The New Population Census Strategy: from Tradition to Innovation

Stefano Falorsi

Head of Division on
Population Census and Integration of Social Surveys
Goals

Focus on relationships among census surveys and census production process

The new census is based on complex infrastructure composed by Register System (SIR) and census surveys: Master Sample - Areal component (A) and Lista component (L)

The role of census surveys is a part of general strategy of continuous census
Why is the Census important and useful?

- Provide a statistical information framework on the characteristics of the population and housing for the whole country and at different territorial levels, up to the "local" level
- Establish the legal population of each municipality
- Provide useful data for the review of the civil status register
Continuous Census Strategy

Traditional Strategy:
Census = Unique big survey

Direct, total and simultaneous survey

Continuous census strategy:
Census = Process that combine administrative and survey data

Continuous flow of administrative data → Istat Registers system (SIR) → Sample survey
The registers utilized for the new census

- Population Register: Sex, Age, Civil Status, citizenship, Municipality of usual residence, place of birth, education level
- Labour Register: Employed classified in economic activities (NACE)
- Places Register: Territory (Enumeration areas, addresses, dwellings)
The census surveys: goals

- More information
- Measurement of registers’ quality

Two different annual surveys with different purposes
The Publication of the results

- National publication with annual frequency
- Complete national publication referred to 2021
- Publication at European level respecting Eurostat rules
The transition to the new census strategy means that much information is taken directly from the administrative data integrated into the SIR. Not all detection units must be subject to an annual survey: cohabitations will be recorded in 2021.
Considering the information available thanks to the Registers and the annual survey that involves about one million and 400 thousand families, the production of a minimum set of tables at municipal level (about 30, remodeled over time) has been suggested to be disseminated annually.

The new annual dissemination is flanked by the traditional Italian ten-year spread, which currently includes around 260 crossings articulated in various levels of classification and territorial detail.

• The possibility of disseminating crossings at sub-municipal level (census areas, micro-zones, ...) starting from the integration between the SIR and the survey data is under study.

As already happens in other European countries (for example in Spain and in France) the hypothesis of a uniform “minimal” diffusion for all the municipalities could be evaluated, regardless of the population size, and of a more extended diffusion for the exclusive use of common areas above a certain population threshold.
How does the ten-year national publication change?

- eliminate tables dedicated to obsolete phenomena;
- eliminate crossings at regional level preferring tables to provincial or municipal details;
- simplify particularly complex intersections in terms of the number of variables and / or modalities associated with the classifications.

In addition to the results of the survey of information content, tools and dissemination strategies adopted in some foreign countries, the results of a study on the use of data from the 2011 Italian Census were evaluated (consultations / downloading of / from dedicated sites) I.Stat, LOD, Territorial bases - requests for personalized elaborations) with the aim of proposing an exhaustive information offer, not redundant and useful to stakeholders for programming and research purposes.

The new 10-year publication plan

More territory, less complex tables, no junctions with non-core variables not used in the inter-period period, focus on evolving phenomena (foreign, new types of family, the elderly, the NEET ...
The design of the first phase will make it possible to obtain direct estimates on the census variables for all the municipalities, strengthening the reliability of the results, and at the same time allowing the setting up and consolidation of the statistical registers of SIR.

Every year about 3 thousand municipalities will be investigated: a part of which (1,100) always involved in the four-year period (so-called AR) and a remaining part of municipalities that will instead be rotated (NAR).

The two sample surveys are related to a total of about 1.5 million households (950 thousand for the survey L and 450 thousand families expected for the survey A).

When fully operational, after 2021, over a ten-year period, a design based on half-yearly surveys will be taken: respectively to 500 thousand families for the survey L and 200 thousand families for the A, with an annual commitment of significantly lower resources.

This lower use of direct investigations will be made possible by an even more massive use of administrative sources, obtained on the one hand by intensifying the use of those already present within the information systems of the Institute, on the other by the integration of new acquisitions to be valued adequately for statistical purposes.
Main goal: to gather information useful for calculating the population of each Italian municipality (estimate of the quality of the Basic Register of Individuals).

- In order to estimate the RBI quality it is necessary to have a measurement of the usual resident population that is independent of the existing lists / registers: therefore a measurement obtained by counting people directly on the territory, "blind" (= without lists).
- This operation on the territory is carried out on specific survey areas, sampled by the Register of Places: in some cases they correspond to Census enumeration areas (portions of municipal territory), in other cases to specific addresses. A specific preliminary phase of verification of the sampled areas is foreseen: at the end of this phase, a validated list of addresses is available in which to survey households and houses.
- Collected data will be used to calculate RBI quality measures, referring to the different profiles of individuals (sub-populations). These measures will be used in the estimate, made from RBI, of the population of each municipality: even those not sampled.

Other goals: detection of unoccupied dwellings; thematic integration of the SIR; acquisition of information on places / territory; obtaining contact information from families for subsequent social investigations.
Main goal: information enrichment of the SIR in relation to socio-economic variables on families, individuals, occupied dwellings and buildings.

- To realize the thematic integration of the SIR by means of surveys it is necessary that the sampled families fill out the questionnaire; the investigative objective therefore translates into obtaining the questionnaire filled in by (almost) all the sampled families, within the timeframe set by the survey.
- This objective is pursued by facilitating the task of respondents through effective communication and the provision of various channels for compilation; in addition, by promoting self-completion.
- The experiments carried out have shown that focusing only on spontaneous restitution is not enough: for this reason, from a certain date, the investigators are sent to obtain a questionnaire filled out even by those who have not filled in autonomously.

Further goals: estimation of the quality of the Basic Register of Individuals (RBI); acquisition of information on places / territory; obtaining contact information from families for subsequent social investigations.
Sample designs to compare...

Sample size of L and A for the period 2018-2021

<table>
<thead>
<tr>
<th>AR/NAR</th>
<th>L</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Municipalities</td>
<td>Households</td>
</tr>
<tr>
<td>NAR</td>
<td>2.274</td>
<td>438.291</td>
</tr>
<tr>
<td>AR</td>
<td>672</td>
<td>561.553</td>
</tr>
<tr>
<td>TOTAL</td>
<td>2.946</td>
<td>999.845</td>
</tr>
</tbody>
</table>

Number of municipalities selected (2018-2021)

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sample size of L and A from 2022 onwards

<table>
<thead>
<tr>
<th>AR/NAR</th>
<th>L</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Municipalities</td>
<td>Households</td>
</tr>
<tr>
<td>NAR</td>
<td>1.970</td>
<td>246.929</td>
</tr>
<tr>
<td>AR</td>
<td>457</td>
<td>247.497</td>
</tr>
<tr>
<td>TOTAL</td>
<td>2.427</td>
<td>494.426</td>
</tr>
</tbody>
</table>

Number of municipalities selected

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DIFFERENCES BETWEEN THE TWO SURVEYS

In most of the sample municipalities, both surveys are expected to be performed: it is important that the detection actors manage and adequately communicate the differences between the two surveys in order to preserve the integrity of the design of each survey.

AREAL SURVEY

- There is no sampling list of responding families and should not be constructed prior to the survey
- The responding units are determined by the detection units found in the sampled areas; it is possible to obtain ex-ante from the registers a “number of expected families”, but the number of those detected can be both greater and less
- It is always the detector who completes the questionnaire (the family, although informed, must “wait” the detector).
- The questionnaires are never pre-filled

FRAME SURVEY

- There is a sample list of responding families
- The responding units are directly contained in the sample list; if the number of families sampled is ex-ante, those actually detected will be at most equal to that number
- It is desirable that the family complete the questionnaire (also by telephone call). The detector intervenes only starting from a certain date (the family does not have to wait for the detector)
- The questionnaires are pre-filled
QUESTIONNAIRE

Single questionnaire for the two surveys: in this way we find all the socio-economic information on a larger sample and not only on the sample L

The questionnaire is structured in a similar way to that of the latest experimental surveys and the long form of the 2011 Census Family Sheet. The contents are relative to:

- typology of the families, characteristics of the registry, of marital status, socio-economic and of territorial mobility of the people in habitual residence;
- type of accommodation, characteristics of occupied dwellings and related buildings.

- Blocking questions are expected (mandatory response)
- In the case of the L survey, some fields of the questionnaire are pre-filled on the basis of the information contained in the SIR for the families sampled
What news of content compared to the past? Some examples...

Some questions were eliminated on the basis of the 2011 results (for some variables a coverage of around 100% was recorded), the analyzes carried out on the use of the data released during the last comprehensive survey and in relation to the growing availability of administrative source data.

- Availability of water and toilets inside the house
- Rooms for professional use
- Mobile phones
- Presence at the date of the census
- Specific qualification
- Number of years required to achieve the qualification

New variables have been added following requests from stakeholders of the census data.

- Type of dwelling
- Specification province of birth of parents
- Type of studied course
- Car garage for rent / property

Other questions have been modified to ensure a higher quality of the information found.

- Family relationship
- Questions on commuting
Preliminary Experimental: Results on the Italian Population and Housing Census Estimation Methods

• The goal of the simulation based on a Monte Carlo experiment is to compare the empirical properties of the estimates in terms of bias and mean squared error (MSE);

• Linear and not linear models for the projection estimator have been fitted;

• Different set of auxiliary variables have been used in the simulation study;

• Furthermore, a composite version of the projection estimator has been considered;

• The indicators have been computed also for the in-sample domains and for the out-of-sample counterpart;

• All the estimators are compared by means of the standard indicators of accuracy of prediction: the Average Absolute Relative Bias (AARB) and Average Relative Root Mean Squared Error (ARRMSE).
Preliminary Experimental: Results on the Italian Population and Housing Census Estimation Methods

<table>
<thead>
<tr>
<th>Target Variable</th>
<th>Synthetic Projection</th>
<th>Composite Projection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ARR</td>
<td>ARR IN</td>
</tr>
<tr>
<td>Employed</td>
<td>5.4</td>
<td>5.2</td>
</tr>
<tr>
<td>Unemployed</td>
<td>31.1</td>
<td>28.5</td>
</tr>
<tr>
<td>Retired</td>
<td>6.9</td>
<td>6.5</td>
</tr>
<tr>
<td>Student</td>
<td>16.3</td>
<td>14.3</td>
</tr>
<tr>
<td>Other</td>
<td>10.1</td>
<td>9.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Target Variable</th>
<th>Synthetic Projection</th>
<th>Composite Projection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ARRMSE</td>
<td>ARRMSE IN</td>
</tr>
<tr>
<td>Employed</td>
<td>5.5</td>
<td>5.2</td>
</tr>
<tr>
<td>Unemployed</td>
<td>31.6</td>
<td>29.0</td>
</tr>
<tr>
<td>Retired</td>
<td>7.0</td>
<td>6.5</td>
</tr>
<tr>
<td>Student</td>
<td>16.5</td>
<td>14.4</td>
</tr>
<tr>
<td>Other</td>
<td>10.2</td>
<td>9.6</td>
</tr>
</tbody>
</table>
OUTLINE OF RESULTS

- The best model is specified by the variables: cross-classification of gender and age (28), marital status(6), educational level (12), citizenship (2), latitude, longitude, census count 2001

- The projection estimator is a very promising approach both in the synthetic and in the composite formulation, even if better results are obtained by the composite estimator. GLM outperforms LM results in terms of AARB and ARRMSE.

- Most of the results seem to be good enough to be published, since the MSE values reported in the previous tables are below the threshold of 33% commonly used to decide if disseminate or not estimates of labour market variables.
CONCLUSIONS

- The census operations are due by law and are essential to provide information on the population and housing for the whole country, up to a high territorial detail.

- The new census strategy puts the statistical registers of the SIR at the center of the production of the results: the surveys become one of the different sources of data acquisition and are samples.

- In the four years of the 'start-up' of the new strategy, the census surveys will be conducted each year on about 3,000 municipalities and will cover about 1.5 million families. This sample will be halved in the period following the 2021 survey.

- The permanent update of the SIR is an information asset that allows the dissemination of part of the census results, at the municipal level, on an annual basis. In particular, the data relating to the population will be produced annually from RBI on all Italian municipalities, based on an estimate that takes into account the quality measures proven by the investigations.
Thanks for your attention!