Causes-of-Death Specific Estimates from Synthetic Health Measure: A Methodological Framework

Andrea Nigri1 · Susanna Levantesi2 · Gabriella Piscopo3

Accepted: 17 December 2021 / Published online: 13 January 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract
Life expectancy at birth has attracted interest in various fields, as a health indicator that measures the quality of life. Its appeal relies on the ability to enclose and summarize all the factors affecting longevity. However, more granular information, provided by social indicators such as cause-of-death mortality rates, plays a crucial role in defining appropriate policies for governments to achieve well-being and sustainability goals. Unfortunately, their availability is not always guaranteed. Exploiting the relationship between life expectancy at birth and cause-of-death mortality rates, in this paper we propose an indirect model to produce estimates of death rates due to specific causes using the summary indicator of life expectancy at birth, thus the general levels of the observed mortality. By leveraging on a constrained optimization procedure, we ensure a robust framework where the cause-specific mortality rates are coherent to the aggregate mortality. The main advantage is that indirect estimations allow us to overcome the data availability problem: very often the cause-specific mortality data are incomplete, whereas data on the aggregate mortality are not. Using data from the Human Cause-of-Death Database, we show a numerical application of our model to two different countries, Russia and Spain, which have experienced a different evolution of life expectancy and different leading causes of death. In Spain, we detected the impact of several public health policies on the lowered levels of cancer deaths and related life expectancy increases. As regards the Russia, our results catch the effects of the anti-alcohol campaign of 1985–1988 on longevity changes.

Keywords Life expectancy · Indirect estimates · Causes of death

Andrea Nigri
andrea.nigri@unibocconi.it
Susanna Levantesi
susanna.levantesi@uniroma1.it
Gabriella Piscopo
gabriella.piscopo@unina.it

1 Department of Social and Political Sciences, Bocconi University, Milan, Italy
2 Department of Statistics, Sapienza University of Rome, Rome, Italy
3 Department of Economics and Statistical Science, University of Naples Federico II, Napoli, Italy
1 Introduction

1.1 Literature Review

The rise in human longevity during the 20th century had a prominent impact on population dynamics, showing a rapid decline in the mortality levels in developed countries. After World War II, scientific achievements against chronic diseases (Vaupel 1997; Rau et al. 2008) lowered the adult age mortality, likewise, medical progress and better living conditions provided a reduction in infant mortality. After decades of gains, researchers have been accustomed to steady longevity improvements (Oeppen and Vaupel 2002), although life expectancy has momentarily declined in several developed countries, showing diverging trends, with stagnations and decelerations (Nigri et al. 2021; Levantesi et al. 2021).

Life expectancy represents an indicator measuring the quality of life by enclosing all the factors explaining longevity. The COVID-19 pandemic is affecting some of these factors, globally causing health, economic, environmental, and social problems to the population. As pointed out by Wang et al. (2021), the advent of the pandemic interested individuals’ lifestyles around the world, and a healthy lifestyle is known to be effective in preventing some important diseases such as diabetes (Moradi et al. 2021). The circulation of an excessive amount of information from social media that occurred during the outbreak of the pandemic has probably negatively affected mental health concerns (Su et al. 2021).

Indeed, the literature has recently recognized that single or groups of causes of death might jeopardize longevity evolution (Ho and Hendi 2018), as the Danish stagnation and sub-sequential improvements (Lindahl-Jacobsen et al. 2016; Aburto et al. 2018), or the decrease in Sweden’s life expectancy related to the increase in cardiovascular disease mortality (Drefahl et al. 2014). Similarly, Mehta et al. (2020) bring evidence of stagnation in cardiovascular disease mortality, holding back the increase of US life expectancy. Therefore, it has become crucial to delve into the study of cause-specific mortality, underlining longevity patterns that should be exploited to achieve more accurate demographic forecasting (Manton et al. 1976; Keyfitz 1977; Manton et al. 1986; Caselli et al. 2006; Dimitrova et al. 2013). The cause-specific mortality rates include more peculiar information with respect to the aggregate mortality, which can shed light on the mortality evolution. Paulson et al. (2021) present different mortality findings by analyzing both cause-specific mortality data and aggregate data. In the last decades, longevity change is most likely an outgrowth of the diversification in causes of death around the globe. Early diagnosis in lockstep with acquired epidemiological knowledge, aimed at primary and secondary prevention, witnessed relevant reductions in shares of deaths from circulatory and infectious diseases. These conditions are likely to change from one birth cohort to another (Lindahl-Jacobsen et al. 2016), alongside social-economic conditions, and lifestyle behaviors, smoking, and diet are prime examples. Different cause-of-death patterns might emerge, posing fundamental challenges for the health care system, pension schemes and the insurance industry, among others. In fact, the overall mortality might be described as a composition of causes of death. This poses a coherence issue during the estimation procedure, ensuring the sum of estimated cause-specific mortality rates is equal to the forecasts of the aggregate mortality rates. Many contributions in the literature have not taken this dependence into account and modeled each cause of death independently of the others. Some of them apply the Lee–Carter model (Lee and Carter 1992), see for instance Wilmoth (1995) among others. The resulting cause-specific mortality estimates are inconsistent with the total mortality and produce implausible forecasts.
Our work fits this line of research in lockstep with the aim of an original indirect estimation model, which produces estimates of cause-specific mortality rates exploiting their relationship with the life expectancy at birth in a log-log scale. According to the definition of the UN manual (United Nations 1967), through the indirect estimation it is possible to derive many parameter estimates on the basis of the available information indirectly related to them. Since the 1960s, scholars have been questioning how it was possible to estimate health measures trends for those countries with poor quality and often unavailable data. Indeed, the indirect estimation models can be applied in those situations when traditional methods are jeopardized by large bias.

The approach we follow is peculiar and new. The novelty lies in the incorporation of causes of death information into indirect mortality estimation models based on life expectancy. Thus, by working on causes of death directly, we fill the gap between cause-specific mortality models and indirect estimation approaches. Our model differs from the canonical models as we indirectly estimate cause-of-death mortality from the observed general levels, consistently with the values of life expectancy at birth.

Specifically, we exploit the relationship between cause-of-death mortality rates and life expectancy at birth to reconstruct, for a given age, the cause-specific mortality over time from a single value of life expectancy at birth. The advantages of using life expectancy mainly reside in the ability of this indicator to enclose all the factors affecting longevity. Besides, life expectancy models offer a high level of forecast accuracy, being parsimonious at the same time (Oeppen and Vaupel (2002), White (2002), Torri and Vaupel (2012), Raftery et al. (2014), Pascariu et al. (2018), Nigri et al. (2020)). To the best of our knowledge, our study is the first attempt to apply an indirect estimation model to address the issue of describing the mortality pattern by causes of death.

1.2 The Study Design

Our contribution to the literature is twofold. First of all, starting from the available values of life expectancy at birth we work with all observed causes of death, ensuring the consistency of our estimations against the total mortality. Secondly, indirect estimations allow us to overcome the problem of incomplete data: very often cause-specific mortality data are incomplete, whereas our approach exploits the univariate series of life expectancy to reconstruct the mortality rates by cause indirectly. This framework may represent a prominent approach to help Governments define appropriate policies for increasing local wellbeing in presence of poor or incomplete information.

The remaining part of this work is organized as follows. In Sect. 2 we describe the longevity measures, in Sect. 3 we introduce the methodology. Sections 4 and 5 are devoted respectively to the description of data and results from the numerical application. Social implications and final remarks are offered in Sects. 6 and 7.

2 Longevity Indicators

2.1 Life Expectancy

The health profile is crucial to define well-being, and among the OECD indicators (OECD (2011)), life expectancy plays a key role in measuring the quality of life. It is indeed extensively used in literature D’Urso et al. (2019, 2020); Alaimo and Maggino (2020); Lind (2021);
Ruscone and Fernandez (2021). It refers to the average number of years a synthetic cohort of newborns would live if they were to experience the death rates observed in a given period throughout their lifespan. Life expectancy is an age-standardized indicator, which makes it appealing for comparisons over time and across countries with populations of different sizes and age structures.

Let \( x \) be the age and \( t \) the calendar year. Let \( S(x, t) \) and \( \mu(x, t) \) be two continuous functions with respect to age and time, respectively representing the survival probability and the force of mortality of an individual aged \( x \) at time \( t \) in a given population. We denote \( e_x(t) \) the life expectancy at age \( x \) and time \( t \), that is defined as follows:

\[
e_x(t) = \frac{\int_x^\infty S(y, t)dy}{S(x, t)}
\]  

(1)

where \( S(x, t) = \exp \left( - \int_0^x \mu(a, t)da \right) \) is the survival function and \( \mu(a, t) \) is the force of mortality at age \( a \) at time \( t \) survival probabilities.

2.2 Causes of Death Rates

More recently other longevity indicators have been introduced contributing to defining health profiles, such as mortality specific for causes (D’Urso et al. 2019), with particular regard to Cancer and Heart diseases. Indeed with the aging of society, national plans aim to measure risks indicators, especially at a population level.

Let suppose to observe \( C \) mutually exclusive causes of death, for \( X \) different ages over \( T \) calendar years. The central death rate for cause \( c \in C \) is defined as:

\[
m_x(c, t) = \frac{D_x(c, t)}{E_x(t)}
\]  

(2)

Where \( D_x(c, t) \) is the number of deaths aged \( x \) in year \( t \) due to cause \( c \), and \( E_x(t) \) are the exposures-to-risk aged \( x \) in year \( t \).

3 Methods

Let’s assume that the number of deaths follows a Poisson distribution, \( D_x(c, t) \sim \text{Poisson}\{E_x(t) \cdot m_x(c, t)\} \). For each age \( x \in X \), the aim of our model is to convert a given value of life expectancy at birth, \( e_0(t) \), into a list of cause-specific death rates \( m_x(c, t) \). Formally, the specification of the model is:

\[
\log m_x(c, t) = \beta_x(c) \log e_0(t) + \delta_x(c) \gamma_x(t)
\]  

(3)

Where, for a given age, \( \beta_x(c) \) represents the sensitivity of cause-specific mortality to the variation of \( \log e_0(t) \), then estimating the relationship between cause-specific mortality rates and life expectancy at birth in log-log scale. \( \delta_x(c) \) provides the cause-of-death mortality structure within the estimation period, in other words for each reference period, it allows for catching the dynamic of mortality composition by causes. Finally, \( \gamma_x(t) \) is the parameter which satisfies the condition that the sum of the estimated cause-specific mortality \( \hat{m}_x(c, t) \) for all the causes of death equals the mortality rates aggregated on all the causes \( m_x(t) \):
\[ \sum_{c \in C} \hat{m}_c(c, t) = m_x(t) \]  

Therefore, \( \gamma_x(t) \), adjusting the cause-specific mortality rates can the fulfillment of the aggregation constraint in Eq. 4 and coherent and proper causes of deaths modeling. Moreover, given a certain age, our model links the time series of life expectancy at birth to the central death rates by cause and time that return the life expectancy at birth level \( e_0(t) \). This ensures that the estimated life expectancy trend is consistent with the observed one, in line with the procedure of Lee and Miller (2001) developed in the framework of general mortality instead of causes of death.

### 3.1 Parameters’ Estimation

Starting from \( m_x(c, t) \) for a given age \( x \in X \) and \( m_x(t) \) as inputs\(^1\), our model is fitted by maximizing the log-likelihood function, \( \log L[\beta_x(c), \delta_x(c), \gamma_x(t)] \), which is given by:

\[
\log L[\beta_x(c), \delta_x(c), \gamma_x(t)] = \sum_{x,t} \{D_x(c, t)[\beta_x(c) \log e_0(t) + \delta_x(c) \gamma_x(t)] - E_x(t) \exp [\beta_x(c) \log e_0(t) + \delta_x(c) \gamma_x(t)] \} + C
\]

Where \( C \) is a constant. We propose to estimate the parameters modifying the scheme proposed by Brouhns et al. (2002) based on the Newton-Raphson algorithm to adapt it to our indirect estimation model. The steps of the procedure we propose, starting from the initial values (iteration 0) \( \beta_x^0(c) = 0, \delta_x^0(c) = 1, \gamma_x^0(t) = 0 \), are as follows:

**Iteration k + 1:**

\[
\hat{\beta}_x^{k+1}(c) = \hat{\beta}_x^k(c) + \frac{\sum_t (D_x(c, t) - \hat{D}_x^k(c, t))(\log e_0(t))}{\sum_t \hat{D}_x^k(c, t)(\log e_0(t))^2},
\]

\[
\hat{\delta}_x^{k+1}(c) = \hat{\delta}_x^k(c),
\]

\[
\hat{\gamma}_x^{k+1}(t) = \hat{\gamma}_x^k(t).
\]

**Iteration k + 2:**

\[
\hat{\delta}_x^{k+2}(c) = \hat{\delta}_x^{k+1}(c) + \frac{\sum_t (D_x(c, t) - \hat{D}_x^{k+1}(c, t))(\log e_0(t))(\hat{\gamma}_x^{k+1}(t))}{\sum_t \hat{D}_x^{k+1}(c, t)(\log e_0(t))^2(\hat{\gamma}_x^{k+1}(t))^2},
\]

\[
\hat{\beta}_x^{k+2}(c) = \hat{\beta}_x^{k+1}(c),
\]

\[
\hat{\gamma}_x^{k+2}(t) = \hat{\gamma}_x^{k+1}(t).
\]

**Iteration k + 3:**

\(^1\)To this end, we may use the observed values for countries where the overall mortality is provided but causes-specific levels are not available or incomplete. In other cases the predicted values of life expectancy at birth and deaths, which constitute the model’s input, can be obtained by a certain extrapolation method or be the target values provided by official forecasting (e.g. WHO).
\[
\hat{\gamma}_{x}^{k+3}(t) = \hat{\gamma}_{x}^{k+2}(t) + \frac{\sum(D_{x}(c, t) - \hat{D}_{x}^{k+2}(c, t))(\log E_{0}(t))(\hat{\delta}_{x}^{k+2}(c))}{\sum(D_{x}(c, t))(\log E_{0}(t))^{2}(\hat{\delta}_{x}^{k+2}(c))^{2}},
\]

\[
\hat{\beta}_{x}^{k+3}(c) = \hat{\beta}_{x}^{k+2}(c),
\]

\[
\hat{\delta}_{x}^{k+3}(c) = \hat{\delta}_{x}^{k+2}(c).
\]

where \(\hat{D}_{x}(c, t)\) is the estimated number of deaths after iteration \(k\).

The second step consists in optimizing \(\hat{\gamma}_{x}(t)\), where the optimized value \(\hat{\gamma}_{x}(t)\) satisfies the condition in eq. 4. Hence, denoting \(\Delta_{x}(c, t) = |\sum_{c \in C} \log \hat{m}_{x}(c, t) - \sum_{c \in C} \log m_{x}(c, t)|\) as the difference between the sum of estimated and observed cause-specific mortality rates in log scale, given a certain age \(x\), the optimization problem can be formalized as follows:

\[
\min_{\gamma_{x}(t)} \Delta_{x}(c, t)
\]

subject to

\[
\sum_{c \in C} \hat{m}_{x}(c, t) = m_{x}(t)
\]

It is a linear problem that can be solved by using the optimization algorithms developed in the R package *optim*.

This model allows to easily obtain the prediction intervals by applying the above estimation procedure to bootstrap samples of the input data.

4 Data

The empirical study concerns Russian and Spanish male mortality. In both countries, the cause-of-death data have been taken from the newly developed Human Cause-of-Death Database (HCD)\(^2\), which provides high-quality data on the cause-specific mortality. It is coded using the international classification of diseases (ICD), providing different aggregation levels: full list, intermediate list, and shortlist. Each classification has been developed using the same criteria for all countries, ensuring homogeneity and comparability. By using these data, we obtain a universal and standardized methodology to redistribute deaths between 104 disease categories in five-year age groups. Avoiding issues regarding the ICD revisions and ensuring cross-country comparability to different coding practices.

We truncated the cause-of-death analysis at age 80 because of classification quality and the presence of comorbidities (Aburto and van Raalte 2018).

We start from the shortlist and further clustering the death classification. For Spain, we consider the following four major causes of death with the indication of the ICD codes:

- (1) Cancer (C00-D48),
- (2) Circulatory (I00-I52, G45, I60-I69, I70-I99),
- (3) Respiratory (J00-J22, U04, J30-J98) and Infectious (A00-B99),
- (4) Others (residual class).

For Russia, we consider the following five major causes of death:

\(^2\) The HCD database can be found at www.causesofdeath.org.
We choose circulatory and cancer causes because they belong to the main causes of death in both countries. Respiratory and infectious diseases are among the most important causes of death in Spain. According to the Global Burden of Disease, remarkable increases in Spanish mortality from 1990 to 2016 were observed due to respiratory and infectious complications (Soriano et al. 2018). Transportation accidents are particularly high in Russia. For instance, according to the World Health Organization (The Global Health Observatory), the road fatality rate in Russia in 2016 was 18 per 100,000 population compared to 4 in Denmark, 4.1 in Spain and 5.5 in France. Therefore, the causes’ choice has been driven by each country’s specific mortality history and population health style behavior. Consequently, the results for Spain and Russia are not directly comparable. They are used to illustrate the proper functioning of the model for different sets of causes of death.

Determining the exact connection between mortality rates and life expectancy is complex and becomes relevant when the latter is used as a summary measure of longevity in the analysis of mortality trends (Pollard (1982)). It is not easy to answer the problem even in the area of causes of death. However, we base our model on the linear relation between cause-of-death mortality and life expectancy at birth, both of them on log scale.

To empirically support our assumption we carry out a correlation analysis between these two indicators. Results are illustrated in Fig. 1 for Russia and Fig. 2 for Spain at fixed ages (20, 40, 60, 80) and years 1993–2012. Plots for other two longer periods (1983–2012 and 1988–2012) are reported in the Appendix. These figures show the relationship between \( \log m_x(c, t) \) and \( \log e_0(t) \) and display the Pearson correlation coefficient \( R \) with a significance level \( p < 0.05 \) for each age.

Although slopes and intercepts differ from case to case, we can affirm that a linear correlation between the logarithmic levels of cause-specific mortality and life expectancy across ages exists. Figures 1 and 2 often show very high values of \( R \). The only exceptions (\( R \) values less than 50\%) are for mortality by cancer at age 80, circulatory diseases at age 20 and wholly attributable to alcohol diseases at age 60 and 80 in Russia, and residual cause 4 at age 80 in Spain. In these cases life expectancy at birth should have a limited explicative power of the cause-specific mortality. However, when the linear relationship of life expectancy and cause-specific mortality on a log–log scale fails, the model may still be able to provide reliable estimates. In fact, the presence of the \( \delta_x(c) \) parameter (see Eq. 9) can allow the model to grasp what is not captured from life expectancy.

5 Results

The analysis is performed rolling the origin of the observed period from 5 to 10 years, then obtaining three time windows of estimation (w) for the model fitting: 1983–2002 (w=1), 1988–2002 (w=2), 1993–2002 (w=3). While the estimation period is set to 2003–2012 and remains the same for all the time windows.

Figure 3 illustrates the estimated values of \( \beta_x(c) \) (left panels) and \( \delta_x(c) \) (right panels) parameters for Russia at ages 20, 40, 60 and 80. Concerning \( \beta_x(c) \), the closer its value
is to zero, the lower the relationship between life expectancy at birth and cause-specific mortality on log–log scale. The values of $\beta(c)$ highlight that there is a stronger negative relationship between $e_0(t)$ and mortality due to cause 3 (transportation accidents) in the adult ages compared to the other causes of death. This behavior holds for all time windows of estimation. However, in absolute terms, the power of $e_0(t)$ in explaining the mortality due to transportation for age 20 is equal to that of age 40, moderately higher than age 60 and more significant than age 80. As regards cause 2 (circulatory diseases), parameter $\beta(c)$ increases as the age increases, demonstrating that circulatory diseases have a greater influence on mortality at older ages. The same behavior of the parameter is found for cause 4 (cancer). The values of $\delta(c)$, which provides the cause-specific mortality structure over the estimation period, evidence some significant changes of mortality composition by causes as the time window of estimation changes. We observe remarkable changes for causes 1 and 2 (alcohol and circulatory diseases, respectively) at age 20 and for all causes at age 80.

Figure 4 shows the estimated values of $\beta(c)$ (left panels) and $\delta(c)$ (right panels) parameters for Spain at ages 20, 40, 60 and 80. The results of $\beta(c)$ are more sensitive to changes in
the time window of estimation, contrary to what happened to Russia, where $\beta_x(c)$’s are substantially independent from the time window. We speculate that the effect of the time window on $\beta_x(c)$ could be due to the dynamics of life expectancy at birth: in Russia, where values of $\beta_x(c)$ are not sensitive to $w$, there were no substantial changes in $e_0(t)$ due to the longevity stagnation, while in Spain, which experienced a growth of $e_0(t)$ over the period studied, $\beta_x(c)$ is most affected from a change in $w$. In view of the relationship between $e_0(t)$ and mortality by cause of death in the log-log scale, the strongest impact is shown for cause 3 (respiratory and infectious) at all ages, in line with the result of the demographic research previously cited. Concerning $\delta_x(c)$, we observe appreciable changes for causes 1 and 2 (cancer and circulatory diseases, respectively) at age 20 and for all causes at age 80 as the time window changes. In general, $\delta_x(c)$ values for Spain show trends that are less sensitive to changes in the time window than those for Russia, except for age 40.

The model’s accuracy is evaluated by the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), which are defined respectively as:

\[
\text{RMSE} : \sqrt{\frac{\sum_{x,t} (m_x(c,t) - \hat{m}_x(c,t))^2}{n}},
\]

\[
\text{MAE} : \frac{\sum_{x,t} \mid m_x(c,t) - \hat{m}_x(c,t) \mid}{n}.
\]
The results of RMSE and MAE calculated on the different time windows of estimation are reported in Table 1 for Russia and Table 2 for Spain. For Russia, our model obtains very low errors, less than 0.2%. Specifically, the errors are even below 0.02% for causes 1 (alcohol) and 3 (transportation accidents). For Spain, errors are lower than 0.08% showing no appreciable differences between the causes.

Figure 5 shows cause-specific mortality rates over the estimation period 2003–2012 for Russia. For this country, we can observe that our model generally well captures the trend of cause-specific mortality at all ages and each time window of estimation. However, it partially deviates at ages 20 and 40 for the residual cause 5 if we consider the most recent time window of estimation (w=3) and at age 40 for cause 2 (circulatory diseases). The time window is also particularly significant in the case of cause 4 (cancer) at age 60 and cause 2 (circulatory diseases) at age 80, where the older period in the first case and latest period in the second case are less effective.

For Spain, Fig. 6 shows that the goodness of the estimation depends, more clearly than Russia, from the time windows of estimation. The motivation lies in what has been observed in the results of the estimated $\beta_c(x)$ parameter. The time window 1993–2002 (w=3) definitely provides the best estimates for all ages and causes, except for cause 4 (other causes) at age 20 and 40 and cause 1 (cancer) at age 40. Finally, for older ages, we highlight model improvements for all causes when moving from an earlier time window of estimation to a more recent one.
Fig. 4 Parameters $\beta(x)$ and $\delta(x)$ for different time windows of estimation, 1983–2002 (w=1), 1988–2002 (w=2), 1993–2002 (w=3). Ages 20, 40, 60, 80. Spain

Table 1 RMSE and MAE over years 2003–2012 for different time windows of estimation. Russia

| Cause          | 1983–2002 | 1988–2002 | 1993–2002 |
|----------------|-----------|-----------|-----------|
|                | RMSE      | MAE       | RMSE      | MAE       | RMSE      | MAE       |
| 1: Alcohol     | 0.000245  | 0.000162  | 0.000236  | 0.000149  | 0.000197  | 0.000127  |
| 2: Circulatory | 0.001405  | 0.000893  | 0.001207  | 0.000752  | 0.001953  | 0.000885  |
| 3: Transp. acc.| 0.000055  | 0.000043  | 0.000056  | 0.000042  | 0.000044  | 0.000032  |
| 4: Cancer      | 0.000898  | 0.000571  | 0.000754  | 0.000479  | 0.000673  | 0.000380  |
| 5: Others      | 0.001514  | 0.000777  | 0.000964  | 0.000506  | 0.000811  | 0.000557  |

Table 2 RMSE and MAE over years 2003–2012 for different time windows of estimation. Spain

| Cause          | 1983–2002 | 1988–2002 | 1993–2002 |
|----------------|-----------|-----------|-----------|
|                | RMSE      | MAE       | RMSE      | MAE       | RMSE      | MAE       |
| 1: Cancer      | 0.000684  | 0.000291  | 0.000400  | 0.000200  | 0.000260  | 0.000121  |
| 2: Circulatory | 0.000776  | 0.000250  | 0.000442  | 0.000176  | 0.000372  | 0.000152  |
| 3: Respir. + Infect. | 0.000424 | 0.000148  | 0.000351  | 0.000138  | 0.000305  | 0.000103  |
| 4: Others      | 0.000487  | 0.000219  | 0.000444  | 0.000266  | 0.000329  | 0.000196  |
5.1 Comparison with the Lee–Carter Model on a Single Cause of Death

The last step of our numerical application focuses on a single cause of death. Indeed, our model is able to reconstruct, for a given age, the cause-specific mortality over time from a single value of life expectancy at birth. The aim is to show its reliability compared to

Fig. 5 Mortality rates by cause over years 2003–2012 for different time windows of estimation, 1983–2002 (w=1), 1988–2002 (w=2), 1993–2002 (w=3). Russia

Fig. 6 Mortality rates by cause for different time windows of estimation, 1983–2002 (w=1), 1988–2002 (w=2), 1993–2002 (w=3). Spain
the widely used Lee–Carter model. Therefore, we highlight the scope differences between these two models: our indirect estimation is useful where mortality data for a specific cause of death are not available and it is not possible to implement a direct estimation, while the Lee–Carter model provides mortality rates estimates for the single cause of death working on past observations. The Lee–Carter model has been developed for modeling total mortality, but has also been applied to cause-specific mortality modeling. The Lee–Carter model working on the generic cause $c$ can be formalized as follows:

$$\log m_c(x, t) = \alpha_c(x) + \beta_c(x) \kappa_c(t)$$  \hspace{1cm} (9)$$

Where, consistently with our model, we assume that the number of deaths follows a Poisson distribution. The models’ performances are compared using RMSE and MAE, and the relative differences of the central death rates between our model/Lee–Carter model and the observed values as

$$\mathcal{M}_c(x, t) = \frac{\hat{m}_c(x, t) - m_c(x, t)}{m_c(x, t)}$$

Starting from age 5 up to 80, we focus the analysis on two relevant causes of death: mortality wholly attributable to alcohol in Russia and cancer in Spain.

### 5.2 Discussion

**Mortality wholly attributable to alcohol in Russia.** According to Rehm et al. (2010), mortality wholly attributable to alcohol relates to the health conditions which, by the ICD definition, recognize alcohol consumption as a principal cause. This cause of death accounts for a high proportion of early mortality in Russia and has been widely analyzed in the literature. Indeed, we found a remarkable relationship between summary longevity levels and mortality attributable to alcohol in Russia. Specifically, our model tested on three-time windows of estimation, always outperforms the Lee–Carter model estimated on the single cause (see Table 3 reporting RMSE and MAE over years 2003–2012 for the different time windows of estimation).

The relative differences of the central death rates between our model, Lee–Carter model and the observed values $\mathcal{M}_c(x, t)$ for $c = 1$ (wholly attributable to alcohol) is illustrated in Fig. 7. It gives evidence of the supremacy of our model over the Lee–Carter one especially at younger ages and from the year 2007, independently from the time window used for estimating the model’s parameters. It is evident that the younger age groups are marginally influenced by this cause and the link with the general trend of alcohol-related mortality produces error in the fitting. Our model instead, captures a more complex relationship. Therefore by leveraging also on the analysis of parameters, we bring evidence of the longevity levels sensitivity to the anti-alcohol campaign of 1985–1988 that showed a

### Table 3

RMSE and MAE over years 2003–2012 for different time windows of estimation. Russia

| Country | Cause        | Model     | 1983–2002 | 1988–2002 | 1993–2002 |
|---------|--------------|-----------|-----------|-----------|-----------|
|         |              | RMSE      | MAE       | RMSE      | MAE       | RMSE      | MAE       |
| Russia  | Alcohol      | Our model | 0.000245  | 0.000162  | 0.000236  | 0.000149  | 0.000197  | 0.000127  |
|         |              | Lee–Carter| 0.000302  | 0.000170  | 0.000367  | 0.000204  | 0.000209  | 0.000134  |
rapid increase in life expectancy in Russia (Shkolnikov and Nemtsov 1997) where male life expectancy recovered to mid-1960s levels. Equally, the termination of the aforementioned measures resulted in a sharp decline in life expectancy at the beginning of the 1990s (Shkolnikov and Cornia 2000; Shkolnikov et al. 2001). In 1994, life expectancy in Russia fell to the lowest levels ever recorded in the country. After reaching these low levels, life expectancy at birth increased rapidly between 1994 and 1998, even if the increase was short-termed, and after 1998, began a new decline albeit not as fast as in the early 1990s.

**Cancer in Spain.** Over the last decade, Spanish life expectancy has increased more rapidly than the EU average, reaching 83.4 years. This achievement has been driven by the reduction in mortality rates from different causes of death (circulatory diseases, notably ischaemic heart disease and cerebrovascular disease (stroke)). Among others, lung, colorectal, and smoking-related cancers, also remain important causes of death, but their burden has decreased at least since 2000. Indeed, our findings bring evidence of the high sensitivity of longevity to changes in mortality levels attributable to cancer. Our model parameters reflect the mortality changes incorporated into different time windows. Specifically, we bring evidence of improvements for cancer mortality confirming the usefulness of screening programs, supported by the OECD report OECD (2019), as well as the anti-tobacco law adopted in 2005 and reinforced in 2010.

---

3 In Spain in 2018 according to the Spanish Statistical Institute, they account for 26.4% of the total deaths (male population), second only to diseases of the circulatory system which represents 28.3% of the total deaths.

4 Deaths by cause of death (short list). Available at www.ine.es.
Even though mortality due to cancer is very important, limited research has been performed to improve the forecasting of cancer deaths. The error measures are reported in Table 4. Similar to what was observed in Russia, our model had always the lowest values of both RMSE and MAE with respect to the Lee–Carter model. Our model clearly highlights its supremacy over the Lee–Carter model for each time window of estimation. It should be noted that the estimates obtained with our model improve as the time window progresses over time, which does not happen with the LC, so that it appears suitable to capture the increase in survival that have been recorded in recent years due to medical advances, which have improved prognosis especially for the middle age groups. And it is precisely for these ages that our model shows a better estimation capacity.

Results of the relative difference $M_c(x, t)$ for $c = 1$ (cancer) are illustrated in Fig. 8.

![Table 4](image)

**Table 4** RMSE and MAE over years 2003–2012 for different time windows of estimation. Spain

| Country | Cause | Model       | 1983–2002     | 1988–2002     | 1993–2002     |
|---------|-------|-------------|---------------|---------------|---------------|
|         |       | RMSE        | MAE           | RMSE          | MAE           |
| Spain   | Cancer | Our model   | 0.000684      | 0.000291      | 0.0000400     | 0.0002000     | 0.0002600     | 0.0001210     |
|         |       | Lee–Carter  | 0.000814      | 0.000456      | 0.0000772     | 0.0004280     | 0.0008030     | 0.0004450     |

![Fig. 8](image)

Fig. 8 Spain, Cancer: relative differences $M_c(x, t)$ over years 2003–2012 for different time windows of estimation
6 Social Implications

By recurring to the life expectancy at birth as a key indicator to monitor the evolution of population health, we have shed light on an important feature that is often overlooked. Despite life expectancy is an indicator that embeds macro-level factors affecting social and health policy more peculiar indicators might be needed. Both Governments and researchers would need detailed information for a more comprehensive social planning, such as age-specific rates for the cause of death, which are very limited compared to overall mortality. Indeed, the proposed approach allows estimating death rates due to specific causes using the summary indicator of life expectancy at birth. This can be valuable in contexts where population-level mortality studies are hindered by financial or time constraints, for national registry systems showing a slow-down data processing, or for others who do not support the open data system. Indirect estimation offers a benchmark with which overall mortality can be compared, thus dealing with very uncertain scenarios. Reliable estimation of health indicators like causes-of-death specific rates allows qualifying more appropriate National and sub-National frameworks to achieve wellbeing and sustainability objectives. Our approach provides a more informative perspective at a social level and it might be also considered a diagnostic tool that helps to draw conclusions about changes in life expectancy at birth, which is susceptible to variations of specific causes of death.

As stated in the introduction, the COVID-19 pandemic created several adverse consequences, which are leading to changes in social behaviors, affecting mental health. The need for the psychological well-being of the population in response to the pandemic has been well highlighted for instance in Abbas (2020) that recommend implementing proper mental health precautions along with physical health precautions, particularly in underdeveloped countries, like Pakistan. The outbreak is also posing challenges at the organizational level to the firms to continue business operations. This issue has been analyzed by Azizi et al. (2021) which advise innovative human resource management strategies to be implemented during the COVID-19 pandemic. Appropriate human resource management strategies would increase employees’ mental well-being, satisfaction, productivity, motivation, and health safety at the workplace. These recommendations should, in turn, influence the future life expectancy dynamics. In this context, our model could be applied to study the effects of the restrictive policies implemented in the various countries and the vaccination campaign on overcoming the post-crisis mortality peaks and longevity changes. However, at the moment, no data are available to conduct this analysis.

7 Conclusions

The proposed model represents an advance in social-health indicators modeling, offering the advantage of an indirect and complementary way to approximate death rates specific for age and causes. We are able to provide a comprehensive picture of specific mortality levels, in lockstep to a summary longevity trend. Results give us many insights into the ongoing trend in terms of the composition of causes of death and while many of them come as no surprise, in some cases, we might find untracked behaviors. In Spain, we detected the impact of several public health initiatives implemented to address important risk factors. Indeed Spain’s lowered levels of cancer deaths and related life expectancy increase can be partly explained by the strong public health policy. Narrowing the analysis down to the
Russian case study, our results provide the effectiveness of historical longevity changes due to political decisions. The proposed model is able to catch the effects of the anti-alcohol campaign of 1985–1988 and its national changes that directly affected life expectancy behaviors. In conclusion, our model provides reliable indirect estimation, outperforming the Lee–Carter model (estimated on the single cause), showing its utility also as a demographic tool. We should bear in mind that our model, as other longevity methods (Lee and Carter (1992), Lee and Miller (2001)), leverages on the log-linear assumption, exploiting the regularities of age patterns of mortality that might not hold for some specific causes.

Appendix

See Figs. 9, 10, 11, 12.

Fig. 9 Relation between \( \log m_x(c, t) \) and \( \log e_0(t) \). Panels contain data for specific ages (20, 40, 60, 80). The Pearson correlation coefficient \( R \) and the significance level \( p \) are reported in each panel. Years 1983–2012. Russia
Fig. 10 Relation between $\log m_x(c, t)$ and $\log e_0(t)$. Panels contain data for specific ages (20, 40, 60, 80). The Pearson correlation coefficient $R$ and the significance level $p$ are reported in each panel. Years 1983–2012. Spain.
Fig. 11  Relation between $\log m_c(t, t)$ and $\log e_0(t)$. Panels contain data for specific ages (20, 40, 60, 80). The Pearson correlation coefficient $R$ and the significance level $p$ are reported in each panel. Years 1988–2012. Russia
Fig. 12 Relation between $\log m_x(c, t)$ and $\log e_{0t}(t)$. Panels contain data for specific ages (20, 40, 60, 80). The Pearson correlation coefficient $R$ and the significance level $p$ are reported in each panel. Years 1988–2012. Spain

**Funding** Susanna Levantesi and Andrea Nigri acknowledge financial support from Sapienza University of Rome, Grant RG11916B7982729D.

**References**

Abbas, J. (2020). The impact of coronavirus (SARS-CoV2) epidemic on individuals mental health: The protective measures of pakistan in managing and sustaining transmissible disease. *Psychiatria Danubina*, 32(3–4), 472–477. https://doi.org/10.24869/psyd.2020.472.

Aburto, J. M., & van Raalte, A. (2018). Lifespan dispersion in times of life expectancy fluctuation: The case of central and eastern Europe. *Demography*, 55, 2071–2096.

Aburto, J. M., Wensink, M., van Raalte, A., & Lindahl-Jacobsen, R. (2018). Potential gains in life expectancy by reducing inequality of lifespans in Denmark: An international comparison and cause-of-death analysis. *BMC Public Health*, 18(1), 831. https://doi.org/10.1186/s12889-018-5730-0.

Alaimo, L. S., & Maggino, F. (2020). Sustainable development goals indicators at territorial level: Conceptual and methodological issues-the Italian perspective. *Social Indicator Research*, 147, 383–419.

Azizi, M. R., Atlasi, R., Ziapour, A., & Naemi, R. (2021). Innovative human resource management strategies during the COVID-19 pandemic: A systematic narrative review approach. *Heliyon*, 7(6), e07233. https://doi.org/10.1016/j.heliyon.2021.e07233.

Brouhns, N., Denuit, M., & Vermunt, J. K. (2002). A poisson log-bilinear approach to the construction of projected life tables. *Insurance: Mathematics and Economics*, 31, 373–393.

Caselli, G., Vallin, J., & Marsili, M. (2006). How useful are the causes of death when extrapolating mortality trends. An update. *Social Insurance Studies from the Swedish Social Insurance*, 4, 9–36.

Dimitrova, D., Haberman, S., & Kaishev, V. (2013). Dependent competing risks: Cause elimination and its impact on survival. *Insurance Mathematics and Economics*, 53(2), 464–477.

Drefahl, S., Ahlbom, A., & Modig, K. (2014). Losing ground: Swedish life expectancy in a comparative perspective. *PLoS ONE*, 9(2).
D’Urso, P., Alaimo, L.S., De Giovanni, L., & Massari, R. (2020). Well-Being in the Italian regions over time. Social Indicators Research.

D’Urso, P., De Giovanni, L., Massari, R., & Sica, F. G. M. (2019). Cross sectional and longitudinal fuzzy clustering of the nuts and positioning of the Italian regions with respect to the regional competitiveness index (RCI) indicators with contiguity constraints. Social Indicator Research, 146, 609–650.

Ho, J. Y., & Hendi, A. S. (2018). Recent trends in life expectancy across high income countries: Retrospective observational study. BMJ, 362, k2562.

Keyfitz, N. (1977). What difference would it make if cancer were eradicated? An examination of the Taubler paradox. Demography, 14, 411–418.

Lee, R. D., & Carter, L. R. (1992). Modeling and forecasting US mortality. Journal of the American statistical association, 87(419), 659–671.

Lee, R., & Miller, T. (2001). Evaluating the performance of the Lee-Carter method for forecasting mortality. Demography, 38(4), 537–549.

Levantesi, S., Nigri, A., & Piscopo, G. (2021). Clustering-based simultaneous forecasting of life expectancy time series through long-short term memory neural networks. International Journal of Approximate Reasoning. https://doi.org/10.1016/j.ijar.2021.10.008.

Lind, N. (2021). Objective inequality indexes joining income with life expectancy through the life quality index of sub-populations. Social Indicators Research, 153, 781–794.

Lindahl-Jacobsen, R., Rau, R., Jeune, B., Canudas-Romo, V., Lenart, A., Christensen, K., & Vaupel, J. W. (2016). Rise, stagnation, and rise of Danish women's life expectancy. Proceedings of the National Academy of Sciences, 113(15), 4015–4020.

Manton, K. G., Stallard, E., & Vaupel, J. W. (1986). Alternative models for the heterogeneity of mortality risks among the aged. Journal of American Statistics Association, 81(395), 635–644.

Manton, K. G., Tolley, H. D., & Poss, S. S. (1976). Life table techniques for multiple-cause mortality. Demography, 13(4), 541–564.

Mehta, N. K., Abrams, L. R., & Myrskylä, M. (2020). US life expectancy stalls due to cardiovascular disease, not drug deaths. Proceedings of the National Academy of Sciences, 117(13), 6998–7000.

Moradi, F., Ziapour, A., et al. (2021). Comparing the associated factors on lifestyle between type 2 diabetic patients and healthy people: A case-control study. International Quarterly of Community Health Education. https://doi.org/10.1177/0272684X211022158.

Nigri, A., Barbi, E., & Levantesi, S. (2021). The relationship between longevity and lifespan variation. Statistical Methods and Applications. https://doi.org/10.1007/s10260-021-00584-4.

Nigri, A., Levantesi, S., & Marino, M. (2020). Life expectancy and lifespan disparity forecasting: a long short-term memory approach. Scandinavian Actuarial Journal. https://doi.org/10.1080/03461238.2020.1814855.

OECD. (2019). Spain: Country Health Profile 2019. State of Health in the EU, OECD Publishing, Paris/ European Observatory on Health Systems and Policies, Brussels. https://doi.org/10.1787/8f834636-en

OECD (2011). Compendium of OECD well-being indicators. http://www.oecd.org/sdd/d74717288.pdf

Oeppen, J., & Vaupel, J. W. (2002). Broken limits to life expectancy. Science, 296(5570), 1029–1031.

Pascariu, M. D., Valdimir, C.-R., & Vaupel, J. W. (2018). The double-gap life expectancy forecasting model. Insurance: Mathematics and Economics, 78, 339–350.

Paulson, K. R., Kamath, A. M., Alam, T., Bienhoff, K., Abady, G. G., et al. (2021). Global, regional, and national progress towards Sustainable Development Goal 3.2 for neonatal and child health: all-cause and cause-specific mortality findings from the Global Burden of Disease Study 2019. The Lancet. https://doi.org/10.1016/s0140-6736(21)01207-1.

Pollard, J. H. (1982). The expectation of life and its relationship to mortality. Journal of the Institute of Actuaries, 9, 225–240.

Preston, S. H. (1976). Mortality patterns in national populations: With special reference to recorded causes of death. New York: Academic Press.

Raftery, A. E., Lalic, N., & Gerland, P. (2014). Joint probabilistic projection of female and male life expectancy. Demographic Research, 30, 795.

Rau, R., Soroko, E., Jas stylionis, D., & Vaupel, J. W. (2008). Continued reductions in mortality at advanced ages. Population and Development Review, 34, 747–768.

Rehm, J., Balunus, D., Borges, G. L., Graham, K., Irving, H., Kehoe, T., & Taylor, B. (2010). The relation between different dimensions of alcohol consumption and burden of disease: An overview. Addiction, 105, 817–843.

Ruscone, M.N., & Fernandez, D. (2021). Dynamics of HDI index: Temporal dependence based on D-vine copulas model for three-way data. Social Indicators Research.
Shkolnikov, V. M., & Cornia, G. (2000). The population crisis and rising mortality in transitional Russia. In G. A. Cornia & R. Paniccia (Eds.), The mortality crisis in transitional economies. New York: Oxford University Press.

Shkolnikov, V., McKee, M., & Leon, D. A. (2001). Changes in life expectancy in Russia in the mid-1990s. Lancet, 357, 917–921.

Shkolnikov, V. M., & Nemtsov, A. V. (1997). The anti-alcohol campaign and variations in Russian Mortality. In J. L. Bobadilla, C. A. Costello, & F. Mitchell (Eds.), Premature deaths in the new independent states (pp. 120–155). Washington: National Academy Press.

Soriano, J.B., Rojas-Rueda, D., Alonso, J., Antó, J.M., Cardona, P.J., Fernández, E., García-Basteiro, A.L., Benavides, F.G., Glenn, S.D., Krish, V., Lazarus, J.V., Martínez-Raga, J., Masana, M.F., Nieuwenhuijsen, M.J., Ortiz, A., Sánchez-Niño, M.D., Serrano-Blanco, A., Tortajada-Girbés, M., Tyrovolas, S., Haro, J.M., Naghavi, M., & Murray, C.J.L. (2018). The burden of disease in Spain: Results from the Global Burden of Disease 2016. Medicina Clinica, 151(5).

Su, Z., McDonnell, D., Wen, J., et al. (2021). Mental health consequences of COVID-19 media coverage: the need for effective crisis communication practices. Global Health, 17, 4. https://doi.org/10.1186/s12992-020-00654-4.

Torri, T., & Vaupel, J. H. (2012). Forecasting life expectancy in an international context. International Journal of Forecasting, 28, 519–31.

United Nations. (1967). Methods for estimating basic demographic measures from incomplete data: Manuals and methods of estimating populations. Manual IV. Population Studies, No. 42. Department of Social Affairs.

Vaupel, J. W. (1997). The remarkable improvements in survival at older ages. Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences, 352, 1799–1804.

Wang, C., Wang, D., Duan, K., & Mubeen, R. (2021). Global financial crisis, smart lockdown strategies, and the COVID-19 spillover impacts: A global perspective implications from Southeast Asia. Frontiers in Psychiatry, 12, 1–14. https://doi.org/10.3389/fpsyt.2021.643783.

White, K. M. (2002). Longevity advances in high-income countries, 1955–1996. Population and Development Review, 28, 59–76.

Wilmoth, J. R. (1995). Are mortality projections always more pessimistic when disaggregated by cause of death? Mathematics Population In Studies, 5, 293–319.

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