Evaluations on list undercoverage bias and possible solutions: the case of ISTAT CATI survey "Trips, holidays and daily life"

Claudia De Vitiis, Paolo Righi[1]

Abstract

The list of fixed line telephone numbers, used as sampling frame for CATI surveys on households, suffers of an increasing undercoverage rate. Some methodological studies conducted on ISTAT CATI household surveys have pointed out the risk of biased estimates for particular phenomena of interest due to this type of frame imperfections. A research group in Istat has been set up and has highlighted a risk of estimation bias due to the undercoverage effect for the Trips, holidays and daily life survey, for which data are collected by means of CATI technique on a sample selected from the phone directories. In the paper two possible approaches to pursue unbiased estimates are described. The first one is based on the enhancement of the estimator aiming at taking into account the frame imperfections. An experiment on real survey data shows a good improvement on the undercoverage bias of the estimates choosing more properly the calibration variables. The second approach defines a different survey strategy, based on a different frame not suffering from undercoverage (or most likely less affected by undercoverage), the set of municipal population registers, together with a new sampling design and a mixed mode data collection.

Keywords: Frame Undercoverage, CATI Surveys, Sample Design and Estimation, Official Statistics

1. Introduction

In recent years the fixed line telephone has lost increasing quotas of population coverage, as shown in several studies about U.S. (Blumberg *et al*. 2008) and the European Countries (Kuusela *et al*., 2008). In Italy the overall percentage of households covered by a landline telephone has decreased from 84.7 in 2001 to 71.9 in 2008, while the mobile phone only households increased from 10.2 in 2001 to 25.6 in 2008 (ISTAT, 2009b and 2009c). Moreover, the rate of non fixed line households is not uniform across population (Callegaro and Poggio, 2005; Istat, 2009a).

These circumstances produce relevant implications for the household sampling surveys carried out by means of the CATI (Computer Assisted Telephone Interview) technique, which are commonly based on a sample selected from the list of fixed line telephone. Telephone surveys are a cost-effective and timely method for conducting household surveys. Nevertheless,

[1] Italian National Statistical Institute (ISTAT) (e-mail: devitiis@istat.it, parighi@istat.it).

a major drawback of this data collection mode is that it covers only households with fixed line telephone, leaving not covered by the sample the sub-population of households without telephone, with exclusive mobile phone or using reserved fixed lines numbers. The estimates produced by these surveys could be affected by a certain degree of bias if referred to the whole population, because of the undercoverage rate of the list and when the population units belonging to the list are different from the units excluded by the list with respect to the target variables (Lessler and Kalsbeek, 1992).

To deal with undercoverage bias, a research group has been established at the Italian National Statistical Institute (Istat). In order to verify and to asses this suspected bias, several analysis have been carried out using the data of the *Multipurpose survey on households: aspects of daily life* (in the following indicated as MPS, Istat 2009c). MPS is based on a sample of households selected from the municipal registers and, for this reason, it can be reasonably considered not affected by severe frame undercoverage and, therefore, producing approximately unbiased estimates. This survey collects some variables which are observed in the main CATI household surveys as well. Moreover, the MPS collects the presence of a fixed line telephone in the sample households; this fact allows to individuate in the overall sample the subsample of the households covered by fixed line telephone, which can be considered similar to the sample contactable through the CATI surveys selected from the telephone list.

The results of the research group highlighted significant differences, reported in the Istat technical report (2009a), among the estimates of some interest variables obtained on the whole MPS sample and the same estimates obtained on the subsample of the phone-owner households, consequently reachable through a survey using the telephone list. In particular, the analysis carried out by the group stressed a suspected bias for a variable that is quite similar to one of the main target variable of the CATI survey *Trips, holidays and daily life* (THS, Istat 2009d).

This paper shows the activities described in the second report of the working group Istat (2009b) and, in particular, the planning of the simulative procedure for analyzing the bias of THS and two solutions to deal with bias caused by frame undercoverage, either proposed by the authors. As far the solutions are concerned, the first one improves the calibration estimation procedure without modifying the frame of the survey (Särndal and Lundström 2005, ch. 14) and the sample selection scheme, the second one consists in a complete redesign of the sampling design, based on the use of a better population list and a mixed mode data collection technique (Brick and Lepkowski, 2008). For the second case it is relevant to stress the CATI technique still remains the preferable way of conducting interviews for the timeliness and the quality of data and for cost reasons as well.

The paper is articulated in the following sections: section 2 describes the definition of the simulative context and the analysis of the undercoverage bias arising from the experiment; section 3 reports the driving steps for defining a new estimator taking more carefully into account the undercoverage bias. The study is supported by an empirical analysis of the property of the new estimator in terms of bias. The experiment is based on MPS data, in which is observed a variable similar to the fundamental variable of interest of the THS. Section 4 is devoted to the definition of a two stage sampling design for the THS, using a more exhaustive selection list, attempting to preserve the estimates from undercoverage bias. Section 4.1 gives an empirical evaluation of the design effect of the estimates based on MPS data, as the current THS is based on a one stage sampling design and the proposed design can possibly produce a negative impact on the sampling errors. In Section 5 some conclusions and a future research path are given.

2. Bias Assessment in the Survey on Trips and Holidays

Bias estimation is quite a complex issue in the sample surveys. Sampling theory gives some general frameworks for avoiding biased estimates and some methodological tools for bounding the bias when the sampling survey is planned out of these frameworks. In the paper we take into account the randomization-assisted survey sampling approach. The inferential approach guarantees the unbiasedness when each population unit has a sample inclusion probability greater than zero. Generally, this condition does not hold when using a fixed line telephone list due to the undercoverage. Theory shows that the undercoverage produces biased estimates unless the undercoverage is MCAR (Little and Rubin, 2007). Moreover, when the undercoverage is MAR the researcher may remove the bias with suitable estimators. Usually, the undercoverage model is more complex and assuming a MAR mechanism the bias may just be bounded.

In the THS the high rate of undercoverage of the list leads to suspect that the estimates might be biased. Nevertheless, the quantifying of bias is impossible in a direct way because the parameter of interest and the undercoverage mechanism are unknown. In these cases the bias can be assessed only indirectly and in a simulative or experimental context.

The bias of the THS has been evaluated by means of an experiment based on the MPS data. The MPS sample of household is drawn according to a two stage sampling design. Data are collected by means of PAPI (Paper And Pencil Personal Interview) technique. MPS uses currently a calibration estimator benchmarking the number of persons by sex and age classes at regional level.

The choice of using MPS data has been suggested by the following aspects:

- the sample is selected from a theoretically exhaustive demographic lists (the municipal registers). In practice, the demographic lists may suffer from undercoverage (or overcoverage or duplication of the population units), but we assume the phenomenon is much less relevant with respect to the undercoverage of a telephone list. Furthermore, we assume these are the best demographic lists, currently, available. Quality enhancement of the lists should require an expensive "ad hoc" procedure;
- MPS collects many socio-economic variables (sex, age, educational level, professional position, etc.) observed also in the THS;
- MPS collects the "*yes*/*no* variable about at least one Overnight Stay lasting Four nights or more (OSF) in a collective accommodation establishment for holiday, during the last twelve months". This variable is quite similar to the most representative variable of the THS. In the experimental study we suppose that if an evidence of bias on the total estimation of the OSF variable is found, then we can infer that a bias on the estimation of THS variable is also likely;
- MPS collects the "*yes*/*no* fixed line phone owner" variable, hereinafter denoted by fixed line phone owner. Therefore, it is possible to distinguish the sub-sample of MPS reachable by a survey using CATI techniques such as in the THS.

Assuming that the municipal registers are affected only marginally by list errors, the sub-sample of fixed line phone owner is representative of the population of the fixed line phone owner households due to the randomization-assisted (or design-based) approach used in the MPS sample strategy. That implies that the estimated frequencies on the fixed line phone owner sub-sample are unbiased estimates of the fixed line phone owner population.

The bias assessment of the OSF estimates is carried out in a simulative context comparing the MPS estimates with the estimates based on the MPS phone owner sub-sample and the THS estimator.[2] Hence, we do not use the current estimates of the MPS but we consider a simulated estimation process performed currently in THS.

In this way, assuming that the MPS produces in general unbiased estimates (hereinafter denoted by reference estimates), the comparison may suggest that the possible differences depend on coverage and estimator only, because we use data which are equal regarding all the other conditions. Moreover, by the randomization-assisted approach the THS estimator produce unbiased estimates when there are not frame errors. Therefore, in the proposed comparison significant differences between couples of estimates should depend on the undercoverage problem of phone owner sub-population. We underline that the THS presents some substantial divergences with respect to MPS not only regarding the frame and estimator. The survey uses different sample design, sample size, survey techniques and a slightly dissimilar definition of the interest variable. If we had compared directly the MPS and THS estimates, the possible significant disagreements could have derived from a lot of causes such as: unlike definition of the interest variable; different survey techniques (respectively PAPI and CATI) having unequal effect on measurement errors and on rate of non response. Basing the analysis on such data it would be difficult to detect the bias due to undercoverage.

A final assumption necessary to obtain the simulative comparison is that the response probability in the MPS must not depend on fixed line phone owner variables. We believe that the assumption is good enough in the sense that is negligible the dependence of the household non response from the fixed phone owner variables.

Table 1 shows the results of the simulative comparison.

Table 1 - Estimates of relative (%) frequency distributions of the number of person per at least one Overnight Stay lasting Four nights or more (OSF). Estimates obtained by means of the MPS estimator on the overall MPS sample (type a estimate) and by means of the current THS estimator on the MPS sub-sample of the phone owner households (type b estimate) (year 2008 data)

OSF	*Type a* estimates and confidence interval		*Type b* estimates and confidence interval	
No	48.7	(47.8-49.5)	44.4	(43.4-45.5)
Yes	50.3	(49.5-51.2)	54.6	(53.5-55.7)
No-response	1.0	(0.9-1.2)	1.0	(0.8-1.2)

Souce: Elaborations on Istat data

The main results are the not overlapping of the confidence intervals computed by means of the two estimation procedures. We explain them as a sign of biased estimates due to undercoverage. These findings have led to define two approaches (see section 3 and 4) for harmonizing the compared estimates.

2 The THS current estimator is calibrated with respect to the following known totals: population by sex and region; population by municipal type; population by age groups; households by size.

3. Calibration approach

The first approach to deal with the bias deriving from the undercoverage of the telephone list leaves substantially unchanged the survey design (sampling design and the selection frame) and aims at improving the estimation procedure. This is done by introducing in the estimator, which is generally a calibration estimator, additional or different auxiliary variables, more correlated with the survey variables or with the undercoverage phenomenon. The approach assumes the existence of a relationship among the target variables and some auxiliary variables observed in the sample units and for which either the totals are known at population level or unbiased estimates with small sampling variance are available. We introduce these variables in the calibration procedure, modifying the current calibration estimator. Adding these benchmarking variables, the final estimates of the totals of the new auxiliary variables will be equal to the known (or estimated) totals.

The use of the calibration approach is one of the possible ways to deal with undercoverage at the estimation phase (Särndal and Lundström 2005, ch. 14). The approach is based on the following assumption: if the new estimator corrects the biased sampling distribution of the explaining (auxiliary) variables of the target variable, then the estimator corrects, at least partially, the bias of the estimates of the target variable as well. Another standard reweighting approach is the propensity score weighting method (Rosembaum and Rubin, 1983; Lee and Valliant, 2008). The approach usually relies on logistic regression modelling and is sometimes referred as logistic regression weighting. Propensity score method requires the knowledge of the auxiliary variables for the overall population and this aspect may restrict its use. On the contrary, calibration approach needs to know the auxiliary variables at sample level and the totals at population level. Cobben and Bethlehem (2005) made an interesting and quite exhaustive application of the two approaches on real survey data.

Starting from the findings of table 1, the calibration approach has been tested on the MPS data referred to 2008. Before defining a new estimator, the following steps have been performed:

- a logistic regression model fitted on MPS data has identified some significant socio-economic variables in explaining the probability to have yes for the OSF variable (Istat 2009a);
- the estimates of the sampling distribution of the significant socio-economic variables obtained on the sub-sample of fixed line telephone owner households were computed using the current THS estimator;
- the comparison among the estimates computed with the current MPS estimator based on the overall household sample and the estimates with the THS estimator on the fixed line telephone owner households has been performed by means of a Chi-square index and taking into account the confidence intervals overlapping;
- finally, some socio-economic variables presenting different estimated distributions when obtained with MPS and THS estimators are introduced in the system of constraints of the THS calibration estimator if they are not already used in the current estimator.

The analysis of the logistic model (accomplished by the research group, Istat 2009a), outlines that there are several variables which are explicative for the yes probability of OSF variable. Therefore, a selection process of the variables to be taken into account has been necessary. Otherwise, with so many benchmarking variables, the calibration procedure would become unfeasible. Firstly the variables not collected in the THS such as income level, possession of durable goods, type of dwelling have been discarded. Then, the variables, marital status, household size, nationality, that do not show significant differences in the estimated distributions by the two compared sampling strategies have been left out. Moreover, the age of the head of the family has not been considered: even though it is a significant variable, the logistic model evidences have a complex interpretation with respect to the connection with the probability of yes for the OSF variable. Hence, the preliminary steps have leaded to include the variables Educational Level (EL) and Professional Position (PP) in the set of the additional calibration variables.

Table 2 and 3 show the estimated distributions of the selected variables obtained by means of the overall MPS sample using the current MPS estimator and the MPS sub-sample of the fixed phone owner using THS estimator. The estimated distributions obtained through the two estimators appear unlike.

Table 2 - Estimates of relative (%) frequency distributions of Educational Level. Estimates obtained by means of the MPS estimator on the overall MPS sample (*type a* estimate) and by means of the current THS estimator on the MPS sub-sample of the phone owner households (*type b* estimate) (year 2008 data)

EDUCATIONAL LEVEL	Type *a* estimates and confidence interval		Type *b* estimates and confidence interval	
Doctorate, degree	8.9	(8.5-9.2)	10.8	(10.1-11.5)
Upper secondary school certificate	26.0	(25.4-26.6)	30.1	(29.0-31.2)
Lower secondary school certificate	35.3	(34.7-35.9)	32.7	(31.6-33.8)
Primary school certificate, no education	29.9	(29.4-30.4)	26.4	(25.4-27.4)

Souce: Elaborations on Istat data

Table 3 - Estimates of relative (%) frequency distributions of Professional Position. Estimates obtained by means of the MPS estimator on the overall MPS sample (*type a* estimate) and by means of the current THS estimator on the MPS sub-sample of the phone owner households (*type b* estimate) (year 2008 data)

PROFESSIONAL POSITION	*Type a* estimate and confidence interval		*Type b* estimate and confidence interval	
Manager	1.0	(0.9-1.2)	0.7	(0.5-0.9)
Executive / Clerk	14.2	(13.8-14.6)	23.5	(22.5-24.5)
Workman / workwoman	12.6	(12.2-13.0)	15.4	(14.6-16.2)
Entrepreneur, Professional	3.3	(3.1-3.5)	2.9	(2.5-3.3)
Coordinated free-lance worker	1.3	(1.2-1.4)	-	-
Self-employed, Collaborator in the family Business	5.7	(5.4-6.0)	5.1	(4.6-5.6)
Not employed	61.8	(61.3-62.2)	52.3	(50.9-53.7)

Souce: Elaborations on Istat data

In many cases, as the confidence intervals do not overlap, we assume that estimated frequencies are significantly different. Furthermore, we assume the estimates based on the overall sample of MPS as unbiased and by the consequence the estimates based on the MPS sub-sample and THS estimator as biased. It is worthwhile to note that these are the only distributive evidences for the variables EL and PP and no indication at population level are available.

The EL and PP variables have been introduced in the set of the calibration variables of the THS estimator.

Table 4 shows the estimated distribution of the OSF variable having introduced the EL variable in the calibration system, based on the sub-sample of the phone owner households (*type c* estimated distribution). The distribution is more similar to the reference estimates with respect to the estimated distribution based on the sub-sample of the phone owner households and the current THS estimator (*type b* estimated distribution of table 1). Table 4 also shows the results of the calibration process on the sub-sample of phone owner households when EL and PP are in the set of calibration variables (*type d* estimated distribution). We may observe a further reduction of the bias with confidence intervals almost overlapping the confidence intervals of the reference estimates. Therefore, table 4 shows that significant improvements of undercoverage bias reduction by using a suitable calibration procedure are possible.

Table 4 - Estimates of relative (%) frequency distributions of the number of person per "at least one Overnight Stay lasting Four nights or more" (OSF), obtained on the MPS sub-sample of the phone owner calibrating on the auxiliary variables currently used in the THS estimator plus EL variable (*type c* estimate) and plus EL and PP variables (*type d* estimate) (year 2008 data)

OSF	*Type c* estimate and confidence interval		*Type d* estimate and confidence interval	
No	46.0	(45.0-47.0)	46.4	(45.4-47.4)
Yes	53.0	(52.0-54.0)	52.6	(51.6-53.6)
Non-response	1.0	(0.8-1.2)	1.0	(0.8-1.2)

Souce: Elaborations on Istat data

Finally, we have to point out that table 4 gives downward biased estimation of the sampling errors of the type c and d estimators, because they are computed assuming the totals of the new benchmarking variables as they were known while they are estimated. Nevertheless, we underline that unbiased variance estimations in table 4 would enlarge the confidence intervals generating a greater overlapping among the confidence intervals obtained with the estimates based on the MPS overall sample and the MPS estimator.

The proposed approach refers to the sample weighting adjustment methods (Kalton and Kasprzyk, 1986). The aim of these methods is to reduce the undercoverage bias. A drawback is the increasing of the variances of the survey estimates. A second issue to deal with regards the possibility to utilize such an estimator in the current THS, which is based on a relatively small sample, on which the convergence of the calibration procedure is not guaranteed. The trade-off between bias and variance and convergence are not faced in the paper, but it will have to be taken into account in an application of the approach.

4. Redesigning a CATI survey: selection from municipal registers by a two stage scheme and mixed data collection technique

In the survey context under examination, an alternative approach to deal with the declining coverage of the fixed telephone list is the use of a different selection frame, guaranteeing a very high level of coverage of the target population. At present, the only list covering theoretically the whole Italian population is the set of municipal registers available at each Italian municipality. Even though this list may suffer from some coverage error (both over and under coverage), we can assume that these failures are much less relevant than those encountered in the telephone list. The municipal lists are currently used for the household surveys carried out through face to face interview. If THS will be based on these lists a new sampling strategy is required. In fact, as a unique list of households is not available and each municipal register is obtainable only in each municipality, the sampling selection scheme has to be similar to those ones commonly used at Istat for household survey with selection from municipal register: a two stage sample scheme, in which the municipalities are the PSUs (primary sampling units) and the households the SSUs (secondary sampling units).

After the selection of the sample households the proposed approach requires a delicate phase of linking the sample units to the list of fixed telephone numbers. The aim is to contact the sample units by means of the telephone as much as possible. In this way multi frame approach, together with a multi-mode collection technique is proposed.

The approach resorting to such a different frame will imply the definition of a different sampling scheme and the use of different interviewing technique to be added to the usual CATI method for those units not reachable through a fixed telephone. Nevertheless, the CATI technique remains the main interviewing mode for its cost effectiveness.

This solution is suggested in Nathan (2001), while a relevant review of multi frame and multi-mode issues is reported in Lepkowski et al. (2008).

The linkage phase will produce a positive result only for a part of the sample households, both for errors of linkage and for real absence of a fixed line. At present, evidences from the Istat Labour Force Survey (Istat 2009b) show that in general the linkage rate between a list of households (selected for the survey from the municipal registers) and the fixed line telephone directories is around 40% at national level, being slightly higher in the North and Center of Italy, while it is lower in the region of the South.

While for the sub-sample of units linked to a fixed line number the CATI mode can be applied, for the remaining ones the interview has to be carried out through a different technique: by CAPI (Computer assisted Personal Interview), which it is the most expensive one, or, more conveniently, the CAWI (Computer Assisted Web Interview). The CAPI technique will be used also for those households not reachable through the linked telephone numbers, in order to avoid the substitution of such units, which is a not recommended practice if the aim is to reduce the bias of the estimates.

In some case an attempt can be done in order to get from the non-linked households the number of a fixed line, asking it directly to them by means of a letter sent to their address. In this way a part of the “uncovered” households could be retrieved and they could be interviewed through the CATI mode. A first experiment of this approach will be performed in the Istat Adult Education Survey, which will be conducted in 2011 on a sample of 6,000 households, following this proposed mixed mode collection technique and the selection of the sample from the municipal registers.

In this context, another relevant issue under discussion is the possibility to conduct CATI interviews through mobile telephone numbers. Although this interviewing mode has been already experimented giving good performances (Bethlehem, 2009), at present a complete list of mobile telephone numbers to be linked to a sample of individuals or households is not available (mostly for legal reasons). Therefore, this possibility cannot be considered in our context and mobile interviewing can be considered only for taking appointments after a first contact through a fixed line.

The multi-mode data collection, mainly CATI mode, addresses the need for timeliness and cost reduction at the same time; using the CAPI technique for the oveall sample indeed, abandoning the telephone interview at all, would be much more expensive. On the other hand, the use of several survey techniques yields different effects on the survey estimates. For instance, because the questionnaire must be modified according to the data collection mode, or because the interviewers have a different role in the mode collection or may not be employed. The mode effect together with advantage and drawbacks of multi-mode survey are discussed in Lepkowski et al. (2008). The focus of the paper is on the evaluations of a survey redesign based on a two stage sampling design using CATI and/or CAWI when possible, with reference mainly to the sampling aspect. This is a first step for planning a multi-mode surveys but not the only one. The subsequent evaluations (questionnaire, interviewer effect, ect.) must be taken into account with field experiences. The results of the Adult Education Survey will provide useful indications.

As far as the sampling aspects are concerned, a relevant element has to be taken into account: a two stage sampling scheme produces in general an inefficiency effect due to the level of similarity of elementary units within the selected clusters, with respect to the survey variables: the more similar to each other the units belonging to the clusters, the higher the increase of inefficiency of the estimates. This effect, known as *design effect* (Kish, 1965), can be limited, in general, by selecting a large number of PSUs and therefore a small number of final units in each PSU. In the case under study, the only reason to limit the number of sample municipalities is due to the necessity to get access to each municipal register. The use of CATI technique, eliminating the need for the interviewers to go personally to the houses of the sample households, allows, on one hand, to widen the sample of PSUs obtaining an improvement of the efficiency of the estimate, and enables, on the other hand, to reduce the cost of the survey, being the CATI much cheaper than a face to face interview. In order to obtain an adjustment of the cost of the redesigned survey, it is possible to operate on three alternatives: selecting a very large number of municipalities keeping the number of sample final units similar to the current one, increasing the number of final units in the sample keeping the number of municipalities small, or balancing both numbers at the same time.

In any case, to fix the parameters of the two stage selection scheme (the size of first and second stage sample and the number of sample households for each sample municipality), it is necessary to carry out evaluations on the efficiency of the estimates to be obtained, which varies from survey to survey depending on the intracluster correlation coefficient of the specific variables.

It is useful to point out that this sampling approach presents very good advantages together with some drawbacks. The advantages are: it mostly resolves the problem of undercoverage bias assuming the municipal registers are less affected by undercoverage; it allows to limit the increase of the costs of the survey as a high quota of CATI interviews is preserved. On the other hand, the disadvantages are: the problems of linkage between sample households selected from municipal registers and the telephone list; the difficult to know in advance the quota of linked sample units while the knowledge of this subdivision of the sample is fundamental for fixing the parameters to plan the two stage sample design (number of PSUs, SSUs per PSU and total sample size); the need to handle differently the nonresponse and measurement errors according to each data collection techniques.

Furthermore, we remark that the proposed approach is based on the use of well consolidated sampling, estimation and data collection techniques in the Official Statistics. These properties do not necessarily improve estimation quality, but they are appreciated from the operative point of view, especially in large scale surveys, where the data production processes are complex and the introduction of not frequently used tools can create some troubles and can worsen the data quality.

4.1. First evaluations on a redesign of the Survey on Trips and Holidays

The THS has resulted to be exposed to the risk of bias deriving from the undercoverage of the selection list, as shown in (Istat 2009a) and in section 2. To follow the redesign approach, the use of an exhaustive population archive is required. As stated in previous section, the most natural population archive is the municipal registers, giving rise to a two stage sampling scheme and a multi-mode data collection, CATI for the part of the sample to which it is possible to link a fixed phone number and CAPI or CAWI for the other part. It is important to underline that for this particular survey the use of the a computer assisted interview is crucial, essentially for timeliness reasons, being the survey obliged to send quarterly data to EUROSTAT; in this context the CATI and the CAWI technique would be better because they would allow also to limit the survey costs.

Through the adoption of such a sample strategy it should be possible to reduce consistently the part of the total error of the estimates due to the undercoverage of the list, and maintaining at the same time the sampling error at the same level as it is currently, by increasing if necessary the overall sample size.

In this context some analysis have been carried out to obtain some evaluation about the required sample size to realize a two stage sample scheme, with the aim not to worsen the precision of the estimates of the survey. The analysis has been made in term of *design effect* of the estimate of the parameter related to the OFS variable, described in the previous section 2. The design effect is a measure of the impact on the sampling variance of an estimate deriving from the use of a complex sample design, in comparison with a simple random sample with the same sample size (Kish, 1965). For a two stage design this quantity is generally greater than unity, owing to the *intracluster correlation coefficient* ρ, which expresses the similarity among units belonging to the same cluster with respect to a given survey variable.

The expected necessity to increase the number of sample units derives from the fact that at present the survey is based on a single stage stratified sample selection, directly from the fixed phone directory, being such a design a very efficient one. For example, in general, a design effect equal to 2 would require an increase of the sample size equal to $\sqrt{2}$, passing from a simple random sample to a two stage design.

From the empirical analysis on the MPS data, which derive from a two stage sample design municipality-household, the first finding is that the ρ shows very high values within the "household" clusters, while the effect of grouping the sample households in sample municipalities is weak. In Table 5 are shown the design effect of the OSF variable in the different considered samples of MPS and THS. Surprisingly, for the analogous OSF variable collected in the THS (based on a single stage design) it can be observed a design effect higher than in MPS: 1.83 for the latter and 1.90 for the THS, both at national level for the last available survey year 2008. Going into details of MPS data separately for the self representative (SR) and non-self representative (NSR) part of the sampled municipalities, the data show that the design effect is around 1.6 for the SR part and around 2 for the NSR part. All this numerical results have been obtained by means of the generalized software GENESEES, currently used at Istat for the calculus of calibrated sample weights and estimates of sampling errors.

Table 5 - Estimates of design effect for the variable "at least one Overnight Stay lasting Four nights or more" (OSF), obtained on MPS (total sample and telephone owner sample) and THS sample data (year 2008 data)

SURVEY DATA	Design effect	Intracluster correlation coefficient	Sample size
Multipurpose Everyday Life Survey			
Total sample- *Type a* estimator	1,83	0,435	48.861
Self Representative part	1,60	1,133	16.620
Non Self Representative part	1,96	0,058	32.241
Telephone owner subsample -*Type b* estimator	1,93	0,520	30.851
Self Representative part	1,83	1,621	10.601
Non Self Representative part	1,99	0,097	20.250
Holidays and Trips (all SR)- four quarters	1,89	1,650	35.870

Souce: Elaborations on Istat data

It is useful to underline the meaning of the comparison between the design effect evaluated on the SR part of the MPS sample and the THS sample. In fact the two sample scheme are similar: THS sample is selected from the fixed telephone list by means of a stratified sample scheme and the SR part of the MPS sample is the part of the sample selected through a single stage design, in which the municipalities constitute one stratum each and are selected in the sample with certainty. The result of the evaluation of design effect states that, with a similar selection scheme, MPS design effect is much lower than the THS one. This result allows to suppose that the impact of undercoverage of the selection list produces a bias (an overestimation) of the design effect. In other words, it seems that in the part of the population covered by fixed telephone the intracluster correlation coefficient is higher than in the whole population, observed through the MPS. This evidence is confirmed by the analysis of MPS data limited to the sub-sample covered by fixed phone: the design effect estimated on this subset of households is higher than the design effect estimated on the whole sample. This result emerges both in the SR and in the NSR part of the sample. Although this analysis is valid with respect to the considered variable OSF, the result is very appealing and would deserve a deeper investigation.

In order to confirm the last findings, the quantities about the design effect have been estimated (Table 6) also with reference to a particular subpopulation, for which the sub-samples of the two surveys are comparable in a better way: the set of metropolitan cities,[3] constituting a domain of estimate for both surveys: for this domain a single stage sample selection scheme is performed in both samples. What emerges from this table is that even for the subpopulation of households living in metropolitan towns, the cluster association is higher for telephone owners and, moreover, the THS seems to over-estimates it.

[3] The metropolitan cities are: Torino, Milano, Venezia, Genova, Bologna, Firenze, Roma, Napoli, Bari, Palermo, Catania, Cagliari.

Table 6 - Estimates of design effect for the variable "at least one Overnight Stay lasting Four nights or more" (OSF), obtained on MPS and THS sample data referred to the municipalities center of metropolitan areas (year 2008 data)

SURVEY DATA	Design effect	Intracluster correlation coefficient	Sample size
Multipurpose Everyday Life Survey			
Total sample- Type a estimator	1,50	1,015	4.829
Telephone owner sub-sample -Type b estimator	1,65	1,345	2.864
Holidays and Trips (all SR)- four quarters	1,93	2,05	3.919

Souce: Elaborations on Istat data

The consequence of these first results about the design effect on the definition of the sample size for a two stage sample design is relevant. In fact, it implies that to reduce the bias produced by the use of the telephone list affected by undercoverage, it is not required an increase in sample size to keep unchanged the sampling error of the estimates. The reason of this fact is that the current THS estimates are already affected by a high design effect due to the strong similarity of the individuals belonging to same household and it is reasonable to expect that, with a two stage design with selection from municipal registers, this design effect will not rise. Therefore, by utilizing a population frame not affected by undercoverage, it would be possible to obtain a remarkable reduction of the total error of the estimates, deriving from the reduction of the bias, together with the maintaining of the level of the current sampling error.

It would be necessary, however, a further experimental phase, through a pilot survey, in order to get estimates of all the unknown parameters about variability and design effect for the other variables of the THS (number of trips, number of nights spent in the trips) not considered in the reported analysis because they are not collected in the MPS, which is at the moment the only source of reference unbiased information.

Finally, a pilot survey should be also fundamental for obtaining evidences about many others issues for the underpinning the proposed redesign. Such issues involve: the quantification of the cost according to the rate of CATI, CAPI and CAWI interviews; the data quality depending on non sampling errors generated by different questionnaires and the interviewer effect acting in the different collection modes.

5. Conclusions

International literature shows extensively that in recent years an increasing proportion of households no longer use the fixed line telephone. This trend potentially saps the reliability of the estimates produced by CATI surveys due to the undercoverage of the fixed line telephone list. In the Italian National Statistical Institute (Istat) to deal with this problem a research group has been established. The first goal of the group has been an assessment of the bias of the estimates produced by some important CATI surveys conducted by Istat. The Trips, holidays and daily life Surveys (THS) has resulted to be exposed to the risk of bias deriving from the undercoverage of the selection list. In the paper two possible approaches to achieve unbiased estimates are proposed. The first approach, maintaining unchanged the sampling design, is based on the enhancement of the estimator to take into account some auxiliary variables, related to the coverage phenomenon. The second approach defines a new sampling design assuming to use a list not suffering from undercoverage and, consequently, a different sampling scheme. We propose to use municipal registers, because although these registers are known having error list problems as well, we believe the undercoverage phenomenon is less intensive with respect to a fixed line telephone list. If the undercoverage of the municipal registers is not negligible is another issue that should be handled in case of practical redesign.

The evaluations of the bias of the THS estimates and the study of the performances for bounding the bias of the two proposed approaches have been made in a simulative and experimental context. No other ways can be run because the interest parameters and undercoverage mechanism are unknown. The experiment has been based on the data of the Multipurpose survey on households: aspects of daily life. These data allow to give some indication about the estimation bias of the most important THS parameter of interest for several reasons: the MPS observes a sample selected from a list with less undercoverage problems (municipal registers); it collects two useful variables, the presence of the fixed line telephone and the yes/no variable about at least Four Overnight Stays, which is quite similar to the most relevant interest variables of the THS. On the MPS data we have compared the estimates obtained by means of the THS estimator computed on the sub-sample of telephone owner households and the estimate obtained on the overall MPS sample with the estimator currently used in this survey. Considering the last ones as unbiased estimates, the significant differences between the two types of estimates are supposed to be caused mainly by the undercoverage of the telephone frame. The two proposed approaches have been applied and the results have been encouraging on both sides. The enhanced estimator produces a negligible difference with respect to the assumed unbiased estimates of the MPS. The analysis of the proposed redesign highlights that a multi-mode survey, privileging the CATI techniques as much as possible, is a practicable road as the redesign does not required an enlargement of the sample size. Finally, it is worthwhile to note that the two approaches are in some senses original. The use of calibration to set about the undercoverage is well known in literature, but the process to define the explicative variables is unusual. The redesign through multi-mode approach represents a new frontier of the sampling practice and it has never been used in Istat, although is potentially feasible. Further analysis must be arranged on the real data of the THS, and the results will give better indications about the proposed solutions to the undercoverage problem.

Among future perspectives, we can mention that Istat at present, following also the suggestions highlighted by the research group, whose results are reported in this paper, decided to include the THS questionnaire as a module in another Istat household sampling survey, the Consumer Expenditure Survey. This survey is based on the municipal registers and a two stage sample design. In the experimental version this survey makes use of the CAPI technique for all the households in the sample. This choice agrees in some sense to second proposal. The aim of a new Istat research group is to study how to implement the new THS according to the renovated sampling design.

Finally, the evidences emerged from the research group have induced Istat to use the CATI-CAPI technique for the Adult Education Survey, and to redesign the Citizen Victimization Survey from CATI to CATI-CAPI data collection mode, according to the sampling strategy described in section 4.

References

Bethlehem J. (2009), “Can We Make Official Statistics with Self-Selection Web Surveys?” *Proceedings of Statistics Canada Symposium 2008*.

Blumberg S.J., Luke J.V., Cynamon, M.L., Frankel M.R. (2008), “Recent Trends in Household Telephone Coverage in the United States”, in Lepkowski J.M., Tucker C., Brick J.M., de Leeuw E.D., Japec L., Lavrakas P.J., Link, M.W., Sangster R.L. (eds) *Advances in Telephone Survey Methodology*, pp. 56-86. Wiley, New Jersey.

Brick J.M., Lepkowski J.M. (2008), “Multiple Mode and Frame Telephone Surveys”, in: Lepkowski Lepkowski J.M., Tucker C., Brick J.M., de Leeuw E.D., Japec L., Lavrakas P.J., Link, M.W., Sangster R.L. (eds) *Advances in Telephone Survey Methodology*, pp. 149-169. Wiley, New Jersey.

Callegaro M., Poggio T. (2004), “Where can I call you? The “Mobile (Phone) Revolution and its impact on survey research and coverage error: discussing the Italian case”, in: *Proceedings ISA RC33 Sixth International Conference on Logic and Methodology: "Recent Developments and Applications in Social Research Methodology"*, Amsterdam.

Cobben F., Bethlehem J. (2005), “Adjusting undercoverage and nonresponse bias in telephone surveys”, *Discussion paper* 05006. Statistics Netherlands Voorburg/Heerlen http://www.cbs.nl/NR/rdonlyres/7FD00F42-15A3-4151-9DAA-2D54566CF59A/0/200506x10pub.pdf.

Istat (2009a), “Documento di sintesi sullo stato delle indagini CATI presso le famiglie e sull'analisi delle criticità legate alla copertura delle indagini basate su liste di telefoni fissi”, Technical Report.

Istat (2009b), “Documento di sintesi sulla prospettiva di ampliamento delle indagini CATI sulle famiglie e sulle possibili soluzioni da adottare per risolvere i problemi di copertura e qualità delle indagini telefoniche”, Technical Report.

Istat (2009c), *La vita quotidiana nel 2008*, Collana Informazioni.

Istat (2009d), *I viaggi in Italia e all’estero, Anno 2006*, Collana Informazioni.

Kalton G., Kasprzyk D. (1986), “The Treatment of Missing Survey Data”, *Survey Methodology*, 12, 1-16.

Kish L. (1965), *Survey Sampling*, Wiley, New York.

Kuusela V., Callegaro M., Vehovar,V. (2008), “The Influence of Mobile Telephones on Telephone Surveys”, in: Lepkowski J.M., Tucker C., Brick J.M., de Leeuw E.D., Japec L., Lavrakas P.J., Link, M.W., Sangster R.L. (eds) *Advances in Telephone Survey Methodology*, pp. 87-112. Wiley, New Jersey.

Lee S., Valliant R. (2008), “Weighting Telephone Samples Using Propensity Scores”, in: Lepkowski J.M., Tucker C., Brick J.M., de Leeuw E.D., Japec L., Lavrakas P.J., Link, M.W., Sangster R.L. (eds) *Advances in Telephone Survey Methodology*, pp. 170-183. Wiley, New Jersey.

Lepkowski J.M., Tucker C., Brick J.M., de Leeuw E.D., Japec L., Lavrakas P.J., Link, M.W., Sangster R.L. (eds) (2008), *Advances in Telephone Survey Methodology*, Wiley, New Jersey.

Lessler J. T., Kalsbeek W. D. (1992), *Nonsampling Error in Surveys*. Wiley, New York.

Little R. J. A. Rubin D. B. (2007), *Statistical Analysis with Missing Data*. Wiley, New York.

Nathan G. (2001), “Telesurvey Methodologies for Household Surveys – A Review and Some ThougTHS for the Future”, *Survey Methodology*, 27, 7-31.

Rosenbaum P. R., Rubin D. B. (1983), “The Central Role of the Propensity Score in Observational Studies for Causal Effects”, *Biometrika*, 70, 41–55.

Särndal C.E., Lundström S. (2005), *Estimation in Surveys with Nonresponse*, Wiley, New York.