS
Program-Brazil of the Centers for Disease Control and Prevention (CDC/GAP), the University of California/San Francisco (USA), Tulane University (USA), and the Oswaldo Cruz Foundation (Fiocruz, Brasil), conducted the transfer of specific sampling methods for hard-to-reach population subgroups. The TSS \(^{18}\) and RDS \(^{13}\) sampling techniques were selected for this transfer, having been adopted in the United States for behavioral surveillance in most-at-risk groups \(^{19}\).

The aim of the current study was to describe the process of transferring methods for sampling populations at greatest risk of HIV, the methodological procedures used in the pilot studies conducted in Brazil, and the lessons learned in the transfer process.

**Methodology: the transfer process**

In October 2004, experts from the Department of STD, AIDS, and Viral Hepatitis and Fiocruz participated in the International Consultation on Sampling Most at Risk Populations (MARPs), held by CDC in Atlanta. During this meeting the most appropriate methods for sampling most-at-risk populations for HIV infection were discussed. For application in Brazilian studies, a commitment was made to transfer the TSS \(^{18}\) and RDS \(^{13}\) methods. In addition to the Department of STD, AIDS, and Viral Hepatitis and Fiocruz, the process also involved Tulane University and the University of California (San Francisco). The entire process received technical and financial support from CDC/GAP-Brazil.

The objective of this initiative was not to obtain results for the most-at-risk groups for HIV, but to test the feasibility of applying TSS and RDS in these population subgroups in Brazil, as well as to train researchers in the HIV/AIDS field to implement the new methods.

The team in charge of the transfer process consisted of two experts from the AIDS Department; two experts from CDC/GAP-Brazil; one researcher from Fiocruz; and one professor each from Tulane University and the University of California (San Francisco). The transfer process took place from November 2004 to September 2006, with the stages described briefly below:

a) A workshop for presenting the state of the art in sampling methods for hard-to-reach populations. American scientists with expertise in RDS and TSS were invited to present and discuss these methods with Brazilian epidemiologists and statisticians and representatives of social movements. Participants were offered the possibility of applying one of the two methods, RDS or TSS, in pilot projects;

b) Selection of ten pilot projects, of which eight proposed to use RDS in a study on MSM (one), drug users (three), IDU (one), and female sex workers (three). Two projects proposed to use TSS, involving MSM and truck drivers. These studies were entirely financed with resources from the cooperative project between the Department of STD, AIDS, and Viral Hepatitis and CDC/GAP-Brazil. Each project was offered a maximum of 50 thousand Brazilian Reais, the equivalent of approximately US$23,000.00 (twenty-thousand US dollars) at the time. The approximate total cost of the ten studies was BR$ 460,000.00 (US$ 209,000.00). Funding was only released for each study after approval of the projects by the local research ethics committees;

c) A workshop was held with the principal investigators of the ten selected projects to prepare the research protocols. This workshop included a detailed discussion of methodological issues, the basic questionnaire to be used by all the studies, and the procedures related to implementation of the fieldwork;

d) The central team visited all pilot study sites before the fieldwork began in order to prepare the local teams to implement the chosen sampling process;

e) The central team conducted supervisory visits to all the pilot study sites to check progress with the fieldwork; and

f) A workshop was held to transfer the data analysis techniques. All the principal investigators participated in this workshop in order to systematize the studies’ main findings.

The projects’ field activities took place from September 2005 to January 2006. When the transfer process was finalized in September 2006, the Department of STD, AIDS, and Viral Hepatitis held a symposium, broadcast live on the Internet, to publicize the information.

To evaluate the transfer process, a questionnaire was submitted to the principal investigators of the pilot projects. The questionnaire contained questions on prior experience with studies in most-at-risk populations for HIV, reasons for choosing the sampling method, evaluation of the transfer process, and difficulties in implementing the pilot studies. Data on the type of sampling method used, target population, sample (planned versus reached), and costs of the pilot projects were collected from the project reports.

**Results: pilot projects and evaluation of the transfer process**

The transfer of sampling methods for most-at-risk subgroups for HIV included ten different...
studies and nine principal investigators (one investigator coordinated two studies).

For each project, Table 1 shows: the sampling method used; target population; municipality in which the research was done; sample size (planned and reached), and cost.

Of the ten pilot studies, only two used TSS as the sampling method, both of which were completed successfully. The main fieldwork problems identified by the researchers that used TSS related to difficulty in updating the target population’s gathering places. They also cited difficulty in accessing the venues frequented by individuals with higher purchasing power.

Among the eight studies that used RDS, only four succeeded in reaching the planned sample sizes. The main reasons cited were limitations in funding and available time for implementing the projects. In the case of the RDS study in Curitiba, Paraná State, with drug users, the planned sample size was not reached because of saturation in the study population, since the study was limited to only two neighborhoods, and the drug-using population had been overestimated by the principal investigator.

The main difficulty in the fieldwork identified by the researchers who used RDS was the fluctuation in the number of participants at the interview site, alternating idle time with moments of heavy demand. Due to the exponential recruitment process, by the end of the study there was a heavy demand at the research site, which often meant a long waiting period for the participants, some of whom even gave up. The complexity of the coupon control system was another problematic issue cited by researchers.

Particularly in the study in Porto Alegre with male and female sex workers, the networks that began with female seeds failed to capture a sufficient number of male sex workers (transvestites and non-transvestites) to reach stochastic equilibrium. Thus, rather than considering the network as having a single component (a theoretical assumption of RDS), it would have been more appropriate to treat male and female sex workers as having independent social networks.

In the pilot studies that used RDS as the sampling method in IDU populations, it proved impossible to identify seeds, which resulted in changes in the target population in the two studies. The initial proposal was to study IDUs, but the target population had to be expanded to include drug users in general. However, the research ended up showing that the difficulty was in identifying IDU seeds, since IDUs appeared in the initial waves in both studies.

Figure 1, by way of illustration, shows the recruitment network obtained in the study on drug users in Manaus, using RDS as the recruitment method. The first waves not only revealed injecting cocaine users, but also captured injecting heroin users, previously “hidden” within the population of drug users.

According to the answers by the principal investigators to the evaluation questionnaire, application of the methods generally entailed a medium level of difficulty. For both sampling methods, implementation of the ethnographic research was identified as an additional difficulty. This was most evident in the two groups who applied TSS, due to the need to identify and update the list of time-space units using ethnographic research.

As regards analysis of data collected with TSS, the main problem encountered by the pilot projects was calculation of the expansion factors, since there are still no well-established procedures for weighting the sample in the two selection stages. One of the authors of the current article (C.L.S.) proposed a sample weighting procedure for the second selection stage, based on the number of persons numbered and interviewed. The expansion factors are calculated based on counting the individuals in each of the time-space units. The proposed weighting is based on the inverse selection probability for each interview site, day, and time and assigned to each individual in the sample. This type of weighting was used in the two studies that used TSS as the sampling method.

For analysis of data collected by RDS, the process proposed for calibrating the data was also problematic. As occurred in international studies, the study on female sex workers in Santos showed that changes in the way questions on the network’s size are formulated substantially modify the mean estimates and range of target parameters, thereby altering the confidence intervals and design effects. Additional limitations were encountered in attempts to conduct multivariate statistical analysis, not available to date in the Respondent Driven Sampling Analysis Tool (RDSAT).

In response to the evaluation questionnaire for the transfer process, the positive points mentioned by the principal investigators were: the instructors’ expertise in the methods discussed and the organization and logical sequence of the themes presented in the workshops. Also cited were the following: the transfer had been conducted didactically; all the stages in the study’s development had been presented step by step; and all stages in the fieldwork had been accompanied by the team in charge. Suggestions for improving the process included increasing the time devoted to the workshop on data analysis, espe-
Table 1

Sampling method, target population group, municipality, planned sample size, sample size reached, and amount spent (in thousands of US dollars) on pilot studies. Brazil, 2006.

<table>
<thead>
<tr>
<th>Method</th>
<th>Population group</th>
<th>Municipality (State)</th>
<th>Sample size</th>
<th>Amount spent (in thousands US$) *</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Planned</td>
<td>Reached</td>
</tr>
<tr>
<td>Time-space sampling</td>
<td>Men who have sex with men</td>
<td>Porto Alegre (RS)</td>
<td>620</td>
<td>718</td>
</tr>
<tr>
<td></td>
<td>Truck drivers</td>
<td>Salgueiro (PE)</td>
<td>500</td>
<td>704</td>
</tr>
<tr>
<td>Respondent-driven</td>
<td>Men who have sex with men</td>
<td>Fortaleza (CE)</td>
<td>400</td>
<td>406</td>
</tr>
<tr>
<td>sampling</td>
<td>Sex workers</td>
<td>Porto Alegre (RS)</td>
<td>420</td>
<td>190</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Manaus (AM)</td>
<td>568</td>
<td>210</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Santos (SP)</td>
<td>200</td>
<td>175</td>
</tr>
<tr>
<td></td>
<td>Injecting drug users</td>
<td>Campinas (SP)</td>
<td>330</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Drug users</td>
<td>Manaus (AM)</td>
<td>245</td>
<td>499</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Curitiba (PR)</td>
<td>392</td>
<td>295</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Recife (PE)</td>
<td>291</td>
<td>184</td>
</tr>
</tbody>
</table>

* US$1 = BR$2.20.

especially in the case of application of RDS, which is more complex and uses a specific analytical application, RDSAT 22.

Final remarks

The transfer of sampling methods for most-at-risk subgroups for HIV infection in Brazil used two sampling techniques, TSS and RDS, in ten separate studies.

Among the difficulties with the TSS method, formative research was emphasized by the principal investigators who used this methodology. In order to identify all the target population's gathering places, ethnographic research is generally difficult, toilsome, and time-consuming. Another limitation of TSS is the incomplete coverage of the target population, since individuals that do not frequent public venues have zero probability of being selected 25. The technique may also exclude individuals who only schedule meetings through the Internet. Recent initiatives to refine TSS have included considering the virtual space of Internet 26.

Limitations of another order refer to researchers' safety in hazardous places and difficulties in approaching subjects when they are having fun, often under the influence of alcohol or other drugs 25.

Another problem with the implementation of TSS in Brazil was the impossibility of performing serological tests at the time of the interview, given the specific standards for places where rapid tests for HIV infection can be performed 27. Since it is necessary to refer the research subjects to health services, this can cause significant losses, thus hindering the estimation of HIV and syphilis prevalence rates with this type of sampling. Such problems can obviously be solved, for example using a mobile health unit, as according to experience in other countries 28.

Meanwhile, in studies that used RDS as the sampling technique, conducting the project in health units was identified as the method’s greatest advantage, with the understanding that this research implementation approach encourages the use of health services and opens opportunities for intervention and performing serological tests in the subgroups at greatest risk of HIV infection. Another benefit of this sampling method is the possibility of finding “hidden” individuals. After capturing the most visible subjects, the extended acquaintance networks end up including hidden individuals. In fact, the study on drug users in Manaus, Amazonas State, revealed participants who injected heroin, a highly uncommon practice in Brazil 29.

However, even a long network does not always mean that the sample captures the diversity
of types of individuals in the population segment under study, which is essential for the sample to be representative of the target population. In the study on female sex workers in Santos, which used the RDS method to select 173 women, 121 worked in nightclubs, of whom 116 (96%) were concentrated in only two clubs.

Another point that merits discussion is the issue of material incentives, provided for at two moments in the RDS deployment. Each participant receives a stipend for participating in the study, called the primary incentive, plus a bonus, called the secondary incentive, per subject recruited who participates successfully in the study. From the practical point of view, in Brazil, monetary incentives to participate are not generally allowed by institutional review boards, but must be limited to providing transportation and food to cover the individual’s expense in order participate in the study. Furthermore, from the theoretical and statistical point of view, the secondary incentive contradicts the assumption of random choice among the recruiter’s friends and acquaintances. If someone receives incentives to bring a new participant, the recruiter will choose the closest potential person, that is, the
person who works in the same place or whom he or she meets frequently. Limiting the coupons to a small number (two or three) means choosing the persons at the beginning of the proximity line, as reported by Heckathorn in his initial model, but it contradicts the assumption of randomness in choosing the recruit in the subsequent model.

Flaws in the principles of randomness in choosing recruits and the structure of dependence among observations created in the RDS recruitment process are issues that were not considered in the method’s original proposal, thus limiting the application of traditional statistical techniques. In fact, the greatest difficulties identified in applying RDS in the Brazilian studies related to data analysis, including calculating the expansion factors, estimating parameters, elaborating confidence intervals, calculating design effects, and applying multivariate statistical techniques. Studies to improve the statistical analysis are being developed in Brazil and other countries, from the perspective of using estimation of parameters by means of generalized regression models with correlated errors.

Despite the difficulties in analyzing data collected with RDS, the pilot studies financed by the project showed that it is feasible to implement these sampling methods in Brazil among population subgroups most at risk for HIV. In 2007, the Department of STD, AIDS, and Viral Hepatitis launched a call for research projects adopting RDS as the sampling method for monitoring most-at-risk populations in Brazil: MSM, drug users, and female sex workers. The research projects are currently being carried out in 10 Brazilian cities under the coordination of Brazilian researchers, thus demonstrating the country’s self-sufficiency in conducting studies with probabilistic sampling methods. Studies done in Brazil with convenience sampling would rarely be accepted today.

The most important outputs generated by the transfer process include human resources training. In less than four years, some 50 professionals were trained during the transfer phase, and approximately 100 researchers are participating in the teams in the three approved projects and are ready to train new local teams and further disseminate the use of RDS.

Training of researchers allowed Brazilian scientists to participate for the first time in international debate forums on sampling methodology for populations most at risk for HIV. Application of the methods allowed a critical vision of the theoretical principles, estimation processes, and interpretation of the results. Importantly, at the international level, Brazilian researchers have coordinated studies in Portuguese-speaking African countries using the RDS methodology.

In short, there is now an installed capacity in Brazil for applying sampling methods in hard-to-reach groups, identifying each method’s advantages and disadvantages, and improving the existing methods according to the groups’ characteristics in Brazil. Meanwhile, the efforts have resulted in a new approach to groups at greatest risk of HIV infection, by the National Department of STD, AIDS, and Viral Hepatitis and the State and Municipal programs, meaning greater knowledge of the dynamics in the spread of HIV infection and allowing the development of evidence-based public policies in these population subgroups.
Resumo

O objetivo deste trabalho é descrever o processo de transferência de dois métodos de amostragem de populações sob maior risco ao HIV: RDS (respondent driven sampling) e o TSS (time space sampling). São descritas as etapas da transferência, os procedimentos metodológicos utilizados nos dez estudos-piloto conduzidos no Brasil e as lições aprendidas desse processo. O processo de transferência foi realizado em seis etapas, que envolveram desde a apresentação do estado da arte dos métodos de amostragem até a elaboração de artigo com os resultados dos estudos-piloto. Os pesquisadores apontaram dificuldades tanto na execução do campo como na análise dos dados dos estudos, independentemente do método de amostragem utilizado. Dentre os resultados mais importantes desse processo destaca-se a capacitação de mais de 100 pesquisadores. O processo de transferência possibilitou a realização de estudos específicos para construção das linhas de base para monitoramento da epidemia nas subgrupos sob maior risco para a infecção pelo HIV, e implicou o maior conhecimento da dinâmica da disseminação da infecção pelo HIV.

HIV; Populações Vulneráveis; Amostragem

Contributors

A. Barbosa Júnior was the article’s principal mentor and prepared the text. A. R. P. Pascom participated in drafting the methodology and results. C. L. Szwarcwald participated in the article’s conceptualization and in writing the introduction and discussion. C. Kendall participated in writing the discussion and revision of the final article. W. McFarland participated in the elaboration of the introduction and discussion.

Acknowledgments

The authors wish to acknowledge the following persons that were actively engaged in the technology transfer process: Bill Brady, Suzanne Westman, Keith Sabin, Eric Voltz, Henry Raymond, John M. Karon, Lisa Johnston, Sanny Chen, Cyprian Wejnert, Elizabeth Moreira; Marcos dos Santos, principal investigators, and teams of the pilot projects; and Marcos Cleuton de Oliveira.

References


