Deriving and Comparing Healthy Longevity Distributions by Gender and Health Prevalence Measures: A Statistical Moments and Maximum Entropy Approach

Rami Cosulich[1], Vanessa di Lego[2], Virginia Zarulli[1]

[1]Department of Statistical Sciences, University of Padova, Padova, Italy

[2] Department of Demography, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

Abstract

Background. The literature on healthy longevity has typically focused on average values (i.e., healthy life expectancy). Despite recent studies starting to shift attention to the whole healthy lifespan distribution, research gaps remain. This study aimed to compare healthy longevity distributions at age 60 between different health measures and sexes.

Data and methods. We used data from the Survey of Health, Ageing and Retirement in Europe merged with the Human Mortality Database. A Markov chain model was used to estimate the first three statistical moments of healthy longevity distributions. The maximum entropy method was then applied to derive the full distributions. The φHL statistic and the Hellinger distance were used to compare distributions between males and females.

Findings. For most health measures, the probabilities of health loss at younger ages were higher for males than for females, and females had a longer healthy life expectancy. Males had more dispersed distributions with a lower mode. The probability for a man to have a longer healthy lifespan than a female was below 50% for various health measures and was the lowest for living free of cardiovascular disease. In contrast, the probability for a man to live free of arthritis or rheumatism for longer than a female was above 50%. The most similar distributions between males and females were observed with life free of any chronic conditions and life with no more than one chronic condition.

Discussion. Further research could investigate healthy longevity distributions by socio-economic status.

Keywords

Healthy longevity distribution, healthy lifespan variation, φ statistic, Hellinger distance, gender health-survival paradox, health inequalities.

Background

Accurately capturing health inequality is one of the most pressing challenges in health research. The literature on healthy longevity has typically focused on average values (i.e., healthy life expectancy), highlighting fundamental differences between subpopulations distinguished by factors like socioeconomic status or gender, but masking information on how individual healthy lifespans are

distributed around a subpopulation mean. In response to the limitations of relying solely on averages, recent studies have begun examining inter-individual variation in healthy longevity. Two studies (Permanyer et al., 2023; Zarulli & Caswell, 2024) addressed this topic globally, using Global Burden of Disease (GBD) data and the standard deviation as an indicator of variation. Moreover, one study assessed not only the standard deviation but also the skewness of healthy longevity distributions (Caswell & Zarulli, 2018), using data from the Survey of Health, Ageing and Retirement in Europe (SHARE). Despite these latest efforts, unanswered questions remain on the shape of distributions of healthy lifespan, how equally (or unequally) individuals lose their health, and how the process of health loss differs by gender and based on the health prevalence measure used to define healthy longevity.

A focus on gender in healthy longevity research is particularly relevant because of the puzzling and persistent male–female health-survival paradox, which is the fact that women live longer but have poorer health than men (Ahrenfeldt et al., 2019; Oksuzyan et al., 2008). Despite extensive study and almost four decades of data confirming this gender gap, no single explanation fully accounts for why women live longer but have sicker lives. This leads to yet another major challenge in the field: the evaluation of healthy lifespan and gender differences is highly sensitive to the choice of health indicator. Estimates can vary widely depending on whether one uses self-rated health, functional limitations (ADL), or chronic conditions—as highlighted in studies by (Di Lego et al., 2020; Reinwarth et al., 2024; Van Oyen et al., 2010; Di Lego & Sauerberg, 2023; Van Oyen et al., 2013).The fact that many indicators have been used depends both on the intrinsic multifaceted concept of health, which rightfully calls for many definitions to be used depending on the research interest, but also on the difficulty of operationalizing it univocally. An additional issue is that, contrary to incidence-based health definitions, definitions of health based on prevalence data score poorly at capturing the real progression of health loss of individuals over age, which is often made of transitions back and forth between the health and unhealthy status (Caswell & Daalen, 2021; Murray et al., 2000). Nevertheless, prevalence-based approaches remain the most used ones due to the abundance of data and the ease of measurement, compared to incidence-based analyses (Schroeder, 2012).

Without the presumption to address the substantive and normative aspects related to the ways health prevalence is operationalized in the demography of health, this study adopted a comparative approach to investigate healthy longevity distributions by gender using seven distinct measures of health prevalence, each representing a different aspect of health. These measures were defined

specifically to minimize the problem of the incapability of prevalence data to capture the dynamic transitions back and forth between healthy and unhealthy states, because they represent health statuses that are very difficult to fully recover from. Given that this paper focuses on chronic conditions, we use the term "health loss" to refer to a person reaching the end of their healthy lifespan, which can occur either by death or by transitioning to an unhealthy state (which is a permanent state). In addition, given our focus on gender differences, we prioritized the inclusion of conditions with known sex-specific impacts on health. With this approach, we can look at how the process of health loss unfolds and accumulates over age, differently for men and women, and for different measures of health.

We provide a full characterization of the healthy longevity distributions through the statistical moments (mean (healthy life expectancy), variance, and skewness), which has been shown to provide a broader picture of health across populations, beyond an exclusive focus on healthy life expectancy (Caswell & Daalen, 2021; Caswell & Zarulli, 2018). We report the standard deviation rather than the variance for ease of interpretability and because it is a standard practice in demography, and we call the standard deviation of healthy longevity (SDHL) "inter-individual variation in healthy longevity", following standard demographic terminology (Permanyer et al., 2023; Zarulli & Caswell, 2024). We also focus on the probabilities of losing health at specific ages and on the mode of the distributions. Moreover, for the first time, we present a formal comparison of gender differences in healthy longevity through the φHL statistic, expressed as the probability for a male to have a longer healthy lifespan compared to a female, and the Hellinger distance, used to measure the dissimilarity between the distributions of healthy years of life of men and women.

By moving beyond averages and analyzing the full distribution of healthy longevity, this study offers deeper insights into the dynamics of health inequality. It provides both a formal quantification of individual variation and a richer understanding of how health loss unfolds unequally between men and women across different dimensions of health.

Methods

The Markov chain model

We applied a Markov chain with rewards model (Caswell & Zarulli, 2018) to estimate the mean, variance (reported as standard deviation for ease of interpretability), and skewness of healthy lifespan from prevalence data. The estimates are generated from two stochastic processes: the risk

of mortality and the chance of developing health conditions, estimated by the prevalence of such conditions at each age. The model (Caswell & Zarulli, 2018) accounts for both. The model requires age-specific mortality and health prevalence data, for which we used different health measures, outlined below.

Data: SHARE and HMD

Data on health prevalence come from various health measures included in the Survey of Health, Ageing and Retirement in Europe (SHARE), which is a study conducted approximately every two years, starting with wave 1, which was conducted in 2004 or in 2004/05 in most countries (SHARE, 2024). We used SHARE data from waves 2, 4, 5, 6, 7, 8, 9 (the citations for the datasets and details on the funding of SHARE are provided in the references and acknowledgments sections). In most countries, wave 2 was conducted in 2006/07, and in all countries, wave 9 was conducted in 2021/22 (SHARE, 2024). We did not include wave 1 because it had no data on whether respondents had a dementia or Alzheimer's diagnosis, which was used to define two cognitive measures used in this study. Similarly, we did not use wave 3 because it had no relevant data. We did not use data from the SHARE Corona surveys because it did not include data on some of the conditions relevant to our analysis.

We combined SHARE's health prevalence data with the annual probability of death for each specific age, sex and country from the Human Mortality Database (HMD) (Human Mortality Database (HMD), 2024). More details on this data merge are reported in Table S2 in the Supplementary Material.

SHARE has been conducted in 28 European countries and in Israel, but only some countries have participated in all waves (SHARE, 2024). We only included 9 countries that participated in all the SHARE waves outlined above and under the condition that they had HMD data for all the relevant years: Austria, Belgium, Czech Republic, Denmark, France, Italy, Spain, Sweden, Switzerland. Germany was excluded, despite participating in all the SHARE waves, because HMD data was not available for Germany for the years 2021 or 2022 (corresponding to wave 9).

SHARE surveys people aged 50 and over but we restricted our analysis to people aged 60 and over because refreshment samples were missing for some country-wave combinations, and in some countries, this occurred for some consecutive waves (Bergmann et al., 2019, 2021; SHARE, 2024).[1]

[1] Table 1 in (Bergmann et al., 2019) reports the inclusion of refreshment samples for each country and wave from wave 1 to 7. A report on the methodology in wave 8 (Bergmann et al., 2021) reports about countries that drew refreshment samples in wave 8 and countries that managed to interview these refreshment samples

This meant that people in their early- and mid-50s were missing or limited to younger partners. Due to sample size limitations, we grouped data with an open-ended interval for ages 90 and above.

Health measures

Because our concern is with differences between women and men, gender-specific health conditions have been the focus in calculating health prevalence. The health measures taken from SHARE to calculate health prevalence are outlined in Table 1 (more details provided in Table S1 in the Supplementary Material), together with why they are specifically important for understanding gender differences in health and mortality.

before the fieldwork interruption due to COVID-19 (see pp. 24 and 28-29). The Release Guide 9.0.0 (SHARE, 2024) mentions that no refreshment samples were drawn in wave 9 due to the uncertainty around the ongoing COVID-19 pandemic (pp.40-41). However, refreshment samples that had been drawn but not fielded in wave 8 were fielded for the first time in wave 9 (SHARE, 2024).

Table 1. Health measures: definitions and rationales for their inclusion.

Health Measure	Definition	Rationale
Life free from major chronic conditions	Free from: high blood pressure, high cholesterol, heart problems, stroke or cerebral vascular disease, diabetes/high blood sugar, chronic lung disease, arthritis/rheumatism, Parkinson's, Alzheimer's/dementia	These conditions show significant gender disparities: women are more affected by arthritis and Alzheimer's, while men have higher prevalence of lung disease and high blood sugar.[1]
Life with no more than one chronic condition	Either no condition or only one from the list in the row above	Focuses on multimorbidity: it is linked to how long individuals remain relatively healthy before developing two or more chronic conditions.
Life free from any cardiovascular disease (CVD)	Free from: heart problems, stroke or cerebral vascular disease	CVD significantly contributes to gender disparities in health outcomes: although men have higher prevalence and mortality at younger ages, women experience greater disability and death burden due to late diagnosis and less aggressive treatment.[2]
Life free from cardiovascular conditions/risk factors	Free from: high blood pressure, high cholesterol, heart problems, stroke or cerebral vascular disease	Includes key risk factors for CVDs to capture earlier stages of cardiovascular health deterioration.
Life free from arthritis or rheumatism	Free from arthritis or rheumatism	Women are more likely than men to develop arthritis and live fewer years free from it; arthritis explains over 30% of the sex difference in ADLs, a key health expectancy measure.[3]
Good cognitive health (general population comparison)	Scoring above -1.5 SD on cognitive tests (time orientation, verbal fluency, memory), relative to all wave-person observations, and no dementia/Alzheimer's diagnosis	It is a relevant indicator in relation to gender because women have a higher risk and earlier onset of Alzheimer's than men.[4]
Good cognitive health (education-adjusted comparison)	Same as the measure above, but scores are evaluated relative to all wave-person observations with the same educational level	Accounts for educational disparities in cognitive test performance, making cognitive health comparisons more equitable.

Table notes: [1] (Crimmins et al., 2002; Namavari et al., 2024; Nebel et al., 2018; Ntritsos et al., 2018).[2] (Townsend et al., 2015; Vogel et al., 2021). [3] (Whitson et al., 2010) [4] Beam et al, (2018); (Nebel et al., 2018).

Data preparation

Men and women were analyzed separately. For each wave, age and sex, we calculated health prevalence and a weighted average of the annual probability of death by pooling data across all countries. For these calculations, we applied the individual calibrated cross-sectional weights available within the SHARE dataset. These weights "may help reduce the potential selection bias generated by nonresponse errors" (SHARE, 2024) (p.42) and enable inference from the responding sample to the target population (SHARE, 2024) (p.45). Each weight takes into account the household design weight, the NUTS1 region (Nomenclature of Territorial Units for Statistics – Level 1), the gender and age class of the respondent (SHARE, 2024) (p.45).

From the Markov chain model to the maximum entropy method

Using the first three statistical moments of the healthy longevity distributions, calculated for each health measure and sex with the Markov chain model (Caswell & Zarulli, 2018), we applied the maximum entropy method to derive the full empirical distribution of health loss over the lifespan. Despite its high predictive power, applications of this method in demography so far have been limited to age at death distributions in the context of mortality forecasting (Pascariu et al., 2019). For the first time, we have used it to model the health loss process, enabling us to examine the detailed shape of healthy lifespan distributions, their evolution over time, and the differences between men and women across varying definitions of health.

Comparison of healthy longevity distributions through the φHL statistic and the Hellinger distance

We compared the healthy longevity distributions between different sexes using both the φHL statistic and the Hellinger distance. φ is an "outsurvival statistic" in studies on lifespan distributions, where φ expresses "the probability that an individual from a population with lower life expectancy outlives an individual from another population with higher life expectancy" (p.2) (Bergeron-Boucher et al., 2022). In our study, φHL expresses the probability for a male to have a longer healthy lifespan than a female, when comparing randomly selected healthy men and women of age 60, assuming independence between the healthy lifespan of men and women (in agreement with the assumptions in Bergeron-Boucher et al., 2022). The complementary value of φHL, i.e., φHL – 1, refers to the probability for a female to have a longer healthy lifespan compared to a male. Note that when φHL is equal to 0.5, the probability for a man to have a longer healthy lifespan than a

woman is exactly the same as the probability for a woman to have a longer healthy lifespan compared to a man.

The Hellinger distance measures the dissimilarity between distributions. It is closely related to the Bhattacharya coefficient (Nilsson et al., 2017), which in turn, is a measure of overlap between different distributions (Guillerme & Cooper, 2016). The higher the Hellinger distance, the more dissimilar the distributions. The lower the Hellinger distance, the more similar the distributions.

We combined the use of φHL and the Hellinger distance as we believed this was the best way to tackle the analysis of the distributional differences. The φHL statistics formally relates two distributions by focusing on the part on which they overlap, while the Hellinger distance not only accounts for the overlap but is also sensitive to distributional differences beyond the overlap. By using these two metrics, we were able to get a complete picture of the differences between the distributions, not only capturing discrepancies or similarities in shape, but also computing a probabilistic indicator of relevant demographic significance (the φHL statistics).

Calculation of φHL

We used a discrete approximation of the main formula for φ (Bergeron-Boucher et al., 2022; Vaupel et al., 2021). These studies applied it to longevity distributions while we applied it to healthy longevity (HL) through the formula presented in equation (1). The formula refers to φHL defined as the probability that an individual from population 1 has longer healthy lifespan than an individual from population 2:

$$\varphi HL \approx \sum_{x=60}^{w}({}_{n}phloss_{x}^{2} * {}_{n}lhealth_{x+n}^{1}) + \frac{\sum_{x=60}^{w}({}_{n}phloss_{x}^{1} * {}_{n}phloss_{x}^{2})}{2} \quad (1)$$

Where ${}_{n}phloss_{x}^{i}$ is the probability of health loss between age x and $x+n$ in population i. n is the width of the age group, so, when n is 1, we are referring to single ages (for example, 70, 71, 72). ${}_{n}lhealth_{x+n}^{i}$ is the probability of being healthy until age $x+n$ for people in population i. For example, in our dataset focused on people who are healthy at 60, if x=70 and n=1, and if population 1 refers to males and population 2 to females, ${}_{1}phloss_{70}^{F}$ refers to the probability of losing health when aged 70 for females who are healthy at 60, and ${}_{1}lhealth_{70+1}^{M}$ refers to the probability of being healthy until age 71 for males who are healthy at 60. w is the final age when anyone who is still healthy loses health. In our dataset, w = 90, which is the age limit artificially imposed by our analysis due to data limitations.

Calculation of the Hellinger distance

We calculated the Hellinger distance applying the formula from Cha (2007), shown in equations (2) and (3).

$$HD = \sqrt{2 * \sum_{x=60}^{w}(\sqrt{{}_n phloss_x^1} - \sqrt{{}_n phloss_x^2})^2} \quad (2)$$

$$= 2 * \sqrt{1 - \sum_{x=60}^{w} \sqrt{{}_n phloss_x^1 * {}_n phloss_x^2}} \quad (3)$$

Where x, n, ${}_n phloss_x^i$, i and w are defined as per equation (1). Note that the order of ${}_n phloss_x^1$ and ${}_n phloss_x^2$ is interchangeable in equations (2) and (3).

After applying the formula above, we validated the calculation using the R package Philentropy to calculate the Hellinger distance.

Accounting for the sampling error around the prevalence point estimates

To incorporate sampling uncertainty around the prevalence point estimates, weighted non-parametric bootstrapping was used, applying the individual calibrated cross-sectional weights available within the SHARE dataset. More specifically, for each combination of health measure, wave, sex and age, weighted sampling with replacement was done 5,000 times, thus obtaining 5,000 prevalence values. The Markov chain model and the maximum entropy code were then run 5,000 times. 95% confidence intervals were obtained from the 5,000 sets of results. More specifically, the 2.5th and 97.5th percentiles of the results were the lower and upper limits of the confidence intervals.

Sample size

For health measures 1 to 5 (i.e., the health measures that referred to chronic conditions rather than to the cognitive test scores), the unweighted sample per wave-sex-age combination ranged from a minimum of 22 males aged 89 in wave 2 to a maximum of 887 females aged 65 in wave 5. For the measure related to cognition, the sample was smaller, ranging from a minimum of 19 males aged 89 in wave 2 to a maximum of 765 females aged 66 and 765 females aged 67 in wave 6. One of the reasons for the more missing data for cognition is that time orientation questions were only asked to baseline respondents in waves 4 and 5, although the same questions were asked to both baseline

and longitudinal respondents in the other waves.[2] Wave 5 had a lower percentage of baseline respondents than wave 4, and so it had the lowest sample size for cognition. Moreover, anyone who responded to a SHARELIFE questionnaire in wave 7 (i.e., anyone who had not previously participated in wave 3 (Bergmann et al., 2019b) was not asked questions on time orientation and verbal fluency in wave 7. See Figure S1 and Tables S4-S7 in the Supplementary Material for more details on the sample size.

Final notes on the methods to help with the interpretation of the results.

Before moving on to the results, it is worth reminding the reader that distributions of healthy longevity from age 60 apply to people who reach age 60 in a healthy state (defined differently depending on the health measure used). It may also help the reader to note that the distributions of healthy longevity can alternatively be interpreted as distributions of the age of health loss, defined as per the introduction section, i.e., the age at the end of healthy lifespan. The analysis is based on prevalence data and, therefore, does not allow for transitions back to a healthy state.

Results

Statistical moments of the healthy longevity distributions: comparison between different health measures

For both males and females, and across all waves, life free of any chronic conditions was characterized by both the shortest healthy life expectancy (HLE) and the lowest standard deviation in healthy longevity (SDHL), i.e., the lowest inter-individual variation. Life free of any cardiovascular conditions had the second-lowest HLE and SDHL. Note that within each sex, SDHL was positively and very strongly correlated with HLE. Figure 1 shows the probability density functions of healthy longevity distributions from age 60, by wave, sex and health measure. Figure 2 shows the first three statistical moments (mean (i.e., HLE), standard deviation, and skewness) of the distributions. Supplementary Figure S2 shows the correlation between the statistical moments. Supplementary Table S8 outlines the definitions of correlation strength in the current work.

[2] This was observed during the data analysis. Also see the SHARE questionnaires for the filters applied in each wave (SHARE, 2006, n.d.c, n.d.a, n.d.b, n.d.d, n.d.e, n.d.f).

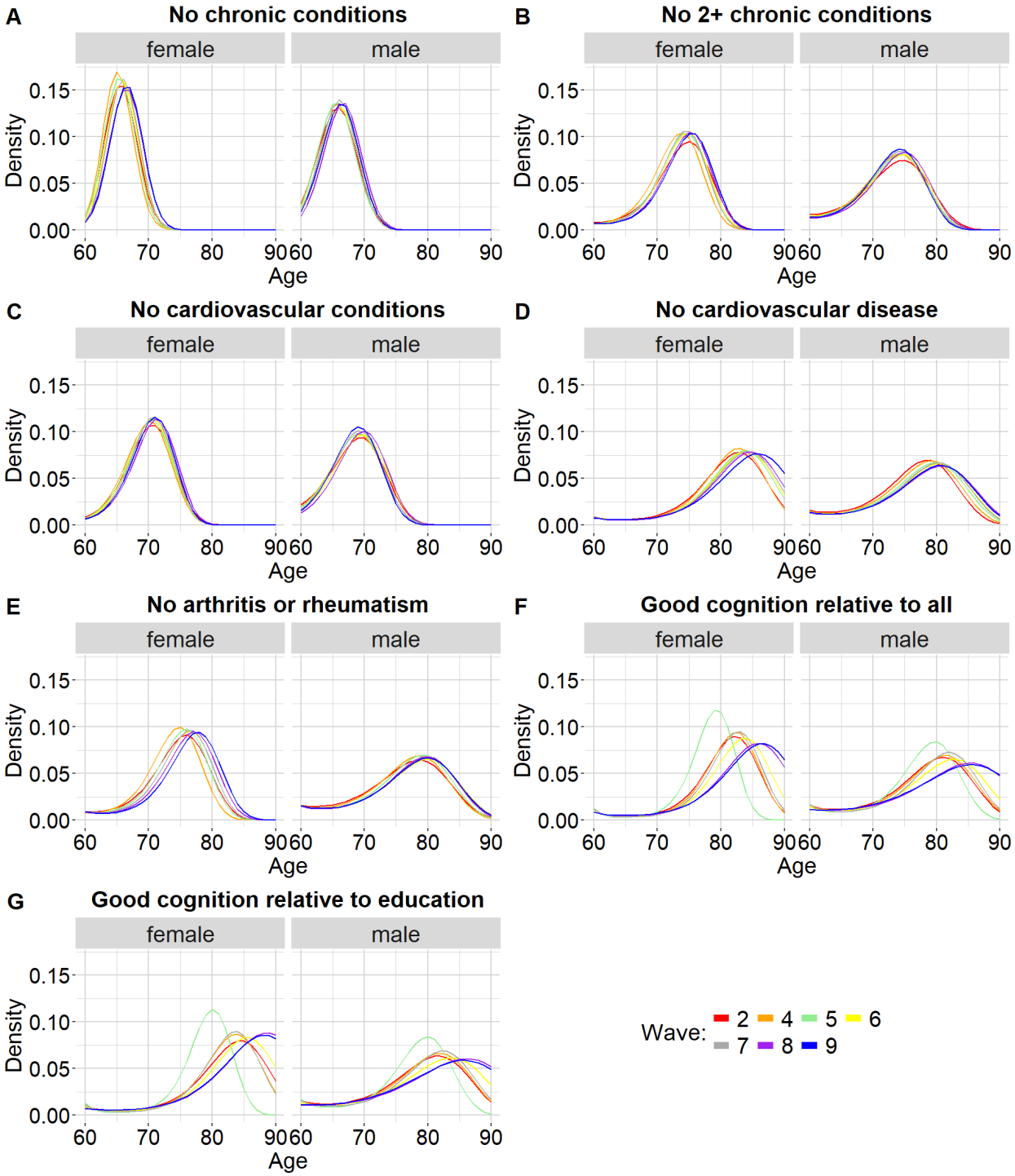

Figure 1. Probability density functions.

Figure notes. Functions based only on point estimates to simplify the visual representation.

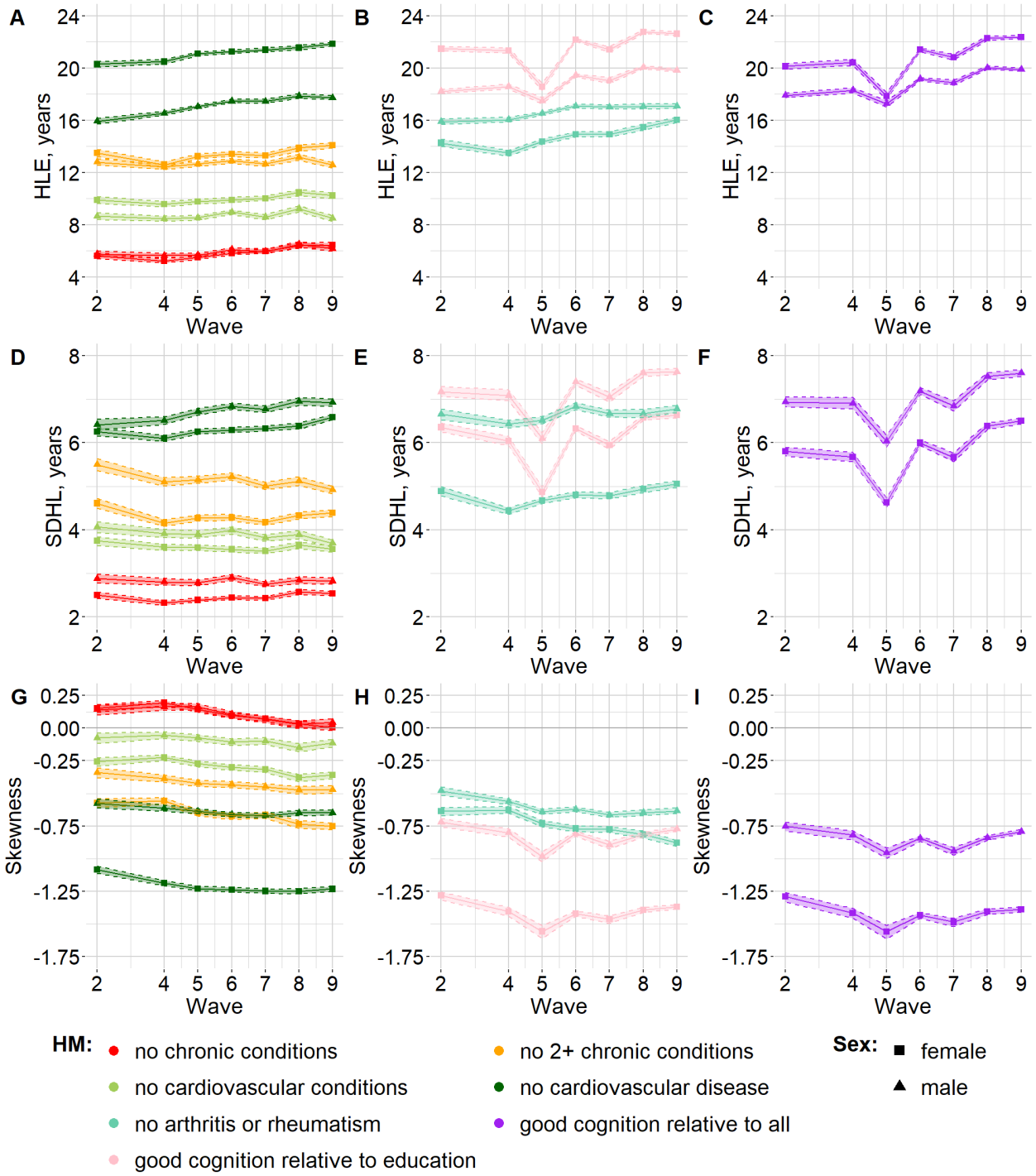

Figure 2. Statistical moments of healthy longevity distributions.

Figure notes. Abbreviations: HLE: healthy life expectancy (mean of the healthy longevity distribution); HM: health measure; SDHL: standard deviation of healthy longevity.

Symbols (squares or triangles) connected by continuous line: point estimates. Dashed lines: 95% confidence intervals.

Figure 2 shows that for all health measures except “no chronic conditions”, skewness was negative across all waves, for both males and females. For each sex, the negative skewness values furthest from 0 were observed for the two “poor cognition” measures. This meant that there was a minority that reached the end of lifespan with good cognition considerably earlier than the majority. Within each sex, skewness was negatively and very strongly correlated with HLE (see Supplementary Figure S2): health measures with a higher HLE (such as the good cognition measures) tended to have more negatively skewed distributions. Consistently with this, the skewness values closest to 0 were observed for two measures with relatively low HLE: “no chronic conditions” and “no cardiovascular conditions”. Life free of any chronic conditions had positive skewness point estimates across all waves, although in some waves the 95% CI ranged from negative to positive. When negative skewness was relatively far from 0 (i.e., with the exception of life free of cardiovascular conditions), the negatively skewed distributions had a mean age of health loss lower than the mode age, i.e., the age with the highest probability of health loss. This held across all waves, for both males and females. For example, for a woman healthy at age 60, in wave 9, the mean age of experiencing either cardiovascular disease or death was 81.9 (95% CI: 81.7 to 82.0) but the age with the highest probability of experiencing either cardiovascular disease or death was age 86 (95% CI: 85 to 87) (the probability of losing health at this mode age rather than at other ages was 7.7%, 95% CI: 7.6% to 7.7%). For wave 9, the point estimates of the HLE age and mode age are visible in Figure 3, which shows the probability mass functions for wave 9, by sex and health measure. Supplementary Figure S14 shows the mode age and the HLE age for all waves, by sex and health measure, with the 95% CIs.

Statistical moments of the healthy longevity distributions: comparison between males and females

For most health measures, HLE was higher for females than for males, across all waves. A notable exception was that life expectancy free of arthritis or rheumatism was higher for males than for females, across all waves. Moreover, life expectancy free of any chronic conditions or with no more than one chronic condition was similar between males and females. For all health measures, the inter-individual variation in healthy longevity (as measured by SDHL) was higher for males than for females, across most waves (there were a couple of exceptions where the 95% CIs overlapped). For all health measures, except for “no chronic conditions”, distributions were more negatively skewed for females than for males.

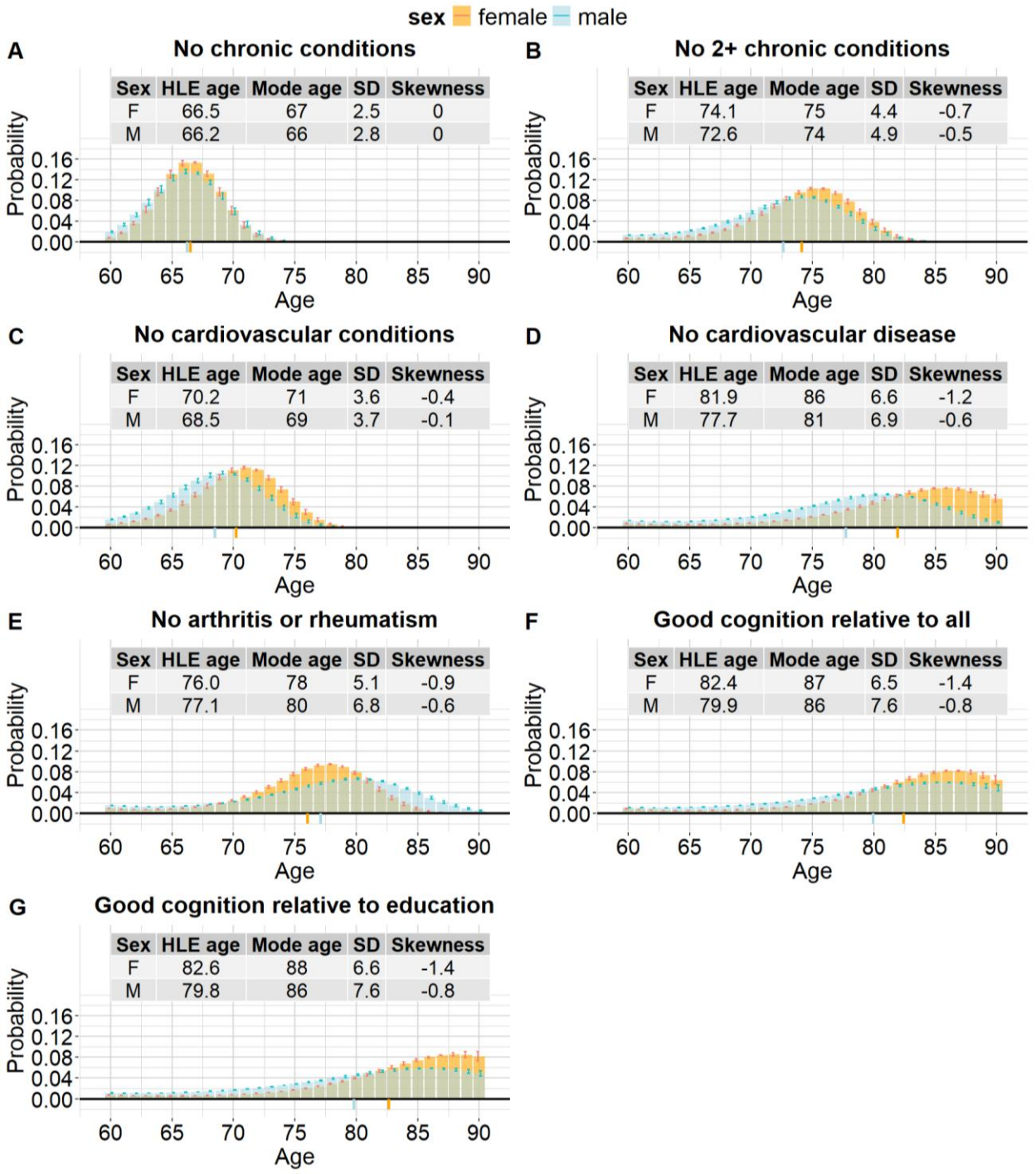

Figure 3. Age at the end of the healthy lifespan: probability mass functions, wave 9.

Figure notes. Abbreviations: F: female; HLE: healthy life expectancy; M: male; SD: standard deviation.
The green colour indicates the overlap between the distributions for males and females.
The coloured lines below the x axis show the mean for each sex.
The HLE age was equal to 60 + HLE at age 60.
The 95% CIs for the probabilities of health loss are shown, but for simplicity, the values reported are only point estimates.

Statistical moments of the healthy longevity distributions: observations over time

Figure 1 and Figure 2 show that the good cognition longevity distributions had the greatest inter-wave variation, compared to other health measures, especially when comparing wave 5 to the other waves. This may be linked to the inter-wave variations in eligibility criteria for some cognition questions. As mentioned in the methods section, wave 5 had the lowest sample size for cognition due to these eligibility criteria.

Figure 2 and Supplementary Figures S3-S5 show that for most health measures, there was no statistical moment with a consistently increasing or decreasing trend over time: there were both increases and decreases between waves. The only notable exception was life expectancy free of cardiovascular disease for females: the point estimate always increased from wave to wave. Comparing wave 9 to wave 2, the point estimate of HLE increased between these two waves, for all health measures (females) or most health measures (males). For both males and females, distributions with negative skewness became more negatively skewed in wave 9 compared to wave 2 (see Supplementary Figure S5).

Probabilities of health loss at specific ages and at mode age

Figure 3 shows that across all health measures, the probabilities of health loss at younger ages were higher for males than for females. However, close to the modes of the healthy longevity distributions, probabilities of health loss were higher for females than for males, and at the oldest ages, probabilities of health loss were either higher for females or similar between sexes (note that for each sex, the sum of all probabilities across all ages is equal to 1). The only exception to this was life free of arthritis or rheumatism, which corresponded to higher probabilities of health loss for males than for females at older ages. Within the same health measure and within the same wave, the probability of health loss at the mode age was always higher for females than for males. Figure 3 shows this for wave 9. For example, in wave 9, for poor cognition relative to all other person-wave observations, the peak probability of health loss was 0.082 (95% CI 0.082 to 0.083) for females (at age 87, 95% CI: 86 to 87) and 0.059 (95% CI: 0.059 to 0.060) for males (at age 86, 95% CI: 85 to 87). Supplementary Figure S12 shows the probability of health loss at mode age for all waves, by health measure and sex. The higher mode for females was consistent with the lower dispersion of the female distributions (which was measured by SDHL). Supplementary Figure S13 shows the relationship between the probability of health loss at mode age and SDHL.

Results from φHL and Hellinger distance

The φHL statistic indicated that across all waves, men had a probability above 50% of living without arthritis or rheumatism for longer than females. Across all waves, for life free of cardiovascular disease, life free of cardiovascular conditions and life with good cognition, men had a probability below 50% of having a longer healthy lifespan than females (there was only one exception when the 95% CI of a good cognition measure crossed the 50% threshold in wave 5). φHL was the lowest for the health measure "no cardiovascular disease" across all waves. In most waves, φHL was closest to 0.5 for "no chronic conditions" and "no 2+ chronic conditions", with the 95% CIs crossing 0.5 in some waves. Figure 4 shows φHL and the Hellinger distance in different waves.

The φHL statistic in wave 9 for cardiovascular disease indicated that if we randomly selected a man and a woman, both without cardiovascular disease at age 60, the man would have a probability of 31% (95% CI: 30% to 32%) of living without cardiovascular disease for a longer time compared to the woman. For arthritis and rheumatism, the φHL statistic in wave 9 indicated that if we randomly selected a man and a woman, both without arthritis or rheumatism at age 60, the man would have a probability of 57% (95% CI: 56% to 59%) of living without arthritis or rheumatism for longer than the woman.

The Hellinger distance was the highest for "no arthritis or rheumatism" and for "no cardiovascular disease", thus indicating the highest dissimilarity between the distributions of males and females when it comes to these health measures. Supplementary Figures S6-S7 show the point estimates of φHL and of the Hellinger distance next to the probability mass functions for wave 2 and wave 9. Figure S8 shows the complementary value of φHL (1- φHL), i.e., the probability of females to have a longer healthy lifespan compared to males, next to the probability mass functions for wave 9.

There was a strong and positive correlation between the distance of φHL from 0.5 and the Hellinger distance (based on point estimates, the Pearson's correlation coefficient was 0.79; see Supplementary Figure S9 for details).

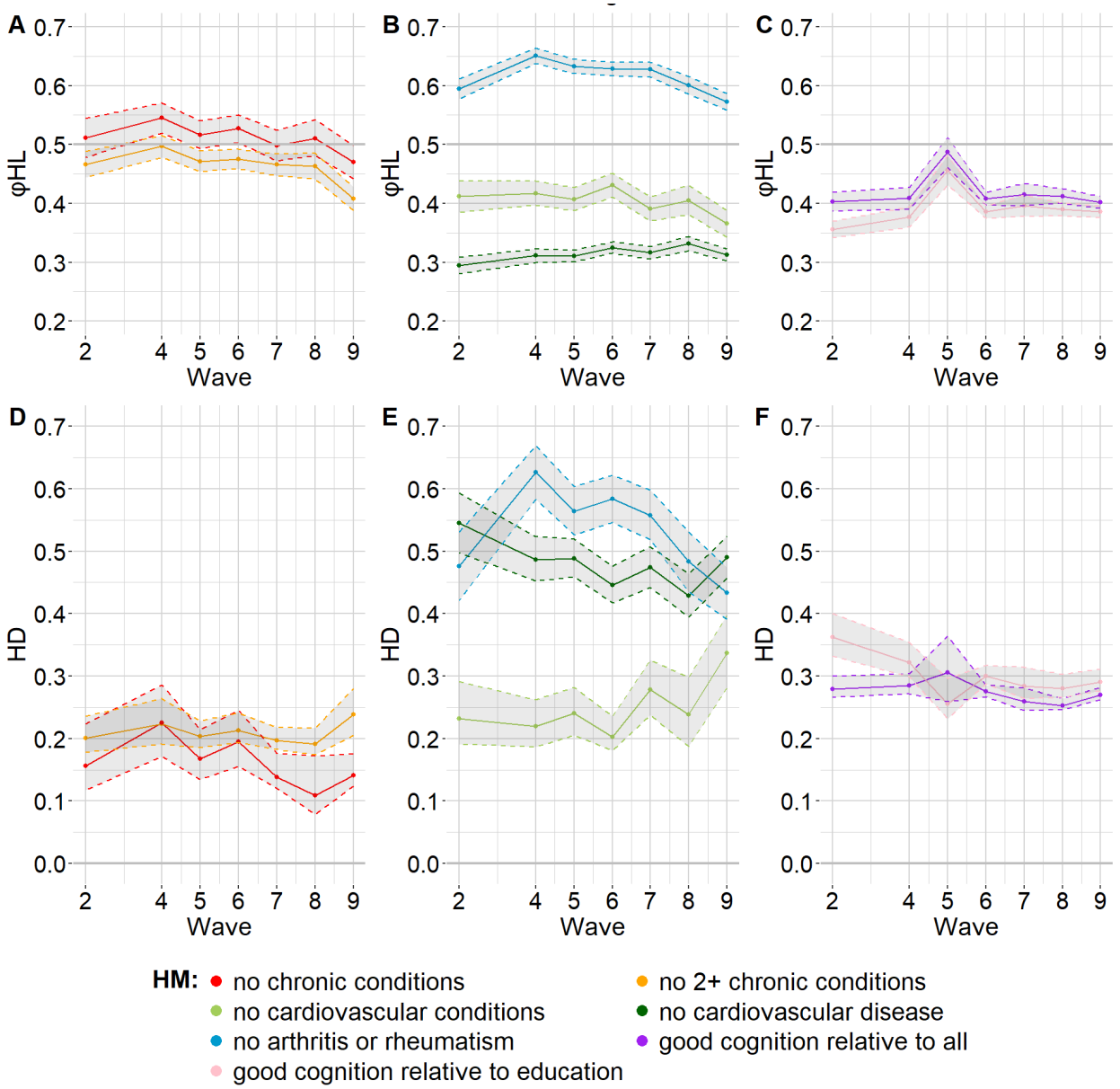

Figure 4. φHL and Hellinger distance over time.

Figure notes. Abbreviations: HD: Hellinger distance; HM: health measure.
The φHL statistic expresses the probability for males to have a longer healthy lifespan than females. The Hellinger distance compares the healthy longevity distributions for males and females.

Discussion

This study compared healthy longevity distributions at age 60 between sexes and between different health measures, focusing on the first three statistical moments, the mode and the probabilities of losing health at specific ages. Moreover, the φHL statistic and the Hellinger distance were used to make formal comparisons between the male and female distributions.

Lifespan free of any chronic conditions corresponded to the lowest HLE, the lowest inter-individual variation and the only measure with positive or null skewness values. For most health measures, healthy longevity distributions for females had higher HLE, lower SDHL, higher mode and more negative skewness compared to the distributions for males, across all waves. An exception was "no arthritis or rheumatism", with higher HLE for males than for females. This was the only health measure for which φHL consistently indicated that men had a probability above 50% of having a longer healthy lifespan than females. Moreover, for two comprehensive health definitions, i.e., "no chronic conditions" and "no 2+ chronic conditions", there was little difference in HLE between males and females, and φHL was the closest to 50% for these two health measures, with the 95% CIs crossing 50% in some waves. The distributions that differed the most between males and females were for life free of arthritis and rheumatism and life free of cardiovascular disease, as indicated by φHL and the Hellinger distance.

To better understand the results, both the differences in mortality and in health prevalence should be considered. Once data had been pooled across all countries, as expected, the annual probability of death was higher for men than for women of the same age, across all waves (Supplementary Figure S11 shows this for wave 2 and wave 9). The prevalence of cardiovascular disease was either similar between males and females or higher for males, depending on age and wave (Supplementary Figure S10 shows health prevalence for wave 9). The combination of these factors resulted in the lowest φHL for life free of cardiovascular disease. Conversely, there was often a higher prevalence of arthritis or rheumatism among women than among men of the same age, which offset the higher male annual probability of death, leading to φHL above 50% for life free of these conditions. For life free of any chronic conditions or life with no more than one chronic condition, prevalence was similar between males and females, although in various wave-age combinations, the point estimates of prevalence were higher for females. When this was combined with the probabilities of death, the distributions of healthy longevity were similar between sexes. Note that the similar prevalence between males and females was somewhat unexpected because a previous study based on SHARE

data had found higher odds of comorbidity among females compared to males (Ahrenfeldt et al., 2019). However, the set of chronic conditions included in the current study only partly overlapped with the set of conditions included in that previous study.

To relate the findings from this study to the female-male health-survival paradox mentioned in the background section, we can note that despite the longer life expectancy of females, lifespan free of any chronic conditions was similar between males and females, and lifespan free of arthritis or rheumatism tended to be longer for males.

The finding that SDHL was higher for males than for females across all health measures and waves was consistent with the findings in (Caswell & Zarulli, 2018), which also used SHARE data but defined healthy longevity differently: either as having no limitations in activities of daily living or based on hand grip strength. In both cases, the SDHL was higher for males than for females at age 55 and 75. In contrast, there were different findings relating to the comparison of SDHL between males and females in two publications based on data from the Global Burden of Disease (GBD) study on 204 countries, from the year 1990 to 2019 (Permanyer et al., 2023; Zarulli & Caswell, 2024). In the GBD data, healthy lifespan was measured based on both years of life lost due to premature mortality and years lived with disability, which were calculated by applying disability weights to a variety of health conditions (Institute for Health Metrics and Evaluation (IHME), 2016). One study (Permanyer et al., 2023) found that at age 65, SDHL was higher for females than for males, on average. Similarly, the other study (Zarulli & Caswell, 2024) found that SDHL at age 65 was higher for females than males in most country-year observations. The opposite findings in studies using SHARE vs. GBD data could be due to the specificities of the GBD definition of healthy longevity.

This study showed that, with the exception of negative skewness values relatively close to 0, negative skewness was reflected in a mode age above the HLE age. At least for some health measures (in particular, those with relatively high probabilities of health loss at age 90), the finding of negative skewness may have been influenced by the artificial limit of age 90 imposed by the analysis: this meant that health loss at very old ages could not be captured by the distributions. Further research with bigger sample sizes could use higher age limits and could attempt to validate the skewness findings versus observational incidence-based studies. Future research could also investigate whether it is more useful to estimate skewness through the Pearson's skewness coefficient rather than through the Markov chain model (Caswell & Zarulli, 2018): the Pearson's coefficient can be calculated through the formula: (HLE age - mode age)/SDHL, so it can be useful to directly relate

different statistical moments and the mode of healthy longevity distributions. Supplementary Figure S15 shows how the skewness values from the Markov chain model compare to the Pearson's skewness coefficients.

Although the present work did not focus on disability-free life expectancy, some of the health measures that we included are related to disability: in particular, previous research found that self-reported arthritis was positively associated with ADL disability, IADL disability and mobility disability (Liu et al., 2024). Moreover, cardiovascular disease can also lead to disability. On the one hand, men had a probability above 50% of living for more years than women without arthritis or rheumatism, but on the other hand, the probability was below 50% when it came to cardiovascular disease, across all waves (see Figure 4). Lee et al. (2021) found no statistically significant differences between sexes in the age gradients of disability incidence in Northern, Central and Eastern Europe in their study focused on people aged 55-89, with disability defined as having any difficulty with ADLs, based on data from the Gateway to Global Aging. They found statistically significant differences in Southern Europe, where higher disability incidence in women was observed in the older age groups (Lee et al., 2021). Further work could investigate how these results relate to the incidence of different chronic conditions by age, sex and region.

Focusing on the inter-wave differences of healthy longevity distributions, there were various waves with unique characteristics, which could have potentially led to distinct results for these waves, but the only wave that stood out visually was wave 5, for the cognition measures. This corresponded to restricted eligibility criteria for the cognition questions and to the lowest sample size. Other peculiarities of specific waves were as follows: narrower eligibility criteria for the cognition questions were implemented not only in wave 5, but also in waves 4 and 7, as mentioned in the methods section; the questions around arthritis and rheumatism were different in wave 2 and 4 compared to the other waves (see Supplementary Table S1 for more details); wave 8 was interrupted prematurely due to the COVID-19 public health emergency (Scherpenzeel et al., 2020). By the time it was interrupted, around 70% of the expected longitudinal interviews and around 50% of the expected refreshment interviews had been completed, although this varied by country (Scherpenzeel et al., 2020). Moreover, for our study, mortality in wave 8 was based on 2019 and 2020, while for wave 9 it was based on 2021 and 2022. Therefore, some Covid-related mortality was incorporated into the healthy longevity estimations. Surprisingly, despite all these aspects, none of these waves stood out visually in the plots created for the current study.

This work has strengths and limitations. One strength is that, to our knowledge, this is the first study to apply the maximum entropy method to derive the full empirical distribution of health loss over the lifespan, and to compare the distributions of healthy longevity between different health measures. We also provide, for the first time, a formalization of the comparison between male and female healthy longevity distributions.

An important limitation is the use of prevalence-based health measures instead of incidence ones. The method assumes that not only death but also the loss of health is an absorbing state and does not allow for transitions back and forth between the healthy and unhealthy state. Because of this, we excluded from the analysis widely-used health measures that do not necessarily capture chronic health issues, such as self-rated health or being free from activity limitations, and we focused on conditions that it is very difficult to recover from, in order to minimize the problem of the possible transition back to full health.

Another limitation is that an average annual probability of death was used in the Markov chain model. However, in practice, there would be different probabilities of death for healthy and non-healthy people.

Moreover, our application of the maximum entropy method seems to underestimate the probabilities of health loss at older ages (towards the right tail of the distributions), and consequently, the probabilities of health loss at some younger ages are likely to be overestimated. See Supplementary Table S9 for details on the validation checks regarding this. Some deviation from empirical data is to be expected, especially because our application of the maximum entropy method only used the first three statistical moments, while the use of additional moments such as kurtosis would improve accuracy (Pascariu et al., 2019). Further research could look into potential ways to estimate and incorporate additional statistical moments.

An additional limitation is that φHL compares randomly-paired individuals, assuming independence between the two populations, however, in practice, some of these individuals would be husbands and wives or cohabiting (Bergeron-Boucher et al., 2022): for people participating from the first wave of SHARE, all household members born in 1954 or earlier were eligible for a SHARE interview. For the new countries entering the survey later and for the refreshment samples, there was only one selected respondent per household, but current partners living in the same household were also interviewed (SHARE, 2024).

Furthermore, it should be noted that the health prevalence data are based on self-report of doctor's diagnoses. The age-specific prevalence used to estimate the healthy longevity distributions may be linked to frequency of attendance at health care services, which may vary by sex. For example, a literature review on attendance at general health checks (including screening for cardiovascular disease) found that women had higher rates of attendance than men (Dryden et al., 2012). In turn, more frequent attendance could lead to earlier diagnosis.

This study is the first of its kind, and opens the way to a variety of research avenues. For example, in future work, it would be important to compare healthy longevity distributions not only between sexes but also between different educational levels. Bergeron-Boucher et al. (2022) point out that "males with a university degree or who are married have a higher chance of outliving females, in particular females with a lower educational level and who are single" (p. 6). Further research could also study coupled men and women. However, this kind of study could be done only on the SHARE countries (or other similar survey data) that have performed record linkage with the mortality registries, in order to use the correct probabilities of death, which refer to the subpopulations of individuals connected by the fact of being in a relation. Additionally, future work could investigate healthy longevity distributions for older ages, specific countries or multi-country European regions.

Conclusions

The maximum entropy method was useful to extend the focus of healthy longevity research beyond the statistical moments of distributions, namely by investigating the mode and by formalizing the comparisons between males and females through the φHL statistic and the Hellinger distance. This study has applied these measures for the first time in the healthy longevity field. As expected, the choice of health measure affected the comparison between sexes, but some findings were persistent across all health measures: males had higher inter-individual variation than females and females had a higher distribution mode. Males and females differed the least in their distributions when all chronic conditions were considered together; they differed the most in life free of cardiovascular disease and life free of arthritis or rheumatism.

Abbreviations

HLE	Healthy life expectancy
HL	Healthy longevity
HD	Hellinger distance
HMD	Human Mortality Database
NUTS1	Nomenclature of Territorial Units for Statistics – Level 1
SDHL	Standard deviation of healthy longevity
SHARE	Survey of Health, Ageing and Retirement in Europe

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets analysed during the current study are available from the SHARE website (https://share-eric.eu/data/) and HMD website (https://www.mortality.org/) .

Competing interests

The authors declare no competing interests.

Funding

This research was supported by the “Differences in Healthy Years of Life Between Men and Women: an innovative perspective on the Gender Health-Survival Paradox” grant awarded by the Fondazione Cariparo (CUP C93C22008360007) to Prof. Virginia Zarulli.

Author contributions

RC: formal analysis, software, data curation, visualization, methodology, writing original draft, writing - review and editing

VDL: writing - review and editing

VZ: conceptualization, funding acquisition, resources, project administration, formal analysis, software, data curation, visualization, methodology, supervision, writing - review and editing

Acknowledgment regarding the use of SHARE data

This paper uses data from SHARE Waves 2, 4, 5, 6, 7, 8 and 9 (DOIs: 10.6103/SHARE.w2.900, 10.6103/SHARE.w4.900, 10.6103/SHARE.w5.900, 10.6103/SHARE.w6.900, 10.6103/SHARE.w6.DBS.100, 10.6103/SHARE.w7.900, 10.6103/SHARE.w8.900, 10.6103/SHARE.w8ca.900, 10.6103/SHARE.w9.900, 10.6103/SHARE.w9ca900, 10.6103/SHARE.HCAP.0) see Börsch-Supan et al. (2013) for methodological details.

The SHARE data collection has been funded by the European Commission, DG RTD through FP5 (QLK6-CT-2001-00360), FP6 (SHARE-I3: RII-CT-2006-062193, COMPARE: CIT5-CT-2005-028857, SHARELIFE: CIT4-CT-2006-028812), FP7 (SHARE-PREP: GA N°211909, SHARE-LEAP: GA N°227822, SHARE M4: GA N°261982, DASISH: GA N°283646) and Horizon 2020 (SHARE-DEV3: GA N°676536, SHARE-COHESION: GA N°870628, SERISS: GA N°654221, SSHOC: GA N°823782, SHARE-COVID19: GA N°101015924) and by DG Employment, Social Affairs & Inclusion through VS 2015/0195, VS 2016/0135, VS 2018/0285, VS 2019/0332, VS 2020/0313, SHARE-EUCOV: GA N°101052589 and EUCOVII: GA N°101102412. Additional funding from the German Federal Ministry of Education and Research (01UW1301, 01UW1801, 01UW2202), the Max Planck Society for the Advancement of Science, the U.S. National Institute on Aging (U01_AG09740-13S2, P01_AG005842, P01_AG08291, P30_AG12815, R21_AG025169, Y1-AG-4553-01, IAG_BSR06-11, OGHA_04-064, BSR12-04, R01_AG052527-02, R01_AG056329-02, R01_AG063944, HHSN271201300071C, RAG052527A) and from various national funding sources is gratefully acknowledged (see www.share-eric.eu).

This paper uses data from the generated easySHARE data set (DOI: 10.6103/SHARE.easy.900), see Gruber (2019) for methodological details. The easySHARE release 9.0.0 is based on SHARE Waves 1, 2, 3, 4, 5, 6, 7, 8 and 9 (DOIs: 10.6103/SHARE.w1.900 for wave 1, 10.6103/SHARE.w3.900 for wave 3, and DOIs mentioned above for the other waves).

References

Ahrenfeldt, L. J., Möller, S., Thinggaard, M., Christensen, K., & Lindahl-Jacobsen, R. (2019). Sex Differences in Comorbidity and Frailty in Europe. International Journal of Public Health, 64(7), 1025–1036. https://doi.org/10.1007/s00038-019-01270-9

Bergeron-Boucher, M.-P., Alvarez, J.-A., Kashnitsky, I., Zarulli, V., 2022. Probability of males to outlive females: an international comparison from 1751 to 2020. https://doi.org/10.1136/bmjopen-2021-059964

Bergmann, M., Kneip, T., De Luca, G., Scherpenzeel, A., 2019a. Survey participation in the Survey of Health, Ageing and Retirement in Europe (SHARE), wave 1-7. https://share-eric.eu/fileadmin/user_upload/SHARE_Working_Paper/WP_Series_41_2019_Bergmann_et_al.pdf

Bergmann, M., Scherpenzeel, A., Börsch-Supan, A., 2019b. SHARE wave 7 methodology: panel innovations and life histories. https://share-eric.eu/fileadmin/user_upload/Methodology_Volumes/SHARE_Methodenband_A4_WEB_Wave7MFRB.pdfBörsch-Supan, A., Brandt, M., Hunkler, C., Kneip, T., Korbmacher, J., Malter, F., Schaan, B., Stuck, S., Zuber, S., 2013. Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). International Journal of Epidemiology 42, 992–1001. https://doi.org/10.1093/ije/dyt088

Caswell, H., & Daalen, S. van. (2021). Healthy longevity from incidence-based models: More kinds of health than stars in the sky. Demographic Research, 45, 397–452. https://doi.org/10.4054/DemRes.2021.45.13

Caswell, H., Zarulli, V., 2018. Matrix methods in health demography: a new approach to the stochastic analysis of healthy longevity and DALYs. Population Health Metrics 16, 8. https://doi.org/10.1186/s12963-018-0165-5

Crimmins, E.M., Kim, J.K., Hagedorn, A., 2002. Life With and Without Disease: Women Experience More of Both. Journal of Women & Aging. https://doi.org/10.1300/J074v14n01_04

Di Lego, V., Di Giulio, P., & Luy, M. (2020). Gender Differences in Healthy and Unhealthy Life Expectancy. In C. Jagger, E. M. Crimmins, Y. Saito, R. T. De Carvalho Yokota, H. Van Oyen, & J.-M. Robine (A c. Di), *International Handbook of Health Expectancies* (pp. 151–172). Springer International Publishing. https://doi.org/10.1007/978-3-030-37668-0_11

Doblhammer, G., & Kytir, J. (2001). Compression or expansion of morbidity? Trends in healthy-life expectancy in the elderly Austrian population between 1978 and 1998. Social Science & Medicine (1982), 52(3), 385–391. https://doi.org/10.1016/s0277-9536(00)00141-6

Dryden, R., Williams, B., McCowan, C., Themessl-Huber, M., 2012. What do we know about who does and does not attend general health checks? Findings from a narrative scoping review. BMC Public Health 12, 723. https://doi.org/10.1186/1471-2458-12-723

Geyer, S., & Eberhard, S. (2022). Compression and Expansion of Morbidity—Secular Trends Among Cohorts of the Same Age. Deutsches Arzteblatt International, 119(47), 810–815. https://doi.org/10.3238/arztebl.m2022.0324

Gruber, S., 2019. Generating easySHARE release 7.0.0. Guidelines, structure, content and programming. Working Paper Series 37-2019. https://share-eric.eu/fileadmin/user_upload/SHARE_Working_Paper/WP_Series_37_2019_Gruber.pdf

Gray, A., Clarke, P., Wolstenholme, J., Wordsworth, S., 2011. Applied methods of cost-effectiveness analysis in health care. Oxford University Press.

Guillerme, T., Cooper, N., 2016. Effects of missing data on topological inference using a Total Evidence approach. Molecular Phylogenetics and Evolution 94, 146–158. https://doi.org/10.1016/j.ympev.2015.08.023

Han, F.-F., Wang, H.-X., Wu, J.-J., Yao, W., Hao, C.-F., Pei, J.-J., 2021. Depressive symptoms and cognitive impairment: A 10-year follow-up study from the Survey of Health, Ageing and Retirement in Europe. European Psychiatry 64, e55. https://doi.org/10.1192/j.eurpsy.2021.2230

Human Mortality Database (HMD), 2024. Period Data. Life Tables—Male. Life Tables—Female. https://www.mortality.org/Data/ZippedDataFiles

Institute for Health Metrics and Evaluation (IHME), 2016. Glossary of Terms. https://www.healthdata.org/sites/default/files/files/policy_report/GBD/2016/IHME_GBD2015_Report_Glossary-of-terms_2016.pdf

Klar, M. K., Geyer, S., Safieddine, B., Tetzlaff, F., Tetzlaff, J., & Sperlich, S. (2021). Trends in healthy life expectancy between 2002 and 2018 in Germany—Compression or expansion of health-related quality of life (HRQOL)? SSM - Population Health, 13, 100758. https://doi.org/10.1016/j.ssmph.2021.100758

Lego, V. di, & Sauerberg, M. (2023). The Sensitivity of the Healthy Life Years Indicator: Approaches for Dealing with Age-Specific Prevalence Data. Comparative Population Studies, 48. https://doi.org/10.12765/CPoS-2023-06

Liu, X., Huang, Y., Fu, J., Mohedaner, M., Danzengzhuoga, Yang, G., et al. Associations of arthritis with functional disability and depressive symptoms in general US adults: NHANES 1988–1994 and 1999–2018. AGING MEDICINE 7, 705–716. https://doi.org/10.1002/agm2.12379

Murray, C. J., Salomon, J. A., & Mathers, C. (2000). A critical examination of summary measures of population health. Bulletin of the World Health Organization, 78(8), 981–994.

Namavari, N., Jokar, M., Ghodsian, A., Jahromi, H. K., & Rahmanian, V. (2024). Menopausal state and rheumatoid arthritis: A systematic review and meta-analysis. BMC Rheumatology, 8(1), 48. https://doi.org/10.1186/s41927-024-00418-2

Nebel, R. A., Aggarwal, N. T., Barnes, L. L., Gallagher, A., Goldstein, J. M., Kantarci, K., Mallampalli, M. P., Mormino, E. C., Scott, L., Yu, W. H., Maki, P. M., & Mielke, M. M. (2018). Understanding the impact of sex and gender in Alzheimer's disease: A call to action. Alzheimer's & Dementia: The Journal of the Alzheimer's Association, 14(9), 1171–1183. https://doi.org/10.1016/j.jalz.2018.04.008

Nilsson, N., Håkansson, B., Ortiz-Catalan, M., 2017. Classification complexity in myoelectric pattern recognition. Journal of NeuroEngineering and Rehabilitation 14, 68. https://doi.org/10.1186/s12984-017-0283-5

Ntritsos, G., Franek, J., Belbasis, L., Christou, M. A., Markozannes, G., Altman, P., Fogel, R., Sayre, T., Ntzani, E. E., & Evangelou, E. (2018). Gender-specific estimates of COPD prevalence: A systematic review and meta-analysis. International Journal of Chronic Obstructive Pulmonary Disease, 13, 1507–1514. https://doi.org/10.2147/COPD.S146390

Oksuzyan, A., Juel, K., Vaupel, J. W., & Christensen, K. (2008). Men: Good health and high mortality. Sex differences in health and aging. Aging Clinical and Experimental Research, 20(2), 91–102. https://doi.org/10.1007/BF03324754

Pascariu, M.D., Lenart, A., Canudas-Romo, V., 2019. The maximum entropy mortality model: forecasting mortality using statistical moments. Scandinavian Actuarial Journal 2019, 661–685. https://doi.org/10.1080/03461238.2019.1596974

Payne, C. F. (2022). Expansion, Compression, Neither, Both? Divergent Patterns in Healthy, Disability-Free, and Morbidity-Free Life Expectancy Across U.S. Birth Cohorts, 1998-2016. Demography, 59(3), 949–973. https://doi.org/10.1215/00703370-9938662

Permanyer, I., & Bramajo, O. (2023). The Race between Mortality and Morbidity: Implications for the Global Distribution of Health. Population and Development Review, 49(4), 909–937. https://doi.org/10.1111/padr.12582

Permanyer, I., Villavicencio, F., Trias-Llimós, S., 2023. Healthy lifespan inequality: morbidity compression from a global perspective. Eur J Epidemiol 38, 511–521. https://doi.org/10.1007/s10654-023-00989-3

Phillips, D., Hu, P., 2021. Disability Incidence Rates for Men and Women in 23 Countries: Evidence on Health Effects of Gender Inequality. The Journals of Gerontology: Series A 76, 328–338. https://doi.org/10.1093/gerona/glaa288

Reinwarth, A. C., Wicke, F. S., Rückert, K. K., Schattenberg, J. M., Tüscher, O., Wild, P. S., Münzel, T., König, J., Lackner, K. J., Pfeiffer, N., & Beutel, M. E. (2024). Change of self-rated physical health predicts mortality in aging individuals: Results of a population-based cohort study. Archives of Public Health, 82(1), 130. https://doi.org/10.1186/s13690-024-01363-9

Scherpenzeel, A., Axt, K., Bergmann, M., Douhou, S., Oepen, A., Sand, G., Schuller, K., Stuck, S., Wagner, M., Börsch-Supan, A., 2020. Collecting survey data among the 50+ population during the COVID-19 outbreak: The Survey of Health, Ageing and Retirement in Europe (SHARE). Survey Research Methods 14, 217–221. https://doi.org/10.18148/srm/2020.v14i2.7738

Schroeder, S. A. (2012). Incidence, prevalence, and hybrid approaches to calculating disability-adjusted life years. *Population Health Metrics*, *10*(1), 19. https://doi.org/10.1186/1478-7954-10-19

Seaman, R., Höhn, A., Lindahl-Jacobsen, R., Martikainen, P., van Raalte, A., & Christensen, K. (2020). Rethinking morbidity compression. European Journal of Epidemiology, 35(5), 381–388. https://doi.org/10.1007/s10654-020-00642-3

SHARE, 2006. SHARE w2 questionnaire version 2.7 2006-09-21 https://share-eric.eu/fileadmin/user_upload/Questionnaires/Q-Wave_2/w2_en_capi_main-Generic.pdf

SHARE, n.d.a Main questionnaire & end of life questionnaire. Wave 5 https://share-eric.eu/fileadmin/user_upload/Questionnaires/Q-Wave_5/w5_en_capi_main-Generic.pdf

SHARE, n.d.b Main questionnaire & end of life questionnaire. Wave 6 https://share-eric.eu/fileadmin/user_upload/Questionnaires/Q-Wave_6/w6_en_capi_main-Generic.pdf

SHARE, n.d.c Main questionnaire. Wave 4. https://share-eric.eu/fileadmin/user_upload/Questionnaires/Q-Wave_4/w4_en_capi_main-Generic.pdf

SHARE, n.d.d Main questionnaire. Wave 7. https://share-eric.eu/fileadmin/user_upload/Questionnaires/Q-Wave_7/w7_en_capi_main-Generic.pdf

SHARE, n.d.e Main questionnaire. Wave 8. https://share-eric.eu/fileadmin/user_upload/Questionnaires/Q-Wave_8/paperverstion_en_GB_8_2_5b.pdf

SHARE, n.d.f Main questionnaire. Wave 9. https://share-eric.eu/fileadmin/user_upload/Questionnaires/Q-Wave_9/paperversion_en_GB_9_2_2a.pdf

SHARE, 2024. Survey of Health, Ageing and Retirement in Europe (SHARE) Release Guide 9.0.0. https://share-eric.eu/fileadmin/user_upload/Release_Guides/SHARE_release_guide_9-0-0.pdf

SHARE, 2022. Survey of Health, Ageing and Retirement in Europe (SHARE) Release Guide 8.0.0 https://share-eric.eu/fileadmin/user_upload/Release_Guides/SHARE_release_guide_8-0-0.pdf

SHARE-ERIC, 2024. easySHARE. Release version: 9.0.0. SHARE-ERIC. Dataset. https://share-eric.eu/data/data-set-details/easyshare

SHARE-ERIC, 2024a. Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 2. Release version: 9.0.0. SHARE-ERIC. Data set. https://share-eric.eu/data/data-documentation/waves-overview/wave-2

SHARE-ERIC, 2024b. Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 4. Release version: 9.0.0. SHARE-ERIC. Data set. https://share-eric.eu/data/data-documentation/waves-overview/wave-4

SHARE-ERIC, 2024c. Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 5. Release version: 9.0.0. SHARE-ERIC. Data set. https://share-eric.eu/data/data-documentation/waves-overview/wave-5

SHARE-ERIC, 2024d. Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 6. Release version: 9.0.0. SHARE-ERIC. Data set. https://share-eric.eu/data/data-documentation/waves-overview/wave-6

SHARE-ERIC, 2024e. Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 7. Release version: 9.0.0. SHARE-ERIC. Data set. https://share-eric.eu/data/data-documentation/waves-overview/wave-7

SHARE-ERIC, 2024f. Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 8. Release version: 9.0.0. SHARE-ERIC. Data set. https://share-eric.eu/data/data-documentation/waves-overview/wave-8

SHARE-ERIC, 2024g. Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 9. Release version: 9.0.0. SHARE-ERIC. Data set. https://share-eric.eu/data/data-documentation/waves-overview/wave-9

Townsend, N., Nichols, M., Scarborough, P., Rayner, M., 2015. Cardiovascular disease in Europe — epidemiological update 2015. European Heart Journal 36, 2696–2705. https://doi.org/10.1093/eurheartj/ehv428

Van Oyen, H., Cox, B., Jagger, C., Cambois, E., Nusselder, W., Gilles, C., & Robine, J.-M. (2010). Gender gaps in life expectancy and expected years with activity limitations at age 50 in the European Union: Associations with macro-level structural indicators. European Journal of Ageing, 7(4), 229–237. https://doi.org/10.1007/s10433-010-0172-2

Vaupel, J.W., Bergeron-Boucher, M.-P., Kashnitsky, I., 2021. Outsurvival as a measure of the inequality of lifespans between two populations. Demographic Research 44, 853–864. https://www.jstor.org/stable/27032937

Vogel, B., Acevedo, M., Appelman, Y., Bairey Merz, C. N., Chieffo, A., Figtree, G. A., Guerrero, M., Kunadian, V., Lam, C. S. P., Maas, A. H. E. M., Mihailidou, A. S., Olszanecka, A., Poole, J. E., Saldarriaga, C., Saw, J., Zühlke, L., & Mehran, R. (2021). The Lancet women and cardiovascular disease Commission: Reducing the global burden by 2030. Lancet (London, England), 397(10292), 2385–2438. https://doi.org/10.1016/S0140-6736(21)00684-X

Whitson, H. E., Landerman, L. R., Newman, A. B., Fried, L. P., Pieper, C. F., & Cohen, H. J. (2010). Chronic medical conditions and the sex-based disparity in disability: The Cardiovascular Health Study. The Journals of Gerontology. Series A, Biological Sciences and Medical Sciences, 65(12), 1325–1331. https://doi.org/10.1093/gerona/glq139

World Health Organization, 2024. Life expectancy at age 60 (years). https://www.who.int/data/gho/data/indicators/indicator-details/GHO/life-expectancy-at-age-60-(years)

Zarulli, V., Caswell, H., 2024. Longer healthy life, but for how many? A stochastic analysis of healthy lifespan inequality. Ann Oper Res. https://doi.org/10.1007/s10479-024-06203-1

Zheng, F., Yan, L., Yang, Z., Zhong, B., Xie, W., 2018. HbA1c, diabetes and cognitive decline: the English Longitudinal Study of Ageing. Diabetologia 61, 839–848. https://doi.org/10.1007/s00125-017-4541-7

Supplementary Material

Table S1. Details on the health measures

Chronic condition outlined in the main text	**More details (citations are from the SHARE questionnaires)** (SHARE, 2006, n.d.c, n.d.a, n.d.b, n.d.d, n.d.e, n.d.f)
Heart problems	The questionnaire asked about: "A heart attack including myocardial infarction or coronary thrombosis or any other heart problem including congestive heart failure"
Chronic lung disease	The questionnaire asked about: "chronic lung disease such as chronic bronchitis or emphysema"
Arthritis or rheumatism	In waves 2 and 4, the questionnaire asked about "arthritis, including osteoarthritis, or rheumatism". In waves 5 to 9, two separate questions were asked, one about "rheumatoid arthritis" and one about "osteoarthritis, or other rheumatism".
Dementia or Alzheimer's diagnosis	The questionnaire asked about "Alzheimer's disease, dementia, organic brain syndrome, senility or any other serious memory impairment".
Cognitive measures	The cognitive tests taken into account were: • Immediate recall (how many words can be recalled after listening to a list of 10 words); • Delayed recall (how many words can be recalled from a list of 10 words after some time); • Orientation in time (four questions about the current year, month, day of the month and day of the week); • Verbal fluency (the person is asked to name as many animals as they can in 60 seconds). For each person, the z-score for each test was calculated. For "poor cognition relative to all person-wave observations", this z-score was relative to all observations across all countries and waves. For "poor cognition relative to those of the same educational level", the z-score was relative to person-wave observations with the same educational level. For each person, the z-scores from different tests were averaged together. Then, this average was standardized into its own z-score. (A similar approach was used in (Zheng et al., 2018)). This was used to assess whether a person's cognitive score was at least 1.5 SDs below the mean. The same threshold of 1.5 SDs below the mean had been used in (Han et al., 2021) to define mild cognitive impairment. People with a cognitive score at least 1.5 SDs below the mean or with a dementia or Alzheimer's diagnosis were classified as with poor cognition. People with a cognitive score above the threshold and with no dementia or Alzheimer's diagnosis were classified as with good cognition. For poor cognition relative to those of the same educational levels, these were the levels: none or primary education; lower secondary education; upper secondary education or above; still in school or other; missing ("don't know/refusal" or "not asked in this wave" or "no information") (this was based on the isced1997_r variable in easySHARE).

Table notes. Note that we included the chronic conditions for which there were data across all relevant SHARE waves. For chronic conditions (health measures 1-6), the questionnaire differed depending on whether the person was a baseline or longitudinal respondent: For baseline respondents, the question was "Has a doctor ever told you that you had any condition ..." whereas for longitudinal respondents the question was "Do you currently have any condition ...". (SHARE, 2024) (p.21).

Table S2. Details on the merge between SHARE and HMD.

Details on the merge	Some SHARE waves were run across different calendar years, so we calculated a weighted average of the annual probability of death (qx) for each country across different calendar years. The weights were the proportions of people aged 50 or over interviewed in each year of the wave for a specific country. For example, in Austria, wave 2 was conducted in 2006/07. 26.6% of people aged 50 and over were interviewed in 2006 and 73.4% in 2007. For a male of age 60, qx was 0.01069 in 2006 and 0.01026 in 2007. The weighted average of qx for males of age 60 in Austria in wave 2 was calculated as follows: 0.01069*0.266+0.01026*0.734 = 0.01037
Exceptions	Exceptions to the above were Spain, Italy and the Czech Republic in wave 9. These countries had SHARE data for wave 9 (2021/2022) but only HMD data for 2021, with no HMD data for 2022, so mortality from 2021 was assigned for wave 9.

Table S3. Weighted average of qx when pooling data across countries.

Calculation	After calculating a weighted average of qx for each country-wave combination, data were pooled across all countries by wave, sex, age. In order to do this, a weighted average (WA) of qx was calculated with this formula: $WA(qx)_{wave\_w, sex\_s, age\_a} =$ $(qx_{wave\_w, sex\_s, age\_a, country\_1} * sample_{wave\_w, sex\_s, age\_a, country\_1} +$ $qx_{wave\_w, sex\_s, age\_a, country\_2} * sample_{wave\_w, sex\_s, age\_a, country\_2} +$ $.... +$ $qx_{wave\_w, sex\_s, age\_a, country\_n} * sample_{wave\_w, sex\_s, age\_a, country\_n})$ $/$ $(sample_{wave\_w, sex\_s, age\_a, country\_1} +$ $sample_{wave\_w, sex\_s, age\_a, country\_2} +$ $.... +$ $sample_{wave\_w, sex\_s, age\_a, country\_n})$
Note	Note that the sample referred to people who had available data on the variables of interest, so it changed depending on whether the focus was on chronic conditions or on poor cognition.
Example	In wave 1, the N of women aged 55 who had data on whether they had a chronic condition was 19 for Austria and 73 for Belgium. So when qx for women aged 55 in wave 1 across all countries was calculated, Austria had a weight of 19 and Belgium of 73. This is consistent with the calculation of health prevalence across all countries, where Belgium also had a bigger weight than Austria.

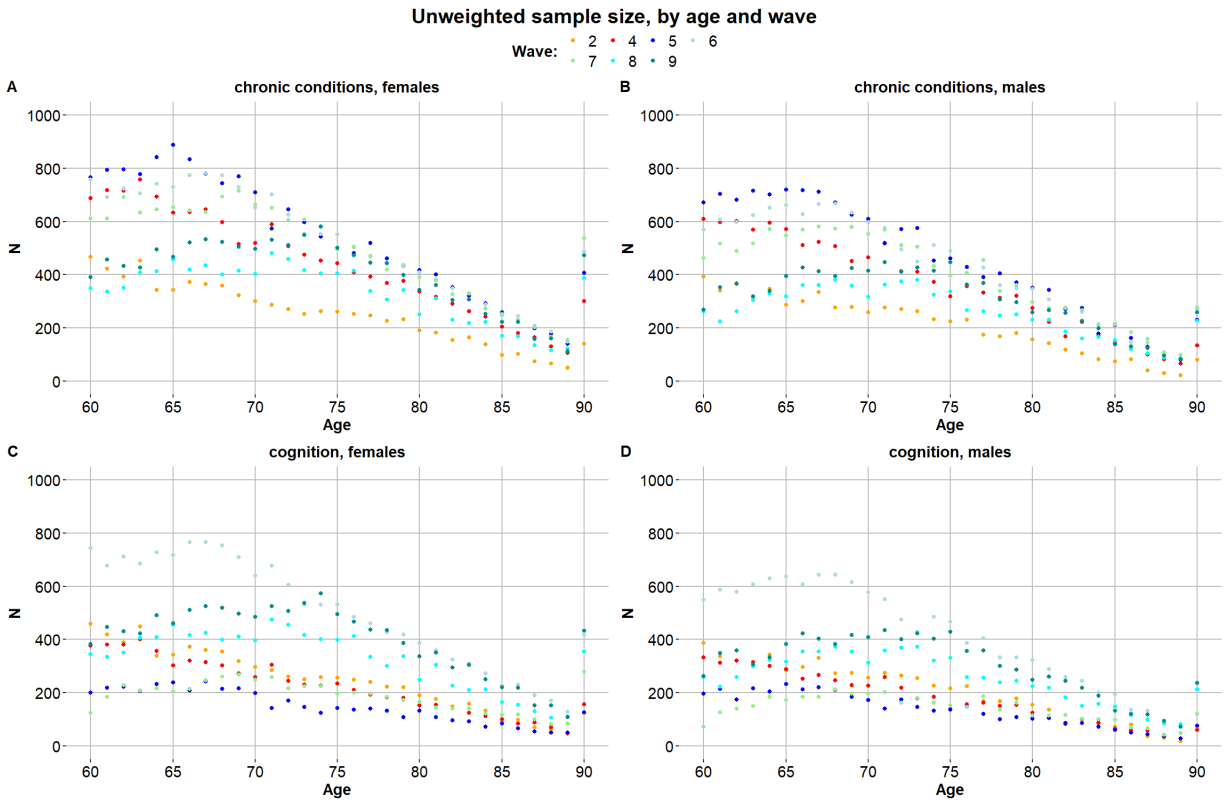

Figure S1. Unweighted sample size by age and wave

Table S4. Unweighted sample size by wave, people with information on chronic conditions.

Wave	Sample size
2	14,240
4	24,734
5	30,731
6	29,516
7	27,507
8	18,323
9	22,228
TOT	167,279

Table S5. Unweighted sample size by wave, people with available data to define poor cognition (i.e., either with a dementia diagnosis or with data on all relevant cognitive tests).

Wave	Sample size
2	13,969
4	12,403
5	8,781
6	28,309
7	10,080
8	17,949
9	21,715
TOT	113,206

Table notes: in waves 4 and 5, differently from the other waves, time orientation questions were only asked to baseline respondents.

Table S6. Baseline vs. longitudinal respondents and data availability on cognition, wave 4 vs. wave 5.

Wave	Baseline respondents		Longitudinal respondents		No classification of baseline or longitudinal respondent			
	With data to define poor cognition	No data to define poor cognition	With dementia	No dementia	With dementia	No dementia but data to define poor cognition	No dementia and no data to define poor cognition	Tot
4	12,085 (48.8%)	312 (1.3%)	317 (1.3%)	11,999 (48.5%)	1	0	20 (0.1%)	24,734 (100%)
5	8,199 (26.7%)	304 (1.0%)	579 (1.9%)	21,631 (70.4%)	1	2	15 (0.05%)	30,731 (100%)

Table S7. People who did vs. did not respond to a SHARELIFE questionnaire in wave 7 and people with available data to define poor cognition.

Did not respond to SHARELIFE and so had regular wave 7 questionnaire		Responded to SHARELIFE and so had condensed wave 7 questionnaire		TOT
Had relevant data to define poor cognition	Had no relevant data to define poor cognition	Had dementia	No dementia	
9531	240	549	17187	27507

Table S8. Definitions of correlation strength for the current work.

Value of ρ	Interpretation of correlation strength
>-0.2 & <0 OR >0 & <0.2	Very weak
>-0.4 & <=-0.2 OR >=0.2 & <0.4	Weak
>-0.6 & <=-0.4 OR >=0.4 & <0.6	Moderate
>-0.8 & <=-0.6 OR >=0.6 & <0.8	Strong
>-1 & <=-0.8 OR >=0.8 & <1	Very strong

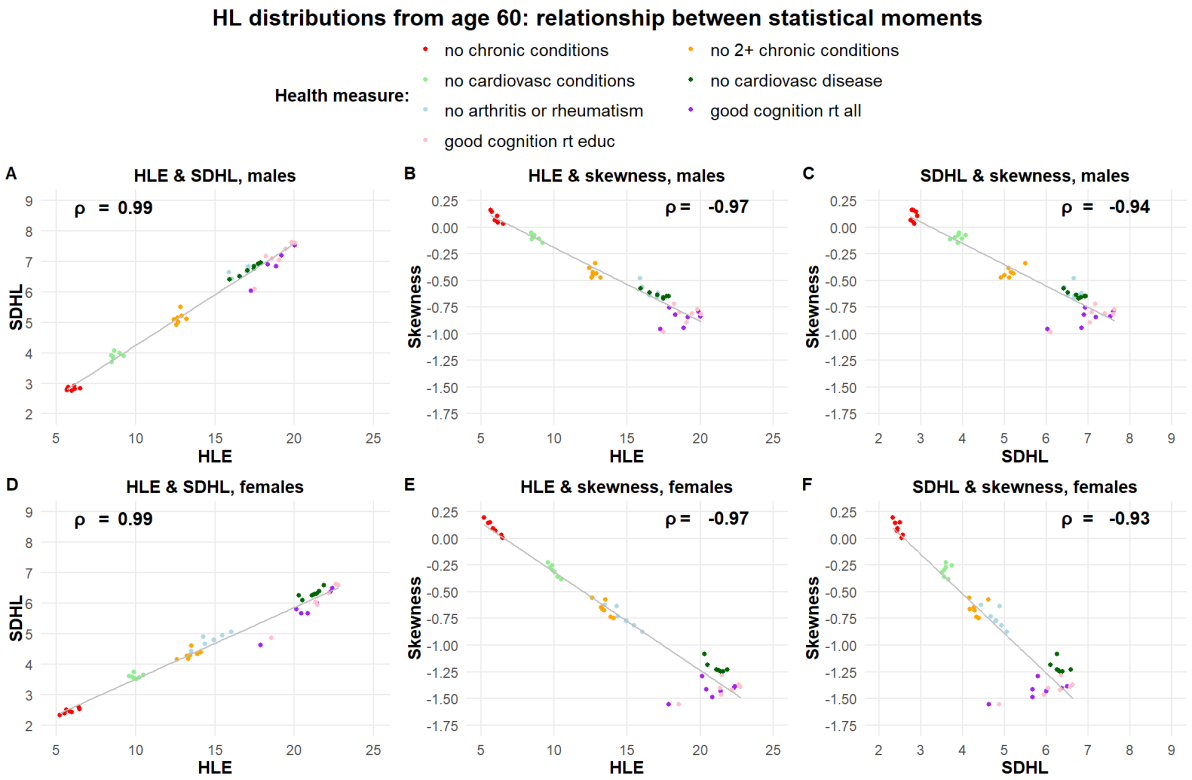

Figure S2. Correlations between the first three statistical moments of healthy longevity distributions, within each sex.

Figure notes. Abbreviations: cardiovasc: cardiovascular; HLE: healthy life expectancy (mean of the healthy longevity distribution); rt all: relative to all person-year observations; rt educ: relative to person-year observations of the same educational level; SDHL: SD of healthy longevity.

ρ refers to the Pearson's correlation coefficient. For each statistical moments, there are 7 observations for males and 7 for females, corresponding to the 7 SHARE waves included in the analysis.

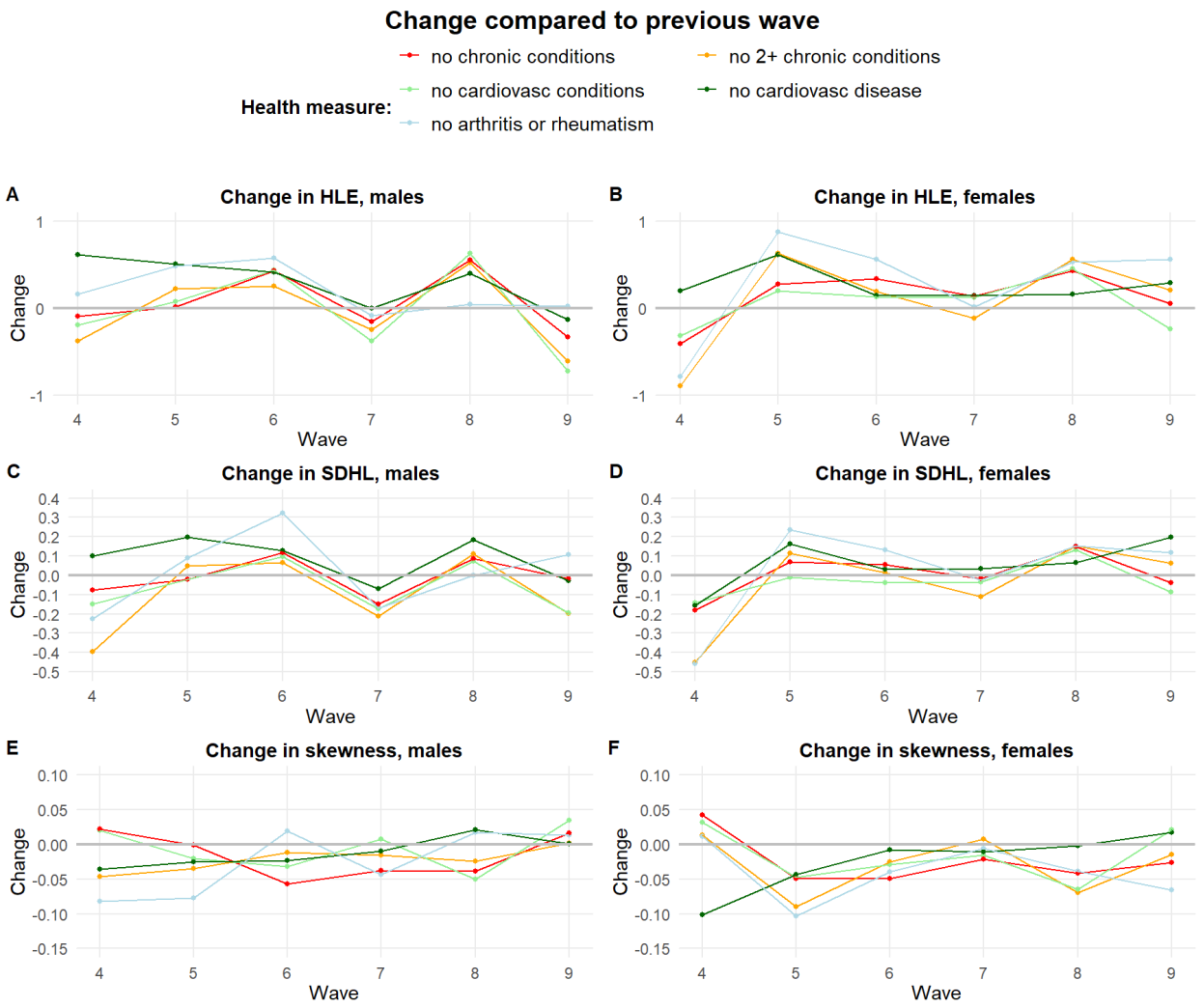

Figure S3. Change in HLE, SDHL and skewness compared to the previous wave, by sex and health measure. All measures except the "good cognition measures".

Figure notes. Abbreviations: Cardiovasc: cardiovascular; HLE: healthy life expectancy; SDHL: standard deviation of healthy longevity. When the change value is positive, there has been an increase compared to the previous wave. When the change value is negative, there has been a decrease compared to the previous wave. Figure based on point estimates.

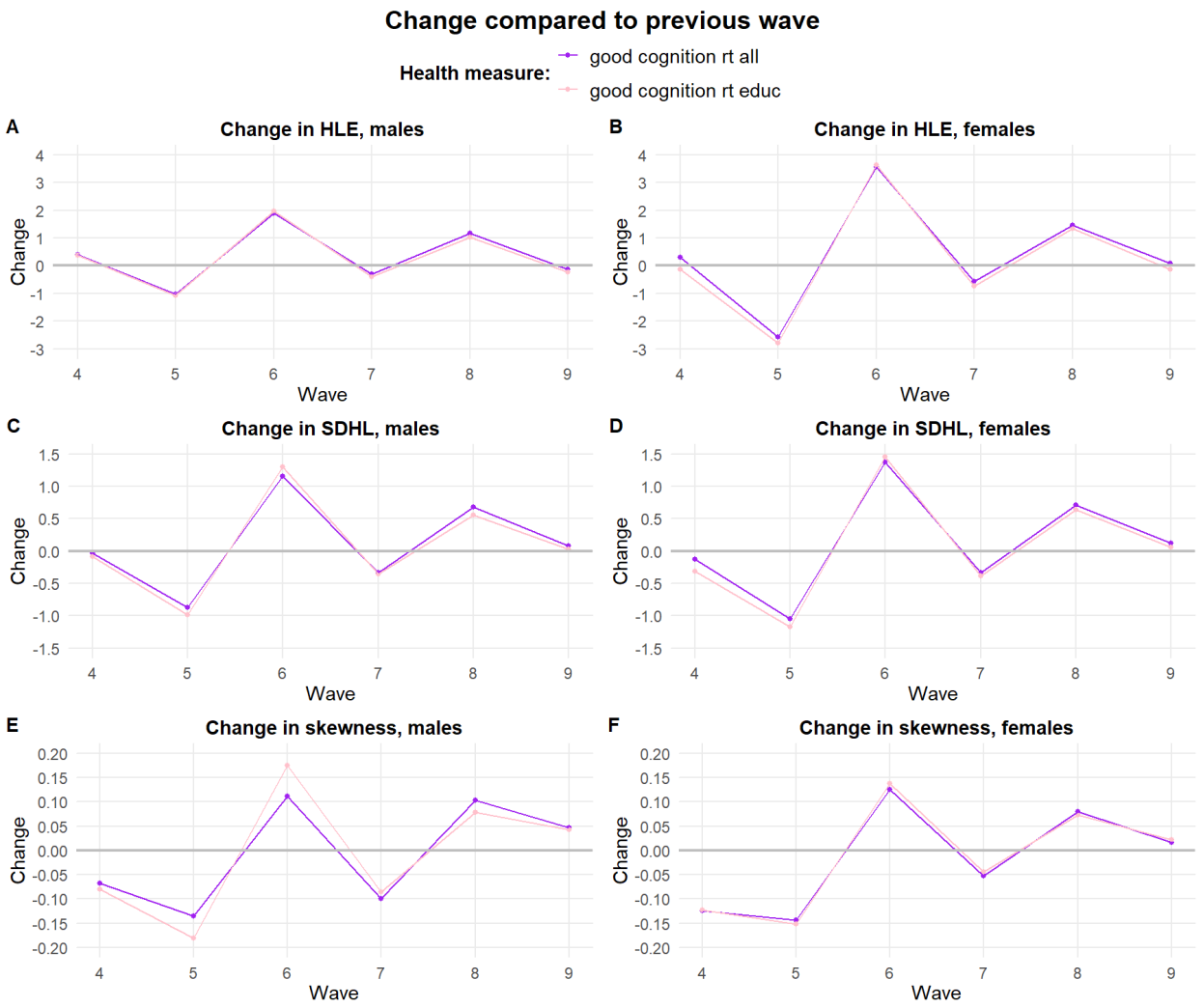

Figure S4. Change in HLE, SDHL and skewness compared to the previous wave, by sex and health measure. Good cognition measures.

Figure notes. Abbreviations: HLE: healthy life expectancy; rt to all: relative to all person-wave observations; rt educ: relative to the person-wave observations of the same educational level; SDHL: standard deviation of healthy longevity. When the change value is positive, there has been an increase compared to the previous wave. When the change value is negative, there has been a decrease compared to the previous wave.
Figure based on point estimates.

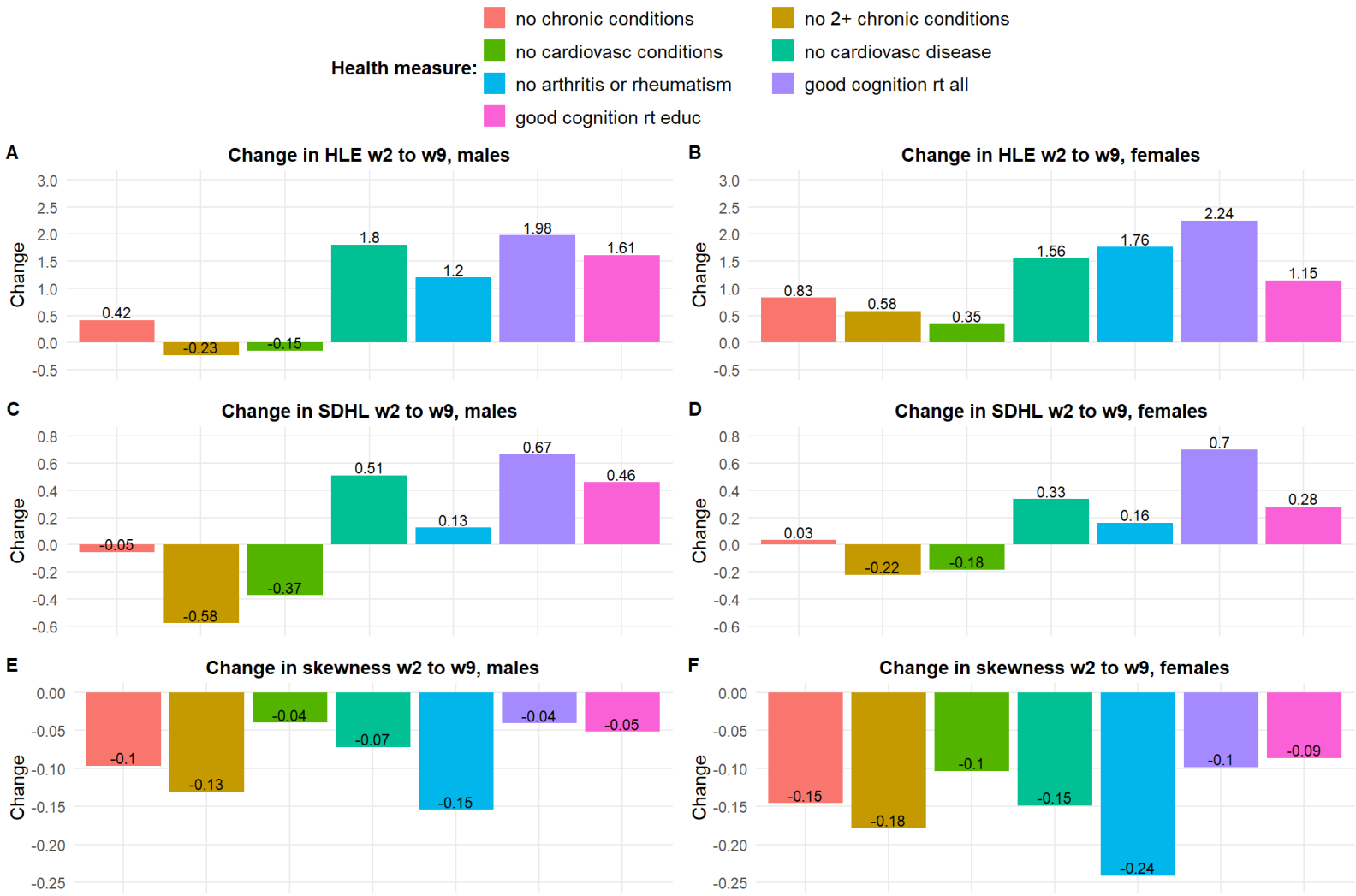

Figure S5. Change in HLE, SDHL and skewness from wave 2 to wave 9, by sex and health measure.

Figure notes. Abbreviations: cardiovasc: cardiovascular; HLE: healthy life expectancy (mean of the healthy longevity distribution); rt all: relative to all person-year observations; rt educ: relative to person-year observations of the same educational level; SDHL: SD of healthy longevity.
When the change value is positive, there has been an increase from wave 2 to 9. When the change value is negative, there has been a decrease from wave 2 to 9. Figure based on point estimates

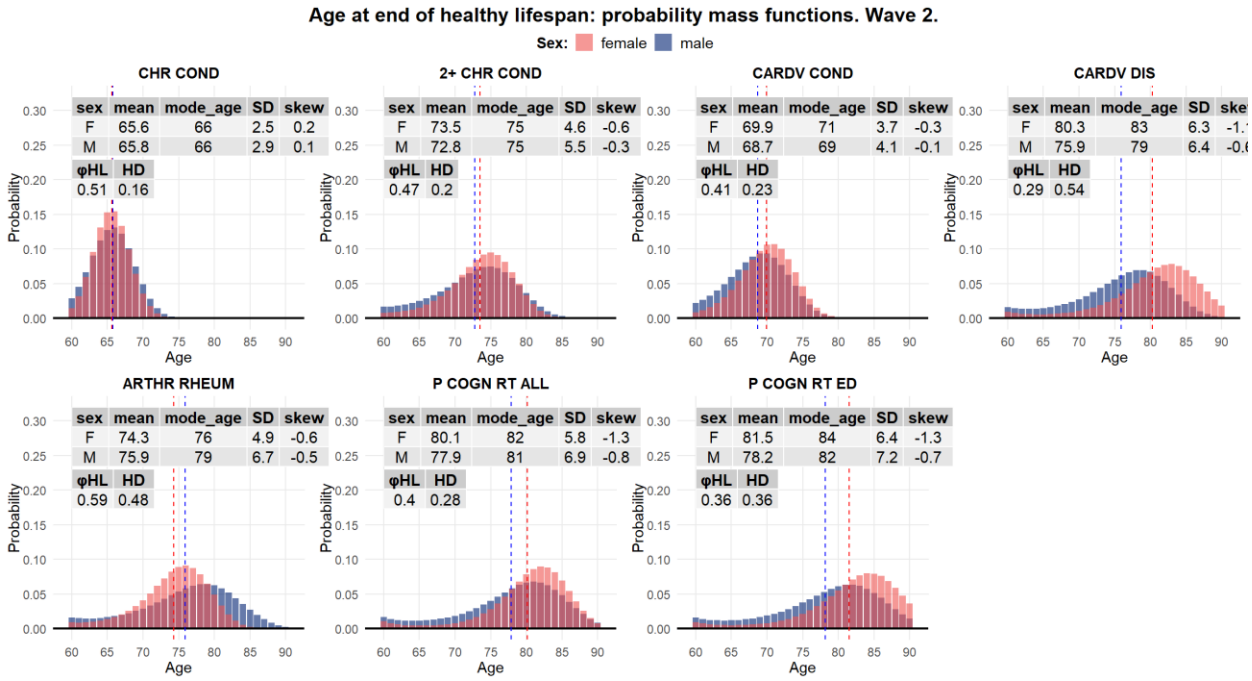

Figure S6. Probability mass functions, point estimates only, wave 2.

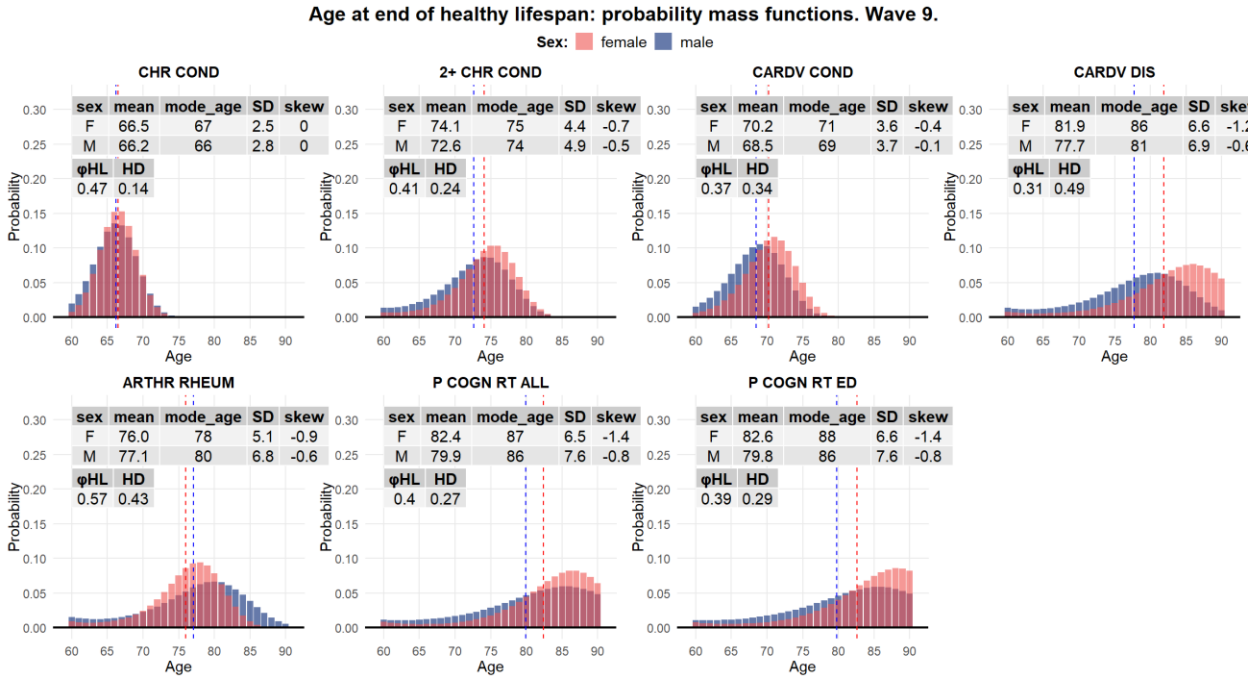

Figure S7. Probability mass functions, point estimates only, wave 9.

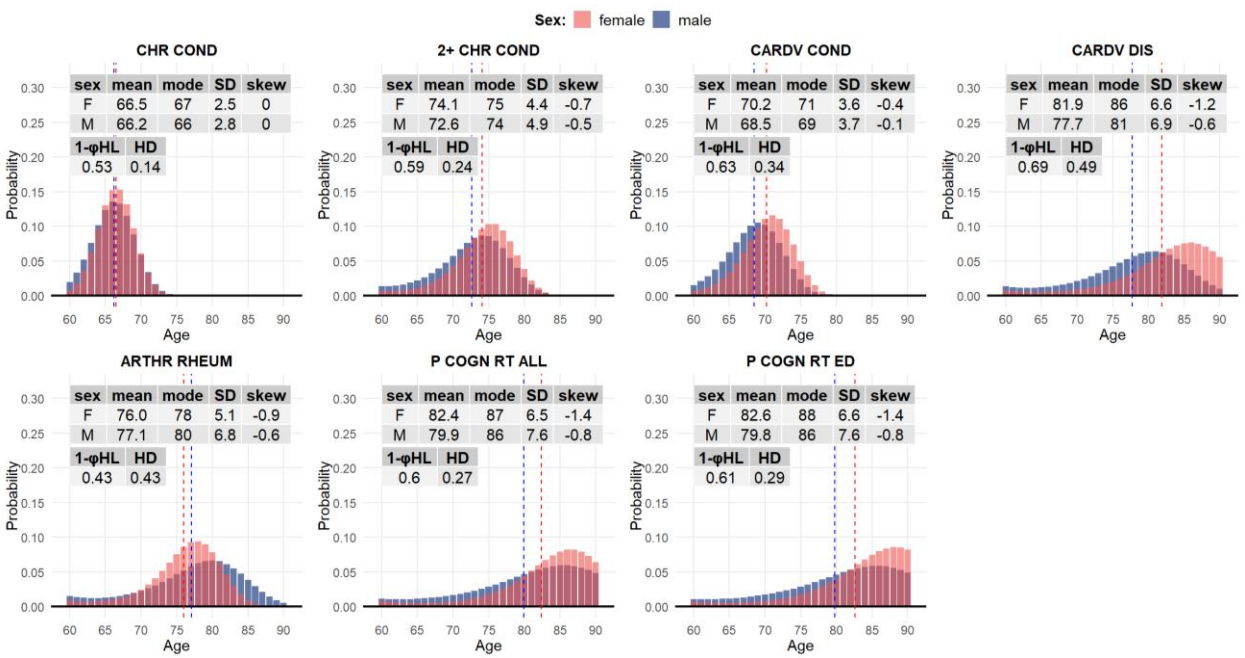

Figure S8. Age at end of healthy lifespan: overlapped probability mass functions for males and females in wave 9, plus the Hellinger distance and the complementary value of φHL (1 - φHL), i.e., the probability for females to have a longer healthy lifespan than males.

Figure notes. Abbreviations: arthr rheum: arthritis or rheumatism; cardv cond: cardiovascular conditions; cardv dis: cardiovascular disease; chr cond: chronic conditions; p cogn rt all: poor cognition relative to all person-year observations; p cogn rt ed: poor cognition relative to the person-year observations of the same educational level.

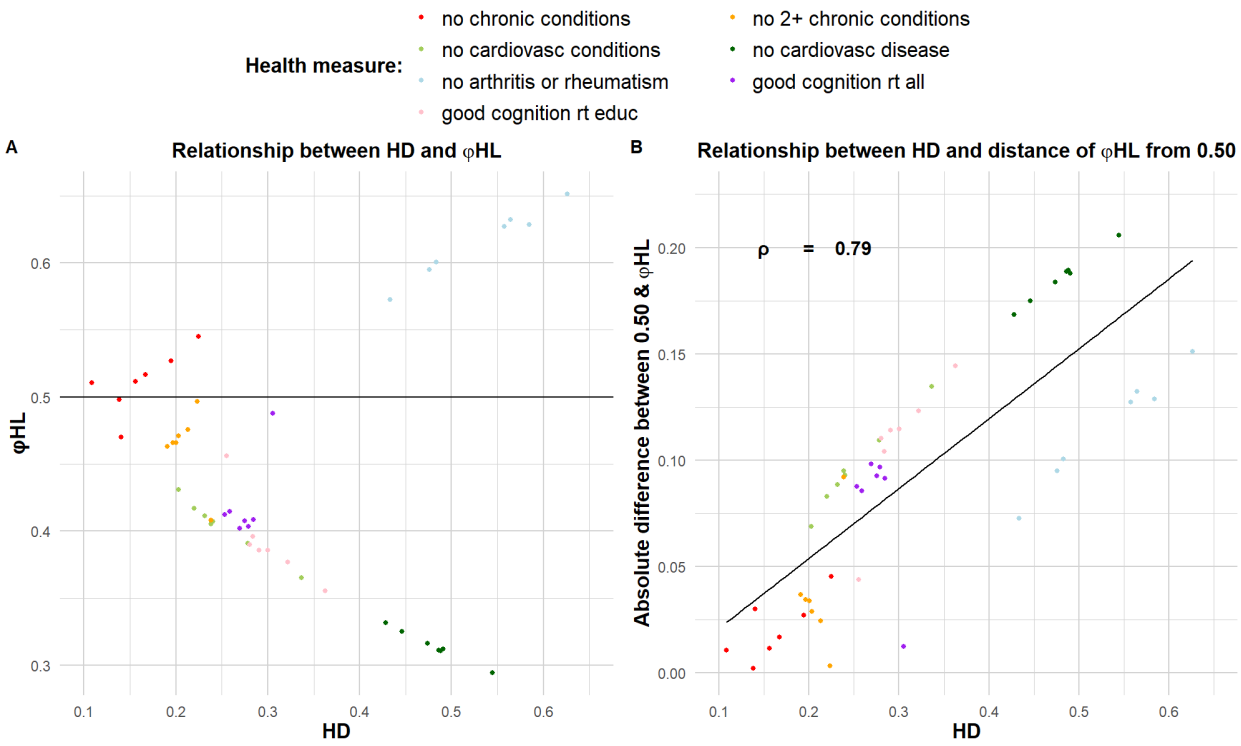

Figure S9. A. Relationship between Hellinger distance (HD) and $\varphi loss$. B. Correlation between HD and the distance of $\varphi loss$ from 0.50.

Figure notes. Abbreviations: cardiovasc: cardiovascular; HD: Hellinger distance; rt all: relative to all person-year observations; rt educ: relative to all person-year observations of the same educational level.
Figure and correlation coefficients based on point estimates of $\varphi loss$ and Hellinger distance.
In plot A, the grey line corresponds to $\varphi loss$ =0.5. In plot B, the grey line corresponds to the least squares model fit.
ρ refers to the Pearson's correlation coefficient.

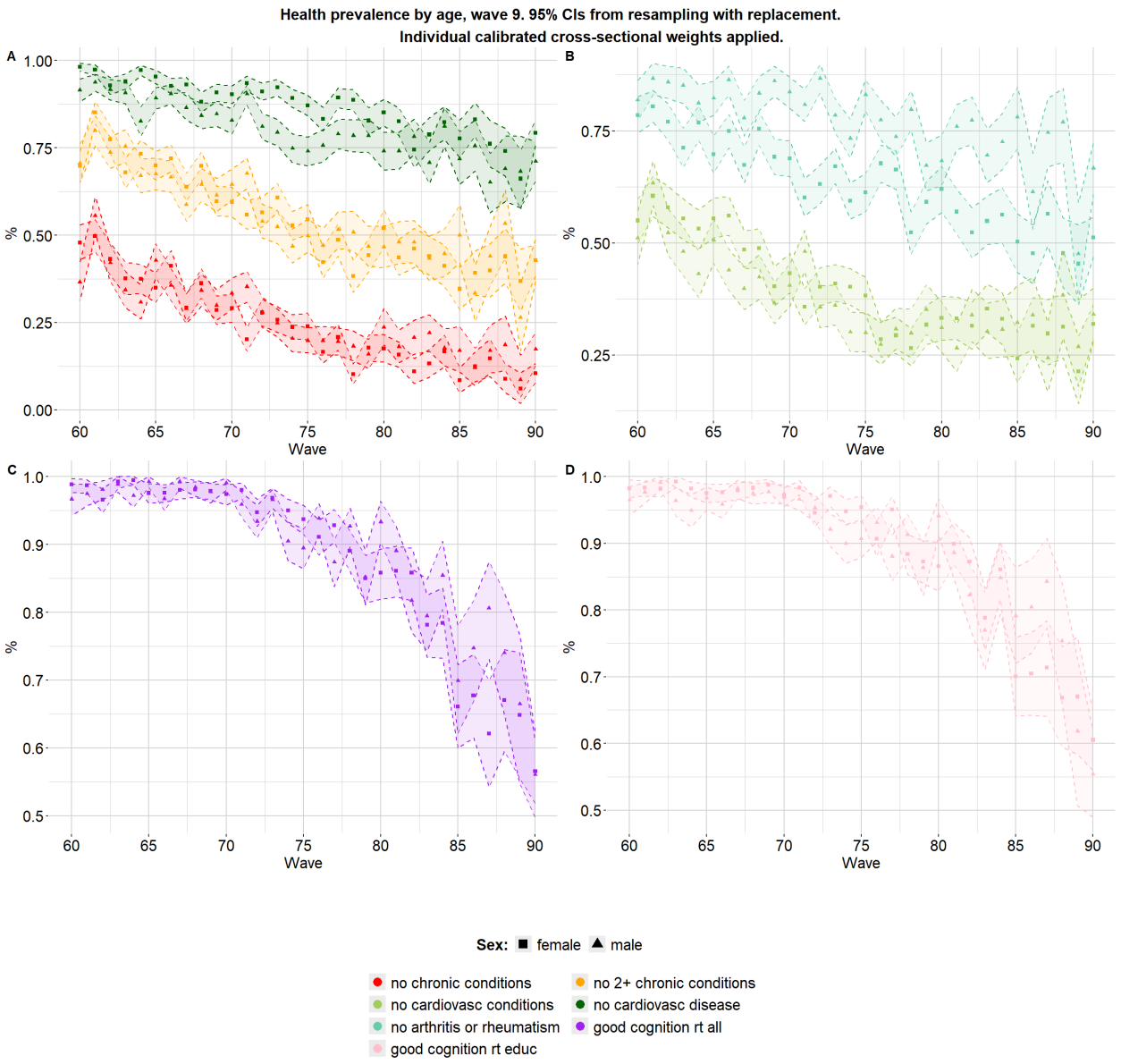

Figure S10. Prevalence with 95% confidence intervals based on weighted bootstrapping with resampling (5,000 iterations).

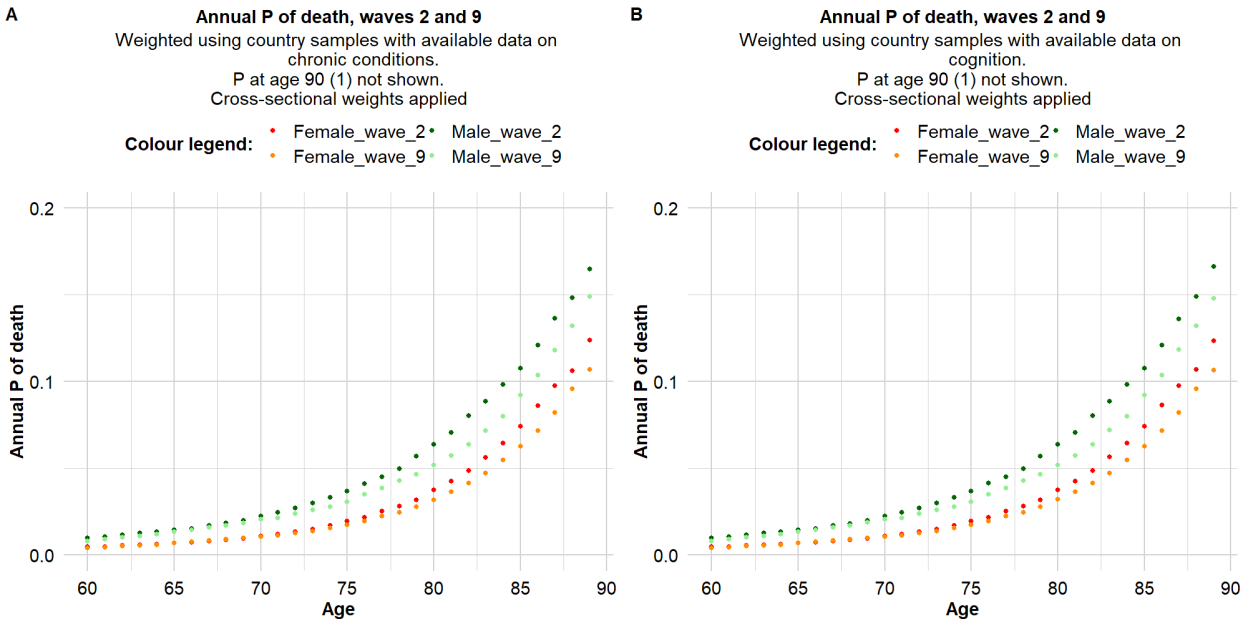

Figure S11. Annual P of death by age (weighted average across all countries)

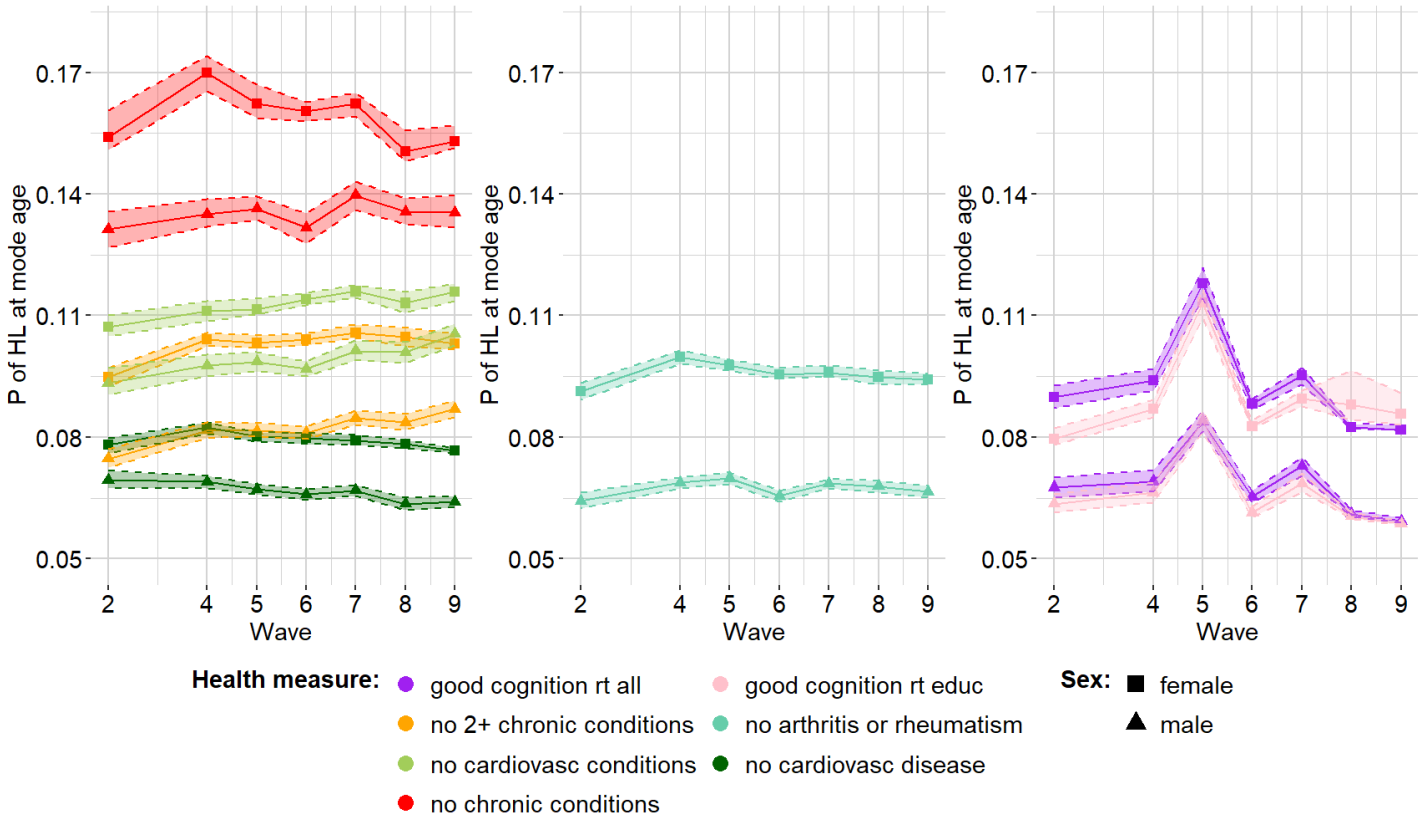

Figure S12. Probability of health loss at mode age

Figure notes. Symbols connected by continuous line: point estimates. Dashed lines: 95% confidence intervals.

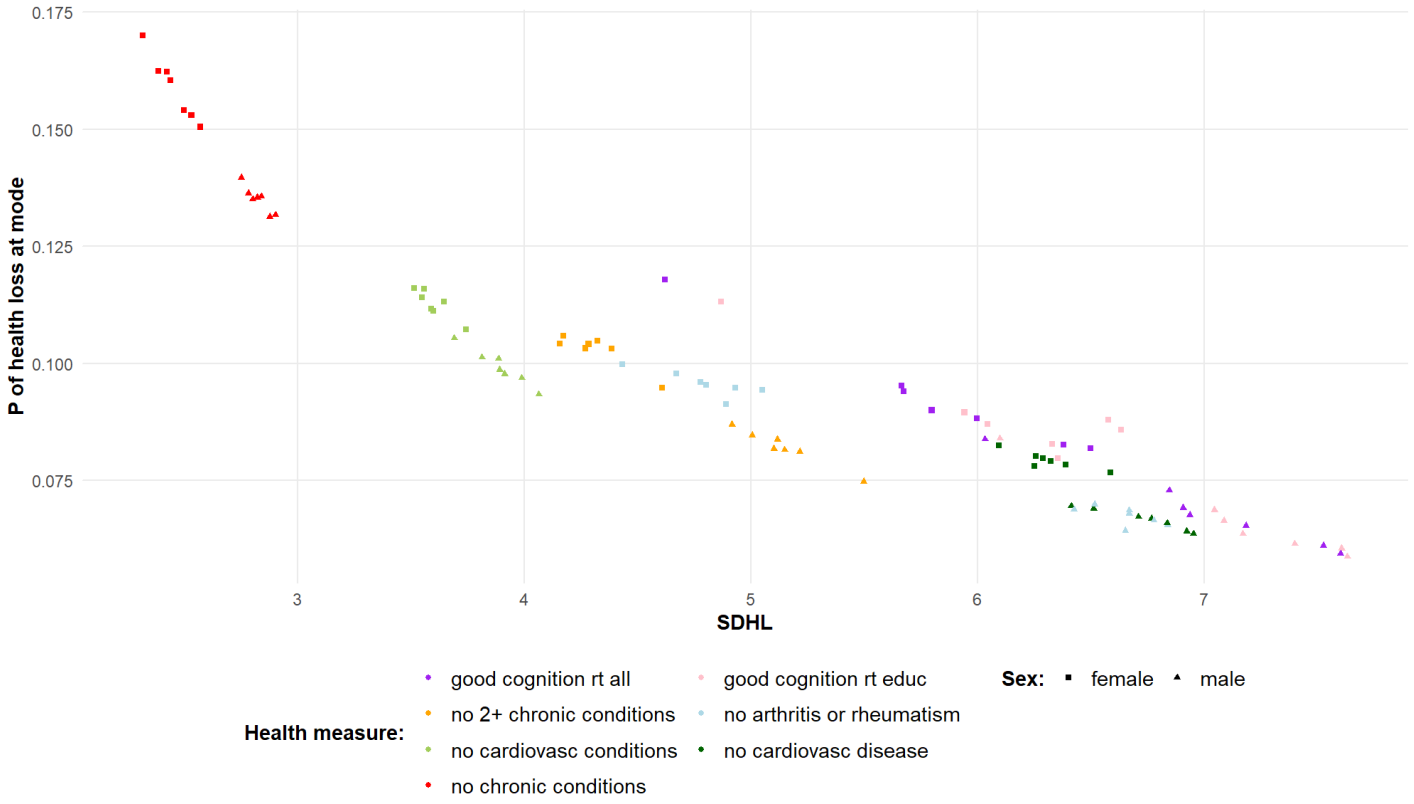

Figure S13. Relationship between the probability of health loss at mode and SDHL.

Figure notes. Abbreviations: P: probability; SDHL: standard deviation of healthy longevity.
For each colour and sex,the figure shows the values observed at different waves (7 waves for each sex-colour combination).
Based on point estimates only.

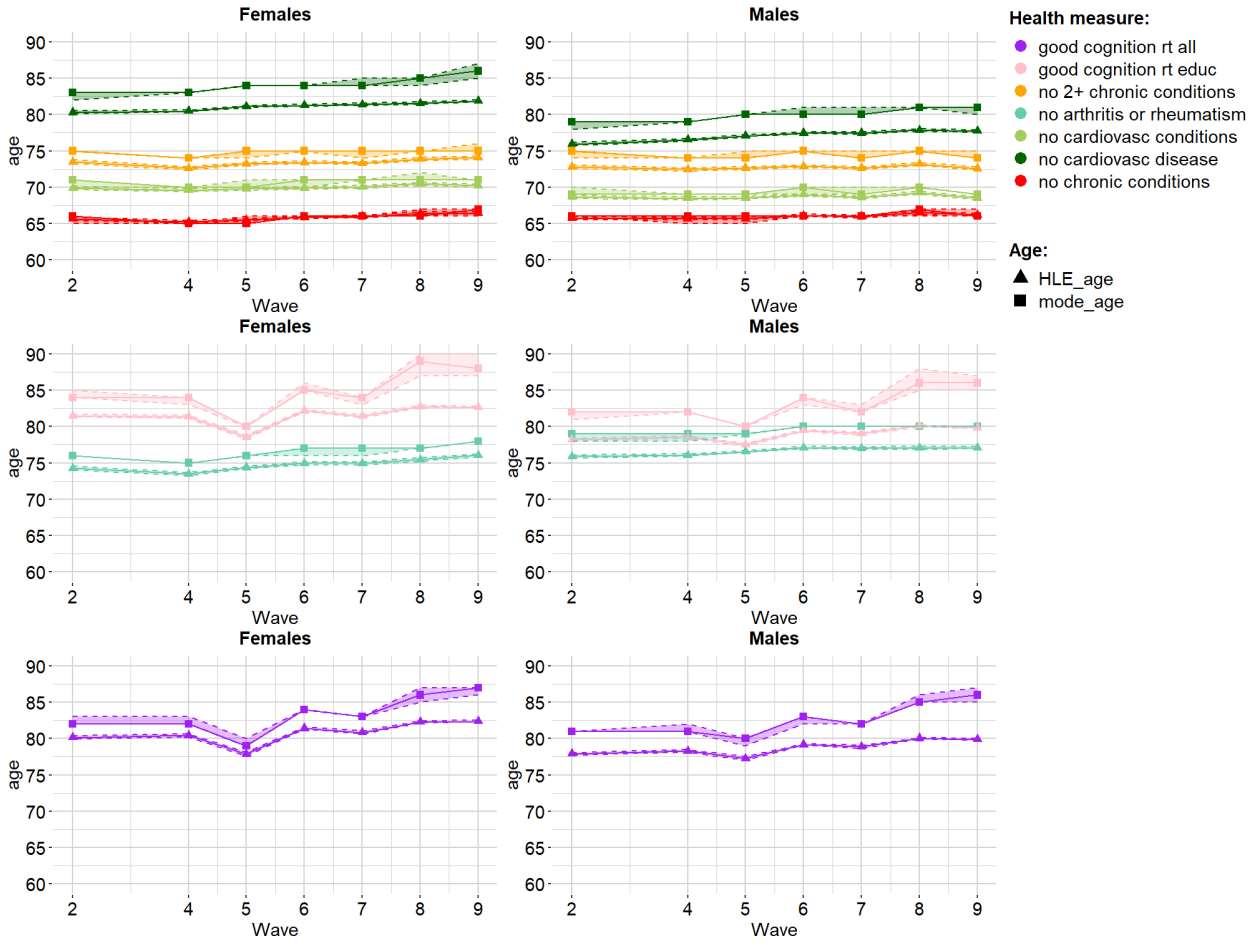

Figure S14. Mode and mean age of health loss.

Figure notes. Abbreviations: HLE: healthy life expectancy.
Symbols (squares or triangles) connected by continuous line: point estimates. Dashed lines: 95% confidence intervals. The confidence intervals for the mode age sometimes correspond the point estimate value: note that in our dataset, mode age was an integer age, while HLE age included decimals.

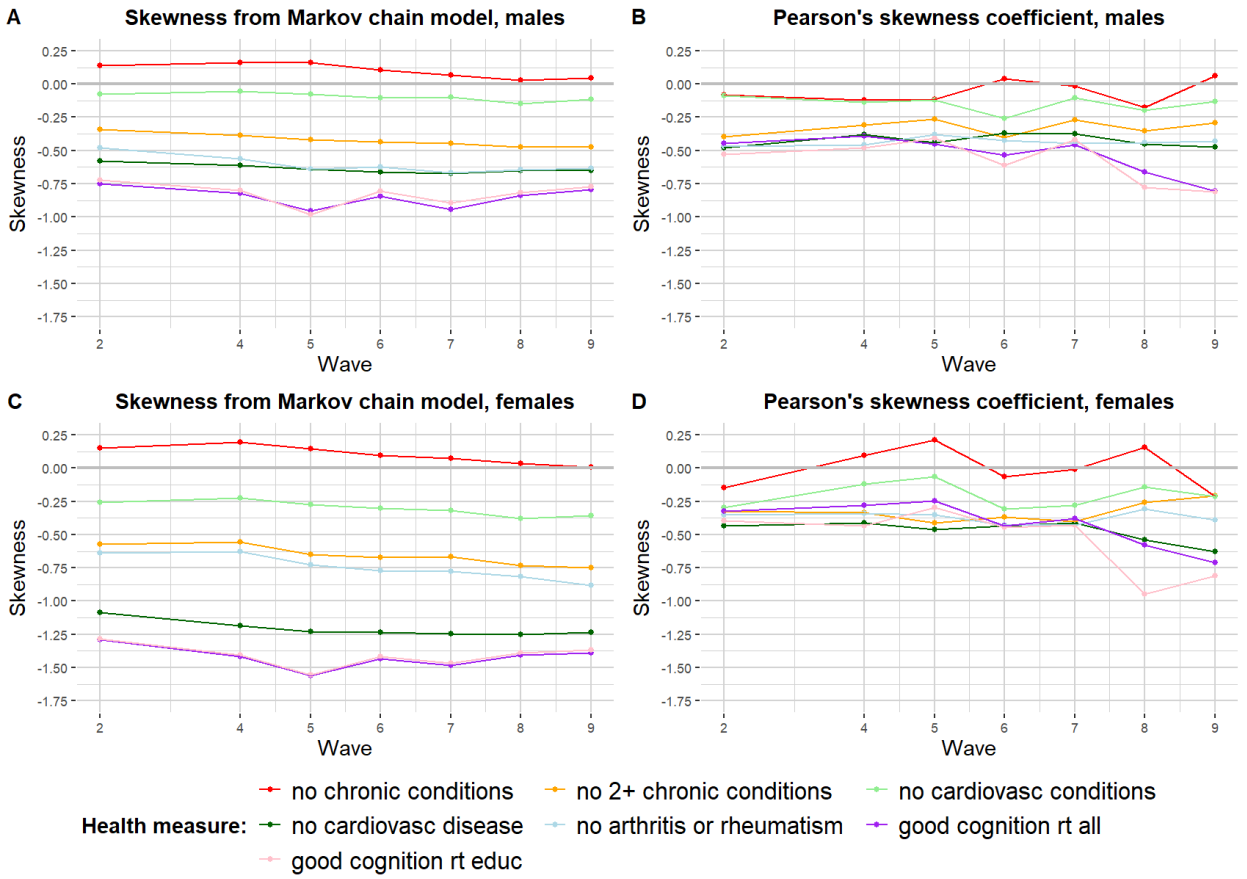

Figure S15. Skewness estimated from the Markov chain model vs. Pearson's skewness coefficient.

Table S9. Validation check on the ages when >=99% cumulative probabilities of health loss were reached based on the maximum entropy method, wave 9.

Health measure	Sex	Estimate from the maximum entropy method	Estimates from the HMD and SHARE data: point estimates (95% CI)		
		Age when the >=99% cumulative probability of health loss was reached: point estimate (95% CI)	Probability of being alive[1] at that age	Prevalence of health (i.e., of being free from the condition of interest) at that age	Probability of being alive and healthy based on the point estimates[2]
No chronic conditions	female	72 (72 to 72)	0.92	0.28 (0.24 to 0.32)	0.26
No chronic conditions	male	73 (72 to 73)	Age 72: 0.84 Age 73: 0.82	Age 72: 0.28 (0.24 to 0.32) Age 73: 0.25 (0.21 to 0.29)	0.21
No 2+ chronic conditions	female	82 (82 to 82)	0.73	0.46 (0.41 to 0.52)	0.34
No 2+ chronic conditions	male	82 (82 to 82)	0.57	0.48 (0.42 to 0.54)	0.27
No cardiovascular conditions	female	77 (77 to 78)	Age 77: 0.85 Age 78: 0.83	Age 77: 0.29 (0.25 to 0.33) Age 78: 0.27 (0.23 to 0.31)	0.25
No cardiovascular conditions	male	76 (76 to 77)	Age 76: 0.75 Age 77: 0.73	Age 76: 0.27 (0.23 to 0.32) Age 77: 0.31 (0.26 to 0.36)	0.20
No cardiovascular disease	female	90 (90 to 90)	Not relevant	Not relevant	Not relevant
No cardiovascular disease	male	89 (89 to 90)	Age 89: 0.28	Age 89: 0.68 (0.58 to 0.78)	0.19
No arthritis or rheumatism	female	85 (85 to 85)	0.63	0.50 (0.43 to 0.57)	0.32
No arthritis or rheumatism	male	89 (89 to 89)	0.28	0.48 (0.37 to 0.59)	0.13
Good cognition relative to all	female	90 (90 to 90)	Not relevant	Not relevant	Not relevant
Good cognition relative to all	male	90 (90 to 90)	Not relevant	Not relevant	Not relevant
Good cognition relative to educ	female	90 (90 to 90)	Not relevant	Not relevant	Not relevant
Good cognition relative to educ	male	90 (90 to 90)	Not relevant	Not relevant	Not relevant

[1] The probability of being alive at a specific age was calculated as follows: (1-p of death at age 60)* (1-p of death at age 61)*(1-p of death at age 62) * [and so on until the p of death at the age prior to the age of interest].

[2] Calculated by multiplying the p of being alive at age x by the p of health prevalence at the same age.

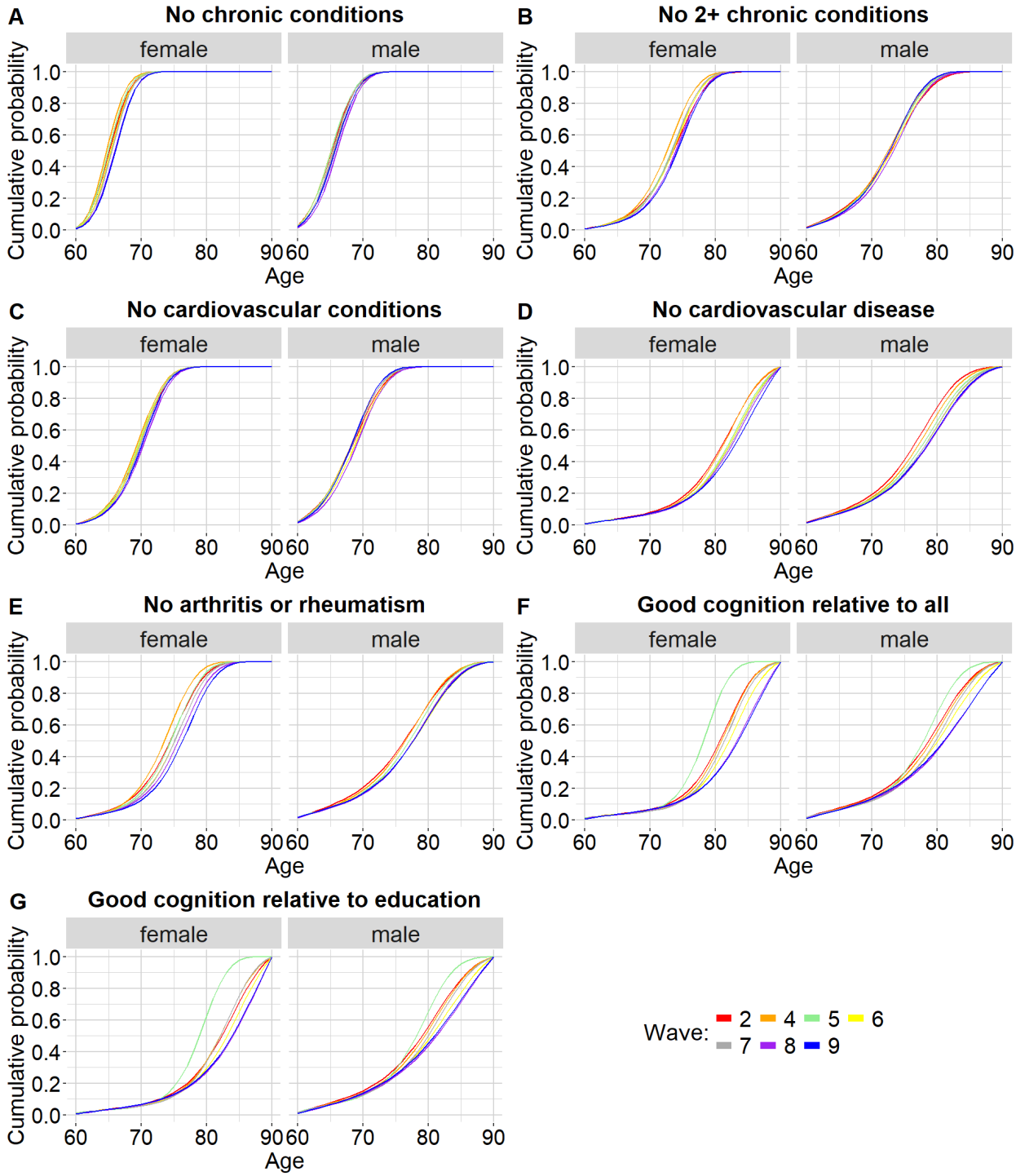

Figure S16. Cumulative distribution functions for different health measures.

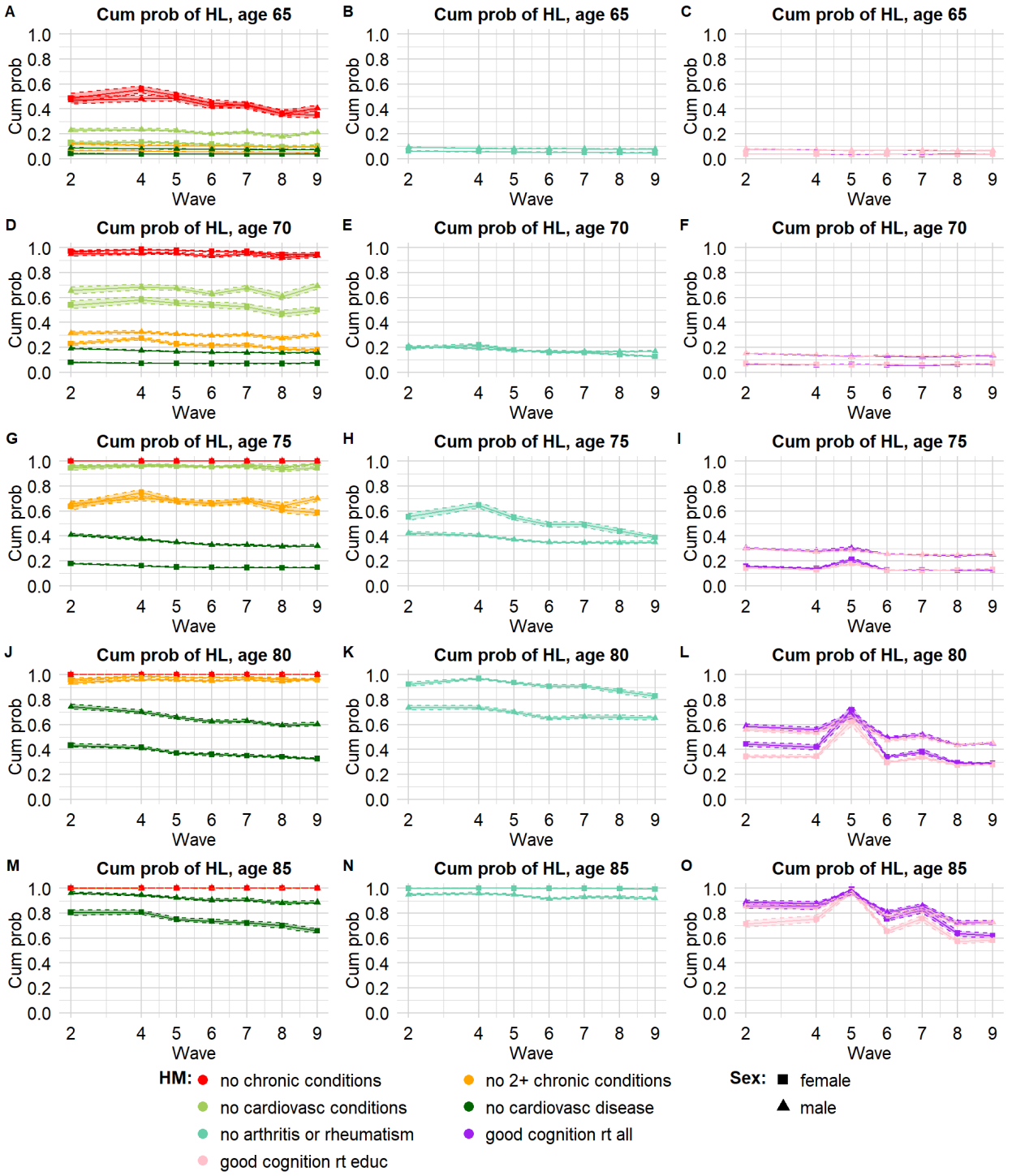

Figure S17. Cumulative probability of health loss at different ages.

Figure notes. Abbreviations: cardiovasc: cardiovascular; cum prob: cumulative probability; HL: health loss; HM: health measure; rt all: relative to all person-year observations; rt educ: relative to person-year observations of the same educational level.

The symbols connected by continuous lines are point estimates. The dashed lines are the 95% confidence intervals.

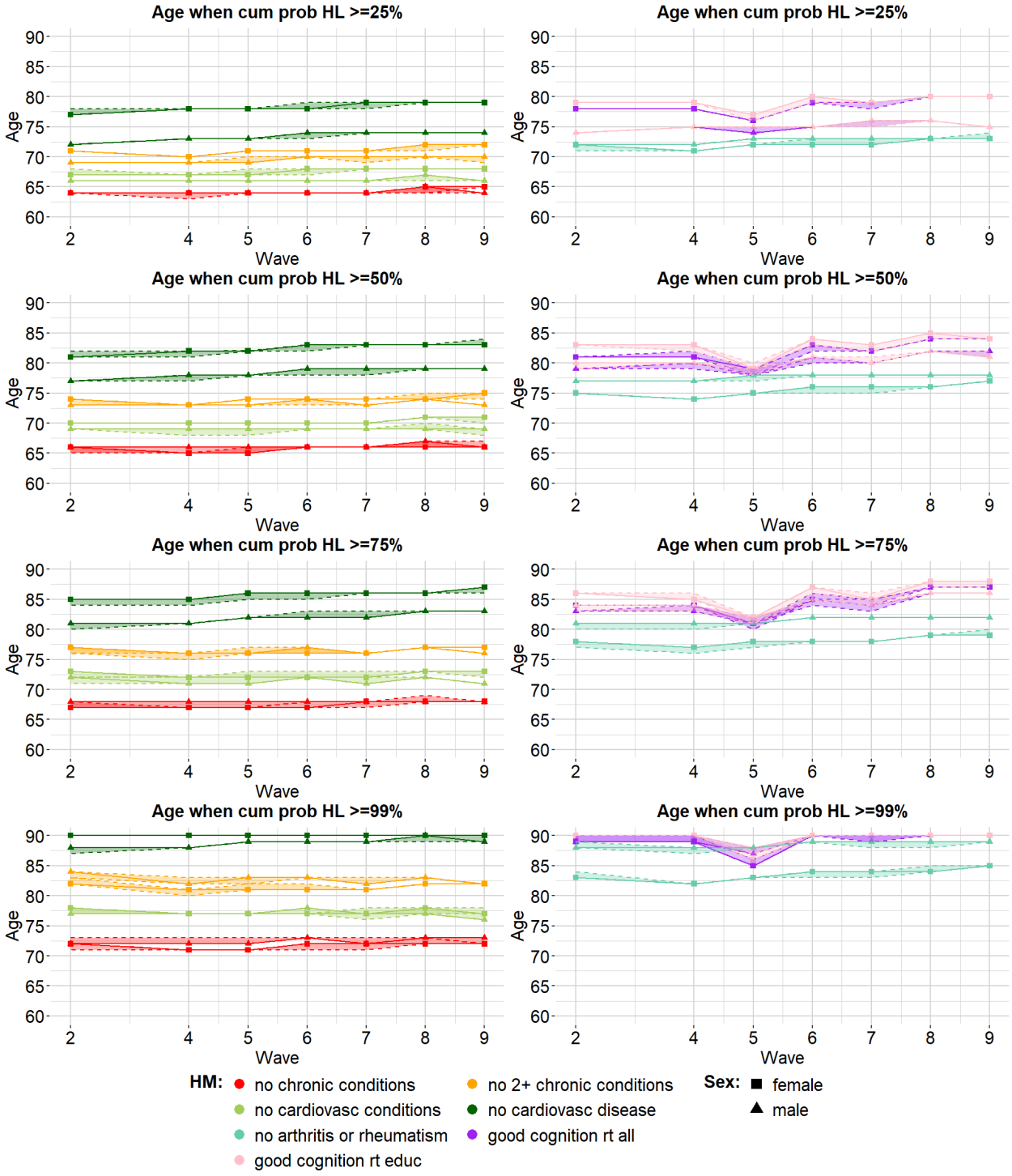

Figure S18. Ages when the cumulative probability of health loss is beyond a specific threshold.

Figure notes. Abbreviations: as per figure S18. Note that the figure shows the age when cumulative probability was *at least* the threshold value, so the cumulative probability could be higher than the threshold at that age.