

A methodological approach based on indirect sampling to survey the homeless population

Claudia De Vitiis¹, Stefano Falorsi², Francesca Inglese³, Alessandra Masi⁴,
Nicoletta Pannuzi⁵, Monica Russo⁶

Abstract

The Italian National Institute of Statistics carried out the first survey on homeless population. The survey aims at estimating the unknown size and some demographic and social characteristics of this population. The methodological strategy used to investigate homeless population could not follow the standard approaches of official statistics usually based on the use of population lists. The sample strategy for the homeless survey refers to the theory of indirect sampling, based on the use of a sampling frame indirectly related to the target population. Following the indirect sampling approach, the estimation is performed through the “weight sharing method”, based on links connecting the frame of services with homeless population.

Keywords: indirect sampling, weight sharing method, link.

1. Introduction

The survey on homeless people, conducted for the first time in Italy in 2011-2012, is an important component of a research project on the condition of people living in extreme poverty. The research project⁷, launched in 2009, was aimed at the dual purpose of building an archive of the system of formal and informal services, public and private, existing in the country aimed to meet the needs of people living in extreme poverty and studying the phenomenon of homeless people spread over the Italian territory.

To achieve these objectives three separate and successive surveys were carried out with the following purposes: (i) the construction of an archive relative to the population of centers (organizations/entities) providing services to people living in extreme poverty (2009-2010), (ii) the acquisition of detailed information on the services provided by each one of the listed centers (2010-2011), (iii) the realization of the sample survey on homeless people (2011-2012).

¹ Istat, e-mail: devitijs@istat.it

² Istat, e-mail: stfalors@istat.it

³ Istat, e-mail: fringles@istat.it

⁴ Istat, e-mail: masi@istat.it

⁵ Istat, e-mail: morusso@istat.it

⁶ Istat, e-mail: pannuzi@istat.it

⁷ The Ministries of Health and Labour and Social Policies, Istat, the Italian Federation of associations for the Homeless (fio.PSD) and Caritas took part in the project.

The views expressed in this paper are solely those of the authors and do not involve the responsibility of Istat.

The homeless population is composed of all individuals who are experiencing housing problems due to the impossibility and/or inability to obtain independently and maintain a home in the true sense (*ETHOS typology*⁸). In this population people living in public spaces (streets, shacks, abandoned cars, caravans, sheds), in night dormitories, in hostels for homeless people or temporary housing, in housing for specific social support interventions (for homeless individuals, couples and groups) are included.

The availability of information about centers providing services to people living in extreme poverty was the fundamental information for the design of the survey on homeless people. In fact, because a direct list of individuals belonging to the target population was not available, the sampling strategy could not be based on a classic direct sampling approach.

In fact, homeless people cannot be captured by means of standard household surveys, as they are not listed in population registers. Investigating this particular population requires, therefore, reaching statistical units in a different way, for example by individuating centers where they regularly go to receive the services they need. This situation can be dealt with in the general context of *hard-to-reach* population methods (Marpsat and Razafindratsima, 2010), in particular by means of sample designs classified *time-location sampling*, in which target population units are reached through the random selection of places (the centers) and instants of time.

The centers chosen for the survey are those providing night shelter and canteen services: the former are by definition services frequented exclusively by homeless people, while the latter are services frequented not only by homeless people but also by people who are living in a state of extreme poverty. However, even for the considered centers the list of users was generally not available (for some of them, especially night shelter, a nominal list existed, but these list are not stable enough), not allowing the direct selection of individuals.

For all the above reasons, the methodology adopted for the survey finds its theoretical basis in the *indirect sampling* which is founded on the idea of using a sampling frame referred to a population that is linked indirectly to the target population. In this approach, therefore, the sampling design is defined on the basis of information contained in a list that refers to a different population from that of interest, assuming an exact correspondence between the units contained in the two populations (provisions and persons) with reference to a specific time interval. An alternative approach proposed for a similar context is the Centre Sampling (Baio *et al.* 2011), which allows to select a sample of individuals from the users of centers and estimate the inclusion probabilities of sample units on the basis of center attendance profiles of interviewed units. Anyway, for the purpose of a nationwide survey in the context of official statistics, we chose to base the sampling strategy on the availability of the list of centers, constructed to this aim in the first two phases of the project, containing crucial information about provisions provided to homeless people, to be used as *reference universe* for the survey.

Thus, the sampling list is represented by provisions, meals and beds, provided by the listed services. In this perspective, the population of interest is restricted to homeless people who frequent services of night shelters and canteen services. This is a similar approach to the one followed for the first French survey on homeless people (Ardilly, Le Blanc – 2001a), which obtained satisfactory results. The method used to estimate the parameters of the population of

⁸ Fonte: http://www.feantsa.org/files/indicators_wg/ETHOS2006/EN_EthosLeaflet.pdf

interest is the *weight sharing method* which is based on the knowledge of the links between units belonging to the two populations. The links between the sample units, i.e. provisions, and survey units, homeless people, are one-to-one, in the sense that each service provided can be associated, at a given instant of time, only to a homeless person.

To ensure a correct sharing of the weights, the map of the links between persons and centers has to be known: for each interviewed person the list of all visited centers had to be collected for a fixed period of time. The length of the observation period for each person has been set in a week, so that the estimate of the number of links refers to an average week. The survey instrument to collect appropriately all the information to map the links was a daily diary, in which the places where people ate and slept in the seven days preceding the interview were collected.

The methodology used for this study presents some limitations: the first is related to the coverage of the phenomenon, because part of the population of interest, consisting in homeless people living in public spaces and which do not use services or to night shelters nor canteen services, is not surveyed; the second is due to the fact that the services, particularly the canteen, are also frequented by people who do not belong to the population of interest. Both problems can introduce bias and/or increased variability (in case of over-coverage) in the estimate of the population of interest, the first implying a problem of under-coverage of the phenomenon and the second a problem of over-coverage due to the fact that the population caught at the centers does not reflect the true target population.

The paper is organized as follows. In section 2 the theoretical context of the survey is outlined. In section 3 the preliminary survey steps are presented to describe the construction of the archive of centers. Section 4 is devoted to the description of the sampling design, while section 5 illustrates the estimation procedure. Finally section 6 reports the main results of the survey about homeless people.

2. The methodological approach

2.1 Outline

As is known, a sample survey can be carried out through the selection of a sample from a list of units belonging to the target population of interest. In the classical theory of sampling, since the selection of the sample is random, it is possible to derive the probability of selection of the sample units. This ensures that the accuracy of the results produced by the survey is assessable.

In many situations, unfortunately, a list pertaining to the target population is not available but a list that is closely linked to it exists. The lack of a sampling frame of the target population is a somewhat widespread in the national statistical institutes (such as when a list of clusters of units is available) and this is the larger context within which lies the theory of indirect sampling, in particular applied for the *hard-to-reach populations*.

In indirect sampling the list from which the sample is selected consists, in general, of a different population from the target population (Lavallée, 2007), but linked to the population of interest. Following this approach, the inclusion probabilities of units surveyed cannot be determined directly and, therefore, an indirect and peculiar way is utilized. The estimation method adopted in indirect sampling is, in fact, the weight sharing method that

uses both the inclusion probabilities of units belonging to the list and the links between the units of the sampling frame and the units of the target population. For a correct application of the method it is necessary to know the links that are established between the units of the two populations.

The estimation method is formalized in the following on the basis of the text by Lavallée (2007), but with reference to a target population not composed of clusters, or better, in which the cluster size is equal to one because homeless population consists of individuals.

2.2 Indirect sampling

In a methodological context based on “indirect sampling” the sampling population, U^A , is considered, different from the target population, U^B , for which the estimates of the parameters of interest must be produced.

From the population U^A containing N^A units, a sample s^A of size n^A is selected in order to produce estimates referred to population U^B containing N^B units.

The link between the two populations is identified through an indicator variable which can take two values:

- $l_{ik}^{AB} = 1$ if unit i belonging to U^A is linked to unit k belonging to U^B ;
- $l_{ik}^{AB} = 0$ if unit i belonging to U^A is not linked to unit k belonging to U^B .

Let π_i^A be the inclusion probability for unit i selected in the sample s^A , $\pi_i^A > 0$ for each $i \in U^A$.

For every unit i selected in the sample s^A , the unit k of U^B which has a correspondence with the unit i , $l_{ik}^{AB} = 1$, is identified. In this way the sample s^B , composed of units belonging to U^B , is obtained as it follows

$$s^B = \{k \in U^B \mid \exists i \in s^A, l_{ik}^{AB} = 1\}.$$

2.3 Estimation based on weight sharing method

In indirect sampling, the method used for calculation of weights of units is the *weight sharing method* or *generalized weight sharing method* (GWSM, Deville and Lavallée 2006). This method allows the reconstruction of the weight of the units of the sample s^B starting from the sampling weight associated to the units selected in the sample s^A . It takes into account the inclusion probability, π_i^A , of the selected sample units through a calculation based on the relationship between the units listed in the sampling frame and the units belonging to the population of interest.

If the links between the units of the two populations are not one-to-one, the procedure for estimating the parameters of interest is very complex. In this case it is necessary to take into account some circumstances: each unit $k \in U^B$ can be associated with multiple units belonging to U^A ; the population U^A can be not perfectly overlapped on U^B , in the sense that some units of U^A can be connected with not eligible units for U^B .

An incorrect reconstruction of the links between units belonging to the population U^B and the units belonging to the population U^A involves the risk of introducing biases in the estimation of the parameters of interest, caused by incorrect counts or by an alteration of the probability selection system. The existence of at least one link between all units of the target population U^B and the units of the population U^A is a necessary condition to ensure the unbiasedness of the estimator based on the weight sharing method.

The weight sharing method provides for each unit k of a final weight that is calculated in several steps:

1. the initial weight is calculated as the sum of the direct weights of the units of U^A selected in the sample s^A having non-zero link with the unit k , $I_{ik}^{AB} = 1$,

$${}_1W_k^B = \sum_{i=1}^{N^A} W_i^A t_i^A I_{ik}^{AB},$$

being

$w_i^A = 1/\pi_i^A$ and π_i^A the inclusion probability of generic unit i selected in s^A defined on the basis of the sample design;

$t_i^A = 1$ if unit $i \in s^A$ and $t_i^A = 0$ if unit $i \notin s^A$;

2. for each unit k of U^B the total number of link with units of population U^A is obtained

$$L_k^B = \sum_{i=1}^{N^A} I_{ik}^{AB}; \quad (2.1)$$

3. the final weight w_k^B is calculated as

$$w_k^B = \frac{{}_1W_k^B}{L_k^B} = \frac{1}{L_k^B} \sum_{i=1}^{N^A} W_i^A t_i^A I_{ik}^{AB}. \quad (2.2)$$

From the expression of the weight w_k^B , it is evident that the estimation of the parameters of interest of the population U^B is based on the sample s^A and on the existence of links between populations U^A and U^B . In order to estimate the total of a generic variable of interest y^B on the target population U^B ,

$$Y^B = \sum_{k \in U^B} y_k^B,$$

the following estimator can be used

$$\hat{Y}^B = \sum_{k \in U^B} w_k^B y_k^B. \quad (2.3)$$

The duality of the estimator \hat{Y}^B with respect to U^A and U^B (demonstrated in Lavallée, 2007) allows to express the estimator of the parameter Y^B as a function of the units belonging to U^A and, at the same time, as a function of the units belonging to U^B . In fact the total Y^B can be written also as

$$Y^B = \sum_{i \in U^A} \sum_{k \in U^B} \frac{I_{ik}^{AB}}{L_k^B} y_k^B. \quad (2.4)$$

This equation allows to define for each unit $i \in U^A$ the variable

$$z_i^A = \sum_{k \in U^B} \frac{I_{ik}^{AB}}{L_k^B} y_k^B, \quad (2.5)$$

from which the equality between the totals of the two variables z_i^A e y_k^B originates

$$Z^A = \sum_{i \in U^A} z_i^A = \sum_{k \in U^B} y_k^B = Y^B.$$

The total $Y^B = Z^A$ can be estimated on variable z_i^A as

$$\hat{Z}^A = \sum_{i \in U^A} w_i^A t_i^A z_i^A = \hat{Y}^B. \quad (2.6)$$

By substituting (2.5) in (2.6) it is possible to obtain

$$\hat{Z}^A = \sum_{i \in U^A} w_i^A t_i^A \sum_{k \in U^B} \frac{I_{ik}^{AB}}{L_k^B} y_k^B = \hat{Y}^B. \quad (2.7)$$

It is clear that the second summation of the formula (2.7) refers to sample s^B and therefore the estimator is

$$\hat{Y}^B = \sum_{s^B} y_k^B \left(\sum_{s^A} w_i^A t_i^A \frac{I_{ik}^{AB}}{L_k^B} \right) = \sum_{s^B} w_k^B y_k^B. \quad (2.8)$$

Properties of GWSM about unbiasedness and variance evaluation using usual Horvitz-Thompson estimators are described and proved by Deville and Lavallée (2006).

3. The first operational phases

The map of - formal or informal, public or private - services to homeless people represented the sample base for the homeless people survey. The list of services represented the sampling frame and also the places where the interviews were conducted (Pannuzi, 2009).

The map has been obtained through two phases: a) a census of the organization and services (centers) offering services to homeless people; b) an in-depth survey on the services providers in order to collect information, both quantitative and qualitative, about their users.

Because the homelessness phenomenon is mainly spread in the wider cities, the census was conducted on a sub-set of Italian municipalities, selected on the base of their demographic size. They are 158 municipalities, including all the municipalities with over 70,000 inhabitants, the provincial capitals with more than 30,000 inhabitants, and the municipalities bordering on the municipality with more than 250,000 inhabitants.

3.1 Census of organizations and services for homeless people

In the 158 municipalities, the census collected information on all the services which provide homeless people supports: for primary needs (one-off financial contributions, drugs distribution, clothing distribution, food distribution, shower facilities and personal hygiene, soup kitchen, street unit), night shelter (self-managed housing, shelter, residential communities, residential communities for night shelter, dormitories, emergency dormitories), day shelter, social secretariat, social support measures.

The purpose was to build a database which, for each detected service, contains all the necessary information: service typology, service details (number of bed spaces, average number of meals provided per day, average number of clients per day), supplier organization denomination, address, phone, possible organizations on behalf of the service is provided, organization representative, organization type. That information is being obtained by a CATI survey, through interviewers selected by fio.PSD and trained by Istat.

Starting from the information contained in the pre-existing Istat, Caritas and fio.PSD databases, the survey updated and completed the picture by adding new organizations, reported by the already interviewed organizations. The added organizations are being interviewed in the same way, with a snowball technique in order to catch the maximum number of centers, even informal, supplying services to the homeless.

3.2 Survey on service providers

Once the database of the centers was complete, the services and, hence, the organizations were surveyed by a CAPI interview. A deep reference frame on the situation of the active services and organizations on the territory for the homeless has been drawn.

The information mainly regarded: the basic organizational and service details (contact details and location); the type of organization (whether municipal or other public bodies, private, NGO, etc.); the geographical area served; the users main characteristics (age, gender, citizenship, household type, presence of any physical or mental restrictions); the service access criteria; the provision of any support to exit from the homelessness; the collection of data by the organization or the service; the funding sources and the share of resources for the homeless; the staff information; the cooperation among the services and the interactions with other organizations, especially with social and health units; the participation of the organization in workshops, seminars about the homeless problems; the client participation in the organizational activities.

The information collected at the services about daily services, number of users and homeless users, opening days and time, represented the key information for the definition of the sampling frame for the survey on homeless people.

4. The sampling design for the survey on homeless people

4.1 Objectives of the survey and reference population

The main objective of the third phase of the project, the survey on homeless people, was to estimate - at national level - the unknown size of the target population, along with some demographic and social characteristics of the homeless people. The survey was planned with the aim of measuring the extent of a phenomenon never observed in our country, if not in local contexts, and to obtain information on the process underlying the causes of social exclusion of homeless people.

The observation field of the survey was delimited within the geographical scope of the 158 municipalities, identified on the basis of population size, and all the centers (more than 800) in which night shelter and canteen services are provided, listed in the archive of centers. The services of night shelters are self-managed housing, shelters, dormitories, emergency dormitories, residential communities and semi-residential community. The canteen services include both those providing a meal for lunch and those providing a meal for dinner.

From the peculiarities of the phenomenon under observation a need arose to use a definition of the population connected to the moment in which the condition of homelessness has occurred and to define an observational period delimited in a precise interval of time. The homeless are, in fact, a population subject to constant changes that may be caused by different phenomena. It is a population that is renewed over time and from one period to another the size and composition of the population may change due to demographic shifts or as an effect of the evolution of society in which, for example, the working condition is strongly precarious. In order to identify correctly the homeless people at the survey stage, it was adopted as reference period of the occurrence of the event that led an individual to sleep in centers that provide night shelters or on the street, the month before with respect to the day of the interview. As the size of the homeless population strongly depends on the period in which the phenomenon is observed, the period of observation has been established in 30 days, a period considered long enough to ensure that the majority of homeless people used at least once the services involved in the investigation. The choice of the reference period of the survey was determined by considerations related to the organization of the network for the data collection and evaluation of experts who assumed that the part of population missed in a winter month was of limited scale. For these reasons, the period of time fixed for the survey of the phenomenon was fixed as the 30 days between November 21st and December 20th 2011.

The delimitation of observation field and reference period for the survey led to a precise definition of the target population. In fact, it consists of all the homeless people who receive at least one provision provided at one of all centers C that provide night shelters or canteen services during the fixed period of time J in the 158 municipalities constituting the territorial scope of reference.

4.2 Sampling design

The sampling list is composed of all the centers - operating in the territorial scope of reference - that provide night shelter and canteen services. The centers are the places where services are provided. The services involved in the survey were classified into three types: nightly accommodation, canteens providing meals for lunch and canteens providing meals

for dinner. In a center there may be multiple services, but in this case each of them were treated as a separate center.

To each center c ($c = 1, \dots, C$) the opening days in a month were associated, in order to build the center-day couples constituting the statistical units of the list. The services thus defined are associated to each center so as to be identified by a unique center code, ensuring the univocal correspondence between provisions and individual, as each individual can receive only one provision in each center during a single time interval (called “day”).

The sampling design adopted for the survey is a two-stage stratified random sampling: the strata are the centers, or, in other words, all centers are included and each of them represents a separate stratum of services. This choice was determined by the limited number of centers, which allowed to visit them all, and from efficiency considerations of the design, since in this way it was avoided a selection stage. The first stage units are the opening days of the centers C in the reference period of the survey J , each corresponding to a cluster of provisions (meals or beds). The final statistical units (or secondary statistical units) are represented by the provisions (each of them connected through a one-to one relationship to one individual) delivered in the opening-days defined in the list.

In order to ensure the coverage of all the days of the reference month, the allocation of the sample days to centers was made randomly, from the list of opening days of each service. The aim of the sample scheme was to achieve a sample of provisions selected with equal probabilities, since the goal of the survey was to produce estimates at national level. Moreover, information on the variables of interest that could lead the definition of a sample design of a different type (for example with non proportional allocation or selection with unequal probabilities) was not available.

4.3 Sample size and allocation

The sample size was determined in relation to the amount of monthly provisions provided to homeless people in the considered centers. This set of provisions constitute the reference universe for the survey, i.e. the population U^A used for reaching the target population U^B . The number of provisions was estimated using information on the services contained in the archive of the centers. In particular, this information concerned the amount of average daily provisions provided by the services, the opening days of the services in the month, the number of total users of the services in a year, the number of homeless people who were users in one year.

The amount of total monthly provisions was estimated through the information deriving from the CAPI second phase survey, considering the average number of daily provisions for each center and the number of opening days of services in a month, while the estimate of monthly provisions provided to homeless has been obtained by applying to the estimate of monthly provisions, the percentage of homeless users.

The sample size was fixed at about 5,400 units after an assessment based on the calculation of the expected sampling error of the estimate of the population size at national level, which was the planned territorial domain. This evaluation was performed, due to the unavailability of prior information about the variability of the individual attendance of services, on the basis of a likely and simplified conjecture on the variability of the individual number of links, assuming that the variability of the number of links is one of the main sources of variability for the estimates, together with the time variability. The

expected sampling cv for the estimate of the population size was around 3.5%.

In order to achieve a self-weighting sample, the sample size to be assigned to each center has been calculated on the basis of a fixed sample fraction defined as

$$f^A(J) = \frac{n^A(J)}{\tilde{N}^A(J)},$$

where $n^A(J)$ indicates the overall sample size and $\tilde{N}^A(J)$ the total estimated monthly number of provisions provided to the population of interest in the period J .

Having thus defined the proportional allocation of the sample between the centers, the total number of provisions attributed to the each service is closely related to the size of the center. The sample size in each center c is thus determined as

$$n_c^A(J) = \tilde{N}_c^A(J) f^A(J).$$

4.4 Assigning the number of sample days to centers

After assigning to each center the total number of provisions to be selected, $n_c^A(J)$, to be sampled in the reference month of the survey J , the number of provisions for every occasion of investigation g and the number of days the sample to be attributed to each center were defined on the basis of fixed rules.

The number of provisions to be selected for every survey occasion has been calculated taking into account the maximum number of interviews that each interviewer would be able to conduct in the selected center-day during opening hours. This number has been fixed at 4 interviews per day per interviewer. For small centers a number of interviews ranging from a minimum of one to a maximum of 4 was attributed; centers of medium size were assigned a number of interviews equal to 4; to larger centers, the number of interviews attributed goes from a minimum of 4 to a maximum of 12 (carried out by 1 to three interviewers).

The number of sample days to centers was assigned by fixing that the maximum number of occasions not exceeded 15 days. In this way two different needs were met: on the one hand to seize variability of the phenomenon through the flow of people in services during the reference period of the survey, and on the other hand to limit the presence of interviewers in services throughout the reference period of the survey.

4.5 Selection of provisions

For each selected primary unit, defined by the couple center-day, users of service were selected at random by the interviewers each time on the spot. The selection was carried out, following a systematic procedure, from the list of the users or from the waiting incoming line, allowing the replacement of the persons selected that do not belong to homeless population.

The interviewer selected at random individuals to be interviewed using a sampling interval defined on the basis of two quantities: the number of people to be interviewed and the number of daily predicted provisions for the specific center, obtained from the second phase survey. Strictly speaking, the sampling interval should be calculated on the actual number of users present on the day of the interview, but the interviewer, especially in the

canteens where the flow is not predictable, did not know this number in advance and had therefore to use “expected” information. However, in order to calculate a posteriori the exact probability of inclusion of selected provisions, at the end of each survey day the interviewer took note of the effective number of users of the service on that occasion.

4.6 Inclusion probabilities

The first stage inclusion probability of the primary units, identified by the couple consisting of the generic g sample day in the center c (during the period J), is defined as

$$\pi_{cg}^A = \frac{d_c(J)}{D_c(J)}, \quad (4.1)$$

where $d_c(J)$ is the number of sample days of the center c and $D_c(J)$ is the number of opening days of the center c in the reference period J . The probability of inclusion of the selected provisions in the center c conditional on the selection of the day g is given by

$$\pi_{i|cg}^A = \frac{n_{cg}^A}{\tilde{N}_{cg}^A}, \quad (4.2)$$

where n_{cg}^A is the number of provisions on the day selected in the center cg and \tilde{N}_{cg}^A is the total number of effective provisions provided to homeless people in the center-day cg .

Finally, the probability of inclusion of the provision i in the center-day cg is defined as

$$\pi_{cgi}^A = \pi_{cg}^A \pi_{i|cg}^A. \quad (4.3)$$

5. Estimation procedure

5.1 Outline

In the estimation phase, for the application of the weight sharing method, some preliminary operations were required to calculate the weights of the provisions in the sample s^A and for an appropriate mapping of links between the two populations.

To compute properly the weights of provisions, the inclusion probability of each provision for each sample center-day (cgi) were calculated and correction factors have been introduced to take account of total non-response, determined by the non-response of centers and survey days. Some centers, in fact, refused to collaborate to the survey (center non-response) and for some others interviewing homeless people in the selected days was not possible (day non-response). Moreover, as during the field work individuals who refused to cooperate were replaced, this last form of non-response was not treated in the estimation phase.

Besides the classic form of non-recording of information on sampling units, in the homeless survey non-response was determined also by failure to collect essential information for the reconstruction of the links between people and provisions, contained in the diary section of the questionnaire. This type of partial non-response constitutes, in fact, a problem for the identification of the link (link non-response).

The two linked populations consist the first of all provisions provided in the set of sample center-days in the period J , $U^A(J) = \bigcup_{cg} U_{cg}^A(J)$, and the second of all the homeless people who visit the set of center-days in the period J , $U^B(J) = \bigcup_{cg} U_{cg}^B(J)$.

In practice, for each sampled person the links with the centers belonging to the reference list were collected over a period of time limited to a week; this was considered, in fact, an observation period in which the behavior of a person about the attendance of the centers can be assumed regular compared for example to that observed in a single day. The collection of information over a period of a week allows to limit the link non-response, more than the observation of a longer period of time.

The links were reconstructed on the basis of information collected in the questionnaire used for the interviews, that contained a diary section in which questions were asked about the centers frequented by the individual in the week prior to the day of the interview: for each day of the week, the interviewed person indicated the center, among those considered as the reference universe, in which he/she had lunch, dinner and/or slept. The reference week is, therefore, not a fixed week but for each individual, is a sliding group of seven days. The availability of such information has made it possible to establish links between the two populations, to associate to each individual k the weights of the provisions received in the sample center-day and determine the total number of links between the individual and population of provisions.

5.2 The sampling weights of selected provisions

The first operation carried out for the construction of the weights of provisions involved the calculus of the inclusion probability of provisions provided to homeless people selected in the sample s^A . To this end some pieces of information have been used, both obtained during survey, regarding the number of provisions provided at each center-day sample cg , and collected in the second phase, regarding the information on the homeless user in centers. The importance of combining these two pieces of information is determined by the fact that the canteen services are also frequented by individuals who are not part of the population of interest. Therefore, the number \tilde{N}_{cg}^A of actual provisions delivered to homeless people in sample center-days was calculated by multiplying the number of total provisions of the specific day by the proportion of homeless users of the center.

It is worth underlining that the high variability of the influx to the centers does not guarantee that the probability of inclusion of the statistical units in the sample is constant, since the number of interviews per day is fixed a priori. This probability, therefore, may actually vary among selected days both within a center and among centers.

5.3 Total non-response treatment

As regards the lack of participation of centers, which in fact constitutes a non-response of strata, a post-stratification of direct weights has been carried out, in order to ensure that the final weights sum to the known total number of provisions provided to homeless population in a generic month. In fact, in the context of indirect sampling, when no known totals on the target population are available, the size of the reference population consists of

the total number of monthly provisions provided in the centers considered, with reference to the period of observation.

The post-strata have been defined through the analysis of data, separately for the two types of centers. For centers that provide night shelter was considered the geographical distribution of the centers and the size of the centers in terms of provisions per month, for canteen services only the size of the centers were taken into account, thus defining the classes by size in terms of provisions.

For each class the post-correction factor was calculated as the ratio between the total archive number of monthly homeless provisions (the known total) ${}_{ps}\tilde{N}^A(J)$ and the number of monthly homeless provisions of respondent centers ${}_{ps}N_R^A(J)$,

$${}_{ps}f = \frac{{}_{ps}\tilde{N}^A(J)}{{}_{ps}N_R^A(J)}.$$

The correction for first stage unit non-response of the sampled days was carried out by the calculation of a correction factor for each center, γ_{cg} , considering the effective number of days of interview $d_c^*(J)$,

$$\gamma_{cg} = \frac{d_c(J)}{d_c^*(J)}.$$

The final weight associated to the provision can be expressed in the form

$$w_{cgt}^A = {}_{ps}f \frac{1}{\pi_{cg}^A} \gamma_{cg} \frac{1}{\pi_{|cg}^A}. \quad (5.1)$$

5.4 Link reconstruction

Proper mapping of the individual links between interviewed sampling units belonging to U^B and the provisions belonging to U^A is an important and delicate phase of the estimation procedure, because through this operation it is possible associate to each unit k selected in the sample s^B the weights of the provisions linked to it.

In the survey, 20 percent of the cases has presented a problem of identification of the link (link non-response). Not all individuals respondents were able to provide the retrospective information requested about their attendance of the centers in the week preceding the interview date. The total number of required links may vary from a minimum of 1 (if the unit frequented only the center where the survey found it) to a maximum of 21 link (3 per day for 7 days). In these cases it was necessary to treat the problem to avoid the risk of multiple counting of the same people.

The imputation of missing data in the diary was obtained through a probabilistic intra-record procedure, based on the fact that: i) the behavior of homeless people in the use of services is regular; ii) the geographical characteristics and socio-demographic characteristics of homeless people with partially completed diary, mostly related to the use of services, are not significantly different than the rest of the population.

This allowed to assign to each individual the weights of the provisions associated with the sample center-day in which he has been caught and to estimate the total number of links connecting the individual to the population of provisions.

5.5 Parameters and estimation

The main parameters of interest in the target population - defined in the survey reference period J - are total of “fixed” variables, i.e. variables that do not vary over time J (age, sex, nationality, etc.).

An important parameter of interest is represented by the unknown size of the population $U^B(J)$.

The total of a generic variable of interest y^B , depending on the period J , is

$$Y^B(J) = \sum_{k \in U^B(J)} y_k^B,$$

while the unknown size $N^B(J)$ of the target population $U^B(J)$ is obtained by setting

$$y_k^B = 1, \forall k \in U^B(J).$$

Defining the application K that relates all the provisions served in the period J in the set of centers C to each individual who has received it

- K : (provisions) \rightarrow (individual)
- $i = K(i)$

the parameter of interest $Y^B(J)$ can be rewritten in the form

$$Y^B(J) = \sum_{k \in U^B(J)} y_k^B = \sum_{i \in U^A(J)} \frac{y_{K(i)}^B}{L_{K(i)}^B(J)}, \tag{5.2}$$

where $L_{K(i)}^B(J) = \sum_{i \in U^A(J)} I_{ik}^{AB}$ is the number of links of the individual k with the provisions received during the period J in centers C (Ardilly, Le Blanc – 2001a).

From (5.2) it is possible to derive that the generic variable of interest y^B takes the same value for all the provision related to the individual k , that is $K(i) = k$. The total $Y^B(J)$, rewritten in the form

$$Y^B(J) = \sum_{k \in U^B(J)} \frac{y_k^B}{L_k^B(J)} \left[\sum_{i \in U^A(J), K(i)=k} 1 \right], \tag{5.3}$$

points clearly out that the quantity in square brackets is the total number of provisions received by unit k in the period J in the set of centers C .

Writing the variable $y_{k(i)}^B$ as y_i^A , the total of variable y^B can be estimated on the population of provisions $U^A(J)$ by the transformed variable z_i^A , defined on the provisions of the population $U^A(J)$, as described in section 2,

$$z_i^A = \sum_{k \in U^B(J)} \frac{I_{ik}^{AB}}{L_k^B(J)} y_k^B \quad (5.4)$$

Rewriting (2.7) in the specific context here considered, we obtain the expression of the estimator of $Y^B(J)$ in the form

$$\hat{Y}^B(J) = \sum_{s^B} y_k^B \left(\sum_{s^A} w_i^A t_i^A \frac{I_{ik}^{AB}}{L_k^B(J)} \right) = \sum_{s^B} w_k^B y_k^B, \quad (5.5)$$

in which, for simplicity, it is omitted to express the provision index in the full form (*cgi*) relative to the two stages of sampling. For each $k \in s^B$

$$w_k^B = \frac{1}{L_k^B(J)} \sum_{s^A} w_i^A t_i^A I_{ik}^{AB},$$

where w_i^A is the weight associated with the unit center-day-provision (*cgi*) selected in the sample s^A with inclusion probability π_{cgi}^A .

Since in practice to each person is given a weekly weight, as it is based on the links referred to a week, the estimation of interest parameters related to the whole reference period, J , was only possible through a strong hypothesis about stability in the behavior of individuals among weeks, each of which can, therefore, be considered representative of the whole period. Under this hypothesis it is possible to expand the links observed for a week to the total period J by simply multiplying the number of weekly links by the number of survey weeks (Ardilly, Le Blanc – 2001b).

In conclusion, the final weight for individual k is expressed as

$$w_k^* = \frac{1}{H L_k^B(\eta)} \sum_{s^A} w_i^A t_i^A I_{ik}^{AB},$$

where

- $L_{k(i)}^B(\eta) = \sum_{i \in U^A(\eta)} I_{ik}^{AB}$
- η is the generic week
- H is the number of weeks in month J
- $U^A(\eta)$ is the part of population $U^A(J)$ restricted to week η .

5.6 Variance estimation

The variance of estimator $\hat{Y}^B(J) = \hat{Z}^A(J)$ can be calculated on sample s^A using the equivalence between y^B and transformed variable z^A defined in (5.4).

Considering the sampling design adopted for homeless people survey, variance estimation for estimator $\hat{Y}^B(J)$ can be obtained through the expression derived for a two stage sampling scheme with equal probabilities, both at first and second stage. This last condition is valid only approximately, as inclusion probabilities are nearly constant.

In order to describe the context in which this expression has been applied it is necessary to introduce the following symbolic notation (for sake of brevity the reference period of the survey J is not indicated)

N_{cg}^A Number of provisions (secondary sample units) belonging to day g (primary sample units - opening days) of center c (stratum);

n_{cg}^A Number of provisions selected belonging to day g of center c ;

z_{cgi}^A Value of the variable z^A on provision i belonging to day g of center c ;

$Z_{cg}^A = \sum_{i=1}^{N_{cg}^A} z_{cgi}^A$ Total of the variable z^A calculated in day g belonging to center c .

The estimator \hat{Z}^A defined in formula (2.6) expressed considering the two sampling stages as

$$\hat{Z}^A = \sum_{c=1}^C \sum_{g=1}^{d_c} \sum_{i=1}^{n_{cg}^A} \frac{z_{cgi}^A}{\pi_{cgi}^A}, \tag{5.6}$$

can be rewritten - substituting in (5.6) the inclusion probabilities defined in (4.3) - in the form

$$\hat{Z}^A = \sum_{c=1}^C \frac{D_c}{d_c} \sum_{g=1}^{d_c} \frac{N_{cg}^A}{n_{cg}^A} \sum_{i=1}^{n_{cg}^A} z_{cgi}^A = \sum_{c=1}^C \frac{D_c}{d_c} \sum_{g=1}^{d_c} \hat{Z}_{cg}^A = \sum_{c=1}^C \hat{Z}_c^A,$$

in which:

$$\hat{Z}_{cg}^A = \frac{N_{cg}^A}{n_{cg}^A} \sum_{i=1}^{n_{cg}^A} z_{cgi}^A \quad \text{and} \quad \hat{Z}_c^A = \frac{D_c}{d_c} \sum_{g=1}^{d_c} \hat{Z}_{cg}^A.$$

An estimator of the variance of \hat{Z}^A (Cicchitelli *et al*, 1999) is given by

$$\hat{V}(\hat{Z}^A) = \sum_{c=1}^C \hat{V}_c(\hat{Z}_c^A) = \sum_{c=1}^C D_c^2 \frac{D_c - d_c}{D_c} \frac{1}{d_c} \frac{\sum_{g=1}^{d_c} \left(\hat{Z}_{cg}^A - \sum_{g=1}^{d_c} \hat{Z}_{cg}^A / d_c \right)}{d_c - 1} + \sum_{c=1}^C \frac{D_c}{d_c} \sum_{g=1}^{d_c} (N_{cg}^A)^2 \frac{N_{cg}^A - n_{cg}^A}{N_{cg}^A} \frac{1}{n_{cg}^A} \hat{S}_{cg}^2, \tag{5.7}$$

where \hat{S}_{cg}^2 is the estimate of variance of the second stage of z^A in sample center-day cg

$$\hat{S}_{cg}^2 = \frac{1}{n_{cg}^A - 1} \sum_{i=1}^{n_{cg}^A} \left(z_{cgi}^A - \sum_{i=1}^{n_{cg}^A} z_{cgi}^A / n_{cg}^A \right)^2.$$

From (5.7) it can be observed that the variance of estimator \hat{Z}^A (equivalent to \hat{Y}^B) depends on first stage variability (deriving from selection of sample days) and on variability of links collected on individuals associated to selected provisions. Actually, the second source of sampling variability is connected to the attendance of centers of homeless people.

6. Results

6.1 Organizations and services

The results of the data collection carried out during the preliminary phases of the project (the census of organizations and the subsequent total survey on service providers, cfr section 3) show that in 158 Italian municipalities, where the survey was carried out, there were 727 organizations providing, in 2010, services to homeless people (Istat, 2011), as summarized in Table 1 where the total number of services and the total number of users are reported for different typologies of services.

These organizations acted in 1,187 locations for a total of 3,125 services, given that each location provided, on average, 2.6 services. A third of the services gives an answer to primary needs (food, clothes or personal hygiene), the 17% provides a night shelter, whereas the 4% offers day shelter. Social secretariat and social support services are very widespread on the territory (24% and 21%, respectively). The users of support services for primary needs are twenty times higher than those using night shelters and they are twice the users of social secretariat and social support services. Public organizations directly provide 14% of the total services and they reach 18% of the total users. The services provided by private organizations with public financing together with the public organizations, represent two thirds of the total services and they reach two thirds of the total users. Among the service of social secretariat and social support, public services reach around one third of the users; the figures rise to 75% and 90%, respectively, if the private organizations with public financing are added. A maximum 10% of the users are reached by public services for primary needs and for night shelter, the services provided by private sector with public financing represent a further 48% among service for primary needs and 58% among night shelters. Services located in Lombardia and Lazio reach, together, almost 40% of the national users (20% and 17% respectively); the users of services in Milano represent 63% of the users in Lombardia, whereas the municipality in Roma serves 91% of the Lazio users. Both Sicilia and Campania reach around 10% of the total users.

Table 1 - Services and users by service macro-typology – Survey on services providers - Year 2010 (percentage composition and absolute values)

	Absolute values		Percentage compositions	
	Services	Users	Services	Users
Support services for primary needs	1,061	1,305,236	34.0	49.9
Night shelter	520	76,657	16.6	2.9
Day shelter	128	47,202	4.1	1.8
Social secretariat	754	568,161	24.1	21.7
Social support services	662	618,734	21.2	23.7
Total	3,125	2,615,990	100.0	100.0

Source: ISTAT

6.2 Homeless people

The third data collection phase, consisting in interviewing the homeless people, estimated that in November and December 2011 47,648 homeless people used a canteen or night-time accommodation service at least once in the main 158 Italian municipalities (Istat, 2013). The confidence interval within which the number of homeless people may vary, with a probability of 95%, is of between 43,425 and 51,872 people, corresponding to an estimated coefficient of variation around 4.5%, evaluated through the variance estimation procedure described in section 5.6 (Istat 2014).

The estimated number of homeless people corresponds to approximately 0.2% of the population regularly registered in the municipalities covered by the survey. However, it should be noted that this group includes individuals that are not registered by town halls or who are officially resident in municipalities other than those where they actually live. The proportion of homeless people out of the total number of residents was highest in the North-west⁹, where homeless people corresponded to approximately 0.35% of the resident population, followed by the North-east with 0.27%, the Central Italy with 0.20%, the Italian islands (0.21%) and the South and the Islands (0.10%).

Most homeless people were male (86.9%) and under the age of 45 (57.9%), two thirds had a maximum level of lower secondary school education and 72.9% stated that they lived alone. The majority were foreign citizens (59.4%) and the most common countries of origin were Romania (11.5% of the total), Morocco (9.1%) and Tunisia (5.7%).

More than half the homeless people who access services (58.5%) live in the North (38.8% in the North-west and 19.7% in the North-east), just over a fifth (22.8%) in the Central Italy and only 18.8% in the South and Islands area (8.7% in the South and 10.1% in the Islands). The result in terms of geographical breakdown, however, stems from the considerably higher concentration of the population in big cities. The higher percentages observed in the North-west and the Central Italy essentially depend on the fact that Milano and Roma account for as much as 71% of the samples surveyed. As many as 44% of homeless people use services based in Roma or Milano: 27.5% in Milano and 16.4% in Roma. Less than 10% of homeless people had problems interacting with the interviewers, and were not able to answer the interview for problems associated with physical or clear disabilities (incapacity, disease or mental disability) and/or dependency issues (7.1%) and for difficulty interacting due to their limited knowledge of the Italian language (2.2%).

⁹ These geographical areas are used here and elsewhere in this paper for the purpose of brevity, and would be more fully described as follows: “in the municipalities in the North-west where the survey was conducted”.

Detailed information was collected on those capable of answering the interview, regarding socio-demographic aspects, relations with family, relatives and friends, type of employment, health conditions, use of services and main source of subsistence¹⁰.

On average, homeless people were aged 42.2; only 5.3% were over 64. Foreign citizens are younger than Italians (36.9 against 49.9 years). The age difference also meant that the duration of homelessness tended to be higher among Italians: around half the foreign citizens (49.7%) had been homeless for less than six months, against a third (32%) of Italians; while "only" 9.3% had been homeless for at least four years, against a quarter (24%) of Italians. Overall, the result was an average duration of homelessness of 2.5 years, lower for foreign citizens (1.6 years) and higher for Italians (3.9 years).

The fact that homeless foreign citizens tended to be younger was also linked to higher average levels of education: as many as 43.1% had at least a secondary school diploma (9.3% had a university degree) against 23.1% of Italians; nonetheless, 6.1% of foreign citizens stated that they were illiterate. More than half the Italian citizens (51.5%) had no more than compulsory schooling (lower secondary school diploma).

Among the homeless, 7.5% stated that they had never had a home; of these, similar numbers stated that before becoming homeless, they had lived with friends and/or relatives, in a travelers' camp or similar or who lived in shared lodgings, institutions for minors, disabled or other. These were mainly foreign citizens (72.3%) or younger people (the average age was of 37.4); 28.8% had been homeless for at least two years, 58.5% lived alone and 30.7% with friends or relatives.

Before becoming homeless, 63.9% lived in their own home, a percentage which rose to 73.2% among Italians. Out of the foreign citizens, 20% were already homeless before arriving in Italy. More than a quarter (28.3%) are in employment particularly occasional or temporary work (24.5%) and in low-qualified jobs. 71.7% of homeless people did not work at all, more than half of homeless people (51.5%) stated that they did not work because they couldn't find a job. Only 6.7% had never had a job (a quarter of these were female, two thirds were foreign citizens and under the age of 35). The majority of homeless people (53.4%) receive financial aid from the support network of family, friends or volunteer associations, which in many cases represent their only source of subsistence; 17.9% of homeless people did not have any source of income. The loss of employment was one of the most important factors in the gradual process of social exclusion that leads to "homelessness", along with separation from spouses and/or children and, to a more limited extent, health issues. As many as 61.9% of homeless people had lost a stable employment position, 59.5% had separated from their spouse and/or children and 16.2% stated they were in bad or very bad health. Moreover, very few had not experienced any or only one of these events, confirming the fact that homelessness is caused by a combination of factors.

The lifestyle of homeless people is reflected in the fact that three quarters lived alone (78.3% of Italians and 71.9% of foreign citizens); the foreign citizens were more likely to live with family other than spouses or children, or with friends (20.5% against 12.1%); while only a very small number lived with their partner, spouse and/or children.

¹⁰ The analyses presented below refer only to homeless people capable of answering the interview.

As many as 78.3% of foreign citizens stated that they were in contact with a family member: however, 35.5% had contact only through the internet or by telephone or letter (essentially with parents, spouses and/or children), only 42.8% stated that they managed to see their family and 21.5% did so less than once a year. The number of Italians who had family contact fell to 58.6%; nonetheless, 50.8% stated that they did see them (only 7.8% did so through the internet, by telephone or letter), 8.8% saw them less than once a year and 14% saw them at least once a week and the same number one or more times each month.

In terms of maintaining relations with people who have their own home, “living” in their country of origin only appeared to have limited benefits for Italians: 76.2% stated they had friends and 66% had non-homeless friends; for foreign citizens, the percentages were 71% and 57% respectively.

In the 12 months prior to the interview, in addition to the services provided where the interview took place, 89.4% of homeless people had used at least one canteen service, 71.2% had used a night-time shelter and 63.1% a shower and personal hygiene service (with lower percentages for the use of medical services, day-time shelters and street units). The foreign citizens made more use of canteen (91.3% against 86.5%) and personal hygiene services (67.5% against 56.7%), also due to the greater frequency with which they were forced to sleep on the street, in other public spaces or in make-shift lodgings.

Nearly half (45%) of the homeless people had used employment services (without any substantial differences between Italians and foreign citizens), while Italians tended to make greater use of social services (53.7% against 30.3% of foreign citizens) and health services (64.1% against 48.2%). In the month before the interview, 61.3% of homeless people had used a night-time shelter and 24.4% had also used a day-time shelter: 41% were forced to sleep in an outdoor public space at least once and 26.7% in an indoor public space; around a quarter had slept in a vehicle, shack or abandoned building. Foreign citizens, more than Italians, were more likely to have been forced to sleep in public spaces (73.5% against 59.1%) or make-shift lodgings (48.7% against 39.0%).

Table 2 - Homeless people¹¹ by citizenship and other characteristics - Year 2011 (percentage composition and absolute values)

	Foreign citizen	Italian	Total
<i>Gender</i>			
Male	87.6	86.2	87.0
Female	12.4	13.9	13.0
<i>Age group</i>			
18-34	46.5	10.4	31.8
35-44	27.7	22.0	25.3
45-54	17.4	30.3	22.7
55-64	7.0	26.5	14.9
65 and over	-	10.9	5.3
<i>Who do you live with?</i>			
Alone	71.9	78.3	74.5
With children and/or spouse/partner	7.6	-	8.4
With other family and/or friends	20.5	12.1	17.1
<i>Where they lived before becoming homeless</i>			
At home	57.5	73.2	63.9
With relatives or friends	18.7	11.5	15.8
Other	23.7	15.3	20.3
<i>Employment</i>			
Employed	27.8	29.2	28.3
Unemployed	72.2	70.8	71.7
- Never been employed	7.7	5.4	6.7
<i>Type of event</i>			
Disease	13.7	19.8	16.2
Separation from spouse and/or children	54.4	67.0	59.5
Loss of stable employment	55.9	70.6	61.9
Total (=100%)	25,658	17,561	43,219

Source: ISTAT

7. Concluding remarks

The adopted approach represents an important innovation for the Italian official statistics because of two main reasons: for the first time the homeless population is surveyed at national level on the whole Italian territory and a new methodological instrument, such as the indirect sampling, is experimented for a large scale survey, obtaining very encouraging results. The analysis of the outcome of the methodological approach has highlighted some features of the implementation which can be taken into account for the improving of future realization of the survey on homeless, which will be carried out in 2014. In particular, one relevant aspect is that for future applications it will be possible to exploit the information about variability of links, associated to interviewed individual, and of the flow of users, in order to plan a more efficient sampling design. Another important aspect to improve is the user selection procedure, especially in the canteens, as the analysis of contact reports showed some difficulties in the selection of homeless people among all users, realized through the replacement of non-homeless persons. These critical aspects are examples of what can be improved on the basis of past experience. Other limitations of the methodological approach, such as coverage issues, can be overcome only by conducting control investigations using different survey techniques.

¹¹ Net of people with interaction difficulties during the interview.

References

- Ardilly, P. and D. Le Blanc. 2001. Sampling and weighting a survey of homeless persons: a French example. *Survey Methodology*. Vol. 27, n.1, 109-118 (2001a).
- Ardilly, P. and D. Le Blanc. 2001. "Échantillonnage et pondération d'une enquête auprès de personnes sans domicile: un exemple français" *Techniques d'enquête*. Vol. 27, n. 1, pp. 117-127 (2001b).
- Baio, G., Blangiardo, G.C. and M. Blangiardo. 2011. Centre Sampling Technique in Foreign Migration Surveys: a Methodological Note. *Journal of Official Statistics*. Vol. 27, n. 3, pp. 451-465.
- Cicchitelli G., A. Herzel and G. E. Montanari. 1992. *Il campionamento statistico*. Bologna: il Mulino.
- De Vitiis C., S. Falorsi, F. Inglese and M. Russo. 2011. "Indirect sampling in the First Italian Survey on Homeless Population". Paper presented at the ITACOSM Conference 2011, Pisa, 27-29 June.
- De Vitiis C., S. Falorsi, F. Inglese, and M. Russo. 2012. "Estimating the Homeless Population through Indirect Sampling and Weight Sharing Method". Paper presented at the SIS Conference 2012, Roma, 19-21 June.
- Deville J.C. and P. Lavallée. 2006. Indirect Sampling: The Foundation of the Generalized Weight Share method. *Survey Methodology*. Vol. 32, n. 2, 165-176.
- Istat. 2011. *Services to homeless people*. 21 november 2011, Available at <http://www.istat.it/en/archive/45837>.
- Istat. 2011. *I servizi alle persone senza dimora*. 3 novembre 2011, Available at <http://www.istat.it/it/archivio/44096>.
- Istat. 2013. *The homeless*. 10 June 2013, Available at <http://www.istat.it/en/archive/92503>.
- Istat. 2014. *The national research on homeless people conditions*. 2014. <http://www.istat.it/it/archivio/127256>.
- 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. 2007. *Indirect Sampling*, Springer, New York.
- Marpsat M, Razafindratsima N. 2010 'Survey methods for hard-to-reach populations: introduction to the special issue' *Methodological Innovation Online*.; 5(2):3-16
- Pannuzi N., A. Masi and I. Siciliani. 2009. *Surveys of homeless persons - the case of Italy, Counting the homeless*. Peer Review, 12-13 November.
- Xiaojian X. and P. Lavallée. 2009. 'Treatments for link nonresponse in indirect sampling', *Survey Methodology*. Vol. 35, n.2, 153-164.