Survey methods for hard-to-reach populations: introduction to the special issue

Maryse Marpsat\textsuperscript{a} and Nicolas Razafindratsima\textsuperscript{b}

\textsuperscript{a}Institut National de la Statistique et des Études Économiques
\textsuperscript{b}Institut National d’Études Démographiques

Abstract

Surveys of hard to reach populations (rare, no known sampling frames) have been, for some years, the object of methodological reflection. Various methods aiming at the production of an ‘extrapolable’ sample of these populations have been proposed: time-space sampling (TSS) or time-location sampling (TLS), respondent driven sampling (RDS), or the ‘capture - recapture’ method. After defining what a hard-to-reach-population is, this article provides an outline of these various approaches before going on to briefly consider the papers contained in this special issue.

Keywords: Hard-to-reach populations; surveys; time-location sampling (TLS); time-space sampling (TSS); respondent driven sampling (RDS); capture-recapture

Introduction

For several years, surveys carried out on hard-to-reach populations have provided subject matter for reflections and seminars - such as the one organised by Statistique Canada [Statistics Canada] in 2004 - or special issues of journals - like those which the journal Population dedicated to methods of collection and analysis for surveys on sensitive issues (No. 2 in 1999 and No. 6 in 2001). This interest is linked to concerns about a certain number of ‘social problems’, such as extreme poverty (homeless persons) or the AIDS epidemic (persons engaging in practices contributing to the spread of the disease such as injecting drugs without taking precautions or having unprotected sex with multiple partners) or the difficulties that persons of foreign origin (immigrants or children of immigrants) may face.

A renewal of interest in social inequalities also leads to the inclusion among hard-to-reach populations of persons with very high income or assets (see the work of Monique Pinçon-Charlot and Michel Pinçon, and the difficulty of obtaining statistical information about the assets or income of the wealthiest people). The methods developed in this way may also be applied to other hard-to-reach populations, whether or not they are stigmatised or rare, such as the jazz musicians studied by Douglas Heckathorn and Joan Jeffri (2001) or tourists (Deville and Maumy-Bertrand, 2005; 2006).

For many years, the Institut National d’Études Démographiques (INED [National Institute of Demographic Studies]) has also been carrying out innovative statistical surveys on populations that are hard to reach:
homeless persons, drug users, victims of female circumcision, etc. The methods developed or adapted by the Institute on those occasions have been borrowed by others, in particular by the Institut National de la Statistique et des Études Économiques (INSEE [National Institute of Statistics and Economic Studies]) for its national survey on the homeless in 2001. In France and abroad, other methods have been devised and tried out in the field. Relying on their expertise, the INED surveys department and the unit ‘Mobility, Housing and Social Networks’ organised an international meeting to encourage exchanges on the questions raised by research conducted on hard-to-reach populations and to help to outline critical knowledge of the chief methods developed in order to answer these questions.

The papers gathered here were presented in research seminars which took place at INED’s headquarters (Paris, France) on 19 May 20091.

What is a hard-to-reach population?

Several types of difficulty (often overlapping) have led researchers to classify a population as ‘hard to reach’:
- the population of interest has relatively low numbers, which makes an investigation throughout the general population very expensive (e.g. those fostered as children, tourists, those with a very high income, etc.);
- members of the population of interest are hard to identify; what they have in common is not easy to detect and is only rarely recorded. This is the case with some homeless persons and injecting drug users and persons who have unprotected sex with multiple partners;
- there is no sampling frame or only a very incomplete one that yields biased results, for example where a survey on the health of homeless people is conducted only at health centres;
- the persons concerned do not wish to disclose that they are members of this population of interest, because their behaviour is illicit (drugs consumption and confusion of statistical survey with police activity), because it is socially stigmatised (prostitution), because they have no desire to revisit a painful past (persons who have been in foster care), or because they refuse to allow any meddling in their affairs (persons with very high income or assets);
- the behaviour of the population of interest is not known, which leads to a poor choice of places in which to approach them (for example, surveying homeless persons only in shelters and thus not including those who sleep rough).

Among the many topics relating to hard-to-reach populations, the meeting on 19 May 2009 focused on constructing probability samples (or rather samples that can be ‘extrapolated’, i.e. that allow the production of unbiased estimators with a calculable variance), where there is no sampling frame, particularly using TLS (time-location sampling), RDS (respondent-driven sampling or probability snowballing) and capture-recapture.

Methods that make it possible to obtain a sample that can be extrapolated in the case of hard-to-reach populations

The question of sampling hard-to-reach populations partly overlaps with that of rare populations for whom there is generally no sampling frame. A review of techniques allowing this type of population to be sampled was recently undertaken by G. Kalton (2009). The two-stage sample – with numerous variations – is one of the techniques mentioned. At the first stage, a ‘filter’ survey is undertaken on a large sample by providing a simplified questionnaire that allows members of the target population to be identified. At a second stage, the sample is selected. The main disadvantage of these techniques is that they are expensive to implement and that they can only be used when the population is fairly stable and easily identifiable, which is not the case with the homeless, for example.

1 Conference presentations can be found: http://difficile_joindre.site.ined.fr/fr/accueil
Other methods, more suitable for hard-to-reach populations, are based on the principle of finding out where persons are present (in specific places, in someone’s network of acquaintances) and establishing an unbiased estimator *a posteriori* for the entire population on the basis of certain questions. These methods belong to the broad category of indirect sampling (Lavallée, 1995; 2002; 2007; Lavallée and Rivest, 2009).

**Time-location sampling**

**General idea:** where the population of interest visits a certain number of places that, conversely, are not much visited by the rest of the population, such as services providing meals, accommodation, clean clothes etc. to the homeless, these places can be sampled after an exhaustive list has been made of them, and then a sample can be taken of the persons visiting those places. Attention must be paid to the times when the population of interest visits those places. The Generalised Weight Share Method (GWSM), which is generally used in indirect sampling allows an unbiased estimator to be established by taking account of the fact that an individual may visit several places so that unequal levels of attendance between one person and another lead to different probabilities of inclusion in the sample.

The graph below gives an example of how weightings are calculated using the weight share method in the case of the National French SD2001 survey on persons visiting hostels and centres providing hot meals intended primarily for homeless persons (Ardilly and Le Blanc, 2001).

**Principle:**
- a complete list is drawn up of the places visited and the times at which they are visited (e.g. the opening hours in the case of a service providing assistance for the homeless);
- place/time pairs are obtained (e.g. services on the one hand, days on the other) from this list;
- persons in the selected places at the selected times are sampled;
- weightings that correct differences between individuals in terms of the frequency of their attendance of the places are calculated.

**Problems:**
- establishing and updating the list is often time-consuming and costly, and the information collected on the locations for the purpose of selecting the sample and collecting data is not always reliable;
- where a high proportion of the population of interest does not visit these places or does so only very rarely, this can lead to a coverage bias;
- calculating the weightings for individuals (e.g. using the weight share method) requires the inclusion of specific questions about attendance of these places. This relies on the memory and can give rise to imprecise replies. However, if this possible multiple use is not taken into account, once again the estimators will be biased;
- there may be data collection problems associated with the places: refusal by managers/owners, rapid departure of users/clients, etc.

This is the type of method that was used by INED and INSEE in surveys on the homeless (Marpsat and Firdion, 2000; Ardilly and Le Blanc, 2001; Brousse *et al.*, 2006), by InVS in its survey on drug users (Jauffret-Roustitde *et al.*, 2008; 2009) and by various authors on men having sex with other men (Pollack *et al.*, 2005; MacKellar *et al.*, 2007).
A sampling example: The arrows represent the links between the services and the individuals. The shaded services were sampled. They point to shaded individuals. Dotted lines represent the links reported by individual 7, which were not used to include the individual in the sample.

Source: Ardilly, Le Blanc, 2001, “Sampling and weighting a survey of homeless persons: a French example”. Survey Methodology, vol 27, n°1, 109-118.
Respondent driven sampling

The RDS method (respondent driven sampling), was introduced by Heckathorn in 1997 in connection with a research action aimed at AIDS prevention among injecting drug users, the project Eastern Connecticut Health Outreach (ECHO). It resembles the snowballing method but under certain conditions allows unbiased estimators to be produced. Let us recall the principle of snowball sampling which, applied in its pure form, runs the risk of producing a sample from which extrapolation is not possible. To begin with a small convenience sample of the population to be studied is accessed. The initial sample then facilitates contact with more people of the right kind to interview and then they in turn provide access to others and so on.

The general idea: in the RDS method the persons initially interviewed receive a limited number of coupons (three or four) which they use to recruit others. Whenever a person is recruited, the recruiter is paid. The coupon has two parts and the recruiter will present his part of the coupon when he comes to claim his remuneration. At that point it can be checked that the other part corresponds to a questionnaire that has been filled in. The person recruited – who is also paid for filling out the questionnaire – in turn receives the same number of coupons, etc. A higher sum can be paid for the recruitment of persons with certain characteristics (e.g. women). The survey is stopped when the size of the sample has been reached and the composition of the sample is stable in terms of those characteristics that form the subject matter of the research. The sample can be collected over a fairly long period (a year in the case of each of the locations of the ECHO project mentioned above).

Unlike snowball sampling, RDS – through the use of weighting – allows unbiased estimators to be obtained. A further difference is that the participants are recruited directly by their peers apart from the first set (the so-called ‘seeds’). Those recruited thus have the freedom to accept or refuse (they are not ‘denounced’ to the researcher as in the case of the snowball sampling method, which is an advantage in the case of populations that are stigmatised or who engage in illicit practices). The incentive to take part is subject to the group’s internal control, the existence and operation of which determines the method’s success.

Principles:

Four elements must be present for a method used to constitute real RDS:

- it must be possible to trace who recruited whom (with the assistance of a number or bar code appearing on the coupon);
- to establish the weightings, it is necessary to collect information concerning the size of each person’s network;
- there must be a maximum number of possible recruitments for one and the same person, in general up to three, which reduces the recruitment differences between individuals who have a wide network and others who have only a small one (as well as taking account of the size of the network in the weightings);
- the recruiter and recruit must have known each other prior to recruitment, each being a member of the other’s social network.

The weighting is established through questions on the size of each interviewee’s social network.
Example of a sample of 18 persons starting with three ‘seeds’

Problems:

- persons with a very poor social network have a lower probability of being reached;
- it is not easy to find out the size of the recruiter’s network in the designated sub-population (even for the recruiter himself). Some solutions to this problem are proposed in a recent paper by McCormick et al. (2010);
- it is necessary that the persons know each other in their capacity as members of the population of interest (thus, two people may know each other as injecting drug users but not know that they are both seropositive);
- since refusals are made to the recruiter, it is difficult to estimate the refusal rate
- the specific data-collection organisation – remuneration etc. – may be complicated;
- it is necessary to check that the person recruited is a member of the population of interest (e.g. the presence of needle marks on an injecting drug user). This is not easy for all populations and may be intrusive;
- if the desired size of the sample is large in terms of the population of interest, recruitment is likely to be difficult for the last waves of interviewees;

Other questions arise. For example, does the hypothesis of random selection by the recruiter in his network stand? How quickly is the stability of the estimators achieved? What is the optimal method of selecting the first set of recruiters in order to achieve rapid stability?

In his groundbreaking article from 1997, Heckathorn showed that, under certain hypotheses, the composition of a sample upon completion does not depend on the initial recruiters if there has been a sufficient number of recruitment waves in order for an equilibrium to be reached (in general fewer than six). He studied the conditions needed to obtain unbiased estimators. In particular, where the population of interest contains several sub-groups (by sex, nationality, etc.), if the tendency to recruit from one’s own group (‘homophilia’) differs from one group to another, the final sample will fail to reproduce the characteristics of the population, with the error increasing in line with the degree to which a recruiter tends to recruit from his own group. In his article from 2002, Heckathorn puts forward other ways of controlling the bias and a method for calculating the precision of estimators.

The two methods mentioned above, RDS and TLS have one thing in common: the probability of each individual being selected is not known a priori and the weightings are calculated a posteriori, after collection,
based on replies to the questions on the use of services/the attendance of the places in the sampling frame (TLS) or on the size of the recruiters’ network (RDS).

**Capture-recapture or contact-recontact**

Capture-recapture methods have been used since the sixteenth century (Cowan, 1991). Their name comes from their use in estimating the numbers of a population of animals in the wild. These models were recently applied to mobile human populations such as migrant agricultural workers. There have been some applications to the homeless, mainly in the United Kingdom under the name of the contact-recontact method (Fisher *et al.*, 1994; Bloor *et al.*, 1994; Shaw *et al.*, 1996; Williams *et al.*, 1995; Williams and Cheal, 2002). The method is also applied in deviance situations, such as taking drugs or prostitution and used to evaluate the rate of cover of a census or exhaustive survey and in epidemiology, where several sources of information coming from the same population about a given disease may be encountered and the number of cases that are not identified by any of the sources is estimated by capture-recapture.

**Principle:** It is assumed that the population will be stable over the observation period, i.e. that there will not be any new members or departures of old ones. The technique is based on at least two independent observations (or sources) of this population. In order to estimate the size N of the population, we need to know: n, the number of persons in the population observed the first time (or in the first source), m, the number observed the second time (or in the second source) and M, the number of persons observed on both occasions. N is then estimated by \((n \times m) / M\). The persons have to be identified (by whatever identifier) in order to be included in M on the second occasion (Sudman *et al.*, 1988).

In the case of homeless persons, the observations are made, for example, where homeless persons make use of services. They then fill out a brief questionnaire at each visit to one of the centres observed. The questionnaires are anonymous but an identifier is formed that allows the persons re-contacted to be identified.

**Problems:** while this method’s underlying concept is simple, the hypotheses that must be met in order for the model to remain valid are fairly restrictive:

- first hypothesis: the observations must be collected in such a way that all individuals in the population have the same chance of being selected in each period (this probability of selection, which is identical from one
individual to another, may vary from one period to another). For example, in the case of the homeless, the population of interest may be extremely heterogeneous in its characteristics and also its behaviour. In order to offset this deviation from the model, each observation (or source) can then be stratified, but it is necessary to know the relevant variables.

- second hypothesis: the observations on two different occasions are independent. Several elements lead to the assumption that the same persons (those living on the sidelines of the assistance services, avoiding contact with the interviewer and managing to hide particularly effectively on the streets) are the ones who elude observation. Moreover, certain types of service direct persons taking them up to other structures and the contacts are not independent of each other.

- third hypothesis: that the reference population does not change between two contacts or captures, i.e. the population remains fixed. This hypothesis can only be relaxed if the number of ‘captures’ increases. Where there are changes in the reference population, the use of the simplest model also leads to biases.

We can in part dispense with these hypotheses in the basic model through log-linear modelling (Fisher et al., 1994; Cormack, 1989), but the data to be input into the model is not easy to establish. Another difficulty may derive from the differences in the definition (e.g. of the homeless) of homeless persons in different sources (a filter questionnaire may then improve the coherence of definitions).

It also appears that the estimations are not very precise but in the case of populations that are not well known, even a rather imprecise number can yield interesting indications.

Finally, it has proved possible to link the capture-recapture method with indirect sampling (Lavallée and Rivest, 2009).

Examples of the application of these different sampling methods

The following texts are grouped into three sections. The first two relate, in order, to the time-location sampling or TLS method, respondent-driven sampling or RDS and the capture-recapture method.

The first section includes two texts. The surveys presented there are based on similar methods: sampling of places visited by persons to be interviewed, then sampling of persons visiting those places and weighting by the generalised weight share method of Pierre Lavallée.

Martine Quaglia and Géraldine Vivier present the surveys on homeless persons that have been carried out in France since 1995. These surveys, conducted by INED and INSEE, are based on indirect sampling of the time-location variety, with the places in which the survey is conducted comprising a sample of the locations where services aimed essentially at the homeless are provided (depending on the case, shelters, kitchens, reception centres, outreach teams, etc). The step from services to persons is undertaken through a set of weightings calculated according to the weight share method. These surveys require drawing up an exhaustive list of the locations concerned and call for a high investment from interviewers who must select the persons to be interviewed (Martine Quaglia and Géraldine Vivier, Construction and field application of an indirect sampling method (time-location sampling): an example of surveys carried out on homeless persons and drug users in France. As the authors mention, these methods were taken up again in interviewing drug users in the course of the Coquelicot survey, which is described in detail in the second contribution.

In order to study the Hepatitis C virus among drug users in France, the Institut national de veille sanitaire (InVS [Health Monitoring Institute]) carried out a survey known as ANRS-Coquelicot between 2004 and 2007. Drug users (DUs) are a difficult population to define, stigmatised by the illegal practice in which they engage and thus hidden, which makes them hard to reach. The survey, which was carried out in five large urban centres, took place in care settings and in health centres, which allowed the survey to reach both DUs whose circumstances were the most precarious (through outreach teams) and the most socially integrated (through GPs). The weighting follows the weight share method. This quantitative work was supplemented by a socio-anthropological section in order to gain a more shaded understanding of the determinants of risk-taking practices (Marie Jauffret et Yann Le Strat, (Design, sampling, organising the field for TLS survey and new developments: Coquelicot, a survey of drug users)

In the second section, Lisa G. Johnston and Keith Sabin demonstrate the limits of the ‘snowball sampling’ method (due in particular to ‘homophilia’, i.e. the tendency to associate with people similar to oneself) and
show how the respondent driven sampling (RDS) allows these limitations to be overcome. After outlining this method and making recommendations on its use, they compare the results obtained from a sample using the RDS method but without weighting with those from a weighted sample. They warn of the necessity of weighting the data, which entails, in particular, recording the size of each interviewee’s network and being able to trace the links between recruited and recruiter. They also insist on the importance of preliminary qualitative investigations, which allow the initial interviewees to be selected more effectively and questions to be formulated in an appropriate manner (Lisa G. Johnston and Keith Sabin, Sampling hard-to-reach populations with Respondent Driven Sampling).

Malcolm Williams uses the capture-recapture method to estimate the number of homeless persons. He first describes the difficulties in defining homelessness and estimating the numbers of homeless persons, particularly due to their mobility – whether geographical or between different housing situations – and their relatively low numbers. He recalls the enumerations of homeless persons (those spending the night in a place not intended for habitation) in the UK and the risks of both overestimation and underestimation as well as the low level of reliability of data from enumerating persons who are visible in the streets at night. Finally, he presents the capture-recapture method and its application in Torbay and Plymouth, where the number of homeless people was estimated on the basis of their visits to various services intended for them. (Malcolm Williams, Can we measure homelessness? A critical evaluation of ‘capture-recapture’).

Finally, the third and last section is dedicated to a comparison of methods through two contributions.

The first, by Salaam Semaan, relates to the three methods mentioned above. Salaam Semaan sets out the advantages and limits of the time-location sampling and respondent-driven sampling methods. The choice of one or the other of these methods depends firstly on the characteristics of the population to be studied (in particular whether there are places that its members frequent) and secondly the project’s objectives and the resources to which the researchers have access. As regards the capture-recapture method used for estimating the size of the population concerned on the basis of multiple sources, it can use a TLS or RDS survey as one of these sources (Salaam Semaan, Time-Space Sampling and Respondent-Driven Sampling with hard-to-reach populations.)

The second contribution in the section describes the experience of the Observatoire Français des Drogues et des Toxicomanies (OFDT [French Monitoring Centre for Drugs and Drug Addiction]) and the methods they use in interviewing drug users. The combination of ethnographic and quantitative methods leads to the drug users being divided into sub-populations whose practices and visibility differ, for which reason the investigative methods used cannot be identical but must be adapted (Agnès Cadet-Taïrou, Catherine Reynaud-Maurupt, Jean-Michel Costes et Christophe Palle, Quantitative surveys in hard-to-reach populations: the experience of the ‘Observatoire français des drogues et des toxicomanies’.

We may reflect on the similarity of these methods to each other: they are all based on the idea that a person may be interviewed though being present in various spheres (lists, networks, services etc.) and that a correction should then be made to take account of his presence in various possible spheres (Lavallée et Rivest, 2009). However, the application of these methods presents a certain number of differences, inter alia in terms of organisation in the field. The gathering of these texts presenting different methods and some of the applications of these methods that have been undertaken, will, we hope, help us to gain a better understanding of the advantages and disadvantages of implementing each method, to plan preliminary qualitative investigations in order to get to know the environment to be studied better and to choose the most suitable method in each case.

Some references relating to these methods

General points on hard-to-reach populations

Riandey, B., and Quaglia, M. (2009) ‘Surveying Hard to Reach Groups’ in P. Bonnel, M. Lee-Gosselin, J. Zmud and J-L. Madre (editors) Transport survey methods: Keeping up with a changing world, Bingley: Emerald
Schiltz, M.-A. (2005) ‘Faire et défaire des groupes : l’information chiffrée sur les ‘populations difficiles à atteindre’”, Bulletin of Sociological Methodology, No. 86, pp.30-54.

**Indirect sampling**

Deville, J-C. (1999) ‘Les enquêtes par panel: en quoi différent-elles des autres enquêtes? Suivi de: comment attraper une population en se servant d’une autre’, Actes des journées de Méthodologie statistique, INSEE méthodes, No. 84-85-86, pp. 63-82.

Lavallée, P. (1995) ‘Cross-sectional weighting of longitudinal surveys of individuals and households using the weight share method’, Survey Methodology, Vol. 21, No. 1, pp. 25-32.

Lavallée, P. (2002) Le sondage indirect, ou la méthode généralisée de partage des poids, Bruxelles: Université Libre de Bruxelles et Paris: Ellipses.

Lavallée, P. (2007) Indirect sampling, New York: Springer edition.

**Time location sampling**

*(weighting through the weight share method is used in the works by INED and INSEE, and by InVS and in surveys on tourism)*

Quaglia, M., and Vivier, G. (2006) « Enquêtes auprès de populations difficiles à joindre : principes méthodologiques et adaptations pratiques », in P. Lavallée and L-P. Rivest (editors) Méthodes d’enquêtes et sondages, pratiques européene et nord-américaine, Paris : Dunod.

Quaglia, M., and Vivier, G. (2006) ‘Conducting surveys among difficult to reach populations : from theory to practice and reality’, European conference on quality in survey statistics, Cardiff, 24-26 April 2006.

**Surveys on tourism**

Deville, J-C., and Maumy-Bertrand, M. (2005) ‘Extensions de la méthode d’échantillonnage indirect et son application à l’enquête dans le tourisme: MORGOAT’, Actes des journées de Méthodologie statistique, http://jms.insee.fr.

Deville, J-C., and Maumy-Bertrand, M. (2006) ‘Extension of the indirect sampling method and its application to tourism’, Survey Methodology, Vol.32, No. 2, pp.177-185.

Deville, J-C., and Maumy-Bertrand, M. MORGOAT : enquête tourisme en Bretagne. Présentation et réalisation.

**INED surveys on the homeless**

Firdion, J-M., and Marpsat, M. (2000) « Une enquête sur échantillon représentatif des utilisateurs des services destinés aux personnes sans domicile », in M. Marpsat and J-M. Firdion (editors) La rue et le foyer. Une recherche sur les sans-domicile et les mal-logés dans les années 90, Paris: PUF/INED

Firdion, J-M., and Marpsat, M. (2007) ‘A Research Program on Homelessness in France’, Journal of Social Issues, numéro spécial sous la direction de P. Toro, International Perspectives on Homelessness in Developed Nations, Vol.63, No. 3, pp. 567-588.

Marpsat, M. (editor) (2008) The INED research on homelessness, 1993-2008, volume 1, Paris : INED, Document de travail No. 156 http://www.ined.fr/fr/ressources_documentation/publications/documents_travail/bdd/publication/1369/ ．

Marpsat, M., and Firdion, J-M. (1999) « The homeless in Paris : a representative sample survey of users of services for the homeless », in D. Avramov (editor) Coping with homelessness: issues to be tackled and best practices in Europe, Aldershot, Brookfield Id USA, Singapore, Sydney: Ashgate Publishing

Quaglia, M., Razafindratsima, N., and Sirot, A. (2003) « Point-in-time statistical surveys of the homeless population in France », CUHP, rencontre de Madrid, www.cuhp.org.
INSEE survey on the homeless (SD2001)
Ardilly, P., and Le Blanc, D. (2001) ‘Sampling and Weighting a Survey of Homeless Persons: A French Example’, *Survey Methodology*, Vol.27, No.1, pp. 109-118.
Brousse, C., Guiot de la Rochère, B., and Massé, E. (2006) *L’enquête sans-domicile 2001*, Paris : INSEE, coll. INSEE-Méthodes, No. 116.
Brousse, C., Guiot de la Rochère, B., and Massé, E. (2002) ‘L’enquête de l’Insee auprès des usagers des services d’hébergement et de distributions de repas chauds. Une méthodologie inédite pour étudier les sans domicile’, *Courrier des statistiques*, No. 104, December 2002. http://www.insee.fr/fr/ffc/docs_ffc/cs104d.pdf

INED SI2002 on the coverage achieved in the INSEE SD2001 survey
Arduin, P., Firdion, J-M., Marpsat, M., Quaglia, M., and Razafindratsima, N. (2006) « Les enquêtes complémentaires » in C. Brousse, B. Guiot De La Rochère and E. Massé, *L’enquête sans-domicile 2001*, Paris : INSEE, coll. INSEE-Méthodes, No116, pp. 193-211.
Marpsat, M., Quaglia, M., and Razafindratsima, N. (2004) « Les sans-domicile et les services itinérants», *Travaux de l’Observatoire 2003-2004*, pp. 255-290.
Quaglia, M., and Razafindratsima, N. (2004) ‘Ined’s 2002 survey of homeless people contacted by itinerant services’, *Proceedings of Statistics Canada Symposium 2004, Innovative Methods for Surveying Difficult-to-reach Populations*, Gatineau, Québec, 3-5 November.

The INVS ‘Coquelicot’ survey (HIV and HCV)
Jauffret-Roustide, M., Le Strat, Y., Razafindratsima, N., Emmanuelli, J., and Desenclos J.C. (2008) « Enquête de prévalence du VIH et du VHC chez les UD, complétée par une recherche socio anthropologique en France métropolitaine 2004-2007 » in P. Guilbert, D. Haziza, A. Ruiz-Gazen and Y. Tillé (editors) *Méthodes de sondage*, Paris: Éditions Dunod
Jauffret-Roustide, M., Le Strat, Y., Couturier, E., Thierry, D., Rondy, M., Quaglia, M., Razafindratsima, N., Emmanuelle, J., Guibert, G., Barin, F., and Desenclos, J-C. (2009) ‘A national cross-sectional study among drug-users in France: Epidemiology of HCV and highlight on practical and statistical aspects of the design’ *BMC Infectious Diseases*, Vol. 9, pp. 113-124.

Surveys on men having sex with men
Pollack L.M., Osmond D.H., Paul J.P. and Catania, J.A. (2005) ‘Evaluation of the Center for Disease Control and Prevention’s HIV behavioral surveillance of men who have sex with men: Sampling issues’, *Sexually Transmitted Diseases*, Vol. 32, No.9, pp. 581-589.
MacKellar, D.A., Gallagher K.M., Finlayson, T. Sanchez, T., Lansky, A. and Sullivan P.S. (2007) ‘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 Reports*, Vol. 122, supplement 1, pp. 39-47.

Survey on the use of drugs in the party scene
Reynaud-Maurupt, C., Chaker, S., Claverie, O., Monzel, M., Moreau C., Evrard, I. and Cadet-Taïrou, A. (2007) *Pratiques et opinions liées aux usages des substances psychoactives dans l’espace festif « musiques électroniques ». Étude de faisabilité d’une enquête quantitative en « Population cachée » à partir d’un plan de sondage ethnographiquement raisonné*, Paris : OFDT.

RDS (respondent driven sampling) surveys
Refer also to [http://www.respondentdrivensampling.org/](http://www.respondentdrivensampling.org/) and [http://www.igh.org/surveillance/other_modules.html#respondent_driven_sampling](http://www.igh.org/surveillance/other_modules.html#respondent_driven_sampling)
2006 Special issue on RDS: *Journal of Urban Health*, Vol. 83, No. 7.
Abdul-Quader, A.S., Heckathorn, D.D., McKnight, C., Bramson, H., Nemeth, C., Sabin, K., Gallagher, K., Des Jarlais, D.C. (2006) ‘Effectiveness of Respondent Driven Sampling for Recruiting Drug Users in New York City: Findings from a Pilot Study’, Journal of Urban Health, Vol. 83, No.3, pp. 459-476.

Des Jarlais D.C., Arasteh K., Semaan S., and Wood, E. (2009) ‘HIV among injecting drug users: Current epidemiology, biologic markers, respondent-driven sampling, and supervised injection facilities’, Current Opinion in HIV and AIDS, Vol 4, Issue 4, pp. 308-313.

Heckathorn, D.D. (1997) ‘Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations’, Social Problems, Vol. 44, no. 2, p. 174-199. (article which first presented the RDS method).

Heckathorn D.D. (2002) ‘Respondent-Driven Sampling II: Deriving Valid Population Estimates from Chain-Referral Samples of Hidden Populations’, Social Problems, Vol. 49, No.1, pp. 11-34. (with a calculation of standard deviations, and post stratification in order to control the differences in network sizes and clustering effects)

Heckathorn, D.D., and Jeffri, J. (2001) ‘Finding the Beat: Using Respondent-Driven Sampling to Study Jazz Musicians’, Poetics.

Heckathorn, D.D. and Rosenstein, J.E. (2002) ‘Group Solidarity as the Product of Collective Action: Creation of Solidarity in a Population of Injection Drug Users’, Advances in Group Processes.

Heckathorn, D.D., Semaan, S., Broadhead, R.S., and Hughes, J.J. (2002) ‘Extensions of respondent-driven sampling: A new approach to the study of injection drug users aged 18-25’, AIDS and Behavior, Vol. 6, No.1, pp. 55-67.

Heimer, R. (2005) ‘Critical Issues and further questions about RDS: comment on Ramirez Valles et al.’, AIDS and behavior, Vol. 9, No.4, pp. 403-408.

Johnston, L.G., Malekinejad, M., Kendall, C., Iuppa, I.M., and Rutherford, G.W. (2008) ‘Implementation Challenges to Using RDS Methodology for HIV Biological and Behavioral Surveillance : Field experiences in international settings’, AIDS and Behavior, Vol. 12, Supplement 1, pp. 131-141.

Johnston, L.G.,, R., Reza, M., Khan, S.I., Banu, S., Alam, M.S., Rahman, M., Azim, T. (2008) ‘The Effectiveness of RDS for Recruiting MSM in Dhaka, Bangladesh’, AIDS and Behavior, Vol 12, No. 2, pp. 294-304.

Malekinejad, M., Johnston, L.G., Kendall, C., Kerr, L.R.F.S., Rifkin, M.R. and Rutherford, G.W. (2008) ‘Using respondent-driven sampling methodology for HIV biological and behavioral surveillance in international settings: A systematic review’, AIDS and Behavior, Vol. 12, Supplement 1, pp. 105-130. (several examples of uses of RDS and their evaluation).

McCormick, T.H., Salganik, M.J., and Zheng, T. (2010) : "How many people do you know? : efficiently estimating personal network size", Journal of the American Statistical Association, Vol . 105, No. 489, pp. 59-70.

Salganik M.J., and Heckathorn, D.D. (2004) ‘Sampling and Estimation in Hidden Populations Using Respondent-Driven Sampling’, Sociological Methodology, Vol. 34, pp. 193-239 (a paper showing that the estimate of the number of persons concerned is unbiased).

Semaan, S., Santibanez, S., Garfein, R.S., , D.D., Des Jarlais, D.C. (2009) ‘Ethical and regulatory considerations in HIV prevention studies employing respondent-driven sampling’, International Journal of Drug Policy, No.20, pp.14-27.

Capture-recapture or contact-reccontact

Berry, B. (2007), ‘A Repeated Observation Approach for Estimating the Street Homeless Population’, Evaluation Review, Vol.31, No.2, pp.166-199.

Bloor, M., Shaw I. and Williamson, H. (1994) The feasibility of Commissioning a Mark-Recapture Study of the Prevalence of Rooflessness in Scotland, Social Research Unit, School of Social and Administrative Studies, University of Wales, College Cardiff.
Cormack, R. (1989) « Log-linear Models for Capture-recapture », *Biometrics*, No. 45, pp. 395-413.

Cowan, C.D. (1991) « Estimating Census and Survey Undercounts Through Multiple Service Contacts », *Housing Policy Debate*, Vol. 2, No. 3, pp. 869-882.

Fisher, N., Turner, S., Pugh R., Taylor C. (1994) ‘Estimating numbers of homeless and homeless mentally ill people in north east Westminster by using capture-recapture analysis’, *British Medical Journal*, Vol. 308, pp. 27-30.

Laska, E.M. and Meisner, M. (1993) « A Plant-Capture Method for Estimating the Size of a Population from a Single Sample », *Biometrics*, Vol. 49, pp. 209-220.

Mazharul, I. (2008) *Estimating the Size of Floating Sex Workers in Selected Urban Areas in Bangladesh: An Application of Capture-recapture Methods*, communication au colloque de l’IUSPP, Marrakech, September 28 -October 2.

Shaw, I., Bloor, M., and Roberts, S. (1996) *Without shelter: estimating rooflessness in Scotland*, Edinburgh : Central Research Unit, Scottish Office.

Sudman, S., Sirken M.G., and Cowan C.D. (1988) « Sampling Rare and Elusive Populations », *Science*, Vol. 40, pp. 991-996.

Vaissade, L. and Legleye S. (2009) ‘Capture-recapture estimates of the local prevalence of problem drug use in six French cities’, *European Journal of Public Health*, Vol. 19, No.1, pp. 32-37.

Williams, M., Cheal, B., and Gomez, S. (1995) *Homelessness in Plymouth: Report of research - Stage Two*. Plymouth, University of Plymouth.

Williams, M., and Cheal, B. (2002) ‘Can we measure homelessness? A critical evaluation of the method of ‘capture-recapture’’, *International Journal of Social Research Methodology*, Vol.5, No.4, pp. 313-331.

**Comparisons of methods**

Kalton, G. (2009) « Methods for oversampling rare subpopulations in social surveys », *Survey Methodology*, Vol. 35, No. 2, pp.125-141.

Kendall, C., Kerr, Ligia, R.F.S., Gondim, R.C., Werneck, G.L., Macena, R.H.M., Pontes, M.K., Johnston, L.G., Sabin, K. and McFarland, W. (2008) ‘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 and Behavior*, Vol. 12, supplement 1, pp. 97-104.

Lavallée, P., and Rivest, L-P. (2009) *Le sondage par capture-recapture et le sondage indirect*, intervention au colloque sur les méthodes de sondage en l’honneur de Jean-Claude Deville, Neuchâtel, Suisse, 24-26 June 2009.

Magnani, R., Sabin, K., Saidel, T., and Heckathorn, D.D. (2005) ‘Review of sampling hard-to-reach and hidden populations for HIV surveillance’, *AIDS*, Vol. 19, Suppl. 2, pp. S67-S72.

Semaan, S., Lauby, J., and Liebman, J. (2002) ‘Street and network sampling in evaluation studies of HIV risk-reduction interventions’, *AIDS Review*, Vol. 4, pp. 213-223. *(compares TLS and RDS).*
