Untangling Mortality: The Role of Causes of Death in Shaping Dispersion around the Mode in High-Income Countries

Viorela Diaconu*, Virginia Zarulli*, and Stefano Mazzuco*

*Department of Statistical Sciences, University of Padova

September 2024

Abstract

Understanding the dispersion of deaths around the late modal age, M, sheds light on shifts in population health and mortality patterns. This dispersion can be divided into variation before the modal age, $SD(M-)$, and after the modal age, $SD(M+)$. While trends in $SD(M+)$ are well known, changes over time in SD(M-) are less understood. This study aims, first, to document and compare trends in $SD(M-)$ and $SD(M+)$ in several high-income countries since 1960. Second, to examine the role of causes of death in the dispersion of deaths at ages below M over time. To do so, we propose a novel decomposition of temporal differences in $SD(M-)$ into two parts: one due to changes in M and the other due to shifts in the age-at-death distribution. The second part of the proposed decomposition will be analyzed from a cause-of-death perspective to better understand epidemiological transitions at ages below M across countries. Preliminary results show a decline in $SD(M-)$ in all countries included, except for males in the United States and the United Kingdom, with the highest values in Eastern Europe. In contrast to the consistent decline in $SD(M+)$, $SD(M-)$ varies more between sexes and countries.

1 Introduction

Improvements in public health, medical advances, and socioeconomic conditions led to significant changes in the distribution of ages at death in high-income countries during the 20th century. Vaccination campaigns, improved hygiene and sanitation, widespread use of antibiotics, and increased access to health care led to impressive reductions in infant and child mortality in the first half of the 20th century. As fewer deaths occurred at younger ages, the distribution of ages at death shifted to adult and older ages. This shift continued in the second half of the 20th century as medical advances in chronic diseases, increased awareness of health risk behaviors such as smoking and physical inactivity, and improved nutrition and living conditions led to a reduction in premature mortality in early and late adulthood. As a result, deaths became increasingly concentrated in a narrower age range, with the majority of deaths occurring at older ages. As old-age mortality declines, the age-at-death distribution became heavily skewed to the left, with a long tail at younger ages and majority of deaths concentrated around the late modal age at death, i.e., the age at which the highest proportion of deaths occur.

Understanding the dispersion of deaths around the late modal age, M, offers valuable insights into shifts in population health and mortality patterns. The dispersion around the mode reflects how deaths are distributed around the most common age at death, providing key information on the variability of mortality. This dispersion can be decomposed into deviation before – $SD(M-)$ – and after the mode – $SD(M+)$. Several studies have analyzed trends in the dispersion of deaths around and above the mode in several high-income countries since the 1960s (Bergeron-Boucher et al., 2015; Canudas-Romo, 2008; Cheung et al., 2008, 2005; Kannisto, 2001, 2007; Lan Karen Cheung and Robine, 2007; Ouellette and Bourbeau, 2011). These studies showed a general decline in $SD(M)$ and $SD(M+)$, suggesting that deaths have become increasingly concentrated in a narrower old-age range, with the majority occurring at M. These studies have also showed that the increase in M was in general accompanied by a decline in $SD(M+)$.The strong dependence of $SD(M)$ on M is one of the limitations of the use of this metric for monitoring the spread of deaths at older ages. While changes in the dispersion of deaths at ages above M have been documented, there is little evidence on the levels and trends in $SD(M-)$ and their comparison to $SD(M+)$.

The aim of this study is twofold. First, to document trends in $SD(M-)$ and its relationship with $SD(M+)$ for several European and non-European countries since the mid-20th century. Second, to identify which causes of death have contributed to changes over time in the dispersion of deaths at ages below the mode. To do so, we propose a novel decomposition of temporal differences in $SD(M-)$ into two components: (1) one that depends on the change in M and (2) the other that depends on the change in the distribution of the ages at death. As noted above, $SD(M+)$ depends strongly the value of M. Consequently, so does $SD(M-)$: as M shifts to higher ages over time, reflecting improved survival conditions, the variability before the mode increases. This complicates the interpretation of $SD(M-)$ as standalone metric. A cause-of-death analysis of the change in the dispersion of deaths that has occurred solely from the second component, in our proposed decomposition, will provide a clearer view of epidemiological transitions across countries and regions.

2 Data & Methods

2.1 Data

Observed death counts (all-causes) and exposure to risk by single year of age and sex, covering the period 1970-2019 are taken from the Human Mortality Database (HMD).

Data on causes of death come from the Human Causes of Death (HCD) database and the WHO mortality database. These cause-specific mortality series provide information on observed deaths by five-year age group, cause of death, sex and calendar year. In these datasets, causes of death are classified according to the World Health Organization's International Classification of Diseases (ICD). As our study period spans from 1970 to 2019, it covers ICD-8, ICD-9, and ICD-10 revision of the International Classification of Disease. The introduction of ICD-10 resulted in a significant increase in the number of codes, due to enhanced classification detail. To ensure consistency of categories across the three ICD revisions and minimize the impact of successive revisions, we limited our analysis to 8 large groups of causes: cerebrovascular diseases, digestive diseases, endocrine diseases, external causes of death, heart diseases, infectious diseases, neoplasms, and respiratory diseases.

The following countries, grouped into four major regions, are included in the analysis:

North Europe: Denmark (DNK), Finland (FIN), Norway (NOR), Sweden (SWE),

West Europe: Austria (AUT), Belgium (BEL), France (FRA), Ireland (IRL), Netherlands (NLD), U.K. (GBR),

East Europe: Estonia (EST), Hungary (HUN), Latvia (LVA), Lithuania (LTU), Poland (POL),

Outside Europe: Australia (AUS), Canada (CAN), Japan (JPN), New-Zealand (NZL).

2.2 Methods

2.2.1 Dispersion indicators

To monitor changes in the distribution of deaths by sex and country over time, we can consider the standard deviation *around* the mode at time t

$$SD(M)_t = \sqrt{\frac{\int_{x_0}^{\omega} (x - M_t)^2 \, f_t(x) \, dx}{\int_{x}^{\omega} f_t(x) \, dx}} \tag{1}$$

where x is the age at death, x_0 is the initial value of x (usually $x_0 > 0$ to disregard infant mortality), ω is the highest age at death and $f_t(.)$ the density function defining the distribution of ages at death at time t. Equation (1) can be decomposed into two terms: the standard deviation *below* the mode, and the standard deviation *above* the mode,

$$SD(M-)_t = \sqrt{\frac{\int_{x_0}^{M_t} (x - M_t)^2 \, f_t(x) \, dx}{\int_{x}^{M_t} f_t(x) \, dx}} \qquad SD(M+)_t = \sqrt{\frac{\int_{M_t}^{\omega} (x - M_t)^2 \, f_t(x) \, dx}{\int_{M_t}^{\omega} f_t(x) \, dx}} \tag{2}$$

2.2.2 Decomposition of $SD(M)$ indicators

Based on the equations presented above, a change in the $SD(M)$s indicators can occur following a change in the modal age at death, M, and/or a change in the density function, $f(x)$. Consider the case where M increases while the density function, $f(x)$, remains the same. As M moves further away from the left tail and closer to the right tail, the dispersion indicator increases at ages below M and decreases at ages above M. The overall effect on $SD(M)$ depends on whether the increase in $SD(M-)$ or the decrease in $SD(M+)$ was larger.

Since we are interested in changes in the density function and the role that causes of death played in those changes, to remove the effect of M on the $SD(M)$ indicators, we decompose the total change in the $SD(M)$ into changes due to M and changes due to $f(x)$ as follows:

The changes in $SD(M)$ between two calendar years, t_2 and t_1, can be written as:

$$\Delta SD(M) = SD(M)_{t_2} - SD(M)_{t_1} \tag{3}$$

$$\Delta SD(M) = \sqrt{\frac{\int_x^\omega (x - M_{t_2})^2 \, f_{t_2}(x)\, dx}{\int_x^\omega f_{t_2}(x)\, dx}} - \sqrt{\frac{\int_x^\omega (x - M_{t_1})^2 \, f_{t_1}(x)\, dx}{\int_x^\omega f_{t_1}(x)\, dx}} \tag{4}$$

By adding and subtracting $\sqrt{\frac{\int_x^\omega (x-M_{t_2})^2 \, f_{t_1}(x)\, dx}{\int_x^\omega f_{t_1}(x)\, dx}}$ to equation (4), the change in $SD(M)$ can be rewritten as:

$$\begin{aligned}\Delta SD(M) = &\underbrace{\sqrt{\frac{\int_x^\omega (x - M_{t_2})^2 \, f_{t_2}(x)\, dx}{\int_x^\omega f_{t_2}(x)\, dx}} - \sqrt{\frac{\int_x^\omega (x - M_{t_2})^2 \, f_{t_1}(x)\, dx}{\int_x^\omega f_{t_1}(x)\, dx}}}_{\Delta f(x)} + \\ &\underbrace{\sqrt{\frac{\int_x^\omega (x - M_{t_2})^2 \, f_{t_1}(x)\, dx}{\int_x^\omega f_{t_1}(x)\, dx}} - \sqrt{\frac{\int_x^\omega (x - M_{t_1})^2 \, f_{t_1}(x)\, dx}{\int_x^\omega f_{t_1}(x)\, dx}}}_{\Delta M}\end{aligned} \tag{5}$$

where $\Delta f(x)$ shows the change in $SD(M)$ due solely to the change in $f(x)$ and ΔM shows the change in $SD(M)$ due solely to the change in M.

This decomposition method can also be applied to $SD(M-)$ and $SD(M+)$. However, the upper and lower bounds of the integral must be adjusted accordingly.

To assess the contribution of causes of death to changes in the dispersion around the modal age, we will focus only on the first two terms of 5. More specifically, we want to determine which causes of death are responsible for the difference between $f_{t_1}(x)$ and $f_{t_2})(x)$ when both densities have a modal age equal to M_{t_2}.

2.2.3 Estimation of M and all-cause and cause-specific $f(x)$

The calculation of the standard deviation indicators depend on M and $f(x)$. By definition, the modal age at death, M, is given by:

$$M = \max_x \, f(x). \tag{6}$$

The density function provides information on the unconditional risk of dying at age, x, and describes the distribution of deaths across ages. It can be written as follows:

$$f(x) = \mu(x)S(x) = \mu(x)exp\Big[-\int_0^x \mu(u)du\Big] \tag{7}$$

where $\mu(x)$ is the hazard function, also known in mortality as the force of mortality.

To maintain continuity in the age patterns of mortality and derive precise density functions and the associated modal ages at death, we use single year age, sex, and calendar mortality data from HMD. For each country, sex, and calendar year, we calculate the force of mortality by smoothing age-specific death rates with penalized B-splines, (P-splines). B-splines ensure flexibility and accuracy, while the penalty, acting on the coefficients of adjacent B-splines, ensures the smoothness of the fit (Camarda, 2012; Currie et al., 2004). The balance between smoothness and precision is controlled by a smoothing parameter, selected using the Bayesian Information Criterion (BIC) (Schwarz, 1978), which has been shown to be the most suitable for mortality data (Currie et al., 2004).

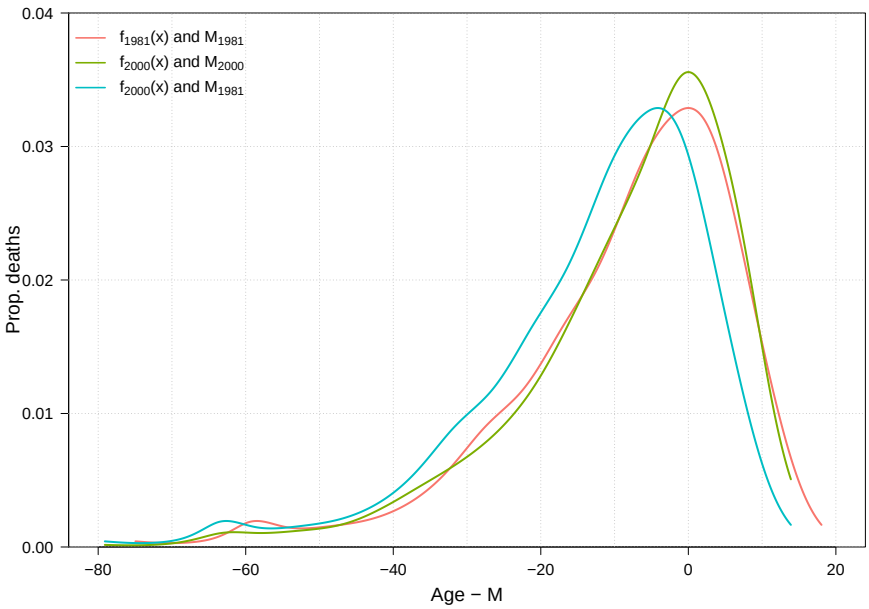
The all-cause $f(x)$ are obtained from the smooth force of mortality using standard numerical techniques. Once the smooth density functions are obtained, the modal age at death, M, can be determined with great numerical precision. Cause-specific density functions will be estimated using the Penalized Composite Link Model (PCLM) to handle five-year age-grouped data, allowing for smooth distributions from coarsely binned data (Rizzi et al., 2015)

2.2.4 Illustration decomposition method

Figure 1 illustrates the smooth distribution of age at death, $f(x)$ for French males in 1981 (red curve) and 2000 (green curve). The modal age at death was 79.9 years in 1981 and 84.1 years in 2000. The blue curve represents the smooth age-at-death distribution of 1981, corresponding to $f_{t_1}(x)$ in equation 5, with the modal age at death of 84.1 years, corresponding to M_{t_2} in equation (5). The change in $SD(M-)$ between 1981 and 2000 due solely to a change in $f(x)$ is shown by the difference between the green and blue curves.

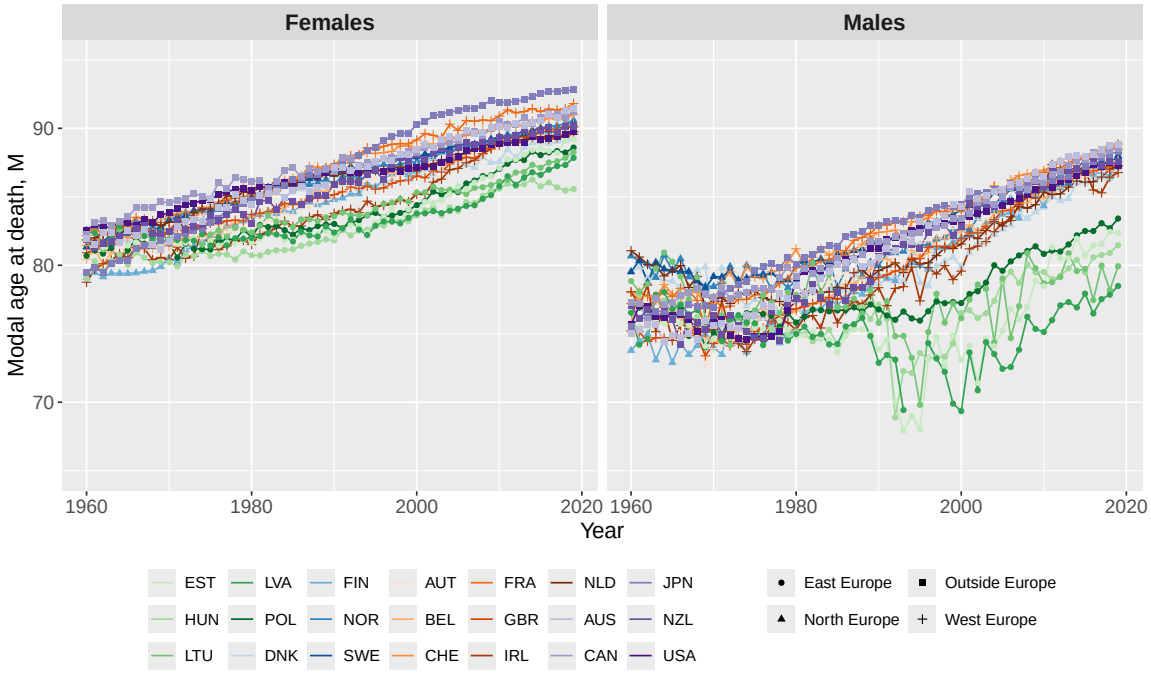
We are interested in identifying the causes of death that have led to the difference between the two curves at ages below M, represented by the shaded area in the figure.

Figure 1: Smooth distribution of age at death, $f(x)$, for French males, 1981 and 2000



Source: (HMD, 2024)
Note: Smoothing was based on ages above 10.

Figure 2: Trends in modal age at death, M, for females and males in several countries in and outside Europe, 1960 to 2019



Source: (HMD, 2024)
Note: Smoothing was based on ages above 10.

3 Preliminary Results

Since the dispersion indicators presented above are relative to the modal age at death, it is important to look at how this metric has evolved over time in the countries included in the study before diving into the $SD(M)$ indicators. Figure 2 illustrates trends in modal age at death, M, spanning from 1960 to 2019 across several European and non-European countries. As shown, M increased over time for both sexes in all countries. In 1960s the female the modal age at death, M, ranged from 78.8 years (Ireland) to 82.5 (Norway) years. By 2019, Japanese women had the highest modal age at death at 93 years, while Hungarian women had the lowest at 85.6 years.

Overall, men consistently had a lower modal age at death than women, indicating earlier-age mortality. In 1960, Finnish males had the lowest M at 73.8 years while Dutch males the highest at 81 years. By 2019, French men took the lead with a modal age of 88.9 years, while Latvian men recorded the lowest at 78.5 years.

Interestingly, the data reveals that country differences in M for men shrank between 1960 and 2019, excluding Eastern European countries, hinting at a convergence in the 'typical' length of life. For women, however, the opposite occurred, with greater variation in modal age at death in 2019 compared to 1960.

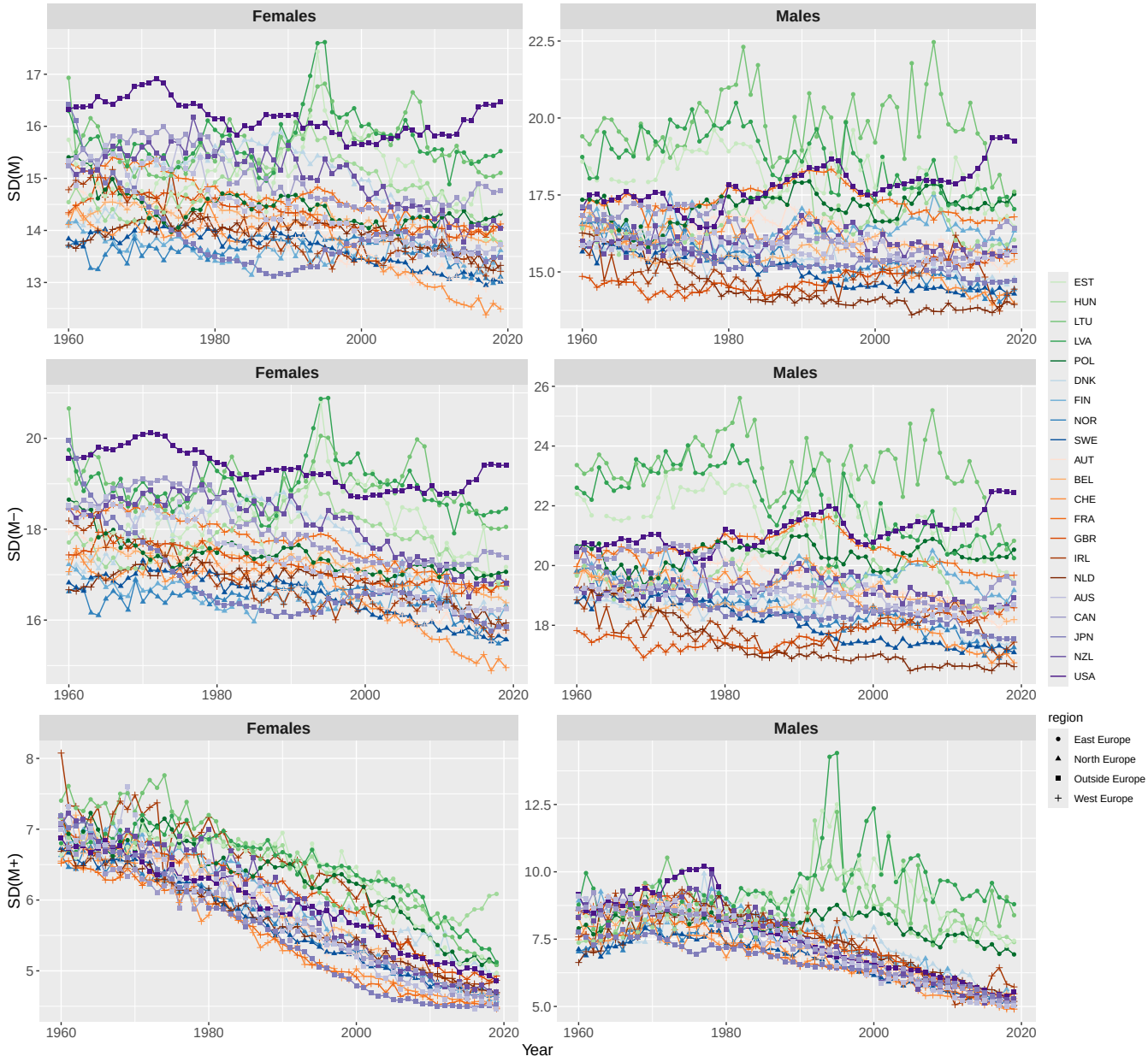
Figure 3 highlights the trends in the standard deviation *around*, *below*, and *above* the modal age, $SD(M)$, $SD(M-)$, and $SD(M+)$, for both sexes across various countries from 1960 to 2019. Over time, these dispersion indicators have generally decreased for both men and women, with the exception of Eastern Europe. Notably, $SD(M+)$ shows a steadier decline compared to $SD(M-)$, indicating that deaths at ages above the mode have become more concentrated, while variation in deaths below the mode remains broader. In addition, there is less variation across countries in the dispersion of deaths at ages above than below the modal age for both sexes, particularly for females. Gender differences emerge as well, with females showing more country variation in deaths above the mode and less variation below it compared to males.

Since $SD(M)$ is influenced by changes both below and above M, we dive deeper into the trends in $SD(M-)$ (Figure 2, middle panels). In 1960, the standard deviation below the mode varied from 16.6 years (Norway) to 20.6 years (Lithuania) for females, and from 17.8 years (U.K.) to 23.4 years (Latvia)

for males. Over time, $SD(M-)$ decreased for both sexes in most countries, except for males in the U.S. and U.K. Eastern European countries, however, showed the greatest variability in age at death for both sexes.

The country comparison reveals higher variability in the age at death in the Eastern European countries compared to the remaining countries included in the study for both males and females.

Figure 3: Trends in standard deviation around the mode, $SD(M)$, below the mode, $SD(M-)$, above the mode, $SD(M+)$, for females and males in several countries in and outside Europe, 1960 to 2019



Source: (HMD, 2024)
Note: Smoothing was based on ages above 10.

As mentioned earlier, changes in the $SD(M)$ indicators can occur following a change in M and/or a change in $f(x)$. If the change is driven by M alone, then $SD(M-)$ will increase over time and $SD(M+)$ will decrease over time. Therefore, the upward trend in M, observed in Figure 1, should result to a negative relationship between these two dispersion indicators.

Figure 4 explores the relationship between $SD(M+)$ and $SD(M-)$ over time and across countries. As

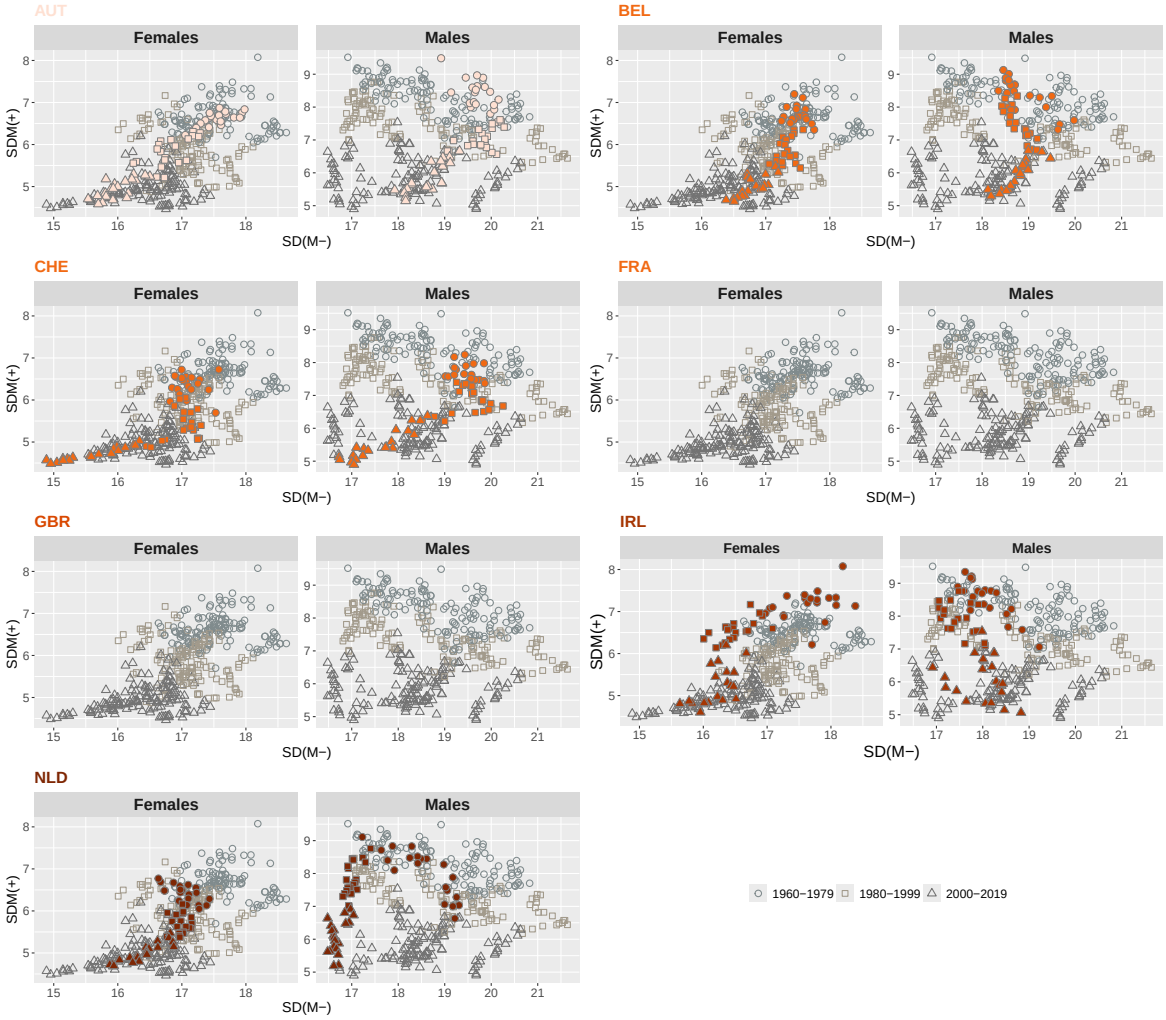
shown, $SD(M+)$ has declined steadily over time and this trend is relatively stable across the countries considered. In contrast, $SD(M-)$ shows much more variability by gender and country. In Western Europe, for example, while $SD(M+)$ dropped, $SD(M-)$ increased or remained stable in certain countries like France and the UK, particularly for males (Fig. 4a). French men and women, notably, showed the highest dispersion in deaths both above and below the mode compared to other Western European countries.

In the Nordic countries, trends varied: Finnish females saw increasing dispersion below M until 2000-2019, while Danish and Norwegian males exhibited relatively stable trends (Fig. 4b). Finnish females consistently had the highest dispersion above and below M, while Danish males had the lowest.

In the Eastern European countries,the decrease in $SD(M+)$ for females in Estonia, Latvia, Lithuania, and Poland was accompanied by a general decrease in $SD(M-)$ (Fig. 4c). For males, the relationship between the two dispersion indicators is not very clear.

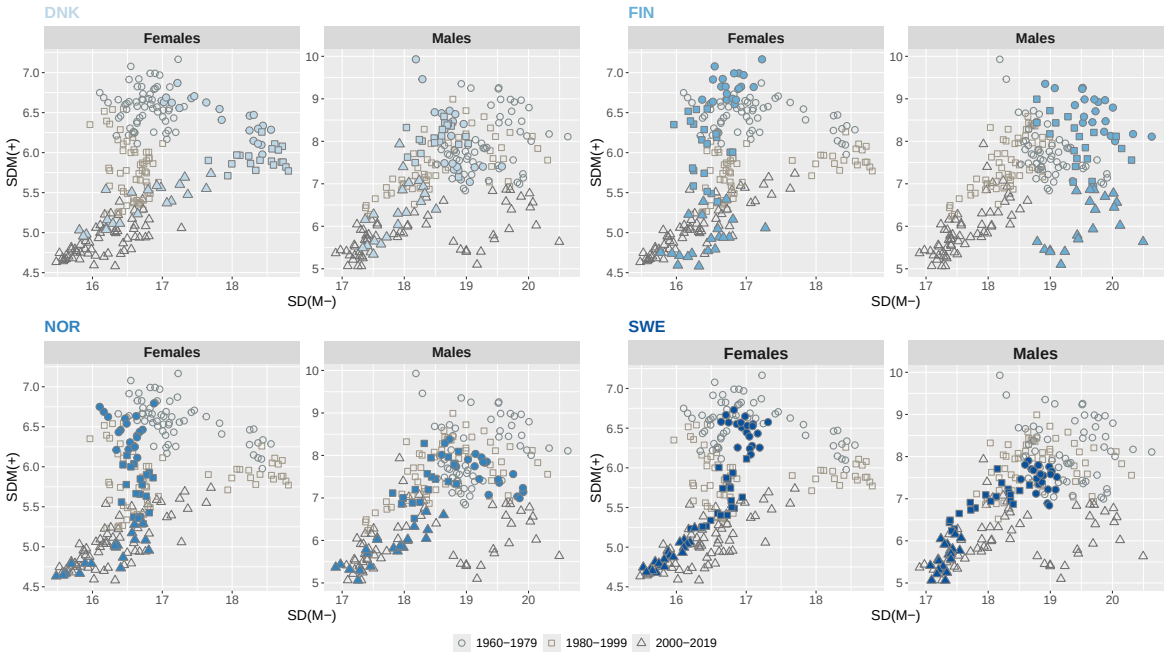
Outside of Europe, most countries have experienced declines in both $SD(M+)$ and $SD(M-)$ since the 1960s, with the notable exception of the U.S., where variability below the mode has increased (Fig. 4d). A closer look at the results shows that Japan and the U.S. have two opposite trends for $SD(M-)$ and a rather similar one for $SD(M+)$. Japanese females stand out for their sharp decline in $SD(M-)$ and their position among the lowest in dispersion below the mode, while U.S. males and females had the highest levels for both $SD(M+)$ and $SD(M-)$.

Figure 4a: Relationship between $SD(M+)$ and $SD(M-)$ for females and males since 1960 in **West European** countries



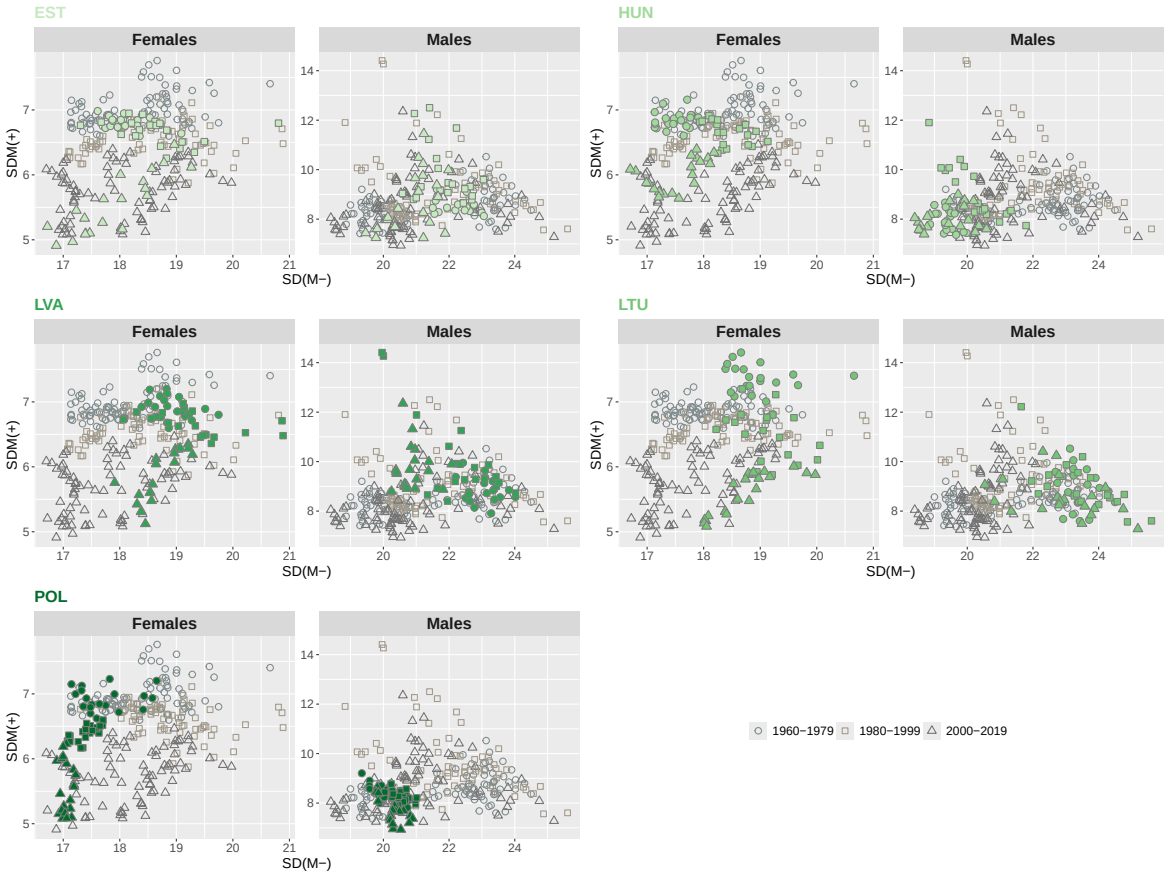
Source: (HMD, 2024)
Note: Smoothing was based on ages above 10.

Figure 4b: Relationship between $SD(M+)$ and $SD(M-)$ for females and males since 1960 in **North European** countries



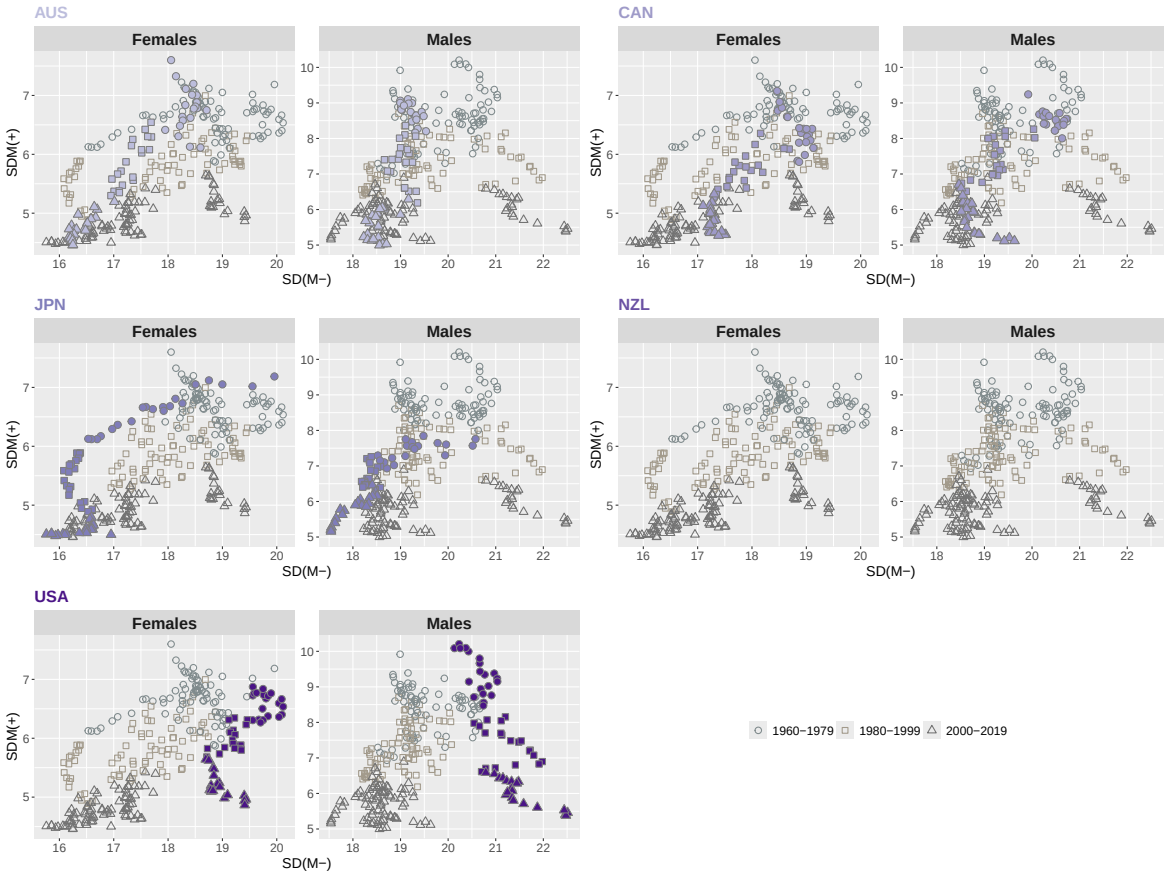
Source: (HMD, 2024)
Note: Smoothing was based on ages above 10.

Figure 4c: Relationship between $SD(M+)$ and $SD(M-)$ for females and males since 1960 in **East European** countries



Source: (HMD, 2024)
Note: Smoothing was based on ages above 10.

Figure 4d: Relationship between $SD(M+)$ and $SD(M-)$ for females and males since 1960 in **non-European** countries



Source: (HMD, 2024)
Note: Smoothing was based on ages above 10.

4 Next Steps

This is an ongoing project. Our next step will focus on estimating smooth age-at-death distributions by cause of death from grouped data. To do this, we will use the pclm method, which allows to obtain smooth age-at-death distributions from binned data. Our experience with the method has shown that some modifications are needed to improve the performance of the method, especially when using causes with small numbers of deaths. Finally, in our last step, we will investigate which causes of death are responsible for the change in the all-cause age-at-death distribution at ages below the mode. This is done for males and females and for all countries included in the study over the period for which data on causes of death is available.

References

Bergeron-Boucher, M.-P., Ebeling, M., and Canudas-Romo, V. (2015). Decomposing changes in life expectancy: Compression versus shifting mortality. *Demographic Research*, 33:391–424.

Camarda, C. G. (2012). Mortalitysmooth: An r package for smoothing poisson counts with p-splines. *Journal of Statistical Software*, 50:1–24.

Canudas-Romo, V. (2008). The modal age at death and the shifting mortality hypothesis. *Demographic Research*, 19:1179–1204.

Cheung, S. L. K., Robine, J. M., and Caselli, G. (2008). The use of cohort and period data to explore changes in adult longevity in low mortality countries. *Genus*, pages 101–129.

Cheung, S. L. K., Robine, J.-M., Tu, E. J.-C., and Caselli, G. (2005). Three dimensions of the survival curve: Horizontalization, verticalization, and longevity extension. *Demography*, 42(2):243–258.

Currie, I. D., Durban, M., and Eilers, P. H. (2004). Smoothing and forecasting mortality rates. *Statistical modelling*, 4(4):279–298.

Kannisto, V. (2001). Mode and dispersion of the length of life. *Population: An English Selection*, pages 159–171.

Kannisto, V. (2007). Central and dispersion indicators of individual life duration: New methods. *Human longevity, individual life duration, and the growth of the oldest-old population*, pages 111–129.

Lan Karen Cheung, S. and Robine, J.-M. (2007). Increase in common longevity and the compression of mortality: The case of japan. *Population studies*, 61(1):85–97.

Ouellette, N. and Bourbeau, R. (2011). Changes in the age-at-death distribution in four low mortality countries: A nonparametric approach. *Demographic Research*, 25:595–628.

Rizzi, S., Gampe, J., and Eilers, P. H. (2015). Efficient estimation of smooth distributions from coarsely grouped data. *American Journal of Epidemiology*, 182(2):138–147.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, pages 461–464.