Geography of Italian centenarians and semi-supercentenarians: statistical evidences in territorial longevity differences

Graziella Caselli*, Marco Battaglini**, Giorgia Capacci**, Silvia Capuano**, Gianni Corsetti**

*Department of Statistical Sciences, Sapienza University of Rome, Italy
** Italian National Institute of Statistics, Rome, Italy

Background and aims

Following the IDL (International Database of Longevity) protocol, the Italian National Institute of Statistics (Istat) has been collecting and validating data about living and dead semi-supercentenarians since 2009, as a first stage of the SemiSuperCentenarian (SSC) survey (Caselli et al. 2018a, 2018b, 2019). On January 1st 2019, ten years after the survey began, 5,866 living and dead individuals aged 105 and over were collected and validated for the cohorts born between 1896 and 1913, 5,151 women (88%), and 715 men (12%).

In a previous study, by using the region as the territorial unit considered, we showed that there are significant geographical differences among semi-supercentenarians, which are partly similar to those already observed in some studies on centenarians (Caselli et al. 2018a, 2018b, 2019). Obviously, in order to analyse these differences net of the size of the population it is necessary, first of all, to bear in mind that Italy is a country where the cohorts of the study were subject to significant migration flows, both to other countries and within the country.

Needing a measurement that takes into account the role played by national and international migration, the semi-supercentenarians rate was calculated for each region as $SSCR_{60}$, which is the relation between the individuals (living or dead) of the cohorts 1896-1913 and the number of survivors of the same cohorts at age 60 -i.e. 45 years earlier, ($SSCR_{60} = {}_{R}N_{105}/{}_{R}N_{60}$ per 100,000). This indicator excludes not only the differential effect of migrations (supposing them to be zero for the cohorts studied after 60 years of age) but also the direct differential effects of the number of births, and of mortality before age 60 (Robine et al. 2006).

If the geography of the $SSCR_{60}$ does not wholly reflect the classical socio-economic one of an Italy divided between North/Centre and South (to the advantage of the former), the geography related to ${}_{M}SSCR_{60}$ calculated for men only aged 105+ reproduces that dichotomy (Caselli et al 2019) as confirmed by the Femininity Ration (FR).

In this study, we aim to move from description to some explanatory hypothesis of geographical differences of semi-supercentenarian characteristics by applying some statistical models at both aggregated and individual level to the extinct cohorts 1896-1907 of the SSC survey.

Data and methods

Until 2009 a group of Italian researchers from the Italian National Institute of Statistics (Istat) and from the Department of Statistical Sciences of Sapienza University in Rome have been involved in collecting data on semi-super and supercentenarians. Only age-validated individuals are included in the database. The database allows to test theories of aging and mortality at extreme ages. Until 2008 the only source of information was the Causes of Death survey realized by Istat, hence only deceased people were included.

Since 2009 Istat has started a new survey "Rilevazione della popolazione Supercentenaria" (hereafter SSC survey) in order to collect data about alive and deceased semi-super and supercentenarians (Caselli, Battaglini et Capacci, 2018a, 2018b; Caselli et al., 2019).

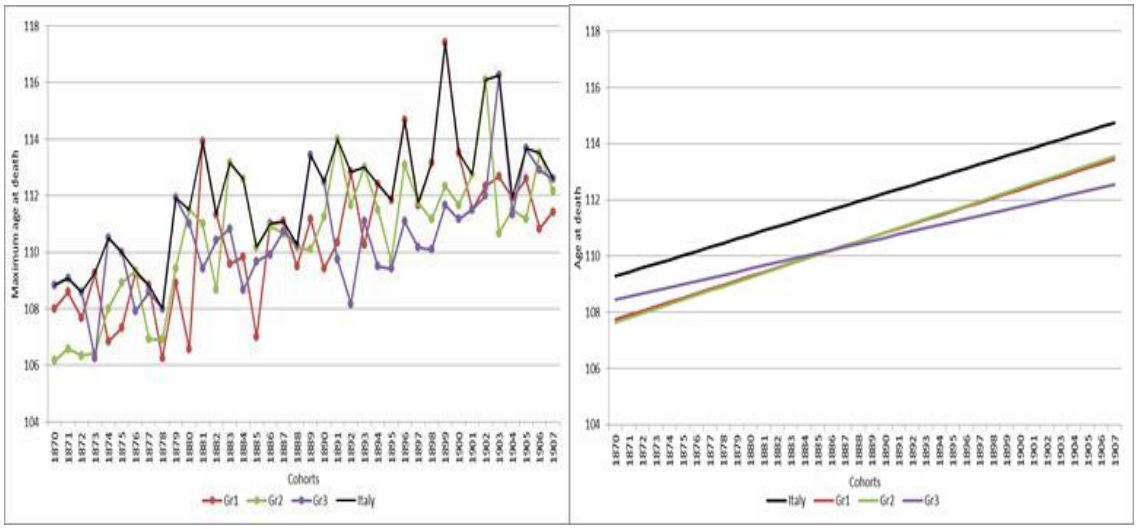
The SSC survey enables the identification of the population above age 105 cleaned from those individuals who cannot be "validated". The initial data source for the validation process is the

"Resident Population by Age, Sex, and Marital Status" (POSAS survey). This is a survey carried out by Istat on the resident population in any of the Italian municipalities (about 8,000), broken down by gender, cohort, and marital status on January 1st of each year. When the information collected by municipalities is received by Istat each municipality, where a person aged 105 or more is present, has contacted in order to collect additional information.
As long as the individual remains alive, the validation process never stops and each single year Istat proceeds to "re-validate" backward the data observed in the past. Therefore, the data quality check for the SSC's database is performed retrospectively day by day.
Regarding centenarian deaths for the cohorts 1870–1913, they were reconstructed by referring to information on the dead individuals over age 105 who had been collected in the SSC survey and to data of the Cause of Death Survey.
In order to study the distribution of the duration of life we have applied to the data of 1870-1907 cohorts the quantile regression. Referring to the statistical multi-regression model used for the explicative analyses, we have considered individual variables, such as sex, year of birth, place of birth, etc., and contextual variables such as hospital emigration, average annual amount per capita of retirement income, etc.

Preliminary results

To better understand the geographical differences of the semi-supercentenarians, it may be useful to start from age 100, also considering a larger number of cohorts that are now extinct. The analysis will refer to male and female deaths in the cohorts 1870-1907 overall, which number 65,171 for the whole of Italy. In the first step, applying the linear regression and considering the maximum age at death, we can see, as was to be expected, that there is an increase everywhere as we move from the first cohort to the last. In the South, for the cohorts 1870-1886 the regression line of the maximum age at death is higher than that of the North and Centre North (which are equal to each other), as the deaths of centenarians are numerically superior.

Figure 1. Maximum age at death for the cohorts 1870–1907, with linear regression and confidence intervals at 95%. Each point represents one individual. Italy and 3 regional groups, as G1-North, G2-Centrer North, and G3-South



Note 1: Grouping created starting from the average and the standard deviation of territorial distribution of the FR. (The classification of regions in each group follows the intensity of the FR from highest to lowest. For the North, *Gr1* include Lombardy, Friuli-Venezia Giulia, Veneto, Piedmont and Valle d'Aosta. For the Centre North and Centre, *Gr2* include Liguria, Trentino-Alto Adige, Tuscany Marche, Lazio, Umbria, Emilia Romagna). For the South *Gr3* include Puglia, Sicilia, Campania, Abruzzo and Molise, Calabria, Sardinia, Basilicata)

The opposite can be seen for the succeeding cohorts, 1887-1907, where there is a gradual deviation of the line as we move from the first to the last of these cohorts, with maximum distances between the cohorts in the SSC survey (1896-1907).
A lower number of centenarians than elsewhere, which, as has been shown in a previous study (Robine et al 2005), does not depend on a greater elimination through death at earlier ages, but on greater emigration from the South probably during adult age.
Returning to the maximum age at death and its lines of regression, it is useful to recall their descriptive and applicative limitations, as linear regression presents a parsimonious representation of the relationship between lifespan and cohort, and so cannot provide important details. The use of quantile regression (Medford A. et al. 2018) has the advantage of studying the distribution of the duration of life, evaluating the trends in more quantiles at the same time and thus obtaining a more complete picture of how age at death and, hence, the entire duration of life changes from one cohort to the next.

Figure 2 Quantile regression at percentiles 10, 30, 50, 70, 90, 95, 96, 97, 98, 99. Each point represents ONE individual at the age of death by cohort of birth. In red the percentiles from 90 on. (Italy and the 3 regional groups indicated in note to Figure 1)

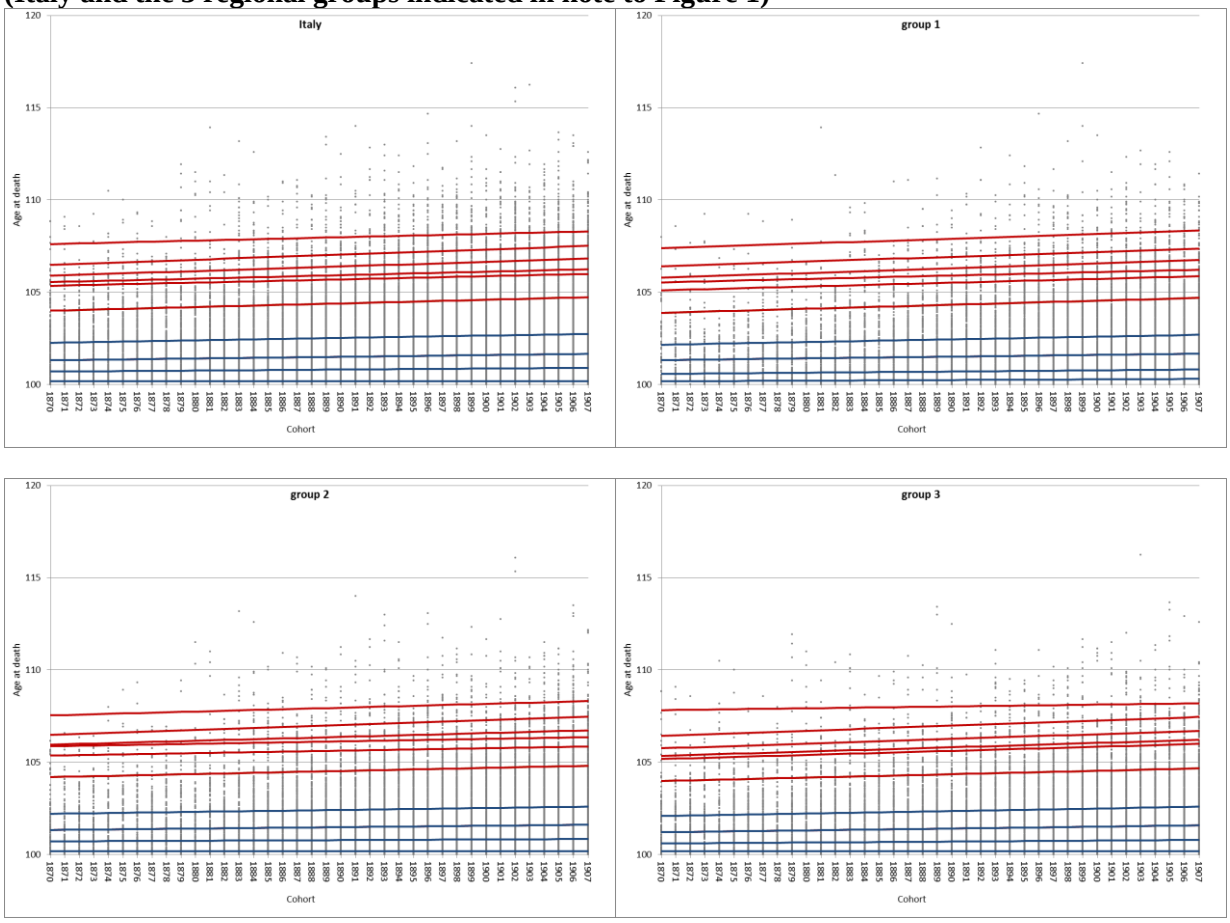


Figure 2 shows the result of applying quantile regression to the distributions of deaths by cohorts 1870-1907, for the three regional groups and for Italy. The percentiles from 10 to 70 are highlighted in blue, with increases of 20%, and the percentiles 90, 95, 96, 97, 98 e 99 in red. It is after percentile 95, corresponding to ages 105-106, that territorial differences are most marked, with values for the South of age 108 for the last percentile, which is more or less the same for all cohorts (the line is close to horizontal), showing the lack of any meaningful relation between age at death and cohort in these regions after age 108. By contrast, the same percentile shows a growing trend, from a cohort to the next, in the regions of the North and Centre. If, at the same

time, we observe the highest percentiles for the most recent cohorts, we see that the horizontalization of the slopes in the South is also present for the cohorts in the SSC survey (1896-1907). The semi-supercentenarians of the South are thus less long-lived than their contemporaries residing in the richer regions of the country.
We believe this descriptive analysis enables us to advance some explanatory hypotheses for the regional differences, in $SSCR_{60}$ and in FR, suggesting we consider in a succeeding multi-regression analyses limited to individuals in the cohorts of the SSC survey, the individual variables deriving from birth and death certificates, as well as other variables both representative of other aspects of SSC's life and socio-economical and health aspects of the regions (some Well-being composite indicators). The analyses confirm the hypotheses that the contextual variables explain the geography of the $SSCR_{60}$, penalizing Southern regions, and the FR dichotomy between North/Centre and South.

Conclusions

We can see that among centenarians those who reach age 105 are fewer in the South show the presence of a greater difficulty for all cohorts to become semi-supercentenarians. Referring to our hypotheses the multiple regression analysis confirmed the role of the context in which the semi-supercentenaries are resident in determining the observed geographical differences. On the contrary, the only individual variables collected by birth and death certificates are not able to give meaningful answers. We believe that, for this reason, it is necessary to start ad hoc surveys similar to those realized in Sardinia (Deiana et al., 1999).

References

- ✓ Caselli G., Battaglini M., Capacci G., Capuano S., Corsetti G. (2019). Italian Centenarians and Semi-supercentenarians Surveys. In: Gu D., Dupre M. (eds), Encyclopedia of Gerontology and Population Aging. Springer, Cham.
- ✓ Caselli G., Battaglini M., Capacci G. (2018a). Beyond one hundred: A cohort analysis of Italian centenarians and semi-supercentenarians. J Gerontol B Psychol Sci Soc Sci, Mar 26, gby033.
- ✓ Caselli G., Battaglini M., Capacci G. (2018b). Cohort Analysis of Gender Gap after One Hundred Years Old: The Role of Differential Migration and Survival Trajectories. J Aging Sci, 6, 199.
- ✓ Deiana L. et al. (1999). AKEntAnnos. The Sardina study of extreme longevity. Aging 11.
- ✓ ISTAT, http://dati.istat.it (Istat database)
- ✓ Medford A, Christensen K, Skytthe A, Vaupel JW (2019) A Cohort Comparison of Lifespan After Age 100 in Denmark and Sweden: Are Only the Oldest Getting Older?, Demography.
- ✓ Robine J-M, Caselli G. (2005). An unprecedented increase in the number of centenarians. Genus LXI(1, Special issue).
- ✓ Robine J-M, Caselli G, Rasulo D, Cournil A (2006) Differentials in the femininity ratio among centenarians: variations between northern and southern Italy from 1870. Popul Stud 60(1).