MAX PLANCK INSTITUTE
FOR DEMOGRAPHIC RESEARCH

Konrad-Zuse-Strasse 1 · D-18057 Rostock · Germany · Tel +49 (0) 3 81 20 81 - 0 · Fax +49 (0) 3 81 20 81 - 202 · www.demogr.mpg.de

MPIDR Working Paper WP 2024-029 I September 2024
https://doi.org/10.4054/MPIDR-WP-2024-029

Arriaga meets Kitagawa. Life expectancy decomposition with population subgroups

Timothy Riffe I tim.riffe@ehu.eus
Rustam Tursun-zade
Sergi Trias Llimós

This working paper has been approved for release by: Marcus Ebeling (ebeling@demogr.mpg.de), Deputy Head of the Laboratory of Population Health.

Arriaga meets Kitagawa. Life expectancy decomposition with population subgroups

Dr. Timothy Riffe[1,2,3*], Rustam Tursun-zade[1*] and Dr. Sergi Trias Llimós[4*]

[1*]OPIK, Department of Sociology and Social Work, University of the Basque Country (UPV/EHU), Leioa, Spain, 48940, Bizkaia, Barrio Sarriena s/n.
[2]Ikerbasque (Basque Foundation for Science).
[3]Laboratory of Population Health, Max Planck Institute for Demographic Research.
[4] Centre d'Estudis Demogràfics, Centres de Recerca de Catalunya (CERCA), Carrer de Ca n'Altayóo, Edifici E2, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain.

*Corresponding author(s). E-mail(s): tim.riffe@ehu.eus;

Abstract

Background: An Arriaga decomposition partitions differences in life expectancy into contributions from mortality rate differences in each age. A Kitagawa decomposition partitions a difference between two weighted means into effects from differences in structure and from differences in each element of the weighted value.
Methods: Life expectancy differences between like-defined subpopulations can be decomposed using the Arriaga method, or a variant thereof. The results of subgroup-specific decompositions can be weighted together using the value component from a Kitagawa decomposition of total life expectancy, given that subpopulations are blended in the radix age. The composition component of the same Kitagawa decomposition gives the effect of subgroup composition on total life expectancy differences.
Results: Notable properties of the method include: (i) it accommodates any number of subpopulations, (ii) it easily incorporates cause-of-death information, and (iii) composition is considered in the radix age. We apply the method to Spanish cause- and education-specific data.

Conclusions: This method can further disentangle the effects of mortality and composition differences, helping to explain or clarify paradoxes and secular change. We give both R code and spreadsheet implementations of the method.

Keywords: Decomposition, Mortality, Cause of death, Population Structure, Mortality Inequalities

Introduction

Population-level differences in life expectancy have been widely analyzed using decomposition methods to understand the impacts of mortality differences in age groups (e.g., Arriaga, 1984; Pollard, 1988), or Kitagawa (1955) to understand better the impact of compositional differences in social strata (Kitagawa, 1977). It has long been known among demographers that improvements in social welfare have the potential to bring large scale improvements in mortality conditions even net of medical improvements (Kitagawa & Hauser, 1968). However, very few demographers have ventured to account for both age-specific effects and compositional effects at the same time (Hendi & Ho, 2021; Su, van Raalte, Aburto, & Canudas-Romo, 2024; Su, Welsh, Korda, & Canudas-Romo, 2023; Torres, Canudas-Romo, & Oeppen, 2019),. Disentangling the compositional effect within a conventional age group decomposition would improve our understanding of life expectancy differences between populations.

The life expectancy decomposition method proposed by Arriaga (1984) separates a change in life expectancy into contributions due to mortality changes at different ages. This technique is designed to be practical in that it is framed in terms of lifetable columns expressed in discrete age groups. Two well-known properties of the method are that mortality changes in different ages do not need to be proportional and that the derived contributions sum exactly to the observed difference in life expectancy. A documented but

lesser-known property of this method is its asymmetry: the absolute values of age-specific contributions depend on whether we compare population one with two, or vice versa. This is why the method is described in terms of changes rather than group differences; the direction of time is clear, so we always decompose an earlier period against a later period. Importantly, the method is designed to work with homogeneous populations, meaning a single lifetable represents each population. Let's call this method and the variant of it that we later apply the "Arriaga method". The Arriaga method has different extensions available to separate the effects of different causes of death within age groups. We seek to extend this method to account for lifetables that are composed of subpopulations with different mortality.

The decomposition method proposed by Kitagawa (1955) separates differences in a weighted mean into contributions from differences in weights (structure) and the value being weighted. Often, the weighted values are rates, but in our case, the weighted values will be the life expectancies of subgroups. This method is well-known to partition differences owing to "structure" and "value" components, which sum exactly to the observed difference in weighted means. The individual elements of the value being weighted (in our case, life expectancies) have identifiable effects. It is not well-known that the individual elements of the structure component do not have identifiable effects. Instead, the structure effects should be summed up as a total marginal effect due to composition differences. We refer to this method as the "Kitagawa" method.

We propose to treat the problem of life expectancy decomposition by age, where a total life expectancy is calculated as the weighted mean of the expectancies of subpopulations. This is only practical when each population is composed of like-defined subpopulations. For example, we may decompose the life expectancy difference between two countries considering differences in education structure, since life expectancy varies by education level (Mackenbach et al., 2019; Trias-Llimós, Spijker, Blanes, & Permanyer, 2023). Or,

one might wish to calculate a national life expectancy as the weighted average of its regions at two points in time since regional weightings and inequalities may have changed over time. In such scenarios, a decomposition should yield the contribution to the overall life expectancy difference stemming from rate differences in each age (and potentially causes of death) within each subpopulation while giving the overall effect of compositional differences. We propose a straightforward analytic method to perform such decompositions.

In short, differences in life expectancy between subpopulations can be decomposed using the Arriaga method, one of its variants, or by any other life expectancy decomposition method that is acceptably precise with discrete data (Andreev, Shkolnikov, & Begun, 2002; Horiuchi, Wilmoth, & Pletcher, 2008; Ponnapalli, 2005). In a second step, the Kitagawa method is then used to rescale group-specific life expectancy decompositions, yielding an extra component telling us the importance of compositional differences. We justify this second-pass rescaling and demonstrate that the results of this procedure align with those of a linear integral reframing of the problem, per Horiuchi et al. (2008). We then apply these methods to decompose the sex gap between Spanish men and women in remaining life expectancy at age 35, considering differences in composition by education attainment around the same age and considering cause-specific mortality rates.

Method

Notation

We use the following variables and scripting, most of which are lifetable columns:

- ${}_nm_a$ mortality rate in the age interval $[a, a+n)$.
- ℓ_a lifetable survivorship at exact age a.
- ${}_nL_a$ lifetable person-years lived in the interval $[a, a+n)$.
- T_a total lifetable person-years lived beyond age a.

- e_a remaining life expectancy at exact age a.
- π_a^s the population fraction for subgroup s in age a.

We use the superscript s to index subpopulations comprising the total population, and the superscript "1" to typically indicate the population with the lower value of e_0, and "2" for the higher value, such that their difference, $\Delta = e_0^2 - e_0^1$, is always positive. Causes of death may be indicated with a c subscript, i.e. $m_{a,c}$. Throughout the manuscript, we assume single-year age groups and omit age interval (left-side) subscripts (n) where they would otherwise be due.

Averaging life expectancy

Vaupel (2002) distinguishes between two major approaches to calculate life expectancy for a total population composed of observed subgroups. The dominant approach is based on averaging the mortality rate in each age by aggregating deaths and exposures over all subpopulations and calculating the rate per equation (1):

$$m_a = \frac{\sum_i D_a^i}{\sum_i E_a^i} = \sum_i \pi_a^i m_a^i \quad , \tag{1}$$

where D and E stand for deaths and exposures, respectively. This is the standard method used implicitly or explicitly by national statistical offices, the Human Mortality Database (Wilmoth et al., 2021), or the World Population Prospects (UN Population Division, 2022), among others. Vaupel (2002) called this approach the *current rates* perspective, which treats a heterogeneous population as a homogeneous one.

A second approach derives independent subgroup-specific lifetables and then weights life expectancies together based on an initial mixing composition to obtain the total life expectancy per equation (2):

$$e_0 = \sum_s \pi_0^s \cdot e_0^s \quad . \tag{2}$$

This method aligns with Vaupel's *current conditions* approach, at least to the extent that heterogeneous conditions are accounted for using observed strata. This second approach is commonly applied in multistate models with strata combined according to mixed initial conditions to obtain the total population (Caswell, de Vries, Hartemink, Roth, & van Daalen, 2018; Caswell & van Daalen, 2021), occasionally also with standard lifetables (Gupta, 1988; Muszynska-Spielauer & Riffe, 2022), or when calculating between-within decompositions of variance or other similar summary measures (Riffe, van Raalte, & Dudel, 2024; Seaman, Riffe, & Caswell, 2019). If the composition is more favorable in younger than in older ages (for example when younger generations have higher average educational attainment), the life expectancy obtained with (2) will be higher than one based on (1). This is the total life expectancy approach that our proposed method is designed to decompose.

Kitagawa decomposition

Equation (2) treats the total life expectancy as a weighted average, allowing us to precisely decompose a difference in e_0, i.e. where $\Delta = e_0^2 - e_0^1$, using the formulas from Kitagawa (1955). We presume that population 2 has the higher life expectancy of the two, such that Δ is positive. Specifically, equation (3) gives an overall effect of differences in composition:

$$\Delta^{(\pi)} = \sum_s \left(\pi^{s,2} - \pi^{s,1}\right) \cdot \overline{e_0^s} \quad , \tag{3}$$

where

$$\overline{e_0^s} = \frac{e_0^{s,1} + e_0^{s,2}}{2} \tag{4}$$

This result is widely known. Note that the the composition effect must be summed this way because group-specific composition (structure) effects are not well-identified and therefore cannot be interpreted in isolation. Equation (5) gives the subgroup-specific effects of differences in life expectancy:

$$\Delta^{(e_0^s)} = \left(e_0^{s,2} - e_0^{s,1}\right) \cdot \overline{\pi^s} \quad , \tag{5}$$

where $\overline{\pi^s}$ is the average composition, analogous to equation (4), and where the superscripts are consistent with those used to calculate the total difference Δ.

Equation (6) states that the observed difference in total life expectancy is the sum of (i) a single component capturing the effect of compositional change and (ii) a set of components giving the contribution of each subgroup's life expectancy difference to the total life expectancy difference. This second effect could equivalently be called the *rate* effect because each life expectancy is ultimately a function of mortality rates.

$$\Delta = \Delta^{(\pi)} + \sum_s \Delta^{(e_0^s)}. \tag{6}$$

Symmetrical Arriaga decomposition

The *rate* effect as per (5) can be viewed as the *net* effect on total life expectancy differences resulting from the age-specific (potentially also cause-specific) differences between like subgroups, as isolated by various life expectancy decomposition methods. In this setting, the choice of method to derive age-specific effects for subgroup-specific changes in life expectancy is not crucial. The Arriaga (1984) decomposition approach is widely favored because it is designed for discrete data, it is analytic (implying computational efficiency), and because decomposition results sum exactly to the observed life expectancy difference. Since our application concerns sex differences rather than changes over time, we consider a *symmetrical* Arriaga decomposition. Since we also include cause-of-death information, we transform the symmetrical decomposition results into an implied sensitivity function. This exercise is repeated for each subgroup to decompose group-specific changes in life expectancy (indicated with the superscript ^{s} on each variable):

$$\Delta^s = e_0^{s,2} - e_0^{s,1} \tag{7}$$

In this equation Δ^s is the subgroup-specific (s) difference in life expectancy being decomposed, which is composed of age-specific contributions, $\overrightarrow{\Delta_a^s}$ or $\overleftarrow{\Delta_a^s}$ depending on whether we decompose from population 1 to 2 (forward) or from 2 to 1 (backward). The forward age-specific values $\overrightarrow{\Delta_a^s}$ can be calculated following Arriaga's decomposition method, consistent with Arriaga (1984) or the presentation in Preston, Heuveline, and Guillot (2000), as outlined in equation (8). We use a lifetable radix of 1, meaning $\ell_0 = 1$, to simplify the formula slightly.

$$\overrightarrow{\Delta_a^s} = \begin{cases} \ell_a^{s,1} \cdot \left(\frac{L_a^{s,2}}{\ell_a^{s,2}} - \frac{L_a^{s,1}}{\ell_a^{s,1}} \right) + T_{a+n}^{s,2} \cdot \left(\frac{\ell_a^{s,1}}{\ell_a^{s,2}} - \frac{\ell_{a+n}^{s,1}}{\ell_{a+n}^{s,2}} \right) & \forall a < \omega, \\ \ell_\omega^{s,1} \cdot (e_\omega^{s,2} - e_\omega^{s,1}) & \forall a = \omega. \end{cases} \tag{8}$$

Equation (8) represents the first pass of our symmetrical decomposition, while (9) is the second pass, which is identical except for swapping population superscripts.

$$\overleftarrow{\Delta_a^s} = \begin{cases} \ell_a^{s,2} \cdot \left(\frac{L_a^{s,1}}{\ell_a^{s,1}} - \frac{L_a^{s,2}}{\ell_a^{s,2}} \right) + T_{a+n}^{s,1} \cdot \left(\frac{\ell_a^{s,2}}{\ell_a^{s,1}} - \frac{\ell_{a+n}^{s,2}}{\ell_{a+n}^{s,1}} \right) & \forall a < \omega, \\ \ell_\omega^{s,2} \cdot (e_\omega^{s,1} - e_\omega^{s,2}) & \forall a = \omega. \end{cases} \tag{9}$$

Importantly,

$$\begin{aligned} -\Delta^s &= \sum_a \overleftarrow{\Delta_a^s} \\ &= e_0^1 - e_0^2. \end{aligned} \tag{10}$$

A symmetrical estimate of Δ_a^s is given by the sign-adjusted average of (8) and (9) for each age, again with a preference for a positive difference:

$$\Delta_a^s = \frac{\left(\overrightarrow{\Delta_a^s} - \overleftarrow{\Delta_a^s} \right)}{2}. \tag{11}$$

Repeat the symmetrical Arriaga steps to derive age-specific contributions, Δ_a^s, for each life expectancy difference between like-defined subgroups. To

take into account information on causes of death, it is best to derive an ad-hoc all-cause sensitivity function, s_a^s, by dividing the decomposition result by the mortality rate difference:

$$s_a^s = \frac{\Delta_a^s}{m_a^{s,2} - m_a^{s,1}} \quad , \tag{12}$$

and then multiply age-cause-specific mortality rate differences into s_a^s to obtain the subgroup-specific decomposition by age and cause per equation (13).

$$\Delta_{a,c}^s = s_a^s \cdot \left(m_{a,c}^{s,2} - m_{a,c}^{s,1}\right). \tag{13}$$

This way of accounting for causes of death is exactly additive. However, if any all-cause mortality rate difference is close to 0, equation (12) is vulnerable, in which case we advise deriving the sensitivity using a more robust (but less precise) approach (Riffe & Atance, 2024)

Rescale Arriaga results

To obtain the *net* impact of Δ_a^s (or $\Delta_{a,c}^s$) on overall life expectancy change we rescale using (14) to match the life expectancy components from equation (5).

$$\Delta_{a,c}^{s,\text{net}} = \Delta^{e_0^s} \cdot \frac{\Delta_{a,c}^s}{\Delta^s} \quad , \tag{14}$$

With the composition effect from equation (3) (Δ^π) and (14), we have all elements of the proposed decomposition:

$$\Delta = \Delta^\pi + \sum_s \sum_a \sum_c \Delta_{a,c}^{s,\text{net}}. \tag{15}$$

Application

We use data from Trias-Llimós et al. (2023) on sex-, age-, education-, and cause-specific death counts for individuals aged 35 and over for the years 2016-21 in Spain, which were obtained by request from the Spanish National Statistics Institute (INE). We use all-cause mortality rates for

each education- and sex-specific subpopulation from the same source. Educational attainment information is not recorded on death certificates, but INE adds this variable through multiple data linkages INE (2020), including municipal population registers ("Padrón") and the 2011 census. We redistributed deaths with missing education codes ($< 2\%$ of deaths) across the four educational groups within age, sex, and cause of death classes, proportional to deaths with non-missing education. We then grouped educational attainment into three categories: Low (primary education or less), Medium (secondary education), and High (tertiary or university education). We grouped causes of death from the original ICD10 4-digit codes into 17 main causes: Infectious (ICD-10 codes: A00-B99), Neoplasms (C00-D48), Blood (D50-D89), Endocrine (E00-E90), Mental (F00-F99), Nervous (G00-G99), Circulatory (00-I99), Respiratory (J00-J99), Digestive (K00-K93), Skin (L00-L99), Musculoskeletal (M00-M99), Genitourinary (N00-N99), Congenital (Q00-Q99), Ill-defined (R00-R94), External (chapters S, T, V, W, X and Y), Other causes (chapters H, O and P), and COVID-19 (U071-U072). Moreover, to separate the pandemic period we combined years into two time periods: 2016-2019 and 2020+. Our main results refer to 2016-2019, whereas extra results for 2020-2021 are given in the materials repository. We smoothed and graduated the original data from 5-year age groups to single ages using a generalized additive model (GAM), including a P-spline over age and population offsets (S. Wood, 2017; S.N. Wood, 2011). The GAM formula is given in equation (16), using the Quasi-Poisson family of models to account for possible overdispersion or Negative Binomial model in cases with a more substantial overdispersion:

$$\log(\mu_a) = \beta_0 + f(a) + ln(pop_a). \tag{16}$$

Results

We present empirical results for the decomposition of the sex difference in life expectancy in 2016-2019. Table 1 presents the components of the Kitagawa decomposition for males and females. In this context, π c represents the educational structure, while $e(35)$ denotes life expectancy at age 35 ("rate" element in original Kitagawa terms). The table also gives the corresponding education structure and life expectancy differences between sexes. The final element, $\Delta^{(e^s_{35})}$, is the decomposition result, which gives the education group contribution. The largest life expectancy gap is observed among individuals with Primary and Secondary education levels, and this gap is lower for those with Higher education.

Education	π^f	π^m	$\pi^f - \pi^m$	$e(35)^f$	$e(35)^m$	$\overline{e(35)}$	$e(35)^f - e(35)^m$	$\Delta^{(e^s_{35})}$
Primary	0.09	0.10	-0.01	48.21	42.47	45.34	5.74	0.54
Secondary	0.41	0.52	-0.11	50.29	44.52	47.40	5.78	2.69
Higher	0.50	0.38	0.12	51.30	46.92	49.11	4.38	1.92

Table 1 Elements of the Kitagawa decomposition, sex gaps in life expectancy in Spain, 2016-19

The corresponding difference in education-specific life expectancy is also illustrated in Figure 1. This figure additionally shows the differences observed for standard aggregated (non-stationary) versus radix-weighted life expectancy (stationary).

Figure 2 is complementary to Figure 1 and represents the education contribution to the $e(35)$ sex gap with the educational composition component contribution, about 3 months, being added to the figure.

Figure 3 illustrates the cause-specific contributions to the life expectancy gap. The majority of this gap is explained by differences in mortality from cancer, circulatory diseases, respiratory conditions, and external causes. Neoplasms alone contribute to almost two full years of life expectancy.

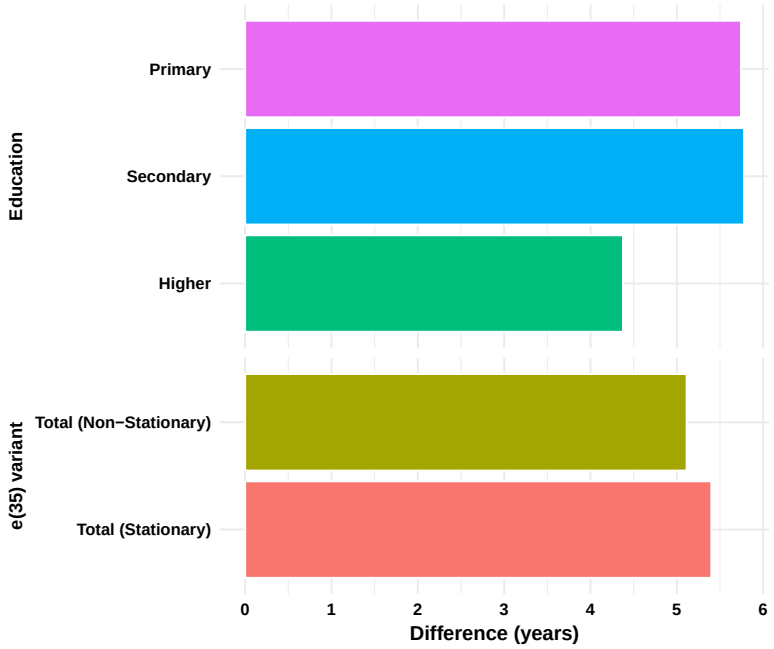

Fig. 1 The Female-Male difference in $e(35)$ by education and population type, Spain 2016-19.

The only categories where males have a slight advantage in mortality are musculoskeletal and other minor causes.

Further decomposition of cause-specific differences for age and education components of composed decomposition is shown in Figure 4. The majority of the difference in neoplasm mortality is concentrated in the older age group of 70-80 years. The circulatory component is observed in slightly younger ages and is more uniformly distributed between ages 60 and 80. In contrast, the respiratory component is more pronounced around age 85. The external causes of death primarily contribute to the difference in younger ages and since our data is bounded below by the age of 35, the full effect cannot be fully observed here.

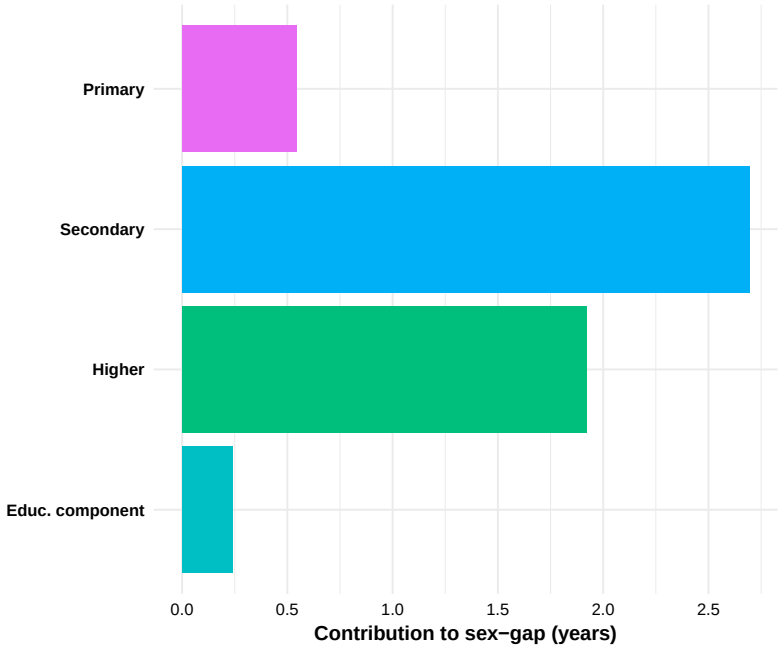

Fig. 2 Education structure contribution to total $e(35)$ sex gap, Spain 2016-19

Discussion

In this paper, we propose a method for decomposing the differences in life expectancy that accounts for the compositional effects of education and causes of death between two populations mixed in the radix. Deriving an analytical solution for decomposition offers several advantages over alternative methods, particularly in terms of simplicity, speed, and repeatability. Unlike the linear integral method Horiuchi et al. (2008), which requires programming expertise, our approach can be executed swiftly and efficiently, even within spreadsheet-like environments (see reproducibility repository for an example). The computational efficiency of our method allows for the calculation of confidence intervals using bootstrapping.

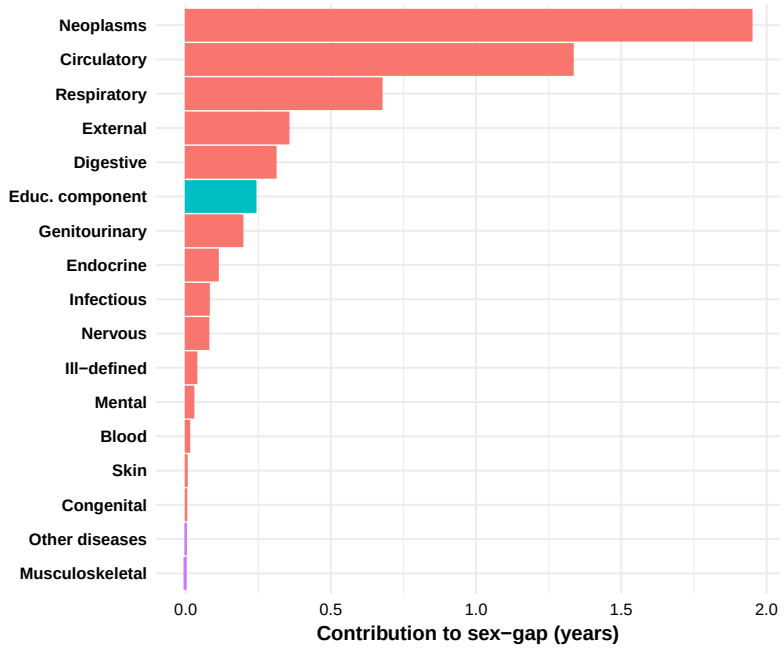

Fig. 3 Cause-specific contribution to sex difference in total $e(35)$, Spain 2016-19. The Education structure contribution is presented in a different colour

Using the Kitagawa decomposition approach (Kitagawa, 1955) to weigh together group-specific paired decompositions, our method can be generalized to address many decomposition problems involving structural compositional components, and it is not bound to mortality analyses. However, this post-weighting approach is most suitable for cases where populations are blended in a radix. In contrast, other approaches in the literature (Hendi & Ho, 2021; Su et al., 2024, 2023; Torres et al., 2019) decompose referring to the standard *current rates* (Vaupel, 2002) method of calculating life expectancy. These approaches treat the age pattern of group prevalence differently, in essence fixing group prevalence (weights) rather than making prevalence depend on mortality. We point out that mortality is often one of the major drivers of how group prevalence weights change over age.

When incorporating information on causes of death, a potential vulnerability of our method arises when the difference in mortality rates in the

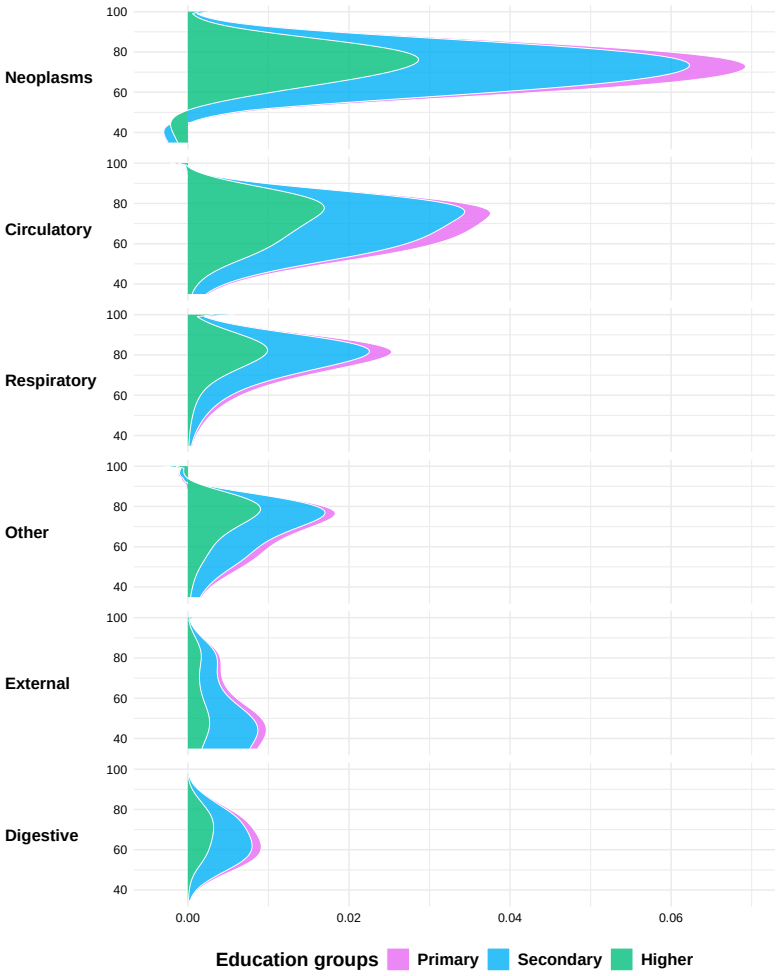

Fig. 4 The contributions of age, education, and major causes of death to the sex gap in total $e(35)$.

denominator of equation (12) approaches zero (compare with Box 4.3 of Preston et al. (2000)), rendering an implausible result. In this case, one may prefer to use a direct approximation or numerical estimate of the life expectancy sensitivity in equation (13).

Our empirical results indicate that the sex gap in life expectancy remains a significant issue in Spain across all educational strata. However, it is about 25% lower in the Higher education group. This can be attributed to several

factors, including females' lower engagement in risky behaviours (Byrnes, Miller, & Schafer, 1999; Cook & Bellis, 2001; Kritsotakis, Psarrou, Vassilaki, Androulaki, & Philalithis, 2016; Olson, Hummer, & Harris, 2017; Ross, Masters, & Hummer, 2012), better jobs and greater health awareness of males with higher education (Lawrence, 2017; McMahon, 2009; Ross et al., 2012; Ross & Wu, 1995). In terms of causes, neoplasms, circulatory diseases, and respiratory diseases account for the largest contributions to the sex gap. Given that the male disadvantage peaks around ages 65-85 (birth cohorts 1934-1951), this difference can be partly explained by the differences in smoking patterns between males and females within each subpopulation (Haeberer et al., 2020). Since our findings are based on education as a marker of socioeconomic status, they refer only to ages above 35. Therefore, we do not here measure the full power of socioeconomic status or different causes of death in driving the overall sex gap in life expectancy. Further studies could use our method to explain the sex gap using geography, socioeconomic status, and causes of death.

Our proposed combination of two analytical decomposition methods can handle the problem of a composed life expectancy, thereby giving an efficient framework for analyzing differences in life expectancy while accounting for population heterogeneity on fixed characteristics. We give both R code and spreadsheet implementations of the method. Our framework is straightforward to use and does not require high computational capacity, spreadsheet implementation in a spreadsheet-like environment while providing results comparable to those of the widely used Horiuchi method. The main advantages of the proposed method are its computational efficiency and ease of implementation over the commonly used Horiuchi method. Further efforts in alleviating the sex gap in mortality are required. The main cause of death contributions suggests that attention to the determinants of neoplasms, cardiovascular, and respiratory diseases could play a substantial role in reducing the sex gap in life expectancy.

References

Andreev, E.M., Shkolnikov, V.M., Begun, A.Z. (2002). Algorithm for decomposition of differences between aggregate demographic measures and its application to life expectancies, healthy life expectancies, parity-progression ratios and total fertility rates. *Demographic Research*, *7*, 499–522,

Arriaga, E.E. (1984). Measuring and explaining the change in life expectancies. *Demography*, *21*, 83–96, https://doi.org/10.2307/2061029

Byrnes, J.P., Miller, D.C., Schafer, W.D. (1999). Gender differences in risk taking: A meta-analysis. *Psychological bulletin*, *125*(3), 367,

Caswell, H., de Vries, C., Hartemink, N., Roth, G., van Daalen, S.F. (2018). Age$\times$ stage-classified demographic analysis: a comprehensive approach. *Ecological monographs*, *88*(4), 560–584,

Caswell, H., & van Daalen, S. (2021). Healthy longevity from incidence-based models: More kinds of health than stars in the sky. *Demographic Research*, *45*, 397–452,

Cook, P.A., & Bellis, M.A. (2001). Knowing the risk: relationships between risk behaviour and health knowledge. *Public health*, *115*(1), 54–61,

Gupta, P.D. (1988). A note on the consistency of life tables for total populations with those for their subpopulations. *Social biology*, *35*(1-2), 158–163,

Haeberer, M., León-Gómez, I., Pérez-Gómez, B., Téllez-Plaza, M., Pérez-Ríos, M., Schiaffino, A., ... Galán, I. (2020). Social inequalities in tobacco-attributable mortality in spain. the intersection between age, sex and educational level. *PLoS One*, *15*(9), e0239866,

Hendi, A.S., & Ho, J.Y. (2021). Immigration and improvements in american life expectancy. *SSM-Population Health*, *15*, 100914,

Horiuchi, S., Wilmoth, J.R., Pletcher, S.D. (2008). A decomposition method based on a model of continuous change. *Demography*, *45*, 785–801, https://doi.org/10.1353/dem.0.0033

INE (2020). *Asignación de nivel educativo, relación con la actividad laboral y ocupación a ficheros de movimiento natural de la población (mnp). método de obtención y advertencias a usuarios.* INE (Spanish National Statistics Institute) Madrid, Spain.

Kitagawa, E.M. (1955). Components of a difference between two rates. *Journal of the american statistical association*, *50*(272), 1168–1194, https://doi.org/10.1080/01621459.1955.10501299

Kitagawa, E.M. (1977). On mortality. *Demography*, *14*(4), 381–389,

Kitagawa, E.M., & Hauser, P.M. (1968). Education differentials in mortality by cause of death: United states, 1960. *Demography*, *5*(1), 318–353,

Kritsotakis, G., Psarrou, M., Vassilaki, M., Androulaki, Z., Philalithis, A.E. (2016). Gender differences in the prevalence and clustering of multiple health risk behaviours in young adults. *Journal of advanced nursing*, *72*(9), 2098–2113,

Lawrence, E.M. (2017). Why do college graduates behave more healthfully than those who are less educated? *Journal of health and social behavior*, *58*(3), 291–306,

Mackenbach, J.P., Valverde, J.R., Bopp, M., Brønnum-Hansen, H., Deboosere, P., Kalediene, R., ... others (2019). Determinants of inequalities in life expectancy: an international comparative study of eight risk factors. *The Lancet Public Health*, *4*(10), e529–e537,

McMahon, W.W. (2009). *Higher learning, greater good: The private and social benefits of higher education*. JHU Press.

Muszynska-Spielauer, M., & Riffe, T. (2022). Mortality rates, life tables and life expectancy under a stationary population composition. https://doi.org/https://doi.org/10.31219/osf.io/rnbf4

Olson, J.S., Hummer, R.A., Harris, K.M. (2017). Gender and health behavior clustering among us young adults. *Biodemography and social biology*, *63*(1), 3–20,

Pollard, J.H. (1988). On the decomposition of changes in expectation of life and differentials in life expectancy. *Demography*, *25*(2), 265–276,

Ponnapalli, K.M. (2005). A comparison of different methods for decomposition of changes in expectation of life at birth and differentials in life expectancy at birth. *Demographic Research*, *12*, 141–172,

Preston, S., Heuveline, P., Guillot, M. (2000). *Demography: measuring and modeling population processes*. Wiley-Blackwell.

Riffe, T., & Atance, D. (2024). coddecomp: Decompose life expectancy into contributions from rates in different ages and potentially also causes of death [Computer software manual]. Retrieved from https://github.com/timriffe/coddecomp (R package version 00.00.003, commit c68f942d8f400c6d04914b88a79d408eebbb10f3)

Riffe, T., van Raalte, A., Dudel, C. (2024). Lifeineq: Calculate indices of lifespan variation [Computer software manual]. Retrieved from https://github.com/timriffe/LifeIneq (R package version 00.05.01, commit b250c45286dfbb47a2e9d236bfe7191e4429cf32)

Ross, C.E., Masters, R.K., Hummer, R.A. (2012). Education and the gender gaps in health and mortality. *Demography*, *49*, 1157–1183,

Ross, C.E., & Wu, C.-l. (1995). The links between education and health. *American sociological review*, 719–745,

Seaman, R., Riffe, T., Caswell, H. (2019). Changing contribution of area-level deprivation to total variance in age at death: a population-based decomposition analysis. *BMJ open*, *9*(3), e024952,

Su, W., van Raalte, A., Aburto, J.M., Canudas-Romo, V. (2024). Subnational contribution to life expectancy and life span variation changes: Evidence

from the united states. *Demographic Research*, *50*, 583–624,

Su, W., Welsh, J., Korda, R.J., Canudas-Romo, V. (2023). Educational composition effect on the sex gap in life expectancy: A research note based on evidence from australia. *Population Studies*, 1–9,

Torres, C., Canudas-Romo, V., Oeppen, J. (2019). The contribution of urbanization to changes in life expectancy in scotland, 1861–1910. *Population studies*, *73*(3), 387–404,

Trias-Llimós, S., Spijker, J.J., Blanes, A., Permanyer, I. (2023). Age and cause-of-death contributions to educational inequalities in life expectancy and lifespan variation in a low-mortality country: A cross-sectional study of 1.67 million deaths in spain (2016–19). *SSM-Population Health*, *23*, 101461, https://doi.org/https://doi.org/10.1016/j.ssmph.2023.101461

UN Population Division (2022). *World population prospects 2022: Methodology of the united nations population estimates and projections.* (Tech. Rep. No. 6). New York: United Nations. Retrieved from https://population.un.org/wpp/Publications/Files/WPP2022$_{M}ethodology.pdf$

Vaupel, J.W. (2002). Life expectancy at current rates vs. current conditions: A reflexion stimulated by bongaarts and feeney's "how long do we live?". *Demographic Research*, *7*, 365–378,

Wilmoth, J., Andreev, K., Jdanov, D., Glei, D., Riffe, T., Boe, C., ... M, B. (2021). Methods protocol for the human mortality database. version 6. *University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock. URL: http://mortality.org [version 26/01/2021]*, ,

Wood, S. (2017). *Generalized additive models: An introduction with r* (2nd ed.). Chapman and Hall/CRC.

Wood, S.N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society (B)*, *73*(1), 3-36,

Acknowledgments. Thanks to Michael Lachanski for posing the question that led to the solution we present here, to Jonas Schoeley for first raising our attention to the composition problem with the Kitagawa method, and to the MPIDR Laboratory of Population Health for posing assorted questions in September 2023 that also led to this work.

Funding

TR acknowledges funding from *la Caixa Foundation* grant nr SR22-00502 from and from Spanish *Ministerio de Ciencia e Innovation* grant number PID2022-142762OA-I00. STL acknowledges research funding from the Ramon y Cajal program of the same Ministry (RYC2021-033123-I).

Conflict of interest

The authors declare no conflict of interest

Availability of data and materials

We cannot share the original data that this study is based on, but these can be requested either from the Spanish INE (https://www.ine.es) or from Trias-Llimós et al. (2023) for purposes of reproducing our results from the source data stage. We share smoothed and graduated single-age mortality rates necessary to reproduce all our results in an open OSF repository (anonymized for peer review)

https://osf.io/xnmv6/?view_only=047c4f8fe11c478a9aa71e8993b73e22

Code availability

All code for this paper is available in an open OSF repository (anonymized for peer review) https://osf.io/xnmv6/?view_only=047c4f8fe11c478a9aa71e8993b73e22