CENSIMENTI PERMANENTI
POPOLAZIONE
E ABITAZIONI

Using Administrative Data for the population count in the Italian Permanent Population Census

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

CENSIMENTI PERMANENTI
L'ITALIA, GIORNO DOPO GIORNO.
POPOLAZIONE
E ABITAZIONI

19-20 May 2021 - *Transitioning to register-based censuses beyond the 2020 round* - 2021 UNECE Online Expert Meeting on Population and Housing Censuses

Istat | Istituto Nazionale di Statistica

The Italian Permanent Population and Housing Census

❖ 2018: start of the first cycle (2018-21) of the Permanent Population and Housing Census (PPHC) = combined census based on registers + *ad hoc* sample surveys, allows the yearly availability of detailed census statistics and replaces intercensal population counts based on administrative data

❖ The Population Base Register (RBI) is at the core of the PPHC. Together with the Statistical Base Register of Addresses (RSBL) and with the thematic registers on education and employment provides the basis for the production of population census data in a combined census design.

❖ *Ad hoc* surveys are used to measure coverage errors of RBI and to collect data for variables non-replaceable (or only partially replaceable) through the registers.

   ✓ Areal survey conducted on a sample of addresses drawn from RSBL in order to count and interview (CAPI technique) every usually resident household

   ✓ List survey conducted with a mixed mode technique (CAWI, CAPI, CATI) on a sample of households drawn from the PBR -

   ✓ Same questionnaire used in both surveys (includes all the hypercubes variables in order to test the quality and the coverage of data already available in registers).

CENSIMENTI PERMANENTI
L'ITALIA, GIORNO DOPO GIORNO.
POPOLAZIONE E ABITAZIONI
Istat | Istituto Nazionale di Statistica

The PPHC population count

Survey data are used to correct RBI data within a Dual System Estimation model aimed at estimating coverage errors of the register.

- RBI = first capture

- second capture = annual sample surveys + 'administrative signs of life' derived by the Integrated Administrative Data Base (AIDA) [*i.e. in order to correct for the undercoverage of the survey, non respondents with strong 'signs is of life' in AIDA are 'considered as enumerated*]

  ✓ Areal survey used for measuring the under-coverage error of RBI for each *i* municipality and *j* individuals profile (individuals usually resident in the municipality who are not included in the PBR) →

  $$p_{ij,under} = \frac{Newly\ Enumerated_{ij}}{Total\ Enumerated_{ij}}$$

  ✓ List survey + "administrative signs of life" for measuring the over-coverage error of RBI for each *i* municipality and *j* individuals profile (individuals included in the register who are no more usually resident in the municipality) →

  $$p_{ij,over} = \frac{Expected\ and\ not\ found_{ij}}{Expected\ and\ not\ found_{ij} + Expected\ and\ Enumerated_{ij}}$$

POPOLAZIONE E ABITAZIONI

CENSIMENTI PERMANENTI
L'ITALIA, GIORNO DOPO GIORNO.

Istat | Istituto Nazionale di Statistica

Population count as a result of PBR correction

❖ The ratio between the two components defines the "raw corrector" of over and under-coverage

$$corr_{ij} = \frac{1 - p_{ij,over-coverage}}{1 - p_{ij,under-coverage}}$$

▪ <u>Calculation of direct estimates</u> calibrated for over and under-coverage for each j profile ('Italian' or 'foreign') for sampled i municipalities.

▪ <u>Calculation of indirect estimates</u>: small areas estimation models are used to reduce direct estimates' variability for sampled municipalities and to calculate estimates for non-sampled municipalities

❖ At the end of the process, a 'weight' is applied to each individual in RBI (according to his/her profile):

✓ If RBI, for a given municipality, is affected by neither over-coverage nor under-coverage errors (or if the two errors compensate each other), the weight applied to residents in the register will be equal to 1.

✓ If under-coverage of RBI is > over-coverage → the corrector applied to each individual of RBI will be > 1 (and the total population will result higher than that of PBR). Vice versa, if under-coverage of PBR is < than over-coverage, the corrector applied to each RBI record will be < 1 (and the total population will be lower than that of RBI).

CENSIMENTI PERMANENTI
L'ITALIA, GIORNO DOPO GIORNO.
POPOLAZIONE
E ABITAZIONI

Istat | Istituto Nazionale di Statistica

Drivers for change and use of admin data

- Due to the pandemic, the 2020 surveys had to be canceled. Nevertheless, according to the Permanent Census design, the 2020 population data have to be released by the end of 2021.

- Need to further reducing costs and minimizing households' response burden.

- Weaknesses identified in the combined design based on the first 2 waves experiences (i.e. controversial interpretation of some of the outcomes registered in the survey monitoring system, on the basis of which non-respondent households are classified according to their "coverage status"; sampling size not allowing to calculate correctors by municipality, sex, age class and citizenship; due to the methodology applied for the correction of RBI, the result is a register with weights not a register of "heads")

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

- Feasibility study in order to try and estimate the municipality population count by sex, age and citizenship solely through the use of administrative sources

POPOLAZIONE E ABITAZIONI

CENSIMENTI PERMANENTI
L'ITALIA, GIORNO DOPO GIORNO.

Istat | Istituto Nazionale di Statistica

Challenges and opportunities of an extensive use of admin data

Is it possible to use only administrative sources (independent from the Population Register) for calculating over/under coverage errors of the Population Register and thus estimate the 2020 population count (i.e. produce reliable estimates at specified territorial domains)?

- **micro-level analysis - linkage between AIDA and RBI** in order to identify individuals eligible for under and over coverage of the Population Register - study of association patterns between RBI data and 'signs of life' from AIDA at the micro level and associations between 'signs of life' and survey data for target populations i.e. "newly enumerated" (under-coverage according to Areal survey) and for "expected not found" (over-coverage according to List survey), to define clustering of municipalities/identify subpopulations at higher risk of coverage errors. also by investigating. These patterns can represent estimation domains, deterministic rules or even useful covariates in a predictive model

- **macro-level analysis** – calculations of correctors based only on administrative data, and study of correlations between these last and those obtained trough the combined approach
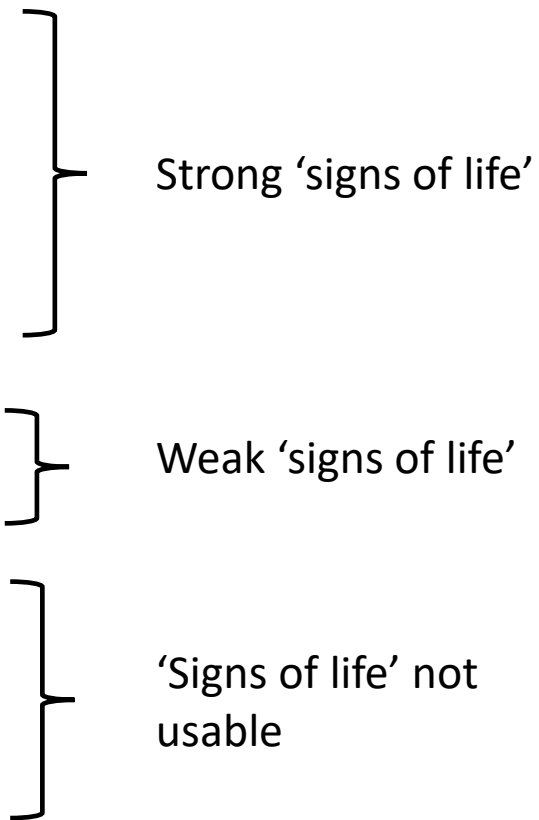
- ✓ define deterministic criteria for determining under/over coverage using the strength of the signals at individual and/or household level

- ✓ identify predictive models of the 'usual place of residence' variable; in this approach, the 'signs of life' coming from administrative data are evaluated and selected according to the type and reliability of the source, to geographical and duration patterns, and to possible associations with other individual characteristics (household relationships), with the aim of defining the place of usual residence of each individual in a model based approach

POPOLAZIONE E ABITAZIONI

CENSIMENTI PERMANENTI
L'ITALIA, GIORNO DOPO GIORNO.

Istat | Istituto Nazionale di Statistica

Individuals in Aida by type of 'sign of life'

Type of sign of life	Absolute values	% values
Steady signs of work/study	32.121.583	51,6
Signs of university ennrollment	1.438.897	2,3
Weak signs of work/study weak	1.359.072	2,2
Signs of work/study episodic = not usable	3.716.311	6,0
Rent contract	468.658	0,8
Rent contract + valid permit of stay	4.634	0,0
Pension	13.251.073	21,3
Pension + valid permit of stay	3.121	0,0
Other income sources	2.660.642	4,3
Valid permit of stay	641.083	1,0
Fiscally dependent family member	4.240.523	6,8
Indirect signs of life - several sources	293.169	0,5
Signs of Italians living abroad (Anagrafe consolare)	112.571	0,2
No signs of life	**1.968.137**	**3,2**
Total population in Aida	**62.279.474**	**100,0**

Trying to «seize» the intention of staying through continuity patterns

Tentative household criteria for evaluation of national over/under coverage

Type of household	Criteria for individuals in RBI eligible to overcoverage	Outcome
One-person household	One-person household with no signs of live over the past 5 years who don't live in border municipalities	*national overcoverage*
Individual in a multi-person household	All household members with no signs of life	*national overcoverage*
	Individual < 14 years of age attending school in the same municipality of RBI	*no national overcoverage*
	If at least one member owns a dwelling (or holds a rent contract) in the same municipality of RBI	*no national overcoverage*
	Husband/wife (even if not fiscally dependent) of worker/student/pension perceiver	*no national overcoverage*

Individuals not in RBI eligible to undercoverage	If all criteria are verified
Signs of work/study for at least 12 months over the relevant period	*national under coverage*
Owner of a dwelling (or holder of a rent contract	
No-border municipality	

Identification of subpopulations eligible to over/under coverage

Tentative classification after application of deterministic criteria		Absolute values	% values
0	Individuals present both in RBI and AIDA in the same municipality ---> confirmed in RBI	58.309.572	93,6
1	Under coverage country level ---> to be added to the count	416.807	0,7
2	National over coverage (country level) ---> to be excluded from the count	368.185	0,6
3	Grey area (potential over/undercoverage but without clear signals ---> need of further investigation	281.679	0,5
4	Local over/under coverage (different province from RBI's) ---> confirmed in a different municipality	12.626	0,0
5	Individuals not in RBI but no under coverage (weak signs of life) ---> not to be included in the population count	895.208	1,4
6	Individuals in RBI with no 'signs of life' ---> need of further investigation	1.599.952	2,6
7	Currently not classifiable - ---> need of further investigation	395.445	0,6
Total population in AIDA		**62.279.474**	**100,0**

Conclusions

Two possible scenarios for calculating over/under coverage errors of the Population Register and estimating the 2020 population count

A – Use of administrative data (deterministic and/or model based approach)

✓ definition of criteria for predicting 'the usual place of residence'

✓ individuation of 'grey areas' (subpopulation or municipalities with specific characteristics) → what to do for subpopulations for which signs of life are NOT sufficiently predictive?

B – Use of corrector (weight applied to each record of RBI) = apply the 2018/2019 corrector for over and under coverage to the 2020 RBI

REFERENCES

Bernardini A., Cibella N., Gallo G., & Al. (2019), "Empirical evidence for population counting: the combined use of administrative sources and survey data". Paper presented at ESS Workshop on the use of administrative data and social statistics, Valencia, June 4th 2019

Chieppa A., Gallo G., Tomeo V. & Al. (2018), "Knowledge discovery for inferring the usually resident population from administrative registers". In Mathematical Population Studies. International Journal of Mathematical Demography, Published online: 27 Jul 2018. http://www.tandfonline.com/loi/gmps20

Chieppa A., Ferrara R., Gallo G., Tomeo V. (2017), «Towards The Register-Based Statistical System: A New Valuable Source For Population Studies". Presentazione al Convegno Statistics and Data Science: New Challenges, New Generations. University of Florence, June 2017

Kim J.K., Rao J.N.K. (2012), Combining data from two independent surveys: a model assisted approach, Biometrika, Vol. 99(1), 85-100.

Nirel, R. , Glickman, H. (2009), Ch. 21 - Sample Surveys and Censuses. In: Rao, C.R. (ed.) Handbook of Statistic, Elsevier.

Pfeffermann D. (2015), METHODOLOGICAL ISSUES AND CHALLENGES IN THE PRODUCTION OF OFFICIAL STATISTICS. Journal of Survey Statistics and Methodology 3, 425–483.

GRAZIE PER L'ATTENZIONE

Bernardini, A. Chieppa, N. Cibella, G. Gallo, D. Zindato

anberna@istat.it, chieppa@istat.it, cibella@istat.it, gegallo@istat.it, zindato@istat.it

CENSIMENTI PERMANENTI
L'ITALIA, GIORNO DOPO GIORNO.

POPOLAZIONE E ABITAZIONI

Istat | Istituto Nazionale di Statistica