

The Strategy for Validating the Ultracentenarians Enumerated in the 2001 Population Census

*La Strategia di Validazione degli Ultracentenari Rilevati al Censimento
della Popolazione del 2001*

Alessandra Nuccitelli, Anna Pezone, Alessandra Reale

Istituto Nazionale di Statistica, Roma

nuccitel@istat.it, pezone@istat.it, reale@istat.it

Riassunto: attualmente, il censimento della popolazione rappresenta la principale fonte disponibile in grado di fornire il quadro più dettagliato della popolazione ultracentenaria abitualmente dimorante in Italia; inoltre, può costituire un importante punto di partenza per l'eventuale costruzione di un archivio specifico e aggiornato mediante integrazione con altre fonti, tra cui quelle anagrafiche. In questo lavoro si descrive la strategia utilizzata per la validazione del numero di ultracentenari censiti e si presentano i principali risultati ottenuti.

Keywords: Data quality, Editing, Linkage, Population register.

1. Introduction

Centenarians represent a relatively small proportion of the total Italian population. Yet, researchers have long been interested in people aged 100 or above in order to secure information that might increase understanding of the ageing process and on the characteristics of persons who succeed in living to an extremely high age. In particular, observation of centenarians can aid those studying the human life span and the trajectory of mortality. However, in Italy, as well as in other countries, the number of people who are 100 years and over is not known exactly. Generating a count of centenarians is hounded by problems that may result in an overestimate of the “true” size of this segment of population. The Census, like any data collection system that involves measuring the elderly population, can potentially be affected by the matter of the accuracy of the recorded date of birth. Specifically, errors may arise either from the phase of completion of the questionnaire or from that of data capture by optical reading. With regard to the former phase, low literacy levels and functional (or cognitive) disability can make it difficult for some elderly to correctly report their own date of birth or necessary to rely on the knowledge of others; in addition, it is well known (Krach and Velkoff, 1999) the tendency on the part of respondents (either the individuals, their relatives, or other cohabitants) to overstate or to round off the age of very old persons. The next section focuses on the strategy used for validating the centenarians enumerated in the 2001 Population Census. The procedure implementing this strategy is an integral part of the process of editing and imputation of the Census data. It is worth noting that the recent introduction of advanced information technologies in the collection and processing of statistical data has played an important role in achieving the reported results. A brief descriptive analysis of the validated centenarians is presented in the section 3 and some conclusions are drawn in section 4.

2. The strategy for validation

The strategy for validating centenarians is a combination of several methodologies directly applied to the raw micro-data. Records are related to individuals usually living in households or in institutional households.

The main key stages of this strategy are as follows:

- 1) automatically match the records which refer to individuals born in the period 1888-1901 and enumerated in the 1991 Census, with the records which refer to individuals born in the same period and enumerated in the 2001 Census, using exact (i.e. deterministic) matching;
- 2) automatically check for internal consistency of records related to the same questionnaire;
- 3) manually check for consistency of the records with the respective questionnaire images which are stored during the process of optical reading.

With regard to the matching stage, two records are labelled as linked if the following conditions are simultaneously satisfied:

- a) records perfectly agree on fields common to both data sets (matching variables);
- b) the used matching key¹ identifies the links uniquely and does not contain missing values.

The individuals (enumerated in the 2001) corresponding to linked records are considered validated. For people living in households, records can be linked:

- *at household level* by keys made up of all or nearly all the following matching variables related to the individual born in the period 1888-1901 and to each relative or cohabitant living with him: *province of residence, municipality of residence, sex, day of birth, month of birth, year of birth*;
- *at individual level* by keys made up of all or nearly all the following matching variables related to the individual born in the period 1888-1901: *province of residence, municipality of residence, sex, day of birth, month of birth, year of birth, month of marriage, year of marriage, birthplace*.

For people living in institutional households, records are linked only *at individual level* by a key made up of all the following matching variables related to the individual born in the period 1888-1901: *province of residence, municipality of residence, sex, day of birth, month of birth, year of birth, marital status, birthplace*.

The unlinked records from the 2001 Census are submitted to a rigorous check for internal consistency. The *year of birth* is checked for consistency with the following data fields related to the individual: *highest educational qualification, occupational status, occupation, place of work or study*; for individuals living in multimember-households, also other variables related to relatives and/or cohabitants are considered (*relationship with the householder, sex, year of birth*). Individuals presenting consistent items are considered validated. In some ambiguous cases, records are clerically checked for consistency with the respective questionnaire images.

⁽¹⁾ A matching key is a combination of matching variables. It is important to note that matching variables have a different identifying power and may contain errors. So, the used matching keys consist of all or nearly all the common fields and are sequentially used in decreasing order according to their identifying power. In fact, a key as the date of birth (composed by *day of birth, month of birth, year of birth*) can take (365 days in a year × 14 years) 5,110 value states for an individual born in the period 1888-1901, but could not impart enough information to identify a link uniquely, because of the large size of the involved data sets (13,200 records for the 2001 Census and 211,869 records for the 1991 Census).

Results of the validation are shown in Table 1. Among the 7,035 validated individuals, 30.6% and 16.1% were respectively linked at household and individual level.

Table 1: *Individuals by year of birth before and after the validation*

<i>Year of birth</i>	<i>Individuals before the validation</i>	<i>Individuals after the validation</i>
1888	63	-
1889	70	8
1890	46	3
1891	54	10
1892	77	11
1893	82	23
1894	88	43
1895	112	57
1896	217	146
1897	296	233
1898	577	504
1899	920	848
1900	4,675	1,978
1901	5,923	(*) 3,171
<i>Total</i>	13,200	7,035

(*) Of these 3,171 individuals born in 1901, 722 turned a hundred only after the Census date.

3. Main results

Characteristics presented are based on the validated data and are related to the 6,313 individuals classified as centenarians at the date of October 21, 2001; of these, only 16 percent were living in institutional households. As the 1991 Census validated 3,345 centenarians out of 56,778,031 inhabitants², within a decade, the ranks of the people aged 100 or above have nearly doubled. As with the elderly population as a whole, women outnumber men; table 2 shows that about five in six centenarians are women. Although both women and men have experienced dramatic improvements in mortality at the oldest ages over the last few decades, gains for females have typically been higher than those for males (Caselli *et al.*, 2000). Table 3 shows that the regions with the greatest number of centenarians are Lombardia (940), Piemonte and Valle D'Aosta (613), Emilia Romagna (597) and Toscana (506). The top two regions with the highest proportion of centenarians among its own population differ from those with the greatest number and are Liguria and Friuli Venezia Giulia with approximately 2 centenarians out of every 10,000 residents. With regard to geographical variations in the sex ratio³ by region of residence (see column 6 in table 3), the North of Italy shows the highest values, with Friuli Venezia Giulia peaking at over 11 females per one male. Instead, for Calabria and Sardegna the sex ratios are more balanced (2.15 and 2.53, respectively).

Table 2: *Ultracentenarians by sex and age*

<i>Age</i>	100	101	102	103	104	105	106	107	108	109	110	111	112	<i>Total</i>
<i>Males</i>	519	306	111	55	37	21	9	10	2	4	2	1	3	1,080
<i>Females</i>	2,401	1,405	658	381	172	105	48	30	12	10	5	3	3	5,233
<i>Total</i>	2,920	1,711	769	436	209	126	57	40	14	14	7	4	6	6,313

⁽²⁾ The total number of residents enumerated in the 2001 Population Census is 56,995,744.

⁽³⁾ The reported sex ratio is calculated by dividing the number of females by the number of males.

Table 3: *Ultracentenarians by region of residence and sex; ultracentenarians per 10,000 residents and female-to-male sex ratio by region of residence*

<i>Region of residence</i>	<i>Males</i>	<i>Females</i>	<i>Total</i>	<i>Ultracentenarians per 10,000 residents</i>	<i>Sex ratio</i>
<i>Piemonte and Valle D'Aosta</i>	94	519	613	1.41	5.52
<i>Lombardia</i>	117	823	940	1.04	7.03
<i>Trentino Alto Adige</i>	12	106	118	1.26	8.83
<i>Veneto</i>	70	472	542	1.20	6.74
<i>Friuli Venezia Giulia</i>	18	208	226	1.91	11.56
<i>Liguria</i>	38	282	320	2.04	7.42
<i>Emilia Romagna</i>	95	502	597	1.50	5.28
<i>Toscana</i>	70	436	506	1.45	6.23
<i>Umbria</i>	18	71	89	1.08	3.94
<i>Marche</i>	26	156	182	1.24	6.00
<i>Lazio</i>	90	309	399	0.78	3.43
<i>Abruzzo</i>	31	140	171	1.35	4.52
<i>Molise</i>	6	38	44	1.37	6.33
<i>Campania</i>	86	300	386	0.68	3.49
<i>Puglia</i>	67	257	324	0.81	3.84
<i>Basilicata</i>	11	32	43	0.72	2.91
<i>Calabria</i>	68	146	214	1.06	2.15
<i>Sicilia</i>	110	302	412	0.83	2.75
<i>Sardegna</i>	53	134	187	1.15	2.53
<i>Italia</i>	1,080	5,233	6,313	1.11	4.85

4. Concluding remarks

The aim of this work is not to provide the definitive answer to the number of people aged 100 or above living in Italy at the date of the 2001 Census, but to take a first step to get more accurate information on this important segment of the population. Because of the rigorous rules adopted for validation, the provided count could even be an underestimate of true size of the centenarian population and, as a consequence, mentioned findings for centenarian characteristics could be biased. As suggestion for future work, centenarians enumerated in the Census could hopefully be ascertained or added by using other sources of data, including birth and death records, with the goal of establishing a complete and up-to-date database. Besides, this would enable to estimate the degree of census under-coverage for this segment of the population. At this moment, only a few developed countries in the world can provide useful information for studying human longevity. In fact, as Siegel and Passel aptly noted in 1976, obtaining high quality data on the oldest old population depends upon a long standing, reliable population register or vital registration system in existence for more than 100 years.

References

- Caselli, G., Vaupel, J. W., Yashin, A. I. (2000) Longevity, Heterogeneity and Selection, in: *Atti della XL Riunione Scientifica della Società Italiana di Statistica*, Firenze, 26-28 aprile 2000, 49-72.
- Krach, C. A., Velkoff, V. A. (1999) *Centenarians in the United States*, Current Population Reports, Series P23-199RV, U. S. Bureau of the Census, U. S. Government Printing Office, Washington, DC.
- Siegel, J. S., Passel, J. S. (1976) New Estimates of the Number of Centenarians in the United States, *Journal of the American Statistical Association*, 71, 355, 559-566.