The integration of administrative data sources in Italy to increase Population Census data availability

Gallo G, Chieppa A., Tomeo V. & Falorsi S.
Italian National Institute of Statistics (Istat)
Population Census and Integration of Social Surveys Division)

Group of Experts on Population and Housing Censuses Eighteenth Meeting
Geneva, 28 - 30 September 2016 Item 5 of the provisional agenda Methods for assessing quality and usability of registers and administrative sources

Censimento PERMANENTE della popolazione e delle abitazioni

Istat Istituto Nazionale di Statistica

The Population Census before 2011

A decennial population census has been taken in Italy since 1861

- complete field enumeration (so called “door-to-door” enumeration)
- census forms delivered and collected by enumerators and self-filled in by respondents
- all information collected and processed on a complete basis
- same economic, human and organisational resources were allocated to every household

The 2011 census strategy

Combining the study of census experiences of other countries with a more effective use of administrative data held by Municipality Population Registers (MPRs) ⇒ completely new strategy

- methodological and technical innovations
- adoption of standardised solutions according to municipality size (necessity to differentiate census organization according to the needs and capacities of the different actors)
- use of new territorial instruments designed to improve coverage and quality of the enumeration

Lessons learnt

- Spontaneous return went very well
- Web return (34,5%) was much higher than expected (no particular emphasis in the communication campaign)
- Logistics of the *mail out/mail back* process proved very complex and with many potential (and actual) points of failure
  - ✓ Not all addresses in the municipal registers could be processed by the contractor system
  - ✓ Many questionnaires could not be delivered for various reasons (about 2,000,000 had to be delivered directly by the enumerators)
  - ✓ Information about questionnaires returned to post offices was not always reported promptly, making it impossible for municipal offices to keep track of which households had to be contacted to prompt completion of the questionnaires

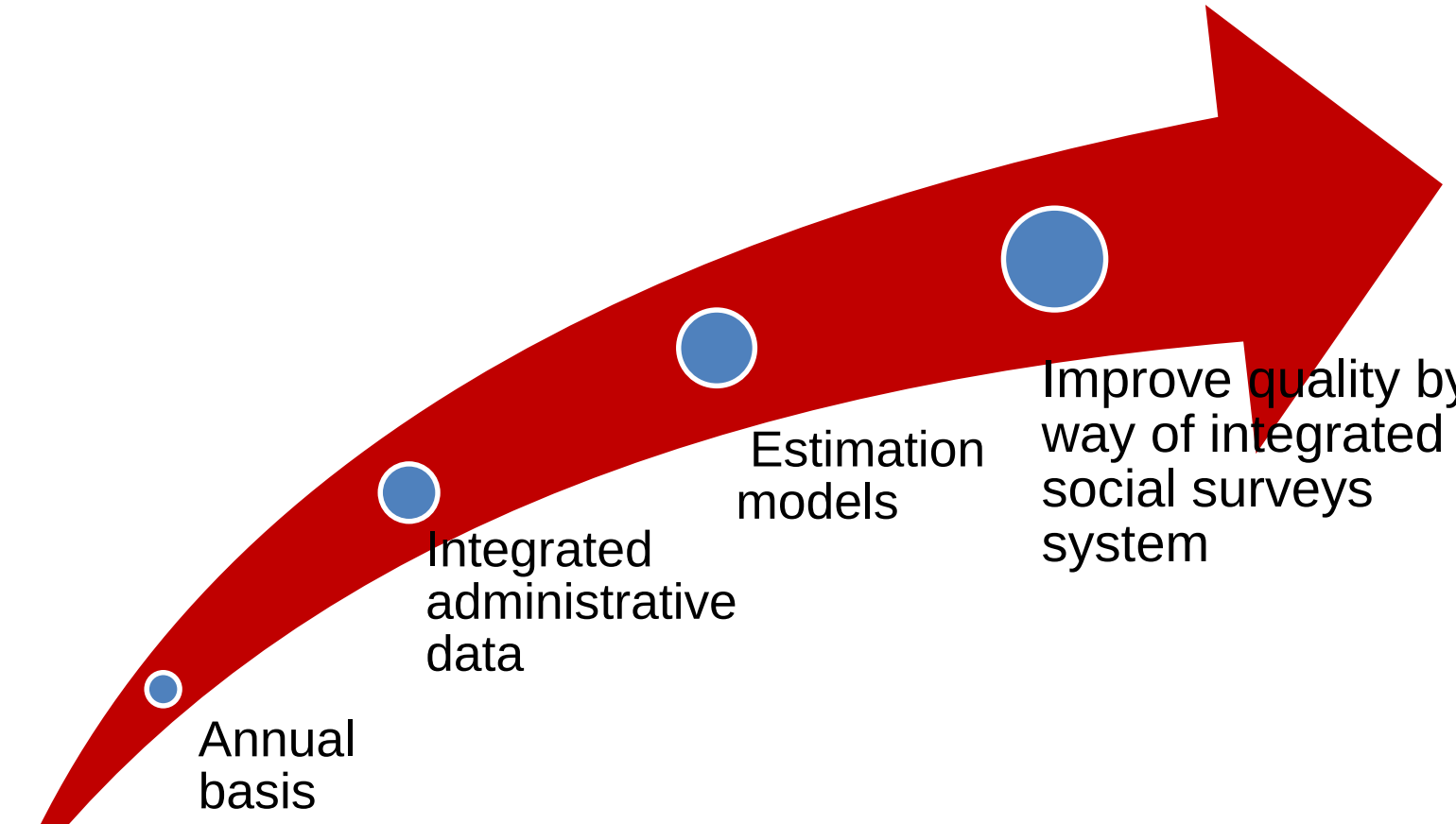
Still need for a change

In spite of their improvements, the innovations designed for 2011 census are not enough to achieve a stable and enduring balance between census costs and benefits.

- Costs remain high and too concentrated in time, while the use of administrative data is not tested to the potential offered by the Italian context.
- Census data continue becoming quickly out-dated, and the supply of highly detailed geographic data remains only decennial.

⇒ A totally different approach seems necessary, combining a most effective use of administrative sources with the spread over years of the census fieldwork.

… a bridge towards change to reduce costs



ADMINISTRATIVE DATA FOR POPULATION CENSUS

- A specific task-force of ISTAT is in charge of investigating how to identify usually resident population from administrative sources in order to lead the transition to a Population Census register-based

- This presentation shows goals, approach in investigation and first results of the tasks carried out by ISTAT

Key idea of the work is the use of a continuously improving process that combines knowledge discovery from databases with expert/domain guidance

Modernization project towards a Register based statistics… continuously updated

- Production and use of a statistical population register
- Base-registers linked at the micro-data level by using a unique ID number
- Continuous update from administrative sources and integrated social surveys system

ISSUES INVOLVED IN THE INVESTIGATION

- To design and build a Data Base with data coming from qualified administrative data, to allow investigations on quality of Population Register
- To define dimensions (attributes, time periods and so on) to be taken into account when assigning usual residence
- To identify relevant population groups when evaluating under and over coverage of the population register
- To highlight crucial points involved in managing such Data Base/attributes/procedures

The results are very important when deciding to which extent administrative databases could be used for the next Population Census round

GENERAL FRAMEWORK AND STARTING POINTS

- EU Regulations on *usually resident* population
- ISTAT Data Base where administrative data are integrated (SIM)
- Experts choice of qualified administrative sources and related attributes, relevant when managing the Population Register
- *Knowledge discovery from databases* (KDD) approach

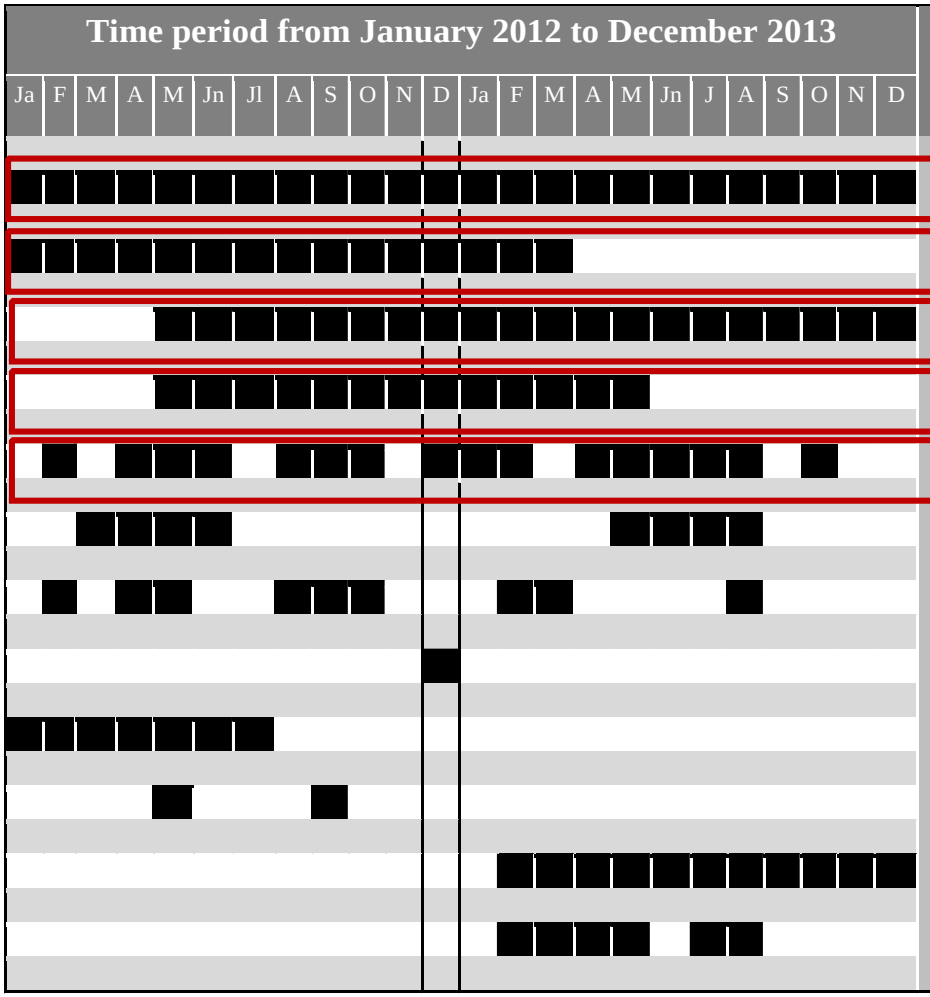
FRAMEWORK: SIM DATABASE

Istat Repository of **integrated administrative microdata** built with the aim of supporting the statistical production processes both for social and economic statistics

- Each object or unit (person; economic unit; places; relations) has been assigned an ID number, which remain the same for different tables/time periods

- Logical and physical relations have been designed, regarding time and spaces, between different sources

Monthly presence scheme of continuity's patterns in job and study activities (direct signals)



Profile of presence in employment and educational registers	
1	Steady
2	Outgoing signals with presence in December 2012
3	Ingoing signals with presence in December 2012
4	Signals of presence around December 2012
5	Not steady, at least 12 months
6	Seasonal signals
7	Less than 12 months, not seasonal
8	Just one signal of presence in December 2012
9	Presence only before December 2012
10	Random signals only before December 2012
11	Ingoing signals after December 2012 and presence till 2013
12	Ingoing signals after December 2012 but not till December 2013

Proceeding scheme and counts of population groups (000)

Step I					Step II			Step III		
PR	LE	PS	Group	Count	RR-NPR	Group	Count	TR	Group	Count
61,068	37,704	3,378			20,764			26,649		
Signals					Signals			Signals		
Yes	Yes	-	**A**	**36,618**						
Yes	No	-	B	24,450	Yes	**E**	**14,485**			
					No	F	9,965	Yes	**G**	**6,939**
								No	H	3,026
No	Yes	-	C	1,086						
No	No	Yes	D	351						

Table legend:

Eligible as usually residents

Uncertain residents

Administrative sources legend:

PR	Population Register
LE	Labor and Education Registers
PS	Permits to stay Register
RR-NPR	Retired and Non-Pension Benefits Registers
TR	Tax Returns Register

Retired people

Dependents

People in PR with signals in other sources

People in PR with NO signals in other sources

People with signals only in administrative sources

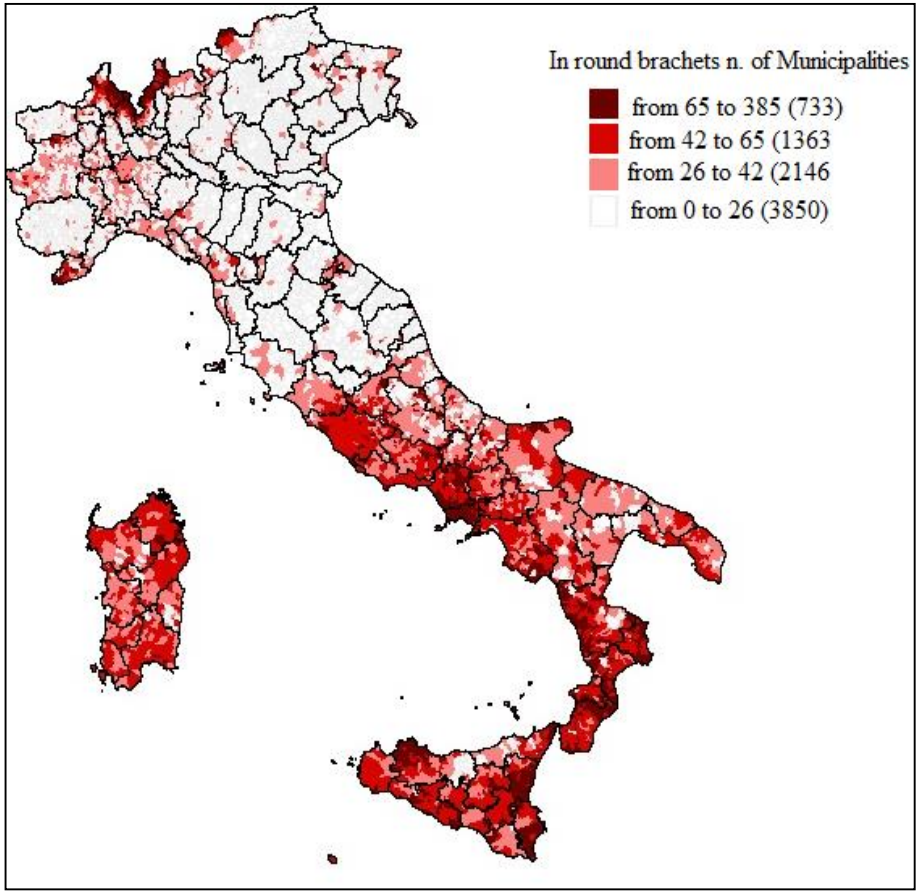
People with only Permits to stay

About 410,000 People with strong – steady signals of usual residence

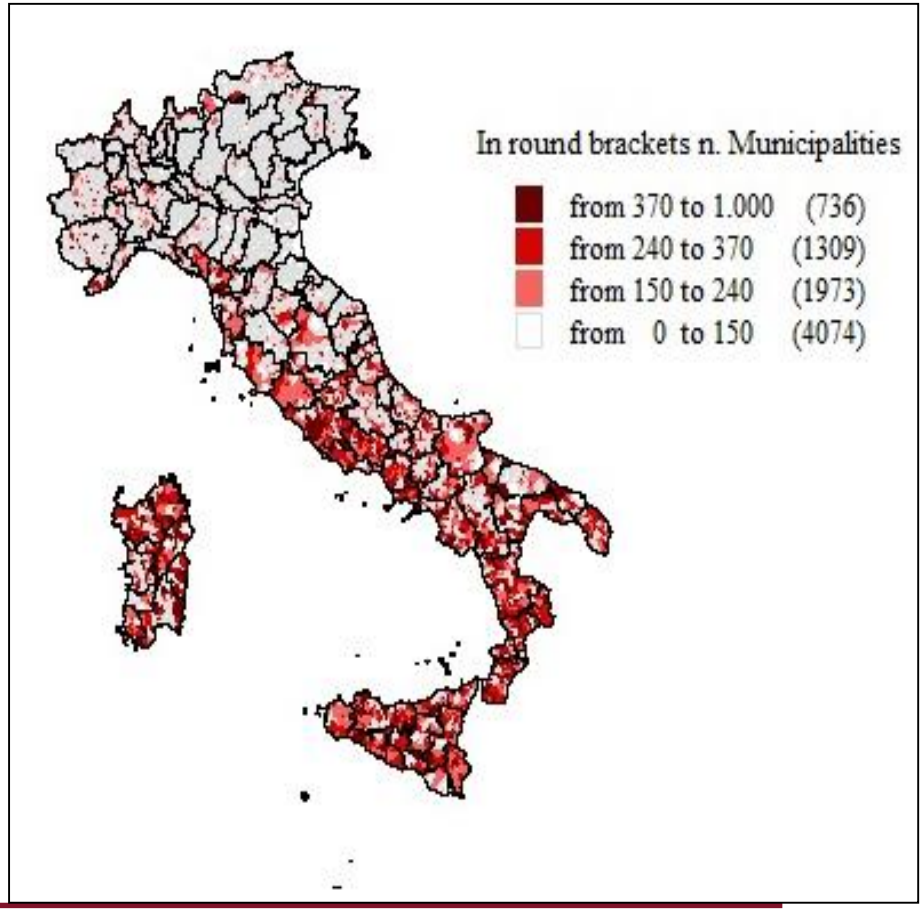
Istat

Subpopulation H: people recorded in the population register without signals from other sources

Italian citizens (about 2,2 million people)
(Values per 1,000 Italian people)



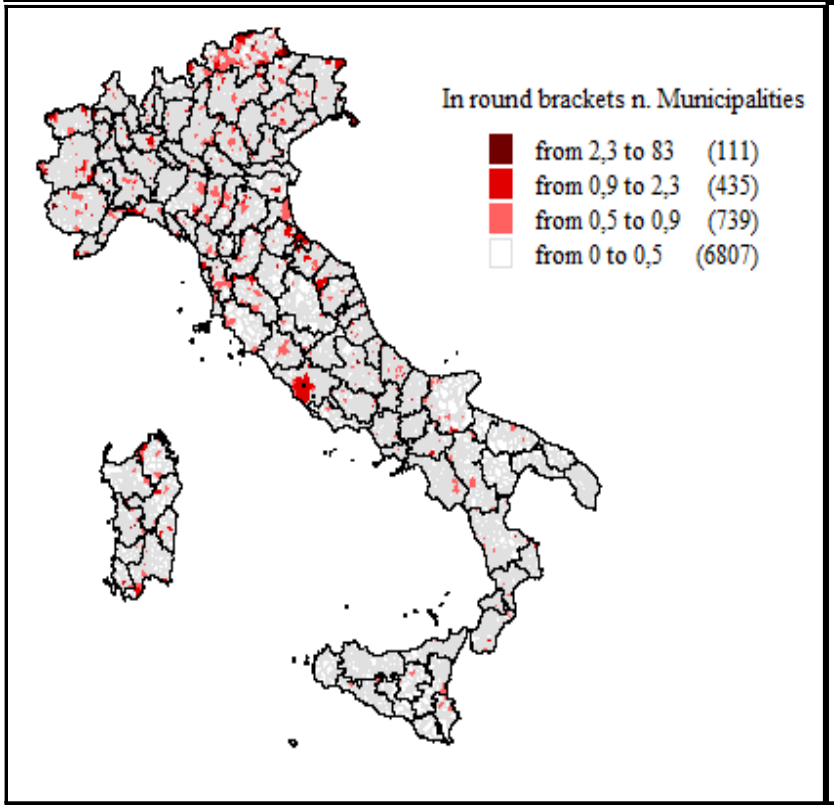
Foreign citizens (about 800,000 people)
(Values per 1,000 Foreign people)

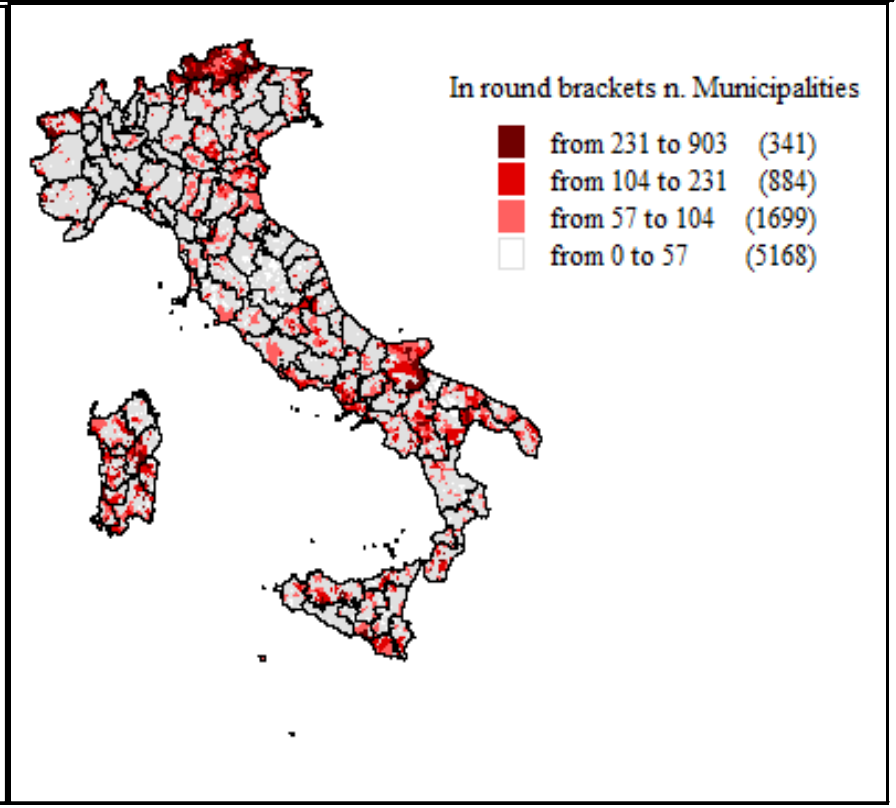


Istat

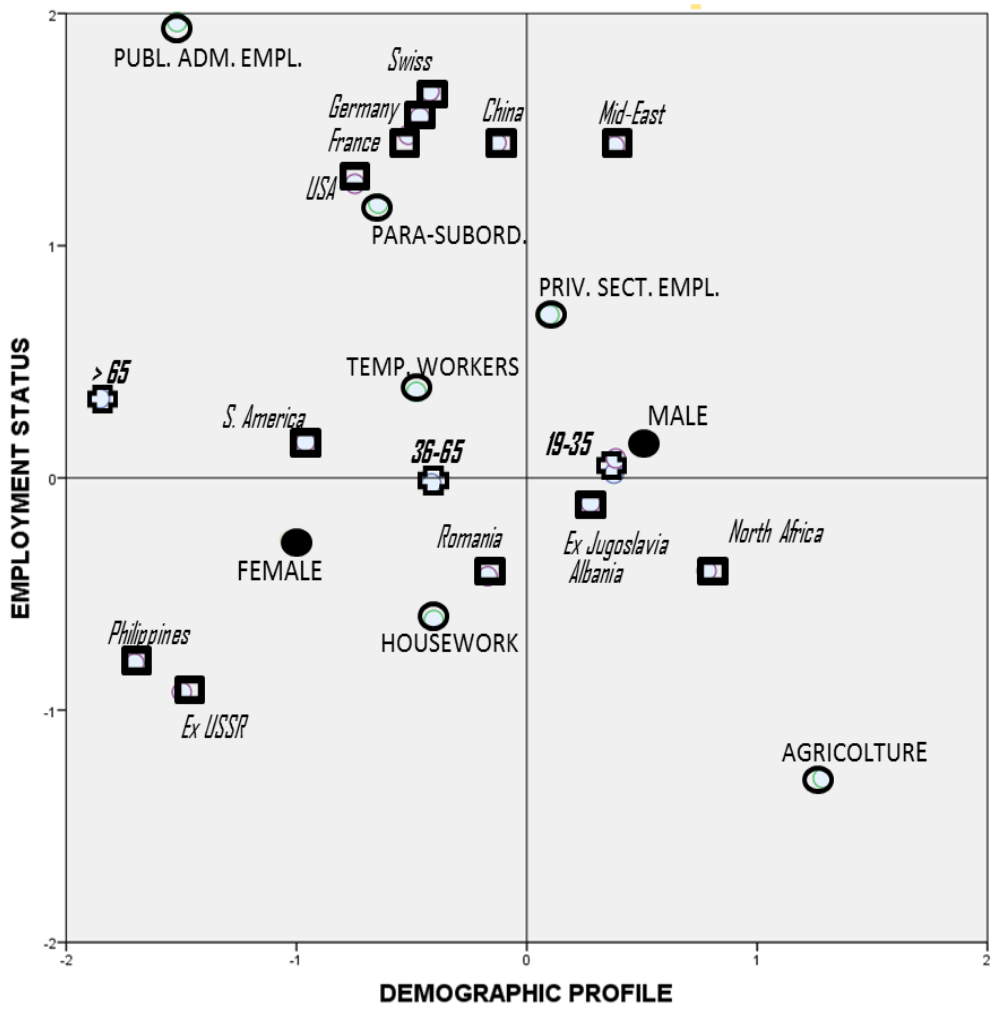
Subpopulation C1: people not recorded in the population register with strong-steady signals of usual residence

Italian citizens *(Values per 1,000 Italian people)*	**Foreign citizens** *(Values per 1,000 Foreign people)*





Subpopulation C1: Foreign workers with strong-steady signals and their association between the categories



Gender

Job Type

Citizenship

Age

Istat

Conclusions(1)

Signals of presence on the territory are extracted from the sources and can be used effectively to improve the quality of population registers

ISTAT defined a preliminary workflow to integrate the use of administrative sources and the official population registers in order to calculate the usually-resident population

Using this workflow, it is possible to define a group of individuals eligible to be included in the usually-resident population of Italy at a given reference date

Conclusions(2)

Demographic variables, especially those of gender, age and country of citizenship as well as the location of the signal on the territory have proved to be very significant variables for defining specific sub-population profiles

The division into sub-groups and clusters allows the identification of subpopulations (for example, the typical foreign communities that elude the registers, but perform specific labor activities)

These same groups can represent the basis for defining a census strategy formulated on the use of mixed techniques that combine specific surveys and appropriate statistical models

The analysis of signal strength based on its continuity over time is only the starting point in the use of longitudinal data that can be processed with administrative sources