Multistate distributions, lifespan inequality and morbidity compression: Advancing the debate on aging and health

Iñaki Permanyer[1,2], Tim Riffe[3,4,5]

Abstract

Background: Longevity increases have been accompanied by a compression in the distribution of ages at which individuals die – thus reducing lifespan inequality and making age-at-death increasingly predictable. Less clear, though, is whether the retreat of mortality towards older ages has been accompanied by commensurate morbidity declines. Assessments of countries' longevity performance should take into consideration the fact that a non-negligible fraction of individuals' lives is spent in less-than-good health.

Aim: While countless studies have investigated the influence of morbidity on *average* longevity, virtually nothing is known about the effect of morbidity on lifespan inequality. In this paper, we study (i) what fraction of lifespan inequality is attributable to the years individuals spend in different health states, and (ii) what implications does this have for the formulation *and* testing of the compression vs expansion of morbidity debate.

Data: We use self-reported health measures from the US Health and Retirement Study by sex and in 5-year periods.

Methods: We model health and mortality transition probabilities and apply recently proposed multistate life table methods to estimate the number of years individuals have spent in different health states at time at death. Standard decomposition techniques are applied to break down lifespan inequality in components with clear demographic interpretations.

Results: The results from this research shed new light to the longstanding compression vs expansion of morbidity debate initiated in the late 1970s – early 80s. Our novel approach offers complementary insights that could not be gained through traditional approaches relying on average morbidity and mortality measures.

Keywords: Mortality, Morbidity, Lifespan Inequality, Compression of Morbidity, Health Inequalities, Ageing and Health, Multistate Distribution

[1] Centre d'Estudis Demogràfics (CED-CERCA), Universitat Autònoma de Barcelona.
[2] ICREA, Passeig de Lluís Companys 23, 08010, Barcelona.
[3] Universidad del País Vasco (UPV/EHU).
[4] Ikerbasque (Basque Foundation for Science).
[5] Max Planck Institute for Demographic Research

1. Introduction

During the last decades, survival prospects for humanity have improved dramatically around the globe. Despite occasional setbacks, the levels of life expectancy have increased in a sustained fashion in most world countries since the turn of the 20th century (Oeppen and Vaupel 2002, Riley 2005). Parallel to these unprecedented changes in average longevity, the distribution of ages at which individuals die has become increasingly compressed over time (Smits and Monden 2009, Vaupel et al 2011). Length of life inequality (from now on 'lifespan inequality') is an important marker of heterogeneity in populations' health (van Raalte et al 2018) that has attracted the attention of demographers and other social scientists, so there is growing consensus that lifespan inequality should be regularly monitored alongside other population health indicators, like life expectancy (van Raalte et al 2018).

The study of longevity dynamics (i.e., investigating not only how efficient populations are in sustaining and extending life but also how years of life are distributed across individuals) has implicitly assumed that "more" is necessarily "better". However, years of life can be spent in "good" or in "less-than-good" health. While the normative desirability of the former is almost universal, it is not clear how desirable the latter is.[6] Given the trade-offs between quantity and quality, assessments of populations' longevity performance should be revisited taking into consideration the fact that non-negligible fractions of individuals' lives can be spent in varying degrees of less-than-good health. In this context, it is fundamental to explore how years lived in good and in less-than-good health have contributed both to the composition of individuals' length of life *and* to the differences in longevity across individuals (i.e., to lifespan inequality).

While countless studies have investigated the influence of morbidity on *average* longevity (i.e., a lot is known about what fraction of life expectancy (LE) is spent in 'good' or 'less-than-good' health through 'Health-adjusted life expectancy' (HALE) or other conceptually-related indicators (Jagger et al 2020)), virtually nothing is known about (1) how healthy and unhealthy years are distributed across *individuals*' lifespans (e.g., among individuals dying at age x, how many years have they accumulated in good and in less-than-good health?), and (2) the effect of morbidity on lifespan inequality (i.e., what is the contribution of years spent in less-than-good health on lifespan inequality?). The existence of these lacunae can be largely attributable to the lack of appropriate methods to answer these important questions – a limitation we want to address in this paper.

The exploration of these issues has direct bearing with the longstanding 'compression vs expansion of morbidity' debate, which investigates whether morbidity is retreating to older ages at a faster or slower pace than mortality does (Fries 1980, Gruenberg 1977,

[6] Having to choose between 'a long yet unhealthy life' and 'a shorter but fully healthy life', it is not entirely obvious that the former would be universally chosen in favor of the latter; see, for instance, Gerstorf et al. (2008), Lawton et al. (1999).

Manton 1982). Traditionally, the standard competing hypotheses of this debate have been tested through the comparison of population health indicators like LE and HALE (details shown below). However, the lack of suitable data and methods has prevented going beyond average-based indicators and investigate whether, *across all possible ages at death* x, the contribution of healthy and unhealthy years to individuals' lifespans has been shifting over time. Another aim of the paper is to revisit the 'compression vs expansion of morbidity' debate taking advantage of the newly proposed analytical tools.

To attain our goals, we plan to make use of the recent multistate modeling techniques proposed in Riffe et al (2023). Such methods allow deriving individual-level multistate distributions estimating the number of years individuals have accumulated in good and in less-than-good health at time at death. Importantly, this allows exploring the association between the "healthy" and "less-than-healthy" years distributions (i.e., whether living more years in good health is associated with more or less years spent in less-than-good health). Empirical analyses are based on data from the US-based Health and Retirement Study (HRS), showing results from 2000 to 2015 for women and men separately.

The ideas and findings presented in this paper can contribute to improving our understanding of contemporary health dynamics. In addition, they can guide the elaboration of policies aiming at reducing health disparities while promoting the sustainability of welfare and health care systems.

2. Background

Lifespan inequality

In recent times, the study of lifespan inequality has attracted a great deal of attention among demographers and other social scientists. Overall, empirical findings suggest that, with increasing life expectancy, age-at-death distributions tend to compress, thus reducing lifespan inequality (Smits and Monden 2009, Vaupel et al 2011, Colchero et al 2016, Permanyer and Scholl 2019, Aburto et al 2020). There are, however, some notable exceptions to the generally strong and negative association between life expectancy and lifespan inequality. Exceptions to that pattern can be found among socially disadvantaged groups (e.g., low SES groups; see van Raalte et al (2014), Sasson (2016), Permanyer et al (2018), van Raalte et al (2018), Seaman et al (2019)), or populations experiencing great shocks, like famines, wars and episodes of socio-economic or political disruption (e.g., countries of the Eastern bloc after the collapse of the Soviet Union); see Aburto and van Raalte (2018), Vigezzi et al (2022)).

There are different measures of lifespan inequality, but in general they tend to be highly correlated (van Raalte and Caswell 2013). Some of them are amenable to specific decompositions that might be useful depending on the research question we are dealing with. In the methods section, we introduce some of the decompositions of lifespan inequality that will be applied in the empirical section of this paper.

Healthy lifespan inequality

The measurement of inequality is implicitly based on the assumption that the outcome variable whose variability we are analyzing is normatively desirable. Economists, for instance, assume that the utility that individuals derive from income can be modeled with a *non-decreasing* concave function; the so-called 'utility function' (Mas-Colell et al 1995). Thus, while such utility function is more sensitive to income gains at the bottom of the distribution, it never decreases with increasingly higher income levels. Yet, there are good reasons to question the validity of this assumption when measuring inequality in individuals' lifespans. A non-negligible part of those lifespans might be spent in very bad health conditions (e.g., experiencing severe neurological disorders, like Alzheimer disease, malignant tumors, strokes, and so on; specially towards the end of individuals' lives), and in some circumstances it might be debatable whether "more" is necessarily "better". Including years spent in good health together with those spent in bad health might muddy the waters when interpreting both longevity performance *and* the extent of variability in individuals' lifespans.

Given the above mentioned reasons, several studies have proposed to measure the extent of variability among healthy lifespans only (a quantity that will be referred to as 'healthy lifespan inequality' (see Permanyer et al 2022)). Focusing on the length of life that is spent in good health, one works with a normatively desirable outcome and avoids the problems associated with having to weigh 'quantity' vs 'quality'. Examples of this approach can be found in Caswell & Zarulli (2018) (who apply the approach to a group of 9 European countries), Seaman et al (2020) (who measure variability in healthy lifespans in Denmark), Permanyer et al (2022) (with an application to education groups in contemporary Spain) and Permanyer et al (2023) (who explore trends in healthy lifespan inequality in all regions around the world).

Yet, current studies following this approach have their own limitations as well. Most of them are based on cross-sectional prevalence data, which is more readily available than incidence data. When this happens, analysts are forced to make stringent assumptions about the reversibility of health states (e.g., no recoveries from unhealthy states are admissible), an issue that can be potentially problematic depending on the health outcome indicator one is working with. In addition, these studies focus on the variation of healthy lifespans, but fail to take into consideration the variation of its "unhealthy counterparts" – which can arguably be considered the other side of the coin. Importantly, currently existing approaches do not allow investigating the *joint* distribution of individuals' healthy *and* unhealthy lifespans. The multi-state based approach presented below allows overcoming such limitations, thus offering an analytically powerful method to improve our understanding about contemporary health dynamics.

A conceptually related but substantively different approach has been recently proposed by Permanyer and Bramajo (2023). In that paper, the authors investigate the role that the country-specific healthy and unhealthy components of life expectancy (HALE and UHLE) have had on recent trends in life expectancy inequality between countries (also

known as ‘International Health Inequality’, or IHI) – a key metric to investigate whether countries’ health status is converging or diverging worldwide. Empirically, they find that, while IHI values have been mostly attributable to the variability in the corresponding HALE levels between 1990 and 2019, the unhealthy component of life expectancy is playing an increasingly prominent role (Permanyer and Bramajo 2023). While related, that paper is based on country-level indicators only, and thus ignores intra-country variation. Significantly, previous studies comparing variation in lifespans across and within countries found that the latter is much higher than the former (Smits and Monden 2009, Edwards 2011, Permanyer and Scholl 2019), so the approach proposed in this paper has the potential to unearth a great deal of health variation that cannot be identified with currently existing methods.

Compression vs expansion of morbidity debate

The unprecedented success in delaying the ages at which individuals die have led many scientists to speculate whether improved survival prospects would be accompanied by concomitant morbidity declines. In this regard, three main hypotheses have been proposed. The so-called ‘compression of morbidity’ hypotheses suggests that, with increases in longevity, the onset of morbidity is gradually compressed towards the last years of life, thus reducing the number of years individuals are expected to live in less-than-good health (Fries 1980). At the opposite extreme, the ‘expansion of morbidity’ hypothesis suggests that, in post-epidemiological transition countries, further gains in longevity would be achieved through the survival of people living in morbid states – thus resulting in more disease in the population (Gruenberg 1977). Between these two extremes, the ‘dynamic equilibrium’ hypothesis proposes that, with increasing survivorship, severe disability decreases but mild and moderate disability increase (Manton 1982).

In the vast majority of cases, competing hypotheses in this debate have been tested by comparing the size of average-based longevity and healthy longevity indicators, like LE and HALE (Robine et al 2020, Jagger et al 2020). Unfortunately, this population-level approach does not take into consideration the distribution of healthy and unhealthy years among individuals’ lifespans – a fundamental limitation we take up in this paper. Taking advantage of the newly proposed methods, we introduce novel approaches to test alternative hypotheses in a debate with enormous consequences for the sustainability of welfare and health care systems.

3. Methods

3.1. Bivariate distributions / Multi-state distributions

In our models, we assume that years of life can be spent either in ‘good’ or in ‘less-than-good’ health states. The precise definitions of what it means to be in one state or the other are context-specific, and in the empirical section of the paper we will show

some illustrations.[7] Drawing from the multi-state life table techniques described in Riffe et al (2023), it is possible to generate bi-variate random variables $L = (H, U)$ measuring the cumulated number of years each individual has lived in 'good' health (H) and in 'less-than-good' health (U) at the time at death (some brief details on how these distributions are arrived at are given in the Appendix).[8] The joint density function associated with L will be denoted as $f(h, u)$ (that is, $f(h, u)$ can be interpreted as the relative likelihood that a randomly chosen individual has accumulated h years in good health and u years in less-than-good health at time at death). Assuming that individuals' lifespans are bounded between 0 and ω (the maximal possible age at death), by definition one has that

$$\int_0^{\omega}\int_0^{\omega-h} f(h,u)dudh = \int_0^{\omega}\int_0^{\omega-u} f(h,u)dhdu = 1.$$

From this bi-variate distribution, it is possible to recover the standard age-at-death distribution X, which is simply defined as $H + U$. For any age at death $x \in [0, \omega]$ let $\mathcal{D}_x := \{(h, u) \in \mathbb{R}^2 | h \geq 0, u \geq 0, h + u = x\}$ be the set of pairs of non-negative values of h and u adding up to x. The elements of $\mathcal{D}_x$ describe all possible combinations of years spent in good and in less-than-good health that add up to x (e.g., $(h = 0, u = x)$, $(h = 1, u = x - 1), \cdots, (h = x, u = 0)$ belong to $\mathcal{D}_x$). The density function of X will be denoted as $\varphi(x)$, and is defined as

$$\varphi(x) = \int_{\mathcal{D}_x} f = \int_0^x f(h, x-h)dh = \int_0^x f(x-u, u)du$$

for any given age at death $x \in [0, \omega]$. By construction,

$$\int_0^{\omega} \varphi(x)\, dx = 1$$

so

$$\int_0^{\omega}\int_0^{x} f(h, x-h)dh\, dx = \int_0^{\omega}\int_0^{x} f(x-u, u)du\, dx = 1$$

Thus, (H, U) can be seen as a generalization of standard "age-at-death distributions" that allows estimating the cumulated number of years individuals have spent in different health states at the end of their lives – rather than merely accounting for their overall length, as is the case in the traditional approach. Figure 1 shows the shape of a hypothetical joint density function $f(h, u)$ associated to (H, U). Being a two-dimensional random variable, the plot of $f(h, u)$ is a 2-dimensional surface embedded

[7] As discussed below, our framework can be easily extended to more sophisticated settings including multiple (i.e., higher than two) health states. For the sake of simplicity, here we will focus our attention on the '2-health-states' case.

[8] Importantly, these models allow for the possibility that, along their life cycle, individuals transition from healthy to unhealthy states and vice-versa. Further details are given in the Appendix.

in the 3-dimensional space. In addition, we illustrate how the values of the density function of X, $\varphi(x)$, are estimated by integrating $f(h,u)$ along the $\mathcal{D}_x$ diagonals defined above.

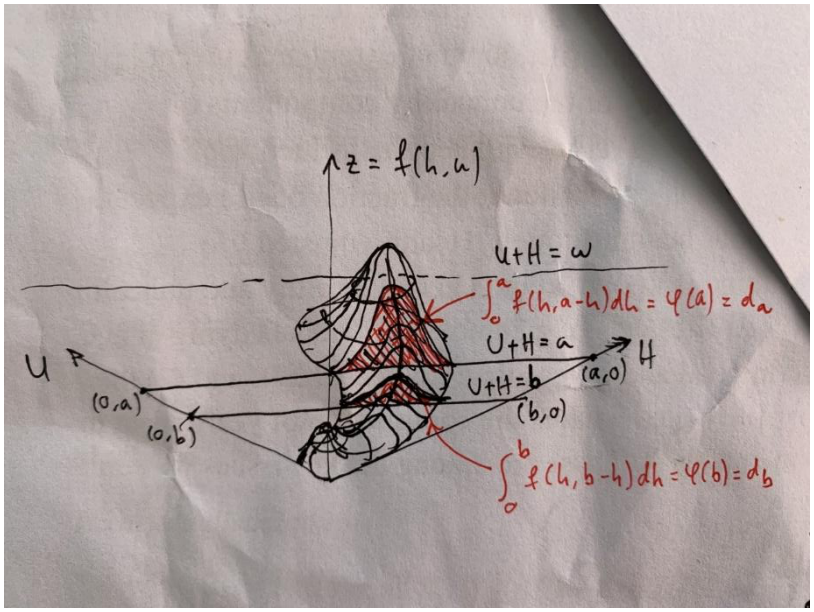


Figure 1. Illustration of a hypothetical joint density function $f(h,u)$, together with a representation of a couple of values of the age at death distribution $\varphi(x)$.

In this setting, we can define 'Health-adjusted life expectancy' as

$$\int_0^{\omega}\int_0^{\omega-h} hf(h,u)dudh = HALE.$$

This is the average number of years individuals have spent in good health throughout their lifetimes. Likewise, we can define 'Unhealthy life expectancy' as

$$\int_0^{\omega}\int_0^{\omega-h} uf(h,u)dudh = UHLE.$$

This is the average number of years individuals have lived in less-than-good health along their lifetimes. Putting together these definitions, we have that

$$HALE + UHLE = \int_0^{\omega}\int_0^{\omega-h} hf(h,u)dudh + \int_0^{\omega}\int_0^{\omega-h} uf(h,u)dudh$$

$$= \int_0^{\omega}\int_0^{\omega-h} (h+u)f(h,u)dudh = \int_0^{\omega}\int_0^{\omega-h} xf(h,x-h)dxdh = LE,$$

which corresponds to the traditional life expectancy.

3.2. Measuring and decomposing lifespan inequality

Traditional indicators of lifespan inequality measure the extent of inequality in age-at-death distributions, which are derived from life tables to account for varying population age-structures and allow comparability (van Raalte et al 2018). In the setting discussed in this paper, age at death (x) is split in two parts: the number of healthy (h) and unhealthy (u) years lived, in such a way that $x = h + u$. The aim of this section is to show examples of how traditional measures of lifespan inequality can be broken down into easily interpretable components describing how the distribution of healthy and unhealthy years contribute to overall inequality levels. The examples include decompositions of the Variance and (absolute and relative versions of) the Gini index.

The variance

The Variance of a life table age-at-death distribution X is calculated as

$$V(X) = \frac{1}{\ell_0}\int_0^{\omega} \varphi(x)(x - e_0)^2 dx$$

where $\varphi(x)$ is the share of individuals who die at age x, ℓ_0 is the initial population at age 0 and e_0 is the life expectancy at birth. It is well-known the variance of a sum of random variables ($X = H + U$) can be written as

$$V(X) = V(H + U) = V(H) + V(U) + 2Cov(H, U)$$

Such decomposition explicitly takes into consideration the association between the healthy and unhealthy components of individuals' lives. Among other things, it reveals that, other factors kept constant, a negative (resp. positive) association between h and u contributes to decrease (resp. increase) overall lifespan inequality as measured by the variance.

The Gini index

Using the same notations as above, the absolute and relative versions of the Gini index in a life table framework can be written as

$$A(X) = \frac{1}{2\ell_0^2}\int_0^{\omega}\int_0^{\omega} \varphi(a)\varphi(b)|a - b|dbda$$

$$G(X) = \frac{1}{2\ell_0^2 e_0}\int_0^{\omega}\int_0^{\omega} \varphi(a)\varphi(b)|a - b|dbda = \frac{A(X)}{e_0}$$

The absolute Gini index (A) is defined as *half* the expected age-at-death difference between two randomly chosen individuals, whereas the relative Gini index is defined as the ratio between the absolute Gini and the life expectancy of the corresponding age-at-death distribution (that is, it puts A in relation to the mean of the age-at-death distribution).

Whenever a certain variable X is defined as a sum of two other variables (as is the case in our setting: $X = H + U$), it is possible to decompose the values of the absolute and relative versions of the Gini index associated to X as the sum of two components, each explaining the contribution of the corresponding variable (i.e., H and U) to overall inequality. That is, one can write

$$A(X) = A_h + A_u$$

$$G(X) = G_h + G_u$$

where A_h (resp. A_u) is the part of total inequality ($A(X)$) that is explained by the variation in the number of healthy (resp. unhealthy) years lived – and the same goes for G_h, G_u and G. To arrive at such decomposition, we apply the method presented by Lerman and Yithzaki (1985) in the context of income inequality (details shown in the appendix)[9].

3.3. Revisiting the compression vs expansion of morbidity debate

In this sub-section, we first show the approach that has been traditionally used to test competing hypotheses in the compression vs expansion of morbidity debate, and then proceed to present the novel approach proposed here based on the analytical setting introduced in the previous sub-sections.

3.3.1. The classical approach

Since their inception, 'compression' or 'expansion of morbidity' hypotheses have been typically tested by comparing the values of LE vis-à-vis those of HALE (Robine et al 2020). Usually, increases (resp. decreases) in the ratio HALE/LE over time lend support to the compression (resp. expansion) of morbidity hypothesis.

Figure 2 illustrates how this 'classical' approach looks like in the analytical setting proposed in this paper. In this Figure, we assume that (1) at a given point in time, say t_1, the values of HALE and UHLE equal H_1 and U_1, respectively; and (2) the values of LE are expected to increase over time. Under those assumptions, Figure 2 shows the combinations of HALE and UHLE that must be observed in time t_2 for the 'expansion' or 'compression of morbidity' to occur. The line separating the opposite conclusions of 'expansion' vs 'compression' is the one satisfying the restriction $\frac{HALE}{UHLE} = \frac{H_1}{U_1}$. As an illustration, consider the values (at time t_1) of $H_1 = 60$ and $U_1 = 10$, so $LE_1 = 70$. If at time t_2 one has that $LE_2 = 80, H_2 = 69$ and $U_2 = 11$, then the classical approach would conclude that a compression of morbidity has occurred, because the fraction HALE/LE has increased from 0.857 to 0.862. Alternatively, if at time t_2 one had that $\widetilde{LE}_2 = 80, \widetilde{H}_2 = 65$ and $\widetilde{U}_2 = 15$, then it would conclude that an expansion of morbidity has occurred, because the fraction HALE declines to 0.812.

[9] In that paper, the authors estimate the contribution of different income sources (e.g., earnings, pensions, capital gains, and so on) to overall income inequality.

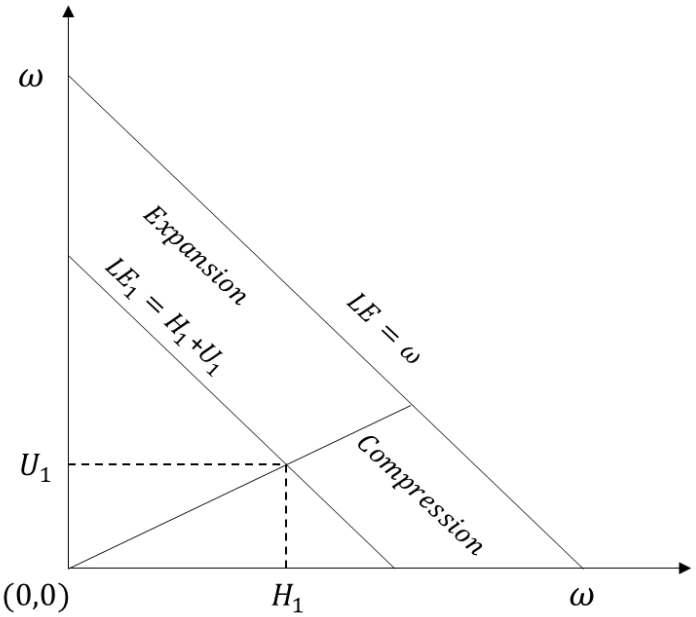


Figure 2. Traditional approach to test compression vs expansion of morbidity. Source: Authors' own elaboration.

In the classical approach, the only piece of information that is needed to reach a conclusion is the change over time in the relative size of the population-based indicators HALE and LE. This ignores the shape of the (H, U) distribution (i.e., whether individuals dying at different ages spend more or less years in good or in less-than-good health), an issue we take up in the following sub-section.

3.3.2. Healthy year curves

The new approach proposed here takes advantage of the fact that, having information about the joint distribution of H and U, we can estimate the number of years individuals have accumulated in good health *at all possible ages at death* x – thus offering a richer and more nuanced picture that goes beyond the classical approach exclusively relying on the average-based indicators LE and HALE. To show how the proposed approach works, we need to introduce some formal definitions.

Definition 1. For each age at death $x \in (0, \omega]$, let

$$M_h(x) := \int_0^x h \left[\frac{f(h, x-h)}{\int_0^x f(a, x-a)da} \right] dh$$

$$M_u(x) := \int_0^x u \left[\frac{f(x-u, u)}{\int_0^x f(x-a, a)da} \right] du$$

The first equation in Definition 1 is simply an average of the values of h for all individuals who died at age x. Thus, $M_h(x)$ measures the average number of years lived

in good health among those who died at age x. Likewise, $M_u(x)$ measures the average number of years lived in less-than-good health among those who died at age x. The curve $M_h(x)$ (resp. $M_u(x)$) will be referred to as 'healthy years curve' (resp. 'unhealthy years curve'). Similarly, it is straightforward to define the relative version of the healthy and unhealthy year curves (i.e., the functions that, for each age at death x measure the *proportion* of years lived in good (resp. less-than-good) health among those who died at age x (see Appendix)). It is easy to check (see Appendix) that, for any age at death $x \in (0, \omega]$,

$$0 \leq M_h(x) \leq x,$$
$$0 \leq M_u(x) \leq x,$$

and

$$M_h(x) + M_u(x) = x$$

That is: among those who die at age x, the average number of years lived in good health and the average number of years lived in less-than-good health add up to x. In Figure 3, we show hypothetical examples of how these $M_h(x)$ and $M_u(x)$ curves could look like.

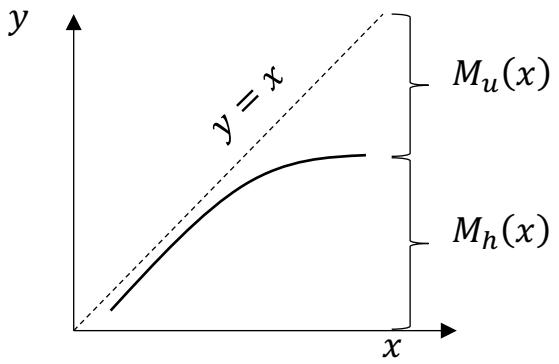


Figure 3. Illustration of healthy year and unhealthy year curves. Source: Authors' own elaboration.

Importantly, the healthy and unhealthy year curves satisfy the following identities (proofs shown in appendix 3).

$$\int_0^{\omega} M_h(x)\varphi(x)dx = HALE$$
$$\int_0^{\omega} M_u(x)\varphi(x)dx = UHLE$$

That is: weighting the mean years lived in good (resp. less-than-good) health among those who die at age x by the share of deaths occurring at that age gives the expected average number of years lived in good (resp. less-than-good) health for the entire population. These identities show how, in our setting, HALE and UHLE can be derived

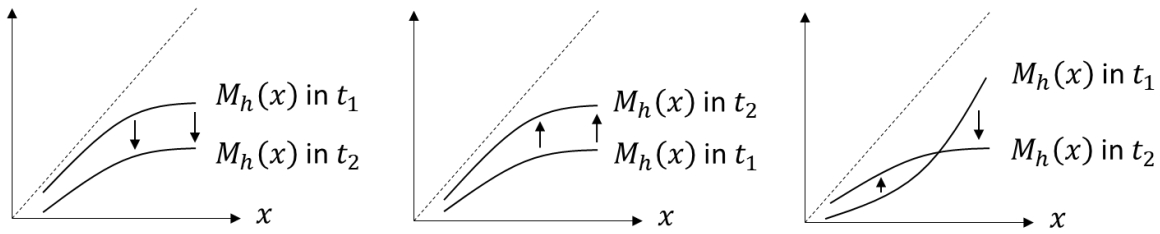
after averaging simpler age-at-death-specific estimates of the number of years individuals spend in good and less-than-good health, respectively.

Having introduced the healthy year curves, we can now present our new criteria to test alternative hypotheses in the compression vs expansion of morbidity debate.

Definition 2. Let $M_h(x)$ and $\widetilde{M}_h(x)$ be the healthy year curves for the population under study at times t_1 and t_2, respectively.

> *Case (i).* Whenever $M_h(x) \geq \widetilde{M}_h(x)$ for each age at death $x \in (0, \omega]$, then we say that there has been an expansion of morbidity between t_1 and t_2.
>
> *Case (ii).* Whenever $M_h(x) \leq \widetilde{M}_h(x)$ for each age at death $x \in (0, \omega]$, then we say that there has been a compression of morbidity between t_1 and t_2.
>
> *Case (iii).* Whenever $M_h(x) < \widetilde{M}_h(x)$ for some values of $x \in (0, \omega]$ but $M_h(x) > \widetilde{M}_h(x)$ for some other values of $x \in (0, \omega]$, then we cannot say whether morbidity has expanded or compressed between t_1 and t_2.

In case (i), the average number of years individuals have spent in good health at time at death has decreased across all possible ages at death. When this happens, it seems reasonable to conclude that morbidity has expanded between the two time periods. Likewise, if the average number of years individuals have spent in good health at time at death has increased across all possible ages at death (case (ii)), then it seems reasonable to conclude that morbidity has compressed between t_1 and t_2. Whenever the average number of years spent in good health have increased for some ages at death but decreased for other ages, then it is not obviously clear whether morbidity has expanded or compressed overall, so no conclusion is reached (case (iii)). These different scenarios are graphically illustrated in the panels Figure 4.



Expansion of morbidity **Compression of morbidity** **Neither expansion nor compression**

Figure 4. Testing the compression vs expansion of morbidity debate using healthy year curves in two time points. Source: Authors' own elaboration.

How do the classical and the new approach proposed here compare vis-à-vis each other? While the classical one compares what fraction of life expectancy is spent in good

health at the population level, the new one performs a similar exercise but across all possible ages at death (i.e., it inspects what fraction of life has been spent in good health among all those who die at a given age x, across all possible ages at death). The finer detail we are working with in the new approach comes at a cost: since we are imposing unanimity in the comparisons across *all* possible ages at death, there might be instances where a firm conclusion cannot be reached (i.e., in case (iii), when the corresponding healthy year curves cross). However, such lack of conclusiveness should not be necessarily seen as a limitation. Using the healthy year curves, we can identify the specific age ranges that have benefited the most (or the least) from health changes over time (see empirical application below).

4. Data

In this article, we use data from the HRS (Health and Retirement Study), a longitudinal panel study carried out in the United States (see https://hrs.isr.umich.edu). More specifically, we use data from the years 2000, 2005, 2010 and 2015. The respondents included in this survey have 50 years of age and above.

To operationalize what it means to be in "good" or in "less-than-good" health, we use the ADL and IADL indicators available from the HRS. That is: an individual is considered to be in "less-than-good" health whenever s/he reports experiencing any of the ADLs or IADLs included in the HRS questionnaires.

5. Results

Standard/Classical indicators

Table 1 shows the values of LE, HALE and UHLE for the US population aged above 50 over time when we use the presence of ADLs as a measure of less-than-good health. HALE and UHLE are the expected values of the bivariate random variable (H, U) associated to the corresponding populations. Results are shown for women and men separately (the results corresponding to the IADL definition of less-than-good health are shown in the Appendix). As can be seen, the three indicators increase between 2000 and 2015, both for women and for men. As expected, LE is higher among women for all years, but the gap with respect to men more than halves from 4.3 years in 2000 to 2.1 years in 2015. While LE, HALE and UHLE tend to increase over time, the rates at which these indicators grow are not the same, so the fraction HALE/LE does not keep constant. According to the classical criterion to test compression vs expansion of morbidity hypotheses presented in section 3.3.1, morbidity is expanding both for women and for men, because the fraction of life expectancy spent in good health decreases over time.

MEASURE	SEX	YEAR	LE	HALE	UHLE	HALE/LE
ADL	F	2000	27.15	23.84	3.31	0.878
ADL	F	2005	31.37	27.4	3.97	0.873
ADL	F	2010	33.01	27.57	5.44	0.835
ADL	F	2015	31.7	26.25	5.45	0.828
ADL	M	2000	22.81	20.71	2.1	0.908
ADL	M	2005	27.98	24.96	3.02	0.892
ADL	M	2010	29.41	25.77	3.64	0.876
ADL	M	2015	29.58	25.55	4.03	0.864

Table 1. Values of LE, HALE and UHLE for women and men between 2000 and 2015 using ADLs as a measure of less-than-good health. Source: Authors' elaboration based on HRS data.

Results for bivariate distributions

In Figure 5 we show the joint density functions $f(h,u)$ associated to the (H,U) distributions for women and men in 2000 and 2015 when less-than-good health is measured via the presence of ADLs (the corresponding results for the IADLs are shown in the Appendix). Inspecting the shape of the plots, it seems that there is an upward-right shift in the bivariate (H,U) distribution over time, suggesting that individuals, overall, tend to live longer lives, and at time at death they tend to accumulate more years both in good and in less-than-good health – both for the case of women and men. Likewise, the shape of the plots seems to indicate that the distribution of years accumulated in different health states becomes more dispersed over time. Comparing the plots between women and men for the same years, it seems that the former distribution is shifted upward and more spread out with respect to the latter, suggesting that women tend to accumulate more years in less-than-good health than men at time at death. The expected (i.e., mean) values of these (H,U) distributions were shown in Table 1.

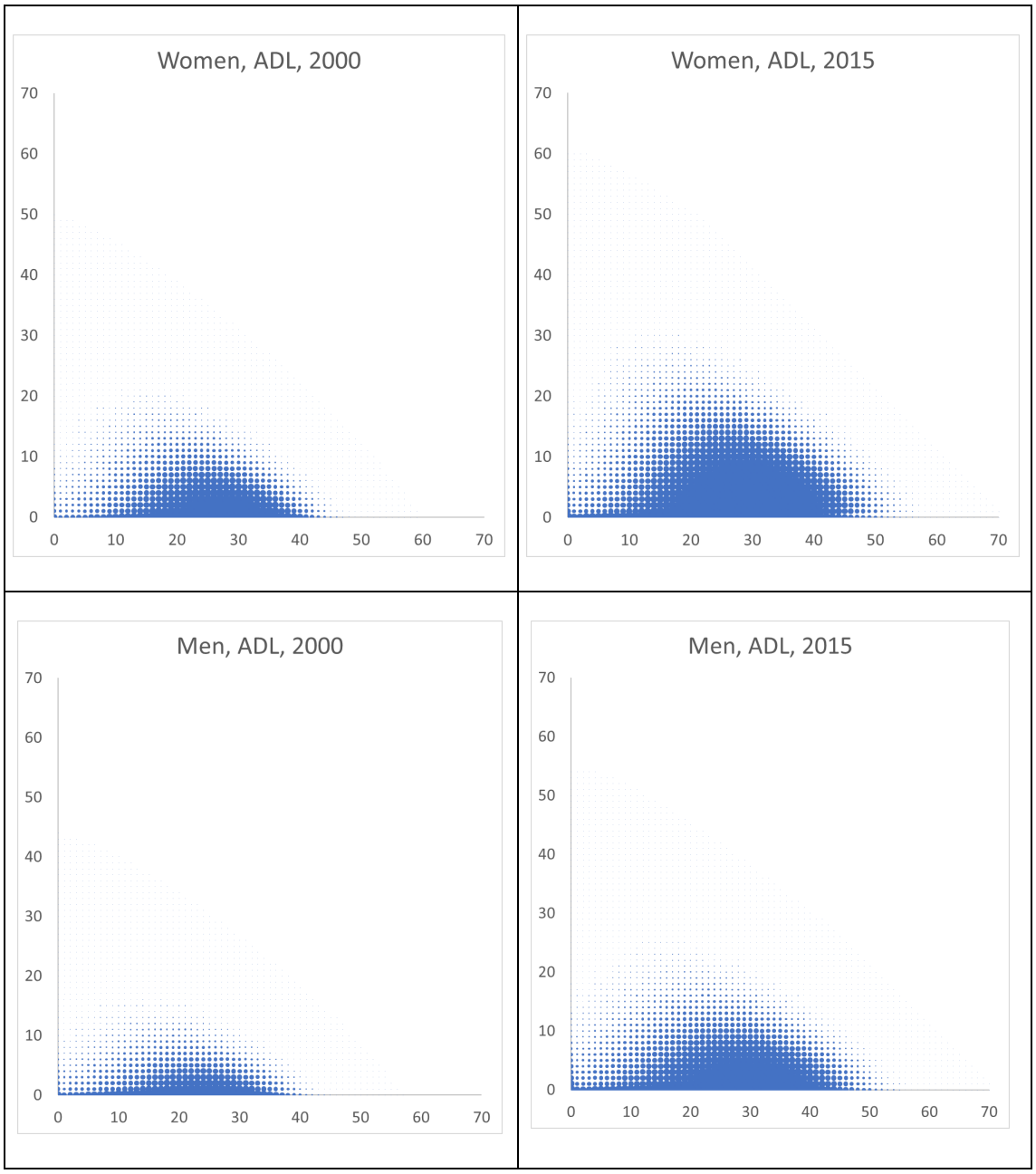


Figure 5. Graphical representation of the joint density functions $f(h,u)$ for women and men in 2000 and 2015 using the existence of ADLs as a measure of less-than-good health. Source: Authors' elaboration based on HRS data.

Going beyond visual inspection, in Table 2 we report the values of different lifespan inequality indicators and their corresponding decompositions presented in section 3.2. Using the Variance (an absolute measure of inequality), we observe that lifespan inequality has increased when moving from 2000 to 2015, both for women and for men. Interestingly, both the variation in healthy and unhealthy years (i.e., $V(H)$ and $V(U)$) also increases over time for both sexes, but the rate of increase of $V(U)$ is much higher than that of $V(H)$. The covariance between H and U is negative for all cases, thus suggesting that higher values of one variable tend to be associated with lower values of the other and vice versa.

Using the Gini index (a relative measure of inequality), we observe increases in lifespan inequality among women (from 0.181 to 0.195) but slight declines among men (0.216 to 0.209) between 2000 and 2015. While the variability in healthy years (G_h) has slightly decreased for both sexes, the variability in unhealthy years (G_u) has clearly increased, specially among women. Overall, the contribution of years lived in good health to overall lifespan inequality (i.e., $100*G_h/(G_h + G_u)$) tends to decline over time, particularly among women (from a 90.8% of the total variation explained by the healthy component in 2000 to 82.8% in 2015).

MEASURE	SEX	YEAR	V	Vh	Vu	Cov(h,u)	G	Gh	Gu	%Ch
ADL	F	2000	76.7	78.4	12.9	-7.32	0.181	0.164	0.017	90.8
ADL	F	2005	89.1	89.9	16.2	-8.57	0.168	0.156	0.012	92.8
ADL	F	2010	118.7	107.5	26.0	-7.38	0.185	0.158	0.027	85.4
ADL	F	2015	121.1	102.6	25.1	-3.32	0.195	0.162	0.034	82.8
ADL	M	2000	75.6	72.5	7.6	-2.28	0.216	0.202	0.013	93.9
ADL	M	2005	93.1	91.3	12.4	-5.33	0.194	0.182	0.012	93.6
ADL	M	2010	111.6	102.6	16.2	-3.60	0.202	0.181	0.021	89.6
ADL	M	2015	120.0	111.1	17.1	-4.10	0.209	0.188	0.021	90.1

Table 2. Levels and decompositions of the Variance and the relative Gini index for women and men between 2000 and 2015 using ADLs as a measure of less-than-good health. Source: Authors' elaboration based on HRS data.

The healthy year curves (i.e., $M_h(x)$) associated to the 2000 and 2015 (H, U) distributions are shown in Figure 6 (women in the left panel, men in the right one). All the curves shown in that figure increase with age at death x, thus implying that the higher the number of years that individuals survive, the higher the number of years they have accumulated in good health (in absolute terms) *on average*. As can be seen, the 2000 $M_h(x)$ curves steadily increase with age at death x until $x \approx 50$, and then they become much flatter (meaning that, for those individuals surviving beyond those ages, the gain in the number of years lived in good health is relatively lower). In contrast, the 2015 $M_h(x)$ curves increase at a more constant pace with increasing age at death x, specially among women.

Interestingly, the 2000 and 2015 healthy year curves cross when age at death x approaches 50. For those dying at ages below that threshold, the number of years accumulated in good health was lower in 2015 than in 2000, and the opposite happens for those dying above the threshold. Stated otherwise, in 2000, those who died at ages below 50 did so having accumulated a higher number of years lived in good health than those who had died at the same ages in 2015. However, the situation reversed for those dying above that threshold. Thus, following the criterion suggested in this paper to test whether morbidity is compressing or expanding over time, we cannot reach a definite conclusion – in contrast to the conclusion arrived at by the classical approach based on HALE/LE ratios (see above).

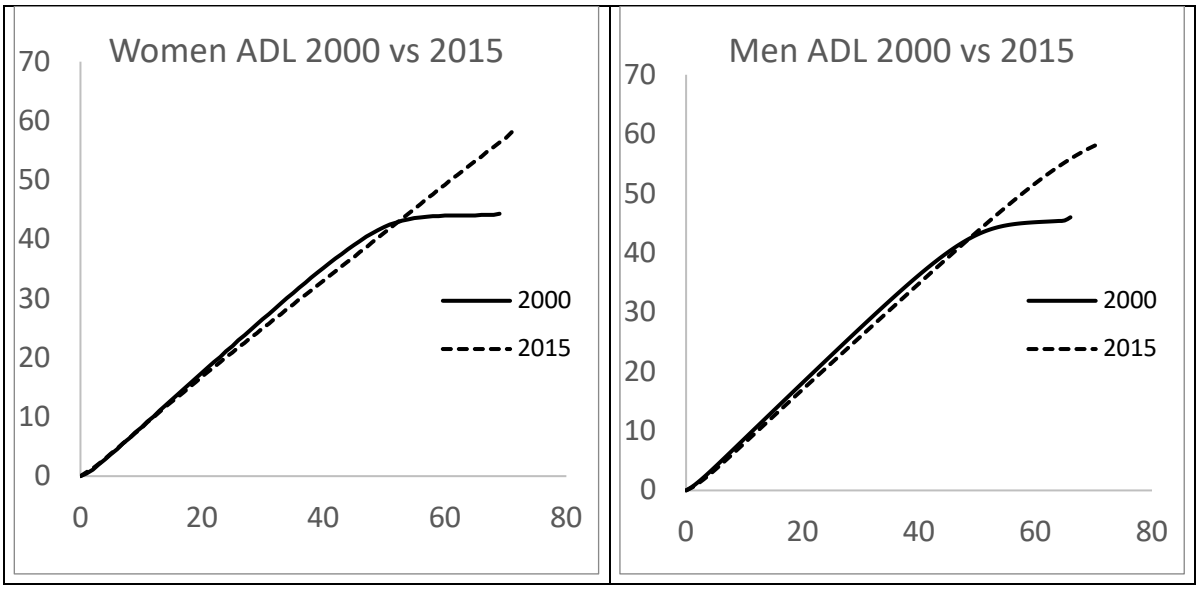


Figure 6. Healthy year curves for women and men in 2000 and 2015 using ADLs as a measure of less-than-good health. Source: Authors' elaboration based on HRS data.

6. Discussion

Studies investigating mortality dynamics typically assume that gains in longevity are desirable no matter what. However, the fact that a non-negligible and potentially large and increasing fraction of individuals' lifespans are composed of years spent in less-than-good health question the validity of this assumption. The trade-offs between the quantity of years of life and the 'health quality' of those years could potentially affect our assessments of populations' and individuals' health performance. This state of affairs cast doubts on the use of standard approaches followed to judge contemporary health dynamics.

The tools presented in this paper allow estimating how *individuals'* lifespans are composed of years spent in good and in less-than-good health at time at death. Analytically, the possibility of breaking down individuals' length of life as the sum of time spent in different health states is a major breakthrough with respect to currently existing approaches to investigate (healthy) population ageing, which are almost exclusively based on average-based indicators (i.e., akin to LE, HALE and UHLE). The new approach opens the possibility of going beyond state-of-the-art methods to advance our knowledge about the complex interplay between mortality and morbidity.

On the one hand, the methods proposed here allow exploring the relationship between the number of years individuals accumulate in good (H) and in less-than-good health (U) at time at death. Inter alia, such relationship has implications for the extent and composition of lifespan variability. Other factors kept constant, whenever the relationship between H and U is strong and negative (i.e., an accumulation of a large number of years spent in good health tends to be accompanied by an accumulation of a small number of years in less-than-good health at time at death, and vice versa), the

variability in the ages at which individuals die tends to decline. Analogously, a strong and positive association between H and U contributes to increase lifespan variability. In the different empirical applications presented here, the association between H and U is negative, but not particularly strong. In future research, it would be interesting to investigate the sign and strength of such association among other populations to gain further insights into the relationship between mortality and morbidity.

On the other hand, the approach discussed in this paper allows investigating what portion of overall variability in length of life is attributable to the number of years individuals have accumulated in good or in less-than-good health. Lifespan variability is an important marker of populations' health heterogeneity (van Raalte et al 2018), so it fundamental to understand what its main sources are. Our findings indicate that while most of the variability in lifespans is attributable to the number of years individuals have accumulated in good health (a normatively desirable indicator), the contribution made by the years spent in less-than-good health (an indicator whose normative desirability is unclear) is non-negligible and becoming increasingly important over time (e.g., among women, the contribution of the latter increased from 10% in 2000 to 18% in 2015)[10]. Therefore, the years lived in morbid states are becoming an increasingly prominent factor explaining the variability in the ages at which individuals die. Future research should determine whether similar results obtain when measuring less-than-good health with alternative indicators.

Lastly, we present a more refined approach to assess whether morbidity is compressing or expanding over time. Using traditional techniques based on the evolution of the HALE/LE ratio, our data suggests that morbidity has expanded among the US population aged 50 and above between 2000 and 2015. The new approach proposed here based on the relative position of the healthy year curves gives a more nuanced picture, suggesting that, between those two years, the average number of years accumulated in good health has decreased at time at death for those dying before age 50, and the reverse happens for those dying after that age. Since the vast majority of deaths occur before that age, the new approach seems to give more support to the expansion of morbidity side of the debate.

These findings cohere with recent studies suggesting that, as survival prospects further improve in low-mortality countries, the health profiles of the elder become an increasingly heterogeneous mix of robust and frail individuals (Engelman et al 2010), with an increasing prominence of the years that are lived in morbid states.

[10] The magnitude of these contributions is similar to the one found in the study of Permanyer and Bramajo (2023), albeit in a different setting (in that paper, the authors investigate the contribution of country-level HALE and UHLE to inequality in LE across world countries and found that the latter contributed to around 10% of that inequality).

Our study has several limitations. The method we have used to estimate the number of years individuals accumulate in different health states at time at death is based on transition probabilities that are measured between two points in time only. Thus, the age-specific transitions are applied to a fictitious cohort of individuals that are subject to the same transition probabilities throughout their lifetimes. Inter alia, this implies that the process has 'no memory', that is: at each age, individuals are exposed to the risks of moving throughout health states independently of what their previous health trajectories were. While unrealistic, this is the kind of simplifying assumption that is built in standard life table methods or in many Markov chain models traditionally used to calculate period life expectancy or other health-related indicators. In future research, it would be interesting to overcome such limitations resorting to richer longitudinal data that allows assessing individuals' *true* (i.e., not simulated) health trajectories over time.

Extensions

In this paper, we have only worked with two health states: 'good' and 'less-than-good' health. It is certainly possible to extend the method to analyze more health states (e.g., 'very bad health', 'very good health', and so on). If we denote by $k \geq 2$ the number of well-defined health states we are dealing with, it is possible to replicate the same approach and create a k-dimensional random variable $L = (H_1, H_2, \cdots, H_k)$, where H_i would measure the number of years accumulated in health state 'i' at time at death. In that setting, the random variable measuring length of life would simply be

$$X = \sum_{i=1}^{k} H_i$$

Using analogous methods, one could generate the corresponding density function $f(h_1, h_2, \cdots, h_k)$ and study not only the association between the different pairs of variables (H_i, H_j), but also the contribution of each H_i to overall lifespan variability levels. Indeed, using the same techniques discussed in section 3.2, it is easy to check that, in this extended setting,

$$V(X) = \sum_{i=1}^{k} V(H_i) + \sum_{i=1}^{k} \sum_{j \neq k} Cov\left(H_i, H_j\right)$$

$$A(X) = \sum_{i=1}^{k} A_i$$

$$G(X) = \sum_{i=1}^{k} G_i$$

where A_i (resp. G_i) is the part of total inequality $A(X)$ (resp. $G(X)$) that is explained by the variation across individuals in the number of years accumulated in health state 'i' at time at death. Such richer models based on more than two health states could be potentially used to test the dynamic equilibrium hypothesis proposed by Manton (1982), which alludes to the varying degrees of severity of diseases and disabilities.

7. Conclusion

By moving from an average-based to an individual-based approach, the ideas and methods presented in this paper are able to reveal a great deal of health variation that cannot be identified with currently existing methods to study population ageing. Importantly, they bridge and combine "mortality analysis" with "health & morbidity analyses" into a unified and coherent framework, thus opening exciting research avenues that hold promise to uncover a much deeper and comprehensive understanding of contemporary health dynamics.

References

J. M. Aburto, F. Villavicencio, U. Basellini, S. Kjærgaard, J. W. Vaupel, Dynamics of life expectancy and life span equality. Proc. Natl. Acad. Sci. U.S.A. 117, 5250–5259 (2020).

Aburto, J. M., & Raalte, A. van. (2018). Lifespan Dispersion in Times of Life Expectancy Fluctuation: The Case of Central and Eastern Europe. Demography, 55(6), 2071–2096. https://doi.org/10.1007/s13524-018-0729-9

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

Edwards, Ryan D. 2011. "Changes in World Inequality in Length of Life: 1970–2000." *Population and Development Review* 37(3): 499–528. https://doi.org/10.1111/j.1728-4457.2011.00432.x

Fries James F. 1980. "Aging, natural death, and the compression of morbidity." *The New England Journal of Medicine* 303(3), 130–135. https://doi.org/10.1056/NEJM198007173030304

Gerstorf, Denis, Nilam Ram, Christina Röcke, Ulman Lindenberger, and Jacqui Smith. 2008. "Decline in Life Satisfaction in Old Age: Longitudinal Evidence for Links to Distance-to-Death." *Psychology and Aging* 23(1): 154–168. https://doi.org/10.1037/0882-7974.23.1.154

Gruenberg, Ernest M. 1977. “The Failures of Success.” The Milbank Memorial Fund Quarterly. Health and Society 55(1), 3–24. https://doi.org/10.2307/3349592

Jagger, C., Crimmins, E. M., Saito, Y., De Carvalho Yokota, R. T., Van Oyen, H., & Robine, J.-M. (2020). *International Handbook of Health Expectancies* (C. Jagger, E. M. Crimmins, Y. Saito, R. T. De Carvalho Yokota, H. Van Oyen, & J.-M. Robine, Eds.; internatio ed.). Springer International Publishing. https://doi.org/10.1007/978-3-030-37668-0

Lawton, M. P., M. Moss, C. Hoffman, R. Grant, T. Ten Have, and M. H. Kleban. 1999. “Health, Valuation of Life, and the Wish to Live.” *The Gerontologist* 39(4): 406–416. https://doi.org/10. 1093/geront/39.4.406

Lerman, Robert I., and Shlomo Yitzhaki. 1985 “Income Inequality Effects by Income Sources: A New Approach and Applications to the United States.” *The Review of Economics and Statistics* 67(1): 151–156.

Mas-Colell, A., Whinston, M. and Green, J. (1995), *Microeconomic Theory*. Oxford, Oxford University Press.

Manton Kenneth G. 1982. “Changing concepts of morbidity and mortality in the elderly population” *The Milbank Memorial Fund quarterly. Health and society* 60(2): 183–244.

Oeppen, Jim, and James W. Vaupel. 2002. “Demography: Broken Limits to Life Expectancy.” *Science* 296: 1029–1031. https://doi.org/10.1126/science.1069675

Permanyer, I. and Bramajo, O. (2023), “The Race between Mortality and Morbidity: Implications for the Global Distribution of Health”, *Population and Development Review*, https://doi.org/10.1111/padr.12582

Permanyer, I., Villavicencio, F. and Trias-Llimós, S. (2023), “Healthy lifespan inequality: morbidity compression from a global perspective”, *European Journal of Epidemiology*, DOI: 10.1007/s10654-023-00989-3.

Permanyer, Iñaki, and Nathalie Scholl. 2019. “Global Trends in Lifespan Inequality: 1950–2015.” *PLoS ONE* 14(5): 1–19. https://doi.org/10.1371/journal.pone.0215742

Permanyer I, Spijker J, Blanes A, Renteria E. Longevity and Lifespan Variation by Educational Attainment in Spain: 1960–2015. *Demography* 2018; **55**(6): 2045-70

Permanyer I, Spijker J, Blanes A. (2022), “On the measurement of healthy lifespan inequality”, *Population Health Metrics*, 20:1. doi:10.1186/s12963-021-00279-8

Riffe, T., Permanyer, I., Tursun-Zade, R. and Muszynska-Spielauer, M. (2023), "Calculating a state occupancy distribution in multistate settings", paper presented at the 34th REVES annual meeting.

Riley, James C. 2005. "Estimates of Regional and Global Life Expectancy, 1800–2001." *Population and Development Review* 31(3): 537–543. https://doi.org/10.1111/j.1728-4457.2005.00083.x.

Robine, Jean-Marie, Carol Jagger, Eileen M. Crimmins, Yasuhiko Saito, and Herman Van Oyen. 2020. "Trends in Health Expectancies." Edited by Carol Jagger, Eileen M. Crimmins, Yasuhiko Saito, Renata Tiene De Carvalho Yokota, Herman Van Oyen, and Jean-Marie Robine, *International Handbook of Health Expectancies* (pp. 19–34). Springer International Publishing. https://doi.org/10.1007/978-3-030-37668-0_2

Sasson I. Trends in Life Expectancy and Lifespan Variation by Educational Attainment: United States, 1990–2010. *Demography* 2016; **53**(2): 269-93.

Seaman R, Höhn A, Lindahl-Jacobsen R, *et al.* Rethinking morbidity compression. *Eur J Epidemiol* 2020;**35**:381–8. doi:10.1007/s10654-020-00642-3

Seaman R, Riffe T, Leyland AH, Popham F, van Raalte A. The increasing lifespan variation gradient by area-level deprivation: A decomposition analysis of Scotland 1981–2011. *Soc Sci Med* 2019; **230**: 147-57.

Smits, Jeroen, and Christaan Monden. 2009. "Length of life inequality around the globe." *Social Science & Medicine* 68: 1114–1123.

van Raalte AA, Martikainen P, Myrskylä M. Lifespan variation by occupational class: compression or stagnation over time? Demography 2014; 51(1): 73-95.

van Raalte, A. A., Sasson, I., & Martikainen, P. (2018). The case for monitoring life-span inequality. Science, 362(6418), 1002–1004. https://doi.org/10.1126/science.aau5811

van Raalte, A. A., & Caswell, H. (2013). Perturbation Analysis of Indices of Lifespan Variability. *Demography*, *50*(5), 1615–1640. https://doi.org/10.1007/s13524-013-0223-3

J. W. Vaupel, Z. Zhang, A. A. van Raalte, Life expectancy and disparity: An international comparison of life table data. BMJ Open 1, e000128 (2011).

Vigezzi, S., Aburto, J.-M., Permanyer, I., & Zarulli, V. (2022). Divergent trends in lifespan variation during mortality crises. *Demographic Research*, *46*, 291–336

Appendices

Appendix 1. Lerman and Ytzhaki decomposition of the Gini index

Appendix 2.

$$M_h(x) + M_u(x) = \int_0^x (h+u)\left[\frac{f(h,x-h)}{\int_0^x f(a,x-a)da}\right]dh = \int_0^x x\left[\frac{f(h,x-h)}{\int_0^x f(a,x-a)da}\right]dh =$$
$$= x\int_0^x \left[\frac{f(h,x-h)}{\int_0^x f(a,x-a)da}\right]dh = x\frac{\int_0^x f(a,x-a)da}{\int_0^x f(a,x-a)da} = x$$

Appendix 3.

Prove that

$$\int_0^\omega M_h(x)\varphi(x)dx = HALE$$
$$\int_0^\omega M_u(x)\varphi(x)dx = UHLE$$

Appendix 4

Relative version of $M_h(x), M_u(x)$

Definition 2. For each age at death $x \in (0, \omega]$, let

$$P_h(x) := \frac{1}{x}\int_0^x h\left[\frac{f(h,x-h)}{\int_0^x f(a,x-a)da}\right]dh = \frac{M_h(x)}{x}$$
$$P_u(x) := \frac{1}{x}\int_0^x u\left[\frac{f(x-u,u)}{\int_0^x f(x-a,a)da}\right]du = \frac{M_u(x)}{x}$$

The functions introduced in Definition 2 simply are the relative version of the 'healthy' and 'unhealthy year curves'. Thus, $P_h(x)$ (resp. $P_u(x)$) measures the *proportion* of years lived in good (resp. less-than-good) health among those who died at age x. Thus, it is easy to check that, for all possible ages at death $x \in (0, \omega]$, the following identity holds

$$P_h(x) + P_u(x) = 1$$

The relative versions of the healthy and unhealthy year curves yield the following identities.

$$\int_0^{\omega} P_h(x)\varphi(x)dx = \int_0^{\omega}\int_0^{x}\left(\frac{h}{h+u}\right)f(h,x-h)dhdx = \int_0^{\omega}\int_0^{x}\frac{h}{x}f(h,x-h)dhdx$$

$$\int_0^{\omega} P_u(x)\varphi(x)dx = \int_0^{\omega}\int_0^{x}\left(\frac{u}{h+u}\right)f(h,x-h)dhdx = \int_0^{\omega}\int_0^{x}\frac{u}{x}f(h,x-h)dhdx$$

The first (resp. second) equation is an average of the fractions of life spent in good (resp. less-than-good) health across all the individuals in the population. Interestingly, this quantity does *not* necessarily coincide with the value of HALE/LE (and the same happens with the second equation and UHLE/LE). Stated otherwise: the average of fractions of life spent in good health across individuals is typically different from the fraction of averages HALE/LE[11].

Appendix 5.

MEASURE	SEX	YEAR	LE	HALE	UHLE	HALE/LE
IADL	F	2000	27.24	24.6	2.64	0.903
IADL	F	2005	31.42	27.71	3.71	0.882
IADL	F	2010	33.12	28.16	4.96	0.850
IADL	F	2015	31.97	26.92	5.05	0.842
IADL	M	2000	22.9	21.13	1.77	0.923
IADL	M	2005	27.84	25.29	2.55	0.908
IADL	M	2010	29.38	26.06	3.32	0.887
IADL	M	2015	29.7	26.33	3.37	0.887

MEASURE	SEX	YEAR	V	Vh	Vu	Cov(h,u)	G	Gh	Gu	%Ch
IADL	F	2000	77.5	77.1	10.6	-5.08	0.181	0.170	0.011	94.1
IADL	F	2005	98.1	100.1	14.9	-8.45	0.177	0.168	0.009	94.9
IADL	F	2010	118.5	107.5	23.4	-6.21	0.185	0.159	0.025	86.4
IADL	F	2015	129.4	111.9	23.6	-3.04	0.201	0.170	0.030	84.9
IADL	M	2000	74.7	69.5	6.4	-0.62	0.213	0.198	0.015	93.0
IADL	M	2005	92.3	88.6	10.3	-3.34	0.194	0.181	0.013	93.4
IADL	M	2010	110.6	99.5	13.2	-1.05	0.202	0.181	0.021	89.7
IADL	M	2015	114.2	106.2	14.8	-3.39	0.203	0.184	0.019	90.8

[11] The reason why these two quantities do not necessarily coincide is because arithmetic averages are additive, while fractions are multiplicative – so to speak. For instance, if $x_1, \cdots, x_n, y_1, \cdots, y_n$ are real non-negative numbers, then $(1/n)\sum_i (x_i/y_i)$ generally differs from $\big((1/n)\sum_i x_i\big)/\big((1/n)\sum_i y_i\big)$.