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

Rami Cosulich and Virginia Zarulli

Department of Statistics, University of Padova, Italy

Abstract

This study explores the distributions of healthy longevity, focusing on differences between men and women and across various measures of health prevalence. Using a Markov chain model with rewards and data from SHARE and Human Mortality Database, we estimate distribution moments—mean, variance, and skewness—based on prevalence of limitations in activities of daily living (ADL), self-rated health (SRH), and cognitive health (COGN). We then derive the empirical density function of health loss using the maximum entropy method.

Preliminary results reveal that healthy life expectancy is increasing across all health measures, but self-rated health remains shorter in duration compared to ADL and COGN. Standard deviations show minor changes, with slight reductions at age 50, except for female cognitive health. Skewness is becoming more negative, particularly in cognitive health, where the most likely age of health loss has shifted to over 90 years. Women have narrower distributions than men and generally higher risks than loosing their health than men. This is particularly true when health is defined with self reported indicators.

To further compare these distributions, we will employ visual and statistical methods. Our findings shed light on gender-specific differences and the evolving shape of healthy longevity distributions across health metrics.

Introduction

The literature on healthy life expectancy has typically focused on average values within specific populations or subgroups. However, recent studies have begun to explore the under-researched subject of healthy lifespan variation between individuals within a population. Two studies (Permanyer et al., 2023; Zarulli and Caswell, 2024) addressed this topic globally, using Global Burden of Disease (GBD) data and the standard deviation as an indicator of variation. Despite these efforts, unanswered questions remain on the shape of distributions of healthy longevity and how these differ by gender or based on the health measure used to define healthy life expectancy. A characterization of the healthy longevity distributions with its indicators of shape and location, such as mean (healthy life expectancy), variation, and skewness, has not yet been investigated, despite its importance in understanding health inequalities. This study aims to compare distributions of healthy longevity across different measures of healthy life expectancy for men and women.

Data and Methods

We apply a Markov chain with rewards model developed by Caswell and Zarulli (2018) to estimate the mean, variance, and other moments of the healthy lifespan length for prevalence-based data. Variation in healthy longevity results from two stochastic processes: the risk of mortality and the chance of encountering disabling conditions, estimated by the prevalence of such conditions at each age. Caswell and Zarulli's model accounts for both.

The model requires age-specific mortality data and health prevalence across ages, for which we use three health measures from the SHARE survey (SHARE), that we coded as follows:

1. Life free from any limitations in activities of daily living (ADL)
2. Life in good, very good, or excellent health self-rated health (SRH)
3. Life in good cognitive health (COGN), based on performance in orientation and memory tasks, with poor cognitive health defined as scoring 1.5 standard deviations below average.

The prevalence rates of the measures cross the years, for men and women, broken down by two broad age groups, are reported in table 1. We combine SHARE's health data with mortality information from the Human Mortality Database (HMD), spanning the years 2004–2022 across 27 European countries and Israel. In this first analysis, we pool the data across all countries and survey waves, calculating average health prevalence and mortality rates, weighted by country sample size. The next phase will involve country-specific or regional analyses. Men and women are assessed separately.

Our analysis starts at age 50, grouping data with an open-ended interval for ages 90 and above. We use the matrix model to estimate not just the mean (healthy life expectancy) but also the variation (reported as standard deviation for ease of interpretability) and skewness over time. These metrics are calculated for each of the three health measures (ADL, SRH, COGN) for both genders. Finally, from the central moments of the healthy longevity distributions, we apply 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 are limited to age at death distributions in the context of mortality forecasting (Pascariu et al. 2019). For the first time, we use 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.

Table 2. Prevalence of the three health measures by year and sex

year	sex	ageg	no_ADL_limit	good_cogn	good_SRH
2004	female	50-69	0.94	0.98	0.73
2004	female	70+	0.73	0.79	0.51
2004	male	50-69	0.95	0.97	0.78
2004	male	70+	0.79	0.84	0.58
2005	female	50-69	0.93	0.97	0.75
2005	female	70+	0.68	0.8	0.52
2005	male	50-69	0.93	0.98	0.78
2005	male	70+	0.77	0.83	0.51
2006	female	50-69	0.93	0.97	0.69
2006	female	70+	0.7	0.81	0.43
2006	male	50-69	0.94	0.96	0.71
2006	male	70+	0.77	0.79	0.49
2007	female	50-69	0.94	0.98	0.69
2007	female	70+	0.72	0.78	0.44
2007	male	50-69	0.95	0.97	0.74
2007	male	70+	0.77	0.8	0.48
2008	female	50-69	NA	NA	0.61
2008	female	70+	NA	NA	0.4
2008	male	50-69	NA	NA	0.63
2008	male	70+	NA	NA	0.45
2009	female	50-69	0.94	0.97	0.67
2009	female	70+	0.68	0.82	0.4
2009	male	50-69	NA	NA	0.72
2009	male	70+	0.78	0.84	0.47
2010	female	50-69	0.9	0.98	0.57
2010	female	70+	0.6	0.81	0.25
2010	male	50-69	0.91	0.97	0.53
2010	male	70+	0.68	0.83	0.32
2011	female	50-69	0.93	0.97	0.66
2011	female	70+	0.73	0.81	0.41
2011	male	50-69	0.93	0.96	0.69
2011	male	70+	0.78	0.81	0.47
2012	female	50-69	0.86	NA	0.55
2012	female	70+	0.61	NA	0.2
2012	male	50-69	0.89	NA	0.45
2012	male	70+	0.72	NA	0.29
2013	female	50-69	0.94	0.98	0.7
2013	female	70+	0.73	0.78	0.42
2013	male	50-69	0.94	0.97	0.71
2013	male	70+	0.78	0.81	0.49
2015	female	50-69	0.93	0.98	0.69
2015	female	70+	0.73	0.82	0.4
2015	male	50-69	0.94	0.97	0.71
2015	male	70+	0.79	0.84	0.48
2017	female	50-69	0.94	0.99	0.68
2017	female	70+	0.76	0.87	0.4

2017	male	50-69	0.94	NA	0.68
2017	male	70+	0.79	0.87	0.47
2018	female	50-69	0.81	NA	0.39
2018	female	70+	0.5	NA	0.22
2018	male	50-69	0.88	NA	0.58
2018	male	70+	0.64	NA	0.28
2019	female	50-69	0.94	0.99	0.7
2019	female	70+	0.78	0.88	0.46
2019	male	50-69	0.93	0.98	0.72
2019	male	70+	0.8	0.89	0.5
2020	female	50-69	0.94	0.99	0.73
2020	female	70+	0.76	0.85	0.5
2020	male	50-69	0.94	0.98	0.76
2020	male	70+	0.81	0.85	0.55
2021	female	50-69	0.93	0.99	0.73
2021	female	70+	0.79	0.87	0.59
2021	male	50-69	0.95	0.98	0.75
2021	male	70+	0.83	0.87	0.6
2022	female	50-69	0.93	0.98	0.72
2022	female	70+	0.79	0.89	0.55
2022	male	50-69	0.92	0.97	0.73
2022	male	70+	0.83	0.89	0.56

Preliminary results

Figure 1 illustrates the changes in the distribution of healthy years of life at ages 50 and 70 for men and women, spanning from 2004 to 2022. These changes are presented across three health measures: life free from limitations in daily activities (ADL), life in good cognitive health, and life in good, very good, or excellent self-rated health (SRH).

Over this period, both men and women experienced a general increase in healthy life expectancy at ages 50 and 70, regardless of the health definition used. However, life expectancy in good SRH consistently remains shorter than life expectancy free from ADL limitations or in good cognitive condition. Additionally, longevity based on self-rated health shows a smaller standard deviation and skewness compared to the other two health measures.

Between 2004 and 2022, the standard deviation of healthy longevity decreased slightly at age 50 across all health measures, with the exception of cognitive health in women. At age 70, however, the standard deviation increased for ADL and cognitive health in both sexes, while it decreased for SRH. The distributions of healthy longevity based on these three health

measures, at both ages, show a trend towards increasing negative skewness. Notably, in 2004, the distribution of SRH for men at age 70 was the only one with a positive skew.

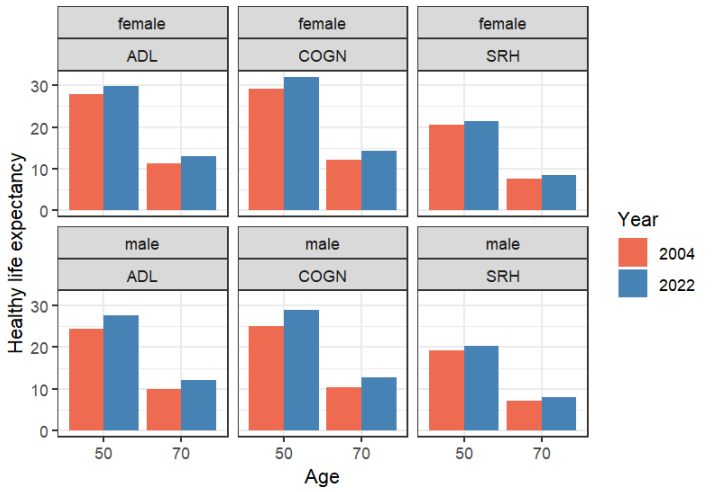
Figure 2 explores a critical question: how likely is it for those healthy at age 50 to lose their health as they age? The figure shows the changes in the empirical density function of health loss at age 50 from 2004 to 2022, across all three health measures. Overall, there has been a noticeable shift in the likelihood of health loss towards older ages, particularly for ADL and cognitive health, with the exception of SRH.

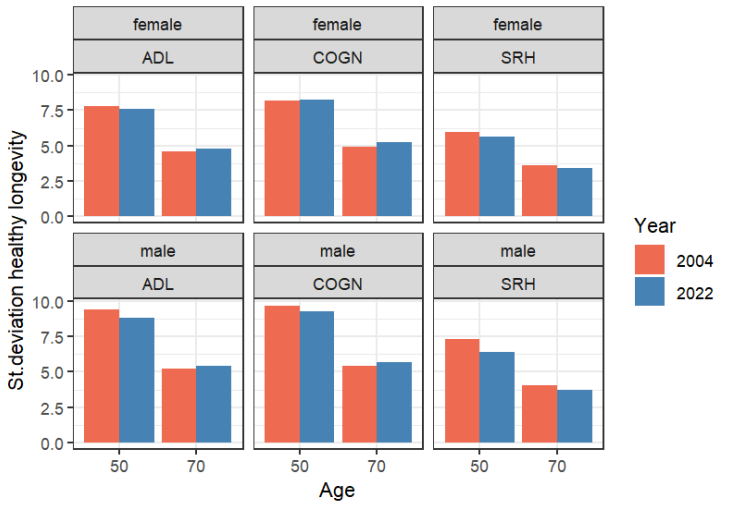
An increase in negative skewness is evident, with the left tail of the distribution extending. This is particularly striking for cognitive health: by 2022, individuals aged 50 are most likely to lose good cognitive health after age 90. While the open-ended age group of 90+ limits further detail, the key takeaway is that for those healthy at 50, cognitive health loss is now more likely to occur after age 90, while physical health loss (as measured by ADL) or loss of perceived good health tends to happen between ages 75 and 85.

Several important points emerge from this analysis:

1. ADL-related distributions are more spread out compared to SRH.
2. ADL and cognitive health distributions have shifted to older ages, but SRH has not followed this pattern.
3. Women have narrower distributions than men
4. The most likely age to lose cognitive health has shifted beyond age 90.

Over time, the distributions of healthy longevity are showing increasing negative skewness, which means that, while most people are living longer healthy lives, some individuals are still losing their health at relatively younger ages. In a negatively skewed distribution, the bulk of the values (in this case, the ages at which health loss occurs) shifts toward older ages, but the left tail of the distribution—representing younger ages—becomes longer and flatter.This pattern indicates that, although overall improvements in healthy life expectancy are pushing the typical age of health loss to later in life, a small but significant group of individuals still experiences health decline at younger ages. As a result, the process of health loss is becoming more stretched out—with a growing gap between those who maintain good health into old age and those who lose their health earlier.





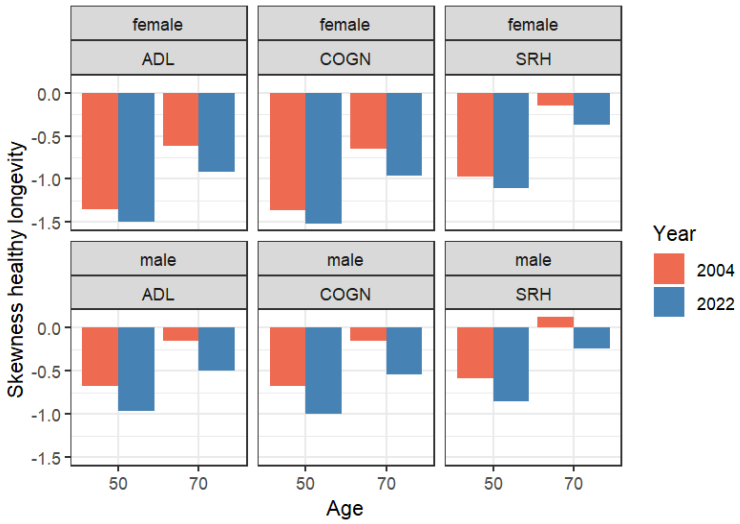


Figure 1 - Change in male and female healthy life expectancy (HLE), standard deviation (STD) and skewness (SKW) of healthy longevity at age 50 and age 70, from 2004 to 2022, for three measures of health.

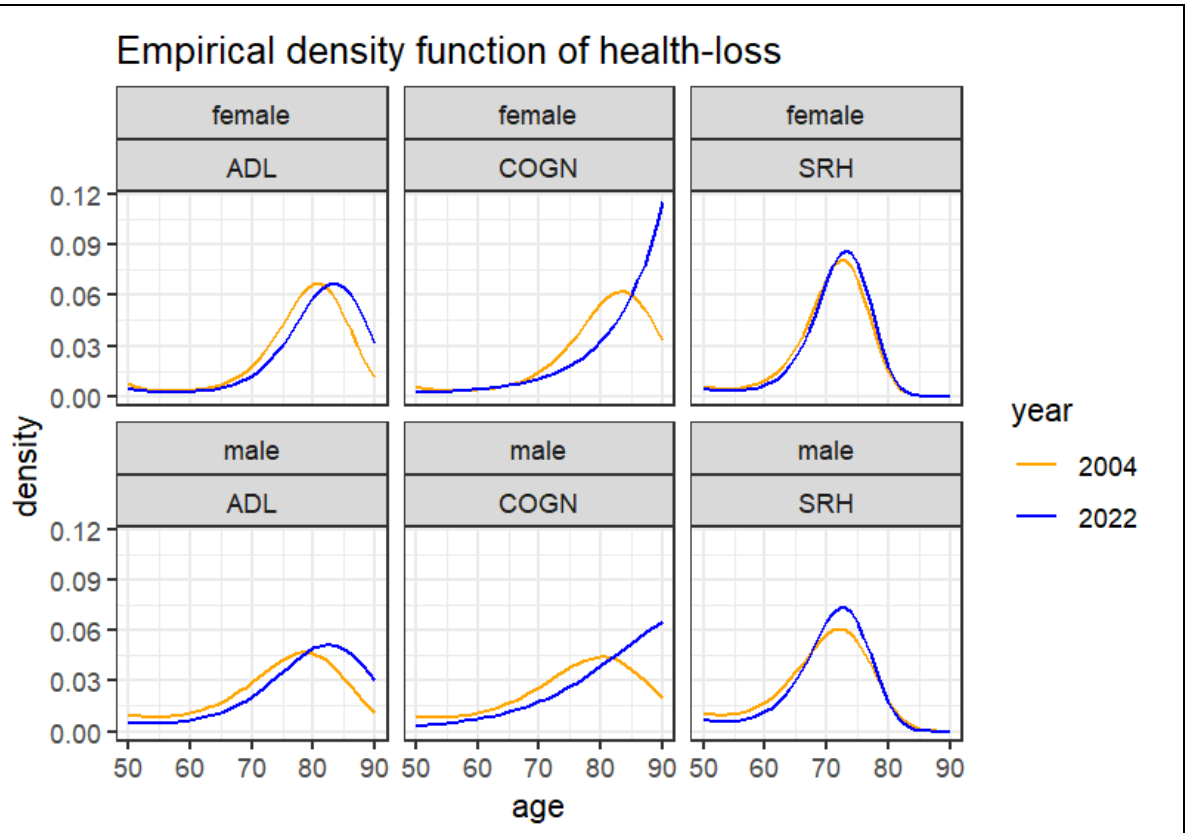


Figure 2 – Density function of health-loss process from age 50, according to three different measures of health expectancy, for men and women, from 2004 to 2022 (SHARE data, all countries pooled together).

Future steps and expected findings

We will compare the distributions of healthy longevity using both visual and statistical methods to uncover key differences in their shapes. Visually, we will overlay density estimates and compare cumulative distribution functions (CDFs) to illustrate how one distribution accumulates values relative to another. To quantify asymmetry, we will calculate summary statistics like mean, variance, and skewness.

To assess overlap and similarity between the distributions, we will use the Hellinger distance, which highlights differences in shape, location, and scale. Additionally, we will employ statistical tests, such as the Kolmogorov-Smirnov test, to determine the significance of the differences between distributions based on different health measures and between men and women.

Our analysis will also extend beyond age 50, exploring healthy life expectancy, standard deviation, and skewness at older ages. We will compare results across health measures and genders using this range of techniques, while adapting the approach as needed. Country-specific and regional comparisons will also be part of this detailed investigation.

This comprehensive analysis will offer valuable insights into gender-specific variations in healthy lifespan and provide a deeper understanding of how different health measures influence longevity patterns.

References

Permanyer, Iñaki, Francisco Villavicencio, and Sergi Trias-Llimós. "Healthy lifespan inequality: morbidity compression from a global perspective." European Journal of Epidemiology 38.5 (2023): 511-521.

Zarulli, Virginia, and Hal Caswell. "Longer healthy life, but for how many? A stochastic analysis of healthy lifespan inequality." Annals of Operations Research (2024): 1-18.

Caswell, Hal, and Virginia Zarulli. "Matrix methods in health demography: a new approach to the stochastic analysis of healthy longevity and DALYs." Population health metrics 16 (2018): 1-14.

Survey of Health, Ageing and Retirement in Europe (SHARE), https://share-eric.eu/data/

HMD. Human Mortality Database. Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France). Available at www.mortality.org .

Pascariu, Marius D., Adam Lenart, and Vladimir Canudas-Romo. "The maximum entropy mortality model: Forecasting mortality using statistical moments." Scandinavian Actuarial Journal 2019.8 (2019): 661-685.