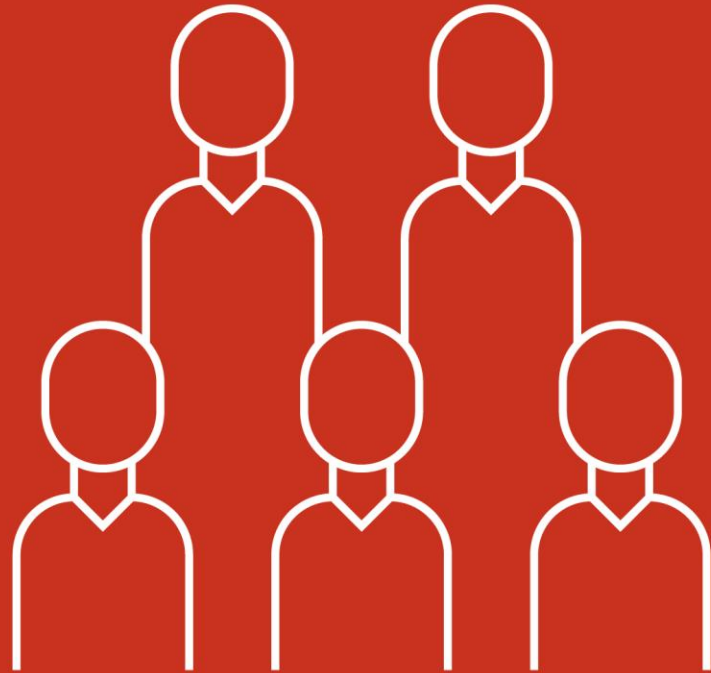CENSIMENTI PERMANENTI
POPOLAZIONE
E ABITAZIONI

Evolution of the Italian Permanent Population Census. Lessons learnt from the first cycle and the design of the Permanent Census beyond 2021

A. Bernardini, A. Chieppa, N. Cibella,
G. Gallo, F. Solari, D. Zindato
Istat, Population Census Division

POPOLAZIONE
E ABITAZIONI

CENSIMENTI PERMANENTI
L'ITALIA, GIORNO DOPO GIORNO.

UNECE Group of Experts on Population and Housing Censuses
Geneva, 21-23 September 2022

Istat | Istituto Nazionale di Statistica

❖The first cycle of the Permanent Population and Housing Census (PPHC): weaknesses and lessons learnt from 2018-19

❖2020 & 2021 fully register-based estimation

❖Second cycle of the PPHC: rethinking of the estimation process and of the quality assessment framework

First cycle of the Permanent Population and Housing Census

To replace the decennial census, in 2018 Istat launched the Permanent Population and Housing Census (PPHC):

- Combination of:

  ❖ registers: Population Register (PR) - Italian acronym RBI, Address Register (AR) - Italian acronym RSBL, thematic registers (educational level and employment status)

  ❖ sample Surveys: Areal Survey and List Survey

    ✓ to collect data for variables not covered by registers

    ✓ for the estimation of coverage errors of the PR

- It allows yearly availability of detailed census statistics (thematic census tables) and replaces intercensal population counts based on administrative data

First cycle of the Permanent Population and Housing Census

Population count: estimation of coverage errors of the PR - Dual System Estimation model

- PR (first capture) + survey data (second capture), differing from a typical PES, the second capture aims at measuring and correcting for both under- and over-coverage

- Areal Survey to estimate PR under-coverage, List Survey to estimate PR over-coverage

- In order to correct for the under-coverage of the List Survey, administrative 'signs of life' (SoL) derived from the Integrated Data Base of Usual Residents (AIDA) were integrated in the estimation process, i.e. non respondents with strong SOL were considered as enumerated

- Small area estimation models were used to reduce direct estimates variability for in-sample municipalities and to produce estimates for out-of-sample municipalities

- Population counts obtained applying correction coefficients to each individual in the PR according to his/her profile (for details see Gallo, G. & Zindato, D. 2021)

Weaknesses and Constraints

o <u>Sampling issues</u>:
- ✓ sampling size not allowing to estimate PR coverage errors by municipality, sex, age class and citizenship

o <u>Fieldwork issues</u>:
- ✓ strong heterogeneity of the fieldwork actors
- ✓ under-coverage issues for the Areal Survey

o <u>Users expectations issues</u>:
- ✓ need for a shared, accountable estimation process / need to produce population data usable by local authorities (register with weights i.e. not a register of "heads")

▪ Due to the pandemic, the 2020 surveys were canceled: in order to produce the 2020 population count, Istat opted for the use of administrative 'signs of life' (SoL) derived from AIDA (SoL Archive) to estimate the coverage errors of the population register

2020 Population Count Based on SoL

Administrative signs of life (SoL) refer to activities performed by individuals from which a durable period of time (e.g., a year) and a place (a municipality) can be clearly identified. Being self-employed or working for a business, being a public employee, having a regular rental contract on a dwelling, attending a school or university are examples of direct signs of life. On the other hand, indirect signs of life are defined as those situations identifying a status or a non-professional condition (i.e., people with universal income, pensions, or children or other relatives as dependents in tax returns).

- Linkage of the PR and the SoL archive (AIDA) results in an **'Extended Population Register' (EPR)** supposed to include the whole target population, i.e. the linked register is supposed to be affected by over-coverage but not by under-coverage

- Admin data have been used to calculate **over-coverage for the linked PR-SoL archive register (EPR)**. In details, deterministic criteria have been defined for determining over-coverage using SoL pattern at individual and/or household level

o Significant innovation ensuring the correspondence between the census count and the individual records (no more weights) but

  ✓ Istat is working to improve the use of SoL in the new cycle (post-2021) of the PPHC. A turning point will be represented by the acquisition of new sources (e.g. archives such as energy consumption - smart meters data - providing objective assessment elements with regard to the actual place of usual residence - Albert, A. & Rajagopal, R., 2013)

  ✓ the misplacement error (i.e. individuals assigned to a municipality by the PR or the SoL but usually resident in a different one) has not been evaluated so far.

Population Count Estimation in 2021

- **In 2021 Istat was able to carry out the census surveys: need to integrate register, admin and survey data**

- The choice of the SoL profiles and the corresponding decision to set each profile as usually resident or not can be:

  - ✓ the result of information provided by experts

  - ✓ suggested by the application of statistical models

  - ✓ a combination between them

- In 2021 the last option was adopted (combination of experts' information and statistical models). The statistical model used to integrate the decisions made by experts was a latent class model.

- **More intensive use of administrative sources** for the post 2021 census

- Crucial role of the **surveys** for the second round of the PPHC starting in 2022

  - ✓ to collect data for not replaceable (or only partially replaceable) variables (List Survey). Furthermore, new questions could be introduced in accordance with the new EU regulation to be approved

  - ✓ to improve expert/model based criteria applied to SoL

  - ✓ to provide quality measures of the fully register based population size estimation

- Need for a rethinking of the statistical framework for the overall estimation processes of the PPHC design

❖ For the second round of the PPHC, the proposed architecture is based on the definition of an **Extended Population Register (EPR)** resulting from an integration process involving the Population Register and administrative archives containing SoL (AIDA).

❖ **It is assumed that the EPR can only be affected by over-coverage** (i.e. it is not affected by under-coverage).

❖ **SoL profiles** are defined in order to identify subpopulations whose individuals are supposed to have a similar over-coverage behavior.

❖ A SoL profile based indicator function is defined (Bernardini, Cibella and Solari, 2022), according to which **all the individuals in each profile are classified to be either included or excluded from the population count**.

❖ This choice can be seen as a dichotomisation approach of fractional counting proposed by Zhang (2019) where instead over-coverage profile probabilities in [0,1] are predicted.

❖ In order to allow (model) unbiased register based estimates, instead of setting the SoL based indicator function in each SoL profile to be a constant value equal to 0 or 1, random values 0 or 1 can be generated starting from some specified predicted probabilities θ_h . Furthermore, in order to have a stable list of over-covered individuals over time, instead of generating independent random numbers every year, permanent random numbers can be assigned to each individual.

❖ Quality measures related to the population size estimation can be assessed by comparing register based estimates and survey estimates (for details see Zhang, 2022b). **This implies using surveys to assess quality measures instead of producing estimates and it is known as Audit Survey approach**.

❖ Following the **Audit Survey approach**, survey inefficiencies, as those encountered in 2018 and 2019, would affect only the estimation of quality measures and not the estimation of the population size.

❖ **The Audit Surveys approach allows a cost reduction compared to standard area surveys,** i.e. the Audit Survey sample size can be smaller than that of a based survey aiming at estimating the population size.

Quality measures and Audit survey approach (2)

❖In addition to the over-coverage error, an EPR can be affected by misplacement, i.e. individuals can be correctly included in the EPR but assigned to a wrong location.

❖An enhanced version of the EPR should be used, **in which records refer to pairs individual-address instead of only individuals** (see Zhang, 2021 and Bernardini, Cibella and Solari, 2022).

❖This would prevent adding bias selection components in survey estimates and, therefore, in the quality measures associated to register based population size estimates (see Bernardini, Cibella and Solari, 2022).

❖ The most appropriate sampling design for the Audit Survey consists in sampling individuals from the EPR or pairs of individuals and addresses from the enhanced version of the EPR.

❖ For some sets of individuals, there might be incomplete or unreliable information about their potential usual residence address: Possible solutions:

- area sampling (if they are supposed to be concentrated in specific portions of territory)

- indirect sampling (e.g., by selecting their work or study SoL addresses in order to try to collect useful information on their true address).
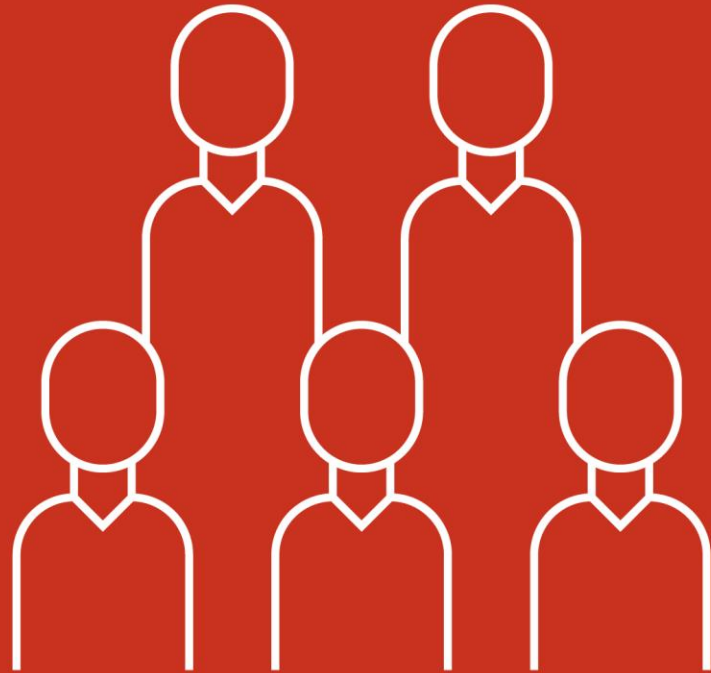
❖ Finally, in order to evaluate a possible under-coverage of the EPR or the EEPR:

- ✓ a small-scale Area Survey

- ✓ introducing into the Audit Survey a reverse record check component (i.e. cross-check for records related to non-respondents).

❖ Though putting together the different surveys could pose organizational issues and fieldwork problems, from a bare statistical point of view they can be linked by using graph sampling (see Zhang, 2022a)

❖ Pushing towards a larger use of administrative data:

  ✓ rethinking of overall estimation processes of the PPHC design (survey data used for the quality measurement of a fully register-based population count estimation)

  ✓ need to improve the use of SoL (improve classification criteria, get new sources)

  ✓ need for a statistical framework for the quality assessment

❖ The processing of 2021 Census data, currently ongoing, will be of great importance, given the availability of both survey data and administrative ones. Comparisons among different estimation models, the integration of administrative and survey data, the evaluation of fieldwork quality are all important areas of investigation to improve the design of the future PPHC cycles.

References

Albert, A. & Rajagopal, R. (2013), Smart Meter Drive Segmentation: What Your Consumption Says About You. IEEE Transactions on power systems: 4019-4030.

Bernardini A., Cibella N. & Solari F. (2022). A Statistical Framework for Register Based Population Size Estimation, Technical Report, Istat Advisory Committee on Statistical Methods 2022 Springtime Meeting, Roma, 19-20 May 2022.

Falorsi, S. (2017), The Italian experience on the Population and Housing Census: the Master Sample, UNECE Meeting, October 4-6 (2017), https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.41/2017/Meeting-Geneva-Oct/Day2_1130_Italy_falorsi_presentation.ppt__1_.pdf

Gallo, G. & Zindato, D. (2018). Annex H. Italy case study, in UNECE, Guidelines on the Use of Registers and Administrative Data for Population and Housing Censuses, Geneva, https://unece.org/guidelines-use-registers-and-administrative-data-population-and-housing-censuses-0.

Gallo, G. & Zindato, D. (2021). Italy: The combined use of survey and register data for the Italian Permanent Population Census count in UNECE, Guidelines for Assessing the Quality of Administrative Sources for Use in Censuses (endorsed by the 69th plenary session of the Conference of European Statisticians), https://unece.org/statistics/publications/CensusAdminQuality.

Istat (2020). Nota tecnica sulla produzione dei dati del Censimento Permanente: la stima della popolazione residente per sesso, età cittadinanza, grado di istruzione e condizione professionale per gli anni 2018 e 2019, https://www.istat.it/it/files/2020/12/NOTA-TECNICA-CENSIPOP.pdf

Istat (2020). Nota tecnica sulla produzione dei dati del Censimento Permanente: la popolazione residente per genere, età, cittadinanza e grado di istruzione al 31.12.2020, https://www.istat.it/it/files//2021/12/NOTA-TECNICA-CENSIMENTO-POPOLAZIONE_2020.pdf

Zhang, L.-C. (2019). On provision of UK neighbourhood population statistics beyond 2021, Report for ONS, https://arxiv.org/pdf/2111.03100.pdf.

Zhang, L.-C. (2021). Discussion at ISI session 'Population statistics using administrative data instead of census', Virtual 63rd ISI World Statistics Congress, 11–16 July 2021.

Zhang, L.-C. (2022a). Graph Sampling, Chapman and Hall/CRC.

Zhang, L.-C. (2022b). Complementarities of survey and population registers, in N. Balakrishnan, T. Colton, B. Everitt, W. Piegorsch, F. Ruggeri and J. L. Teugels (eds.), Wiley StatsRef: Statistics Reference.

UNECE Group of Experts on Population and Housing Censuses, Geneva, 21-23 September 2022