Proceedings of Q2008
European Conference on Quality in Official Statistics

Innovations on methods and survey process for the 2011 Italian population census

Giancarlo Carbonetti, Marco Fortini, Fabrizio Solari
Italian National Statistical Institute ISTAT, Rome, Italy

1. Introduction

Censuses are huge operations which need to be concentrated in a very short period of time This feature tends to be in contrast with new social trends where people are increasingly hard to be surveyed and is one of the main emerging problems in Censuses carried out by different Countries (Desplanques, 2000). In addition, National Statistical Institutes need both to reduce costs and to improve timeliness. For these reasons a great deal of new methods and techniques have been applied by many countries so to support conventional census or even in their place (United Nations, 2007).
The 2001 Italian census evaluation program showed satisfactory accuracy and adequate timeliness for data dissemination purposes. Instead, organizational difficulties met by the municipalities during the field work were highlighted, in connection with the great differences among the Italian municipalities both in terms of their populations and their work efficiency.
Taken also into account the time restrictions that the Eurostat regulation is introducing for the census dissemination results, the main goal of the next Italian census round is to simplify the municipalities activity by two major innovations:

- a smaller workload on the field so as to decrease the number of enumerators;
- a greater use of administrative data in supporting the data collection phase.

These key points will affect the planning of census methods and process workflow, with important effects on quality issue.
Section 2 presents an outline of next census round strategy and of the legislation and society factors behind these changes. Section 3 focuses on sampling strategy planned in order to collect information through short and long forms, giving some insight on experimentation carried out so to evaluate possible alternatives. In section 4 are described the two options being considered so as to correct census counts for undercoverage errors and, finally, last section is devoted to final remarks.

2. Operational, social and legislation factors influencing new census strategy

Italian Population Census is traditionally carried out by means of conventional data collection techniques. In addition, it is used information coming from municipal population registers after the census collection phase in order to prevent as much as possible census undercoverage errors.
Population registers are managed locally by each municipality since in Italy a unique centralised register is not yet available. Municipalities population registers enlist people and households dwelling in the municipalities and record every vital and

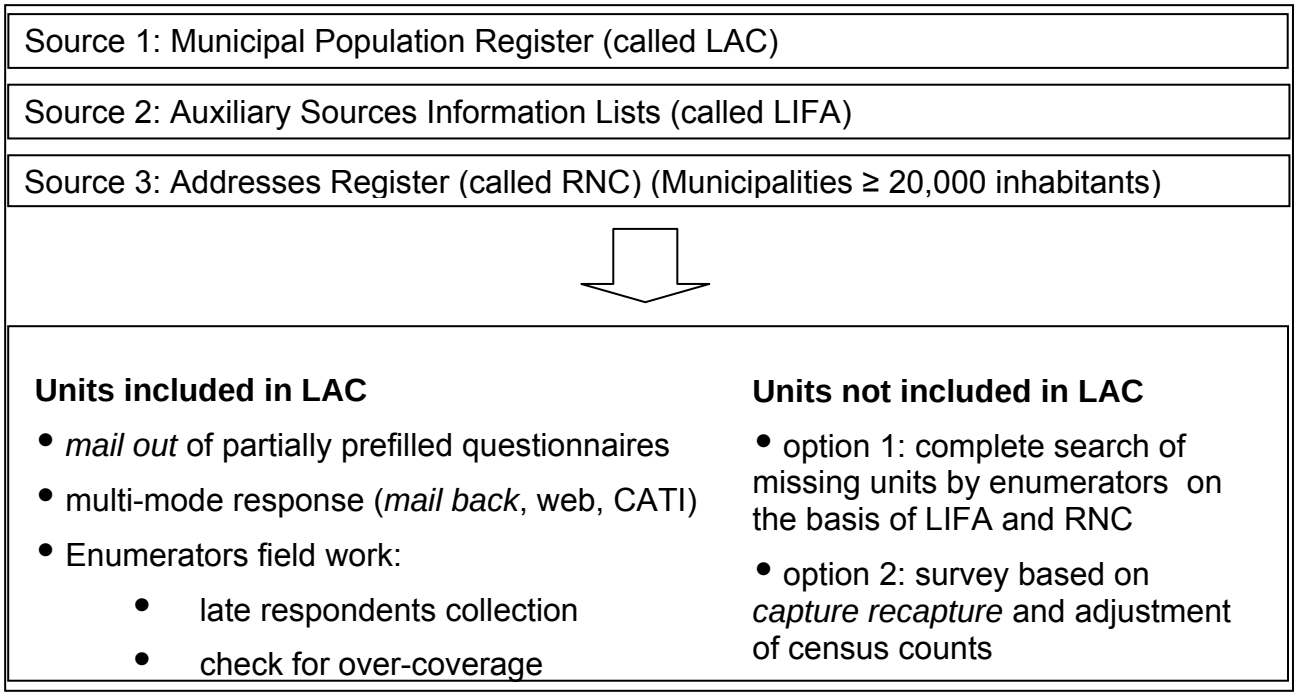
migration events either internal or from/to abroad. Moreover, population registers are integrated into the vital statistics system and help in updating legal population in intercensal years.
We can summarize the achievement of the past Italian Population Census in the following steps: first a complete search for people and households inside each enumeration area is accomplished; then a check of the population register is carried out for detecting eventually missed people and households and finally a follow up is made in order to check for their actual presence. Therefore, the relationship between census and registers contributes to enhance the quality of both of them.
As far as social changes occurred in Italy during the past years, the appropriateness of conventional methodology for taking the population and housing census has to be questioned. In fact, timeliness of data dissemination could be affected by changes in the population life-style, especially for people living in larger municipalities. Moreover, censuses regarding information already available from registers can affect the willingness to take part in the survey by respondents, who could not consider the quality issues regarding those sources. Finally, an increasing concern for confidentiality by respondents could also cause an increase of difficulties in enumeration tasks.
Another source of concerns comes from the analysis of the past population censuses which displayed some critical points to be considered by Istat in planning the next census strategy. First of all too many administrative offices (municipalities, provinces, chambers of commerce) have been involved in census operations resulting in inefficiency due to overlapping tasks. For instance, deliver and collection of questionnaires are only partially accomplished in time because of institutional inefficiencies (legislative delays, delays in providing funds) and administrative problems (delivering of questionnaires, problem with the collection system). In particular, the need of a massive number of enumerators (about 108,000 people) caused problems in personnel selection, training and turn-over.
A great effort in conducting the Italian census falls on municipalities and above all on large municipalities. Figure 1 displays the amount of Italian population in four main classes of municipalities depending on their population size.
It seems evident that a consistent part of the population dwells in large municipalities, which, therefore, are in charge of a consistent amount of census operations. Consequently, population size affects the process organisation and has a great impact on the accuracy and timeliness of the results.
In addition it is necessary to take into account recent change happened in European regulation. In fact next censuses must satisfy the *Recommendations for the 2010 Censuses of Population and Housing*, suggested by *United Nations Economic Commission for Europe* (UNECE, 2006) in cooperation with *Statistical Office of European Communities* (Eurostat) which compels the National Statistical Institutes to improve the timeliness of data dissemination. Therefore, all the census outputs must be supplied to Eurostat within 27 months from the end of the census year which for Italy means March 31th 2014. Then the organization and the methodological structure of the next census must guarantee an improvement in timeliness respect to the previous experiences.
In order to attain UNECE/Eurostat recommendations, it will be considered a modularity for the innovation scheme which depends on the municipalities population size. More precisely the municipalities will be divided into four categories defined in table 1 and for each of them a different census scheme will be followed.

Table 1: Distribution of the Italian population for type of municipalities

TYPE	Size (inhabitants)	Municipalities	% Population (01.01.07)
A	> 50,000 and chief towns	165	35.8
B	20,000 – 49,999	339	17.0
C1	5,000 – 19,999	1,859	29.6
C2	< 5,000	5,738	17.6

In figure 1 it is displayed the next census strategy which aims in achieving a better integration between census and registers.
To this aim, the first step consists in the preparation of lists to be used during the field activities and this task is planned to start more or less one year before census date. Then the Municipal Population Registers (called LACs) will be supported by auxiliary sources information lists (called LIFA) such as taxation files and by the list of street numbers obtained through a survey carried out before the census (called RNC and expected only for municipalities with a population size of at least 20,000 inhabitants)). In order to reduce the workload in municipalities, the questionnaires will be delivered by mail and the household are able to choose between different forms of returning: by mail, via web or delivering them directly to municipal collection centres. Furthermore, the municipalities will monitor the return of questionnaires and enumerators will be sent on the field only to check for the register overcoverage errors (i.e. actual absence of residents) and eventually collecting questionnaires from late respondents.

Figure 1: Census strategy for municipalities



The questionnaire will not be unique: part of the households will receive the complete questionnaire (long form), while the remaining will be mailed a smaller questionnaire (short form) containing the main structural information. The long forms will be delivered only to samples of household and will include all the socio-economic information we need to satisfy national and international demand.
Since population registers are affected also by undercoverage census strategy is planned to take into account for this type of error and a detailed description of this issue is given in section 4.

3. Use of sample of households for long form enumeration

In order to achieve high spontaneous response rates in census survey, length of questionnaires can be reduced as much as possible by introducing the so called short form. As a consequence, it is expected an increase of spontaneous return rates and a reduction of the average return time of questionnaires.
Since this approach would imply information loss, the selection of a sample of households to be provided with long form would be a solution so to preserve the richness of census information.
The sample strategy will then consist in simultaneous use of short and long forms, where short forms are sent to the most part of the households and contribute in reducing the enumerators workload, whereas long forms are submitted to a sample of households.
To this purpose, the overall set of census variables is divided in two subsets: the first one regarding demographic variables, the second set concerning educational level, occupational status and commuting. The short form will account for the first set of variables whereas the long form for the whole set.
In so doing, data related to the demographic variables will be collected on the whole population while the information related to the remaining variables will be surveyed on a sample of households.
Since sampling strategy is not effective for small population size, it is planned only for larger municipalities. An issue is related to the smallest municipality population for which the sampling strategy can be adopted. An option, which tries to maximise savings, considers sampling for municipalities with at least 5,000 inhabitants, whereas for municipalities smaller than 5,000 inhabitants, long forms will be submitted to the whole population.
In order to produce accurate estimates for small territorial references, new sub-municipal domains have been defined aggregating neighbouring and homogeneous Enumeration Areas. As a consequence, a second option would limit the use of sampling to municipalities of at least 20,000 inhabitants since sampling variances are expected to be too large in comparison with the requested precision of sub-municipal results.
In both cases sample strategy will be considered only in municipal urban areas while long forms will be delivered to all the households dwelling in rural areas.
The most appropriate sampling strategy has been chosen through a Monte Carlo simulation study on Italian 2001 Census data.
Simple Random Sampling of Households with sampling rates of 10%, 20% and 33% were tested on 40 municipalities of different size and properly chosen among the

Table 2: Distribution of municipalities by geographical area and demographic size.

Geographical areas	Classes of population size (a)			Total
	10,000-20,000	20,000-100,000	more than 100,000	
North	4	6	6	16
Center	2	3	3	8
South	4	6	6	16
Total	10	15	15	40

(a) It has been considered the legal (official) population date referred to the 2001 Census of Population.

three Italian macro regions. Simple random sampling was chosen because outperforming area sampling in terms of variance (see Carbonetti and Fortini, 2008). The distribution of municipalities according to their population size and geographical area was considered so to take into account the strong differences among the Italian municipalities (see Table 2).
About 90 cell counts resulting from census multiway tables were estimated through calibration methods. Cell counts were computed for each one of the 457 tested sub-municipal areas between 5,000 and 15,000 people. The properties of sampling estimates were assessed by means of CV statistic computed on 1,000 sampling replications for each tested sampling rate.
Table 3 describes the amount of units involved in the simulation. In particular, the study considered a little more than 10 percent of households and a little less than 10 percent of individuals in Italy.
Table 4 shows the main simulation results. It reports, for each class of cell counts and for each tested sampling ratio, the average and the maximum CV estimated over the sub-municipal areas. For instance, for sampling ratio of 10% the average CVs fall under 10% when counts go over 1,000; when sampling ratio increases up to 33% this threshold decreases to 250. Higher values of CV for small frequencies are expected. However, higher values of CV for small counts corresponds to smaller differences in absolute terms.

Table 3: Counts of areas, enumeration areas, households and individuals involved in the simulation

	Sampled Units	Universe	%
Areas	497	3,347(*)	14.85%
Enumeration areas	30.890	382.534	8.08%
Households	2.243.511	21.810.676	10.29%
Individuals	5.537.582	56.594.021	9.78%

(*) Estimated counts

Table 4: Distribution of average and maximum CV% for classes of cell counts for the three tested sampling ratios

Classes of cell count	sampling ratio = 10%		sampling ratio = 20%		sampling ratio = 33%	
	Average Cv%	Max CV%	Average CV%	Max CV%	Average CV%	Max CV%
<10	**143.3**	191.8	**101.4**	123.7	**66.5**	95.8
10 ├30	**75.9**	85.1	**48.4**	54.6	**33.8**	38.5
30 ├50	**51.8**	57.1	**31.8**	37.1	**23.4**	25.6
50 ├100	**38.6**	41.3	**22.3**	28.4	**17.4**	19.1
100 ├250	**25.4**	28.5	**15.7**	19.6	**11.4**	12.8
250 ├500	**16.1**	18.3	**10.4**	12.5	**7.5**	8.1
500 ├1,000	**11.8**	12.8	**7.5**	8.2	**5.3**	5.9
1,000 ├2,500	**7.5**	8.9	**4.7**	5.9	**3.3**	3.9
2,500 ├5,000	**4.9**	5.4	**3.0**	3.6	**2.0**	2.5
5,000 ├10,000	**3.2**	3.8	**2.0**	2.5	**1.3**	1.9

As expected, the most accurate estimates are obtained for largest sampling ratios. In general, the gain of efficiency measured as relative difference of CVs, is about 33%-38% when sampling ratio going from 10% to 20% and it is about 53%-58% in case sampling ratio increases from 10% to 33%.
Moreover, CVs decline for largest sub-municipal domains. After classifying the areas in three groups of population size (less than 10,000 or small; between 10,000 and 12,000 or medium; more than 12,000 or large), it has been observed a decrease of CVs of about 14%-20% from small to medium areas and of about 22%-28% from small to large areas. This result suggested planning areas of about 15,000 people as sub-municipal census domains.
Table 5 shows the percentage of cell counts estimates falling into a given interval of CV for each sampling ratio. For example, the percentage of cell counts showing a CV less than 10% increases from 29.80% to 65.42% when the sampling ratio increases from 10% to 33%. Conversely, incidence of cell counts that exhibit CV higher than 50% decreases from 15.97% to 5.05%.
However, many cell counts show CVs higher than 10% even when the largest sampling ratio is considered. This is not surprising since census multivariate tables tend to concentrate the most part of the character in relatively few cells. Large improvement in efficiency can be gained by marginalising tables or increasing the domain of interest.
It can be observed that estimates referred to municipal or larger domains obtain an extra improvement of accuracy as far as part of their territory is completely surveyed by means of long forms. For instance, rural areas within municipalities interested by sampling for estimation at municipal level or municipalities under 5,000 inhabitants for estimation at regional or greater level.

Table 5: Distribution of cell counts in classes of percentage CVs (only for areas larger than 12,000 inhabitants). Comparison for the three tested sampling ratios

Classes of CV%	Sampling ratio		
	10%	20%	33%
< 10%	29.80	51.07	65.42
10% - 50%	54.23	40.29	29.52
> 50%	15.97	8.64	5.05

Finally, in order to improve the accuracy of the estimates, small area estimation methods seem to be reasonable for the estimation of:
- smallest territorial levels in which the sample could be not representative;
- very small cell counts in larger domains.
Standard small area estimators is being analysed so to improve efficiency (Eurarea Consortium, 2004). Early results from ongoing experiments seems encouraging since, for percentages under 1% a reduction of CV between 40% and 80% is obtained. For percentages over 1%, it comes out that estimates are affected by bias. Further work will deal with accuracy improvement of estimation of percentages between 1% and 3%.

4. Undercounting evaluation and integration of missing units

In order to ensure the production of unbiased population counts, a census based on municipal population registers needs to take into account for both overcoverage and undercoverage of the administrative lists. Overcoverage can be investigated during survey field activities by excluding from the enumeration process those who are enlisted into a municipal register without actually dwelling in the municipal area.
In so far as the administrative source is affected by undercount, a *register supported* census has to be corrected also for those units actually residing but not enlisted into the population register for the same area. This result can be achieved either by means of a complete and labour intensive search on the field or through a capture recapture model based on a sample survey.
Early results of analyses on administrative data, regarding changes of residence of people moving either between other Italian municipalities or toward/from abroad after 2001 census, show that people enlisted in population registers because of census enumeration are more or less 235,000.
This value can be assumed as a rough measure of the effect of census on amendment for registers undercount even though it is probably only an underestimate of the whole undercount since not all the municipalities corrected their population registers after census results.
Since, at present, the debate about the best choice to make is still in progress, this section discusses both approaches in order to enlighten strengths and weaknesses.

4.1 Complete search of missing units

Here the main goal is to identify all those units that despite residing on the municipal territory are not enlisted into the population register. Indeed, Italian census regulation requires the registration of these units in the population register and their inclusion in the census counts.
Auxiliary lists are used in order to obtain signals from not registered people or households actually living on the territory.
A first list is represented by the households to be contacted because of their inclusion in administrative registers, such as central or local tax contributors, domestic utility users and benefits recipients, which give evidence of their existence on the municipal territory (*LIFA*[1]).
The second list is an inventory of street numbers, called *RNC*[2], and it is obtained six months before census through a survey based on previously available lists of postal

[1] Lista Integrativa Famiglie in Anagrafe
[2] Rilevazione Numeri Civici

addresses. Among data contained for each street number in RNC there are included both the whole number of housing units found at the street number and, coming from the LAC, the number of housing units that are dwelled by legally resident people, arranged for each street number as well. The RNC will help enumerators in pointing out those buildings which have to be surveyed in search for housing units that are eventually occupied by not legally resident people. In fact, for a given street number, search for people who are eligible for inclusion in the population register will be required only if the difference between the above cited pair of numbers in RNC is positive.
Another piece of information available with respect to each street number is the list of households that already returned their form. In order to attach this information to each housing units so to avoid the enumerators contacting those households that already answered, a real time process control system is needed. It is worthwhile underlining the importance of this kind of knowledge since, compelling the enumerators to contact twice the respondents, it could cause an increase in the cost of collection and could make more probable the inflating of census counts due to double enumerations.
Municipalities with population lower than 20,000 inhabitants will not carry out a RNC survey and, as a consequence, their enumerators are asked to go through their assigned enumeration areas equipped with the list of addresses dwelled by people who already sent back their questionnaires and check for any not registered dwellings. In fact it is considered that for small municipalities this procedure could adequately fit with the requested reliability even if less supported by previous information.

4.2 Complete search of missing units: critical issues

Field allocation of enumerators and forms has to be carefully considered since the expected number of households to be contacted does not depend only on the expected missing units on the field but also on the number of empty housing units and on the *mail out* response rate. This issue, if not adequately considered, could result either in process bottlenecks or in diseconomies.
A complete search can be also used to obtain the best possible result when the entire field search is considered unfeasible. In this case, however, other actions have to be planned in order to estimate the expected downward bias resulting from the reduced field work.
Currently, territorial distribution of buildings and housing units on 2001 census data is being analysed. For instance, in Florence 72% of buildings have all housing units occupied by resident households. This result seems to be hopeful, considering also that about 90% of the residual buildings enclose up to 3 vacant housing units.
Less appealing is the mean number of housing units within these buildings which is 8.5; such a value is in fact larger than 5.6, which is the average housing units for buildings in the whole Florence. This fact probably will make more difficult the search for not resident people inside each building.
If the enumerators are not provided with all the available information on respondents for each household they will have to contact every housing unit included in a given street number, entailing a larger probability to cause person and household duplications together with a greater fieldwork to be carried out.
The use of LIFA list can make easier the detection of not registered households, but on the other hand, could cause duplications of respondent units. This problem might be emphasized if the same household is included both in LAC and LIFA and is not

recognized as the same unit by the enumerator. Appropriate actions should be planned with the aim of facing this circumstance by means of both enumerators and back office operations.
Duplicates detection activities based on record linkage should be carried out on census respondents in order to recognise those units which could have been counted twice because living at addresses different than their legal residences. These cases, even if not large in number, are expected to be found within the same municipality or in two different municipalities. In the latter case they could be hard to be identified without asking the legal residence to those met on the field.

4.3 Adjustment of census counts by a capture recapture model

A thorough field search is not the only option for detecting missed people and households to correct census counts. As an alternative, the number of people living on the territory without being enlisted into the population register can be estimated through a sample survey in order to simplify field activities and keep under control the costs. A sampling supplementary survey based on capture recapture methods could in fact limit municipalities and territory involved into the missing units retrieval operations so employing fewer enumerators and a smaller amount of resources. In this way it would be possible to estimate the number of missed units at municipal and sub-municipal area levels.
The approach consists in the selection of a sample of enumeration areas (or other proper clusters of reference units) based on an area frame design, the enumeration of people and households laying in the selected territorial units and the subsequent matching to the records coming from the population registers.
Therefore, an estimate of the population dwelling in a given area could be obtained by the equation

$$\hat{P} = A\left(1 - t_{over}\right) / \left(1 - \hat{t}_{under}\right), \quad (1)$$

where:
$\hat{P}$ is an estimate of the true population count;
A is the number of people enlisted into the population register;
t_{over}, estimates the over-count affecting the population register given by the rate of not eligible people that are enlisted into the population register;
$\hat{t}_{under}$, is the rate of dwellers non enlisted into the population register over people enumerated during the supplementary survey
The equation (1) estimates the amount of the usually resident population according to the standard hypotheses stated for capture recapture models (Wolter, 1989). In this case, some valuable characteristics result from the use of population registers which will be available before the survey reference date. The diagram in figure 2, represents a municipality partitioned into areas and sketches the census in a generic municipality. White areas are those involved in mail out from population register and checked for overcoverage. The grey areas are instead selected for undercoverage estimation and inspected by a conventional census enumeration, carried out independently from the population registers.
Following this scheme, the main survey based on mail out and the supplementary survey could be carried out on different parts of the territory in order to keep distinct the two field survey processes.

In doing so, the estimation of $\hat{t}_{under}$ and t_{over} would be kept independent since the former relies on administrative data whereas the latter is carried out through field search.
Moreover, since both the surveys are run at the same time, the closed population hypothesis required for capture recapture estimators is fulfilled by definition. In addition this approach will save time because the demanding phase of matching records between the two sources will not have to wait for the ending of a second round of data collection. Finally, being the first list represented by the population register, households are surveyed only once either in the white or in the grey areas, so reducing the statistical burden in comparison with the traditional implementation of a capture-recapture experiment.
The sample design will rely on two stages represented respectively by municipalities and by either enumeration areas or street numbers, depending on the class of population size in which the sampled municipality is included. The second stage sampling units are not the same for all municipalities since lists of street numbers will be available only for municipalities with at least 20,000 inhabitants whereas in smaller municipalities the finest territorial partition will correspond to the enumeration areas. Of course, where a reliable list of street numbers is available, a sample based on these units should be preferred since the smaller housing units enclosed in each street number entails a less significant cluster effect than that implied by the use of enumeration areas.
The supplementary survey sampling design aims to give more accurate estimates for larger municipalities where higher level of under coverage errors are expected. Specific analyses based on the past Italian census and post enumeration survey data are underway in order to find the right balance between the sample size and its allocation given the expected efficiency of the estimates. Insofar, as the sample strategy prefers the estimates variance reduction rather than cost and operational savings, a greater number of first stage units should be sampled, each one including only a few second stage units. On the contrary the opposite choice should be taken only when savings are preferred to efficiency.
Sample size is not expected to exceed one million people as a whole. Municipalities over 50,000 inhabitants will be considered self-representative, whereas smaller municipalities will be sampled with probabilities proportional to their population size. From early computations it results that reliable direct estimation cannot be guaranteed until municipal and sub-municipal levels without a sample size of about ten million people, so vanishing the expected savings of human and asset resources.

Figure 2. A simplified scheme of the survey in a municipality

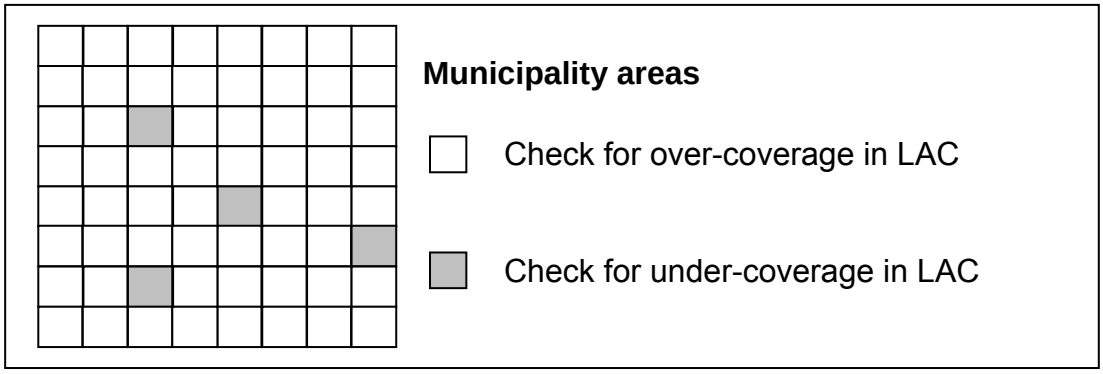
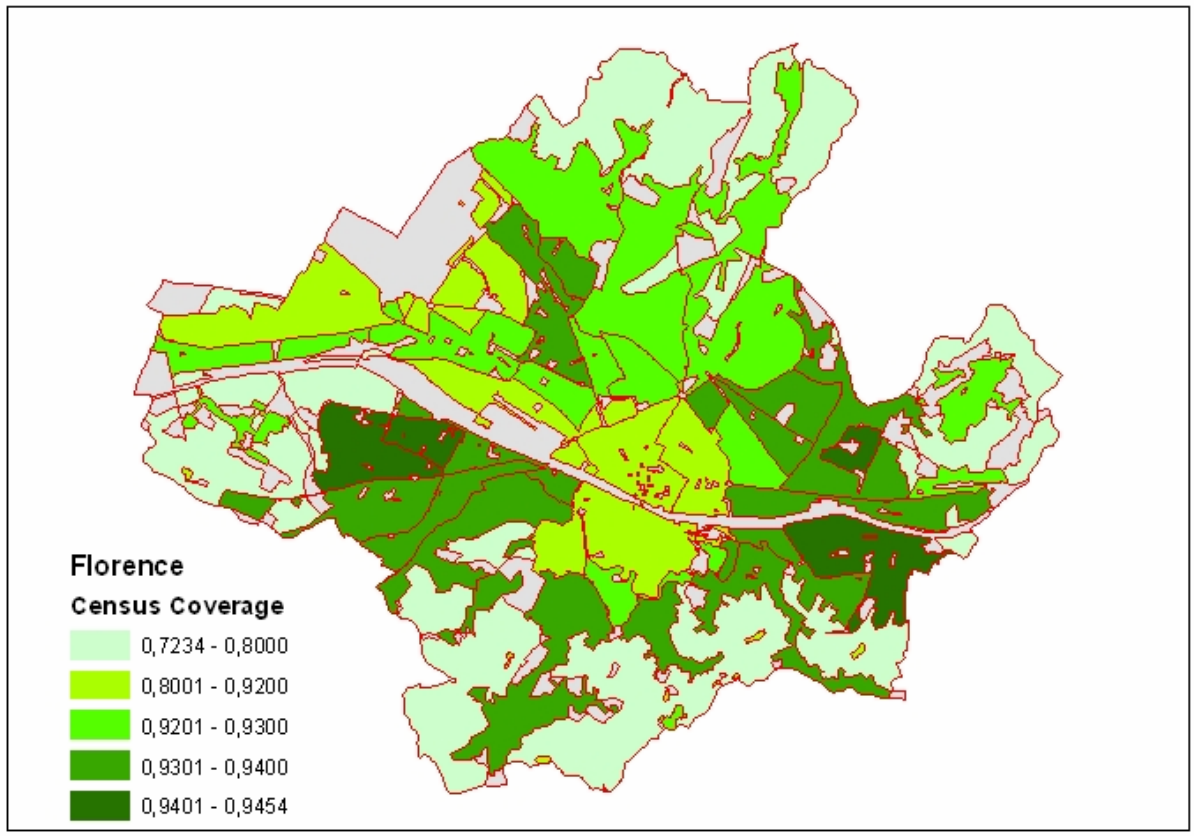


Figure 3. 2001 Census Undercoverage rate for 34 sub-municipal areas of Florence



For this reason direct estimates are planned until regional domains whereas model based estimators will be considered at municipal and sub-municipal levels.
Analysis of undercoverage small area estimators based on 2001 Census Post Enumeration Data is just started. At present we are testing the synthetic estimator proposed by Wolter (1986) on PES data by fitting a logistic model in order to estimate individual enumeration probability. A jackknife method will be used to evaluate sampling variance. In the next step a comparison with area level estimators such as the estimator proposed by Dick (1995) is scheduled.
Preliminary results are available for the 34 sub-municipal areas in the city of Florence, which have an average number of 10,000 inhabitants (See figure 3). The distribution of census coverage rate across these areas shows a certain degree of spatial correlation. Moreover, when summed up to municipal level, the estimate is perfectly coherent to the direct census coverage estimate of about 92%.

4.4 Capture recapture approach: critical issues

The most important theoretical requirement of the capture recapture model is the statistical independence between the captures. In this context it means that the probability to be enumerated by the supplementary survey for two people having the same characteristics has to be the same whether they are registered into the population register or not.
In case that independence assumption does not hold, the capture recapture estimates will result in either a positive or a negative bias. For instance, a higher tendency of people to avoid the supplementary survey, among those that are already included into the administrative file, causes an overestimate of the population counts. On the contrary, an underestimate is determined if not registered people tend to be missed by the supplementary survey more frequently than those who have their legal

residence in the municipality. For this reason the possible existence of subpopulations affecting the independence assumption has to be carefully considered.
Since survey operational features can be considered one of the main causes of dependence between the different captures, particular attention have to be paid in order to maintain as more separate as possible the two captures in terms of involved human resources and survey phases. Actually, the assumption is expected to hold in the case considered here, since the procedures of registration into the population registers are unrelated to those involved by the supplementary survey.
Even considered the difficulty to check for independence assumption, an opportunity would be given by the availability of a third list regarding at least some important subpopulations of interest. This list for the Italian case could be represented by the tax files.
A second requirement asks for the equivalence of each single unit in terms of its probability to be captured at a given occurrence. This condition do not hold for the whole population since of course people with different occupational status, age group or other characteristics will experience different probabilities to be enumerated. Many methods are available in literature to face the lack of this hypothesis, the most widely applied consisting in making estimation inside properly designed post-strata in which the capture probabilities can be considered as constant. For this reason it is essential that all the characteristics useful in defining the post-strata are collected during the different captures. Being the information stored into the population register given in advance, an accurate assessment of its capability in defining appropriate post-strata should be carefully evaluated.
The accuracy of estimates is also affected by the ability to identify without errors how many times each unit has been captured during the study. Since in our case this information is acquired by means of a record linkage between the population registers and the survey data, the occurrence of record linkage errors can cause either positive or negative bias in the census counts, depending on the prevalence of missed or false matches in the data. The availability of consistent record linkage procedures is directly related to the use of discriminating and accurate key variables, such as personal identification variables and dwelling addresses. At the same time, adequate methods and technical equipment have to be addressed so as to handle with large data sets. Closely related to these aspects, it is very important to consider how much treatment of individual data for record linkage purposes can affect privacy issues.
Timeliness is another issue to be considered when capture recapture approach is adopted in order to correct census counts for missed people and households. Indeed, the additional record linkage tasks required by the method, if not well calibrated, can affect survey timeliness for combining information coming from population registers together with those obtained by the supplementary survey. In order to meet timeliness, deployment of human, hardware and software resources should be carefully taken into account.
Applying small area estimation at municipal and sub-municipal levels implies indirect estimates taking strengths from information regarding other (similar) territories. This statement will concern all the municipalities. In fact, the most part of the smaller ones will not be included into the sample while the largest ones (self-representative) could not have enough sample data to assure accurate direct estimates. Apart from statistical aspects, this issue has to be considered with respect to its legal

consequences given that state financial resources are assigned to municipalities on the basis of the population determined by the census (legal population).
The capture recapture method is not free from the risk of duplications for the enlisted people into the municipal population register at a given address but actually living elsewhere either in the same municipality or even in a different one. Indeed, these people could answer to the mailed questionnaires after having received them at their registered residence. At the same time they could also be detected and counted on the field during the supplementary survey. Categories of people at risk are those who live in more than one dwelling as those students and workers legally residing out of town. People with a fictitious legal residence for administrative or taxation reasons risk the duplication of their census record for the same reasons. Even if it can be considered as negligible, past censuses were also affected by this difficulties with the additional drawback of not being explicitly tackled because Istat was not allowed to collect and process personal identification variables. Differently, in the next census round, duplicates will be recognised through the record linkage of personal identification variables among records if additional information about legal residence will be collected for people enumerated by supplementary survey.
Finally, since the capture recapture approach can amend population counts for missed units without giving a complete micro-level information, a full correction of population registers at the single record detail is precluded as a result of this procedure. In this respect, such a strategy can be certainly considered reductive in comparison with the past censuses whereof, at least in principle, a whole amendment of population registers was achieved. On the contrary, sample strategy could supply the municipalities with target population counts concerning the main categories for which population registers are in need of updating, thus leaving municipalities the duty of searching and adding-on missing people and households by their own proper sources of information. It has to be noted that whereas in past censuses only a minor part of Italian municipalities completed on time the amendment of their populations registers, the strategy planned for next census round aims to achieve a timely, effective and extensive correction of the population registers as far as over-coverage is concerned. On the contrary, only an advice about the improvement needed to correct undercoverage will be given, since municipalities are expected to be willing to spend their own resources for improving their population registers through a capillary field search and consequently better supporting their requests for funds.

5. Final remarks

This paper outline the main innovations on statistical methods implied by the increasing use of administrative data, i.e. municipal population registers, so to support census operation planned for the next Italian census round. These innovations regard:

- the use of a *short/long form* sampling strategy in order of both reducing response burden and simplifying municipality organisation;
- the need of taking under control population registers for undercoverage so to achieve the required accuracy of census counts.

Tests and experiments carried out on previous census so to evaluate *short/long form* sampling strategy showed that satisfactory estimates can be obtained with a simple

random sampling of households from population registers and a sampling ratio of 33%. Rural areas and smallest municipalities will be instead exhaustively surveyed. Since more accurate estimates are observed for census areas larger than 12,000 inhabitants, census areas of about 15,000 people has been planned for next census round.
High sampling ratios imply high accuracy of estimates, but entailing a great economical and organizational effort. For this reason, the final choice should take into account the trade-off between the needed financial savings and the accuracy required at the various territorial domains.
Ongoing studies are evaluating the application of small area estimation in order to improve the efficiency in small domains (e.g. sub-municipal areas) and tables with sparse data.
Concerning undercoverage of population registers, two different approaches have been considered. A first solution consists in searching on territory, by enumerators, for people eligible for inclusion in registers through the use of auxiliary lists of people and of addresses. A different approach considers a capture recapture study in order to correct census counts. As discussed in the paper, the former approach implies, at least theoretically, a complete recognition of people to be added in registers at a price of higher expected cost. Conversely, the latter method allows for a less expensive and intrusive sampling strategy, but it is more dependent by statistical hypotheses, it is affected by sampling variance and it does not allow for a micro-data correction of population registers.
As a final point, 2011 Italian census can be viewed as a bridge from a census carried out by conventional methodologies to a census based on centralised population registers. This represents a complicate road for statistical, technological and political reasons. Therefore, a great deal of methodological work is required in areas such as sampling, modelling, imputation, record linkage, etc. The right balance between innovations, comparability of results and stakeholder needs has to be found in order to meet the required quality of census results.
In order to acquire an insight into the best possible census planning a pilot survey is scheduled on April 2009 with the purpose of evaluating the whole census strategy and testing some technical alternatives. Moreover, the census pilot survey is expected to be an important trial for work organization between Istat, Municipal Statistical Offices and the other local authorities involved in the census process.

References

Desplanques G. (2000). “Le recensement rénové de la population en France”. Colloque Insee-Eurostat sur les recensements après 2001. Paris, novembre 2000.

Carbonetti, G., Fortini, M. (2008), “Sample results expected accuracy in the Italian Population and Housing Census”, *Proceedings of the Joint UNECE/Eurostat Meeting on Population and Housing Censuses*, Geneva, downloadable from http://www.unece.org/stats/documents/ece/ces/ge.41/2008/4.e.pdf

Dick, P. (1995), “Modelling Net Undercoverage in the 1991 Canadian Census”, *Survey Methodology*, Vol. 21, No. 1, pp. 45-54

Deville, J.C. and Särndal, C.E. (1992) “Calibration Estimators in Survey Sampling”, *Journal of the American Statistical Association,* vol. 87, pp. 367-382.

Eurarea Consortium, (2004), “Project Reference Volume”, downloadable from http://www.statistics.gov.uk/eurarea

UNECE - United Nations Economic Commission for Europe and Statistical Office of the European Communities (2006), “Recommendations for the 2010 Censuses of Population and Housing”, *Proceedings of the Conference of European Statisticians*, ECE/CES/STAT/NONE/2006/4

Wolter, K. M. (1986), Some Coverage Error Models for Census Data. *Journal of American Statistical Association*, Vol. 81, No. 394, pp. 338-346

Rao, J.N.K. (2005), *Small Area Estimation,* Wiley, New York

United Nations (2007). “Principles and Recommendations for Population and Housing Censuses - Revision 2”, *Expert Group Meeting on the 2010 World Programme on Population and Housing Censuses*: