Università di Pisa
Sant'Anna Scuola Universitaria Superiore Pisa
Consiglio Nazionale delle Ricerche
Università degli Studi di Cagliari

Book of Short Papers

SIS 2021

SIS 2021
PISA, 21-25 GIUGNO 2021

SIS Società Italiana di Statistica

Editors: Cira Perna, Nicola Salvati and Francesco Schirripa Spagnolo

PUBLISHED BY PEARSON

WWW.PEARSON.COM

ISBN 9788891927361

Analysing contributions of ages and causes of death to gender gap in life expectancy using functional data analysis

Analisi dei contributi per età e cause di morte alla differenza di genere nell'aspettativa di vita attraverso l'analisi dei dati funzionali

Alessandro Feraldi, Virginia Zarulli, Stefano Mazzuco and Cristina Giudici

Abstract The work consists of application of functional data analysis (FDA) to demographic data: it analyses the contribution of ages and causes of death to gender gap in life expectancy in 14 European and non-European countries between 1998 and 2016. Causes-of-death data and life tables were retrieved from the Human Causes-of-Death Database (HCD) and from the Human Mortality Database (HMD). Our analysis allows to identify two main components that capture most of the variability and which captures the extent of the cause-specific gender differences and the age pattern, respectively. Over time, an increase in the most relevant contributions is observed, especially around the modal age and a shift of the contributions towards older age.

Abstract Il lavoro consiste in un'applicazione dell'analisi dei dati funzionali (FDA) a dati demografici: si analizza il contributo delle età e cause di morte alle differenze di genere nella speranza di vita in 14 paesi europei ed extraeuropei nel periodo compreso tra il 1998 e il 2016. I dati sui decessi per causa provengono dallo Human Cause-of-Death Database (HCD), mentre le tavole di mortalità sono tratte dallo Human Mortality Database (HMD). L'analisi consente di individuare due componenti principali che colgono gran parte della variabilità e che descrivono rispettivamente l'entità delle differenze di genere specifiche per causa e i contributi età-specifici. Nel tempo, si osserva un aumento dei contributi più rilevanti soprattutto intorno all'età modale ed uno spostamento degli stessi verso l'età avanzata.

Key words: gender gap, life expectancy, causes of death, functional data analysis

---

Alessandro Feraldi, Department of Statistical Sciences, Sapienza University of Rome, Piazzale Aldo Moro, 5 (00185) Rome, e-mail: alessandro.feraldi@uniroma1.it

Virginia Zarulli, Interdisciplinary Centre on Population Dynamics, CPOP, University of Southern Denmark. Campusvej 55. Odense M DK-5230, e-mail: vzarulli@sdu.dk

Stefano Mazzuco, Department of Statistical Sciences, University of Padova, Via Cesare Battisti 241, 35121 Padova, Italy, email: stefano.mazzuco@unipd.it

Cristina Giudici, MEMOTEF Department, Sapienza University of Rome, Via del Castro Laurenziano, 9, 00161 Rome, Italy, e-mail: cristina.giudici@uniroma1.it

1 Introduction

On average and worldwide, women live longer than men and the absolute difference between male and female mortality risk reaches its maximum at old ages. From the beginning of the 1920s, in most industrialized countries the gap in life expectancy between the two sexes widened until the 1970s, when the difference started to narrow (Austad (2006); Zarulli et al. 2020)). The literature shows that where some convergence has taken place, men have experienced more rapid gains in survival than women (Meslé & Vallin (2011)).

To explain the recent narrowing in the sex gap in LE, several studies have focused on the causes of death that contributed to the gender gap in mortality rates and, thereby, either narrowed or widened this gap (Klenk et al. (2016)). In most of the studies, the topic of the variability in the gender gap in mortality has usually been tackled comparing age trajectories of cause-specific death rates between men and women by fitting specific parametric models on cause-specific life table death rate for women and men separately (Horiuchi et al. (2013)). Other studies used life table or aggregated mortality indicators to provide summary measures of mortality levels (e.g. life expectancy) and dispersion (e.g. lifespan variation) both separately for men and women and with the decomposition of the difference between sexes according to age and causes of death (Trias-Llimós & Janssen (2018)).

Although numerous studies have decomposed the sex gap in life expectancy according to age and causes of death, they did not study the principal components of the contributions of age and causes of death from a functional perspective, which has been shown to be more informative approach (Ramsay & Silverman (2002); Léger & Mazzuco (2020)). To fill this gap, we study absolute and relative contributions of age and causes of death to the gender gap in life expectancy (GGLE) for several countries, using the Functional Data Analysis (FDA) (Ramsay & Silverman (2002)). Following this approach, we consider age- and cause-specific contributions as functions, and therefore we analyse curves rather than scalar data. More specifically we propose a Functional Principal Component Analysis (FPCA) of the contribution profiles of several countries, in order to identify the main components of the distribution of age-specific contributions according to causes of death. To the best of our knowledge, this is the first study analysing age- and cause-specific contributions to the gender gap in life expectancy with a functional data analysis approach.

2 Data and Method

2.1 Data

Cause-specific mortality data were retrieved by gender, 5-year age interval and year from the Human Cause-of-Death Database (HCD (2019)) and life tables were taken from the Human Mortality Database (HMD (2019)). HMD and HCD are open-source projects: the former offers harmonized data on constructed series of mortality rates,

life tables, death counts and population exposures; the latter contains reconstructed long-term trends in cause-specific mortality for sixteen countries over time. Since this study aims at analysing patterns of causes of death across several countries, we focussed on the last 15 years available for each country, within the period 1998-2016. Countries were grouped according to geographical areas: Eastern Europe (EE) (i.e. Russia, Ukraine, Poland, the Czech Republic, Estonia, Latvia, Lithuania and Belarus); Western Europe (WE) (i.e. France, Spain, Germany and the United Kingdom (UK)), and extra-European countries (i.e. Japan and the United States (US)). Romania and Moldova were excluded from the analysis because no data were available in the HMD. Within this time frame all the causes of death are coded in each country according to the 10th version of International Classification of Disease (ICD-10). This allows us to avoid problems related to differences in the classification and to obtain comparable information for all the countries under study. We further restricted our focus on the short list of the ICD-10, in which all the causes are grouped into sixteen major categories, each including a set of similar diseases (e.g. heart diseases, neoplasms, external causes, respiratory diseases etc.) (HCD (2019)). Finally, cause-specific mortality data for all the ages above 85 were grouped in the open-end age interval 85+, to avoid problems related to the data quality, which are particularly common at very old ages.

2.2 Analysis

Age- and cause-specific contributions to the GGLE (female - male) were obtained for each country and over time applying Arriaga's age- and cause-specific decomposition technique combining life tables from the HMD and cause-specific mortality data from the HCD (Arriaga (1984)). FDA was applied to the age-specific relative contributions to the GGLE, separately for each cause of death. Discrete age-specific relative contribution data $x(t_1)$, ..., $x(t_N)$ were assumed to be independent realizations drawn from the same continuous stochastic process $X(t)$ (Ramsay & Silverman (2002)). To obtain the functional representation, each $X(t)$ was approximated by using a basis expansion of cubic B-splines functions (1) and the B-splines basis coefficients were estimated by the ordinary least squares method minimizing the sum of squared residuals (Léger & Mazzuco (2020)). Therefore,

$$X(t) = \sum_{j=1}^{p} \gamma_j \psi_j(t) \quad (1)$$

where ψ_j are p known basis functions and γ_j are the corresponding coefficients to be estimated. In order to maintain the data structure, we used a sequence of $p = 19$ equally distributed knots (i.e. one for 5-years age interval). Afterwards, we performed FPCA separately for each cause of death in order to synthetize the variability of the curves and to identify the main components of the distributions of age-specific contributions according to causes of death, across countries. FPCA is the extension of the more classical multivariate PCA to functional data: for a generic curve $x_i(t)$ we can obtain the approximation (2)

$$x_i(t) = \sum_{k=1}^{\infty} c_{i,k}\phi_k(t) \quad (2)$$

where $c_{i,k}$ are the principal component scores and $\phi_k(t)$ are the eigenfunctions or harmonics. Therefore, the information on the curve $x_i(t)$ were then synthesized by the first q terms. All the analyses were conducted using the R package fda (Ramsay et al. (2011)).

3 Results

Decomposition results confirmed that neoplasms, heart diseases and external causes of death made the largest contributions to the GGLE in all the countries, explaining together more than two third of the overall gap. Additionally, the largest contributions to the GGLE were given by old ages for most of the causes of death over the entire period (Meslé & Vallin (2011); Trias-Llimós & Janssen (2018)). Results of FPCA focus on the three most relevant causes of death.

Most of the variability in the age-specific contributions is explained by the first two principal components for each cause of death (e.g. 97%, 94% and 95% for neoplasm, heart diseases and external causes, respectively). The first FPC mainly captures the extent of the cause-specific gender differences, while the second FPC captures the age pattern. A classical way to interpret the FPCs is to plot the group mean function (solid curve in Figure 1) as well as the functions obtained by adding (+ curve) and subtracting (- curve) to the mean function twice the square root of the principal component variance (Ramsay & Silverman (2002); Léger & Mazzuco (2020)). Regarding neoplasm (Figure 1), for the first FPC, the variability is concentrated at 40 years' age and older and especially around the modal age. A high score on this component suggests an above-average contribution. The second FPC corresponds to a shift of the curves with respect to the overall contribution mean towards older ages. The (+) curve has a higher contribution than the (-) curve with respect to the mean curve before 70 years, lower afterwards. A low score on this component suggests an above-average shift of the distribution towards older ages.

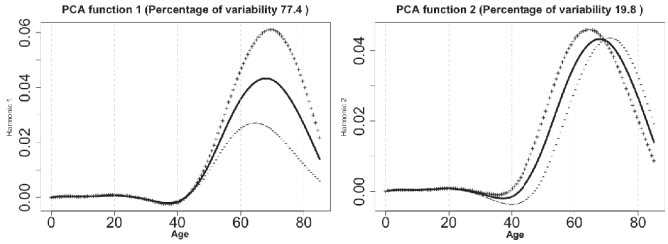

Figure 1. Effect of the first two FPCs on age-specific contributions to GGLE for neoplasm: overall mean and mean ± a suitable multiple of the principal component weight function.

In order to study the evolutions of age-cause contributions over time, the scores of the two first components for each country are plotted at every 3 years. With regards to neoplasm (Figure 2), the first axis indicates that, throughout the whole period, relative contributions to the GGLE were higher in France, Spain (I quarter) and in Japan (IV quarter) than in the other countries, followed by UK, Germany, Poland, Czech Republic and US. The lowest contributions were shown in Ukraine, Estonia, Latvia, Lithuania, Belarus and in Russia (II quarter). The second axis denotes that the distribution of age-specific contributions to the GGLE was more concentrated at older ages in UK than in the other countries. Furthermore, distributions were more concentrated at older ages in US and in Japan (scores of the second FPC < 0) than in the overall mean distribution (i.e. for all the countries).

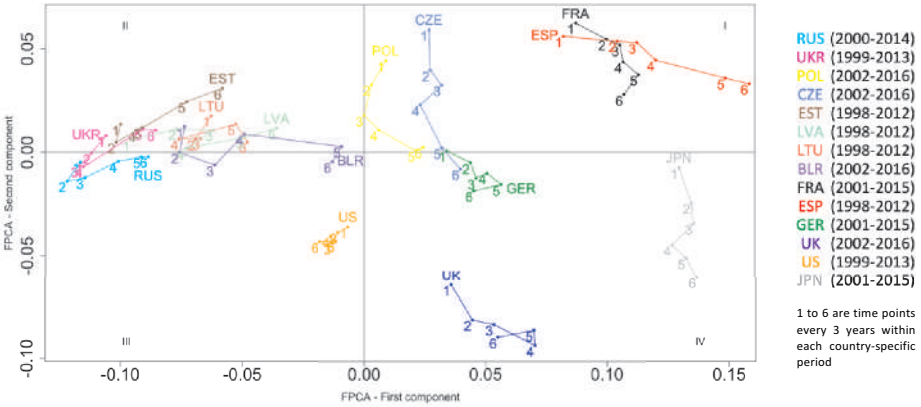

Figure 2. Principal component subspace - neoplasm.

On the contrary, in the remaining countries distributions were more concentrated at younger ages than in the overall mean distribution. Moreover, the increasing trends in the first FPC scores, especially in Spain, together with the decreasing trends in second FPC scores showed also in France and in UK, indicate increasing relative contribution of neoplasm over time as well as a shift of the distribution towards older ages. Similarly, in Poland, Czech Republic and in Japan, although the relative contributions of neoplasm to the GGLE stagnated over time (small variations in first FPC scores), the decreasing trends of the second FPC scores denote that the distributions of age-specific contributions shifted towards older ages over time. A small increase in the score of the first component in Estonia and in Belarus (especially at the second half of the period) suggests a slight increase in the relative contribution of neoplasm to the GGLE. Finally, stagnation in both FPC scores was reported in Germany, Russia, Ukraine and in US over time. Therefore, independent from trends in relative contributions of neoplasm to the GGLE, in most of the countries the distributions shifted towards older ages over time.

For the sake of brevity, we do not display the results of FPCA for the other two main causes, however the analysis shows similar patterns across countries and over time.

Conclusion

This work gives a deeper insight of the main contributing factors to the gender differences in life expectancy, analysing components of the relative age-specific contributions according to the most relevant causes of death in determining the gap. The study also aims at illustrating the demographic application of FDA, a new method for demographic analysis but which could prove useful to deepen our understanding of complex demographic phenomena. Our results allow to identify two main components which capture most of the variability. The first component mainly captures the extent of the cause-specific gender differences, while the second FPC captures the age patterns. Over time, an increase in the most relevant contributions is observed, especially around the modal age and a shift of the contributions towards older age. The analysis confirms that FDA allows to highlight country-specific patterns in the context of the epidemiological transition which need to be further analysed. Further analyses include functional cluster analysis to group countries according to age-cause contributions and study the evolution of the contributions in each cluster over time. Finally, following the FDA approach we also apply other statistical analyses (i.e. regression and hypothesis tests) to the same data and suggest to increase the use of such approach in population studies.

References

1. Arriaga, E. E. (1984). Measuring and explaining the change in life expectancies. Demography, 21(1), 83–96.
2. Austad, S. N. (2006). Why women live longer than men: Sex differences in longevity. Gender Medicine. Official Journal of the Partnership for Gender-Specific Medicine at Columbia University, 3(2), 79–92.
3. Horiuchi, S., Ouellette, N., Cheung, S. L. K., & Robine, J.-M. (2013). Modal age at death: Lifespan indicator in the era of longevity extension. Vienna Yearbook of Population Research, 37–69.
4. Human Cause-of-Death Database. (2019). French Institute for Demographic Studies (France) and Max Planck Institute for Demographic Research (Germany) (downloaded on 2019).
5. Human Mortality Database. (2019). Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany) (downloaded on 2019).
6. Klenk, J., Keil, U., Jaensch, A., Christiansen, M. C., & Nagel, G. (2016). Changes in life expectancy 1950-2010: Contributions from age- and disease-specific mortality in selected countries. Population Health Metrics, 14, 20.
7. Léger, A. E., & Mazzuco, S. (2020). What can we learn from functional clustering of mortality data? An application to HMD data. *arXiv preprint arXiv:2003.05780*.
8. Meslé, F., & Vallin, J. (2011). Historical trends in mortality. In International handbook of adult mortality (pp. 9–47). Springer.
9. Ramsay, J. O. and Silverman, B. W. (2002). Applied functional data analysis: methods and case studies. Springer Series in Statistics, Springer-Verlag.
10. Ramsay, J., Wickham, H., Graves, S., and Hooker, G. (2011). fda: Functional Data Analysis. R package.
11. Trias-Llimós, S., & Janssen, F. (2018). Alcohol and gender gaps in life expectancy in eight Central and Eastern European countries. European Journal of Public Health, 28(4), 687–692.
12. Zarulli, V., Lindahl-Jacobsen, R., & Vaupel, J. W. (2020). Onset of the old-age gender gap in survival. Demographic Research, 42, 727–740.