



# Multi-country clustering-based forecasting of healthy life expectancy

Susanna Levantesi<sup>1</sup> · Andrea Nigri<sup>2</sup> · Gabriella Piscopo<sup>3</sup>  · Alessandro Spelta<sup>4</sup>

Accepted: 30 December 2022 / Published online: 14 January 2023  
© The Author(s), under exclusive licence to Springer Nature B.V. 2023

## Abstract

Healthy life expectancy (HLE) is an indicator that measures the number of years individuals at a given age are expected to live free of disease or disability. HLE forecasting is essential for planning the provision of health care to elderly populations and appropriately pricing Long Term Care insurance products. In this paper, we propose a methodology that simultaneously forecasts HLE for groups of countries and allows for investigating similarities in their HLE patterns. We firstly apply a functional data clustering to the multivariate time series of HLE at birth of different countries for the years 1990–2019 provided by the Global Burden of Disease Study. Three clusters are identified for both genders. Then, we carry out the HLE simultaneous forecasting of the populations within each cluster by a multivariate random walk with drift. Numerical results and the statistical significance of the parameters of the identified multivariate processes are shown. Demographic evidences on the different evolution of HLE between countries are commented.

**Keywords** Healthy life expectancy · Forecasting · Functional data clustering

---

✉ Gabriella Piscopo  
gabriella.piscopo@unina.it

Susanna Levantesi  
susanna.levantesi@uniroma1.it

Andrea Nigri  
andrea.nigri@unifg.it

Alessandro Spelta  
alessandro.spelta01@universitadipavia.it

<sup>1</sup> Department of Statistics, Sapienza University of Rome, Rome, Italy

<sup>2</sup> Department of Economics, Management and Territory, University of Foggia, Foggia, Italy

<sup>3</sup> Department of Economic and Statistical Science, University of Naples Federico II, Naples, Italy

<sup>4</sup> Department of Economics and Management, Pavia University, Pavia, Italy

## 1 Introduction

The rising in longevity is one of the most astonishing achievements of modern society, it is a result of continuous progress in medicine, nutrition, and technology (Riley 2001). This leads to a looming and rapid population aging, and a consequently increased prevalence of chronic diseases, especially among the elderly. Among bio-demographic theories, we found that morbidity was generally associated with diseases, and longevity with biological senescence (Manton 1982). In its seminal work, Fries (1980) theorized that the age at the onset of chronic degenerative diseases could be postponed compressing morbidity into a smaller portion of the life span, thus without significantly changing average life expectancy.

The rapid decrease in mortality at advanced ages suggested that a biological limit was not close, without indication of a deterioration in the health status of older people despite significant increases in life expectancy. By aiming at investigating the nature of these scenarios, it is useful to monitor both mortality and the incidence/ prevalence of morbidity over time. To address these issues, relevant research avenues have evolved recently, promoting the creation of health expectancy indicators (Sanders 1964; Sullivan 1965, 1971) that measure the number of years that individuals at a given age are expected to live in good health under prevailing mortality and morbidity conditions. More precisely, HLE represents the expected number of years of remaining disability-free life a member of the life table cohort would experience if cohort age-specific rates of mortality and disability prevailed throughout his/her lifetime. The basic idea is to merge the period life table, with the age-specific disability prevalence obtained from cross-sectional survey data. In particular, Sullivan's method simply partitions the total number of person-years lived into the disability and disability-free life expectancy.

According to Imai and Soneju (2007), similarly to life expectancy at birth, we are able to define HLE, denoted by  $h(a, t)$ , which represents the expected remaining disability-free (DF) life of an individual age  $a$  born at time  $t$ . Formally, given the survival function  $l(a, t)$ , and let  $\pi(a, t)$  be the proportion disabled at age  $a$  for the cohort born at time  $t$ , HLE is given by:

$$h(a, t) = \frac{1}{l(a, t)} \int_a^{\infty} [1 - \pi(r, t)] l(r, t) dr$$

Since the theoretical definition is given within the continuous-time framework and the data are typically recorded in a discrete form, we exploit the period life table model in order to approximate the continuous-time mortality process and use Sullivan's method. We then obtain the discrete counterpart of  $h(a, t)$ , denoted by  $h_{a,t}$ . A reliable estimation of HLE is fundamental for understanding whether additional years of life are spent in good health and whether life expectancy is increasing faster than the decline of disability. Besides Levantesi et al. (2022) who contributed to underlined heterogeneity in longevity indicators from a global perspective, it is a common practice in life expectancy literature to model each population independently. Indeed, significant studies have been proposed on life expectancy forecasting, see, e.g., Torri (2011) that bring forward a Geometric Brownian motion-based model. Another relevant contribution comes from Raftery et al. (2013) who propose a hierarchical Bayesian model, finally Nigri et al. (2021a) bring forward a neural network approach to forecast longevity measures accounting for long and short term dynamics.

Nevertheless, after a long period of increasing life expectancy, researchers recorded some diverging trends in longevity, whit a situation of stagnation and deceleration (Nigri et al. 2022, 2021b; Ho and Hendi 2018). Regularities have been observed also in HLE (Permanyer et al. 2022a) with a less optimistic picture obtained looking at life expectancy

trends. Therefore, it seems appropriate to identify clusters of countries having common characteristics detected by life expectancy before conducting the forecast by a multi-population model. Aiming at providing for the first time an HLE forecasting, we analyze the patterns of HLE at birth time series by identifying different longevity phases and transitions that allow clustering analysis of countries in line with their longevity dynamics. Leveraging a functional cluster analysis, which allows us to analyze curves rather than scalar data we try to track similarities in HLE among countries. Indeed, if some countries share macrolevel common factors, such as similar improvements in public health or economic circumstances, joint forecasting can be appropriate. Therefore, using the obtained cluster information, we then perform a simultaneous HLE forecasting inside each specific cluster by implementing a multivariate random walk with drift, which is an econometric model able to provide multivariate simultaneous forecasting.

This work is structured as follows. Data and healthy measures are depicted in Sect. 2. The methodology is described in Sect. 3. The numerical application is offered in Sect. 4. Food for thought and Conclusions follow.

## 2 Data and healthy measures

HLE is a measure used by the World Health Organization (WHO) for assessing the health of a population in a country. According to the definition of the Global Burden of Disease Study (GBD), HLE<sup>1</sup> is the number of years that a person at a given age can expect to live in good health, if the rates of all-cause mortality and all-cause disability would be constant into the future.

HLE is calculated by subtracting the Disability Adjusted Life Years (DALY) from life expectancy. DALYs are the sum of years of life lost (YLLs) and years lived with disability (YLDs). DALYs are the lost years of healthy life due to an early death caused by an illness or disability while still alive. YLLs are years lost due to premature mortality and are calculated by subtracting the age at death due to a given illness from the life expectancy at that age. Finally, YLDs are years lived with disability. This latter indicator is calculated as the product of the prevalence of a given condition that causes disability and the disability weight reflecting that condition's severity.

If the comorbidity is ignored, the sum of YLDs across all causes may overestimate the total loss of health, especially at older ages. For this reason, in 2010, the Global Burden of Diseases, Injuries, and Risk Factors Study 2010 (Horton 2012) implemented adjustments for comorbidity so that the sum of YLDs across causes was equal to the sum of the overall lost health at a given age.

At first, the GBD calculated the YLDs following an incidence perspective, so that the number of incident cases in a given period was multiplied by the average duration of the disease and a weight reflecting the severity of the disease on a scale from 0 (perfect health) to 1 (dead). This approach had several drawbacks. For example, data related to the average duration of the disease are often unavailable, and taking into account comorbidities is rather complicated than a prevalence approach. While the incidence rate is the ratio of new cases of a disease divided by the number of exposure to the risk in a specific population over a particular period, the prevalence rate is the ratio of the number of cases of a disease divided by the number of exposure to the risk in a specific population over a particular period.

<sup>1</sup> GBD usually refers to as Healthy Adjusted Life Expectancy (HALE).

**Table 1** List of the countries used in this study and related International Standards Organization (ISO) 3-digit alphabetic codes

Country	ISO3 code	Country	ISO3 code	Country	ISO3 code
Australia	AUS	France	FRA	Norway	NOR
Austria	AUT	Germany	DEU	Poland	POL
Belarus	BLR	Hungary	HUN	Portugal	PRT
Belgium	BEL	Iceland	ISL	Russian Federation	RUS
Bulgaria	BGR	Ireland	IRL	Slovakia	SVK
Canada	CAN	Italy	ITA	Spain	ESP
Czechia	CZE	Japan	JPN	Sweden	SWE
Denmark	DEN	Latvia	LVA	Switzerland	CHE
Estonia	EST	Lithuania	LTU	Ukraine	UKR
Finland	FIN	Netherlands	NLD	United Kingdom	GBR
				United States of America	USA

In 2010, GBD and WHO switched to a prevalence-based approach to the calculation of YLDs, estimating comorbidities under the assumption of independence within age-sex groups (see WHO (2020) for a detailed description of the methods and the weight's calculation).

In this paper, we refer to the HLE data of the Global Burden of Disease Study (Wang et al. 2020) over the period 1990-2019 for the countries listed in Table 1. They have been downloaded at Global Health Data Exchange, <http://ghdx.healthdata.org/gbd-results-tool> (Institute for Health Metrics and Evaluation 2022).

### 3 Methodology

In this section, we describe the model for the multivariate forecasting of the HLE based on a multi-country clustering. We explore the evolution of HLE over time in different countries to outline a comparative analysis. We firstly implement a functional data clustering of the HLE trends, which can be described as curves over time observed for each country. The functional approach allows the clustering of HLE trends and the identification of countries that are evolving according to similar patterns. Then, we provide the fundamentals of the multivariate random walk with drift that we use to simultaneously forecast the populations' HLE within each cluster.

#### 3.1 Functional HLE data clustering

Describing data in a functional curve consists in assuming the existence of a continuous function for which the observed data constitute a discretization. Jacques et al. (2014a) propose the first clustering procedure for multivariate functional time series to catch the similarities among curves. One of the methods proposed in the literature to perform the functional clustering is based on a two-step functional algorithm Abraham et al. (2003). In the first step, the functional curve is derived from discrete data, through the filtering step (James and Sugar 2003) aiming at approximating the curves by using a basis expansion of cubic B-splines functions. In particular, B-splines are widespread because they allow the analysis of the non-linear effects in the covariates and are locally susceptible to data (De

Boor 1978). In the second step, the selected clustering tool is implemented. We refer to the two-step functional algorithm. We consider the basis expansion of  $h^j(t)$ :

$$h^j(t) = \sum_{l=1}^L \beta_{jl} \cdot B_l(t) \quad (1)$$

where  $B_l(t)$  with  $L \geq 1$  ( $L \in \mathbb{N}$ ) is the selected number of basis functions,  $\beta_{jl} \in \mathbb{R}$  are the basis coefficients estimated from the observed data using the classical least square estimation. We observe that Eq. 1 provides a functional representation of the observed values of HLE.

Given the discrete observations  $h^{jk}$  of each sample path  $h^j(t)$  at a finite set of knots  $\{t_{jk} : k = 1, \dots, m_j\}$ , we reconstruct the HLE functional form through the functional predictor:

$$h^{jk} = h^j(t_{jk}) + \varepsilon_{jk} \quad k = 1, \dots, m_j$$

with  $\varepsilon_{jk}$  independent and identically zero mean distributed errors. The spline coefficients of each sample path  $h^j(t)$  are estimated by

$$\hat{\beta}_j = [(B^j)' B^j]^{-1} (B^j)' \hat{h}^j$$

with  $\hat{\beta}_j = (\hat{\beta}_{j1}, \dots, \hat{\beta}_{jL})'$ ,  $B^j = (B_l(t_{jk}))_{1 \leq k \leq m_j, 1 \leq l \leq L}$  and  $\hat{h}^j = (\hat{h}^{j1}, \dots, \hat{h}^{jm_j})'$ . The approximation of the functional curves of HLE for each population is performed through the *fda* package (version 5.1.9) of the R software using a cubic spline, where the number of knots is equal to the number of the available observations.

Once the functional form of each curve is reconstructed, we implement the k-means cluster to the fitted basis coefficients of all expanded curves. As it is well known, the k-means is an iterative clustering method, according to which a data point is assigned to a cluster minimizing the squared distance between each data point and the arithmetic mean of all data points in the cluster. We frame our choice within the existing literature, starting from the contributions of Singhal and Seborg (2005) and Ieva et al. (2011) that use a k-means algorithm on distances between multivariate functional data. Kayano et al. (2010) implement Self-Organizing Maps on the coefficients of orthonormalized Gaussian basis expansions of multivariate curves. Jacques et al. (2014a) consider the use of non orthonormal basis functions. In literature, other functional data clustering approaches have been proposed (James and Sugar 2003; Tarpey and Kinader 2003; Chiou et al. 2007; Bouveyron and Jacques 2011), like the hierarchical non-parametric clustering (Ferraty and Vieu 2006) or the model-based approach that assumes a given density probability on a finite number of parameters describing the curves (Bouveyron and Jacques 2011). In this paper, we have selected the two-step filtering procedure to describe the HLE multivariate time series, to show how similar dynamic behaviours emerge among countries, by performing clustering on the time series. On the other hand other types of probabilistic models could be useful to address the aforementioned problem. For instance, Mixture Models (see Consonni et al. 1995; Giudici 2003), are often used for data clustering being an efficient implementation framework by means of the Expectation–Maximization algorithm.

### 3.2 HLE multivariate modeling

Considering the groups of populations obtained through the functional clustering, we now work on the cluster-related HLE discrete observations.

Let  $C$  the number of clusters obtained from the previous step, and  $n_i$  the number of populations within the generic cluster  $i \in \{1, 2, \dots, C\}$ . Hence, the total number of populations is the sum of the populations belonging to each cluster,  $n = \sum_{i=1}^C n_i$ . We denote  $h_t^{(i,j)}$  the observed value of HLE at birth (for simplicity, age  $a$  has been omitted) for population  $j \in [1, \dots, n]$  in cluster  $i$  for year  $t \in [t_1, t_2, \dots, t_\tau]$ . For the generic cluster  $i$ , we define the matrix of HLE as:

$$H_t^{(i)} = \begin{bmatrix} h_{t_1}^{(i,1)} & h_{t_2}^{(i,1)} & \dots & h_{t_\tau}^{(i,1)} \\ h_{t_1}^{(i,2)} & h_{t_2}^{(i,2)} & \dots & h_{t_\tau}^{(i,2)} \\ \vdots & \vdots & \ddots & \vdots \\ h_{t_1}^{(i,n_i)} & h_{t_2}^{(i,n_i)} & \dots & h_{t_\tau}^{(i,n_i)} \end{bmatrix} \tag{2}$$

where the row indicates the population in the cluster  $i$  and the column indicates the time. Therefore,  $(H_t^{(i)})_{t \in \{t_1, \dots, t_\tau\}}$  is the multivariate time series of HLE at birth of the populations in cluster  $i$  that we simultaneously model by a multivariate random walk with drift (MRWD):

$$H_t^{(i)} = H_{t-1}^{(i)} + \delta^{(i)} + \epsilon_t^{(i)}, \quad \epsilon_t^{(i)} \sim N(\mathbf{0}, \Sigma^{(i)}) \tag{3}$$

where  $\delta^{(i)}$  is the vector of drift parameters driving the dynamics of the populations in cluster  $i$ , and  $\Sigma^{(i)}$  is the  $n_i \times n_i$  variance-covariance matrix of the multivariate white noise  $\epsilon_t^{(i)}$  in cluster  $i$ .

Using data in the in-sample period, we calculate one-step-ahead and two-step-ahead and five-step-ahead point forecasts. We follow an expanding window approach, increasing the in-sample period by one year (two or five years) and calculating one-step-ahead (two-step-ahead or five-step-ahead) forecasts. We determine the out-of-sample forecast accuracy by comparing the forecasts with the actual out-of-sample data. A similar approach has been followed by, e.g., Shang and Yang (2021) for jointly modeling Australian sub-populations using one-step-ahead and five-step-ahead forecast, and Shang and Hyndman (2017) that provided one- to ten-step-ahead forecasts for grouped functional time series using the regional age-specific mortality rates in Japan.

Considering  $t_\tau$  the forecast origin and  $r$  the forecast horizon, the  $r$ -step-ahead forecast of HLE of the populations in the generic cluster  $i$  is given by:

$$\widehat{H}_{t_\tau}^{(i)}(r) = E \left[ H_{t_\tau+r}^{(i)} \mid H_{t_\tau}^{(i)}, H_{t_\tau-1}^{(i)}, \dots, H_1^{(i)} \right] \tag{4}$$

The  $r$ -step-ahead forecast for MRWD in Eq. 3 can be written as:

$$\widehat{H}_{t_\tau}^{(i)}(r) = H_{t_\tau}^{(i)} + \delta^{(i)}r \tag{5}$$

### 3.3 Performance measures

We measure the performance of the model by the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE). Given  $t_\tau$  the final year of the forecast, we calculate MAE and RMSE for each population  $j$  within cluster  $i$  as:

$$MAE^{(i,j)} = \frac{\sum_{t=t_\tau+1}^{t_T} | h_t^{(i,j)} - \widehat{h}_t^{(i,j)} |}{(t_T - t_\tau - 1)}, \tag{6}$$

$$RMSE^{(i,j)} = \sqrt{\frac{\sum_{t=t_\tau+1}^{t_T} (h_t^{(i,j)} - \hat{h}_t^{(i,j)})^2}{(t_T - t_\tau - 1)}}. \quad (7)$$

And, their average values for each cluster  $i$  as:

$$MAE^{(i)} = \frac{1}{n_i} \sum_{j=1}^{n_i} MAE^{(i,j)}, \quad (8)$$

$$RMSE^{(i)} = \frac{1}{n_i} \sum_{j=1}^{n_i} RMSE^{(i,j)}. \quad (9)$$

Furthermore, to compare the predictive accuracy of two competing forecasts we use the Diebold Mariano (DM) test (see Diebold and Mariano 1995). In a nutshell, given the forecast error  $e_{it}$  defined as the difference between the values predicted by a certain model  $i$  and the actual values at time  $t$ , the loss associated with forecast from model  $i$  is assumed to be  $g(e_{it}) = e_{it}^2$ . The two forecasts have equal accuracy if and only if the loss differential between the two forecasts,  $d_t = g(e_{1t}) - g(e_{2t})$ , has zero expectation for all  $t$ . The DM test statistic is defined as follows:

$$DM = \frac{\bar{d}}{\sqrt{\frac{s}{N}}} \quad (10)$$

where  $\bar{d}$  is the sample mean of the loss differential,  $s$  is the variance, and  $N$  the sample size. The null hypothesis of this test is that the models have the same forecast accuracy, i.e.  $H_0 : E[d_t] = 0, \forall t$ , while the alternative hypothesis is that  $H_1 : E[d_t] \neq 0, \forall t$ . If  $H_0$  is true, then the DM statistic is asymptotically distributed as a normal standard normal distribution with 0 mean and standard deviation equal to 1.

## 4 Numerical application

We remind that the proposed methodology consists of two stages. The first one is the functional data clustering of the HLE at birth of the countries listed in Table 1. The second one consists of simultaneously forecast HLE at birth of the countries inside the cluster by a multivariate random walk with drift.

We apply this methodology to the data on HLE for the period 1990-2019, where the first 15 years, 1990-2009, represent the in-sample period, and years 2010-2019 the out-of-sample period.

The optimal number of clusters  $C$  is chosen according to the Elbow and Silhouette methods (see Figs. 14-16 in Appendix A). From Figs. 15-16, we have chosen  $C = 3$  as the optimal number of clusters for both genders.

The results of the application of the functional data clustering to the HLE data can be visualized in the world map in Fig. 1 for females and in Fig. 2 for males. Focusing on female populations, cluster 1 shows a strong geographical connotation as mainly collects former Soviet Union countries (Belarus, Latvia, Russia, and Ukraine). The other two

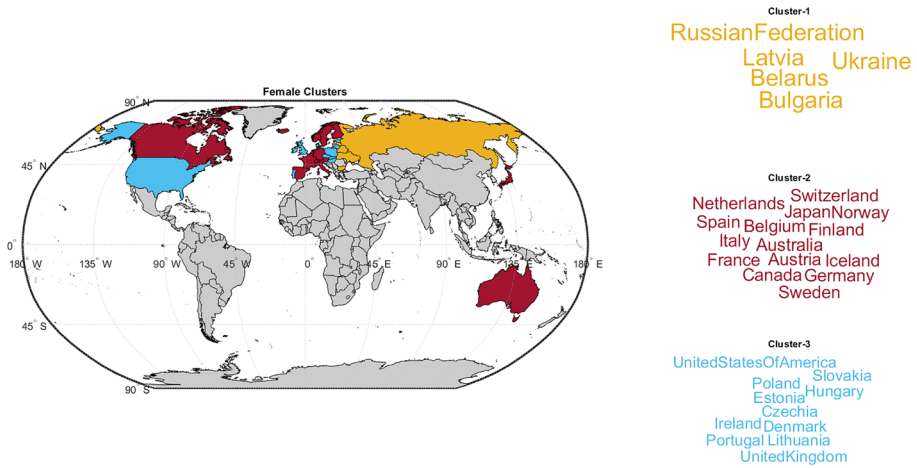


Fig. 1 Clustered populations. Females

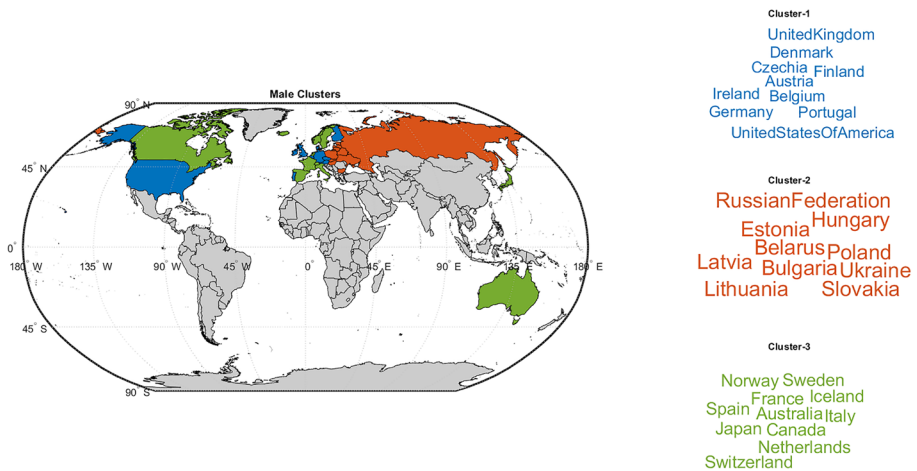


Fig. 2 Clustered populations. Males

clusters collect both European countries and North America, while Australia is in cluster 2. Looking at male populations, cluster 2 gathers the same countries as cluster 1 for females plus the remaining Baltic States (Estonia, Lithuania) and some Eastern Europe countries (Hungary, Poland, Slovakia). Male cluster 3 is similar to female cluster 2, collecting the same countries except for Austria, Belgium, Germany, and Finland.

To test the statistical significance of the drift, we used the Augmented Dickey-Fuller Test (ADF) Unit Root Test, analyzing whether the drift term is needed in the regression model. The null hypothesis means that our process is a random walk with drift. The results of the ADF test are reported in Table 2 and Table 3 for males and females, respectively. We can deduce that the values of test statistic are greater than critical values at levels 1%, 5%, and 10% for all the clusters and both genders with the exception of cluster 2 males and cluster 1 females. Therefore, we conclude that the null hypothesis (the process is a random walk



**Table 2** List of male populations by cluster number, results of the ADF test (null: random walk with drift), and drift parameter

Population	ISO3 code	Cluster	Test statistic	Drift value
Austria	AUT	1	74.9166	0.003011
Belgium	BEL	1	60.5093	0.002979
Czechia	CZE	1	32.5735	0.002932
Denmark	DEN	1	29.1165	0.003018
Finland	FIN	1	84.5433	0.003235
Germany	DEU	1	66.6388	0.003337
Ireland	IRL	1	27.8705	0.003347
Portugal	PRT	1	34.8772	0.003158
United Kingdom	GBR	1	75.0947	0.002847
United States of America	USA	1	18.8464	0.002514
Belarus	BLR	2	2.5781***	–
Bulgaria	BGR	2	1.2940***	–
Estonia	EST	2	0.8428***	–
Hungary	HUN	2	6.5325	–
Latvia	LVA	2	0.3435***	–
Lithuania	LTU	2	0.5894***	–
Poland	POL	2	9.2137	–
Russia	RUS	2	1.7521***	–
Slovakia	SVK	2	10.1347	–
Ukraine	UKR	2	2.3312***	–
Australia	AUS	3	76.2945	0.003232
Canada	CAN	3	95.0936	0.003108
France	FRA	3	55.8375	0.002990
Iceland	ISL	3	9.9812	0.002903
Italy	ITA	3	81.6075	0.002785
Japan	JPN	3	20.5208	0.002817
Netherlands	NLD	3	52.9665	0.002908
Norway	NOR	3	85.3316	0.003073
Spain	ESP	3	50.0970	0.003222
Sweden	SWE	3	114.2692	0.003406
Switzerland	CHE	3	53.1291	0.003570

Years 1990–2009. Critical values for test statistics: \*\*\* 7.88 for  $p < 0.01$ , \*\* 5.18 for  $p < 0.05$ , \* 4.12 for  $p < 0.1$ .

with drift) can be accepted in all the cases except for cluster 2 males and cluster 1 females. Tables 2–3 also show the values of the drift parameters for the populations whose HLE time series follows a random walk with drift, estimated over the in-sample period 1990–2009. We observe positive values of  $\delta$  for all the countries. The  $\delta$  values for male populations are higher than those for female populations highlighting an improvement in the men’s health in the period 1990–2009. Indeed, men live less than women but spend a lower proportion of their total life expectancy in poor health. This phenomenon is known as the “male-female health-survival paradox” (Oksuzyan et al. 2009), and is not constant and universal, depending on different health domain indicators as argued by Di Lego et al. (2020).

**Table 3** List of female populations by cluster number, results of the ADF test (null: random walk with drift), and drift parameter

Population	ISO3 code	Cluster	Test statistic	Drift value
Belarus	BLR	1	0.1716***	–
Bulgaria	BGR	1	2.5783***	–
Latvia	LVA	1	0.8040***	–
Russia	RUS	1	0.8103***	–
Ukraine	UKR	1	0.2365***	–
Australia	AUS	2	49.2964	0.001927
Austria	AUT	2	49.7918	0.001921
Belgium	BEL	2	25.1095	0.001930
Canada	CAN	2	64.1297	0.001944
Finland	FIN	2	28.2506	0.001949
France	FRA	2	25.3753	0.002047
Germany	DEU	2	59.5543	0.002176
Iceland	ISL	2	24.8225	0.002225
Italy	ITA	2	76.4003	0.002204
Japan	JPN	2	57.9446	0.002173
Netherlands	NLD	2	20.7369	0.002105
Norway	NOR	2	27.6603	0.002016
Spain	ESP	2	72.4959	0.001929
Sweden	SWE	2	41.4797	0.001851
Switzerland	CHE	2	31.0529	0.001779
Czechia	CZE	3	13.1897	0.002860
Denmark	DEN	3	17.7256	0.002881
Estonia	EST	3	2.0890***	0.002861
Hungary	HUN	3	13.2208	0.002864
Ireland	IRL	3	38.4853	0.002704
Lithuania	LTU	3	0.6153***	0.002703
Poland	POL	3	26.8252	0.002680
Portugal	PRT	3	36.0128	0.002688
Slovakia	SVK	3	7.3036	0.002277
United Kingdom	GBR	3	52.0672	0.001763
United States of America	USA	3	6.7399	0.001223

Years 1990–2009. Critical values for test statistics: \*\*\* 7.88 for  $p < 0.01$ , \*\* 5.18 for  $p < 0.05$ , \* 4.12 for  $p < 0.1$

Finally, we apply the ADF test for testing whether HLE time series is a random walk (without drift) for the populations included in cluster 2 males and cluster 1 females. The results, provided in Tables 4 and 5, show that for both the clusters above mentioned a random walk is appropriate.

In conclusion, we model all the clusters with a MRWD, with the exception of cluster 2 males and cluster 1 females that are modeled as a MRW.

The results of the out-of-sample test over the years 2010–2019 for each female population cluster are depicted in Fig. 3, 4 and 5 for cluster 1, 2, and 3 respectively.

**Table 4** List of male populations by cluster number, results of the ADF test (null: random walk)

Population	ISO3 code	Cluster	Test statistic
Belarus	BLR	2	-0.4945
Bulgaria	BGR	2	1.5253
Estonia	EST	2	1.3047
Hungary	HUN	2	3.6627
Latvia	LVA	2	0.7710
Lithuania	LTU	2	0.4015
Poland	POL	2	4.2912
Russia	RUS	2	-0.1592
Slovakia	SVK	2	4.2340
Ukraine	UKR	2	-0.0744

Years 1990–2009. Critical values for test statistics: \*\*\* -2.66 for  $p < 0.01$ , \*\* -1.95 for  $p < 0.05$ , \* -1.6 for  $p < 0.1$ .

**Table 5** List of female populations by cluster number, results of the ADF test (null: random walk). Years 1990–2009. Critical values for test statistics: \*\*\* -2.66 for  $p < 0.01$ , \*\* -1.95 for  $p < 0.05$ , \* -1.6 for  $p < 0.1$ 

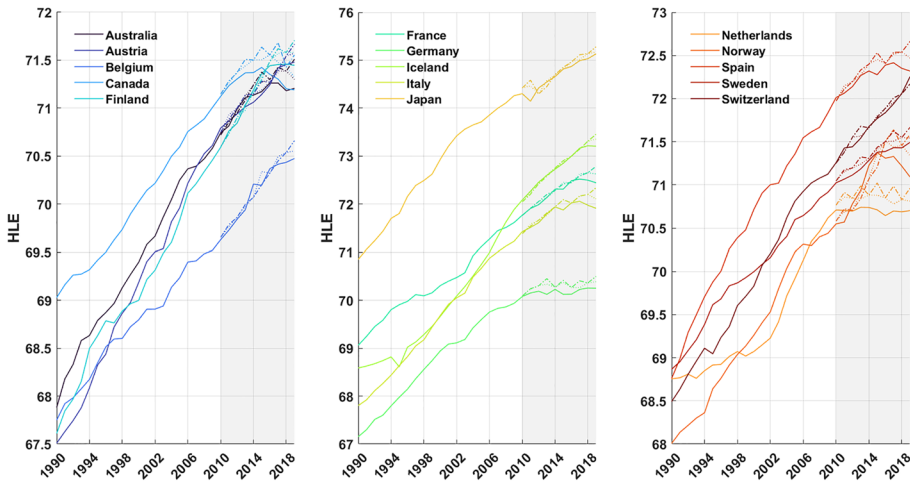
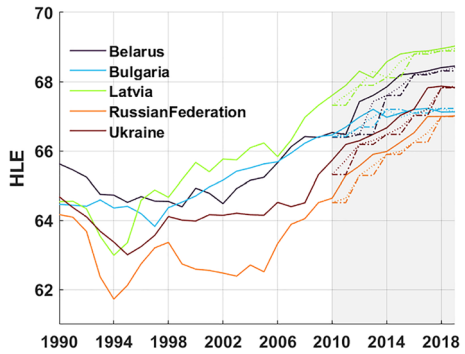
Population	ISO3 code	Cluster	Test statistic
Belarus	BLR	1	0.5952
Bulgaria	BGR	1	2.0904
Latvia	LVA	1	1.3048
Russia	RUS	1	0.1463
Ukraine	UKR	1	0.4444

The results of the out-of-sample test for each male population cluster are depicted in Fig. 6, 7 and 8 for cluster 1, 2, and 3 respectively.

We can trace some reflections starting from the analysis of Figs. 3–8. We choose to provide specific qualitative comments for U.S. and Russia, which show some interesting health and longevity dynamics supported by the literature. The health disadvantage of the United States (National Research Council 2013) is clear for both males and females: HLE in the U.S. is significantly lower than that of other countries. In particular, from the WHO data, it emerges that the U.S. is the only developed country in the world that has experienced a decrease in healthy life expectancy since 2010, despite it having the highest GDP in the world and is one of the wealthiest countries in per capita GDP. The decreasing trend in HLE occurs parallel to the decreasing trend in life expectancy. As early as the 1990s, a slowdown in improving longevity was observed and the life expectancy of Americans fell below the average for developed countries. This trend continued until around 2010 when the average death age in the U.S. stabilized. In 2011, the U.S. recorded the highest health expenditure as a proportion of GDP but the lowest coverage (Lorenzoni et al. 2014). This is mainly due to the fragility of a health system strongly biased in favor of private assistance and to the inequalities that it generates in health care for the poorest sections of the population. In the literature, it has been asked whether the deficiency in the U.S. health system is the sole cause of health disadvantage. McGinnis and Foege (1993) emphasized the important role of health behaviors. Mokdad et al. (2005) highlighted that 40 percent of all deaths in the U.S. are associated with four poor health behaviors: tobacco use, unhealthy diet, physical inactivity, and drinking problems.

Another evidence that emerges from Figs. 3–8 is the particular volatility that characterized the trend in health expectancy in the Russian Federation and Eastern European countries. Russia is another exception to the well-known Preston curve relationship. After an

**Fig. 3** Out-of-sample results for **cluster 1**, females. Observed values (solid line), one-step-ahead forecast (dotted line), and two-step-ahead (dash-dotted line)



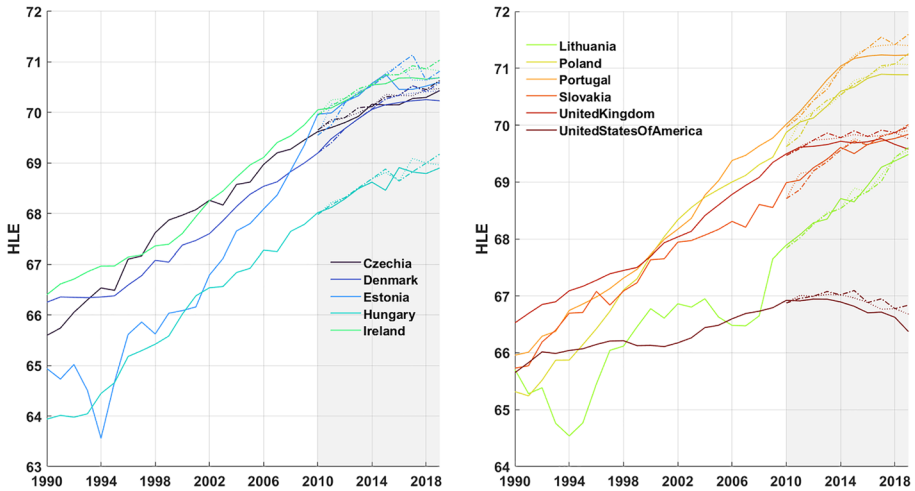
**Fig. 4** Out-of-sample results for **cluster 2**, females. Observed values (solid line), one-step-ahead forecast (dotted line), and two-step-ahead (dash-dotted line)

improvement in health conditions in 1985-87 due to Gorbachev’s anti-alcohol campaign, the 1990s saw a return to easy access to alcohol and economic tensions with the end of the Soviet Union. Unlike what happened in the U.S., however, a growing trend has been consolidated in these countries since 2010. Despite this, the great volatility recorded in the previous period actually has repercussions on higher forecast errors for these countries for both genders as shown in Figs. 9–10, which illustrate the values of the error measures on the single population for females and males, respectively.

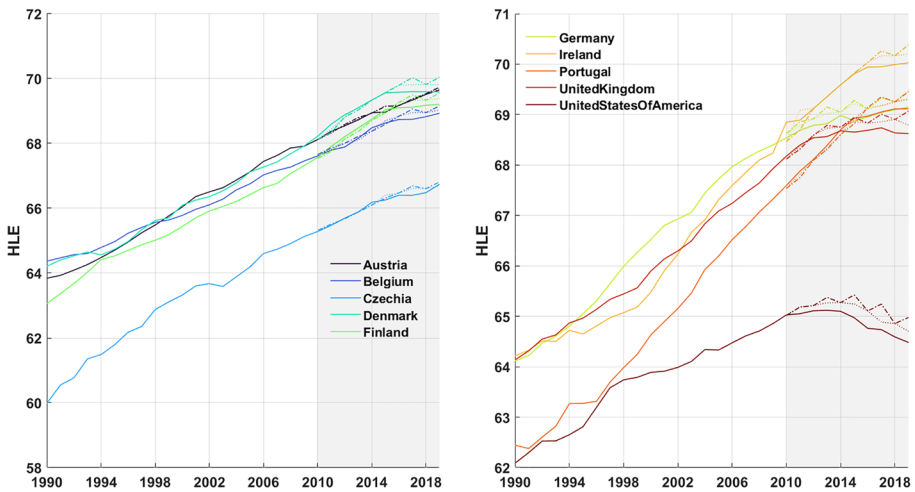
In Table 6, we show the average RMSE and MAE values, by cluster, of both the one-step-ahead and two-step-ahead forecasts. On average one-step-ahead forecast outperforms the two-step-ahead prediction in all the clusters (both for males and females).

Figures 11 and 12 displays the boxplots of the distribution of the performance measures for females and males, respectively.

The two-step-ahead prediction interquartile range is larger than one of the one-step-ahead forecasts, consistent with the errors reported in Table 6. This is especially true for the male cluster 3, while for the other two clusters this quantity is comparable across the two forecasting horizons. For females, the cluster 3 interquartile range is the lowest for both forecasting



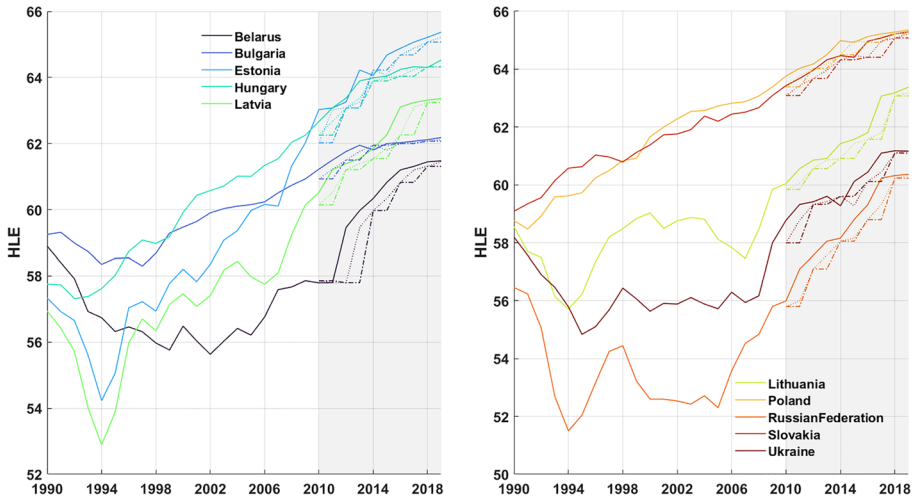
**Fig. 5** Out-of-sample results for **cluster 3**, females. Observed values (solid line), one-step-ahead forecast (dotted line), and two-step-ahead (dash-dotted line)



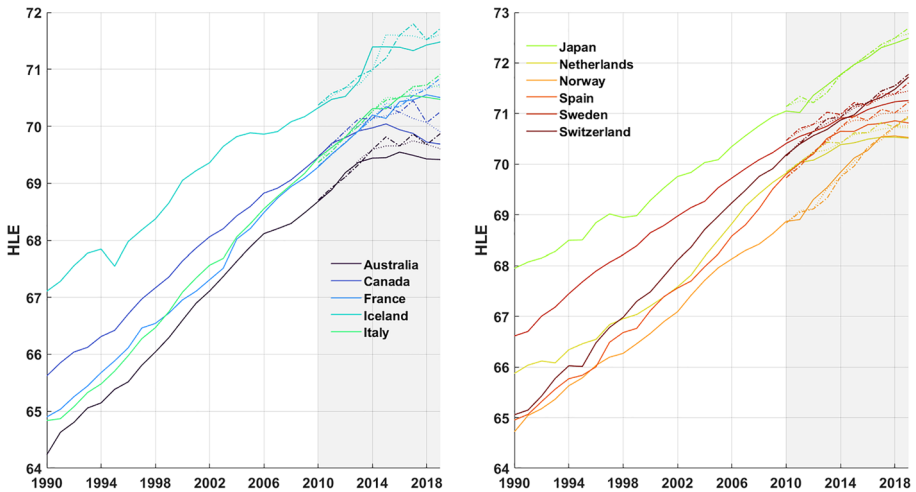
**Fig. 6** Out-of-sample results for **cluster 1**, males. Observed values (solid line), one-step-ahead forecast (dotted line), and two-step-ahead (dash-dotted line)

horizons. On average in all box plots, we observe larger upper-whiskers compared with the lower ones suggesting higher asymmetries for larger values of the forecast distributions.

We apply the DM test for comparing the accuracy of forecast performance between 1-step ahead, 2-step ahead and 5-step ahead forecasts. The results of the DM test are reported in Table 7. When the null hypothesis is accepted at the certain level of significance, the differences between the two models compared are significant and the forecasting accuracy of first model is higher than that of the second model. Therefore, from Table 7, we observe that the accuracy of 1-step-ahead forecasts is always better than the accuracy of the other two models, and that 2-step-ahead forecasts are more accurate than 5-step-ahead ones.



**Fig. 7** Out-of-sample results for **cluster 2**, males. Observed values (solid line), one-step-ahead forecast (dotted line), and two-step-ahead (dash-dotted line)



**Fig. 8** Out-of-sample results for **cluster 3**, males. Observed values (solid line), one-step-ahead forecast (dotted line), and two-step-ahead (dash-dotted line)

### 5 Discussion

While life expectancy is an essential indicator for pension and health care systems to plan health spending to manage disease events that occur as the population ages, HLE is also an important measure to plan health care costs to prevent the onset of diseases. If the first indicator necessarily leads to considerations of coexistence between the lengthening

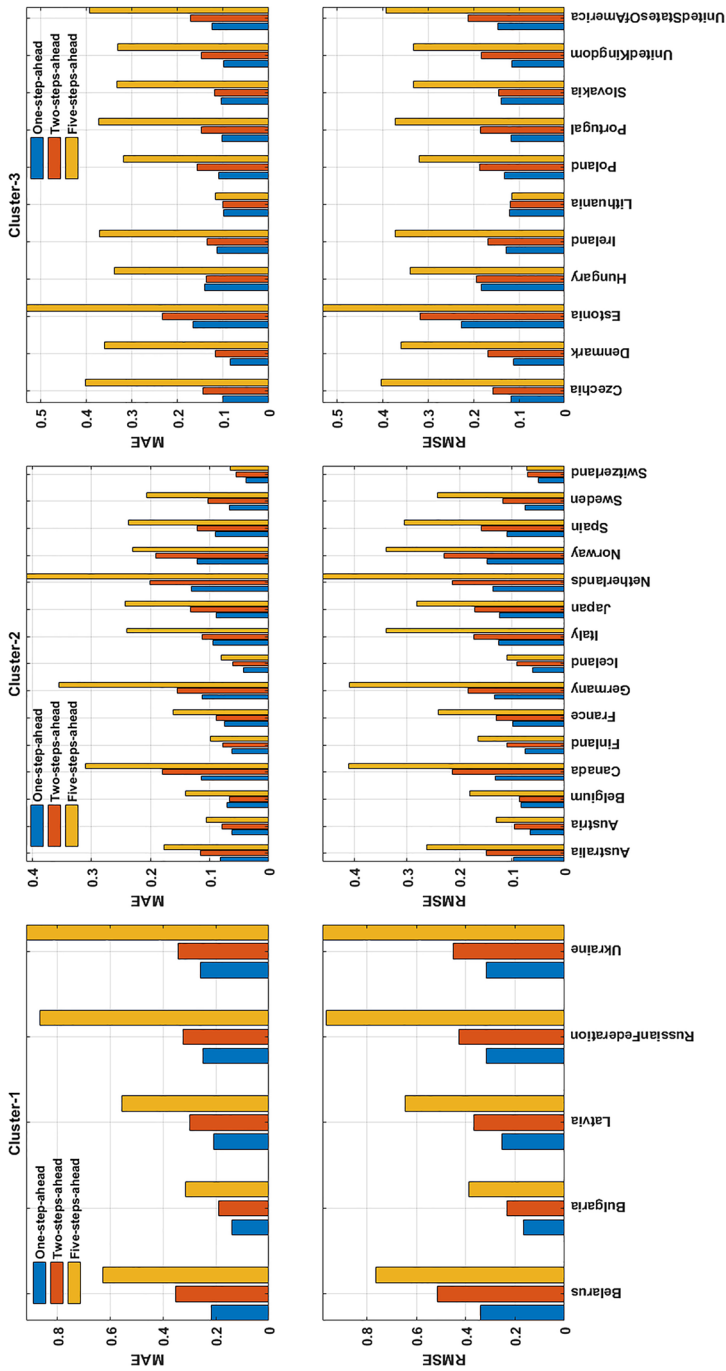


Fig. 9 Error measures, Females

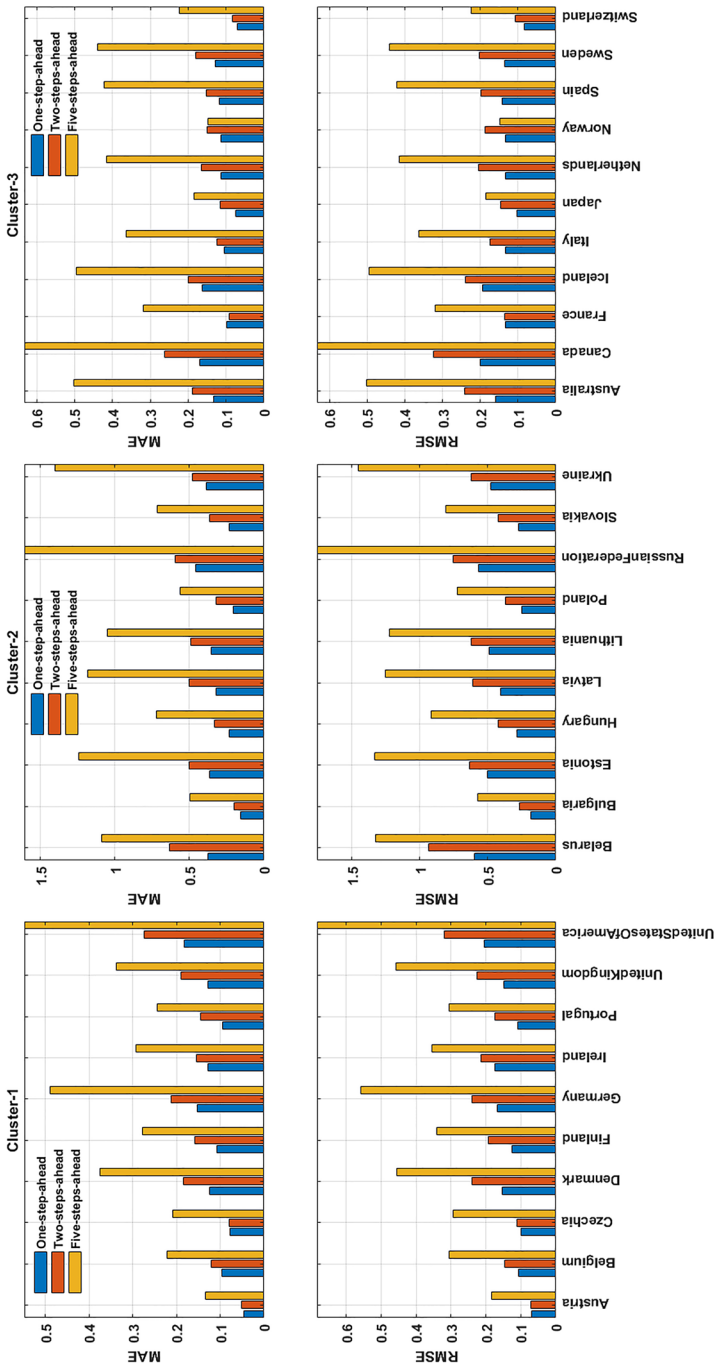


Fig. 10 Error measures. Males



**Table 6** Average RMSE and MAE values by cluster, years 2010–2019. One-step-ahead and two-step-ahead forecasts

Cluster	Model	Males		Females	
		RMSE	MAE	RMSE	MAE
1	One-step-ahead	0.13	0.11	0.35	0.24
	Two-step-ahead	0.20	0.27	0.42	0.33
	Five-step-ahead	0.29	0.34	0.77	0.62
2	One-step-ahead	0.47	0.25	0.09	0.08
	Two-step-ahead	0.57	0.48	0.14	0.11
	Five-step-ahead	1.40	1.17	0.20	0.25
3	One-step-ahead	0.13	0.11	0.12	0.10
	Two-step-ahead	0.19	0.26	0.18	0.14
	Five-step-ahead	0.29	0.41	0.27	0.35

of average life and morbidity, the second focuses on the years of life lived in full health. A choice of economic and social planning is thus posed: there is a trade-off between the expenditure destined to improve healthy life expectancy and that necessary to manage morbidity in senescence. Correct management of the former can have reductive effects on the latter.

In this work, we have focused on healthy life expectancy and have proposed a combined model based on functional clustering and multivariate time series forecasting to offer an overall picture of the evolution of life expectancy on a global level and coherently depict the country-specific life expectancy based on the cluster mortality profiles analyzing its possible future evolution. As regards the cluster methodology, numerous insights can be gathered from recent literature contributions. Since Mixture Model-based clustering, usually applied to multidimensional data, has become a popular approach in many data analysis problems, both for its good statistical properties and for the simplicity of implementation of the Expectation–Maximization algorithm, our future research will involve the development of a framework for dealing with mortality data time series which combines function data analysis and mixture models, where each cluster will be described by a regression in which the polynomial coefficients vary according to a discrete hidden process. Regarding the modeling and forecasting, as in the classical longevity approach, we have followed the assumption according to which the extrapolation of past trends of healthy life expectancy can provide accurate forecasting. Despite the simplicity of the forecasting model, the performance measures showed satisfactory accuracy results. Future proposals may refer to model flexibility, assuming forecasting also conditionally on a cluster latent variable. This would lead to a mixture of a random walk-with-drift model, thus embedding the two steps in a one-stage estimation as in Fraley and Raftery (2002). Moreover, the valuable contribution of this work can be framed in the 2030 Agenda for the achievement of Sustainable Goal n. 3 “Good Health and Well Being”, under which countries are required to obtain universal health coverage, access to essential health services of quality, and accessible to all to pursue the improvement of healthy living conditions globally. Starting from the updated data that will be available in the next future, further research will be able to monitor the degree of satisfaction of the Sustainable Goal leveraging the analysis outlined in this work on the healthy life expectancy.

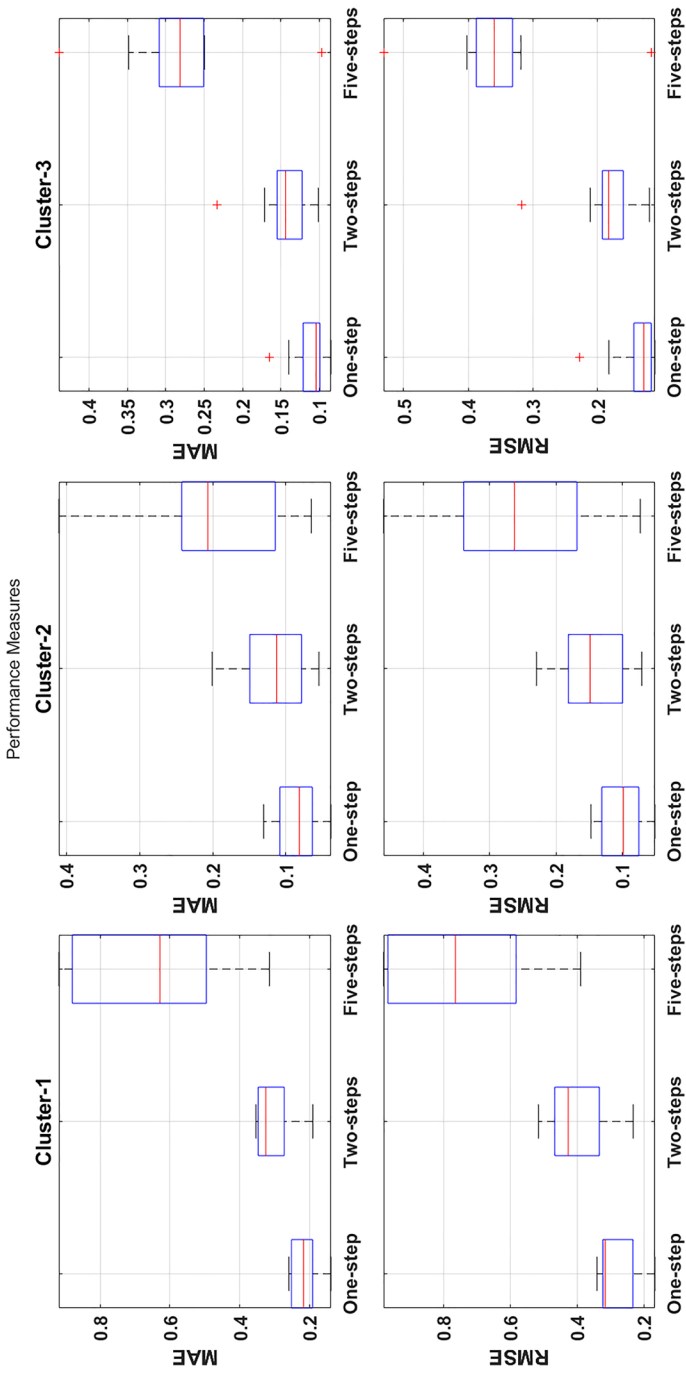


Fig. 11 Boxplots of the point forecast performance measures, Females

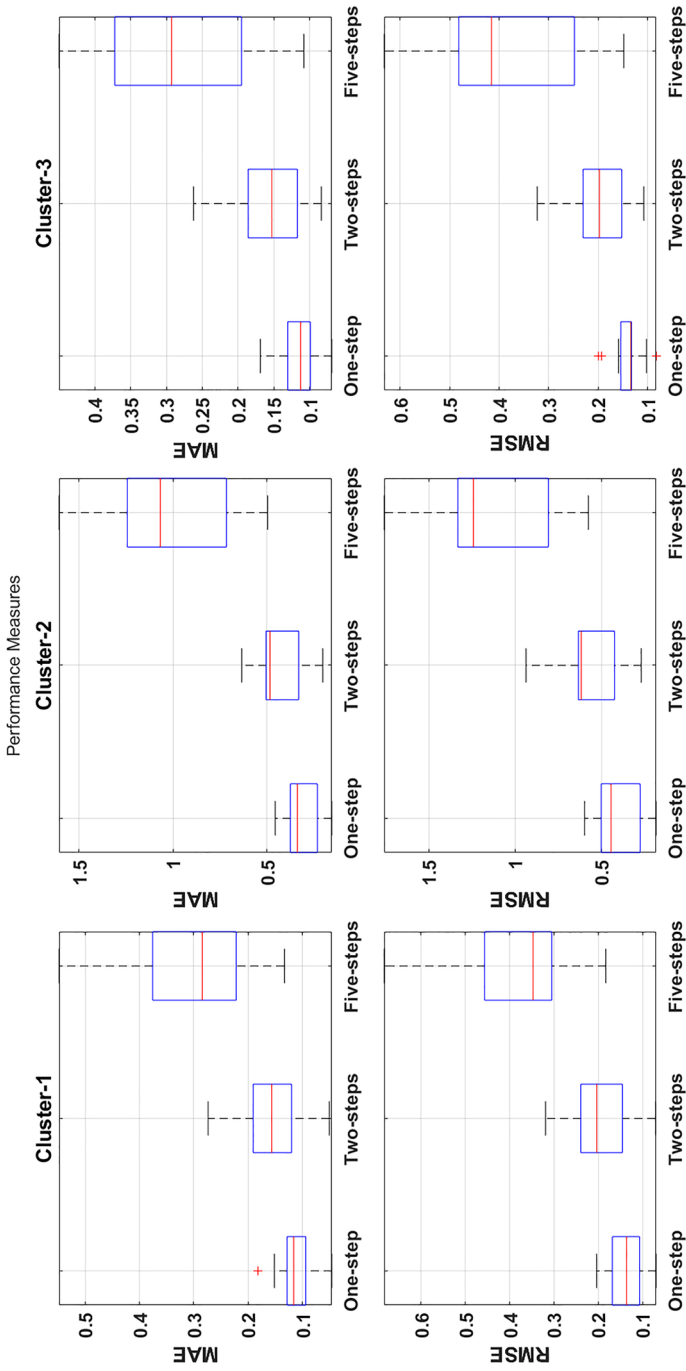


Fig. 12 Boxplots of the point forecast performance measures. Males

**Table 7** DM test by cluster, gender, and the forecast horizon *h*

Cluster	r-step-ahead models' comparison	Males			Females		
		<i>h</i> = 1	<i>h</i> = 2	<i>h</i> = 5	<i>h</i> = 1	<i>h</i> = 2	<i>h</i> = 5
1	1-step vs. 2-step	-4.75***	-3.36***	-	-2.37**	-2.27**	-
	1-step vs. 5-step	-5.52***	-	-2.41***	-4.39***	-	-2.59***
	2-step vs. 5-step	-	-3.67***	-2.29**	-	-2.69***	-2.28**
2	1-step vs. 2-step	-2.06**	-2.03**	-	-5.21***	-3.89***	-
	1-step vs. 5-step	-6.03***	-	-3.84***	-6.51***	-	-3.08***
	2-step vs. 5-step	-	-4.86***	-3.47***	-	-4.76***	-2.97***
3	1-step vs. 2-step	-4.41***	-3.37***	-	-3.30***	-2.96***	-
	1-step vs. 5-step	-5.11***	-	-2.34**	-6.16***	-	-2.49***
	2-step vs. 5-step	-	-3.37***	-2.22**	-	-3.78***	-2.38***

Evaluation period 2010–2019. Asterisks \*\*\* and \*\* indicate two-side significance at 1% and 5% level, respectively

### 6 Conclusions

Although the demographic literature is rich in contributions on the analysis and forecasting of life expectancy, only recently has attention been paid to the HLE indicator, focusing on the descriptive analysis of the variable rather than on the forecast. The proposed work tries to help bridge this gap by extending the predictive analysis in a multi-population perspective to obtain more accurate information by exploiting the similarities between countries that have shown similar trends. Our paper thus opens the way to a line of research with interesting practical implications, which can act as a guide both for governments, grappling with public spending, and for private operators, such as insurers and pension funds, which aim to grasp the challenges of the market of health and life insurance policies.

### Appendix A: Optimal number of clusters

See Figs. 13, 14, 15 and 16.

**Fig. 13** Silhouette Females

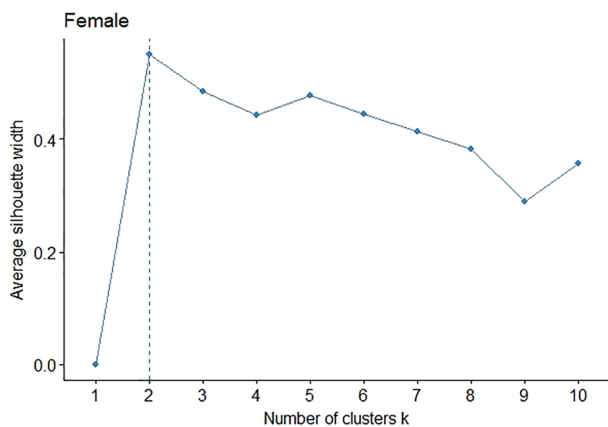


Fig. 14 Silhouette Males

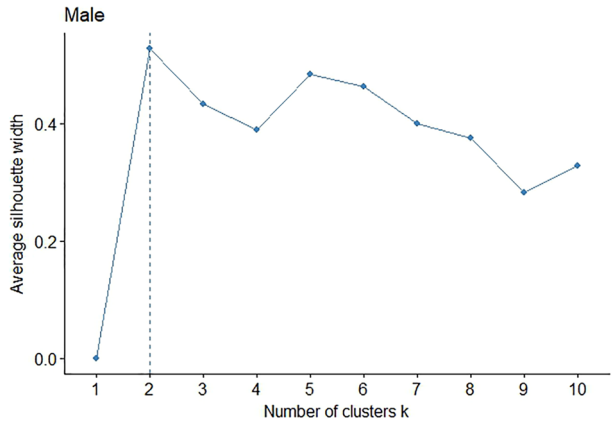


Fig. 15 Elbow Female

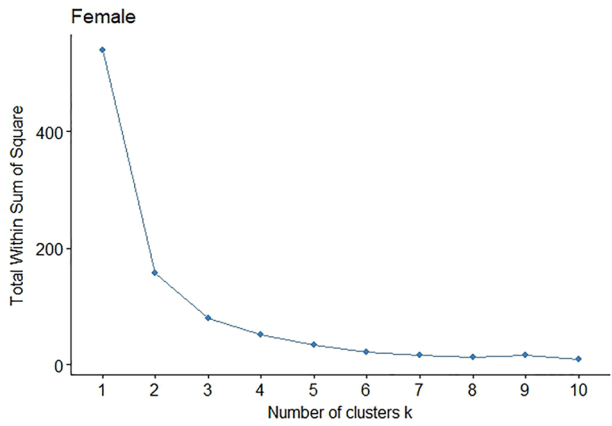
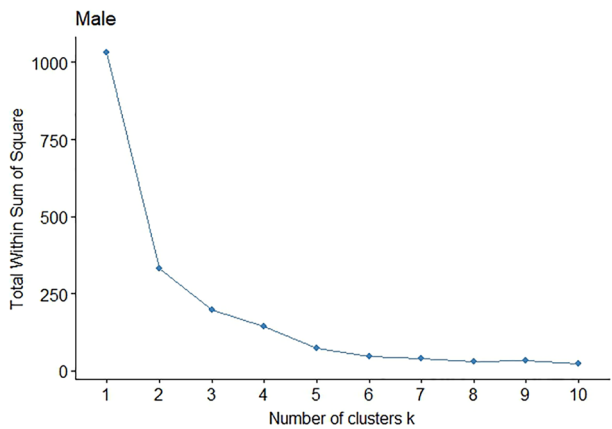


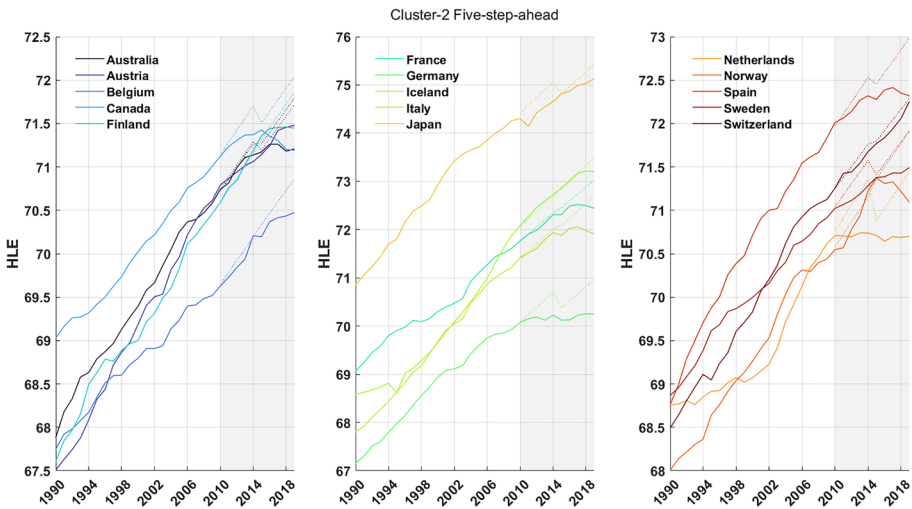
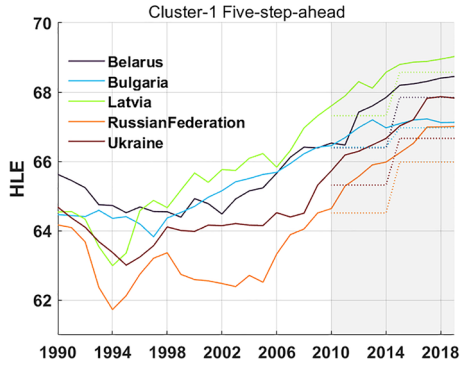
Fig. 16 Elbow Males



## Appendix B: Five-step-ahead forecast

See Figs. 17, 18, 19, 20, 21 and 22.

**Fig. 17** Out-of-sample results for **cluster 1**, females. Observed values (solid line), five-step-ahead forecast (dotted line)



**Fig. 18** Out-of-sample results for **cluster 2**, females. Observed values (solid line), five-step-ahead forecast (dotted line)

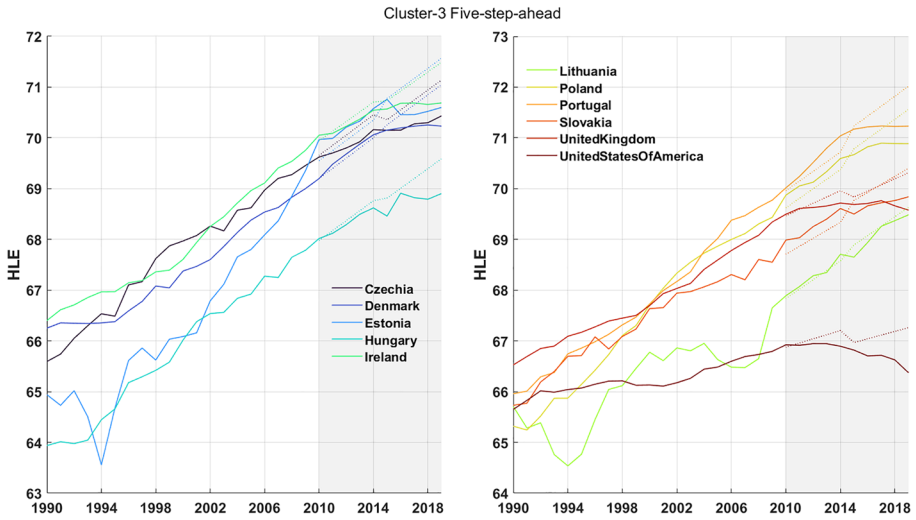


Fig. 19 Out-of-sample results for cluster 3, females. Observed values (solid line), five-step-ahead forecast (dotted line)

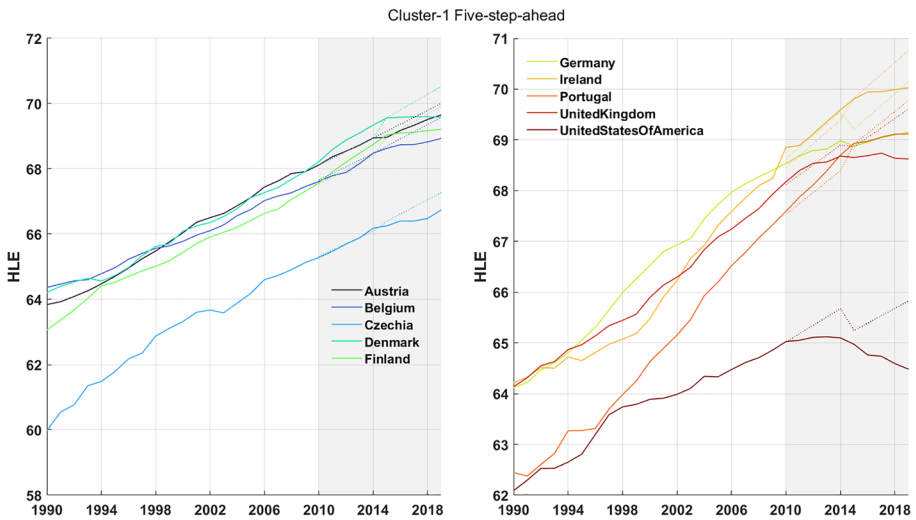
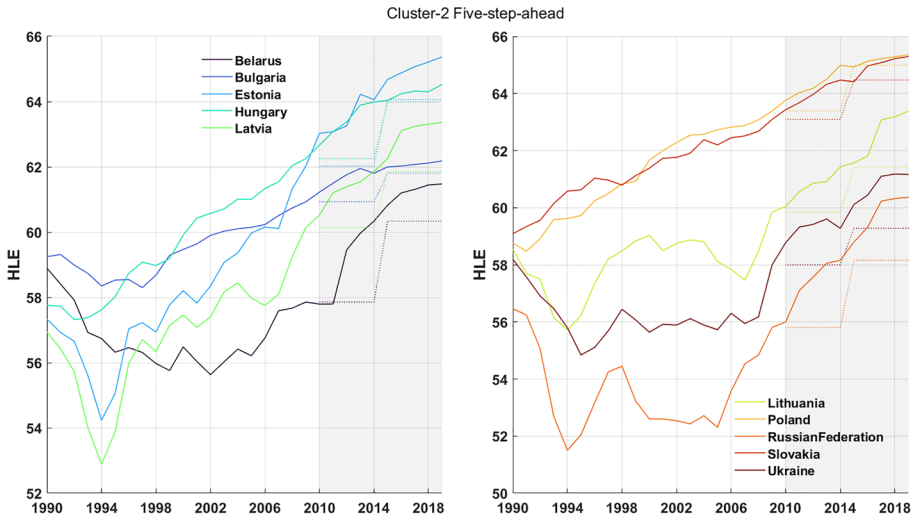
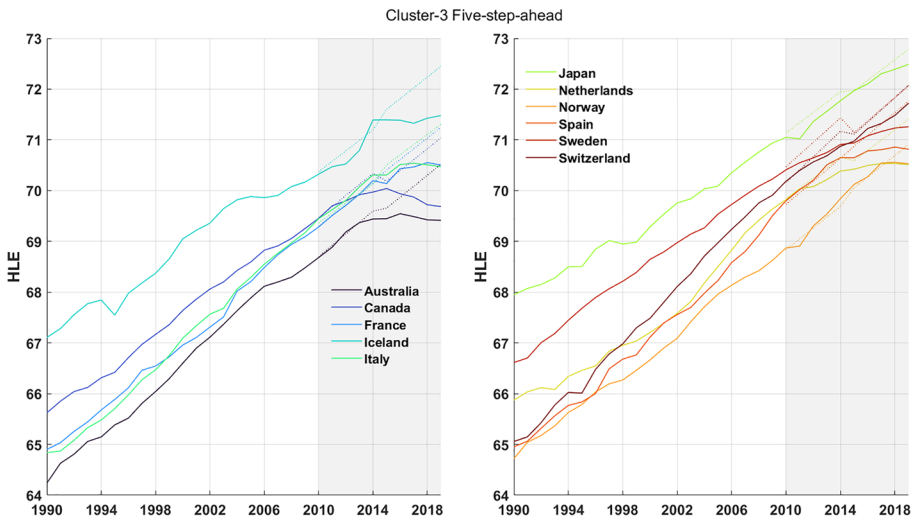


Fig. 20 Out-of-sample results for cluster 1, males. Observed values (solid line), five-step-ahead forecast (dotted line)



**Fig. 21** Out-of-sample results for **cluster 2**, males. Observed values (solid line), five-step-ahead forecast (dotted line)



**Fig. 22** Out-of-sample results for **cluster 3**, males. Observed values (solid line), five-step-ahead forecast (dotted line)

**Acknowledgements** The authors gratefully acknowledge the European Union’s Horizon 2020 research and innovation program “PERISCOPE: Pan European Response to the ImpactS of COVID-19 and future Pandemics and Epidemics”, under the grant agreement No. 101016233, H2020-SC1-PHE-CORONAVIRUS-2020-2-RTD.

**Funding** The author Spelta Alessandro gratefully acknowledges the European Union’s Horizon 2020 research and innovation program “PERISCOPE: Pan European Response to the ImpactS of COVID-19 and future Pandemics and Epidemics”, under the grant agreement No. 101016233,



H2020-SC1-PHE-CORONAVIRUS-2020-2-RTD. The other authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

## Declarations

**Conflict of interest** The authors have no relevant financial or non-financial interests to disclose.

## References

- Abraham, C., Cornillon, P.A., Matzner-Loeber, E., Molinari, N.: Unsupervised curve clustering using B-splines. *Scand. J. Stat.* **30**(3), 581–595 (2003). <https://doi.org/10.1111/1467-9469.00350>
- Bouveyron, C., Jacques, J.: Model-based clustering of time series in group-specific functional subspaces. *Adv. Data Anal. Classif.* **5**(4), 281–300 (2011). <https://doi.org/10.1007/s11634-011-0095-6>
- Chiou, J.M., Li, P.L.: Functional clustering and identifying substructures of longitudinal data. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **69**(4), 679–699 (2007). <https://doi.org/10.1111/j.1467-9868.2007.00605.x>
- Consonni, G., Veronese, P.: A Bayesian method for combining results from several binomial experiments. *J. Am. Stat. Assoc.* **90**(431), 935–944 (1995)
- De Boor, C.: *A Practical Guide to Splines*. Springer-Verlag, New York (1978)
- Diebold, F.X., Mariano, R.S.: Comparing Predictive Accuracy. *J. Bus. Econ. Stat.* **13**, 253–263 (1995)
- Di Lego, V., Di Giulio, P., Luy, M.: Gender Differences in Healthy and Unhealthy Life Expectancy. In: Jagger, C., Crimmins, E., Saito, Y., De Carvalho Yokota, R., Van Oyen, H., Robine, J.M. (eds) *International Handbook of Health Expectancies*. International Handbooks of Population, vol 9. Springer, Cham (2020)
- Ferraty, F., Vieu, P.: *Nonparametric Functional Data Analysis*. Springer Series in Statistics, Springer Verlag, New York (2006). <https://doi.org/10.1007/0-387-36620-2>
- Fraley, C., Raftery, A.E.: Model-based clustering, discriminant analysis, and density estimation. *J. Am. Stat. Assoc.* **97**(458), 611–631 (2002)
- Fries, J.F.: Aging, natural death, and the compression of morbidity. *N. Engl. J. Med.* **303**, 130–135 (1980)
- Giudici, P., Mezzetti, M., Muliere, P.: Mixtures of products of Dirichlet processes for variable selection in survival analysis. *J. Stat. Plan. Inference* **111**(1–2), 101–115 (2003)
- Global Burden of Disease Collaborative Network: *Global Burden of Disease Study 2019 (GBD 2019) Results*. Seattle, United States: Institute for Health Metrics and Evaluation (IHME) (2020). Available from <http://ghdx.healthdata.org/gbd-results-tool>
- Hebrail, G., Huguency, B., Lechevallier, Y., Rossi, F.: Exploratory analysis of functional data via clustering and optimal segmentation. *Neurocomput./EEG Neurocomput.* **73**(7–9), 1125–1141 (2010). <https://doi.org/10.1016/j.neucom.2009.11.022>
- Ho, J.Y., Hendi, A.S.: Recent trends in life expectancy across high income countries: retrospective observational study. *BMJ* **362**, k2562 (2018)
- Horton, R.: GBD 2010: understanding disease, injury, and risk. *Lancet* (2012). [https://doi.org/10.1016/S0140-6736\(12\)62133-3](https://doi.org/10.1016/S0140-6736(12)62133-3)
- Ieva, F., Paganoni, A., Pigoli, D., Vitelli, V.: ECG signal reconstruction, landmark registration and functional classification. In: 7th Conference on Statistical Computation and Complex System (2011)
- Imai, K., Soneji, S.: On the estimation of disability-free life expectancy: Sullivan' method and its extension. *J. Am. Stat. Assoc.* **102**(480), 1199–1211 (2007). <https://doi.org/10.1198/016214507000000040>
- Institute for Health Metrics and Evaluation: *GBD Results Tool*. Global Health Data Exchange. <http://ghdx.healthdata.org/gbd-results-tool>. Accessed 3 January 2022
- Jacques, J., Preda, C.: Model-based clustering for multivariate functional data. *Comput. Stat. Data Anal.* **71**, 92–106 (2014)
- Jacques, J., Preda, C.: Functional data clustering: a survey. *Adv. Data Anal. Classif.* **8**(3), 231–255 (2014). <https://doi.org/10.1007/s11634-013-0158-yf>
- James, G., Sugar, C.: Clustering for sparsely sampled functional data. *J. Am. Stat. Assoc.* **98**(462), 397–408 (2003). <https://doi.org/10.1198/016214503000189>
- Kayano, M., Dozono, K., Konishi, S.: Functional cluster analysis via orthonormalized gaussian basis expansions and its application. *J. Classif.* **27**, 211–230 (2010)

- Levantesi, S., Nigri, A., Piscopo, G.: Clustering-based simultaneous forecasting of life expectancy time series through Long-Short Term Memory Neural Networks. *Int. J. Approx. Reason.* **140**, 282–297 (2022)
- Lorenzoni, L., Belloni, A., Sassi, F.: Health-care expenditure and health policy in the USA versus other high-spending OECD countries. *Lancet* **384**(9937), 83–92 (2014)
- Manton, K.G.: Changing concepts of morbidity and mortality in the elderly population. *The Milbank Memorial Fund Quarterly. Health Soc.* **60**(2), 183–244 (1982)
- McGinnis, J.M., Foege, W.H.: Actual causes of death in the United States. *J. Am. Med. Assoc.* **270**(18), 2207–2212 (1993)
- Mokdad, A.H., Marks, J.S., Stroup, D.F., Gerberding, J.L.: Correction: Actual causes of death in the United States, 2000. *J. Am. Med. Assoc.* **293**, 293–294 (2005)
- National Research Council (US); Institute of Medicine (US); Woolf SH, Aron L, editors. *U.S. Health in International Perspective: Shorter Lives, Poorer Health*. Washington (DC): National Academies Press (US) (2013). 3, Framing the Question. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK154478/>
- Nigri, A., Levantesi, S., Marino, M.: Life expectancy and lifespan disparity forecasting: a long short-term memory approach. *Scand. Actuar. J.* **2**, 110–133 (2021). <https://doi.org/10.1080/03461238.2020.1814855>
- Nigri, A., Barbi, E., Levantesi, S.: The relationship between longevity and lifespan variation. *Stat. Methods Appl.* (2021b). <https://doi.org/10.1007/s10260-021-00584-4>
- Nigri, A., Barbi, E., Levantesi, S.: The relay for human longevity: country-specific contributions to the increase of the best-practice life expectancy. *Qual. Quant.* (2022)
- Oksuzyan, A., Petersen, I., Stovring, H., et al.: The male-female health-survival paradox: a survey and register study of the impact of sex-specific selection and information Bias. *Ann. Epidemiol.* **19**, 504–511 (2009)
- Permanyer, I., Spijker, J., Blanes, A.: On the measurement of healthy lifespan inequality. *Popul. Health Metrics* **20**, 1 (2022)
- Permanyer, I., Trias-Llimós, S., Spijker, J.: Best-practice healthy life expectancy vs. life expectancy: Catching up or lagging behind? *Pnas*, 118(46), (2021)
- Raftery, A., Chunn, J., Gerland, P., Ševčíková, H.: Bayesian probabilistic projections of life expectancy for all countries. *Demography* **50**(3), 777–801 (2013). <https://doi.org/10.1007/s13524-012-0193-x>
- Ramsay, J.O., Silverman, B.W.: *Functional data analysis*, 2nd edn. Springer Series in Statistics. Springer, New York (2005)978-0-387-22751-1
- Ramsey, S.A., Klemm, S.L., Zak, D.E., Kennedy, K.A., Thorsson, V., et al.: Uncovering a macrophage transcriptional program by integrating evidence from motif scanning and expression dynamics. *PLOS Comput. Biol.* **4**(4), (2008)
- Riley, J.: *Rising Life Expectancy: A Global History*. Cambridge University Press, Cambridge (2001)
- Robine, J.M.: Ageing populations: We are living longer lives, but are we healthier? United Nation Population Division (2021). UN DESA/POP/2021/TP/NO.2
- Sanders, B.S.: Measuring community health levels. *Am. J. Public Health* **54**(7), 1063–1070 (1964)
- Shang, H.L., Hyndman, R.J.: Grouped functional time series forecasting: an application to age-specific mortality rates. *J. Comput. Graph. Stat.* **26**(2), 330–343 (2017). <https://doi.org/10.1080/10618600.2016.1237877>
- Shang, H.L., Yang, Y.: Forecasting Australian subnational age-specific mortality rates. *J. Popul. Res.* **38**, 1–24 (2021). <https://doi.org/10.1007/s12546-020-09250-0>
- Singhal, A., Seborg, D.: Clustering multivariate time-series data. *J. Chemom.* **19**, 427–438 (2005)
- Sullivan, D.F.: Conceptual problems in developing an index of health: National Center for Health Statistics. *Vital Health Stat.* **2**(17), 1–18 (1965)
- Sullivan, D.F.: A single index of mortality and morbidity. *Health Serv. Rep.* **86**, 347–354 (1971)
- Tarpey, T., Kinader, K.: Clustering functional data. *J. Classif.* **20**(1), 93–114 (2003). <https://doi.org/10.1007/s00357-003-0007-3>
- Torri, T.: Building blocks for a mortality index: an international context. *Eur. Actuar. J.* **1**, 127 (2011). <https://doi.org/10.1007/s13385-011-0014-4>
- Wahba, G.: *Spline models for observational data*. SIAM, Philadelphia (1990). ISBN 0989712440
- Wang, A., Abbas, K.M., Abbasifard, M., et al.: Global age-sex-specific fertility, mortality, healthy life expectancy (HALE), and population estimates in 204 countries and territories, 1950–2019: A comprehensive demographic analysis for the Global Burden of Disease Study 2019. *Lancet* **396**, 1160–1203 (2020)
- WHO methods and data sources for global burden of disease estimates 2000–2019. Department of Data and Analytics Division of Data, Analytics and Delivery for Impact WHO, Geneva (2020)

---

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.