



Functional data analysis approach in population studies: an application to the gender gap in life expectancy

Alessandro Feraldi^{1,5} · Virginia Zarulli² · Stefano Mazzucco³ · Cristina Giudici⁴

Accepted: 18 July 2023 / Published online: 23 August 2023
© The Author(s) 2023

Abstract

This work analyses the contribution of ages and causes of death to gender gap in life expectancy in 20 European and non-European countries between 1959 and 2015, using Functional Data Analysis. Data were retrieved from the WHO Mortality Database and from the Human Mortality Database. We propose a Functional Principal Component Analysis of the age profiles of cause-specific contributions, to identify the main components of the distribution of the age-specific contributions according to causes of death, and to summarize them with few components. Our findings show that the narrowing gender gap in life expectancy was mainly driven by decreasing differences in cardiovascular diseases. Additionally, the study reveals that the age cause contributions act almost entirely on only two dimensions: level (extent of the cause-specific contribution to the overall mortality gender gap) and age pattern (location of the curves across ages). Notably, in the last period, it is not the "quantum" of the cause-specific contributions that matters, but the "timing", i.e. location across the age spectrum. Moreover, our results show that in the most recent period the gender gap in life expectancy is affected by composition of the causes of death more than it was in previous periods. We emphasise that Functional Data Analysis could prove useful to deepen our understanding of complex demographic phenomena.

Keywords Gender gap · Life expectancy · Causes of death · Functional data analysis · Age- and cause-specific decomposition

✉ Alessandro Feraldi
feraldi@demogr.mpg.de

¹ Department of Statistical Sciences, Sapienza University of Rome, Piazzale Aldo Moro, 5, 00185 Rome, Italy

² Interdisciplinary Centre on Population Dynamics, CPOP University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark

³ Department of Statistical Sciences, University of Padua, Via Cesare Battisti, 241/243, 35121 Padua, Italy

⁴ MEMOTEF Department, Sapienza University of Rome, Via del Castro Laurenziano, 9, 00161 Rome, Italy

⁵ Present Address: Max Planck Institute for Demographic Research, Konrad-Zuse-Straße, 1, 18057 Rostock, Germany

1 Introduction

Throughout the twentieth century, mortality in most industrialized countries became concentrated at older ages and noncommunicable diseases became the prevailing cause of death. A female survival advantage emerged (Goldin and Lleras-Muney 2019; Hollingshaus et al. 2019) and the male–female gap widened from the beginning of the 1920s until the 1970s and then it narrowed (Austad 2006; Zarulli et al. 2020, 2021). The reduction in the gender gap was primarily due to narrowing male–female differences in cardiovascular disease-related mortality, smoking-related mortality, and mortality due to accidents, violence, and suicide, especially in European countries (Meslé 2004; Trovato & Heyen 2006; Nusselder et al. 2010; Horiuchi et al. 2013; Klenk et al. 2016). Existing literature has also shown that, with convergence taking place, men have experienced more rapid gains in survival than women (Oksuzyan et al. 2010; Meslé and Vallin 2011; Thorslund et al. 2013); furthermore, although decreasing gender differences in life expectancy persist over time, in the last decades, the contribution of age older than 80 years to the gender difference in life expectancy is increasing (Beltrán-Sánchez et al. 2015; Zarulli et al. 2020, 2021). However, whether the shift toward older age translates into a more compressed or more dispersed age- and cause-specific gender gap in life expectancy is still not clear. On the one hand, some studies argue that old-age deaths are decreasing at advanced ages; on the other hand, others argue that old-age deaths are becoming more dispersed with age (Bongaarts 2005; Canudas-Romo 2008; Zuo et al. 2018; Janssen and de Beer 2019). The overall scenario is even more complicated because different causes of death may have different age-specific distributions in their contribution to the gender gap in survival.

This study aimed to investigate the presence of common underlying patterns, across multiple countries, over time. However, this type of analysis faces the aforementioned problems of complexity and multidimensionality. Therefore, we adopted a Functional Data Analysis (FDA) approach (Ramsay and Silverman 2005), a method that only recently appeared in demographic analyses and is very promising with respect to the analysis of highly multidimensional problems. The basic idea behind the FDA is that discrete observations arising from data can be expressed in the form of a function (functional data) that represents the entire measured function as a single observation, and then modeling and prediction information can be drawn from a collection of functional data by applying statistical concepts from multivariate data analysis (Ramsay and Silverman 2005; Ullah and Finch 2013). Such data are typically interpreted to reflect the influence of certain smooth functions that are assumed to underlie and generate observations. In contrast to other statistical techniques, FDA takes advantage of additional information that could be implied by the smoothness of underlying functions (Ramsay and Silverman 2005; Ullah and Finch 2013). FDA methods for modelling and forecasting data across a range of health and demographic issues have significant advantages for better understanding trends, risk factor relationships, and the effectiveness of preventive measures (Erbas et al. 2007; Hyndman and Ullah 2007; Ullah and Finch 2013). This approach has become popular in demographic applications, especially in forecasting mortality, fertility, and migration (Erbas et al. 2007; Hyndman and Ullah 2007; Hyndman and Booth 2008; Hyndman and Shang 2009; Ullah and Finch 2010). Recent studies have used FDA to explore changes in age-specific mortality over time in low-mortality countries (Léger and Mazzuco 2021) and developing countries (Shair et al. 2019) as well as to analyze cause-specific mortality trends across countries (Stefanucci and Mazzuco 2022), successfully enhancing the accuracy of the analyses and providing innovative insights.

In this study, FDA was applied to the relative age-cause contributions to the gender gap in life expectancy to study the changes in age- and cause-specific mortality of several countries over time. More specifically, we applied a Functional Principal Component Analysis (FPCA) to the age profiles of cause-specific contributions to identify the main components of the distribution of age-specific contributions according to causes of death and thereby summarize them with a few components that will be easier to interpret.

2 Causes-of-death contributions to the gender gap in life expectancy: review of the literature

Examining gender differences in cause-specific mortality is important because it provides additional insights into the mechanisms that create a gender gap in overall mortality (Rogers et al. 2010). With respect to women, men experience a higher risk of death from almost all causes (Rogers et al. 2010). Men are more likely than women to engage in risky and aggressive behaviors, such as tobacco and alcohol consumption, as well as risky driving and eating habits; these differences generally decrease with age (Gjonça et al. 2005; Rogers et al. 2010). For instance, compared with women, men's greater exposure to smoking for longer periods of time contributes to various smoking-related diseases, such as lung cancer, emphysema, and cardiovascular diseases, which increase the risk of overall mortality.

The three most acknowledged causes of death that determine the overall gender gap in mortality are cancer, cardiovascular diseases, and external causes of death.

Lung cancer is one of the most common cancers, and smoking has played a crucial role in its prevalence (Ferlay et al. 2018). Smoking patterns vary across Europe, with men generally starting smoking earlier and having higher prevalence rates than women (Janssen 2020). This trend has shown variations in countries such as the United States and Sweden (Beltrán-Sánchez et al. 2015). The effect of smoking on gender differences in mortality has been significant, with smoking contributing to a large proportion of the gender gap in European mortality. However, the contribution of smoking to gender gap has declined over time (Janssen 2020; Janssen and van Poppel 2015; Luy and Wegner-Siegmundt 2015; Meslé 2004). The prevalence of lung cancer among men has either stabilized or declined, whereas it has continued to increase among women in recent years (Stefanucci and Mazzuco 2022). In addition to lung cancer, breast, colorectal, and prostate cancers have contributed to the overall cancer burden in Europe (Ferlay et al. 2018). Trends in cancer incidence and mortality vary by gender, age group, period, and country (Zatoński and Didkowska 2008; Gjonça et al. 2005).

In addition to cancer-related mortality, cardiovascular diseases are a major contributor to the gender gap in survival, with excessive male mortality rates observed in several countries (Klenk et al. 2016; Beltrán-Sánchez et al. 2015). However, recent studies have reported a decline in cardiovascular diseases-related mortality in both men and women across Europe (Di Girolamo et al. 2020; Sanders 2018; Conti et al. 2003). Alcohol consumption has been closely linked to cardiovascular disease mortality, particularly in Eastern European countries, and has contributed to gender differences in life expectancy (Pinkhasov et al. 2010; Trias-Llimós et al. 2018; Tarkiainen et al. 2012; Meslé et al. 2002).

External causes of death, such as accidents and unhealthy lifestyles, also play a substantial role in the mortality differences between women and men. Studies examining the external causes of death in Eastern European countries have identified unhealthy lifestyle and accidents account for more than one third of the overall gender gap in life expectancy

(Luy and Wegner-Siegmundt 2015; Trias-Llimòs and Janssen 2018). Finally, women have a higher lifetime prevalence of mental health disorders, such as depression, whereas men are more likely to suffer from substance use and impulsive disorders (Ho and Hendi 2018; Ramsay et al. 2020; Kuehner 2017; Boyd et al. 2015).

Several studies have focused on the contributions of the causes of death to the gender gap in mortality rates to explain the recent narrowing of the gender gap in life expectancy (Glei et al. 2010; Yang et al. 2010; Nusselder et al. 2010; Désesquelles et al. 2012; Zatonski and Bhala 2012; Rasul 2013; Klenk et al. 2016; Chisumpa and Odimegwu 2018). Numerous studies have decomposed the gender gap in life expectancy according to age and cause of death (Meslé 2004; Canudas-Romo 2008; Luy and Wegner-Siegmundt 2015; Trias-Llimòs and Janssen 2018; Hollingshaus et al. 2019; Aburto et al. 2020; Sagna et al. 2020). A few studies have analyzed the evolution of the distributions of the contributions of both ages and causes, and how these would determine the overall gender gap in life expectancy over time (Gómez-Redondo and Boe 2005; Tarkiainen et al. 2012; Sundberg et al. 2018; Sagna et al. 2020; Feraldi and Zarulli 2022). Nevertheless, the gender gap in life expectancy has undergone complex and poorly understood evolution.

Arriaga's age- and cause-specific decomposition (Arriaga 1984) is one of the most common decomposition techniques and represents a useful tool for estimating mortality differences for specific causes of death and age groups, contributing to the total difference in life expectancy (Preston et al. 2001, pages 84–86). However, traditional demographic methods of decomposition, such as Arriaga's method, when simultaneously applied to a broad array of countries over many years, provide an innumerable set of data points that are difficult to summarize in a compact but complete picture. Hence, the custom focuses only on a subset of the data or produces repeated similar analyses and plots by time and/or by country for the age-cause combinations. This makes interpretation and investigation of the underlying general patterns cumbersome.

Even the Truncated Cross-sectional Average Length of life (TCAL) indicator (Canudas-Romo and Guillot 2015; Nepomuceno and Canudas-Romo 2019; Canudas-Romo et al. 2020; Sauerberg and Canudas-Romo 2022), a common measure that summarizes the evolution of cohort- and period-specific information on cause-specific mortality, does not explain time trends (i.e., reducing dimensionality by collapsing one component).

3 Aim and contribution

If we consider the time trend of the relative age-cause contributions, we have not only points but also functions. Therefore, in this study, we analyzed the changes in the contributions of age and causes of death (which were estimated using Arriaga's age- and cause-specific composition tool) from a functional perspective, which has been shown to be a more informative approach and has significant advantages (Ramsay and Silverman 2002; Léger and Mazzucco 2021).

First, FDA has the advantage of generating models that can be described by continuous smooth dynamics, representing a useful tool for expressing curves in a desirable reduced-dimensional space, which allows for accurate estimates of parameters for use in the analysis phase (Ullah and Finch 2013; Ramsay 1988). The FDA has been used to model age-cause contributions to the gender gap because it has the advantage of being less affected by outliers than traditional statistical methods. The FDA uses all the information in the data to generate a smooth linear function, which helps reduce the impact of period shocks and

other random fluctuations in the data. This makes it easier to identify trends and patterns in the data that may not be apparent using other methods.

Second, FDA is a useful tool not only for data smoothing, but also because, by transforming the observed discrete data into functional data, it is possible to apply FPCA to synthesize the temporal dynamics by cause and age of the GGLE with few components, which is not as simple as traditional methods, such as Arriaga's decomposition.

Therefore, FPCA allows the synthesis of large datasets, thereby increasing the interpretability of the data. We propose to apply FDA to the relative age-cause contributions to the Gender Gap in Life Expectancy (GGLE) to describe the global trend of age- and cause-specific mortality in several countries over time. We defined the proportional variable as the age profiles of the cause-specific contributions to the GGLE for each country and year. FDA was applied to the age profiles of each cause of death. Using relative rather than absolute contributions makes it possible to appreciate the changes (increase/decrease) in the age-cause contributions regardless of the overall trend in the GGLE. Considering the GGLE varies over time (increase/decrease), the absolute age-cause contributions (i.e., years) mostly follow the trend of the overall gender gap. However, the contributions of certain causes and age groups may show the opposite trend in relative terms with respect to the trend in the overall GGLE. Additionally, relative contributions allow to appreciate differences between countries and over time in a standardized way, thus abstracting from the potential differences in extent related to absolute measurements.

More specifically, we proposed an FPCA of the age profiles of the relative contribution of 20 countries to identify the main components of the distribution of age-specific contributions according to the causes of death.

Notably, the cause-of-death data suffer from several irregularities. Small numbers that can be observed for less frequent causes of death raise statistical issues concerning accuracy and thereby question the usefulness of these data (Ullah and Finch 2013). Data on the causes of death also suffer from variations in coding protocols. The underlying causes of death were identified from the death certificates and coded according to the International Classification of Diseases (ICD). The coding of an underlying cause of death depends on (i) the information available on the death certificate, (ii) the interpretation of explicit coding or selection rules, and (iii) the individual deliberations of the coders (Treurniet et al. 2004; Harteloh et al. 2010). Additionally, the ICD varies with time and place; therefore, coding could hinder the interpretation of trends in time or differences between populations. Through curve smoothing, FDA facilitates effective and appropriate data noise reduction (Ramsay 1988; Ullah and Finch 2013; Léger and Mazzuco 2021). Although methods to harmonize different ICDs already exist (see Meslé and Vallin 1996), some discontinuities (see Mazzuco 2022, Figure 5) and missing data (see Stefanucci and Mazzuco 2022) persist in the causes-of-death data, even within the same ICD. Dealing with the causes of death likely leads to the handling of sparse data, and consequently, irregularities in estimates. In such cases, the smoothing properties of FDA become useful. Finally, by applying FPCA, the age pattern of cause-specific contributions to GGLE can be summarized with a few components that are easier to interpret. This also allows for an analysis of how these components evolve over time for the considered countries, similar to what Léger and Mazzuco (2021) performed with overall age-specific mortality. They showed that between 1960 and 2018, age-specific mortality rates in 32 developed countries followed the same evolution pattern (i.e., reduction in infant mortality, increase in premature mortality, and shift and compression of deaths), but with different timings. FPCA is useful for dimension reduction of the curves. In particular, the eigenfunctions allowed the identification of the main directions of variability of the age-cause contribution profiles with respect to the mean curve.

The corresponding scores of the functional components were used to characterize the age-cause contribution profiles in a reduced dimensional space.

Since the most recent period is characterized by low mortality for both men and women, as well as greater gender differences in old age, we can expect that the overall GGLE would be mainly determined by a different distribution of mortality across ages rather than by overall differences between women and men in specific causes of death. Additionally, in light of the ongoing reduction in the female–male mortality gap in the most recent period, the distinctive GGLEs observed across countries could be determined by different compositions of causes of death.

The results of this study contribute to research related to the measurement of population changes, providing important insights into the main contributing factors of gender differences in mortality, specifically highlighting the changes in the role of age and causes of death in determining the overall GGLE across European and extra-European countries over a long period. This study illustrates the advantages derived from the demographic application of FDA. To the best of our knowledge, this is the first attempt to analyze age- and cause-specific contributions to the gender gap in life expectancy using an FDA approach.

4 Data and methods

4.1 Data

Cause-specific mortality data were retrieved according to gender, 5-years age interval, and year from the WHO Mortality Database (WHO 2022), and life tables were obtained from the Human Mortality Database (HMD 2022). The HMD and WHO Mortality Databases are open-source projects: the former offers harmonized data on a constructed series of mortality rates, life tables, death counts, and population exposures, whereas the latter is an archive of the causes of death information for several countries. Although the longest time series for many countries started in 1950 and ended in 2019, information is available only from 1959 to 2015. To consider the same time window for each country, we focused on the years 1959–2015 and only considered countries for which data were available in this time frame. Our final analysis focused on 20 countries: Denmark, Finland, Norway, Iceland, Sweden, Austria, Belgium, Switzerland, France, Ireland, the Netherlands, the United Kingdom (UK), Italy, Spain, Hungary, Poland, Australia, Canada, Japan, and the United States (US). These attempts differ in several respects. However, the aim of this study was to find a common underlying pattern in the age-related contribution to GGLE, even though these countries may have experienced changes in the composition of GGLE at different times.

To deal with the different classification of causes of death overtime (from ICD 7th to ICD 10th revision), we followed Canudas-Romo et al. (2020) and use broad classes of causes which are consistent across the revisions (Stefanucci and Mazzuco 2022): (1) lung cancer, (2) other neoplasms, (3) diseases of circulatory system, (4) external causes of death, (5) respiratory diseases, (6) diseases of digestive system, (7) endocrine, nutritional and metabolic diseases, and (8) infectious and parasitic diseases (see Table 2 in the Appendix for the details). Given that old age are the most important age in determining the difference in life expectancy (Zarulli et al. 2020; Zarulli et al. 2021 PNAS; Feraldi and Zarulli 2022) and to avoid problems related to the cause of death data quality (Ullah and Finch 2013), which are particularly common at very old ages, the analysis was restricted to age groups from 50 to 85 years.

4.2 Functional principal component analysis (FPCA)

Absolute 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 WHO mortality database (Arriaga 1984). First, the difference in life expectancy was decomposed according to the contribution of each age group. In the second step, the contributions from each age group were divided into contributions from each specific cause of death. The total contribution from any given cause was obtained by summing the contributions of each cause of death across all age groups. Accordingly, the total gender gap in life expectancy was obtained by summing the contributions across all age groups and causes of death.

The FDA was applied to the age-specific relative contributions to the GGLE separately for each cause of death. Relative contributions were computed by dividing the absolute age- and cause-specific contributions by the overall GGLE (years). Accordingly, the sum of the relative contributions across all age groups and causes of death is equal to one, the sum of the absolute contributions across all age groups and causes of death is equal to the overall GGLE (in years). Discrete age-specific relative contribution data $x(t_1), \dots, x(t_N)$ were assumed to be independent realizations drawn from the same continuous stochastic process $X(t)$ (Ramsay and Silverman 2002). To obtain the functional representation, each $X(t)$ was approximated using a basis expansion of cubic B-spline functions (1), and the B-spline basis coefficients were estimated using the ordinary least squares method, minimizing the sum of squared residuals (Ramsay and Silverman 2002). B-spline functions are commonly used for nonperiodic functional data because they divide the interval over which the function is approximated into subintervals separated by knots. Each subinterval has a polynomial of a specified order and adjacent polynomials join smoothly at knots. The B-splines ensure the continuity of the curve at the knots. The underlying functions are usually smoothed to capture the structural components of the data and reduce noise. The number of knots was used in the regression splines as a regulation parameter with fewer knots resulting in smoother curves.

A relatively large number of knots and a roughness penalty acting on the associated give the B-spline functions great flexibility in approximating complex curves with a relatively small number of parameters (see Léger and Mazzucco (2021) and Ramsay et al. (2009) for more details on B-spline functions). B-spline functions capture the underlying structure of the data while minimizing the effect of noise more accurately than conventional splines using a monomial basis; consequently, they are generally preferred over other bases (Ramsay et al. 2009). Therefore,

$$X(t) = \sum_{j=1}^p c_k \phi_k(t) \quad (1)$$

where ϕ_k are p known basis functions and c_k are the corresponding coefficients to be estimated. In order to maintain the data structure, we used a sequence of $p=8$ equally distributed knots (i.e. one for each 5-years age interval between 50 and 85). Afterward, we performed FPCA separately for each cause of death to synthesize 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)

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

where $c_{i,l}$ are the principal component scores and $\phi_l(t)$ are the eigenfunctions or harmonics. Therefore, the information on the curve $x_i(t)$ was then synthesized by the first q terms.

In addition, the selected FPCs were included as covariates in a multiple regression analysis of GGLE to determine how different causal compositions of GGLE impact the overall gender differences in life expectancy (LE). All analyses were conducted using R package *fda* (Ramsay et al. 2011).

5 Results

5.1 Changes in the overall GGLE over time

Figure 1 shows the absolute GGLE (years) over time for 20 countries and highlights the different patterns in the GGLE across countries. As expected, in all the countries the GGLE trend occurred in an inverted U-curve fashion, suggesting that GGLE undergoes three phases of growth, peak and stability, and decline (Liu et al. 2013).

However, the timing of these phases varies by country. The largest changes in the GGLE were observed around 1975 and 1995 (three time-intervals: 1959–1975, 1976–1995, and 1996–2015). Based on when GGLE started to decrease, we identified three patterns across countries: (a) “Early decreasing GGLE,” (b) “Late decreasing GGLE,” and (c) “Very late decreasing GGLE”.¹

While the GGLE in the first group of countries (Fig. 1 panel a) started to decrease shortly after 1975, the GGLE in the second group of countries (Fig. 1 panel b) was quite constant between 1975 and 1995 and started to decrease thereafter. Owing to the small population size, the GGLE in Iceland showed large variability. In this case, the smoothing properties of the FDA proved to be useful. In Spain, Japan, Hungary, and Poland (Fig. 1 panel c), the GGLE increased until 1995 and started to decrease very later than other countries. In a separate analysis, the results of which are not reported for the sake of brevity, the different trends in the absolute GGLE over time across countries were confirmed by computing correlation coefficients between GGLE and time, within each time interval separately.

The results of this study highlight the narrowing gender gap in life expectancy in recent decades; gains in life expectancy have been larger for males than for females. The reduction in the gender gap occurred at different times and at different pace across countries.

5.2 Contributions of causes of death to the GGLE

Lung cancer, other neoplasms, diseases of the circulatory system, and external causes of death (minor) made the largest contributions to the GGLE in all countries, accounting for more than 75% of the overall gap. Figure 2 shows the average relative cause-specific

¹ The first group of countries included Denmark, Finland, Sweden, the UK, Canada, Australia, and the US. The second group included Austria, Belgium, Switzerland, France, Ireland, the Netherlands, Italy, and Norway. The third group includes Spain, Japan, Hungary, and Poland.

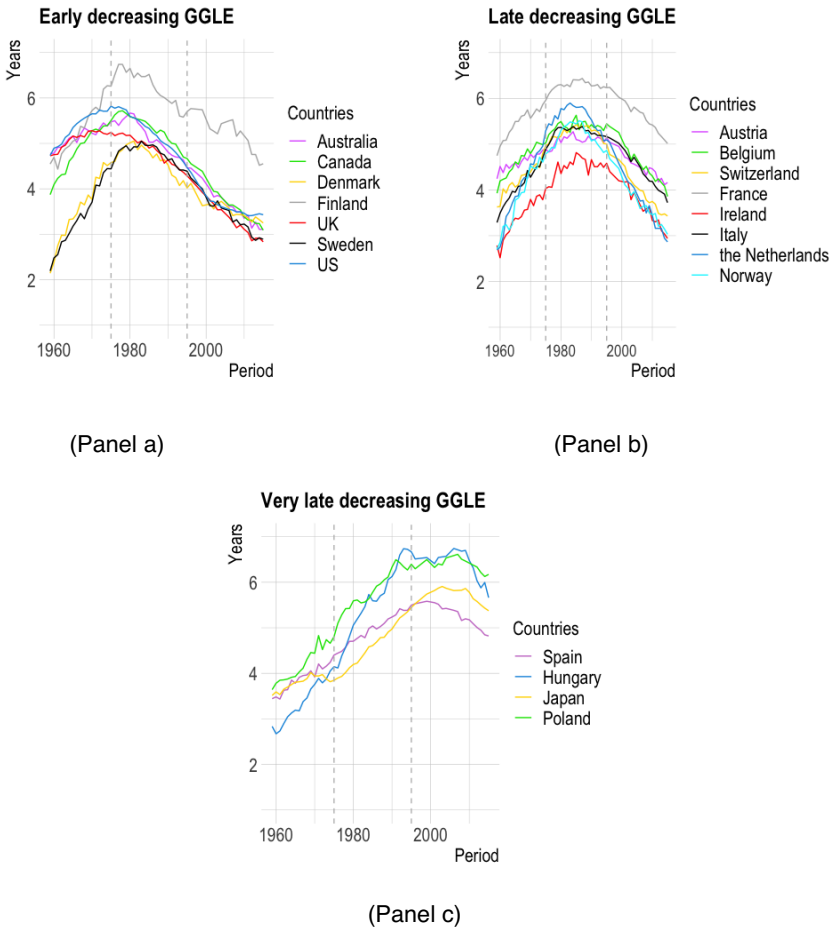


Fig. 1 Absolute (years) GGLE at age 50 over time, by country (dashed lines represent the years 1975 and 1995). (Color figure online)

contribution to the GGLE over time for the years in which data were available for all countries. Between 1959 and 2015, diseases of the circulatory system made the largest contribution (approximately 43.0% on average), followed by lung cancer and other neoplasms (12.2–12.9%), and external causes of death and respiratory diseases (6.9–9.8%). Each remaining group of causes contributed less than 5.0% of the overall GGLE over the entire study period (see Table 3 in the Appendix for the list of causes included in each group).

The contribution of diseases of the circulatory system slightly increased within the first interval (1959–1975) but strongly decreased afterward, from 48.9% on average between 1959 and 1995 to 36.7% in 2015. The average contribution of lung cancer to the GGLE increased until 1975 and decreased within the last interval (1996–2015), whereas it remained constant at approximately 14.3% from 1976 to 1995 (intermediate time interval). Starting around 1975, the average contribution of other neoplasms to the GGLE significantly increased from 6.7% (on average in 1959–1975) to 21.8% in 2015. The contribution

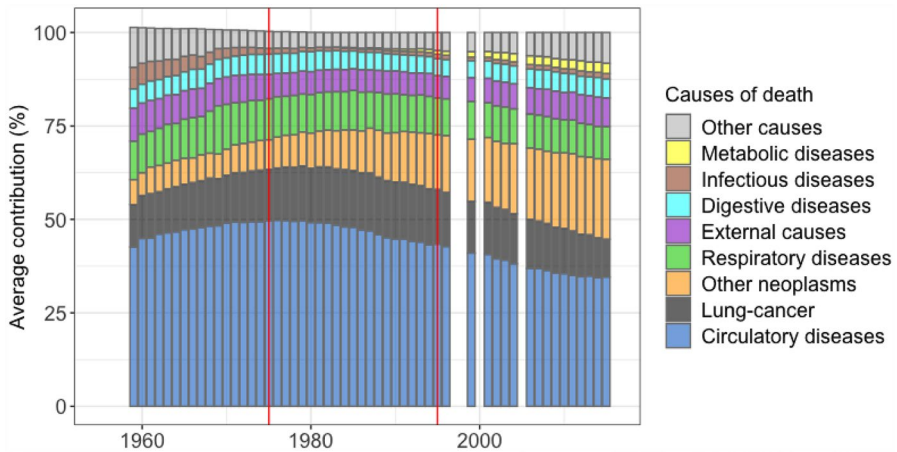


Fig. 2 Relative cause-specific contribution to the GGLE over time: average of 20 countries (red lines represent the years 1975 and 1995). *Note* Bars in 1997, 1998, 2000, and 2005 are missing because data were not available for all countries

of external causes of death slightly decreased within the first two intervals (from approximately 9.0% in 1959 to less than 6.0% in 1985), but slightly increased afterward (approximately 8.0% in 2015). The contributions of respiratory and digestive diseases remained constant over time, whereas the differences owing to infectious diseases decreased. Finally, the contribution of endocrine, nutritional, and metabolic diseases to the GGLE appeared very late, with a small contribution of 2.9% in 2015. These trends were confirmed by computing the correlation coefficients between the relative cause-specific contributions across countries and time for each time interval. We do not present these results here for brevity.

This analysis suggests that the narrowing of the gender gap in LE is mainly driven by decreasing differences in circulatory system diseases. Gender differences in mortality due to external causes and infectious diseases also contributed to the narrowing of the gap, although their contribution to GGLE was lower. Conversely, gender differences in mortality due to other neoplasms contributed to the widening of the gap throughout the analyzed period. Finally, lung cancer widened the gap until 1975, whereas its contribution to GGLE decreased within the last period (1996–2015).

5.3 Functional principal component analysis

Almost the entire variability in age-specific contributions to the GGLE was explained by the first two functional principal components for all causes of death (more than 90%). The first component describes the level (greater/smaller contribution); that is, the extent of the cause-specific contribution to the overall gender gap in mortality. The second component captures the age pattern (location of the curves across ages) and describes the shift in age-specific contributions to older age. A classical way to interpret FPCs is to plot the group mean function (solid curve in Fig. 3) 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 and Silverman 2002).

Regarding lung cancer (Fig. 3), the first FPC captured the extent of the contribution of lung cancer mortality to the overall gender differences. With respect to the overall

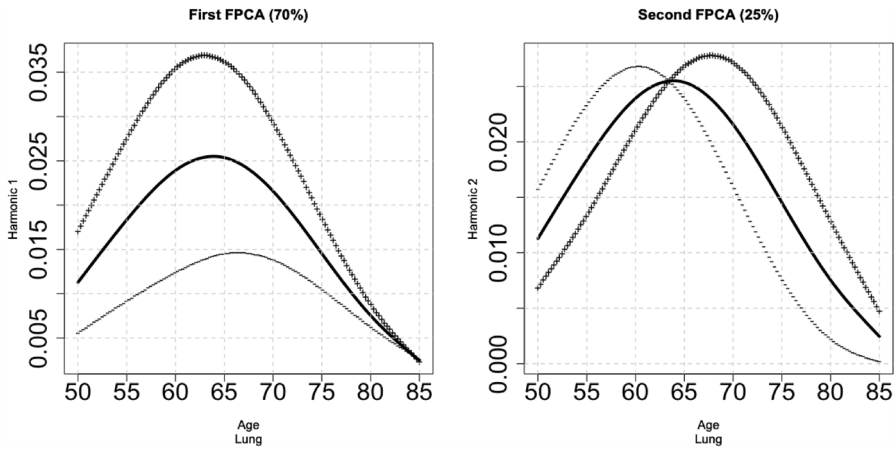


Fig. 3 Effect of the first two FPCs on age-specific contributions to the GGLE for lung cancer: overall mean and mean \pm a suitable multiple of the principal component weight function

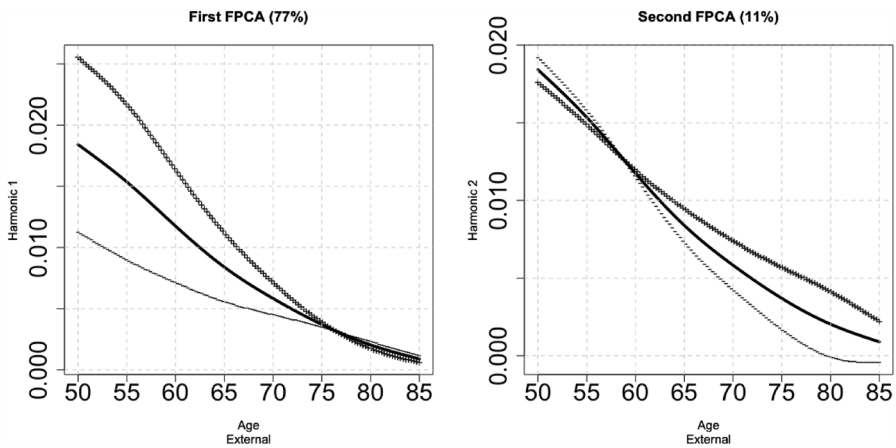


Fig. 4 Effect of the first two FPCs on age-specific contributions to the GGLE for external causes of death: overall mean and mean \pm a suitable multiple of the principal component weight function

mean contribution (solid line), the (+) curve shows a higher cause-specific contribution, whereas the (-) curve shows a lower cause-specific contribution. The largest contribution (i.e., modal age at contribution) was observed around the age of 65 years, and variations in FPC1 involved all ages. For a generic country, a high score for this component suggests a high contribution of lung cancer to the GGLE. The second FPC describes the locations of the curves across ages. With respect to the overall contribution mean (solid line), the (+) curve denotes higher contributions of older age and lower contributions from younger age. Conversely, the (-) curve denotes lower contributions of older age and higher contributions of younger age. Indeed, the modal age at contribution was young in the (-) curve and older in the (+) curve. Similarly, Fig. 4 shows that larger (smaller) values of the first FPC denote a higher (lower) contribution of external causes of death to the GGLE. Compared with lung cancer (Fig. 3), the contributions of external causes of death were mainly

concentrated at younger ages. Variations in FPC2 mainly involve ages older than 60 years; therefore, the larger the score of the second component, the larger is the contribution of external causes of death at older ages. Contributions at younger ages (below 60 years) were similar regardless of the score of the second component. As the contributions of external causes of death are mainly located at younger ages, a high score of the second component indicates that the contribution of external causes of death spread over a wider age range (Fig. 4), that is, an expansion of the differences.

With regard to lung cancer, for which the largest age contributions are located at older ages, the shifts toward older ages might be accompanied by a concentration of contributions over a narrower age range (Fig. 3), that is, compression of the differences. For the other causes of death, FPCA revealed similar results; the first and second components captured the entire variability of the contributions and described two distinctive dimensions (level and age pattern). The FPCA results related to all other causes of death are reported in the Appendix.

5.4 Multiple regression analyses

To analyze the effect of FPCs on GGLE simultaneously, we estimated multiple linear regression models between GGLE (dependent variables) and both the first and second FPCs for each cause of death (independent variables). For each of the three time intervals (1959–1975, 1976–1995, and 1996–2015), regression models were estimated separately and time (year) was included as a covariate to control for overall GGLE trends. Accordingly, these models allowed us to interpret the effect of FPC1 (extent of cause-specific contribution) and FPC2 (shift of age-specific contributions toward older ages) regardless of the overall trend over time in the GGLE in each period. Table 1 presents the estimated parameters and *p*-values (significance level=0.01) of the three models. All models captured a large portion of the variability in the GGLE, especially in the last study period (*adj-R*²=65.0%, 69.0%, and 86.0% for the first, second, and last periods, respectively).

In the first period (between 1959 and 1975), GGLE significantly increased over time (0.06, *p*<0.01), whereas no significant variation was observed during the second time interval between 1976 and 1995 (*p*=0.22). The GGLE significantly decreased over the last period, that is, 1996–2015 (− 0.04, *p*-value ≤ 0.01).

In the first period, positive contributions to the GGLE were observed for circulatory and endocrine diseases (positive parameters for FPC1: 1.67 and 31.38, respectively), whereas negative contributions were observed for other neoplasms, external causes of death, and infectious diseases (− 9.54, − 5.53, and − 13.29, respectively).

Compared with the external causes of death, the contributions of other neoplasms, circulatory diseases, and infectious diseases were mainly concentrated in older age groups (see Fig. 4 and figures in the Appendix). Negative parameters for FPC2 (− 21.32, − 4.74, and − 19.03 for other neoplasms, circulatory diseases, and infectious diseases, respectively) suggest that the shift toward older ages in the distributions of contributions of these causes involved a comparison of the differences over a narrower age range, which in turn corresponded to a reduction in the overall difference. Conversely, the distribution of the external causes of death contributions were mainly located at younger ages, and the shift in the distributions involved an expansion of the differences over a larger age range. Indeed, the FPC2 scores for external causes of death significantly contributed to a widening of the overall gender gap (17.17, *p*-value ≤ 0.01). Finally, changes in the distribution of lung

Table 1 Multiple linear regression between GGLE and both first and second FPC, for each cause of death, in three time intervals

		1959–1975		1976–1995		1996–2015	
		Estimate	<i>p</i> -Value	Estimate	<i>p</i> -Value	Estimate	<i>p</i> -Value
	Intercept	− 104.41	0.00	16.85	0.08	77.89	0.00
	Time	0.06	0.00	− 0.01	0.22	− 0.04	0.00
Lung cancer	FPC1	1.46	0.10	3.84	0.00	0.95	0.35
	FPC2	5.76	0.13	− 0.01	1.00	− 5.59	0.00
Other neoplasms	FPC1	− 9.54	0.00	− 1.20	0.22	− 0.53	0.60
	FPC2	− 21.32	0.00	− 8.91	0.00	− 15.83	0.00
Circulatory	FPC1	1.67	0.01	− 0.25	0.65	− 1.96	0.02
	FPC2	− 4.74	0.00	− 5.60	0.00	− 4.45	0.00
External	FPC1	− 5.53	0.01	3.66	0.01	13.83	0.00
	FPC2	17.17	0.01	9.39	0.02	2.63	0.54
Infectious	FPC1	− 13.29	0.00	− 44.38	0.00	− 49.56	0.00
	FPC2	− 19.03	0.00	− 3.59	0.65	19.42	0.00
Endocrine	FPC1	31.38	0.00	14.52	0.00	− 22.05	0.00
	FPC2	− 3.18	0.77	− 6.01	0.37	− 6.18	0.28
Respiratory	FPC1	− 2.39	0.04	− 6.96	0.00	− 5.21	0.01
	FPC2	− 5.97	0.04	− 5.02	0.01	6.22	0.00
Digestive	FPC1	0.64	0.68	− 9.49	0.00	− 4.93	0.02
	FPC2	10.78	0.10	1.24	0.68	14.33	0.00
	<i>adj-R</i> ²	0.65		0.69		0.86	

Statistically significant parameters in bold (significance level: 0.01)

cancer contributions, as well as respiratory and digestive diseases, did not significantly affect GGLE in 1959–1975 (*p*-values > 0.01 for both FPCs).

In the intermediate as well as in the last period, while the effects of other neoplasms and diseases of circulatory system became not statistically significant, the age pattern of the contributions of these causes still contributed to narrow the overall GGLE. The contributions of external causes of death switched from having a negative effect to having a positive effect on the GGLE.

Between 1976 and 1995, respiratory and digestive diseases showed negative effects. Since the contributions of other respiratory diseases were already mainly concentrated at older ages, the shift in the distributions of age-specific contributions toward older ages involved a slight compression of the distributions over a narrower age interval (Fig. 7 in the Appendix), thereby lowering the overall GGLE. However, in the last period (1996–2015), the shift in the distribution of age-specific contributions toward older ages widened the GGLE (FPC2: 6.22). During this period, while the effects of lung cancer were not statistically significant, the shift toward older age and the slight compression of the age-specific consequences of lung cancer contributed to the narrowing of the overall GGLE.

Notably, the evolution of GGLE was affected not only by the first component (the overall cause-specific contribution) but also by the second component (the age-specific location of the overall contribution) for other neoplasms, circulatory system diseases, external causes of death, and respiratory diseases. Additionally, the role of age-specific

contributions varied over time; for some causes, the first or second component had a negative impact (e.g., infectious diseases) or a positive one (e.g., external causes).

6 Discussion and conclusion

Our results provide new insights into the ongoing dynamics of the considered populations in terms of the contribution of age and cause of death to the overall GGLE. While some of these came as no surprise, we found evidence that has not been highlighted in the literature.

The existing literature on the evolution of the gender gap in LE demonstrates that from the beginning of the 1920s, in most industrialized countries, the gap widened until the 1970s and then started to narrow (Glei and Horiuchi 2007; Oksuzyan et al. 2008, 2010). The increase and decrease in the mortality–gender gap were mainly due to the smoking epidemic (Preston and Wang 2006; Pampel 2010). In addition to some exceptions (e.g., Sweden and the US), men began smoking considerably earlier than women (Meslé and Vallin 2011; Beltrán-Sánchez et al. 2015). Differences in life expectancy between women and men have begun to narrow down in many countries in Northern Europe, North America, and Oceania (Meslé 2004). Compared to other European countries, in most Eastern European countries, GGLE generally declined later (around mid-2000s) (Meslé 2004; Trias-Llimòs and Janssen 2018). Japan constituted a special case, since differences in LE between women and men increased gradually and started to decline only in the most recent period (Meslé 2004; Glei and Horiuchi 2007; Liu et al. 2013; Katanoda et al. 2015; Feraldi and Zarulli 2022). In accordance with previous studies, our findings highlight that, in the most recent period, gains in LE have been larger for males than for females, narrowing the gender gap (Oksuzyan et al. 2009; Meslé and Vallin 2011; Thorslund et al. 2013). In Spain, Japan, and most Eastern European countries, the life expectancy of men has improved at a slower pace than that of women (García González and Grande 2018; Gjonça et al. 2005). Additionally, our findings showed that the narrowing of the gender gap in LE is mainly driven by decreasing differences in cardiovascular diseases. Progress in reducing mortality due to cardiovascular diseases is faster in men than in women, although it is slower at older ages than at younger ones (Meslé 2004; Gjonça et al. 2005; Feraldi and Zarulli 2022). This reduction in cardiovascular mortality may be attributed to medical advances and improved lifestyle habits. Medical and technical advances related to cardiovascular health and mortality may have favored survival among men to a greater extent than among women, but less so at the oldest age (Thorslund et al. 2013). However, trends in gender ratios of cardiovascular mortality are worse in men than in women (Gjonça et al. 2005).

The average contribution of lung cancer to the GGLE increased until 1975 and started to decrease within the last time interval (1996–2015). Some authors have suggested that mortality from lung cancer (mainly caused by smoking) follows a pattern with three phases: a period of diffusion, a peak, and then a period of decline (Pampel 2003a, 2003b). Traditionally, men adopted smoking behaviors earlier than women; therefore, the consequences of smoking first appeared among men. When the GGLE was at its maximum, men were more affected than women; thus, the gender differences in mortality were larger. Mortality rates for women with lung cancer have increased one or two decades later, resulting in a reduction in the gap. Although countries experienced smoking epidemics at different times (e.g., earlier in England and Wales and later in France), in the 1990s, there was an overall

reduction in lung cancer mortality among men than women (Pampel 2003a, 2003b; Gjonça et al. 2005).

Observational comparative analyses of worldwide mortality trends have shown that, although cardiovascular disease mortality rates were twice as high as cancer mortality rates in 1970, cardiovascular disease mortality rates have declined significantly in the past three decades, whereas cancer mortality rates have been on the rise (Weir et al. 2016; Taksler and Rothberg 2017). Accordingly, our study also showed the increasing contribution of other neoplasms to the overall GGLE over time. This is commonly associated with improved early detection, prevention, and treatment measures for cardiovascular diseases coupled with the increasing longevity of the general population, which has led to a high prevalence of cancer (a disease of aging). In particular, it has been shown that males are at higher risk of all cancer incidence and mortality for a vast majority of sites (ReFaey et al. 2021). Neoplasms, (including lung cancer) are the second leading causes of death worldwide (Ferlay et al. 2013; Sitki Copur 2019). Excess male deaths are evident for colorectal, esophageal, liver, and skin cancers, which suggest disproportionate increase in relevant risk factors such as high-fat diet, alcohol consumption, Hepatitis B virus and/or herpes simplex virus infections, and UV exposure among men (Yang et al. 2012). Consequently, the contribution of cancer mortality to the gender gap in life expectancy has become more critical over time, especially in recent decades (Beltrán-Sánchez et al. 2015; Weir et al. 2016; Taksler and Rothberg 2017).

As found in other studies on OECD countries (Waldron et al. 2005; Backhans et al. 2012), our results highlight a reduction in the contribution of external causes to the GGLE over time. The combination of the convergence of gender roles and women's role expansion led to increased risk exposure and more risk-prone behaviors (e.g., regarding drug use and improvements in medical care or public health measures), which in turn determined increase in external-cause mortality among women. However, although very small, we find slightly increasing contributions in the last period.

The contributions of endocrine, nutritional, and metabolic diseases to the gender gap in life expectancy began in the late 1990s. This was also found in a recent study based on 22 European countries from 1959 to 2015 (Stefanucci and Mazzuco 2022), which showed that endocrine, nutritional, and metabolic disease-related deaths have increased in recent times for men (suggesting that poor dieting and alcohol consumption are increasingly impacting men's health). Finally, the contribution of infectious diseases to the overall gender gap may be due to the widely reported decline in mortality from infectious diseases (independent of increases in national income) in both women and men over the last century (Mackenbach 2013; Mackenbach and Looman 2013).

This study reveals that age- and cause-specific contributions to GGLE act almost entirely on two dimensions: level (greater/smaller contribution) and age pattern (location of the curves across ages), even considering the heterogeneity of space and time contained in the dataset. Remarkably, these two components explain almost all the variability in these contributions. Notably, in the last period, it is not the "quantum" of the cause-specific contributions that matters, but the "timing," that is, location across the age spectrum. Overall, in the last period, GGLE was affected more by the composition of the causes of death than in previous periods. This is plausible if one considers how far the mortality reduction from the main causes of death has gone for both genders, and that, until the next revolutionary medical breakthrough, further changes in the GGLE are more likely to occur from a redistribution over the age range of mortality differences.

This study demonstrated the advantages of the demographic application of the FDA. One of the key advantages of the FDA is its ability to smooth complex data, such as the relative contributions of age and causes of death to the GGLE in several countries over time, helping reduce the impact of random fluctuations and other irregularities that can arise when dealing with the causes of death and variations in coding protocols. Changes in the ICD can make it difficult to analyze mortality data owing to different causes of death over time. The FDA approach through statistical smoothing can help mitigate the impact of these changes by reducing the noise in the data. This makes it easier to identify trends and patterns. Additionally, by transforming discrete data into functional data, it is possible to apply FPCA to synthesize the temporal dynamics by cause and age of the GGLE for several countries with few components, which is not possible with traditional methods (e.g., Arriaga).

When analyzing a functional dataset, it is favorable to have a large number of observations for each sample curve (Aguilera and Aguilera-Morillo 2013). Considering that the utility of FPCA is to synthesize complex patterns, it may be beneficial to include multiple attempts to obtain a sufficiently sizeable amount of information. However, it is worth mentioning that applying the FDA to countries with different characteristics simultaneously could lead to the pooling of countries with different patterns. However, separating the analyses during these three periods may mitigate this issue. Conversely, by applying the FDA to demographic data, researchers can gain a more accurate and nuanced understanding of trends and patterns in population dynamics over time.

This study has a number of limitations. Comparisons of cause-specific mortality between countries over time may be affected by differences in coding practices. We attempted to overcome this by choosing broad classes of causes that were consistent across ICD revisions (Canudas-Romo et al. 2020). Additionally, due to the lack of reliable cause-of-death data for age groups older than 85 years, the analysis was restricted to age groups from 50 to 85 years. Thus, shifts in the distributions toward very old age might not be completely captured. Since these very old ages are becoming increasingly important in determining the GGLE (Zarulli et al. 2020), disaggregated mortality data at ages older than 85 years would help better focus on the relevant age contributions to the gender gap. Finally, regression models were conducted for three time intervals (1959–1975, 1976–1995, and 1996–2015) that included time (years) as a covariate to control for the overall GGLE trend. The purpose of the models was to isolate the effects of FPCs in each period regardless of the overall trend over time in the GGLE. Although this approach allowed for a more accurate interpretation of the impact of FPCs in each period, since countries transition during different time points, we might mix attempts with different patterns. However, this issue applies only to the second period (as shown in Fig. 1).

Despite these limitations, the added value of this study is that it analyzes, for the first time, age- and cause-specific contributions to the GGLE using an FDA approach, which could prove useful in deepening our understanding of complex demographic phenomena. Therefore, we encourage the application of this approach to population studies.

Appendix

See Tables 2 and 3

Table 2 Classes of causes of death with related International Classification of Diseases (ICD) codes for each revision

Cause group	ICD-7	ICD-8	ICD-9	ICD-10
Lung cancer	A050	A051	B101	C33-C34
Other neoplasms	A044-A060, B018-B019	A045-A050, A052-A061	B08-B09, B100, B109, B11-B17	C00-C32, C34-D48
Diseases of circulatory system	A070, A079-A086, B022, B024-B029	A080-A088	B25-B30	I00-I99
External causes of death	A138-A150, B047-B050	A138-A150	B47-B56	V00-Y89
Respiratory diseases	A087-A097, B030-B032	A089-A096	B31-B32	J00-I98
Diseases of digestive system	A098-A107, B033-B037	A097-A104	B33-B34	K00-K93
Endocrine, nutritional and metabolic diseases	A061-A064, B020	A062-A066	B18-B19	E00-E88
Infectious and parasitic diseases	A001-A043, A104, A132, B001-B017, B043	A001-A044	B01-B007	A00-B99

Table 3 List of causes of death in each group

Lung cancer	Respiratory diseases
Malignant neoplasms of trachea, bronchus and lung	Influenza, Pneumonia, Other acute respiratory infections, Asthma
Diseases of circulatory system	Other neoplasms
Rheumatic heart diseases, Essential hypertension, Hypertensive disease (heart, kidney and secondary), Acute myocardial infarction, Other IHD, Pulmonary heart diseases, Non rheumatic valve disorders, Cardiac arrest, Heart failure, Other heart diseases, Intracranial haemorrhage, Cerebral infarction, occlusion, and stenosis, Other cerebrovascular diseases, Sequelae of cerebrovascular disease, Diseases of arteries, arterioles and capillaries, Other circulatory diseases	Malignant neoplasms of lip, oral cavity and pharynx, esophagus, stomach, colon, rectum and anus, liver and intrahepatic bile ducts, pancreas, digestive system, larynx, skin, breast, cervix uteri, uterus, ovary, prostate, other genital organs, bladder, kidney and other urinary organ, meninges, brain and other parts of central nervous system, Leukemia, Other lymphoid, hematopoietic and related tissue, independent (primary) multiple sites, Other cancer, In situ neoplasms, benign neoplasms and neoplasms of uncertain or unknown behavior
External causes of death	Infectious and parasitic diseases
Transport accidents, Accidental falls, Accidental drowning and submersion, Accidental exposure to smoke, fire and flames, Accidental poisoning by alcohol, Accidental poisoning by other substance, Other accidental threats to breathing, Suicide and self-inflicted injury, Assault, Event of undetermined intent, Complications of medical and surgical care, Other accidents and late effects of accidents (remainder)	Other specified intestinal infections, Diarrhoea and gastroenteritis of presumed infectious origin, Septicaemia, Other bacterial diseases, HIV disease, Viral hepatitis, Other viral diseases, Other and unspecified infectious and parasitic diseases
Diseases of digestive system	Endocrine, nutritional and metabolic diseases
Gastric and duodenal ulcer, Hernia, Enteritis, colitis and other intestinal diseases, Alcoholic cirrhosis of liver, Other cirrhoses of liver, Other diseases of liver, Cholelithiasis and other disorders of biliary tracts, Diseases of pancreas, Other digestive diseases	Diabetes mellitus, Malnutrition, Other endocrinologic and metabolic diseases

Functional principal component analysis

See Fig. 5

Regarding other neoplasms, the first component captures the extent of the contribution of other neoplasms' mortality to the overall gender differences. According to FPC1, the modal age of contribution was approximately 70 years. A high score for this component suggests an above-average contribution of other neoplasms to GGLE. The second FPC describes the location of the curves across ages. With respect to the overall mean contribution (solid line), the (+) curve denotes considerably higher contributions at older ages and lower contributions at younger ages. Conversely, the (-) curve indicates lower contributions at older ages and higher contributions at younger ages. Indeed, the modal age at the contribution of the (+) curve is older than that of the (-) curve. A high score for this component

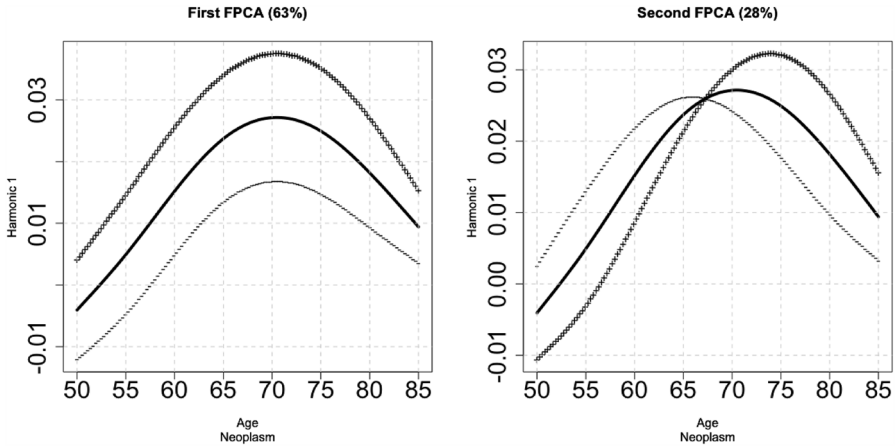


Fig. 5 Effect of the first two FPCs on age-specific contributions to the GGLE for other neoplasms: overall mean and $\text{mean} \pm$ a suitable multiple of the principal component weight function

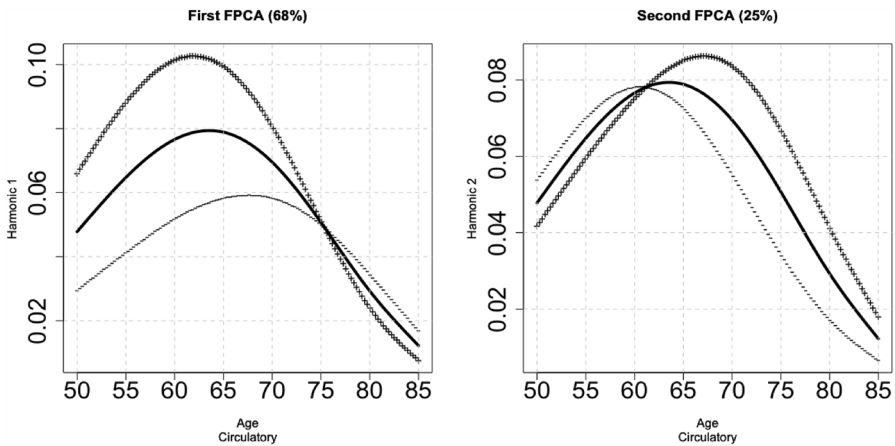


Fig. 6 Effect of the first two FPCs on age-specific contributions to the GGLE for diseases of circulatory system: overall mean and $\text{mean} \pm$ a suitable multiple of the principal component weight function

suggests an above-average concentration of the distribution of other neoplasm contributions at older ages.

See Fig. 6.

Similar to lung cancer, the first FPC related to diseases of the circulatory system shows a modal age of contribution at age 60–65 years, and a high score on this component suggests an above-average contribution of diseases of the circulatory system to the GGLE. According to the second FPC, a high score indicated an increasing contribution of circulatory system diseases at older ages.

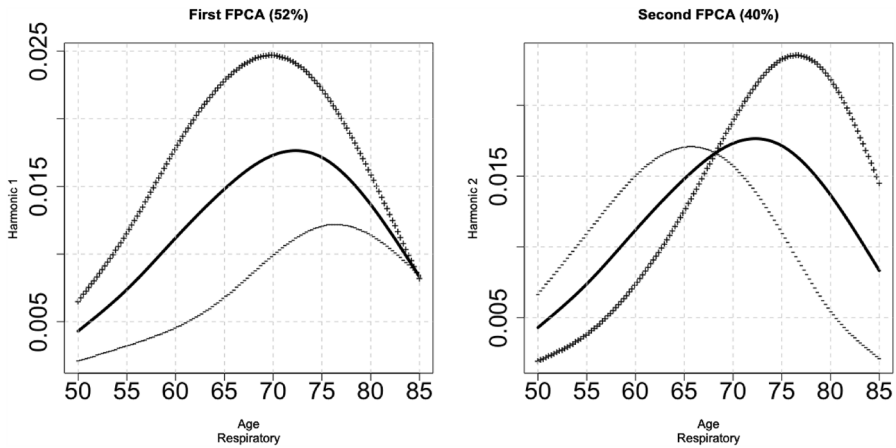


Fig. 7 Effect of the first two FPCs on age-specific contributions to the GGLE for respiratory diseases: overall mean and $\text{mean} \pm$ a suitable multiple of the principal component weight function

See Fig. 7.

FPC1 related to respiratory diseases showed a modal age at contribution of 70–75 years. A high score for this component suggests a significant contribution of respiratory diseases to the GGLE. According to the second FPC, a high score denotes an increasing contribution of respiratory diseases at older ages (as well as a decreasing contribution at younger ages) and a compression of the differences over a narrower age range.

Regarding digestive, infectious, and metabolic diseases, only the first functional principal component was sufficient to explain the largest part of the variability in age-specific contributions (between 84 and 93%). The second functional principal component explained less than 8% of the variability (Figs. 8, 9, and 10).

Figure 8 (digestive diseases) and Fig. 9 (infectious diseases) show that a high score for the first component suggests an above-average contribution of digestive and infectious

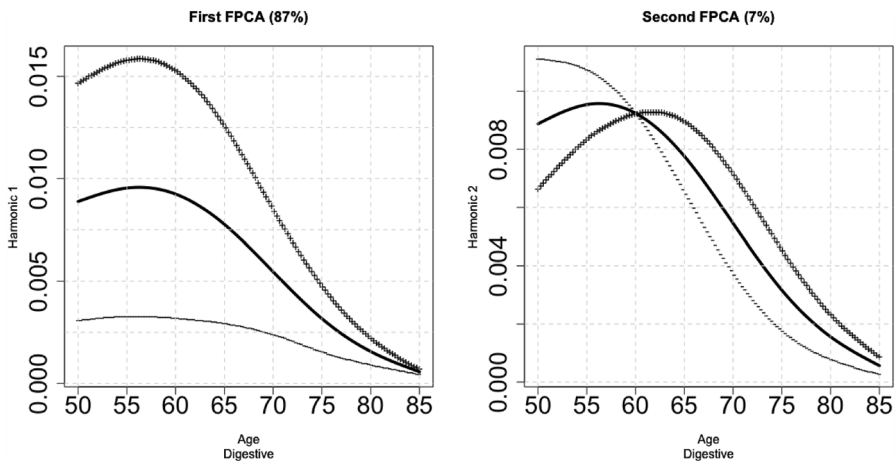


Fig. 8 Effect of the first two FPCs on age-specific contributions to the GGLE for digestive diseases: overall mean and $\text{mean} \pm$ a suitable multiple of the principal component weight function

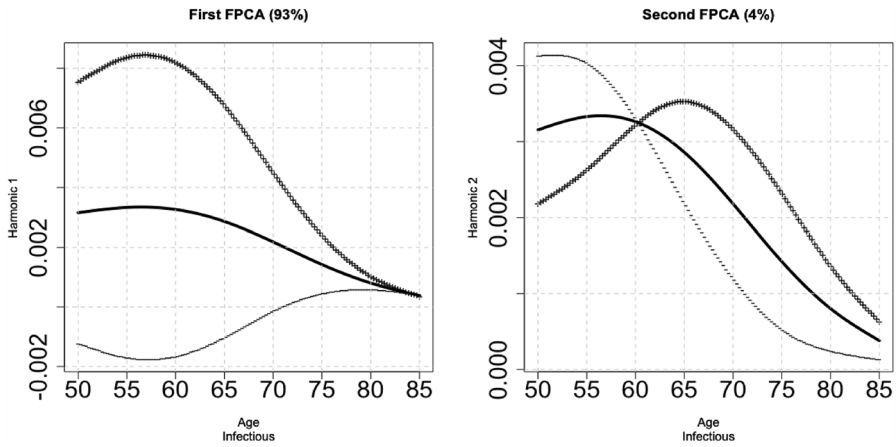


Fig. 9 Effect of the first two FPCs on age-specific contributions to the GGLE for infectious diseases: overall mean and mean \pm a suitable multiple of the principal component weight function

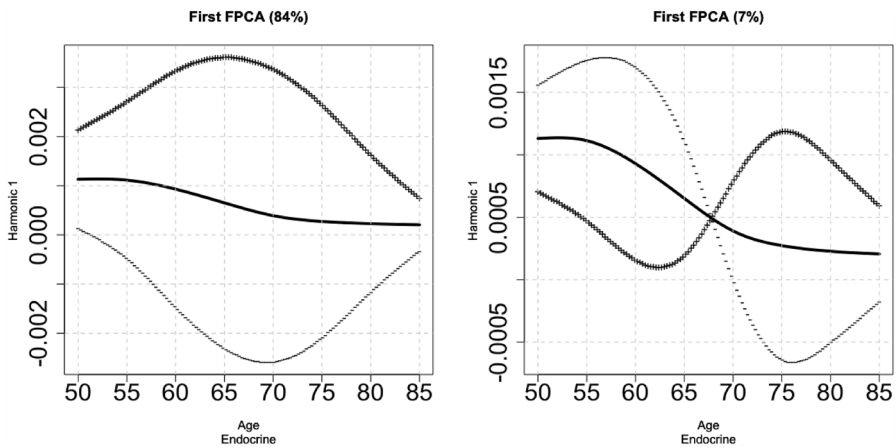


Fig. 10 Effect of the first two FPCs on age-specific contributions to the GGLE for metabolic diseases: overall mean and mean \pm a suitable multiple of the principal component weight function

diseases to the GGLE. The contributions of endocrine, nutritional, and metabolic diseases to the GGLE are mainly concentrated at older ages (60–75 years), and a high score for the first component suggests an above-average contribution of metabolic diseases to the GGLE (Fig. 10).

Author contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by AF, with input from VZ, SM and CG. The first draft of the manuscript was written by AF and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Availability of data and materials The data analysed during the current study is freely available from the Human Mortality Database (<https://www.mortality.org/>) and from the WHO Mortality Database (http://www.who.int/healthinfo/mortality_data/en/).

Declarations

Conflict of interest The authors declare no conflicts of interest.

Ethical approval The authors confirm that the ethical policies of the journal, as noted on the journal's author guidelines page, have been adhered to.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aburto, J.M., Villavicencio, F., Basellini, U., Kjærgaard, S., Vaupel, J.W.: Dynamics of life expectancy and life span equality. *Proc. Natl. Acad. Sci. U. S. A.* **117**(10), 5250–5259 (2020). <https://doi.org/10.1073/pnas.1915884117>
- Aguilera, A.M., Aguilera-Morillo, M.C.: Penalized PCA approaches for B-spline expansions of smooth functional data. *Appl. Math. Comput.* **219**(14), 7805–7819 (2013)
- Arriaga, E.E.: Measuring and explaining the change in life expectancies. *Demography* **21**(1), 83–96 (1984)
- Austad, S.N.: Why women live longer than men: sex differences in longevity. *Sex Med.* **3**(2), 79–92 (2006). [https://doi.org/10.1016/s1550-8579\(06\)80198-1](https://doi.org/10.1016/s1550-8579(06)80198-1)
- Backhans, M., Burström, B., de Leon, A.P., Marklund, S.: Is gender policy related to the gender gap in external cause and circulatory disease mortality? A mixed effects model of 22 OECD countries 1973–2008. *BMC Public Health* **12**(1), 1–12 (2012)
- Beltrán-Sánchez, H., Finch, C.E., Crimmins, E.M.: Twentieth century surge of excess adult male mortality. *Proc. Natl. Acad. Sci. U.S.A.* **112**(29), 8993–8998 (2015). <https://doi.org/10.1073/pnas.1421942112>
- Bongaarts, J.: Long-range trends in adult mortality: models and projection methods. *Demography* **42**(1), 23–49 (2005). <https://doi.org/10.1353/dem.2005.0003>
- Boyd, A., Van de Velde, S., Vilagut, G., De Graaf, R., Florescu, S., Alonso, J., EUWMH Investigators: Gender differences in mental disorders and suicidality in Europe: results from a large cross-sectional population-based study. *J. Affect. Disord.* **173**, 245–254 (2015)
- Canudas-Romo, V.: The modal age at death and the shifting mortality hypothesis. *Demogr. Res.* **19**, 1179–1204 (2008)
- Canudas-Romo, V., Guillot, M.: Truncated cross-sectional average length of life: a measure for comparing the mortality history of cohorts. *Popul. Stud.* **69**(2), 147–159 (2015)
- Canudas-Romo, V., Adair, T., Mazucco, S.: Reflection on modern methods: cause of death decomposition of cohort survival comparisons. *Int. J. Epidemiol.* **49**(5), 1712–1718 (2020)
- Chisumpa, V.H., Odimegwu, C.O.: Decomposition of age- and cause-specific adult mortality contributions to the sex gap in life expectancy from census and survey data in Zambia. *SSM Popul. Health* **5**, 218–226 (2018). <https://doi.org/10.1016/j.ssmph.2018.07.003>
- Conti, S., Farchi, G., Masocco, M., Minelli, G., Toccaceli, V., Vichi, M.: Gender differentials in life expectancy in Italy. *Eur. J. Epidemiol.* **18**(2), 107–112 (2003)
- Copur, M.S.: State of cancer research around the globe. *Oncology* **33**(5), 181–185 (2019)
- Désesquelles, A.F., Salvatore, M.A., Pappagallo, M., Frova, L., Pace, M., Meslé, F., Egidi, V.: Analysing multiple causes of death: which methods for which data? An application to the cancer-related mortality in France and Italy. *Eur. J. Popul.* **28**(4), 467–498 (2012)

- Di Girolamo, C., Nusselder, W.J., Bopp, M., Brønnum-Hansen, H., Costa, G., Kovács, K., Leinsalu, M., Martikainen, P., Pacelli, B., Rubio Valverde, J., Mackenbach, J.P.: Progress in reducing inequalities in cardiovascular disease mortality in Europe. *Heart (British Cardiac Soc.)* **106**(1), 40–49 (2020). <https://doi.org/10.1136/heartjnl-2019-315129>
- Erbas, B., Hyndman, R.J., Gertig, D.M.: Forecasting age-specific breast cancer mortality using functional data models. *Stat. Med.* **26**(2), 458–470 (2007)
- Feraldi, A., Zarulli, V.: Patterns in age and cause of death contribution to the sex gap in life expectancy: a comparison among ten countries. *Genus* **78**(1), 1–22 (2022)
- Ferlay, J., Steliarova-Foucher, E., Lortet-Tieulent, J., Rosso, S., Coebergh, J.W.W., Comber, H., Forman, D., Bray, F.: Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur. J. Cancer* **49**(6), 1374–1403 (2013). <https://doi.org/10.1016/j.ejca.2012.12.027>
- Ferlay, J., Colombet, M., Soerjomataram, I., Dyba, T., Randi, G., Bettio, M., Gavin, A., Visser, O., Bray, F.: Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. *Eur. J. Cancer* **103**, 356–387 (2018). <https://doi.org/10.1016/j.ejca.2018.07.005>
- García González, J.M., Grande, R.: Cambios en las diferencias por sexo en la esperanza de vida en España (1980–2012): descomposición por edad y causa. *Gac. Sanit.* **32**, 151–157 (2018). <https://doi.org/10.1016/j.gaceta.2017.03.004>
- Gjonça, A., Tomassini, C., Toson, B., Smallwood, S.: Sex differences in mortality, a 568 comparison of the United Kingdom and other developed countries. *Health Stat. Q.* **569**(26), 6–16 (2005)
- Glei, D.A., Meslé, F., Vallin, J.: Diverging trends in life expectancy at age 50: a look at causes of death. In: *International Differences in Mortality at Older Ages: Dimensions and Sources*, 17–67 (2010)
- Glei, D.A., Horiuchi, S.: The narrowing sex differential in life expectancy in high-income populations: effects of differences in the age pattern of mortality. *Popul. Stud.* **61**(2), 141–159 (2007). <https://doi.org/10.1080/00324720701331433>
- Goldin, C., Lleras-Muney, A.: XX > XY?: the changing female advantage in life expectancy. *J. Health Econ.* **67**, 102224 (2019)
- Gómez-Redondo, R., Boe, C.: Decomposition analysis of Spanish life expectancy at birth: evolution and changes in the components by sex and age. *Demogr. Res.* **13**, 521–546 (2005)
- Harteloh, P., De Bruin, K., Kardaun, J.: The reliability of cause-of-death coding in The Netherlands. *Eur. J. Epidemiol.* **25**(8), 531–538 (2010)
- Ho, J.Y., Hendi, A.S.: Recent trends in life expectancy across high income countries: retrospective observational study. *bmj* **362**, k2562 (2018)
- Hollingshaus, M., Utz, R., Schacht, R., Smith, K.R.: Sex ratios and life tables: historical demography of the age at which women outnumber men in seven countries, 1850–2016. *Hist. Methods J. Quant. Interdiscip. Hist.* **52**(4), 244–253 (2019)
- Horiuchi, S., Ouellette, N., Cheung, S.L.K., Robine, J.-M. Modal age at death: lifespan indicator in the era of longevity extension. In: *Vienna Yearbook of Population Research* 37–69 (2013)
- Human Mortality Database.: Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany) (downloaded on 2022) (2022)
- Hyndman, R.J., Booth, H.: Stochastic population forecasts using functional data models for mortality, fertility and migration. *Int. J. Forecast.* **24**(3), 323–342 (2008)
- Hyndman, R.J., Shang, H.L.: Forecasting functional time series. *J. Korean Stat. Soc.* **38**(3), 199–211 (2009)
- Hyndman, R.J., Ullah, M.S.: Robust forecasting of mortality and fertility rates: a functional data approach. *Comput. Stat. Data Anal.* **51**(10), 4942–4956 (2007)
- Janssen, F.: Similarities and differences between sexes and countries in the mortality imprint of the smoking epidemic in 34 low-mortality countries, 1950–2014. *Nicotine Tob. Res.* **22**(7), 1210–1220 (2020)
- Janssen, F., de Beer, J.: The timing of the transition from mortality compression to mortality delay in Europe, Japan and the United States. *Genus* **75**(1), 10 (2019)
- Janssen, F., van Poppel, F.: The adoption of smoking and its effect on the mortality gender gap in Netherlands: a historical perspective. *BioMed Res. Int.* (2015)
- Katanoda, K., Hori, M., Matsuda, T., Shibata, A., Nishino, Y., Hattori, M., Soda, M., Ioka, A., Sobue, T., Nishimoto, H.: An updated report on the trends in cancer incidence and mortality in Japan, 1958–2013. *Jpn. J. Clin. Oncol.* **45**(4), 390–401 (2015). <https://doi.org/10.1093/jjco/hyv002>
- Klenk, J., Keil, U., Jaensch, A., Christiansen, M.C., Nagel, G.: Changes in life expectancy 1950–2010: contributions from age- and disease-specific mortality in selected countries. *Popul. Health Metr.* **14**, 20 (2016). <https://doi.org/10.1186/s12963-016-0089-x>
- Kuehner, C.: Why is depression more common among women than among men? *Lancet Psychiatry* **4**(2), 146–158 (2017)

- Léger, A.E., Mazzucco, S.: What can we learn from the functional clustering of mortality data? An application to the human mortality database. *Eur. J. Popul.* **37**(4), 769–798 (2021)
- Liu, Y., Arai, A., Kanda, K., Lee, R.B., Glasser, J., Tamashiro, H.: Gender gaps in life expectancy: generalized trends and negative associations with development indices in OECD countries. *Eur. J. Public Health* **23**(4), 563–568 (2013)
- Luy, M., Wegner-Siegmundt, C.: The impact of smoking on sex differences in life expectancy: more heterogeneous than often stated. *Eur. J. Public Health* **25**(4), 706–710 (2015). <https://doi.org/10.1093/eurpub/cku211>
- Mackenbach, J.P.: Political conditions and life expectancy in Europe, 1900–2008. *Soc. Sci. Med.* **82**, 134–146 (2013)
- Mackenbach, J.P., Looman, C.W.: Life expectancy and national income in Europe, 1900–2008: an update of Preston's analysis. *Int. J. Epidemiol.* **42**(4), 1100–1110 (2013)
- Mazzucco S.: Measuring mortality crises: a tool for studying global health. Oxford Research Encyclopedia of Global Public Health (2022). <https://oxfordre.com/publichealth/view/https://doi.org/10.1093/acrefore/9780190632366.001.0001/acrefore-9780190632366-e-434>. Accessed 23 Dec 2022
- Meslé, F.: Sex gap in life expectancy: the reasons for a reduction of female advantage. *Revue D'epidemiologie Et De Sante Publique* **52**(4), 333–352 (2004). [https://doi.org/10.1016/s0398-7620\(04\)99063-3](https://doi.org/10.1016/s0398-7620(04)99063-3)
- Meslé, F., Vallin, J.: Reconstructing long-term series of causes of death: the case of France. *Hist. Methods J. Quant. Interdiscip. Hist.* **29**(2), 72–87 (1996)
- Meslé, F., Vallin, J.: Historical trends in mortality. In: *International Handbook of Adult Mortality*, pp. 9–47. Springer (2011)
- Meslé, F., Vallin, J., Andreyev, Z.: Mortality in Europe: the divergence between east and west. *Population* **57**(1), 157–197 (2002)
- Nepomuceno, M.R., Canudas-Romo, V.: A cohort survival comparison between Central Eastern European and high-longevity countries. *Population* **74**(3), 281–302 (2019)
- Nusselder, W.J., Looman, C.W.N., Van Oyen, H., Robine, J.M., Jagger, C.: Sex differences in health of EU10 and EU15 populations: the double burden of EU10 men. *Eur. J. Ageing* **7**(4), 219–227 (2010). <https://doi.org/10.1007/s10433-010-0169-x>
- Oksuzyan, A., Juel, K., Vaupel, J.W., Christensen, K.: Men: Good health and high mortality. Sex differences in health and aging. *Aging Clin. Exp. Res.* **20**(2), 91–102 (2008). <https://doi.org/10.1007/BF03324754>
- Oksuzyan, A., Petersen, I., Stovring, H., Bingley, P., Vaupel, J.W., Christensen, K.: The male-female health-survival paradox: a survey and register study of the impact of sex-specific selection and information bias. *Ann. Epidemiol.* **19**(7), 504–511 (2009). <https://doi.org/10.1016/j.annepidem.2009.03.014>
- Oksuzyan, A., Brønnum-Hansen, H., Jeune, B.: Sex gap in health expectancy. *Eur. J. Ageing* **7**(4), 213–218 (2010). <https://doi.org/10.1007/s10433-010-0170-4>
- Pampel, F.C.: Declining sex differences in mortality from lung cancer in high-income nations. *Demography* **40**(1), 45–65 (2003a)
- Pampel, F.C.: Sex differences in mortality from lung cancer in population and development review. *Demography* **40**(1), 45–65 (2003b)
- Pampel, F.C.: Divergent patterns of smoking across high-income nations. In: *International Differences in Mortality at Older Ages: Dimensions and Sources*, pp. 132–163 (2010)
- Pinkhasov, R.M., Shteynshlyuger, A., Hakimian, P., Lindsay, G.K., Samadi, D.B., Shabsigh, R.: Are men shortchanged on health? Perspective on life expectancy, morbidity, and mortality in men and women in the United States. *Int. J. Clin. Pract.* **64**(4), 465–474 (2010). <https://doi.org/10.1111/j.1742-1241.2009.02289.x>
- Preston, S.H., Wang, H.: Sex mortality differences in the United States: the role of cohort smoking patterns. *Demography* **43**(4), 631–646 (2006)
- Preston, S.H., Heuveline, P., Guillot, M.: [BOOK REVIEW] demography, measuring and modeling population processes. *Popul. Dev. Rev.* **27**(2), 365–367 (2001)
- Ramsay, J.O., Silverman, B.W.: *Applied Functional Data Analysis: Methods and Case Studies*. Series in Statistics. Springer (2002)
- Ramsay, J.O., Silverman, B.W.: Principal components analysis for functional data. In: *Functional Data Analysis*, pp. 147–172 (2005)
- Ramsay, J., Minton, J., Fischbacher, C., Fenton, L., Kaye-Bardgett, M., Wyper, G.M., McCartney, G.: How have changes in death by cause and age group contributed to the recent stalling of life expectancy gains in Scotland? Comparative decomposition analysis of mortality data, 2000–2002 to 2015–2017. *BMJ Open* **10**(10), e036529 (2020)

- Ramsay, J.O., Hooker, G., Graves, S.: Introduction to Functional Data Analysis (pp. 1–19) (2009). https://doi.org/10.1007/978-0-387-98185-7_1
- Ramsay, J., Wickham, H., Graves, S., Hooker, G.: fda: functional data analysis. R package (2011)
- Ramsay, J.O.: Monotone regression splines in action. In: Statistical Science, pp. 425–441 (1988)
- Rasul, I.: Reducing risky behaviour through the provision of information Research report (2013)
- ReFaey, K., Tripathi, S., Grewal, S.S., Bhargava, A.G., Quinones, D.J., Chaichana, K.L., Quinones-Hinojosa, A.: Cancer mortality rates increasing vs cardiovascular disease mortality decreasing in the world: future implications. *Mayo Clin. Proc. Innov. Qual. Outcomes* **5**(3), 645–653 (2021)
- Rogers, R.G., Everett, B.G., Onge, J.M.S., Krueger, P.M.: Social, behavioral, and 723 biological factors, and sex differences in mortality. *Demography* **47**(3), 555–578 (2010). <https://doi.org/10.1353/dem.0.0119>
- Sagna, A.O., Kemp, M.L.S., DiNitto, D.M., Choi, N.G.: Impact of suicide mortality on life expectancy in the United States, 2011 and 2015: age and sex decomposition. *Public Health* **179**, 76–83 (2020)
- Sanders, S.: Changing trends in mortality: a cross-UK comparison, 1981 to 2016. Office for National Statistics (2018)
- Sauerberg, M., Canudas-Romo, V.: Comparing cohort survival in good health: a research note on decomposing sex differentials in the United States. *Demography* **59**, 2013 (2022)
- Shair, S.N., Zolkifi, N.A., Zulkafi, N.F., Murad, A.: A functional data approach to the estimation of mortality and life expectancy at birth in developing countries. *Pertanika J. Sci. Technol.* **27**(2), 797 (2019)
- Stefanucci, M., Mazzuco, S.: Analysing cause-specific mortality trends using compositional functional data analysis. *J. R. Stat. Soc. Ser. A Stat. Soc.* **185**, 61 (2022)
- Sundberg, L., Agahi, N., Fritzell, J., Fors, S.: Why is the sex gap in life expectancy decreasing? The impact of ageand cause-specific mortality in Sweden 1997–2014. *Int. J. Public Health* **63**(6), 673–681 (2018). <https://doi.org/10.1007/s00038-018-1097-3>
- Taksler, G.B., Rothberg, M.B.: Assessing years of life lost versus number of deaths in the United States, 1995–2015. *Am. J. Public Health* **107**(10), 1653–1659 (2017)
- Tarkiainen, L., Martikainen, P., Laaksonen, M., Valkonen, T.: Trends in life expectancy by income from 1988 to 2007: decomposition by age and cause of death. *J. Epidemiol. Community Health* **66**(7), 573–578 (2012). <https://doi.org/10.1136/jech.2010.123182>
- Thorslund, M., Wastesson, J.W., Agahi, N., Lagergren, M., Parker, M.G.: The rise and fall of women's advantage: a comparison of national trends in life expectancy at age 65 years. *Eur. J. Ageing* **10**(4), 271–277 (2013). <https://doi.org/10.1007/s10433-013-0274-8>
- Treurniet, H.F., Boshuizen, H.C., Harteloh, P.P.M.: Avoidable mortality in Europe (1980–1997): a comparison of trends. *J. Epidemiol. Community Health* **58**(4), 290–295 (2004)
- Trias-Llimós, S., Janssen, F.: Alcohol and sex gaps in life expectancy in eight Central and Eastern European countries. *Eur. J. Pub. Health* **28**(4), 687–692 (2018). <https://doi.org/10.1093/eurpub/cky057>
- Trovato, F., Heyen, N.B.: A varied pattern of change of the sex differential in survival in the G7 countries. *J. Biosoc. Sci.* **38**(3), 391–401 (2006). <https://doi.org/10.1017/S0021932005007212>
- Ullah, S., Finch, C.F.: Functional data modelling approach for analysing and predicting trends in incidence rates—an application to falls injury. *Osteoporos. Int.* **21**(12), 2125–2134 (2010)
- Ullah, S., Finch, C.F.: Applications of functional data analysis: a systematic review. *BMC Med. Res. Methodol.* **13**, 43 (2013). <https://doi.org/10.1186/1471-2288-13-43>
- Waldron, I., McCloskey, C., Earle, I.: Trends in gender differences in accidents mortality: relationships to changing gender roles and other societal trends. *Demogr. Res.* **13**, 415–454 (2005)
- Weir, H.K., Anderson, R.N., King, S.M.C., Soman, A., Thompson, T.D., Hong, Y., Leadbetter, S.: Peer reviewed: heart disease and cancer deaths—trends and projections in the United States, 1969–2020. *Prevent. Chronic Disease* **13**, e157 (2016)
- WHO, M.: World Health Organization Mortality Database. WHO (2022). http://www.who.int/healthinfo/mortality_data/en/ (accessed 2022).
- Yang, S., Khang, Y.-H., Harper, S., Davey Smith, G., Leon, D.A., Lynch, J.: Understanding the rapid increase in life expectancy in South Korea. *Am. J. Public Health* **100**(5), 896–903 (2010). <https://doi.org/10.2105/AJPH.2009.160341>
- Yang, Y., Li, T., Nielsen, M.E.: Aging and cancer mortality: dynamics of change and sex differences. *Exp. Gerontol.* **47**(9), 695–705 (2012)
- Zarulli, V., Lindahl-Jacobsen, R., Vaupel, J.W.: Onset of the old-age sex gap in survival. *Demogr. Res.* **42**, 727–740 (2020)
- Zarulli, V., Kashnitsky, I., Vaupel, J.W.: Death rates at specific life stages mold the sex gap in life expectancy. *Proc. Natl. Acad. Sci.* **118**, 20 (2021)

- Zatonski, W.A., Bhala, N.: Changing trends of diseases in Eastern Europe: Closing the gap. *Public Health* **126**(3), 248–252 (2012). <https://doi.org/10.1016/j.puhe.2011.11.017>
- Zatoński, W., Didkowska, J.: Closing the gap: cancer in Central and Eastern Europe (CEE). *Eur. J. Cancer* **44**(10), 1425–1437 (2008)
- Zuo, W., Jiang, S., Guo, Z., Feldman, M.W., Tuljapurkar, S.: Advancing front of old-age human survival. *Proc. Natl. Acad. Sci. u.s.a.* **115**(44), 11209–11214 (2018). <https://doi.org/10.1073/pnas.1812337115>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.