Early estimation of life expectancy using weekly deaths data

Giampaolo LANZIERI(*)
Eurostat, Unit F5 "Education, health and social protection"
Luxembourg
giampaolo.lanzieri@ec.europa.eu

Abstract

This paper proposes a simple method to obtain early estimates of life expectancy at any age using data on weekly deaths. Although tailored on weekly deaths data from the Statistical Office of the European Union (Eurostat), the logic of the method is applicable to any timely data on infra-annual deaths counts, would they be quarterly, monthly, or other. When the method is applied to a time period still to complete, it is in substance a nowcasting technique whose reliability increases as new data become available, provided a correct specification of the model. It is also presented an application to 30 European countries for the years 2022 and 2023, returning a provisional estimate of life expectancy at birth much earlier than from official statistics. These early estimates show that in Europe the process of recovery in life expectancy at birth from the COVID-19 pandemic will be practically completed in 2023, unless unexpected mortality crises will occur in the second half of the year.

Keywords: mortality, life expectancy, weekly deaths, Europe, nowcasting.

(*) The information and views set out in this paper are those of the Author and do not necessarily reflect the official opinion of the European Commission / Eurostat.

1. Introduction

Life expectancy is one of the most important indicators – if not *the* main measure – of mortality and it is indeed regularly used in numerous policy reports and analyses. In Europe, it is usually estimated by statistical offices at an annual frequency, using data referring to a single calendar year. Data processing requires however at least a few months before official statistics on life expectancy can be released by national statistical offices. For internationally comparable estimates it is necessary to wait even further, as their input data are generally transmitted at a later time to international organizations. For instance, in accordance with the regulations in force in the European Union (EU) on European demographic statistics, the countries transmit population and deaths data to Eurostat (the Statistical Office of the European Union) only by the end of the year following the reference period. Taking into account the time needed for the processing of these data, internationally comparable official estimates of life expectancy are produced about 15 months after the end of the reference year.

The COVID-19 pandemic started in March 2020 has further increased the interest on mortality measures. Consequently, several statistical offices or national health authorities have begun to release higher-frequency – although provisional – mortality data at a much shorter periodicity, sometimes even every week. Since April 2020, Eurostat has collected and disseminated timely data on weekly deaths on European countries. These data offers the opportunity to gain a first insight on contemporary mortality, an information often appreciated especially in times of ongoing health crisis, such as during the COVID-19 pandemic.

In this paper it is proposed a methodology to compute, using weekly deaths data, a preliminary measure of life expectancy at any age as comparable as possible to the regular international estimates on previous years. The method requires only timely infra-annual deaths counts data and life tables of reference for a past time period adopted as benchmark. Because weekly deaths data are freely available for many countries from the Eurostat database, the methodology for the

life expectancy early estimate is adapted to the one currently used by Eurostat. This methodology is applied to a large set of European countries, deriving early estimates of the life expectancies at birth for the years 2022 and 2023. Such estimates based on weekly deaths are provisional in nature as they are meant to be replaced by the official regular estimates once these become available, but they do provide timely indications of ongoing mortality.

2. Data

Weekly deaths data broken down by sex and 5-year age group up to age 90+ are available for European countries from the Eurostat public database (Eurostat 2023a). Eurostat warns the users that, in order to ensure the highest timeliness possible, data are made available as reported by the countries and work on their quality assessment is under development (Eurostat 2023b). The week is defined according to the ISO 8601 standard by the International Organization for Standardization (2019), that is from Mondays to Sundays and starting from the week with the first Thursday of the year in it. The reporting countries also inform Eurostat about the estimated under-coverage of the weekly deaths counts, which is usually dependent upon the lag from the latest weekly data transmitted (under-coverage of death events is more likely for more recent data). Due to data limitations, the countries here considered are the European Union (EU) Member States and the European Free Trade Association (EFTA) countries, except respectively Ireland[1] and Liechtenstein[2]. Additionally, computations are here made also for the European Union as a whole and therefore the application presented below covers 30 geographic entities (see Table 1 for the list). The latest week of reference in 2023 is different across countries (Table 1), but for all of them - except Germany - detailed data are available for all weeks from 2015 to 2022. In particular, data for Germany are by 10-year age group (instead of 5-year groups) for ages under 40, and without breakdown by sex. At the time of this writing, official Eurostat

[1] Ireland does not provide weekly deaths counts with the necessary age and sex detail.
[2] The Eurostat population projections EUROPOP2023, a needed input (see below), do not cover Liechtenstein.

annual deaths data for the year 2022 broken down by sex and age are not yet available, as annual mortality data are available only after more than one year from the end of the calendar year of reference (thus in 2024).

Before their use for life table computations, the weekly deaths data have undergone several transformations: breakdown by 5-year age group of data for Germany (only for ages under 40); aggregation of death counts for ages 85 and over; redistribution of deaths of unknown age; redistribution of deaths of unknown week of occurrence; and correction for under-coverage of death counts. The 10-year age groups of weekly deaths of Germany have been broken down by applying the average proportion of the 5-year age groups (within that same 10-year age group) as from the reported annual deaths in the years 2015-2019. Because according to the methodology currently applied by Eurostat the life table is closed at age 85, in order to obtain provisional estimates as close as possible to the regular Eurostat estimates of the life expectancies, the death counts within the age groups 85-89 and 90+ have been merged in the open-end age group 85+. The redistribution of the sex-specific deaths of unknown age has been implemented proportionally over the sex-specific deaths occurring in the same week; likewise, the redistribution of the sex-specific deaths of unknown week of occurrence has been implemented proportionally over the sex- and age-specific deaths occurring in the same year. In order to correct for under-coverage, the actual counts have been inflated by a factor k, derived as the reciprocal value of the estimated share of coverage $\hat{C}_{c,w} \in [0, 1]$ in the country c for the week w being considered. Therefore, the weekly death count $D_{c,w}$ corrected for under-coverage is $\widehat{D}_{c,w} = \left(1/\hat{C}_{c,w}\right) \cdot D_{c,w}$. However, as at the time of this writing several months have already passed since the end of 2022, the impact of such corrections is nil or negligible for all the countries for deaths occurred in 2022, whilst it can still be noticeable for 2023 (see Table 1), in particular for the latest weeks. On the contrary, no correction has been applied to match the time period covered by the ISO weeks with the calendar year. In fact, the “year” as defined by compiling weekly data may overlap by a few days with the previous and/or the next calendar year (see Table 2).

Table 1: data availability and correction for under-coverage of weekly deaths counts in 2023 as of end July 2023

Country	Code	Latest week	Deaths	Corrected counts	Difference	Relative difference (%)
Austria	AT	27	46 191	46 481	290	0.63
Belgium	BE	30	64 567	64 592	25	0.04
Bulgaria	BG	26	50 424	50 651	227	0.45
Croatia	HR	26	25 799	25 941	142	0.55
Cyprus	CY	28	3 455	3 494	39	1.13
Czechia	CZ	25	54 607	54 745	138	0.25
Denmark	DK	28	31 370	31 391	21	0.07
Estonia	EE	29	8 992	8 992	0	0.00
Finland	FI	28	31 574	32 231	657	2.08
France	FR	27	330 541	330 541	0	0.00
Germany	DE	28	555 190	555 190	0	0.00
Greece	EL	26	65 129	65 501	372	0.57
Hungary	HU	27	66 871	67 101	230	0.34
Iceland	IS	22	1 062	1 062	0	0.00
Italy	IT	22	286 967	286 967	0	0.00
Latvia	LV	26	14 300	14 309	9	0.06
Lithuania	LT	26	19 168	19 168	0	0.00
Luxembourg	LU	26	2 239	2 239	0	0.00
Malta	MT	28	2 179	2 179	0	0.00
Netherlands	NL	29	94 799	94 867	68	0.07
Norway	NO	30	24 668	24 744	76	0.31
Poland	PL	30	234 617	235 593	976	0.42
Portugal	PT	26	60 449	60 468	19	0.03
Romania	RO	26	121 820	123 059	1 239	1.02
Slovakia	SK	25	25 780	26 102	322	1.25
Slovenia	SI	28	11 590	11 636	46	0.40
Spain	ES	26	225 118	225 118	0	0.00
Sweden	SE	29	51 616	51 809	193	0.37
Switzerland	CH	28	37 016	37 539	523	1.41

Table 2: first and last ISO week per calendar year

Calendar year	Nr. of ISO weeks	First ISO week	Last ISO week
2015	53	Monday 29.12.2014 – Sunday 04.01.2015	Monday 28.12.2015 – Sunday 03.01.2016
2016	52	Monday 04.01.2016 – Sunday 10.01.2016	Monday 26.12.2016 – Sunday 01.01.2017
2017	52	Monday 02.01.2017 – Sunday 08.01.2017	Monday 25.12.2017 – Sunday 31.12.2017
2018	52	Monday 01.01.2018 – Sunday 07.01.2018	Monday 24.12.2018 – Sunday 30.12.2018
2019	52	Monday 31.12.2018 – Sunday 06.01.2019	Monday 23.12.2019 – Sunday 29.12.2019
2020	53	Monday 30.12.2019 – Sunday 05.01.2020	Monday 28.12.2020 – Sunday 03.01.2021
2021	52	Monday 04.01.2021 – Sunday 10.01.2021	Monday 27.12.2021 – Sunday 02.01.2022
2022	52	Monday 03.01.2022 – Sunday 09.01.2022	Monday 26.12.2022 – Sunday 01.01.2023
2023	52	Monday 02.01.2023 – Sunday 08.01.2023	Monday 25.12.2023 – Sunday 31.12.2023

In addition to the mismatch between calendar year and the annual time period covered by ISO weeks, there may be further differences between the sum of weekly deaths referring to a selected year and the corresponding reported annual count due to the use of alternative data sources and / or to different statistical production processes for weekly deaths. The Table 3 shows the relative difference between sum of weekly deaths and annual deaths for both sexes. The highest deviations are mostly in the years with 53 ISO weeks, which is in 2015 and in 2020[3], and all deviations are in a range between -2.6% and +3.6%.

[3] In the ISO system, the extra week is attributed to years that have Thursday on the 1 January (here the year 2015) and to leap years that start on a Wednesday 1 January (here the year 2020). Consequently, even though 2016 is a leap year, it does not have 53 ISO weeks. The next leap year will be 2024, but the next year with an extra ISO week will be the year 2026.

Table 3: relative difference between sum of weekly deaths and annual deaths (%)

	2015	2016	2017	2018	2019	2020	2021	2022
EU	1.3	-0.9	-0.8	-0.9	-0.8	0.9	-0.9	-0.9
AT	-0.1	-2.3	-2.4	-2.3	-2.2	-0.4	-2.6	-1.9
BE	1.7	-0.5	-0.4	-0.3	-0.3	1.2	-0.3	-0.2
BG	1.8	-0.5	-0.4	-0.3	-0.2	0.9	-0.2	-0.5
HR	1.7	-0.6	-0.4	-0.3	-0.4	1.6	-0.1	-0.1
CY	2.1	-0.4	0.2	0.5	-0.3	3.6	0.2	-0.6
CZ	1.8	-0.5	-0.4	-0.3	-0.3	1.7	-0.5	-0.3
DK	1.7	-0.5	-0.3	-0.3	-0.3	1.5	-0.2	-0.3
EE	1.6	-0.5	-0.4	-0.4	-0.4	1.6	-0.9	-0.7
FI	1.5	-0.5	-0.5	-0.3	-0.4	1.3	-0.1	-0.3
FR	1.8	-0.5	-0.4	-0.3	-0.3	1.5	-0.4	0.9
DE	1.7	-0.5	-0.3	-0.3	-0.3	1.6	-0.5	-0.2
EL	1.8	-0.6	-0.3	-0.3	-0.3	1.5	-0.2	0.2
HU	1.8	-0.6	-0.4	-0.5	-0.5	1.4	-0.7	-0.6
IS	2.1	-0.6	-0.3	-0.2	-0.5	1.8	0.2	-0.2
IT	3.2	1.4	1.2	0.9	1.3	2.2	0.7	-0.3
LV	1.7	-0.5	-0.3	-0.3	-0.3	1.7	-0.4	-1.5
LT	1.8	-0.6	-0.4	-0.3	-0.3	1.8	-0.6	-0.7
LU	1.6	-0.6	-0.3	-0.4	0.1	1.3	0	-0.1
MT	2.3	-1	-0.4	-0.4	-0.1	1.5	-0.7	-0.9
NL	1.8	-0.5	-0.3	-0.3	-0.3	1.5	-0.5	-0.7
NO	2.1	-0.7	-0.3	-0.3	-0.4	1.5	-0.2	-0.4
PL	1.8	-0.3	-0.1	-0.7	-0.5	1.7	-0.2	-0.3
PT	2	-0.5	-0.4	-0.3	-0.3	1.6	-0.5	-0.2
RO	1.8	-0.5	-0.4	-0.7	0.1	1.1	-0.4	-0.7
SK	1.7	-0.5	-0.4	-0.3	-0.2	2	-0.7	-0.8
SI	1.8	-0.6	-0.4	-0.3	-0.2	1.6	-0.7	-0.2
ES	1.8	-0.5	-0.3	-0.3	-0.3	1.3	-0.3	0
SE	1.8	-0.5	-0.4	-0.3	-0.2	1.5	-0.5	-0.3
CH	1.7	-0.5	-0.4	-0.3	-0.4	1.3	-0.8	-0.4
Average	**1.8**	**-0.5**	**-0.4**	**-0.3**	**-0.3**	**1.5**	**-0.4**	**-0.4**

NOTES: higher intensity of red / blue correspond respectively to higher / lower values. EU not including Ireland. Average not including EU (average of the countries).

3. Method

The main idea of the method proposed in this article is to adjust the early estimate of the life expectancy on the basis of the similarity of weekly deaths patterns with the mortality observed in a benchmark period. There are however three main points where the computation of the early life expectancy differs from a conventional method for a life table of a calendar year of interest such as the one described in Preston, Heuveline, and Guillot (2001): the first is the estimation of the person-years of exposure based on the available weekly data, usually covering less than a calendar year; the second is the estimation of the average time spent in life by the persons dying in the time interval; and the third is a correction applied to the life expectancy as obtained from the (usually abridged) life table. These points aim to align early estimates with past official estimates of the life expectancy, adopted as benchmark.

3.1. Population and exposure to the risk of dying

For life table computations, in addition to the death counts it is necessary to have estimates of the population exposure to the risk of dying. This is usually done by averaging the population sizes at the beginning and at the end of the year of interest. The sex- and age-specific person-years of exposure to the risk of dying are then estimated as:

[1] $$ {}_{n}^{s}E_{x}^{t} = 0.5 \cdot ({}_{n}^{s}P_{x}^{1.1.t} + {}_{n}^{s}P_{x}^{1.1.t+1}), $$

where ${}_{n}^{s}\hat{P}_{x}^{1.1.t}$ is the population of sex s and age $[x, x+n)$ on 1 January of the year t. When the available data do not cover the entirety of the reference year, the person-years of exposure are reduced proportionally:

[2] $$ {}_{n}^{s}E_{x}^{[w_1,\ldots,w_l]} = w_l \cdot \frac{7}{g^t} \cdot {}_{n}^{s}E_{x}^{t}, $$

where w_1 is the first week of the year of interest, $w_l \epsilon \mathbb{Z}[1, \dots, 53]$ is the latest week for which weekly death counts are available, and g^t is the number of days in the calendar year *t* (365 or 366). At this point, it is possible to estimate the annualized mortality rates in the year *t* based on the available data covering until week w_l as:

[3] $$ {}_n^s M_x^t = \frac{{}_n^s \widehat{D}_x^{[w_1,..,w_l]}}{{}_n^s E_x^{[w_1,...,w_l]}} $$

where ${}_n^s \widehat{D}_x^{[w_1,...,w_l]} = \sum_1^{w_l} \widehat{D}_{w,x}$ is the sum of observed weekly death counts, corrected for under-coverage, from week 1 to the latest available week w_l. When data cover an entire year (that is, $w_l = 52$ or $w_l = 53$), there is still a slight mismatch between the calendar year and the time covered by the weekly deaths (see

Table ***2***) and, consequently, numerator and denominator in the occurrence-exposure mortality rate above refer to (marginally) different periods in time. When the data do not cover the entire year, the mortality rate estimated in the sub-period is assumed to apply for the remaining part of the year.

3.2. Methodology for the life table

The computation of the life table for the year(s) of interest should adopt to the largest possible extent the methodology used for the estimates of reference. In this article, it is based on a conventional methodology for life table as in Preston, Heuveline, and Guillot (2001), but it has a specific formula for the estimation of one of its parameters, an adaption implemented to align as much as possible the early estimates to the official estimates used as benchmark.

The first step is the conversion from mortality rates to sex-specific probabilities ${}_n^s q_x$ of dying between exact ages *x* and *x*+*n*:

[4] $$ {}_{n}^{s}q_x = \frac{n \cdot {}_{n}^{s}M_x}{1 + (n - {}_{n}^{s}a_x) \cdot {}_{n}^{s}M_x} $$

The parameter ${}_{n}^{s}a_x$ is the sex- and age-specific average time spent living by those dying in the time interval. It may happen that the age interval n of the infra-annual deaths counts is larger than in the annual data used as benchmark ($n > n_b$). In order to be consistent with the total person-years lived by those dying in the same time interval as obtained in the benchmark life tables (usually available by single age), the parameters ${}_{n}^{s}a_x$ are estimated as average of the single-age parameters ${}_{1}^{s}a_{b,x}$ weighted with the corresponding synthetic deaths ${}_{1}^{s}d_{b,x}$ from the benchmark life tables:

[5] $$ {}_{n}^{s}a_x^t = \frac{\sum_{i=0}^{n-1}\left(i + {}_{1}^{s}a_{b,x+i}^t\right) \cdot {}_{1}^{s}d_{b,x+i}^t}{\sum_{i=0}^{n-1} {}_{1}^{s}d_{b,x+i}^t}, $$

where the index i within the parentheses allows to take in account the completed years of life lived by those dying in the interval of width n. If $n = 1$, that is infra-annual mortality data are by single age, the formula above returns ${}_{1}^{s}a_{b,x}^t$ and consequently there is perfect consistency with the benchmark values. In order to obtain an estimate of the parameters ${}_{n}^{s}a_x^t$ to be applied to any year, the formula [5] can be applied to a selected set of years and the annual estimates then averaged:

[6] $$ {}_{n}^{s}\bar{a}_x = \frac{\sum_{t=1}^{T} {}_{n}^{s}a_x^t}{T} \quad \forall s $$

Once estimated the ${}_{n}^{s}\bar{a}_x$ and thus the ${}_{n}^{s}q_x$, the remaining computations of the life table can be carried out like in the standard methodology (Preston, Heuveline, and Guillot 2001) or anyway following the methodology of the life table used for the estimates of reference.

3.3. Correction of the life expectancy estimate

Estimating the life expectancy of a reference year using data from a shorter period as well as using input data different from those in the benchmark estimation may return biased values. To correct such a potential bias, the estimate of life expectancy for the year of interest is shifted by the average of the differences between the estimates of the life expectancy obtained for a benchmark period using the same time interval (that is, the same number of weeks) and the final estimates (that is, using data covering the entire years):

[7] $$\hat{e}_{t,x}^{w_l} = e_{t,x}^{w_l} - \frac{\sum_{b=1}^{B}\left(e_{b,x}^{w_l} - e_{b,x}\right)}{B}$$

where $b \in \lfloor 1, \dots, B \rfloor$ indicates the year in the time period adopted as benchmark, w_l is the highest week for which data are available for the year of interest, and $\hat{e}_{t,x}^{w_l}$ is the corrected estimate of the life expectancy at age *x* for the year of interest *t* using weekly data up to the week w_l. In a nutshell, if using the same amount of information and the same input data the estimates of the life expectancy in the benchmark period deviates on average by a certain quantity from the corresponding final estimates, that same average difference is used to correct the early estimate for the year of interest *t*. By doing so, it is implicitly assumed that the distribution of deaths in the year *t* will mimic those observed in the benchmark period. The choice of the benchmark period is therefore fundamental to obtain early estimates as accurate as possible. In time of 'regular' mortality, the benchmark period should be chosen such that mortality does not present atypical values; in times of peculiar mortality (such as during the COVID-19 pandemic), a correction based on a 'regular-mortality' benchmark period is likely to be less accurate. However, when the weekly deaths data cover the entire year of interest, in both cases (periods of regular and irregular mortality) the correction returns enough accurate estimates. In fact, the risk of biased estimates applies as well - although to a lesser extent - using data covering an entire year because of several reasons: mismatch between calendar year and ISO weeks, potential

different data production processes, and inaccuracies in the coverage coefficients. Nonetheless, the correction based on the formula [7] above allows obtaining estimates very close or even identical to the past estimates in the benchmark period.

4. Application to the years 2022 and 2023

The method proposed above has been applied to obtain early estimates of life expectancy at birth in the years 2022 and 2023 in Europe, for which official statistics are not yet available at the time of this exercise. These two years represents the two typical cases of application, that is when weekly deaths data cover the entire year of interest (here 2022) and the case of incomplete data because the period of interest has not yet elapsed (here 2023). In this latter case, the proposed method has in fact the feature of nowcasting, which could be used as well as input to projections exercises.

The computations are run in two phases: the first for the year 2022 using data covering the entire years; the second for the year 2023, where for each country data are limited to the latest week with available data (Table 1). Therefore, if a country has weekly deaths data for the year 2023 until week 26, to obtain the estimates for the year 2023 data for all the other years will also be limited to the same week. At the time of this exercise, the data cover between 5 and 7 months of the year 2023 and about half of the countries have not yet reached the mid-year as for the weekly data availability, which should induce a certain caution on the nowcasting estimates. Data availability for the European Union is limited to the week 22, the lowest week among EU Member States. Estimates for the European Union as a whole are computed without Ireland for the life expectancy for both sexes, and without Germany and Ireland for the sex-specific life expectancies. Higher caution must therefore be applied using the EU life expectancy estimates, particularly those broken down by sex.

For all the countries here considered, the population size on 1 January is available from the Eurostat database until the year 2023 (Eurostat 2023c). An estimate of the population size on 1 January 2024 is instead available from the Eurostat population projections 2023-based (Eurostat 2023d). Estimates of the population exposure to the risk of dying are thus computed using the formula [2].

As benchmark period, it is here adopted the period 2015-2019. This benchmark period is not far from the years of interest, it is well covered by the data at disposal, it does not present excessive deviations from an expected "regular" mortality, and it is wide enough to absorb occasional mortality peaks through averaging. More sophisticated methods of selection of the benchmark period could be applied (e.g., based on time series correlations) but for the sake of the current exercise the criteria above are considered sufficient. Official life tables for the benchmark period are available from Eurostat (2023e), together with a description of their methodology (Eurostat 2023f).

Because weekly deaths data are by 5-year age group, the [4] becomes:

[8] $$ {}_5^s q_x = \frac{5 \cdot {}_5^s M_x}{1 + (5 - {}_5^s a_x) \cdot {}_5^s M_x} \quad \forall x \in \{0, 5, .., 80, 85+\} $$

The parameter ${}_5^s a_x$ is the sex- and age-specific average time spent living by those dying in the 5-year time interval. In the methodology of the life table currently applied by Eurostat, the single age values for this parameter are:

[9] $$ {}_1^s a_x = \begin{cases} 0.2 & x = 0 \\ 0.5 & x \epsilon [1, \dots, 84] \end{cases} \quad \forall s \ . $$

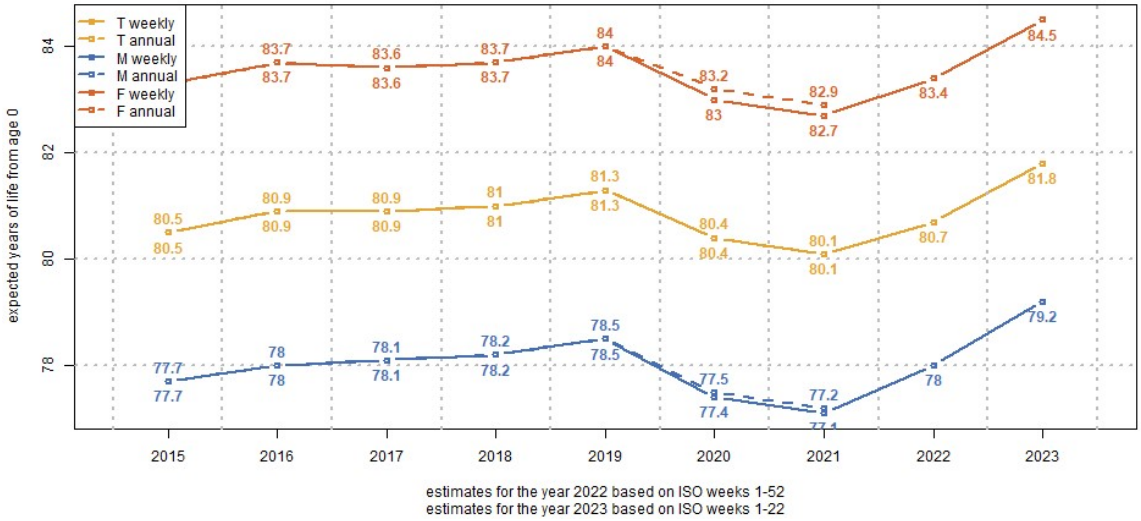
The different value for ${}_1^s a_0$ is meant to capture the higher mortality in the first weeks of life, which leads to a shorter average time lived by those dying at age zero. For abridged life tables, it

is common practice to keep the mortality at age zero separated from the other age groups because of its peculiarity. The available weekly deaths data do not allow such a distinction and thus there is no specific treatment here for the age zero; however, by using the formulas [5] and [6], it is derived a value for ${}_5^s a_0$ that aims to take into account the higher mortality at age zero as in the Eurostat life table methodology. If theoretical deaths were equally distributed within a 5-year interval (that is, ${}_1^s d_{x+i}^t = {}^s d_x^t$), it would follow:

$$ {}_5^s a_x^t = \frac{{}^s d_x^t \cdot \sum_{i=0}^{4} {}_1^s a_{x+i}^t + 10 \cdot {}^s d_x^t}{5 \cdot {}^s d_x^t} = \begin{cases} \frac{2.2}{5} + 2 = 2.44 & x = 0 \\ \frac{2.5}{5} + 2 = 2.50 & x \epsilon [5,10, \dots ,80] \end{cases} \quad \forall s, $$

where t is the year and $n = 5$ the width of the age interval. The set of parameters for $x > 0$ would then take a value commonly applied in 5-year abridged life tables, but the assumption of equal distribution of deaths would return unlikely values for the first age group (0-4 years). Taking into account the unequal distribution of the theoretical deaths within the 5-year interval returns a better estimate of the total person-years aligned with official Eurostat estimates. The estimation of the average persons-years ${}_5^s \bar{a}_x$ lived by those dying in the interval over the benchmark period 2015-2019 returns the values reported in the supplementary material Annex 1. With these inputs it is possible to compute the early estimates of the life expectancy at any age for the years 2022 and 2023, which are then corrected using the formula [7]. For instance, for Austria the uncorrected early estimates are consistently higher than the final official estimates, and vice versa for Italy, but the issue is solved by applying the correction above. The results for life expectancy at birth are reported in the Table 4 and shown with the time series from 2015 for each country in the Annex 2 and, as example, in the Figure 1 for the European Union.

Figure 1: life expectancy at birth, EU(*), 2015-2023



(*): excluding IE for T, and DE and IE for F and M

The circles along solid lines in the graphs are the life expectancies estimated using the method proposed here ("weekly"), while those along dashed lines are from the regular Eurostat life tables using annual deaths data ("annual"), available for the years 2015-2021. When a dashed line with the regular Eurostat estimates is not visible, it indicates a full overlap by the solid line of the early estimates. It should be kept in mind that estimates for the European Union are obtained with partial data, particularly for the breakdown by sex (about one fifth of the observations are missing), which suggests a higher caution.

For all the countries, the graphs reported in the Annex 2 show also (see the full dots) the estimates of life expectancy for 2022 and 2023 as from the latest Eurostat projections (Eurostat

2023g)[4]. In general across countries, estimates for the year 2022 point to a recovery of life expectancy, which is basically achieved in the year 2023. In 2022, early estimates indicate that there were still 10 countries whose life expectancy at birth had decreased from the previous year. By 2023, all countries would have a positive change in life expectancy at birth as compared to the previous year. Compared to the pre-pandemic period (year 2019), in 2023 only Iceland would have a lower life expectancy (but by 0.1 years only), a signal that the recovery in life expectancy at birth from the levels in years when mortality was characterized by the COVID-19 pandemic is essentially completed.

[4] The life expectancy projected by Eurostat is only available for each country and for each sex. Therefore, neither values for the European Union nor for both sexes are reported in the graphs.

Table 4: early estimates of life expectancy at birth, 2022 and 2023, as of end July 2023

	Both sexes		Women		Men	
	2022	**2023**	**2022**	**2023**	**2022**	**2023**
European Union (*)	80.7	81.8	83.4	84.5	78.0	79.2
Austria	81.3	82.0	83.7	84.4	78.9	79.6
Belgium	81.8	82.7	83.9	84.9	79.7	80.5
Bulgaria	74.4	76.4	78.1	80.1	70.8	72.7
Croatia	77.7	78.9	80.8	82.1	74.6	75.7
Cyprus	81.8	83.7	83.7	85.6	79.9	81.8
Czechia	79.1	80.4	82	83.3	76.2	77.3
Denmark	81.3	82.1	83.2	84	79.5	80.2
Estonia	78.2	79.3	82.4	83.6	73.7	74.6
Finland	81.2	82.2	83.8	84.7	78.7	79.6
France	82.3	83.4	85.2	86.3	79.4	80.5
Germany	80.7	81.5	NA	NA	NA	NA
Greece	80.7	81.8	83.3	84.3	78.1	79.2
Hungary	76.1	77.2	79.5	80.6	72.7	73.8
Iceland	82.1	83.1	83.4	85.2	81.0	81.2
Italy	83.0	84.2	85.0	86.2	80.9	82.1
Latvia	74.8	76.0	79.6	81.1	69.8	70.7
Lithuania	76.0	77.6	80.3	82.1	71.4	72.8
Luxembourg	83.0	83.1	85.3	85.1	80.9	81.3
Malta	82.7	84.2	84.8	86.0	80.7	82.3
Netherlands	81.8	82.2	83.2	83.8	80.3	80.7
Norway	82.6	83.5	84.2	85.0	81.0	82.0
Poland	77.4	78.9	81.3	82.7	73.6	75.1
Portugal	81.7	82.7	84.5	85.6	78.8	79.6
Romania	75.3	76.9	79.3	80.8	71.5	73.1
Slovakia	77.2	78.8	80.6	82.0	73.7	75.5
Slovenia	81.3	82.4	84.1	85.4	78.6	79.5
Spain	83.2	84.3	85.9	86.9	80.4	81.6
Sweden	83.1	83.9	84.8	85.6	81.5	82.2
Switzerland	83.7	84.9	85.6	86.6	81.8	83.1

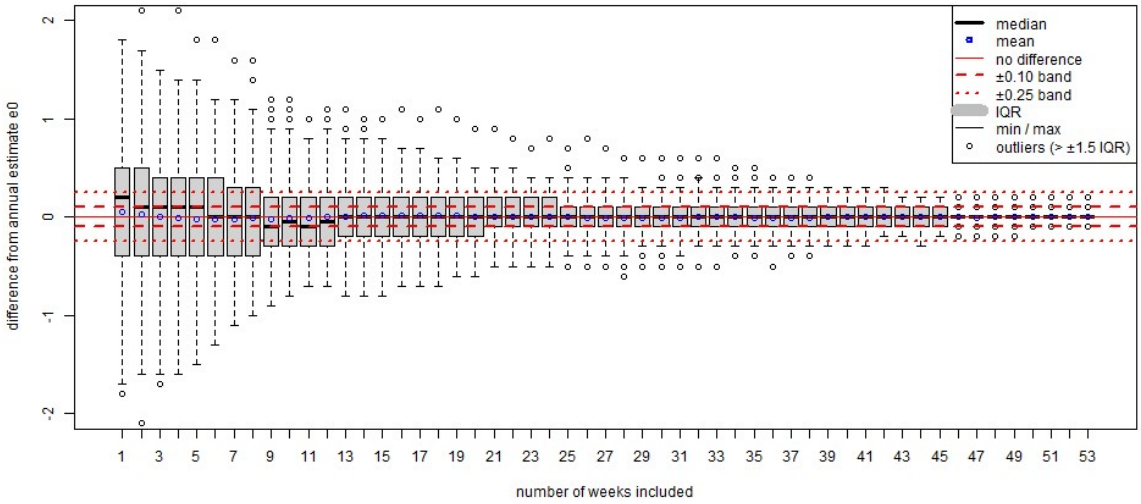
(*): excluding Ireland for both sexes, and Germany and Ireland for women and men

5. Discussion

While in the case of complete data coverage for a year of interest, differences between early estimates and final official estimates are to be found in *ex-post* data revision and – to a much lesser extent – in the difference in outcomes between abridged and complete life tables (the former being used for the early estimation and the latter for the official statistics), a question is how reliable are early estimates based on incomplete information. Intuitively, the larger the time interval with available weekly deaths, the closer the early estimate should be to the official estimate. An empirical assessment of the speed of convergence of the early estimates to the official estimate has been obtained by applying the method to all the years with fully available data. This exercise has been applied separately to years with regular mortality patterns and to the years affected by the COVID-19 pandemic. The former case represents a sort of ideal situation, where the mortality patterns in the years of interest are well associated with the benchmark period; the latter is instead an example of worst-case scenario, as the mortality patterns in the years of interest are very different from those of the benchmark period. The Figure 2 and Figure 3 show the boxplots of the differences between early estimates of life expectancy at birth for both sexes and the official estimates based on annual data, computed week by week as data availability increases. For each week there are respectively 150 and 60 observations (combinations of country and year). While on average the early estimates are since the beginning very close to the official estimates, there is a variability that decreases as more weekly data become available. In periods of regular mortality (Figure 2), 50% of the estimates in this example (those within the interquartile range – IQR) are already within a range of ±0.25 years starting from week 13, which is when data covering until the end of March are available. When the data availability encompasses the first 25 weeks of the year (which is before the mid-year), that range shrinks to ±0.10 years. By the week 43, for the first time all the early estimates are within a range of ±0.25 years from the official estimates, and with three more weeks of data

availability (thus with data covering until about mid-November) they are basically equal to the official estimates, apart a few outliers.

Figure 2: boxplot of differences between early and official estimates of life expectancy at birth, 30 European countries, 2015-2019 (regular mortality years), both sexes

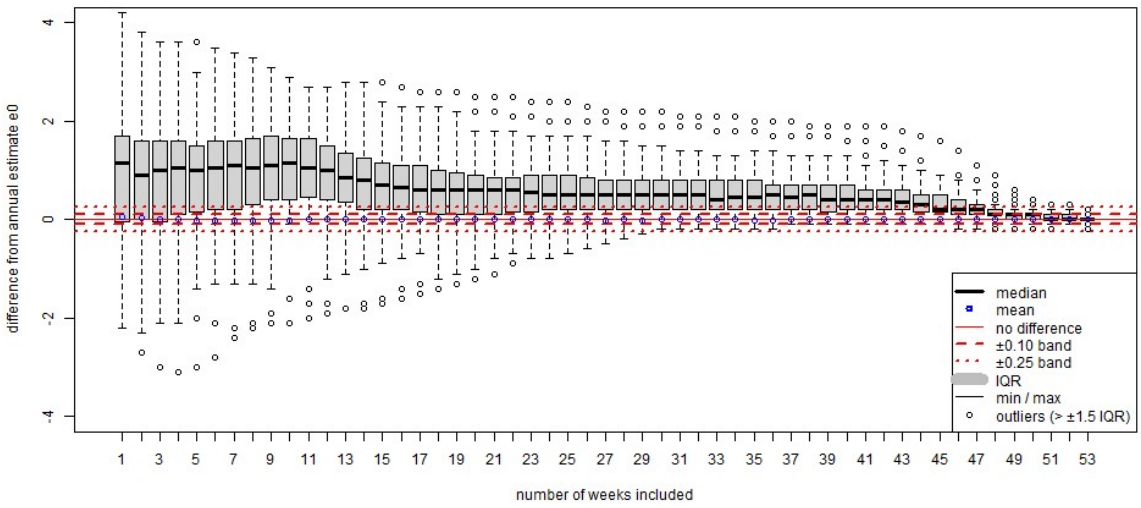


Note: three outliers in week 1 out of frame.

When the same benchmark period is used in times of irregular mortality (Figure 3), higher caution is needed. Data covering until beginning of December are required to have all early estimates for both sexes within a range of ±0.10 years from the official estimates. In fact, the extent of the differences shrinks over time, but most of the early estimates remains biased until an almost complete set of data is available. However, early estimates would still allow a gain of several months in the release of life expectancy estimates. The right side of Figure 3 also shows

that when the complete set of weekly data is of good quality, differences from official estimates are within a very restricted range of values regardless of the similarity of mortality patterns.

Figure 3: boxplot of differences between early and official estimates of life expectancy at birth, 30 European countries, 2020-2021 (pandemic years), both sexes



Note: two outliers in week 1 out of frame.

If the conclusions from this analysis of the speed of convergence hold and there are no forthcoming surprises in mortality patterns, then early estimates for 2022 should be very close to the final official estimates of life expectancy at birth and a good number of the early estimates for 2023 for both sexes reported in the Table 4 should be within a restricted range from the final estimates.

Quality of the input data is an important element for the accuracy of the early estimates. Considering that weekly deaths data are usually produced with high timeliness and perhaps with statistical processes different from those applied for annual deaths data, it is possible that weekly

data might be inaccurate, particularly those referring to the most recent weeks. In order to counterbalance such a risk, data producers may accompany the release of weekly deaths data with a coefficient estimating the extent of their coverage of the events, which may depends also on the freshness of the data (data referring to earlier weeks should be more accurate). This is in fact the approach taken for the data disseminated by Eurostat, where the coverage coefficients are provided in the metadata and here applied for an *ex-ante* correction of the input data. The use of correction coefficients applied to input data should not be neglected, as otherwise the information - particularly if drawn from the latest weeks - has a higher risk of being misleading.

In all countries the total number of weekly deaths has been higher in the year 2022 than in the pre-pandemic years and in many even higher or almost than in one or both of the COVID-19 pandemic years. Ballister *et al.* (2023) have estimated in 70,000 excess deaths the impact of heatwaves in Europe in the same year. At national level, an unusual peak of deaths in the summer of 2022 is noticeable particularly in Southern European countries like Italy, Portugal, and Spain, and in the European Union as a whole. Other countries report numbers of weekly deaths higher than in the pandemic years in the weeks between end 2022 and beginning 2023 (Denmark, Finland, Germany, Luxembourg, and Norway) or in the end of February 2022 (Estonia and Iceland).

Early estimates based on incomplete data should be taken with higher caution when the observed patterns are deviating from those in the benchmark period. A visual inspection of the cumulative time series of weekly deaths can reveal features of interest for speculations about mortality in the remaining weeks of the year 2023. The first weekly data signals that a number of deaths higher than usually may occur (again) in Austria, Denmark, Finland, France, Germany, Iceland, Luxembourg, the Netherlands, and Norway. It also remains the incognita of the heatwaves that have hit Southern Europe in the summer 2023. Further unexpected mortality crises may occur later in the year, breaking the patterns observed so far. However, these factors would affect any forecasting method based on past regularities.

The early estimates of life expectancy at birth in 2022 in the European countries are lower on average by 0.2-0.3 years than the values projected in the latest Eurostat projections (Eurostat 2023g), but they are larger by 0.6 years in 2023. If the early estimates for 2023 would be confirmed, the projected life expectancy would have underestimated the extent of recovery from the pandemic levels, with a potential consequence for the values projected further in the future as they may be too conservative. It is however, at the time of this writing, too early to draw any conclusion on this regard because, with the current data availability for 2023 limited to data covering not more than 7 months, early estimates may still fall outside a range of ±0.25 years from the final official estimate of life expectancy at birth. Nonetheless, as this example shows, early estimates can be a valuable additional input to mortality nowcasting and short-term projections, as they may signal changes of trend or unusual levels that might be missed otherwise.

The use of weekly deaths data is typical of health surveillance exercises, where the early detection of mortality anomalies is the basic aim, and therefore the week is the time periodicity with which death data can be usually found in a timely manner. For the early estimation of life expectancy, turning to monthly deaths would be equally good, as the most important feature is the timeliness of the data. Actually, revolving to monthly data or alike would avoid the mismatch between ISO weeks and calendar year and the complicacies deriving from years with an extra week. The same reasoning applies to quarterly or biannual data, even though in the latter the advantage of progressive refinement of the early estimates as data availability approaches the end of the year would be lacking.

6. Conclusions

The method for the early estimation using timely weekly deaths data here presented allows to produce accurate estimates of life expectancy at any age with a temporal gain of more than one year as compared to the release of internationally comparable official statistics. While here

developed with regard to data and statistics disseminated by Eurostat and therefore applied to European countries, the method is valid in any other situation where timely infra-annual data on deaths counts are available – the timelier the data, the larger the temporal gain. If the weekly deaths data does not cover the entirety of the year of interest, the method returns a nowcast of the life expectancies, whose accuracy improves as more data becomes available and with a speed depending on the similarity of the selected benchmark period with the latest mortality pattern. When the weekly death data covers the entire year of interest or almost, the method is accurate regardless of the selected benchmark period, as slight differences in the life table methodology and in the time period effectively covered (mismatch between ISO weeks and calendar year) do not have a noticeable impact. Large differences between early and official estimates of life expectancies are in general attributable to quality of the input weekly deaths data and/or unexpected deviation in the mortality patterns not yet covered by observation. However, issues of input data quality, such as under-coverage of death events due to the timeliness of the reporting, can be dealt with correction coefficients to be applied to the original weekly death data.

The application of the method for early estimation to weekly deaths data on 30 European countries available from Eurostat as of end of July 2023 and covering until the first 7 months of 2023 at the best has allowed to compute a first estimate of the life expectancy at birth for the years 2022 and 2023, this latter being a nowcast. Results show that the recovery from the mortality crisis in the COVID-19 pandemic years is likely to be completed in 2023, provided that no unusual mortality peak will occur in the remaining part of the year 2023 (e.g., due to heatwaves or severe influenza). With data covering about half of the year 2023, the European country with the highest life expectancy at birth is expected to be Spain for women (86.9 years) and Switzerland for men (83.1 years). Despite the lack of data from Germany and Ireland, results are also available for the European Union as a whole and show an increase to 80.7 years in 2022 for both sexes, further to 81.8 in 2023.

7. References

Ballester, J., Quijal-Zamorano, M., Méndez Turrubiates, R.F. *et al.* (2023). Heat-related mortality in Europe during the summer of 2022. *Nat Med* 29:1857–1866. https://doi.org/10.1038/s41591-023-02419-z

Eurostat (2023a). Deaths by week, sex and 5-year age group (online data code: demo_r_mwk_05) [electronic resource]. Luxembourg: Eurostat. https://ec.europa.eu/eurostat/data/database. Last accessed on 10 August 2023.

Eurostat (2023b). Deaths by week – special data collection (online data code: demomwk). Luxembourg: Eurostat. https://ec.europa.eu/eurostat/cache/metadata/en/demomwk_esms.htm. Accessed on 10 August 2023.

Eurostat (2023c). Population on 1 January by age and sex (online data code: demo_pjan) [electronic resource]. Luxembourg: Eurostat. https://ec.europa.eu/eurostat/data/database. Last accessed on 10 August 2023.

Eurostat (2023d). Population on 1st January by age, sex and type of projection (online data code: proj_23np) [electronic resource]. Luxembourg: Eurostat. https://ec.europa.eu/eurostat/data/database. Last accessed on 10 August 2023.

Eurostat (2023e). Life table by age and sex (online data code: demo_mlifetable) [electronic resource]. Luxembourg: Eurostat. https://ec.europa.eu/eurostat/data/database. Last accessed on 10 August 2023.

Eurostat (2023f). Description of the Eurostat method for the calculation of the life expectancies at all ages. Luxembourg: Eurostat. https://ec.europa.eu/eurostat/cache/metadata/Annexes/demo_mor_esms_an1.pdf.

Eurostat (2023g). Projected life expectancy by age (in completed years), sex and type of projection (online data code: proj_23nalexp) [electronic resource]. Luxembourg: Eurostat. https://ec.europa.eu/eurostat/data/database. Last accessed on 10 August 2023.

International Organization for Standardization (2019). ISO 8601-1:2019 Date and time — Representations for information interchange — Part 1: Basic rules. https://www.iso.org/standard/70907.html

Preston, S.H., P. Heuveline, and M. Guillot (2001). *Demography: measuring and modelling population processes*. Blackwell Publishing.