BMJ Open

Not all females outlive all males: A new perspective on lifespan inequalities between sexes

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-059964
Article Type:	Original research
Date Submitted by the Author:	07-Dec-2021
Complete List of Authors:	Bergeron-Boucher, Marie-Pier; Syddansk Universitet, Interdisciplinary Center on Population Dynamics Álvarez, Jesús-Adrian; University of Southern Denmark, Interdisciplinary Centre on Population Dynamics Kashnitsky, Ilya; Syddansk Universitet, Interdisciplinary Centre on Population Dynamics Zarulli, Virginia; Syddansk Universitet, Interdisciplinary Center on Population Dynamics
Keywords:	PUBLIC HEALTH, STATISTICS & RESEARCH METHODS, EPIDEMIOLOGY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Not all females outlive all males:
A new perspective on lifespan inequalities between sexes

Marie-Pier Bergeron-Boucher, corresponding author, mpbergeron@sdu.dk
Jesús-Adrián Álvarez, alvarez@sdu.dk
Ilya Kashnitsky, ikashnitsky@sdu.dk
Virginia Zarulli, vzarulli@health.sdu.dk

Danish Center for Population Research, Interdisciplinary Center on Population Dynamcis,
University of Southern Denmark, Odense, Denmark

ABSTRACT

Objective To determine the probability of males outliving females.

Design International comparison of national sex-specific life tables from the Human Mortality Database.

Setting 44 populations since the middle of the 18th century.

Primary outcome measure We use the outsurvival statistic (φ) to measure inequality between sexes, which is here interpreted as the probability of males to outlive females.

Results In random pairs at age zero of a male and a female, the probability of the male outliving the female varies between 25% and 50% for lifetables in almost all years and across almost all populations. We show that φ is positively correlated with sex differences in life expectancy and negatively correlated with the level of lifespan variation. The important reduction of lifespan inequality observed in recent years has made it less likely for a male to outlive a female.

Conclusions Although male life expectancy is generally lower than female life expectancy and male death rates are usually higher at all ages, males have a substantial chance of outliving females. These findings challenge the general impression that "men do not live as long as women" and reveal a more nuanced inequality in lifespans between females and males.

Keywords: Lifespan, Inequality, Sex differences, Outsurvival statistics

Strengths and limitations of this study

- This study acknowledges that the lifespan distributions of females and males overlap.
- It is the first study using the outsurvival statistics to quantify the probability of males outliving females.
- The study shows how the outsurvival statistics is influenced by more traditional measures of lifespan inequalities: differences in life expectancy and standard deviation.
- The outsurvival statistic does not account for dependance between individuals.

1. INTRODUCTION

The female survival advantage has been observed over time across many human populations and is rooted in a complex combination of biological, environmental and behavioral factors [1-5]. For example, males tend to have more risky behaviors, such as smoking and heavy drinking; but estrogen could also be preventive against certain diseases [6]. Females have been found to have longer survival and lower death rates than men at all ages and in most modern populations [2, 4, 7-10] and so even under extreme mortality conditions [11].

Sex differences in mortality and longevity are often identified by comparing life expectancy between females and males, which summarizes the average length of life[1]. These differences are often interpreted as "men do not live as long as women". Such an interpretation is simplistic as it does not account for the variation around the means (life expectancies) and potential overlap between female and male lifespan distributions. Despite females having higher life expectancy than males, not all females outlive all males. The lifespan distributions of females and males substantially overlap.

Lifespan variation, i.e. differences in lifespans within a population, has been receiving an increasing attention in the literature [12]. Various indicators reveal heterogeneity in the length of life, beyond what life expectancy indicates. Studies have compared lifespan variation between two populations, focusing on which populations exhibit more inequalities [13-15]. It has been shown that females systematically experience lower lifespan variation than males [13]. However, how this variation around the means leads to potential overlap between two lifespan distribution has been overlooked.

[1] When analyzing period data, life expectancy is the average length of life for a hypothetical cohort of individuals - i.e. it is the average length of life if individuals born a given year experienced the age-specific death rates observed that same year.

To measure the overlap between two distributions, previous research has suggested investigating how different two lifespan distributions are using, for example, the Kullback-Leibler (KL) divergence [16]. Stratification indexes, based on how much two lifespan distributions overlap, have also been used to study mortality differences between socioeconomic groups [17]. The interpretation of these indexes can, however, be cumbersome. For example, the KL divergence can be interpreted as the amount of "effort" needed to transform the male's lifespan distribution into the female's distribution.

In this article, we use a more straightforward measure, the *outsurvival statistic* [18], which quantifies the probability that an individual from a population with lower life expectancy outlives an individual from another population with higher life expectancy. If the two populations are males and females, the statistic captures the correctness of the assertion that males' lifespans are lower than females' lifespans. We aim to 1) quantify the probability that males outlive females over time and across populations; and 2) assess the sensitivity of the outsurvival statistic to changes in life expectancy and lifespan variation. We computed the outsurvival probability to study sex differences in mortality in 44 populations covering over 200 years of data. Despite sometimes large differences in life expectancy, we show that there are substantial overlaps between males' and females' lifespan distributions.

2. METHOD

2.1 OUTSURVIVAL STATISTIC

Consider two populations with mean and standard deviation (SD) specified in panel A of Figure 1. The first population (in red) has a smaller mean lifespan and larger SD than the second population (in blue). An inference from the means would be that individuals in the first population are worse off than individuals in the second. However, there is an important overlap between the two distributions, with some individuals in the first population outliving some individuals in the second population. The outsurvival probability, φ (phi), captures this dimension by measuring the probability that an individual from a population with high mortality will outlive an individual from a population with low mortality [18]. Let $d_i(x), i = 1,2$, denote the lifespan distribution at age x in two populations. The cumulative distributions are represented by $D_i(x)$, such that such $D_i(x) = \int_0^x d_i(t)dt$ and the survivorship is denoted by $\ell_i(x)$, with $\ell_i(x) = 1 - D_i(x)$. The probability that an individual from the first population (males) will outlive an individual from the second population (females) is [18]:

$$\varphi = \int_0^\infty d_2(x)\ell_1(x)dx. \quad (1)$$

In scenario A of Figure 1, φ is 40%.

[Figure 1]

In Appendix A, we show that the outsurvival statistic relates to the joint probability density function of two lifespan distributions, which gives the probability of realizations of two lifespans and thus is related to the overlap of the two distributions.

2.2 RELATION TO LIFE EXPECTANCY AND LIFESPAN VARIATION

Consider the two populations in scenarios B and C of Figure 1. The difference in mean lifespan is the same in both cases, i.e. 15 years. However, in scenario C the first population has a larger SD, which implies more individuals surviving to older ages, despite greater inequalities, and thus a greater potential to outlive individuals from the second population. Indeed, φ is higher in scenario C (19%) than in scenario B (14%). Now compare scenario B to scenario D. This time, the second population in D has a smaller SD, with fewer individuals dying at younger ages, making it more difficult for individuals in the first population to outlive them. This reduces φ to 12%. Thus, for the same difference in life expectancy, larger lifespan variation in both populations generally results in larger φ . The comparison of scenarios A and C also shows that small differences in life expectancy lead to larger value of φ.

Equation (1) is not new and is equivalent to the expected failure probability in a stress-strength interference (SSI) model, which assesses the probability that the stress (population 1) exceeds the strength (population 2) of a material [19][2]. If the distributions of both populations follow a Normal distribution with mean μ_i and standard deviation s_i, the probability of failure is P(Z) with $Z = -\frac{\mu_2 - \mu_1}{\sqrt{s_1^2 + s_2^2}}$ [20]. This relation formalizes what is illustrated in section in Figure 1: φ is sensitive to the difference in the means and to the level of variation in both distributions, with smaller mean differences (numerator) and larger variance (denominator) leading to larger P(Z). However, lifespan distributions are not normally distributed and additional moments could also affect the value of φ. To

[2] The outsurvival statistics can also relate to the Mann-Whitney U test and the probability of superiority.

better understand this relation, we analyzed the correlation between φ and life expectancy as well as between φ and lifespan variation.

2.3 DISCRETE APPROXIMATION

In a discrete time setting, similar equivalences to equation (1) can be found. Let ${}_nd_x^i$ be the life table deaths between age x and $x+n$ in population i and ${}_nD_x^i$ the cumulative distribution. For a given age-group width of n, the probability of individuals in the first population outliving those in the second population can be found by:

$$\varphi \approx \sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nD_{x-n}^2 + \overline{d} = \sum_{x=0}^{\omega} {}_nd_{x-n}^2 \, {}_nl_x^1 + \overline{d} \qquad (2)$$

with $\overline{d} = \frac{\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2}{2}$ and $\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2$ being the probability that individuals in both populations died in the same age-group. The latter statistic is sensitive to the width of the age-groups such that smaller age-groups result into smaller values, with $\lim_{n \to 0} \sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2 = 0$. In the Appendix B, we compared the discrete and continuous approaches and find that both approaches yield comparable results.

Equations (1) and (2) are equivalent to matching random individuals from each population and calculating the proportions of individuals from the first population outliving the paired individual from the second. We performed such analysis via simulations of individuals from a specific lifespan distribution and estimated the corresponding statistics (see Appendix B). Equivalent results were found.

3. DATA

We used life tables by sex for all available countries and years from the Human Mortality Database [21]. Subnational data were used for Germany with separate analysis for East and West Germany; and the United Kingdom for England-Wales (total population), Scotland and Northern Ireland, totalizing 44 populations. The earliest year with available data was 1751 (for Sweden). Information about the available populations and years are provided in Appendix C. We compared females and males' life tables in each country/region.

4 RESULTS

Figure 2 shows the outsurvival probability of males over females (φ) since 1850 for all HMD countries. The probability of males outliving females has, at all points in time and across all populations, varied between 25% and 50%, with only one exception: Iceland in 1891 (51.3%), an exceptional year when female life expectancy was also lower than male life expectancy. Before the First World War, φ was slowly decreasing, on average from 47.3% in 1850 to 46.0% in 1913. Afterwards, φ declined faster. In 1930, the mean φ across populations was 45.4% (ranging from 42.8% (France) to 48.4% (Netherlands)). By 1985, the mean φ was 35.3% (ranging from 31.2% (Russia) to 42.8% (Israel)). The value of φ started increasing around the 1980s for some countries, but continued to decrease in others until the 2000s, especially in Eastern European countries. The mean in 2017 was 37.1%, with values varying between 28.7% (Belarus) and 42.5% (Iceland).

[Figure 2]

Figure 3 shows that φ is negatively correlated with the differences in life expectancy and positively correlated with females' standard deviations (similar results were found when males' standard deviations were used, due to the strong correlation between females and males' standard deviations). This relation empirically demonstrates the formal relation in section 2.2. The correlation between φ and the standard deviation is weaker in recent years, due to reduction in sex differences in life expectancy, which is also driving changes in φ . Even though both life expectancy and lifespan variation affect φ, the statistic appears more sensitive to the differences in life expectancy than to the level of lifespan variation. We found similar results for cohort data (see Appendix D).

[Figure 3]

The same φ value can then be found for different combinations of sex differences in life expectancy and levels of lifespan variation. For example, the same φ of 36.1% was found in France in 1962 and in 2018 (Figure 3). However, the sex difference in life expectancy was 6.9 in 1962 and 5.9 in 2018 and the standard deviation for females was 18.1 in 1962 and 13.6 in 2018.

Figure 4 shows the same relations as shown in Figure 3 but for survivors to age 50. Lifespan variation at age 50 has stayed roughly constant over time [22] and comparing φ from this age can help assessing the sensitivity of the measure to changes in lifespan variation. The relation between φ and differences in life expectancy is stronger and more linear from age 50 (correlation coefficient of -0.99) than when using the full age-range, increasing predictive ability. For example, for a difference in life expectancy at age 50 of 3 years, males have around 42% probability to outlive females. Note that φ in France was 35.9% in 1962 and 36.3% in 2018.

Similar to the distribution from birth, the probability of males outliving females from age 50 has, in almost all periods and populations, varied between 28% and 50%, with only few exceptions. In recent years, the φ statistics from birth and from age 50 are similar.

[Figure 4]

5. DISCUSSION

Our analysis reveals important overlaps between the distributions of lifespans of females and males. Due to a mixture of cultural, social, epidemiological and biological factors, males tend to die earlier, on average, compared to females [1, 3]. Still, we found that despite sometimes large differences in life expectancy, between one and two men out of four outlived a randomly paired woman in almost all points in time and across 44 populations. For example, a sex difference in life expectancy at birth of 10 years can come with a probability of males outliving females as high as 40%. These findings challenge the general impression that "men do not live as long as women" and reveal a more nuanced inequality in lifespan between females and males.

Another important result of our analysis is that the smaller the difference in life expectancy and the larger the standard deviation, the higher the probability that males outlive females. On the one hand, the narrowing sex difference in mortality in the last decades [23] would lead to bigger proportions of males outliving females. On the other hand, the important reduction in lifespan variation observed over time for both sexes [13, 24, 25] reduces the probability of males outliving females. Therefore, within the same country and given the same sex difference in life expectancy, recent values of φ are often smaller than the ones observed in earlier years, due to smaller lifespan variation observed in recent years.

The reduction in lifespan variation represents a major achievement of the mortality improvement process that has reduced inequalities in longevity between individuals. However, the reduction of lifespan inequality has also made it less likely for males to outlive females. This is partly explained by the fact that lifespan variation reduction has been driven by mortality declines at younger ages [25]. When looking at the lifespan distribution (as in Figure 1, scenario D), survival improvements at younger ages narrowed the left tails of the distribution for both sexes. By reducing the left tail of female distribution, without increasing the right tail of the male distribution, the overlapping area is reduced. In other words, as the number of short-lived females decreases over time, males must increasingly shift their distribution to higher ages to be able to outlive females.

Trends over time in φ are consistent with the reversed trends in sex differences in life expectancy [23]: the probability of males outliving females decreased until the 1970s, after which it gradually increased in all populations. H Beltran-Sanchez, CE Finch and EM Crimmins [7] showed that the increase in sex differences in mortality emerged in cohorts born after 1880, which is consistent with our analysis of φ (see Appendix D).

As previously discussed, the metric we used expresses the probability of males outliving females among randomly paired individuals and assumes independence between populations. However, males and females in a population are generally not random pairs but often couples, whose health and mortality patterns have been found to be positively correlated due a strong effect of social ties on health and longevity [26]. Coupled individuals also influence each other's health [27], and this is particularly true for males, who benefit more than females from being in a stable relationship [28]. The HMD data and the φ statistic do not permit the estimation of the probability of males outliving females for not randomly paired individuals.

The outsurvival statistics can be helpful in some social and political debates [18]. Governments develop public health programs to reduce lifespan inequalities at different levels (e,g, socioeconomic status, race, sex, etc.). It would be wrong to say that half of the population is disadvantaged by sex differences in lifespan. If $\varphi = 0.4$, as it is in many modern populations (mean of 0.37 in 2017), then 40% of males lives longer than females. It could then be argued that, if a policy aiming at reducing inequalities between sexes target the male population as a whole, some of the efforts and investments would be misallocated. Such policy could be more efficient if, for example, $\varphi = 0.1$. Inequalities in lifespan between sexes are attributable to some part of (each) population and not to the whole. Indeed, M Luy and K Gast [9] found that male excess mortality is mainly caused by some specific subpopulations of males with particularly high mortality. Being able to better identify the

characteristics of the short-lived men could help tackle more efficiently male-female inequality. The statistic φ is applicable to other (sub-)populations or to a combination of them. For example, differences in life expectancy between females and males tend to decrease with increasing number of education years [29]. Knowing the probability of males outliving females in a specific educational group can contribute to developing more informed policies aimed at reducing specific inequalities in mortality between sexes.

Lifespan inequality between sexes emerges from a complex combination of factors which affects individuals unequally, even from the same sex. Comparing life expectancy between females and males provide a simplistic view of lifespan inequalities between sexes. Using measures of overlap between two distributions, as the outsurvival statistic, complement these summary measure and offer a more comprehensive understanding of inequalities.

Acknowledgement

The authors are grateful to James W. Vaupel for his useful comments and edits.

Funding statement

The research and publication of this paper were supported by the AXA Research Fund, through the funding for the "AXA Chair in Longevity Research".

Author contributions

MPBB designed and conceptualized of the study. MPBB and JAA produced the results and did the analysis. MPBB, JAA, IK and VZ contributed to the interpretation of the results; and the drafting, revision and approval of the manuscript.

Data statement

The data are publicly available at http://www.mortality.org . The R code to reproduce the results will be made public upon acceptance of the paper.

Patient and Public Involvement

No patient involved.

Patient consent for publication

Not applicable

Conflicts of interests

None declare

Ethical approval

Not applicable.

Word count

3215

REFERENCES

1. Crimmins EM, Shim H, Zhang YS, Kim JK: **Differences between men and women in mortality and the health dimensions of the morbidity process**. *Clinical chemistry* 2019, **65**(1):135-145.
2. Austad SN, Fischer KE: **Sex differences in lifespan**. *Cell metabolism* 2016, **23**(6):1022-1033.
3. Rogers RG, Everett BG, Saint Onge JM, Krueger PM: **Social, behavioral, and biological factors, and sex differences in mortality**. *Demography* 2010, **47**(3):555-578.
4. Gjonça A, Tomassini C, Toson B, Smallwood S: **Sex differences in mortality, a comparison of the United Kingdom and other developed countries**. *Health Statistics Quarterly* 2005, **26**:6-16.
5. Preston SH, Wang H: **Sex mortality differences in the United States: The role of cohort smoking patterns**. *Demography* 2006, **43**(4):631-646.
6. Oksuzyan A, Juel K, Vaupel JW, Christensen K: **Men: good health and high mortality. Sex differences in health and aging**. *Aging clinical and experimental research* 2008, **20**(2):91-102.
7. Beltran-Sanchez H, Finch CE, Crimmins EM: **Twentieth century surge of excess adult male mortality**. *Proceedings of the National Academy of Sciences* 2015, **112**(29):8993-8998.
8. Barford A, Dorling D, Smith GD, Shaw M: **Life expectancy: women now on top everywhere**. *BMJ* 2006, **332**(7545):808.

9. Luy M, Gast K: **Do women live longer or do men die earlier? Reflections on the causes of sex differences in life expectancy**. *Gerontology* 2014, **60**(2):143-153.
10. Zarulli V, Kashnitsky I, Vaupel JW: **Death rates at specific life stages mold the sex gap in life expectancy**. *Proceedings of the National Academy of Sciences* 2021, **118**(20):e2010588118.
11. Zarulli V, Jones JAB, Oksuzyan A, Lindahl-Jacobsen R, Christensen K, Vaupel JW: **Women live longer than men even during severe famines and epidemics**. *Proceedings of the National Academy of Sciences* 2018, **115**(4):E832-E840.
12. van Raalte AA, Sasson I, Martikainen P: **The case for monitoring life-span inequality**. *Science* 2018, **362**(6418):1002-1004.
13. Colchero F, Rau R, Jones OR, Barthold JA, Conde DA, Lenart A, Nemeth L, Scheuerlein A, Schoeley J, Torres C *et al*: **The emergence of longevous populations**. *Proceedings of the National Academy of Sciences* 2016, **113**(48).
14. Permanyer I, Scholl N: **Global trends in lifespan inequality: 1950-2015**. *PloS one* 2019, **14**(5):e0215742.
15. Sasson I: **Trends in life expectancy and lifespan variation by educational attainment: United States, 1990–2010**. *Demography* 2016, **53**(2):269-293.
16. Edwards RD, Tuljapurkar S: **Inequality in life spans and a new perspective on mortality convergence across industrialized countries**. *Population and Development Review* 2005, **31**(4):645-674.
17. Shi J, Aburto JM, Martikainen P, Tarkiainen L, van Raalte AA: **The Stratification Index for Comparing Between-Group Differences in Mortality: Evidence from Finnish Income Groups** In: *SocArXiv.* 2021.
18. Vaupel JW, Bergeron-Boucher M-P, Kashnitsky I: **Outsurvival as a measure of the inequality of lifespans between two populations**. *Demographic Research* 2021, **44**(35):853-864.
19. An Z-W, Huang H-Z, Liu Y: **A discrete stress–strength interference model based on universal generating function**. *Reliability Engineering & System Safety* 2008, **93**(10):1485-1490.
20. EL-Sayed S A, Constantin C: **Probabilistic Simulation Approach to Evaluate the Tooth-Root Strength of Spur Gears with FEM-Based Verification**. *Engineering* 2011, **12**(3):1137-1148.

21. HMD: **Human Mortality Database**. In.: University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany), Available at www.mortality.org (Accessed on July 15) . 2021.
22. Engelman M, Caswell H, Agree EM: **Why do lifespan variability trends for the young and old diverge? A perturbation analysis**. *Demographic Research* 2014, **30**:1367.
23. Glei DA, Horiuchi S: **The narrowing sex differential in life expectancy in high-income populations: effects of differences in the age pattern of mortality**. *Population studies* 2007, **61**(2):141-159.
24. Vaupel JW, Zhang Z, van Raalte AA: **Life expectancy and disparity: an international comparison of life table data**. *BMJ open* 2011, **1**(1):e000128.
25. Aburto JM, Villavicencio F, Basellini U, Kjærgaard S, Vaupel JW: **Dynamics of life expectancy and life span equality**. *Proceedings of the National Academy of Sciences* 2020, **117**(10):5250-5259.
26. Rasulo D, Christensen K, Tomassini C: **The influence of social relations on mortality in later life: A study on elderly Danish twins**. *The Gerontologist* 2005, **45**(5):601-608.
27. Perelli-Harris B, Hoherz S, Addo F, Lappegård T, Evans A, Sassler S, Styrc M: **Do marriage and cohabitation provide benefits to health in mid-life? The role of childhood selection mechanisms and partnership characteristics across countries**. *Population research and policy review* 2018, **37**(5):703-728.
28. Staehelin K, Schindler C, Spoerri A, Stutz EZ, Group SNCS: **Marital status, living arrangement and mortality: does the association vary by gender?** *Journal of Epidemiology and Community Health* 2012, **66**(7):e22.
29. Molla MT, Madans JH, Wagener DK: **Differentials in adult mortality and activity limitation by years of education in the United States at the end of the 1990s**. *Population and Development Review* 2004, **30**(4):625-646.

Figure legends

Figure 1 Four scenarios of interactions between lifespan distributions and corresponding statistics.

Figure 2 Probability of males outliving females since 1850 for five countries and the range for all countries in the HMD in grey.
Source: HMD [21] and authors' own calculations using Equation (2).

Figure 3 Relation between φ and (a) the sex differences in life expectancy and (b) the standard deviation of lifespans for females for HMD period data, with France highlighted (red triangles).
Source: HMD [21] and authors' own calculations.

Figure 4 Relation between φ and (a) the sex differences in life expectancy and (b) the standard deviation for females for HMD period data, conditional to survival to age 50, with France highlighted (red triangles).
Source: HMD [21] and authors' own calculations.

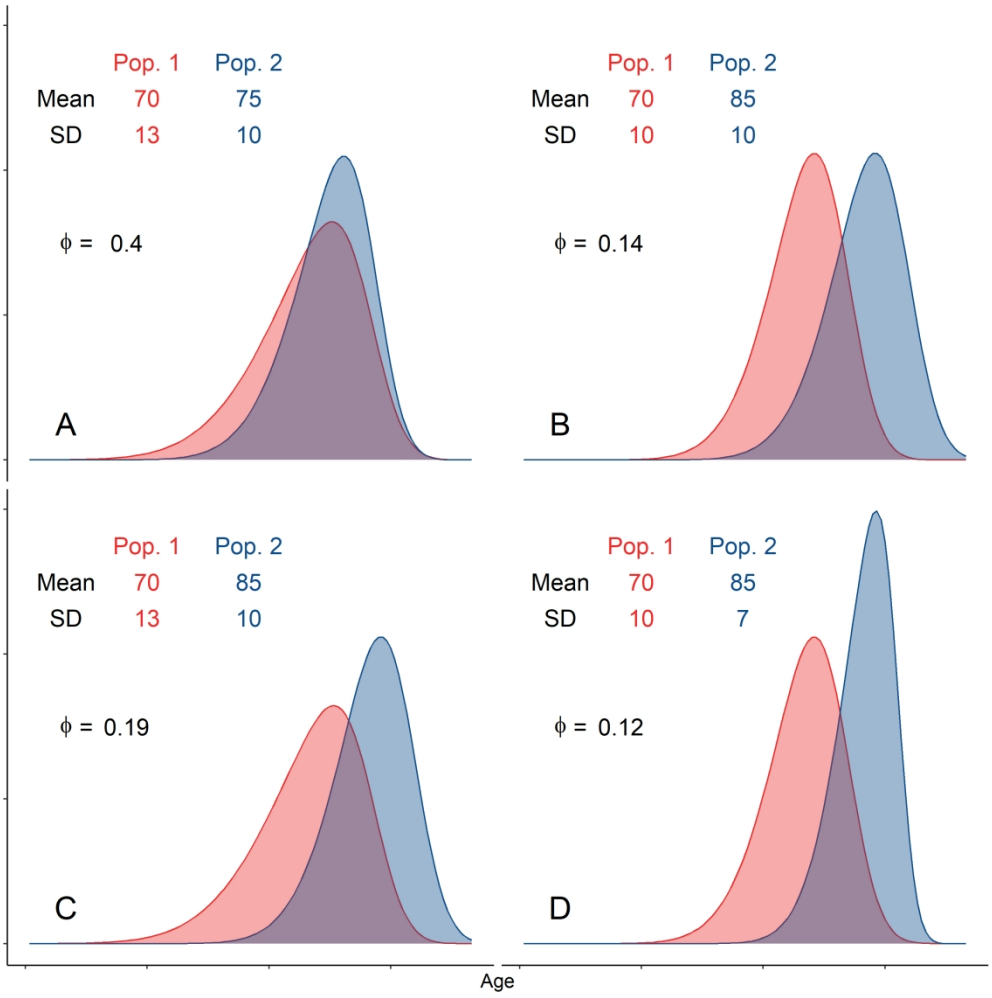

Figure 1 Four scenarios of interactions between lifespan distributions and corresponding statistics.

645x645mm (118 x 118 DPI)

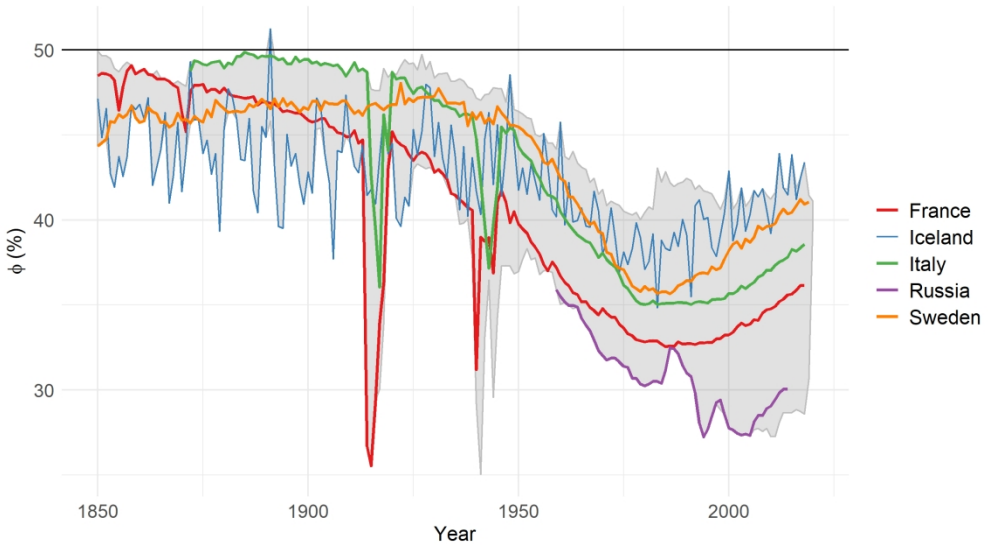

Figure 2 Probability of males outliving females since 1850 for five countries and the range for all countries in the HMD in grey.
Source: HMD (2021) and authors' own calculations using Equation (2).

581x322mm (118 x 118 DPI)

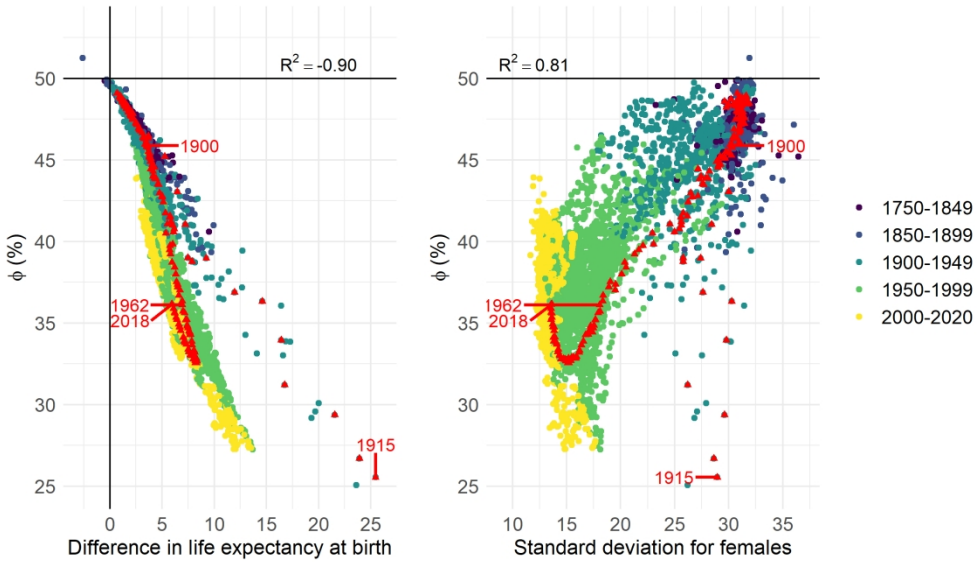

Figure 4 Relation between φ and (a) the sex differences in life expectancy and (b) the standard deviation of lifespans for females for HMD period data since 1751, with France highlighted (red triangles). Source: HMD [23] and authors' own calculations.

1761x1002mm (38 x 38 DPI)

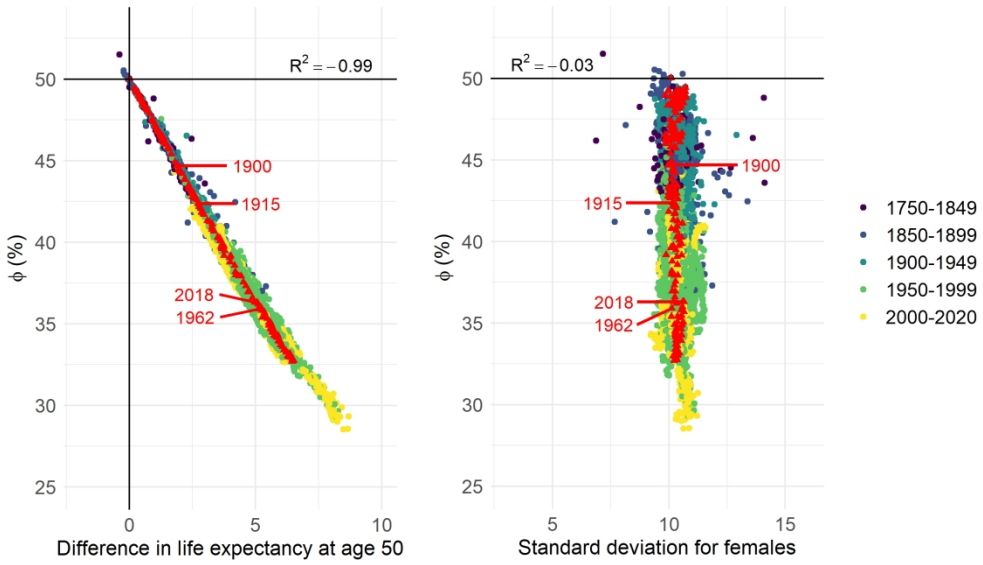

Figure 5 Relation between φ and (a) the sex differences in life expectancy and (b) the standard deviation for females for HMD period data since 1751, conditional to survival to age 50, with France highlighted (red triangles).
Source: HMD [23] and authors' own calculations.

1761x1002mm (38 x 38 DPI)

A. FORMAL RELATION

It can be shown that the outsurvival statistic relates to the joint probability density function of two lifespan distributions, which gives the probability of realizations of two lifespans and thus is related to the overlap of the two distributions. Assume two populations of individuals, with ages at death x and y, respectively. Assume the two populations are independent, meaning that the length of life x does not depend on the length of life y and vice versa. This implies that the joint probability density function, $d_{1,2}(x,y)$, equals the product of the marginal densities so that $d_{1,2}(x,y) = d_1(x)\, d_2(y)$. We are interested in calculating the probability (φ) of individuals in the first population outliving those in the second population. This implies that $0 \leq y < x$ so:

$$\begin{aligned} \varphi \quad &= \int_0^\infty \int_0^x d_{1,2}(x,y)dydx \;\; = \int_0^\infty d_1(x) \int_0^x d_2(y)dydx \\ &= \int_0^\infty d_1(x)D_2(x)dx \\ &= \int_0^\infty d_1(x)[1 - l_2(x)]dx \;\; = 1 - \int_0^\infty d_1(x)l_2(x)dx \\ &= \textstyle\int_0^\infty d_2(x)\ell_1(x)dx. \end{aligned} \tag{A1}$$

Following the same approach, we can find the complement of φ, labeled φ', which is the probability of individuals in the second population to outlive those in the first:

$$\begin{aligned} \varphi' \quad &= \int_0^\infty \int_0^y d_{1,2}(x,y)dxdy \\ &= \int_0^\infty d_2(y) \int_0^y d_1(x)dxdy \\ &= \int_0^\infty d_2(y)D_1(y)dy \\ &= \int_0^\infty d_2(y)[1 - l_1(y)]\, dy \\ &= 1 - \int_0^\infty d_2(y)l_1(y)\, dy. \end{aligned} \tag{A2}$$

From Equations (B1) and (B2) it can be shown that $\varphi + \varphi' = 1$. Thus, φ is also equal to:

$$\begin{aligned} \varphi \quad &= 1 - \varphi' \\ &= 1 - \left[1 - \int_0^\omega d_2(x)l_1(x)\, dx\right] \\ &= \int_0^\omega d_2(x)l_1(x)\, dx. \end{aligned} \tag{A3}$$

B. SIMULATIONS AND DISCREATE APPROXIMATION

We simulated age at death distributions, using the Gompertz model, using various scale (M) and shape (β) parameters (Missov et al., 2015). The distributions were first found using an age width (n)

of 0.0001, after which the data were aggregated within 1-year and 5-years age-groups. The probability that individuals in both population died within the same age-group, $\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2$, was then redistributed between φ and φ' based on two assumptions: equal (equation B1) and proportional redistributions (equation B2). The results are presented in Table B1.

$$\varphi \approx \sum_{x=0}^{\omega} {}_nd_x^1 \; {}_nD_{x-n}^2 + \frac{\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2}{2} \tag{B1}$$

$$\varphi \approx \sum_{x=0}^{\omega} {}_nd_x^1 \; {}_nD_{x-n}^2 + \sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2 \frac{{}_nd_x^1 \; {}_nD_{x-n}^2}{{}_nd_x^1 \; {}_nD_{x-n}^2 + {}_nd_x^2 \; {}_nD_{x-n}^1} \tag{B2}$$

The simulations show that equally redistributing $\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2$ between the two other statistics provide very similar results to the continuous data (*n=0.0001*), especially for the 1-year age-group. More differences are found when aggregating by 5-years age-groups, but the difference in φ between the different age-width remains less than 1 percentage point, when equally redistributing $\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2$.

Table B1. Assumptions to redistribute $\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2$ for different mortality scenarios.

	φ	φ'	$\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2$	Eq. B1	Eq. B2
Gompertz: $M_A = 61, M_B = 65, \beta_A = 0.12, \beta_B = 0.14$					
Continuous	**36.3**	63.7	0.0	-	-
1-year	34.8	62.2	3.0	**36.3**	35.9
5-years	28.2	55.8	15.0	**36.7**	34.3
Gompertz: $M_A = 61, M_B = 70, \beta_A = 0.10, \beta_B = 0.14$					
Continuous	**23.6**	76.4	0.0	-	-
1-year	22.5	75.2	2.3	**23.6**	23.0
5-years	18.5	70.0	11.3	**24.2**	20.9
Gompertz: $M_A = 68, M_B = 70, \beta_A = 0.13, \beta_B = 0.14$					
Continuous	**42.8**	57.2	0.0	-	-
1-year	41.2	55.5	3.3	**42.8**	42.6
5-years	34.9	48.8	16.3	**43.0**	41.7
Gompertz: $M_A = 69, M_B = 70, \beta_A = 0.10, \beta_B = 0.12$					
Continuous	**46.1**	53.9	0.0	-	-
1-year	44.7	52.6	2.7	**46.1**	46.0
5-years	39.4	47.2	13.4	**46.1**	45.5

To further test the model and the redistribution of $\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2$, we simulated 100,000 individual lifespans from an exponential distribution with piece-wise constant rates (Willekens, 2009). We performed this procedure for every population and by sex using as an input empirical death rates retrieved from the HMD (2021). Then we randomly paired males and females and calculated the proportions of males outliving the paired female. Table 3 compares the discrete approach introduced in the main document (eq. B1) and the continuous approach based on simulations. Both approaches provided very similar results.

Table B2. Proportions of males outliving females based on a discrete and continuous approach (simulations).

	% males outliving females		% females outliving males	
	Continuous	Discrete	Continuous	Discrete
Denmark				
1850	46.1	46.4	53.9	53.6
1900	45.8	45.8	54.2	54.2
1950	46.2	46.3	53.8	53.7
2016	40.7	40.6	59.3	59.4
France				
1850	48.6	48.5	51.4	51.5
1915	25.6	25.5	74.4	74.5
1950	39.8	39.8	60.2	60.2
2016	35.9	36.0	64.1	64.0
Japan				
1950	44.1	44.2	55.9	55.8
2016	33.8	33.7	66.2	66.3
Russia				
1960	35.6	35.5	64.4	64.5
2014	30.2	30.0	69.8	70.0

C. DATA

Table C1. Countries/ regions and years with available data in the HMD

Country\region	Years	Country\region	Years
Australia	1921-2018	Japan	1947-2019
Austria	1947-2019	Latvia	1959-2019
Belarus	1959-2018	Lithuania	1959-2019
Belgium	1841-2018	Luxembourg	1960-2019
Bulgaria	1947-2017	Netherlands	1850-2019
Canada	1921-2018	New Zealand	1948-2013
Chile	1992-2017	Norway	1846-2020
Croatia	2001-2019	Poland	1958-2019
Czechia	1950-2019	Portugal	1940-2018
Denmark	1835-2020	Republic of Korea	2003-2018
Estonia	1959-2019	Russia	1959-2014
Finland	1878-2019	Slovakia	1950-2017
France	1816-2018	Slovenia	1983-2017
Germany-East	1956-2017	Spain	1908-2018
Germany-West	1956-2017	Sweden	1751-2019
Greece	1981-2017	Switzerland	1876-2018
Hong Kong	1986-2017	Taiwan	1970-2019
Hungary	1950-2017	UK – England and Wales	1841-2018
Iceland	1838-2018	UK- Scotland	1855-2018
Ireland	1950-2017	UK- Northern Ireland	1922-2018
Israel	1983-2016	USA	1933-2019
Italy	1872-2018	Ukraine	1959-2013

D. COHORT ANALYSIS

Similar relations as those for period data were also found for cohorts (Figure D1). In the HMD, life table for cohorts were only available for 11 countries: Denmark, England and Wales, Finland, France, Iceland, Italy, Netherlands, Norway, Scotland, Sweden and Switzerland. For cohorts with complete mortality history, the proportions of males outliving females varied between 35% and 49%. Only small changes in φ were observed for cohorts born prior to 1870-1890, with φ varying around 46.5%. For the cohorts born afterwards, φ decreases, reaching a mean of 38.4% for the cohort born in 1925, with values varying between 35.3% (Finland) and 40.4% (Scotland).

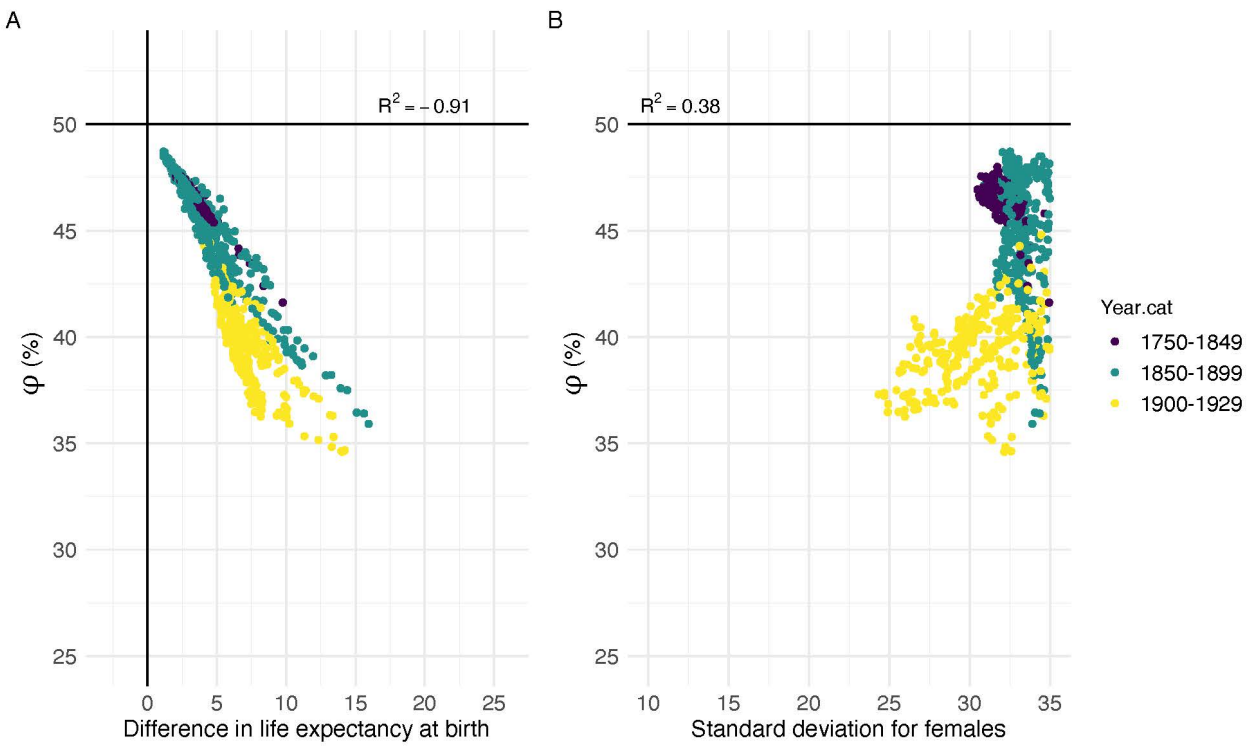

Figure D1 Relation between φ and (a) the sex differences in life expectancy and (b) the standard deviation for females for HMD cohort data.
Source: HMD (2021) and authors' own calculations.

REFERENCES

HMD. (2021). Human Mortality Database. In: University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany), Available at www.mortality.org (Accessed on July 15) .

Missov, T. I., Lenart, A., Nemeth, L., Canudas-Romo, V., & Vaupel, J. W. (2015). The Gompertz force of mortality in terms of the modal age at death. *Demographic Research, 32*, 1031-1048.

Willekens, F. (2009). Continuous-time microsimulation in longitudinal analysis. In A. Harding & A. Zaidi (Eds.), *New frontiers in microsimulation modelling* (pp. 353-376). London: Routledge.

BMJ Open

Probability of males to outlive females, an international comparison from 1751 to 2020

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-059964.R1
Article Type:	Original research
Date Submitted by the Author:	22-Apr-2022
Complete List of Authors:	Bergeron-Boucher, Marie-Pier; Syddansk Universitet, Interdisciplinary Center on Population Dynamics Álvarez, Jesús-Adrian; University of Southern Denmark, Interdisciplinary Centre on Population Dynamics Kashnitsky, Ilya; Syddansk Universitet, Interdisciplinary Centre on Population Dynamics Zarulli, Virginia; Syddansk Universitet, Interdisciplinary Center on Population Dynamics
Primary Subject Heading:	Research methods
Secondary Subject Heading:	Epidemiology, Public health
Keywords:	PUBLIC HEALTH, STATISTICS & RESEARCH METHODS, EPIDEMIOLOGY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Probability of males to outlive females, an international comparison from 1751 to 2020

Marie-Pier Bergeron-Boucher, corresponding author, mpbergeron@sdu.dk
Jesús-Adrián Álvarez, alvarez@sdu.dk
Ilya Kashnitsky, ikashnitsky@sdu.dk
Virginia Zarulli, vzarulli@sdu.dk

Danish Centre for Population Research, Interdisciplinary Center on Population Dynamics, University of Southern Denmark, Odense, Denmark

ABSTRACT

Objective To measure sex differences in lifespan based on the probability of males to outlive females.

Design International comparison of national and regional sex-specific lifetables from the Human Mortality Database and the World Population Prospects.

Setting 199 populations spanning across all continents, between 1751 and 2020.

Primary outcome measure We use the outsurvival statistic (φ) to measure inequality in lifespan between sexes, which is here interpreted as the probability of males to outlive females.

Results In random pairs at age zero of a male and a female, the probability of the male outliving the female varies between 25% and 50% for lifetables in almost all years since 1751 and across almost all populations. We show that φ is negatively correlated with sex differences in life expectancy and positively correlated with the level of lifespan variation. The important reduction of lifespan inequality observed in recent years has made it less likely for a male to outlive a female.

Conclusions Although male life expectancy is generally lower than female life expectancy and male death rates are usually higher at all ages, males have a substantial chance of outliving females. These findings challenge the general impression that "men do not live as long as women" and reveal a more nuanced inequality in lifespans between females and males.

Keywords: Lifespan, Inequality, Sex differences, Outsurvival statistics

Strengths and limitations of this study

- It is the first study using the outsurvival statistic to quantify the probability of males outliving females.
- The outsurvival statistic shows that lifespan inequalities between sexes have more nuances that cannot be captured by comparisons made with classic summary demographic measures (e.g. life expectancy).
- Our results challenge the general impression that men do not live as long as women and provide evidence that shows a large overlap between females' and males' lifespan distribution.
- The outsurvival statistic does not account for dependance between individuals, such as couples whose health and mortality patterns are positively correlated due a strong effect of social ties on health and longevity.

1. INTRODUCTION

The female survival advantage has been observed over time across many human populations and is rooted in a complex combination of biological, environmental and behavioral factors [1-5]. For example, males tend to engage in more risky behaviors, such as smoking and heavy drinking; but estrogen could also be preventive against certain diseases [6]. A study on cloistered population reveals a constant female survival advantage of around 0.2 years. The author attributes the remaining of the sex differences in life expectancy in the general population to differences in lifestyles and socioeconomic burden [7]. However, even among populations where men and women differ less in terms of key lifestyle factors, such as Mormons, sex differences in life expectancy still exist [8]. In 2019, the World sex differences in life expectancy was 4.4 years, with large variation across countries [9]. Females have been found to have longer survival and lower death rates than men at all ages and in most modern populations [2 4 10-13] and so even under extreme mortality conditions [14].

Sex differences in survival are often identified by comparing life expectancy between females and males, which summarizes the average length of life. These differences are often interpreted as "men do not live as long as women". Such an interpretation is simplistic as it does not account for the variation around the means (life expectancies) and potential overlap between female and male lifespan distributions. Despite females having higher life expectancy than males, not all females outlive all males. On the contrary, a sizeable portion of male might live longer than a sizeable portion of females, even if the life expectancy shows a female advantage. This is because the lifespan

distributions of females and males partly overlap, i.e. they share a common range of ages at death. The extent of the overlapping indicates how likely it is for males to outlive females and, ultimately, how sizeable is the portion of males living longer than females.

Lifespan variation, i.e. differences in lifespans within a population, has been receiving an increasing attention in the literature [15]. Various indicators reveal heterogeneity in the length of life, beyond what life expectancy indicates. Studies have compared lifespan variation between two populations, focusing on which populations exhibit more inequalities [16-18]. It has been shown that females systematically experience lower lifespan variation than males [16]. However, it is unclear how this variation around the means leads to potential overlap between two lifespan distribution.

Only a few studies have used measures of overlap or distance to study inequalities between populations. A previous study has investigated how different two lifespan distributions are using the Kullback-Leibler (KL) divergence [19]. The indicator is interpreted as the amount of "effort" needed to transform the male's lifespan distribution into the female's distribution. A disadvantage of this indicator is that it is not symmetric, meaning that the effort needed to transform male's distribution into female's is not the same as the effort needed to transform female's distribution into male's. Stratification indexes, based on how much two lifespan distributions overlap or not-overlap, have also been used to study mortality differences between socioeconomic groups [20]. The larger the overlap, the more likely are the individuals in two populations to survive to the same age. This index is meant to reflect unequal distribution at the societal level, with values varying between 0 (no overlap) and 1 (perfect overlap). A related measure is the outsurvival statistics, which quantifies the probability that an individual from a population with lower life expectancy outlives an individual from another population with higher life expectancy [21]. The main difference with the stratification index is on the interpretation, which focuses on the individuals. If the two populations are males and females, the outsurvival statistic captures the correctness of the assertion that males' lifespans are lower than females' lifespans. If both populations have equal lifespan distribution, the outsurvival statistic is equal to 0.5. Unlike the other two measures, the outsurvival statistic also informs directly on which of the compared populations has an advantage (values above 0.5) or a disadvantage (values below 0.5).

In this article, we use the outsurvival statistic [21] to study lifespan inequalities between females and males. We aim to 1) quantify the probability that males outlive females over time and across populations; and 2) assess the sensitivity of the outsurvival statistic to changes in life expectancy and lifespan variation. We computed the outsurvival probability to study sex differences in mortality in

199 populations over 200 years. Despite sometimes large differences in life expectancy, we show that there are substantial overlaps between males' and females' lifespan distributions.

2. METHOD

2.1 OUTSURVIVAL STATISTIC

Consider two populations with mean and standard deviation (SD) of the age of death (see [22] [23] for more details on the SD calculation) specified in panel A of Figure 1. The first population (in red) has a smaller mean lifespan and larger SD than the second population (in blue). An inference from the means would be that individuals in the first population are worse off than individuals in the second. However, there is an important overlap between the two distributions, with some individuals in the first population outliving some individuals in the second population. The outsurvival probability, φ, captures this dimension by measuring the probability that an individual from a population with high mortality will outlive an individual from a population with low mortality [21]. Let $d_i(x)$, $i = 1,2$, denote the lifespan distribution at age x in two populations. The cumulative distributions are represented by $D_i(x)$, such that such $D_i(x) = \int_0^x d_i(t)dt$ and the survivorship is denoted by $\ell_i(x)$, with $\ell_i(x) = 1 - D_i(x)$. The probability that an individual from the first population (males) will outlive an individual from the second population (females) is [21]:

$$\varphi = \int_0^\infty d_2(x)\ell_1(x)dx. \quad (1)$$

In scenario A of Figure 1, φ is 40%.

[Figure 1]

In the supplementary materials, we show that the outsurvival statistic relates to the joint probability density function of two lifespan distributions, which gives the probability of realizations of two lifespans and thus is related to the overlap of the two distributions.

2.2 RELATION TO LIFE EXPECTANCY AND LIFESPAN VARIATION

Consider the two populations in scenarios B and C of Figure 1. The difference in mean lifespan is the same in both cases, i.e. 15 years. However, in scenario C the first population has a larger SD, which

implies more individuals surviving to older ages, despite greater inequalities, and thus a greater potential to outlive individuals from the second population. Indeed, φ is higher in scenario C (19%) than in scenario B (14%). Now compare scenario B to scenario D. This time, the second population in D has a smaller SD, with fewer individuals dying at younger ages, making it less likely for individuals in the first population to outlive them. This reduces φ to 12%. Thus, for the same difference in life expectancy, larger lifespan variation in both populations generally results in larger φ . The comparison of scenarios A and C also shows that small differences in life expectancy lead to larger value of φ.

Equation (1) is not new and relates to the Mann-Whitney U test, the probability of superiority and to the expected failure probability in a stress-strength interference (SSI) model. The later assesses the probability that the stress (population 1) exceeds the strength (population 2) of a material [24]. If the distributions of both populations follow a Normal distribution with mean μ_i and standard deviation s_i, the probability of failure is P(Z) with $Z = -\frac{\mu_2 - \mu_1}{\sqrt{s_1^2 + s_2^2}}$ [25]. This relation formalizes what is illustrated in section in Figure 1: φ is sensitive to the difference in the means and to the level of variation in both distributions, with smaller mean differences (numerator) and larger variance (denominator) leading to larger P(Z). However, lifespan distributions are not normally distributed, and additional moments could also affect the value of φ. To better understand this relation, we analyzed the correlation between φ and life expectancy as well as between φ and lifespan variation.

2.3 DISCRETE APPROXIMATION

Similar equivalences to equation (1) can be develop in a discrete time setting. Let ${}_nd_x^i$ be the life table deaths between age x and $x+n$ in population i and ${}_nl_x^i$ the survival probability to age x. For a given age-group width of n, the probability of individuals in the first population outliving those in the second population can be found by:

$$\varphi \approx \sum_{x=0}^{\omega} {}_nd_{x-n}^2 \, {}_nl_x^1 + \bar{d} \qquad (2)$$

with $\bar{d} = \frac{\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2}{2}$ and $\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2$ being the probability that individuals in both populations died in the same age-group. The latter statistic is sensitive to the width of the age-groups such that smaller

age-groups result into smaller values, with $\lim_{n \to 0} \sum_{x=0}^{\omega} {}_n d_x^1 \, {}_n d_x^2 = 0$. In the supplementary materials, we compared the discrete and continuous approaches and find that both approaches yield comparable results.

Equations (1) and (2) are equivalent to matching random individuals from each population and calculating the proportions of individuals from the first population outliving the paired individual from the second. We performed such analysis via simulations of individuals from a specific lifespan distribution and estimated the corresponding statistics (see supplementary materials). Equivalent results were found.

PATIENT AND PUBLIC INVOLVEMENT

No patient involved.

3. DATA

The method is applied to three demographic datasets. First, we used life tables by sex for all available countries and years from the Human Mortality Database [26]. The HMD is freely available, provides comparable long time-series for 41 countries with high quality data. Data are provided by single-year age-groups. We used subnational data for Germany with separate analysis for East and West Germany; and the United Kingdom for England-Wales (total population), Scotland and Northern Ireland, totalizing 44 studied populations. The earliest year with available data was 1751 (for Sweden) and the latest was 2020. Information about the available populations and years are provided in the supplementary materials. We compared females and males' life tables in each country/region.

Second, we used abridged life tables from the World Population Prospects 2019 Revision (WPP) [27]. This dataset is also freely available and provides sex-specific life tables for 199 countries by 5-year age-groups and 5-year period from 1950-1954 to 2015-2019. This database covers the whole World, but the data quality varies greatly between countries [28]. The HMD and WPP data are used to compare the outsurvival statistic over time and across multiple populations.

Finally, we computed the outsurvival statistic for subpopulations of females and males using US data in 2015-2019. We compared the probability of males to outlive females by education level and marital status to assess if the sex differences emerge from specific subpopulations. We calculated sex-specific life tables by education level and marital status using death counts from the Multiple Cause of Death dataset (MCDD) from the National Vital Statistics System of the National Center for Health Statistics [29] and population counts from the American Community Service (ACS) from the

United States Census Bureau [30]. The MCDD provides deaths counts by single-year age groups, sex, marital status, and education level. The ACS provides data by similar variables and single-year age-groups until age 96. However, it is worth noting that the ACS data exhibit an important age-heaping at age 95. We therefore ungrouped the population counts from age 90+ using the Penalized Composite Link Model (PCLM) model [31] to obtain the population counts from age 90 to 110 by single-year of age.

4 RESULTS

4.1 HISTORICAL VALUES AND TRENDS in φ

Figure 2a shows the outsurvival probability of males over females (φ) since 1850 for all HMD countries and Figure 2b for all WWP countries since 1950-1955. The probability of males outliving females has, at all points in time and across all populations, varied between 25% and 50%, with only few exceptions with values above 50%: Iceland in 1891; Jordan in 1950-1954; Iran in 1950-1964, Iraq in 1960-1969; before 1985 in Bangladesh, India and the Maldives; and between 1995 and 2010 in Bhutan.

[Figure 2]

For the HMD countries, φ was slowly decreasing before the First World War, on average from 47.3% in 1850 to 46.0% in 1913. Afterwards, φ declined faster. In 1930, the mean φ across populations was 45.4%, ranging from 42.8% (France) to 48.4% (Netherlands). By 1985, the mean φ was 35.3%, ranging from 31.2% (Russia) to 42.8% (Israel). The value of φ started increasing around the 1980s for some countries, but continued to decrease in others until the 2000s, especially in Eastern European countries. The mean in 2017 was 37.1%, with values varying between 28.7% (Belarus) and 42.5% (Iceland).

For the WWP countries, we observed a decrease in φ in all regions since 1950, except in Europe, Northern America and Oceania increasing from the 1980s, as also shown in the analysis of the HMD data. In 1950-1955, φ was 46.1% on average worldwide, with values ranging between 35.3% (in Kazakhstan) and 52.6% (in Iran). By 2015-2019, φ declined to 41.2% with values ranging between 28.8% (in Belarus) and 49.9% (in Bhutan). Figure 3 shows φ across the World in different time periods. In recent years, the outsurvival of females was particularly low in Eastern Europe and Northeast Asia and was particularly high in Southern Asia and in Western and Middle Africa. Males

from many Southern Asian countries had an especially high chance to outlive females, with φ above 50% before 1970.

[Figure 3]

4.2 RELATION TO LIFE EXPECTANCY AND LIFESPAN INEQUALITY

Figure 4 shows that φ is negatively correlated with the differences in life expectancy and positively correlated with females' SD (similar results were found when males' SD were used, due to the strong correlation between females and males' SDs). Figure 4 is based on the HMD data, but the same relation is found when using the WPP data (see supplementary materials). This relation empirically demonstrates the formal relation in section 2.2. The correlation between φ and the standard deviation is weaker in recent years, due to reduction in sex differences in life expectancy, which is also driving changes in φ . Even though both life expectancy and lifespan variation affect φ, the statistic appears more sensitive to the differences in life expectancy than to the level of lifespan variation. We also found similar results for cohort data (see supplementary materials).

[Figure 4]

The same value for φ can be found for different combinations of sex differences in life expectancy and levels of lifespan variation. For example, the same φ of 36.1% was found in France in 1962 and in 2018 (Figure 4). However, the sex difference in life expectancy was 6.9 in 1962 and 5.9 in 2018 and the standard deviation for females was 18.1 in 1962 and 13.6 in 2018.

Figure 5 shows the same relations as shown in Figure 4 but for survivors to age 50. Lifespan variation at age 50 has stayed roughly constant over time [32] and comparing φ from this age can help assessing the sensitivity of the measure to changes in lifespan variation. The relation between φ and differences in life expectancy is stronger and more linear from age 50 (correlation coefficient of -0.99) than when using the full age-range, increasing predictive ability. For example, for a difference in life expectancy at age 50 of 3 years, males have around 42% probability to outlive females. Note that φ in France was 35.9% in 1962 and 36.3% in 2018.

[Figure5]

Similar to the distribution from birth, the probability of males outliving females from age 50 has, in almost all periods and populations, varied between 28% and 50%, with only few exceptions. In recent years, the φ statistics from birth and from age 50 are similar.

4.3 SEX DIFFERENCES BY EDUCATION AND MARITAL STATUS

Tables 1 and 2 show the φ statistic for some subpopulations of males and females in the United States. For the period 2015-2019, the probability of males to outlive females was 40% in the total US population. However, this statistic varies depending on marital status and education level, being higher among the subpopulations with a beneficial characteristic: the probability of males to outlive females was 39% for the married individuals and 37% for the unmarried (Table 1); 43% for individuals with a university degree and 39% for those without a high school diploma (Table 2).

Table 1. Outsurvival statistics by sex and marital status in the United States, 2015-2019

		Female	
		Married	Unmarried
Male	Married	0.39	0.52
	Unmarried	0.26	0.37

Source: MCDD [29], ACS [30] and authors' own calculations using Equation (2).

Table 2. Outsurvival statistics by sex and education level in the United States, 2015-2019

		Female		
		University Degree	High School Diploma	No High School Diploma
Male	University Degree	0.43	0.51	0.53
	High School Diploma	0.32	0.39	0.42
	No High School Diploma	0.30	0.37	0.39

Source: MCDD [29], ACS [30] and authors' own calculations using Equation (2).

Furthermore, these results highlight that males with beneficial characteristics (being married and having a university degree) have an advantage over women with detrimental characteristics (being unmarried and having a high school diploma or less).

5. DISCUSSION

Our study reveals a nuanced inequality in lifespan between females and males, with between one and two men out of four outliving a randomly paired woman in almost all points in time across 199 populations. These results complement the picture given by the comparisons based on life expectancy, which is a summary measure with no information on variation. A blind interpretations of life expectancy differences can sometimes lead to a distorted perception of the actual inequalities. Not all females outlive males. For example, a sex difference in life expectancy at birth of 10 years can come with a probability of males outliving females as high as 40%, indicating that at 40% of males have a longer lifespan than that of a randomly paired female. Therefore, not all males have a disadvantage of 10 years, something that is overlook by solely making comparisons of life expectancy. An even higher proportion of males outliving females are found among advantaged groups (married and with university degree). Our findings challenge the general impression that men do not live as long as women.

Trends over time in φ are consistent with the reversed trends in sex differences in life expectancy [33]: in developed countries, the probability of males outliving females decreased until the 1970s, after which it gradually increased in all populations. Studies showed that the increase in sex differences in mortality emerged in cohorts born after 1880 [10 34], which is consistent with our analysis of φ (see supplementary materials). The increase and decrease in sex differences in life expectancy was mainly attributed to the smoking epidemic and other behavioral differences between sexes [7 13 35].

The φ values are generally higher in low- and middle-income countries. However, this should not be interpreted as a sign of greater gender equality in survival. Southern Asian countries had very high φ values, above 50% in the 1950s and 1960s. Studies for India showed that mortality below age five was higher for females than males and remained higher for females in recent years [36 37]. However, females had a growing mortality advantage above age 15 since the 1980s "balancing out" the disadvantage at younger ages. The reasons for the higher φ and decreasing trends in developing regions vary across countries. It is outside the scope of this study to provide a detailed explanation for the trends in each country.

As previously discussed, the φ metric expresses the probability of males to outlive females among randomly paired individuals, assuming independence between populations. However, males and females in a population are generally not random pairs but often couples, whose health and mortality have been found to be positively correlated due a strong effect of social ties on health and longevity [38]. Coupled individuals also influence each other's health [39], and this is particularly true for

males, who benefit more than females from being in a stable relationship [40]. The datasets used for the analysis do not permit the estimation of the probability of males outliving females for not randomly paired individuals. However, the outsurvival statistic relates to the probability of the husbands to outlive their wives, and even though such measure accounts for the difference in age between husband and wife, it has been shown to generally be between 30% and 40% [41-43], values that are quite close to φ.

Other measures of overlap and distance between distributions could have been used. In the supplementary materials, we compare the outsurvival statistic with a stratification index used by Shi and colleagues [20] and the Kullback-Leibler divergence. We found that all three indicators are strongly correlated and using one or the other would not have change the general conclusions from this article. However, unlike the other indicator, φ directly informs when males live longer than females, which we found in a few instances.

The outsurvival statistic can be informative for public health interventions [21]. Governments develop public health programs to reduce lifespan inequalities at different levels (e,g, socioeconomic status, race, sex, etc.). It would be misleading to say that half of the population is disadvantaged by sex differences in lifespan. The inequalities are more nuanced. If 40% of males live longer than females, it could be argued that, if a policy aiming at reducing inequalities between sexes targets the full male population, some of the efforts and investments would be misallocated. Such policy could be more efficient if φ approaches 0, indicating that sex would explain a large part of the lifespan inequalities within the population. Whereas a φ closer to 0.5 indicates that other characteristics (e.g. socioeconomic and marital statuses) are involved in creating inequalities. We showed that some subpopulations of males have a high probability (above 50%) to outlive some subpopulations of females. Males who are married or have a university degree tend to outlive females who are unmarried or do not have a high school diploma. Inequalities in lifespan between sexes are attributable to some individuals within each population and not to the whole population. Indeed, Luy and Gast [12] found that male excess mortality is mainly caused by some specific subpopulations of males with particularly high mortality. Being able to better identify the characteristics of the short-lived men could help tackle more efficiently male-female inequality.

Another important result of our analysis is that the smaller the standard deviation in the age at death, the smaller the φ. The reduction of lifespan inequality observed over time has then made it less likely for males to outlive females. This is partly explained by the fact that lifespan variation reduction has been driven by mortality declines at younger ages [44]. When looking at the lifespan distribution

(as in Figure 1, scenario D), survival improvements at younger ages narrowed the left tails of the distribution for both sexes. By reducing the left tail of female distribution, without increasing the right tail of the male distribution, the overlapping area is reduced. In other words, the number of females with shorter lifespan, easier to outlive, decreased over time. Indeed, it has been shown that mortality declined at a faster pace for females than males below age 50, especially in the first half of the 20th century [45 46]. This finding implies that it requires more efforts today than in the past to reduce these inequalities. While inequalities were mainly attributable to infant and child mortality, they are today increasingly attributable to older and more broad age-groups. Men are more prone to accidents and homicides in their 20s and 30s than females and they tend to smoke and drink more leading to higher cancer prevalence and deaths in their 60s. At the same time, women benefited from reduced maternal mortality, and recorded faster mortality decline at older ages. Efforts in reducing lifespan inequalities must thus target diverse factors, causes and ages [13 45 47].

A decrease of φ might indicate a discrepancy in the causes of death that affect males and females. External mortality due to accidents and suicide has become more relevant in shaping sex differences in survival in recent years in high income populations [12]. Another example is observed in Latin American populations, where homicides and violent deaths have an increased burden among males in comparison to females since the 1990s [48 49]. In Mexico, for example, the increase in homicide mortality, especially among men between 20 and 40, contributed to increase the gap in mortality between females and males [50]. This phenomenon is reflected in the decrease over time in the overlapping of lifespan distributions, directly informing healthcare systems of emerging inequalities.

However, one might ask if a wider overlapping is necessarily better for health care systems? On the one hand, a larger overlapping means less inequality between sexes, but on its own it does not ensure that there is more “health justice”. For example, if the overlapping areas are large, this still shows a situation of great uncertainty in lifespan for both groups. One health evaluator actor could even prefer a situation where there is a small gap between groups but less inequality within the groups. In the case of sex differences, there might always be a between-group differences due to biological factors [2 51], but more health equity could be reached by reducing within-group inequalities. We argue that the outsurvival statistic is a new tool to evaluate health inequalities between groups within a population, by uncovering underlying dynamics that are otherwise hidden when looking only at conventional indicators. Therefore, it can inform health care systems of the subsequent directions to reach the preferred goal.

6. CONCLUSION

Comparing life expectancy between females and males provide a simplistic view of lifespan inequalities between sexes. Using measures of overlap between two distributions of lifespans complement these summary measure and offer a more comprehensive understanding of inequalities.

Acknowledgements

The authors are grateful to James W. Vaupel, Jim Oeppen and the two reviewers for their useful comments and edits.

Funding statement

The research and publication of this paper were supported by the AXA Research Fund, through the funding for the "AXA Chair in Longevity Research".

Author contributions

MPBB designed and conceptualized of the study. MPBB, JAA and IK produced the results and did the analysis. MPBB, JAA, IK and VZ contributed to the interpretation of the results; and the drafting, revision and approval of the manuscript.

Data statement

The data are publicly available at http://www.mortality.org and https://population.un.org/wpp/. The R code to reproduce the results will be made public upon acceptance of the paper.

Patient consent for publication

Not applicable

Conflicts of interests

None declare

Ethical approval

Not applicable.

Word count

4188

REFERENCES

1. Crimmins EM, Shim H, Zhang YS, et al. Differences between men and women in mortality and the health dimensions of the morbidity process. *Clinical chemistry* 2019;65(1):135-45.
2. Austad SN, Fischer KE. Sex differences in lifespan. *Cell metabolism* 2016;23(6):1022-33.
3. Rogers RG, Everett BG, Saint Onge JM, et al. Social, behavioral, and biological factors, and sex differences in mortality. *Demography* 2010;47(3):555-78.
4. Gjonça A, Tomassini C, Toson B, et al. Sex differences in mortality, a comparison of the United Kingdom and other developed countries. *Health Statistics Quarterly* 2005;26:6-16.
5. Preston SH, Wang H. Sex mortality differences in the United States: The role of cohort smoking patterns. *Demography* 2006;43(4):631-46.
6. Oksuzyan A, Juel K, Vaupel JW, et al. Men: good health and high mortality. Sex differences in health and aging. *Aging clinical and experimental research* 2008;20(2):91-102.
7. Luy M. Causes of male excess mortality: insights from cloistered populations. *Population and Development Review* 2003;29(4):647-76.
8. Lindahl-Jacobsen R, Hanson HA, Oksuzyan A, et al. The male–female health-survival paradox and sex differences in cohort life expectancy in Utah, Denmark, and Sweden 1850–1910. *Annals of epidemiology* 2013;23(4):161-66.
9. Organization WH. World health statistics overview 2019: monitoring health for the SDGs, sustainable development goals: World Health Organization, 2019.
10. Beltran-Sanchez H, Finch CE, Crimmins EM. Twentieth century surge of excess adult male mortality. *Proceedings of the National Academy of Sciences* 2015;112(29):8993-8. doi: 10.1073/pnas.1421942112 [published Online First: 2015/07/08]
11. Barford A, Dorling D, Smith GD, et al. Life expectancy: women now on top everywhere. *BMJ* 2006;332(7545):808.
12. Luy M, Gast K. Do women live longer or do men die earlier? Reflections on the causes of sex differences in life expectancy. *Gerontology* 2014;60(2):143-53.
13. Zarulli V, Kashnitsky I, Vaupel JW. Death rates at specific life stages mold the sex gap in life expectancy. *Proceedings of the National Academy of Sciences* 2021;118(20):e2010588118. doi: 10.1073/pnas.2010588118
14. Zarulli V, Jones JAB, Oksuzyan A, et al. Women live longer than men even during severe famines and epidemics. *Proceedings of the National Academy of Sciences* 2018;115(4):E832-E40.
15. van Raalte AA, Sasson I, Martikainen P. The case for monitoring life-span inequality. *Science* 2018;362(6418):1002-04.
16. Colchero F, Rau R, Jones OR, et al. The emergence of longevous populations. *Proceedings of the National Academy of Sciences* 2016;113(48) doi: 10.1073/pnas.1612191113
17. Permanyer I, Scholl N. Global trends in lifespan inequality: 1950-2015. *PloS one* 2019;14(5):e0215742.
18. Sasson I. Trends in life expectancy and lifespan variation by educational attainment: United States, 1990–2010. *Demography* 2016;53(2):269-93.
19. Edwards RD, Tuljapurkar S. Inequality in life spans and a new perspective on mortality convergence across industrialized countries. *Population and Development Review* 2005;31(4):645-74.

20. Shi J, Aburto JM, Martikainen P, et al. A distributional approach to measuring lifespan stratification. *Population Studies* 2022;(Accepted)
21. Vaupel JW, Bergeron-Boucher M-P, Kashnitsky I. Outsurvival as a measure of the inequality of lifespans between two populations. *Demographic Research* 2021;44(35):853-64. doi: 10.4054/DemRes.2021.44.35
22. Ouellette N, Bourbeau R. Changes in the age-at-death distribution in four low mortality countries: A nonparametric approach. *Demographic Research* 2011;25:595-628.
23. Wilmoth JR, Horiuchi S. Rectangularization revisited: Variability of age at death within human populations. *Demography* 1999;36(4):475-95.
24. An Z-W, Huang H-Z, Liu Y. A discrete stress–strength interference model based on universal generating function. *Reliability Engineering & System Safety* 2008;93(10):1485-90.
25. EL-Sayed S A, Constantin C. Probabilistic Simulation Approach to Evaluate the Tooth-Root Strength of Spur Gears with FEM-Based Verification. *Engineering* 2011;12(3):1137-48.
[dataset] 26. HMD. Human Mortality Database: University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany), Available at www.mortality.org (Accessed on July 15, 2021) . 2021.
[dataset] 27. United Nations. World Population Prospects 2019, Online Edition. Rev. 1.: Department of Economic and Social Affairs, Population Division, Available at https://population.un.org/wpp/Download/Standard/Mortality/ (Accessed on May 21, 2022), 2019.
28. United Nations, Economic Do, Affairs S. World population prospects 2019: methodology of the United nations population estimates and projections. 2019
[dataset] 29. NCHS. National Center for Health Statistics, Multiple Cause of Death File 2015-2019: National Vital Statistics System, Washington D.C., United States, Available at https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm#Mortality_Multiple (Accessed on January 13, 2022) 2022.
[dataset] 30. U.S. Census Bureau. 2015-2019 American Community Survey 5-year estimates - Public Use Microdata Sample: Available at https://data.census.gov/mdat/#/search?ds=ACSPUMS5Y2019 (Accessed on Janúary 13, 2022), 2022.
31. Rizzi S, Gampe J, Eilers PH. Efficient estimation of smooth distributions from coarsely grouped data. *American Journal of Epidemiology* 2015;182(2):138-47.
32. Engelman M, Caswell H, Agree EM. Why do lifespan variability trends for the young and old diverge? A perturbation analysis. *Demographic Research* 2014;30:1367.
33. Glei DA, Horiuchi S. The narrowing sex differential in life expectancy in high-income populations: effects of differences in the age pattern of mortality. *Population studies* 2007;61(2):141-59.
34. Rigby JE, Dorling D. Mortality in relation to sex in the affluent world. *Journal of Epidemiology & Community Health* 2007;61(2):159-64.
35. Wensink M, Alvarez J-A, Rizzi S, et al. Progression of the smoking epidemic in high-income regions and its effects on male-female survival differences: a cohort-by-age analysis of 17 countries. *BMC Public Health* 2020;20(1):1-8.
36. Borah G. Gender gap in life expectancy in India and role of age groups: A comparison between before and after male–female life expectancy at birth crossover. *PloS one* 2021;16(12):e0260657.
37. Canudas-Romo V, Saikia N. Gender Gap in Life Expectancy in India, 1970-2006: Institute of Economic Growth, University of Delhi 2013.

38. Rasulo D, Christensen K, Tomassini C. The influence of social relations on mortality in later life: A study on elderly Danish twins. *The Gerontologist* 2005;45(5):601-08.
39. Perelli-Harris B, Hoherz S, Addo F, et al. Do marriage and cohabitation provide benefits to health in mid-life? The role of childhood selection mechanisms and partnership characteristics across countries. *Population research and policy review* 2018;37(5):703-28.
40. Staehelin K, Schindler C, Spoerri A, et al. Marital status, living arrangement and mortality: does the association vary by gender? *Journal of Epidemiology and Community Health* 2012;66(7):e22.
41. Compton J, Pollak RA. The life expectancy of older couples and surviving spouses. *PloS one* 2021;16(5):e0250564.
42. Goldman N, Lord G. Sex differences in life cycle measures of widowhood. *Demography* 1983;20(2):177-95. doi: 10.2307/2061234
43. Statistical measures in the marital life-cycles of men and women. Population; 1959. INST NAT ETUD DEMOGRAPHIQUES 27 RUE DU COMMANDEUR, 75675 PARIS CEDEX 14, FRANCE.
44. Aburto JM, Villavicencio F, Basellini U, et al. Dynamics of life expectancy and life span equality. *Proceedings of the National Academy of Sciences* 2020;117(10):5250-59.
45. Cui Q, Canudas-Romo V, Booth H. The Mechanism Underlying Change in the Sex Gap in Life Expectancy at Birth: An Extended Decomposition. *Demography* 2019;56(6):2307-21.
46. Andreev K. Sex differentials in survival in the Canadian population, 1921–1997: a descriptive analysis with focus on age-specific structure. *Demographic Research* 2000;3
47. Bergeron-Boucher M-P, Aburto JM, van Raalte A. Diversification in causes of death in low-mortality countries: emerging patterns and implications. *BMJ global health* 2020;5(7):e002414.
48. Alvarez J-A, Aburto JM, Canudas-Romo V. Latin American convergence and divergence towards the mortality profiles of developed countries. *Population studies* 2020;74(1):75-92.
49. García J, Aburto JM. The impact of violence on Venezuelan life expectancy and lifespan inequality. *International journal of epidemiology* 2019;48(5):1593-601.
50. Aburto JM, Beltrán-Sánchez H, García-Guerrero VM, et al. Homicides in Mexico reversed life expectancy gains for men and slowed them for women, 2000–10. *Health Affairs* 2016;35(1):88-95.
51. Lemaître J-F, Ronget V, Tidière M, et al. Sex differences in adult lifespan and aging rates of mortality across wild mammals. *Proceedings of the National Academy of Sciences* 2020;117(15):8546-53.

Figure legends

Figure 1 Four scenarios of interactions between lifespan distributions and corresponding statistics.

Figure 2 Probability of males outliving females (A) since 1850 for five countries and the range for all countries in the HMD in grey and (B) since 1950-1955 by World regions and the range for all countries in the WPP in grey.

Source: HMD [26], WPP [27] and authors' own calculations using Equation (2).

Figure 3 Probability of males outsurviving females across the World, 1950-1954 to 2015-2019.
Source: WPP [27] and authors' own calculations using Equation (2).

Figure 4 Relation between φ and (a) the sex differences in life expectancy and (b) the standard deviation of lifespans for females for HMD period data since 1751, with France highlighted (red triangles).
Source: HMD [26] and authors' own calculations.

Figure 5 Relation between φ and (a) the sex differences in life expectancy and (b) the standard deviation for females for HMD period data since 1751, conditional to survival to age 50, with France highlighted (red triangles).
Source: HMD [26] and authors' own calculations.

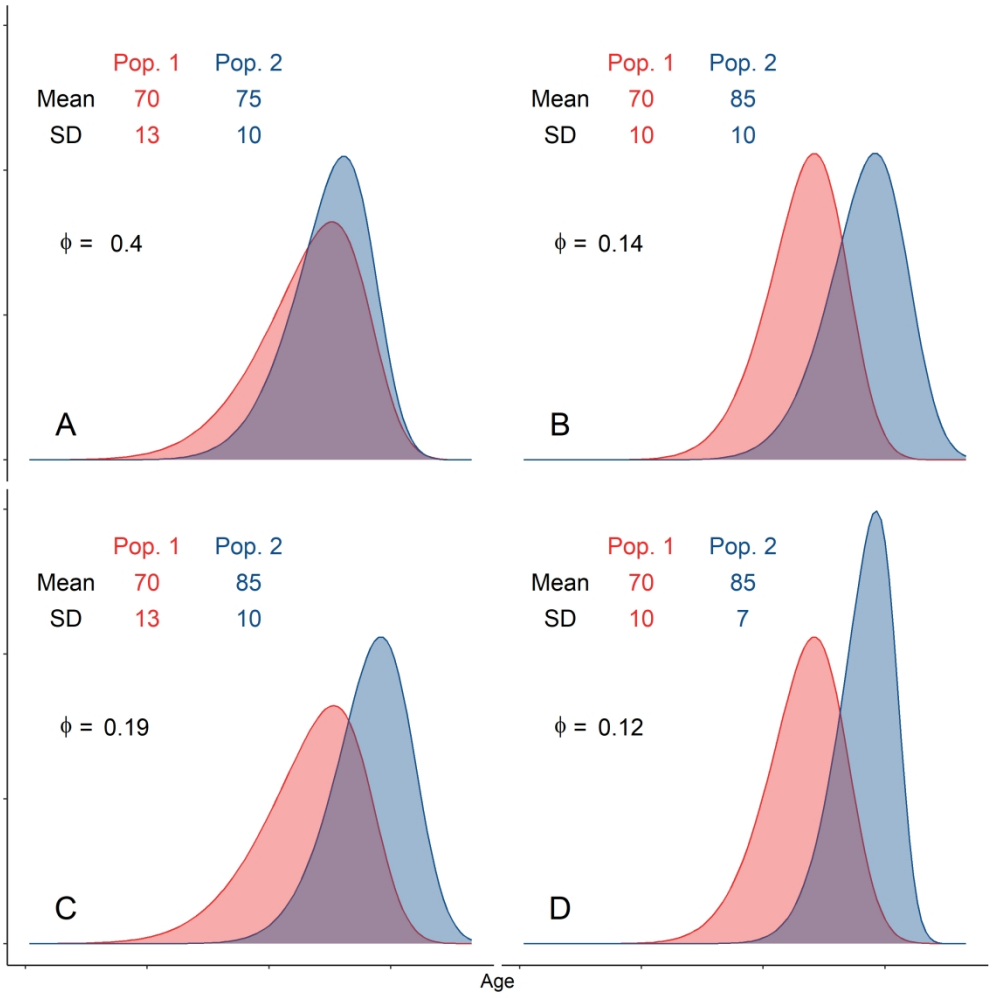

Figure 1 Four scenarios of interactions between lifespan distributions and corresponding statistics.

645x645mm (118 x 118 DPI)

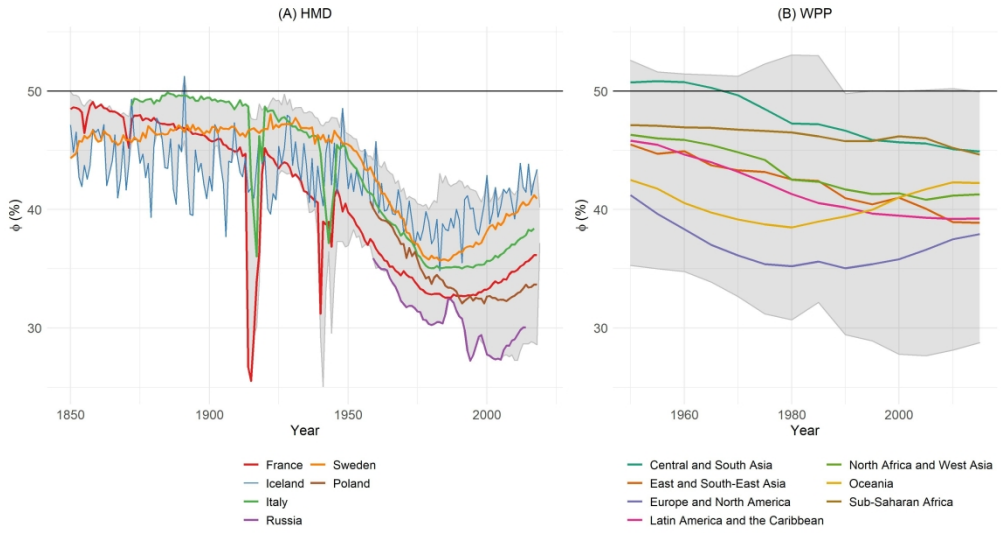

Figure 2 Probability of males outliving females (A) since 1850 for five countries and the range for all countries in the HMD in grey and (B) since 1950-1955 by World regions and the range for all countries in the WPP in grey.
Source: HMD [23], WPP [24] and authors' own calculations using Equation (2).

330x177mm (300 x 300 DPI)

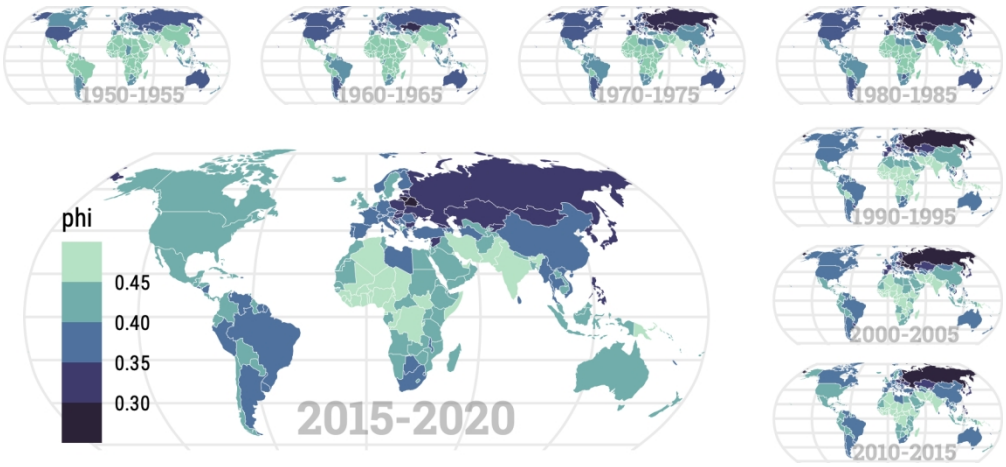

Figure 3 Probability of males outsurviving females across the World, 1950-1954 to 2015-2019. Source: WPP [24] and authors' own calculations using Equation (2).

151x69mm (300 x 300 DPI)

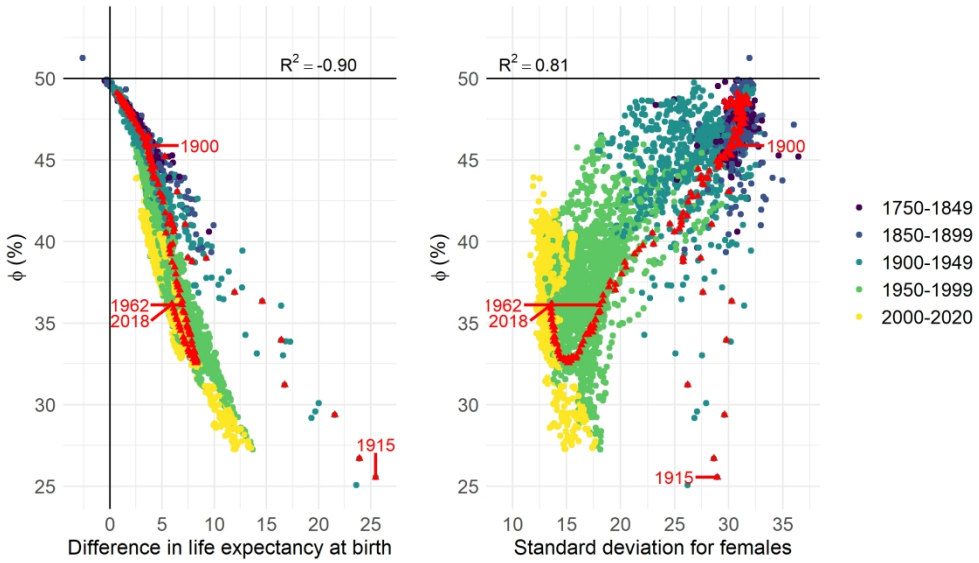

Figure 4 Relation between φ and (a) the sex differences in life expectancy and (b) the standard deviation of lifespans for females for HMD period data since 1751, with France highlighted (red triangles). Source: HMD [23] and authors' own calculations.

1761x1002mm (38 x 38 DPI)

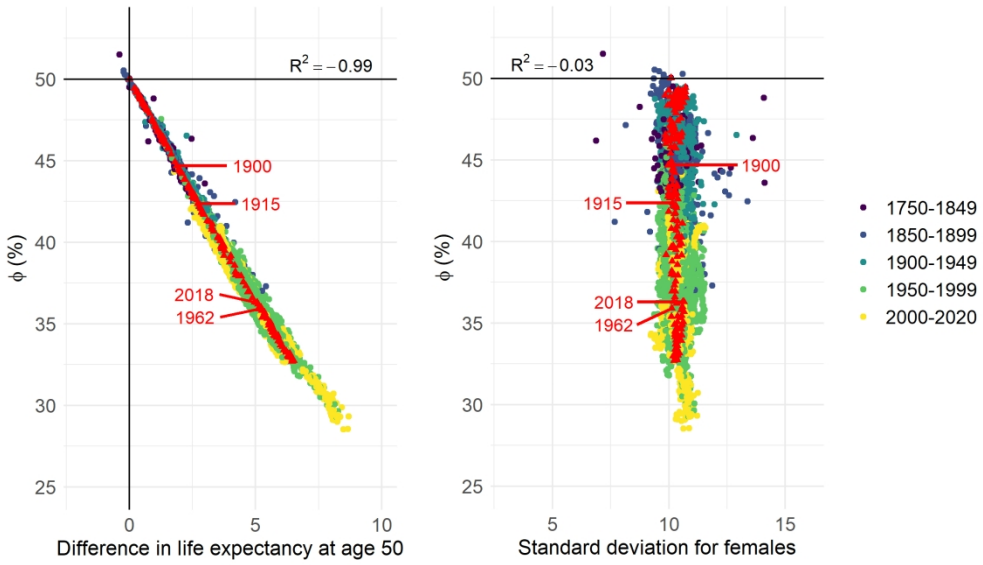

Figure 5 Relation between φ and (a) the sex differences in life expectancy and (b) the standard deviation for females for HMD period data since 1751, conditional to survival to age 50, with France highlighted (red triangles).
Source: HMD [23] and authors' own calculations.

1761x1002mm (38 x 38 DPI)

A. FORMAL RELATION

It can be shown that the outsurvival statistic relates to the joint probability density function of two lifespan distributions, which gives the probability of realizations of two lifespans and thus is related to the overlap of the two distributions. Assume two populations of individuals, with ages at death x and y, respectively. Assume the two populations are independent, meaning that the length of life x does not depend on the length of life y and vice versa. This implies that the joint probability density function, $d_{1,2}(x,y)$, equals the product of the marginal densities so that $d_{1,2}(x,y) = d_1(x)\, d_2(y)$. We are interested in calculating the probability (φ) of individuals in the first population outliving those in the second population. This implies that $0 \leq y < x$ so:

$$\begin{aligned} \varphi \quad &= \int_0^\infty \int_0^x d_{1,2}(x,y)dydx \;\; = \int_0^\infty d_1(x) \int_0^x d_2(y)dydx \\ &= \int_0^\infty d_1(x)D_2(x)dx \\ &= \int_0^\infty d_1(x)[1 - l_2(x)]dx \;\; = 1 - \int_0^\infty d_1(x)l_2(x)dx \\ &= \int_0^\infty d_2(x)\ell_1(x)dx. \end{aligned} \tag{A1}$$

Following the same approach, we can find the complement of φ, labeled φ', which is the probability of individuals in the second population to outlive those in the first:

$$\begin{aligned} \varphi' \quad &= \int_0^\infty \int_0^y d_{1,2}(x,y)dxdy \\ &= \int_0^\infty d_2(y) \int_0^y d_1(x)dxdy \\ &= \int_0^\infty d_2(y)D_1(y)dy \\ &= \int_0^\infty d_2(y)[1 - l_1(y)]\, dy \\ &= 1 - \int_0^\infty d_2(y)l_1(y)\, dy. \end{aligned} \tag{A2}$$

From Equations (B1) and (B2) it can be shown that $\varphi + \varphi' = 1$. Thus, φ is also equal to:

$$\begin{aligned} \varphi \quad &= 1 - \varphi' \\ &= 1 - \left[1 - \int_0^\omega d_2(x)l_1(x)\, dx\right] \\ &= \int_0^\omega d_2(x)l_1(x)\, dx. \end{aligned} \tag{A3}$$

B. SIMULATIONS AND DISCREATE APPROXIMATION

We simulated age at death distributions, using the Gompertz model, using various scale (M) and shape (β) parameters [1]. The distributions were first found using an age width (n) of 0.0001, after

which the data were aggregated within 1-year and 5-years age-groups. The probability that individuals in both population died within the same age-group, $\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2$, was then redistributed between φ and φ' based on two assumptions: equal (equation B1) and proportional redistributions (equation B2). The results are presented in Table B1.

$$\varphi \approx \sum_{x=0}^{\omega} {}_nd_x^1 \; {}_nD_{x-n}^2 + \frac{\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2}{2} \tag{B1}$$

$$\varphi \approx \sum_{x=0}^{\omega} {}_nd_x^1 \; {}_nD_{x-n}^2 + \sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2 \frac{{}_nd_x^1 \; {}_nD_{x-n}^2}{{}_nd_x^1 \; {}_nD_{x-n}^2 + \; {}_nd_x^2 \; {}_nD_{x-n}^1} \tag{B2}$$

The simulations show that equally redistributing $\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2$ between the two other statistics provide very similar results to the continuous data ($n=0.0001$), especially for the 1-year age-group. More differences are found when aggregating by 5-years age-groups, but the difference in φ between the different age-width remains less than 1 percentage point, when equally redistributing $\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2$.

Table B1. Assumptions to redistribute $\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2$ for different mortality scenarios.

	φ	φ'	$\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2$	Eq. B1	Eq. B2
Gompertz: $M_A = 61, M_B = 65, \; \beta_A = 0.12, \; \beta_B = 0.14$					
Continuous	**36.3**	63.7	0.0	-	-
1-year	34.8	62.2	3.0	**36.3**	35.9
5-years	28.2	55.8	15.0	**36.7**	34.3
Gompertz: $M_A = 61, M_B = 70, \; \beta_A = 0.10, \; \beta_B = 0.14$					
Continuous	**23.6**	76.4	0.0	-	-
1-year	22.5	75.2	2.3	**23.6**	23.0
5-years	18.5	70.0	11.3	**24.2**	20.9
Gompertz: $M_A = 68, M_B = 70, \; \beta_A = 0.13, \; \beta_B = 0.14$					
Continuous	**42.8**	57.2	0.0	-	-
1-year	41.2	55.5	3.3	**42.8**	42.6
5-years	34.9	48.8	16.3	**43.0**	41.7
Gompertz: $M_A = 69, M_B = 70, \; \beta_A = 0.10, \; \beta_B = 0.12$					
Continuous	**46.1**	53.9	0.0	-	-
1-year	44.7	52.6	2.7	**46.1**	46.0
5-years	39.4	47.2	13.4	**46.1**	45.5

To further test the model and the redistribution of $\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2$, we simulated 100,000 individual lifespans from an exponential distribution with piece-wise constant rates [2]. We performed this procedure for every population and by sex using as an input empirical death rates retrieved from the HMD [3]. Then we randomly paired males and females and calculated the proportions of males outliving the paired female. Table 3 compares the discrete approach introduced in the main document (eq. B1) and the continuous approach based on simulations. Both approaches provided very similar results.

Table B2. Proportions of males outliving females based on a discrete and continuous approach (simulations).

	% males outliving females		% females outliving males	
	Continuous	Discrete	Continuous	Discrete
Denmark				
1850	46.1	46.4	53.9	53.6
1900	45.8	45.8	54.2	54.2
1950	46.2	46.3	53.8	53.7
2016	40.7	40.6	59.3	59.4
France				
1850	48.6	48.5	51.4	51.5
1915	25.6	25.5	74.4	74.5
1950	39.8	39.8	60.2	60.2
2016	35.9	36.0	64.1	64.0
Japan				
1950	44.1	44.2	55.9	55.8
2016	33.8	33.7	66.2	66.3
Russia				
1960	35.6	35.5	64.4	64.5
2014	30.2	30.0	69.8	70.0

C. DATA

Table C1. Countries/ regions and years with available data in the HMD

Country\region	Years	Country\region	Years
Australia	1921-2018	Japan	1947-2019
Austria	1947-2019	Latvia	1959-2019
Belarus	1959-2018	Lithuania	1959-2019
Belgium	1841-2018	Luxembourg	1960-2019
Bulgaria	1947-2017	Netherlands	1850-2019
Canada	1921-2018	New Zealand	1948-2013
Chile	1992-2017	Norway	1846-2020
Croatia	2001-2019	Poland	1958-2019
Czechia	1950-2019	Portugal	1940-2018
Denmark	1835-2020	Republic of Korea	2003-2018
Estonia	1959-2019	Russia	1959-2014
Finland	1878-2019	Slovakia	1950-2017
France	1816-2018	Slovenia	1983-2017
Germany-East	1956-2017	Spain	1908-2018
Germany-West	1956-2017	Sweden	1751-2019
Greece	1981-2017	Switzerland	1876-2018
Hong Kong	1986-2017	Taiwan	1970-2019
Hungary	1950-2017	UK – England and Wales	1841-2018
Iceland	1838-2018	UK- Scotland	1855-2018
Ireland	1950-2017	UK- Northern Ireland	1922-2018
Israel	1983-2016	USA	1933-2019
Italy	1872-2018	Ukraine	1959-2013

D. EXTRA ANALYSIS: COHORTS AND WWP

Similar relations as those for period data were also found for cohorts (Figure D1). In the HMD, life table for cohorts were only available for 11 countries: Denmark, England and Wales, Finland, France, Iceland, Italy, Netherlands, Norway, Scotland, Sweden and Switzerland. For cohorts with complete mortality history, the proportions of males outliving females varied between 35% and 49%. Only small changes in φ were observed for cohorts born prior to 1870-1890, with φ varying around 46.5%. For the cohorts born afterwards, φ decreases, reaching a mean of 38.4% for the cohort born in 1925, with values varying between 35.3% (Finland) and 40.4% (Scotland).

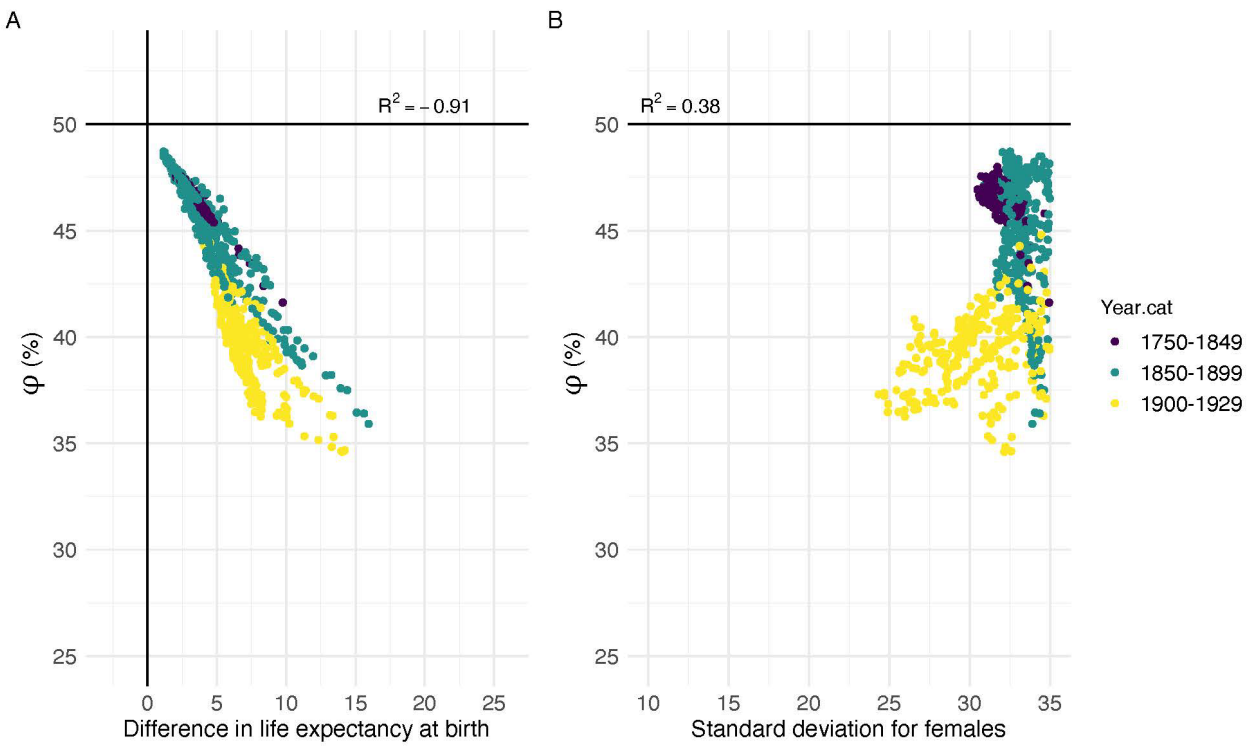

Figure D1 Relation between φ and (a) the sex differences in life expectancy and (b) the standard deviation for females for HMD cohort data.
Source: HMD [3] and authors' own calculations.

Figure D2 also shows the relation between φ and (A) the sex differences in life expectancy and (B) the standard deviation for females for countries in the WWP from 1950-55 to 2015-19. The relation between φ and the two measures is similar to that shown in the main text using the HMD data.

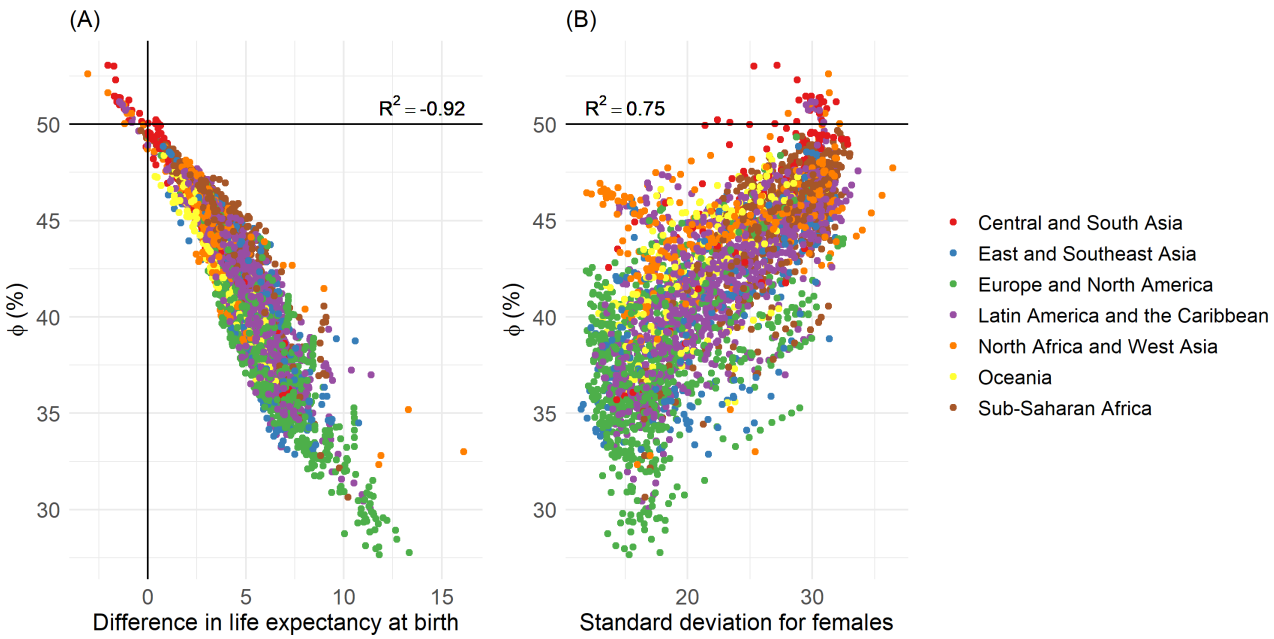

Figure D2 Relation between φ and (a) the sex differences in life expectancy and (b) the standard deviation for females for WPP data.
Source: WPP [4] and authors' own calculations.

E. OTHER MEASURES OF OVERLAP

Figure E1 shows the relation between φ and the stratification index used by Shi et al. [5]. Both indicators are strongly correlated with a correlation coefficient of 0.98. Figure E2 shows a similar relation between φ and the Kullback-Leibler divergence, with a correlation coefficient of -0.93.

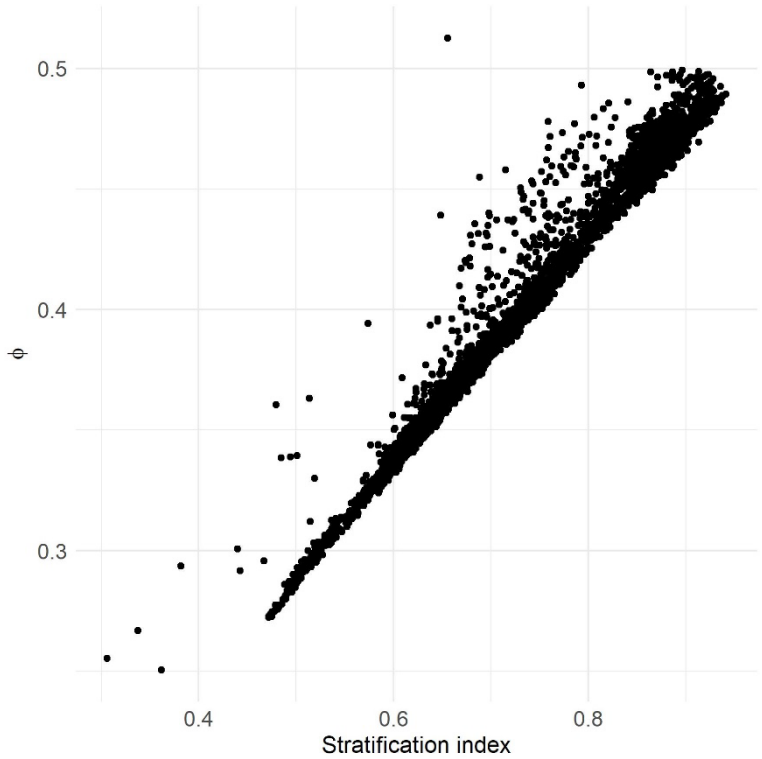

Figure E1. Relation between φ and the stratification index for HMD period data. Source: HMD [3] and authors' own calculations.

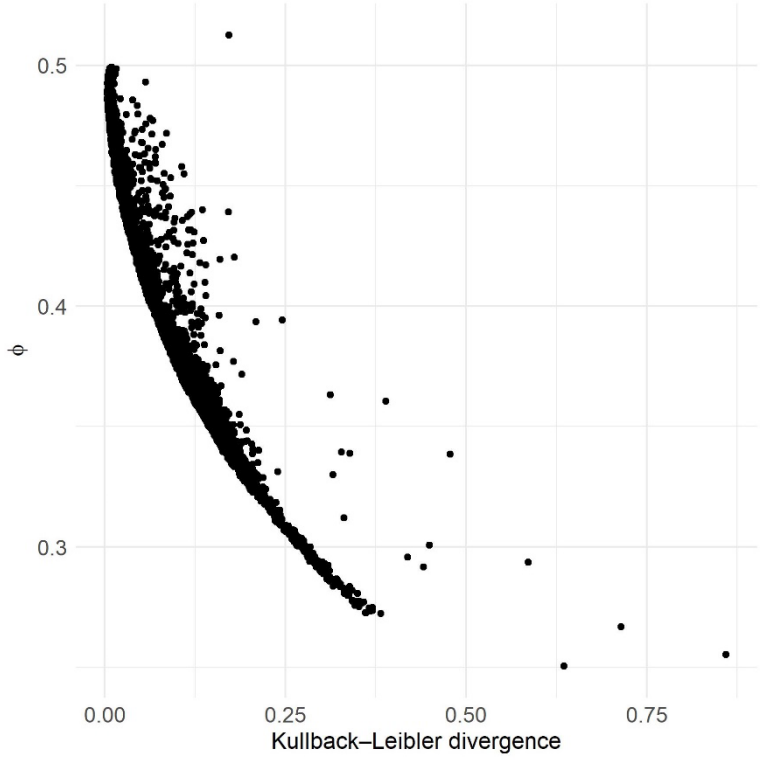

Figure E2. Relation between φ and the Kullback-Leibler divergence for HMD period data. Source: HMD [3] and authors' own calculations.

REFERENCES

1. Missov TI, Lenart A, Nemeth L, Canudas-Romo V, Vaupel JW: **The Gompertz force of mortality in terms of the modal age at death**. *Demographic Research* 2015, **32**:1031-1048.
2. Willekens F: **Continuous-time microsimulation in longitudinal analysis**. In: *New frontiers in microsimulation modelling.* edn. Edited by Harding A, Zaidi A. London: Routledge; 2009: 353-376.
3. HMD: **Human Mortality Database**. In.: University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany), Available at www.mortality.org (Accessed on July 15) . 2021.
4. United Nations: **World Population Prospects 2019, Online Edition. Rev. 1.** In.: Department of Economic and Social Affairs, Population Division, Available at https://population.un.org/wpp/Download/Standard/Mortality/ (Accessed on May 21); 2019.
5. Shi J, Aburto JM, Martikainen P, Tarkiainen L, van Raalte AA: **A distributional approach to measuring lifespan stratification**. *Population Studies* 2022, **(Accepted)**.

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract Reported on p. 1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found p.1
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported p.2-3
Objectives	3	State specific objectives, including any prespecified hypotheses p.3-4
Methods		
Study design	4	Present key elements of study design early in the paper Observational studies. p.4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection Publicly available database. P.6-7
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants Not applicable
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable p. 6-7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group p.6-7
Bias	9	Describe any efforts to address potential sources of bias None
Study size	10	Explain how the study size was arrived at Full population used. P. 6-7
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why p.4-7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding p.4-6
		(*b*) Describe any methods used to examine subgroups and interactions p.4-6 (same methods used for subgroups as for the total population)
		(*c*) Explain how missing data were addressed Handled within the publicly available databases.
		(*d*) If applicable, describe analytical methods taking account of sampling strategy Not applicable
		(*e*) Describe any sensitivity analyses p. 4-5
Results		

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed Not applicable
		(b) Give reasons for non-participation at each stage Not applicable
		(c) Consider use of a flow diagram Not applicable
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders Not applicable
		(b) Indicate number of participants with missing data for each variable of interest Not applicable
Outcome data	15*	Report numbers of outcome events or summary measures Not applicable
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included p. 7-8
		(*b*) Report category boundaries when continuous variables were categorized Not applicable
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period Not applicable
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses p. 8-9
Discussion		
Key results	18	Summarise key results with reference to study objectives p.10
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias p. 10-11
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence p.10-11
Generalisability	21	Discuss the generalisability (external validity) of the study results p.11-12
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based p.13

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely

available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Probability of males to outlive females, an international comparison from 1751 to 2020

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-059964.R2
Article Type:	Original research
Date Submitted by the Author:	03-Jun-2022
Complete List of Authors:	Bergeron-Boucher, Marie-Pier; Syddansk Universitet, Interdisciplinary Center on Population Dynamics Álvarez, Jesús-Adrian; University of Southern Denmark, Interdisciplinary Centre on Population Dynamics Kashnitsky, Ilya; Syddansk Universitet, Interdisciplinary Centre on Population Dynamics Zarulli, Virginia; Syddansk Universitet, Interdisciplinary Center on Population Dynamics
Primary Subject Heading:	Research methods
Secondary Subject Heading:	Epidemiology, Public health
Keywords:	PUBLIC HEALTH, STATISTICS & RESEARCH METHODS, EPIDEMIOLOGY

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Probability of males to outlive females, an international comparison from 1751 to 2020

Marie-Pier Bergeron-Boucher, corresponding author, mpbergeron@sdu.dk
Jesús-Adrián Álvarez, alvarez@sdu.dk
Ilya Kashnitsky, ikashnitsky@sdu.dk
Virginia Zarulli, vzarulli@sdu.dk

Danish Centre for Population Research, Interdisciplinary Centre on Population Dynamics, University of Southern Denmark, Odense, Denmark

ABSTRACT

Objective To measure sex differences in lifespan based on the probability of males to outlive females.

Design International comparison of national and regional sex-specific life tables from the Human Mortality Database and the World Population Prospects.

Setting 199 populations spanning all continents, between 1751 and 2020.

Primary outcome measure We used the outsurvival statistic (φ) to measure inequality in lifespan between sexes, which is interpreted here as the probability of males to outlive females.

Results In random pairs of one male and one female at age zero, the probability of the male outliving the female varies between 25% and 50% for life tables in almost all years since 1751 and across almost all populations. We show that φ is negatively correlated with sex differences in life expectancy and positively correlated with the level of lifespan variation. The important reduction of lifespan inequality observed in recent years has made it less likely for a male to outlive a female.

Conclusions Although male life expectancy is generally lower than female life expectancy, and male death rates are usually higher at all ages, males have a substantial chance of outliving females. These findings challenge the general impression that 'men do not live as long as women' and reveal a more nuanced inequality in lifespans between females and males.

Keywords: Lifespan, Inequality, Sex differences, Outsurvival statistics

Strengths and limitations of this study

- This is the first study using the outsurvival statistic to quantify the probability of males outliving females.
- The outsurvival statistic shows that lifespan inequalities between sexes have more nuances that cannot be captured by comparisons made with classic summary demographic measures (e.g. life expectancy).
- The outsurvival statistic does not account for dependence between individuals, such as couples whose health and mortality patterns are positively correlated due to a strong effect of social ties on health and longevity.

1. INTRODUCTION

The female survival advantage has been observed over time across many human populations and is rooted in a complex combination of biological, environmental and behavioural factors [1-5]. For example, males tend to engage in more risky behaviours, such as smoking and heavy drinking, but estrogen could also be preventive against certain diseases [6]. A study on cloistered populations reveals a constant female survival advantage of around 0.2 years. The author attributes the remaining sex differences in life expectancy in the general population to differences in lifestyle and socioeconomic burden [7]. However, even among populations where men and women differ less in terms of key lifestyle factors, such as Mormons, sex differences in life expectancy still exist [8]. In 2019, the sex difference in life expectancy was 4.4 years on average worldwide, with large variation across countries [9]. Females have been found to have longer survival and lower death rates than men at all ages and in most modern populations [2 4 10-13] and even under extreme mortality conditions [14].

Sex differences in survival are often identified by comparing life expectancy between females and males, which summarises the average length of life. These differences are often interpreted as 'men do not live as long as women'. Such an interpretation is simplistic as it does not account for the variation around the means (life expectancies) and potential overlaps between female and male lifespan distributions. Despite females having a higher life expectancy than males, not all females outlive all males. On the contrary, a sizeable portion of males might live longer than a sizeable portion of females, even if the life expectancy shows a female advantage. This is because the lifespan distributions of females and males partly overlap, i.e. they share a common range of ages at death.

The extent of the overlapping indicates how likely it is for males to outlive females and, ultimately, how sizeable the portion is of males living longer than females.

Lifespan variation, i.e. differences in lifespans within a population, has been receiving increasing attention in the literature [15]. Various indicators reveal heterogeneity in the length of life, beyond what life expectancy indicates. Studies have compared lifespan variation between two populations, focusing on which populations exhibit more inequalities [16-18]. It has been shown that females systematically experience lower lifespan variation than males [16]. However, it is unclear how this variation around the means leads to potential overlap between the two lifespan distributions.

Only a few studies have used measures of overlap or distance to study inequalities between populations. A previous study has investigated the extent to which two lifespan distributions differ using the Kullback-Leibler (KL) divergence [19]. The indicator is interpreted as the amount of 'effort' needed to transform the male's lifespan distribution into the female's distribution. A disadvantage of this indicator is that it is not symmetrical, meaning that the effort needed to transform the male's distribution into the female's is not the same as the effort needed to transform the female's distribution into the male's. Stratification indexes, based on how much two lifespan distributions overlap or do not overlap, have also been used to study mortality differences between socioeconomic groups [20]. The larger the overlap, the more likely the individuals in two populations are to survive to the same age. This index is meant to reflect unequal distribution at the societal level, with values varying between 0 (no overlap) and 1 (perfect overlap). A related measure is the outsurvival statistic, which quantifies the probability that an individual from a population with lower life expectancy outlives an individual from another population with higher life expectancy [21]. The main difference with the stratification index is the interpretation, which focuses on the individuals. If the two populations are males and females, the outsurvival statistic captures the correctness of the assertion that males' lifespans are lower than females' lifespans. If both populations have equal lifespan distribution, the outsurvival statistic is equal to 0.5. Unlike the other two measures, the outsurvival statistic also explicitly reveals which of the compared populations has an advantage (values above 0.5) or a disadvantage (values below 0.5).

In this article, we use the outsurvival statistic [21] to study lifespan inequalities between females and males. We aim to 1) quantify the probability that males outlive females over time and across populations; and 2) assess the sensitivity of the outsurvival statistic to changes in life expectancy and lifespan variation. We computed the outsurvival probability to study sex differences in mortality in

199 populations over 200 years. Despite sometimes large differences in life expectancy, we show that there are substantial overlaps between males' and females' lifespan distributions.

2. METHOD

2.1 OUTSURVIVAL STATISTIC

Consider two populations with mean and standard deviation (SD) of the age of death (see [22] [23] for more details on the SD calculation) specified in panel A of Figure 1. The first population (in red) has a smaller mean lifespan and larger SD than the second population (in blue). An inference from these means would be that individuals in the first population are worse off than individuals in the second. However, there is an important overlap between the two distributions, with some individuals in the first population outliving some individuals in the second population. The outsurvival probability, φ, captures this dimension by measuring the probability that an individual from a population with high mortality will outlive an individual from a population with low mortality [21]. Let $d_i(x), i = 1,2$ denote the lifespan distribution at age x in two populations. The cumulative distributions are represented by $D_i(x)$, such that $D_i(x) = \int_0^x d_i(t)dt$ and the survivorship is denoted by $\ell_i(x)$, with $\ell_i(x) = 1 - D_i(x)$. The probability that an individual from the first population (males) will outlive an individual from the second population (females) is [21]:

$$\varphi = \int_0^\infty d_2(x)\ell_1(x)dx. \quad (1)$$

In scenario A of Figure 1, φ is 40%.

[Figure 1]

In the supplementary materials, we show that the outsurvival statistic relates to the joint probability density function of two lifespan distributions, which gives the probability of realisations of two lifespans and is thus related to the overlap of the two distributions.

2.2 RELATION TO LIFE EXPECTANCY AND LIFESPAN VARIATION

Consider the two populations in scenarios B and C of Figure 1. The difference in mean lifespan is the same in both cases, i.e. 15 years. However, in scenario C the first population has a larger SD, which implies more individuals surviving to older ages, despite greater inequalities, and thus a greater

potential to outlive individuals from the second population. Indeed, φ is higher in scenario C (19%) than in scenario B (14%). Now compare scenario B to scenario D. This time, the second population in D has a smaller SD, with fewer individuals dying at younger ages, making it less likely for individuals in the first population to outlive them. This reduces φ to 12%. Thus, for the same difference in life expectancy, larger lifespan variation in both populations generally results in larger φ . The comparison of scenarios A and C also shows that small differences in life expectancy lead to larger value of φ.

Equation (1) is not new and relates to the Mann-Whitney U test, the probability of superiority and to the expected failure probability in a stress-strength interference (SSI) model. The latter assesses the probability that the stress (population 1) exceeds the strength (population 2) of a material [24]. If the distributions of both populations follow a Normal distribution with mean μ_i and standard deviation s_i, the probability of failure is P(Z) with $Z = -\frac{\mu_2 - \mu_1}{\sqrt{s_1^2 + s_2^2}}$ [25]. This relation formalises what is illustrated in section in Figure 1: φ is sensitive to the difference in the means and to the level of variation in both distributions, with smaller mean differences (numerator) and larger variance (denominator) leading to larger P(Z). However, lifespan distributions are not normally distributed, and additional moments could also affect the value of φ. To better understand this relation, we analysed the correlation between φ and life expectancy as well as between φ and lifespan variation.

2.3 DISCRETE APPROXIMATION

Similar equivalences to equation (1) can be developed in a discrete time setting. Let ${}_nd_x^i$ be the life table deaths between age x and $x+n$ in population i and ${}_nl_x^i$ the survival probability to age x. For a given age group width of n, the probability of individuals in the first population outliving those in the second population can be found by:

$$\varphi \approx \sum_{x=0}^{\omega} {}_nd_{x-n}^2 \, {}_nl_x^1 + \bar{d} \qquad (2)$$

with $\bar{d} = \frac{\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2}{2}$ and $\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2$ being the probability that individuals in both populations died in the same age group. The latter statistic is sensitive to the width of the age groups such that smaller age groups result in smaller values, with $\lim_{n \to 0} \sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2 = 0$. In the supplementary materials, we

compared the discrete and continuous approaches and found that both approaches yield comparable results.

Equations (1) and (2) are equivalent to matching random individuals from each population and calculating the proportions of individuals from the first population who outlive the paired individual from the second. We performed such analyses via simulations of individuals from a specific lifespan distribution and estimated the corresponding statistics (see supplementary materials). Equivalent results were found.

PATIENT AND PUBLIC INVOLVEMENT

No patients were involved.

3. DATA

The method was applied to three demographic datasets. First, we used life tables by sex for all available countries and years from the Human Mortality Database (HMD) [26]. The HMD is freely available, provides comparable long time-series for 41 countries with high quality data. Data are provided by single-year age groups. We used subnational data for Germany, with separate analysis for East and West Germany, and for the United Kingdom, with separate analysis for England-Wales, Scotland and Northern Ireland, amounting to 44 studied populations. The earliest year with available data was 1751 (for Sweden) and the latest was 2020. Information about the available populations and years is provided in the supplementary materials. We compared females' and males' life tables in each country/region.

Second, we used abridged life tables from the World Population Prospects 2019 Revision (WPP) [27]. This dataset is also freely available and provides sex-specific life tables for 199 countries by 5-year age groups and 5-year periods from 1950-1954 to 2015-2019. This database covers the whole world, but the data quality varies greatly between countries [28]. The HMD and WPP data are used to compare the outsurvival statistic over time and across multiple populations.

Finally, we computed the outsurvival statistic for subpopulations of females and males using US data in 2015-2019. We compared the probability of males to outlive females by education level and marital status to assess if the sex differences emerge from specific subpopulations. We calculated sex-specific life tables by education level and marital status using death counts from the Multiple Cause of Death dataset (MCDD) from the National Vital Statistics System of the National Center for Health Statistics [29] and population counts from the American Community Service (ACS) from the

United States Census Bureau [30]. The MCDD provides death counts by single-year age groups, sex, marital status and education level. The ACS provides data by similar variables and single-year age groups until age 96. However, it is worth noting that the ACS data exhibit an important age-heaping at age 95. We therefore ungrouped the population counts from age 90+ using the Penalized Composite Link Model (PCLM) model [31] to obtain the population counts from age 90 to 110 by single-year of age.

4 RESULTS

4.1 HISTORICAL VALUES AND TRENDS in φ

Figure 2a shows the outsurvival probability of males over females (φ) since 1850 for all HMD countries and Figure 2b for all WWP countries since 1950-1955. The probability of males outliving females has, at all points in time and across all populations, varied between 25% and 50%, with only few exceptions with values above 50%: Iceland in 1891; Jordan in 1950-1954; Iran in 1950-1964, Iraq in 1960-1969; before 1985 in Bangladesh, India and the Maldives; and between 1995 and 2010 in Bhutan.

[Figure 2]

For the HMD countries, φ was slowly decreasing before the First World War, on average from 47.3% in 1850 to 46.0% in 1913. After the War, φ declined faster. In 1930, the mean φ across populations was 45.4%, ranging from 42.8% (France) to 48.4% (Netherlands). By 1985, the mean φ was 35.3%, ranging from 31.2% (Russia) to 42.8% (Israel). The value of φ started increasing around the 1980s for some countries, but continued to decrease in others until the 2000s, especially in Eastern European countries. The mean for all countries was 37.1% in 2017, with values varying between 28.7% (Belarus) and 42.5% (Iceland).

For the WWP countries, we observed a decrease in φ in all regions since 1950, except in Europe, Northern America and Oceania, which increased from the 1980s, as is also shown in the analysis of the HMD data. In 1950-1955, φ was 46.1% on average worldwide, with values ranging between 35.3% (in Kazakhstan) and 52.6% (in Iran). By 2015-2019, φ declined to 41.2% with values ranging between 28.8% (in Belarus) and 49.9% (in Bhutan). Figure 3 shows φ across the world in different time periods. In recent years, the outsurvival of females was particularly low in Eastern Europe and Northeast Asia and was particularly high in Southern Asia and in Western and Middle

Africa. Males from many Southern Asian countries had an especially high chance of outliving females, with φ above 50% before 1970.

[Figure 3]

4.2 RELATION TO LIFE EXPECTANCY AND LIFESPAN INEQUALITY

Figure 4 shows that φ is negatively correlated with the differences in life expectancy and positively correlated with females' SD (as shown in the supplementary materials, similar results were found when males' SD was used, due to the strong correlation between females and males' SDs). Figure 4 is based on the HMD data, but the same relation is found when using the WPP data (see supplementary materials). This relation empirically demonstrates the formal relation in section 2.2. The correlation between φ and the standard deviation has been weaker in recent years, due to a reduction in sex differences in life expectancy, which is also driving changes in φ . Even though both life expectancy and lifespan variation affect φ, the statistic appears more sensitive to the differences in life expectancy than to the level of lifespan variation. We also found similar results for cohort data (see supplementary materials).

[Figure 4]

The same value for φ can be found for different combinations of sex differences in life expectancy and levels of lifespan variation. For example, the same φ of 36.1% was found in France in 1962 and in 2018 (Figure 4). However, the sex difference in life expectancy was 6.9 in 1962 and 5.9 in 2018, and the standard deviation for females was 18.1 in 1962 and 13.6 in 2018.

Figure 5 shows the same relations as shown in Figure 4 but for survivors to age 50. Lifespan variation at age 50 has stayed roughly constant over time, [32] and comparing φ from this age can help to assess the sensitivity of the measure to changes in lifespan variation (similar results were found when using males' SD, see supplementary materials). The relation between φ and differences in life expectancy is stronger and more linear from age 50 (correlation coefficient of -0.99) than when using the full age range, increasing predictive ability. For example, for a difference in life expectancy at age 50 of 3 years, males have around 42% probability of outliving females. Note that φ in France was 35.9% in 1962 and 36.3% in 2018.

[Figure5]

Similar to the distribution from birth, the probability of males outliving females from age 50 has, in almost all periods and populations, varied between 28% and 50%, with only few exceptions. In recent years, the φ statistics from birth and from age 50 are similar.

4.3 SEX DIFFERENCES BY EDUCATION AND MARITAL STATUS

Tables 1 and 2 show the φ statistic for some subpopulations of males and females in the United States. For the period 2015-2019, the probability of males to outlive females was 40% in the total US population. However, this statistic varies depending on marital status and education level, being higher among the subpopulations with beneficial characteristics: the probability of males to outlive females was 39% for married individuals and 37% for unmarried individuals (Table 1); 43% for individuals with a university degree and 39% for those without a high school diploma (Table 2).

Table 1. Outsurvival statistics by sex and marital status in the United States, 2015-2019

		Female	
		Married	Unmarried
Male	Married	0.39	0.52
	Unmarried	0.26	0.37

Source: MCDD [29], ACS [30] and authors' own calculations using Equation (2).

Table 2. Outsurvival statistics by sex and education level in the United States, 2015-2019

		Female		
		University Degree	High School Diploma	No High School Diploma
Male	University Degree	0.43	0.51	0.53
	High School Diploma	0.32	0.39	0.42
	No High School Diploma	0.30	0.37	0.39

Source: MCDD [29], ACS [30] and authors' own calculations using Equation (2).

Furthermore, these results highlight that males with beneficial characteristics (being married and having a university degree) have an advantage over women with detrimental characteristics (being unmarried and having only a high school diploma or less).

5. DISCUSSION

Our study reveals a nuanced inequality in lifespan between females and males, with between one and two men out of four outliving a randomly paired woman in almost all points in time across 199 populations. These results complement the picture given by the comparisons based on life expectancy, which is a summary measure with no information on variation. A blind interpretation of life expectancy differences can sometimes lead to a distorted perception of the actual inequalities. Not all females outlive males, even if a majority do. But the minority that do not is not small. For example, a sex difference in life expectancy at birth of 10 years can be associated with a probability of males outliving females as high as 40%, indicating that at 40% of males have a longer lifespan than that of a randomly paired female. Not all males have a disadvantage of 10 years, which is overlooked by solely making comparisons of life expectancy. However, a small number of males will live very short lives to result in that difference. For example, more baby boys die than baby girls in most countries.

The length of the lifespan of an individual results from a complex combination of biological, environmental, and behavioural factors. Being male or female does impact lifespan, but it is not the only determinant contributing to inequalities. Lifespan has been shown to be influenced by marital status, income, education, race\ethnicity, urban\rural residence, etc. [33] As we only disaggregated the population by sex and because of this complex interaction, lifespan distributions of females and males overlap. This nuance is captured by the φ metric. Males with a lower education level or who are unmarried have a particularly low chance of outliving a female. But males with a university degree or who are married have a higher chance of outliving females, in particular females with a lower education level and who are single.

As previously discussed, the φ metric expresses the probability of males to outlive females among randomly paired individuals, assuming independence between populations. However, males and females in a population are generally not random pairs but often couples, whose health and mortality have been found to be positively correlated due to a strong effect of social ties on health and longevity [34]. Coupled individuals also influence each other's health [35], and this is particularly true for males, who benefit more than females from being in a stable relationship [36]. The datasets used for the analysis do not permit the estimation of the probability of males outliving females for non-

randomly paired individuals. However, the outsurvival statistic relates to the probability of the husbands to outlive their wives, and even though such a measure accounts for the difference in age between husband and wife, it has been shown generally to be between 30% and 40% [37-39], values that are quite close to φ.

Other measures of overlap and distance between distributions could have been used. In the supplementary materials, we compare the outsurvival statistic with a stratification index used by Shi and colleagues [20] and the Kullback-Leibler divergence. We found that all three indicators are strongly correlated and using any one of these would not have changed the general conclusions from this article. However, unlike the other indicators, φ directly indicates when males live longer than females, which we found in a few instances.

Trends over time in φ are consistent with the reversed trends in sex differences in life expectancy [40]: in developed countries, the probability of males outliving females decreased until the 1970s, after which it gradually increased in all populations. Studies showed that the increase in sex differences in mortality emerged in cohorts born after 1880 [10 41], which is consistent with our analysis of φ (see supplementary materials). The increase and decrease in sex differences in life expectancy was mainly attributed to the smoking epidemic and other behavioural differences between sexes [7 13 42].

The φ values are generally higher in low- and middle-income countries. However, this should not be interpreted as a sign of greater gender equality in survival. Southern Asian countries had very high φ values, above 50% in the 1950s and 1960s. Studies for India showed that mortality below age five was higher for females than males and remained higher for females in recent years [43 44]. However, females had a growing mortality advantage above age 15 since the 1980s, 'balancing out' the disadvantage at younger ages. The reasons for the higher φ and decreasing trends in developing regions vary across countries. It is outside the scope of this study to provide a detailed explanation for the trends in each country.

The outsurvival statistic can be informative for public health interventions [21]. Governments develop public health programmes to reduce lifespan inequalities at different levels (e.g. socioeconomic status, race, sex, etc.). It would be misleading to say that half of the population is disadvantaged by sex differences in lifespan. The inequalities are more nuanced. If 40% of males live longer than females, it could be argued that if a policy aiming at reducing inequalities between sexes targeted the full male population some of the efforts and investments would be misallocated. Such a policy could be more efficient if φ approaches 0, indicating that sex would explain a large part of the

lifespan inequalities within the population, whereas a φ closer to 0.5 indicates that other characteristics (e.g. socioeconomic and marital statuses) are involved in creating inequalities. We showed that some subpopulations of males have a high probability (above 50%) of outliving some subpopulations of females. Males who are married or have a university degree tend to outlive females who are unmarried or do not have a high school diploma. Inequalities in lifespan between sexes are attributable to some individuals within each population and not to the whole population. Indeed, Luy and Gast [12] found that male excess mortality is mainly caused by some specific subpopulations of males with particularly high mortality. Being able to better identify the characteristics of the short-lived men could more efficiently help tackle male-female inequality.

An important result of our analysis is that the smaller the standard deviation in the age at death, the smaller the φ. The reduction of lifespan inequality observed over time has then made it less likely for males to outlive females. This is partly explained by the fact that lifespan variation reduction has been driven by mortality declines at younger ages [45]. When looking at the lifespan distribution (as in Figure 1, scenario D), survival improvements at younger ages narrowed the left tails of the distribution for both sexes. By reducing the left tail of female distribution, without increasing the right tail of the male distribution, the overlapping area is reduced. In other words, the number of females with shorter lifespan, easier to outlive, decreased over time. Indeed, it has been shown that mortality declined at a faster pace for females than males below age 50, especially in the first half of the 20th century [46 47]. This finding implies that more efforts are required today than in the past to reduce these inequalities, for a same difference in life expectancy. While inequalities were mainly attributable to infant and child mortality, they are today increasingly attributable to older and broader age groups. Men maintained their disadvantage at younger ages, but also faced an increasing disadvantage at older ages. Men are more prone to accidents and homicides in their 20s and 30s than females, and they tend to smoke and drink more leading to higher cancer prevalence and death in their 60s. At the same time, women benefited from reduced maternal mortality and recorded faster mortality decline at older ages. Efforts in reducing lifespan inequalities must thus target diverse factors, causes and ages [13 46 48].

A decrease of φ might indicate a discrepancy in the causes of death that affect males and females. External mortality due to accidents and suicide has become more relevant in shaping sex differences in survival in recent years in high income populations [12]. Another example is observed in Latin American populations, where homicides and violent deaths have had an increased burden among males in comparison to females since the 1990s [49 50]. In Mexico, for example, the increase in

homicide mortality, especially among men between 20 and 40, contributed to increasing the gap in mortality between females and males [51]. This phenomenon is reflected in the decrease over time in the overlapping of lifespan distributions, directly informing healthcare systems of emerging inequalities.

However, one might ask if a wider overlapping is necessarily better for healthcare systems. On the one hand, a larger overlapping means less inequality between sexes, but on its own it does not ensure that there is more 'health justice'. For example, if the overlapping areas are large, this still shows a situation of great uncertainty in lifespan for both groups. One health evaluator actor could even prefer a situation where there is a small gap between groups but less inequality within the groups. In the case of sex differences, there might always be between-group differences due to biological factors [2 52], but more health equity could be reached by reducing within-group inequalities. We argue that the outsurvival statistic is a new tool to evaluate health inequalities between groups within a population by uncovering underlying dynamics that are otherwise hidden when looking only at conventional indicators. Therefore, it can inform healthcare systems of the subsequent directions to reach the preferred goal.

6. CONCLUSION

Comparing life expectancy between females and males provides a simplistic view of lifespan inequalities between sexes. Using measures of overlap between two distributions of lifespans complements these summary measures and offers a more comprehensive understanding of inequalities.

Acknowledgements

The authors are grateful to James W. Vaupel, Jim Oeppen and the two reviewers for their useful comments and edits.

Funding statement

The research and publication of this paper were supported by the AXA Research Fund, through the funding for the "AXA Chair in Longevity Research". In addition, this publication is a part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 884328 – Unequal Lifespans).

Author contributions

MPBB designed and conceptualized of the study. MPBB, JAA and IK produced the results and did the analysis. MPBB, JAA, IK and VZ contributed to the interpretation of the results and the drafting, revision and approval of the manuscript.

Data statement

The data are publicly available at http://www.mortality.org and https://population.un.org/wpp/. The R code to reproduce the results will be made public upon acceptance of the paper.

Patient consent for publication

Not applicable

Conflicts of interests

None to declare

Ethical approval

Not applicable.

Word count

4457

REFERENCES

1. Crimmins EM, Shim H, Zhang YS, et al. Differences between men and women in mortality and the health dimensions of the morbidity process. *Clinical chemistry* 2019;65(1):135-45.
2. Austad SN, Fischer KE. Sex differences in lifespan. *Cell metabolism* 2016;23(6):1022-33.
3. Rogers RG, Everett BG, Saint Onge JM, et al. Social, behavioral, and biological factors, and sex differences in mortality. *Demography* 2010;47(3):555-78.
4. Gjonça A, Tomassini C, Toson B, et al. Sex differences in mortality, a comparison of the United Kingdom and other developed countries. *Health Statistics Quarterly* 2005;26:6-16.
5. Preston SH, Wang H. Sex mortality differences in the United States: The role of cohort smoking patterns. *Demography* 2006;43(4):631-46.
6. Oksuzyan A, Juel K, Vaupel JW, et al. Men: good health and high mortality. Sex differences in health and aging. *Aging clinical and experimental research* 2008;20(2):91-102.

7. Luy M. Causes of male excess mortality: insights from cloistered populations. *Population and Development Review* 2003;29(4):647-76.
8. Lindahl-Jacobsen R, Hanson HA, Oksuzyan A, et al. The male–female health-survival paradox and sex differences in cohort life expectancy in Utah, Denmark, and Sweden 1850–1910. *Annals of epidemiology* 2013;23(4):161-66.
9. Organization WH. World health statistics overview 2019: monitoring health for the SDGs, sustainable development goals: World Health Organization, 2019.
10. Beltran-Sanchez H, Finch CE, Crimmins EM. Twentieth century surge of excess adult male mortality. *Proceedings of the National Academy of Sciences* 2015;112(29):8993-8. doi: 10.1073/pnas.1421942112 [published Online First: 2015/07/08]
11. Barford A, Dorling D, Smith GD, et al. Life expectancy: women now on top everywhere. *BMJ* 2006;332(7545):808.
12. Luy M, Gast K. Do women live longer or do men die earlier? Reflections on the causes of sex differences in life expectancy. *Gerontology* 2014;60(2):143-53.
13. Zarulli V, Kashnitsky I, Vaupel JW. Death rates at specific life stages mold the sex gap in life expectancy. *Proceedings of the National Academy of Sciences* 2021;118(20):e2010588118. doi: 10.1073/pnas.2010588118
14. Zarulli V, Jones JAB, Oksuzyan A, et al. Women live longer than men even during severe famines and epidemics. *Proceedings of the National Academy of Sciences* 2018;115(4):E832-E40.
15. van Raalte AA, Sasson I, Martikainen P. The case for monitoring life-span inequality. *Science* 2018;362(6418):1002-04.
16. Colchero F, Rau R, Jones OR, et al. The emergence of longevous populations. *Proceedings of the National Academy of Sciences* 2016;113(48) doi: 10.1073/pnas.1612191113
17. Permanyer I, Scholl N. Global trends in lifespan inequality: 1950-2015. *PloS one* 2019;14(5):e0215742.
18. Sasson I. Trends in life expectancy and lifespan variation by educational attainment: United States, 1990–2010. *Demography* 2016;53(2):269-93.
19. Edwards RD, Tuljapurkar S. Inequality in life spans and a new perspective on mortality convergence across industrialized countries. *Population and Development Review* 2005;31(4):645-74.
20. Shi J, Aburto JM, Martikainen P, et al. A distributional approach to measuring lifespan stratification. *Population Studies* 2022;(Accepted)
21. Vaupel JW, Bergeron-Boucher M-P, Kashnitsky I. Outsurvival as a measure of the inequality of lifespans between two populations. *Demographic Research* 2021;44(35):853-64. doi: 10.4054/DemRes.2021.44.35
22. Ouellette N, Bourbeau R. Changes in the age-at-death distribution in four low mortality countries: A nonparametric approach. *Demographic Research* 2011;25:595-628.
23. Wilmoth JR, Horiuchi S. Rectangularization revisited: Variability of age at death within human populations. *Demography* 1999;36(4):475-95.
24. An Z-W, Huang H-Z, Liu Y. A discrete stress–strength interference model based on universal generating function. *Reliability Engineering & System Safety* 2008;93(10):1485-90.
25. EL-Sayed S A, Constantin C. Probabilistic Simulation Approach to Evaluate the Tooth-Root Strength of Spur Gears with FEM-Based Verification. *Engineering* 2011;12(3):1137-48.
26. [dataset] HMD. Human Mortality Database: University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany), Available at www.mortality.org (Accessed on July 15, 2021) . 2021.
27. [dataset] United Nations. World Population Prospects 2019, Online Edition. Rev. 1.: Department of Economic and Social Affairs, Population Division, Available at

https://population.un.org/wpp/Download/Standard/Mortality/ (Accessed on May 21, 2022), 2019.
28. United Nations, Economic Do, Affairs S. World population prospects 2019: methodology of the United nations population estimates and projections. 2019
29. [dataset] NCHS. National Center for Health Statistics, Multiple Cause of Death File 2015-2019: National Vital Statistics System, Washington D.C., United States, Available at https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm#Mortality_Multiple (Accessed on January 13, 2022) 2022.
30. [dataset] U.S. Census Bureau. 2015-2019 American Community Survey 5-year estimates - Public Use Microdata Sample: Available at https://data.census.gov/mdat/#/search?ds=ACSPUMS5Y2019 (Accessed on Janúary 13, 2022), 2022.
31. Rizzi S, Gampe J, Eilers PH. Efficient estimation of smooth distributions from coarsely grouped data. *American Journal of Epidemiology* 2015;182(2):138-47.
32. Engelman M, Caswell H, Agree EM. Why do lifespan variability trends for the young and old diverge? A perturbation analysis. *Demographic Research* 2014;30:1367.
33. Marmot M, Wilkinson R. Social determinants of health: Oup Oxford 2005.
34. Rasulo D, Christensen K, Tomassini C. The influence of social relations on mortality in later life: A study on elderly Danish twins. *The Gerontologist* 2005;45(5):601-08.
35. Perelli-Harris B, Hoherz S, Addo F, et al. Do marriage and cohabitation provide benefits to health in mid-life? The role of childhood selection mechanisms and partnership characteristics across countries. *Population research and policy review* 2018;37(5):703-28.
36. Staehelin K, Schindler C, Spoerri A, et al. Marital status, living arrangement and mortality: does the association vary by gender? *Journal of Epidemiology and Community Health* 2012;66(7):e22.
37. Compton J, Pollak RA. The life expectancy of older couples and surviving spouses. *PloS one* 2021;16(5):e0250564.
38. Goldman N, Lord G. Sex differences in life cycle measures of widowhood. *Demography* 1983;20(2):177-95. doi: 10.2307/2061234
39. Statistical measures in the marital life-cycles of men and women. Population; 1959. INST NAT ETUD DEMOGRAPHIQUES 27 RUE DU COMMANDEUR, 75675 PARIS CEDEX 14, FRANCE.
40. Glei DA, Horiuchi S. The narrowing sex differential in life expectancy in high-income populations: effects of differences in the age pattern of mortality. *Population studies* 2007;61(2):141-59.
41. Rigby JE, Dorling D. Mortality in relation to sex in the affluent world. *Journal of Epidemiology & Community Health* 2007;61(2):159-64.
42. Wensink M, Alvarez J-A, Rizzi S, et al. Progression of the smoking epidemic in high-income regions and its effects on male-female survival differences: a cohort-by-age analysis of 17 countries. *BMC Public Health* 2020;20(1):1-8.
43. Borah G. Gender gap in life expectancy in India and role of age groups: A comparison between before and after male–female life expectancy at birth crossover. *PloS one* 2021;16(12):e0260657.
44. Canudas-Romo V, Saikia N. Gender Gap in Life Expectancy in India, 1970-2006: Institute of Economic Growth, University of Delhi 2013.
45. Aburto JM, Villavicencio F, Basellini U, et al. Dynamics of life expectancy and life span equality. *Proceedings of the National Academy of Sciences* 2020;117(10):5250-59.

46. Cui Q, Canudas-Romo V, Booth H. The Mechanism Underlying Change in the Sex Gap in Life Expectancy at Birth: An Extended Decomposition. *Demography* 2019;56(6):2307-21.
47. Andreev K. Sex differentials in survival in the Canadian population, 1921–1997: a descriptive analysis with focus on age-specific structure. *Demographic Research* 2000;3
48. Bergeron-Boucher M-P, Aburto JM, van Raalte A. Diversification in causes of death in low-mortality countries: emerging patterns and implications. *BMJ global health* 2020;5(7):e002414.
49. Alvarez J-A, Aburto JM, Canudas-Romo V. Latin American convergence and divergence towards the mortality profiles of developed countries. *Population studies* 2020;74(1):75-92.
50. García J, Aburto JM. The impact of violence on Venezuelan life expectancy and lifespan inequality. *International journal of epidemiology* 2019;48(5):1593-601.
51. Aburto JM, Beltrán-Sánchez H, García-Guerrero VM, et al. Homicides in Mexico reversed life expectancy gains for men and slowed them for women, 2000–10. *Health Affairs* 2016;35(1):88-95.
52. Lemaître J-F, Ronget V, Tidière M, et al. Sex differences in adult lifespan and aging rates of mortality across wild mammals. *Proceedings of the National Academy of Sciences* 2020;117(15):8546-53.

Figure legends

Figure 1 Four scenarios of interactions between lifespan distributions and corresponding statistics.

Figure 2 Probability of males outliving females (A) since 1850 for five countries and the range for all countries in the HMD in grey and (B) since 1950-1955 by World regions and the range for all countries in the WPP in grey.
Source: HMD [26], WPP [27] and authors' own calculations using Equation (2).

Figure 3 Probability of males outsurviving females across the World, 1950-1954 to 2015-2019.
Source: WPP [27] and authors' own calculations using Equation (2).

Figure 4 Relation between φ and (a) the sex differences in life expectancy and (b) the standard deviation of lifespans for females for HMD period data since 1751, with France highlighted (red triangles).
Source: HMD [26] and authors' own calculations.

Figure 5 Relation between φ and (a) the sex differences in life expectancy and (b) the standard deviation for females for HMD period data since 1751, conditional to survival to age 50, with France highlighted (red triangles).

Source: HMD [26] and authors' own calculations.

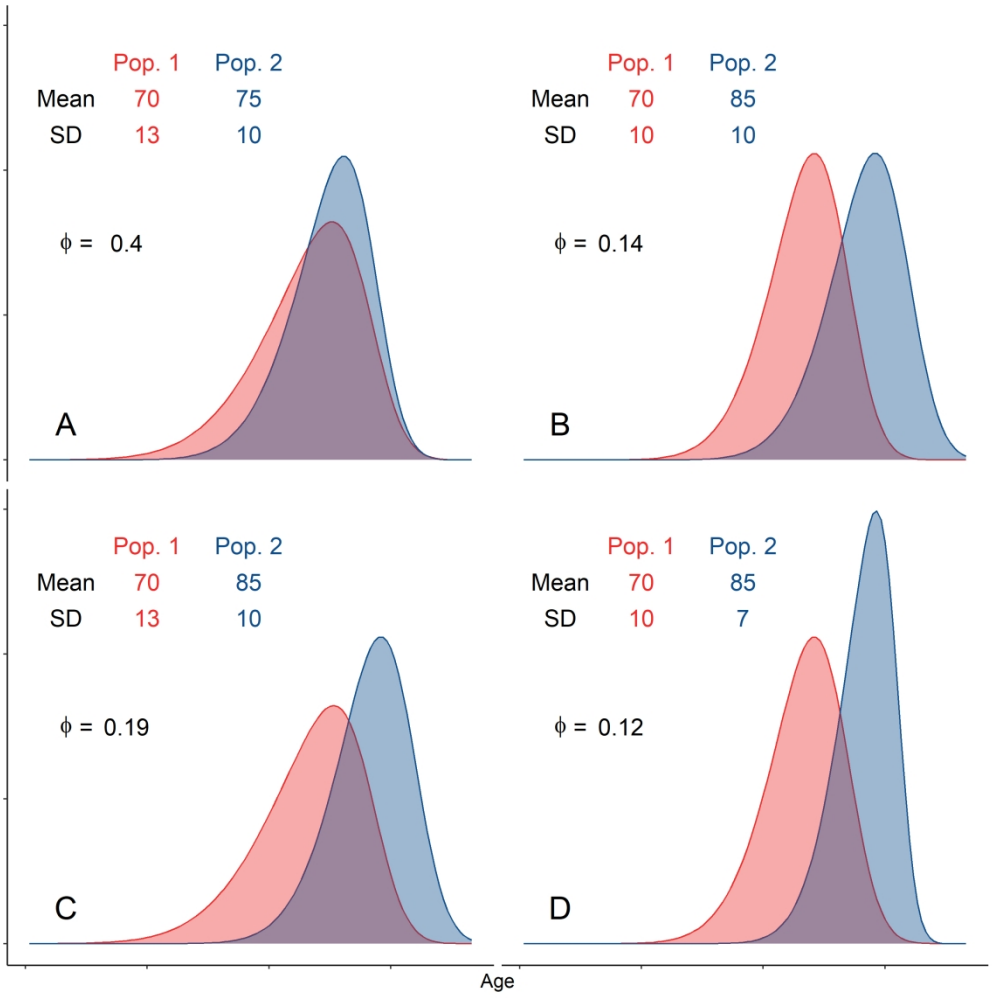

Figure 1 Four scenarios of interactions between lifespan distributions and corresponding statistics.

645x645mm (118 x 118 DPI)

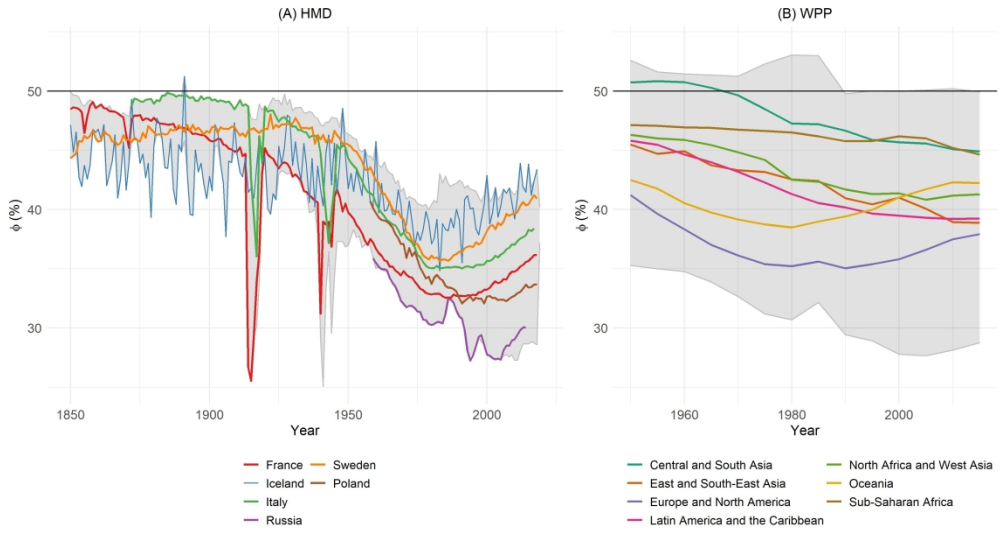

Figure 2 Probability of males outliving females (A) since 1850 for five countries and the range for all countries in the HMD in grey and (B) since 1950-1955 by World regions and the range for all countries in the WPP in grey.
Source: HMD [23], WPP [24] and authors' own calculations using Equation (2).

330x177mm (300 x 300 DPI)

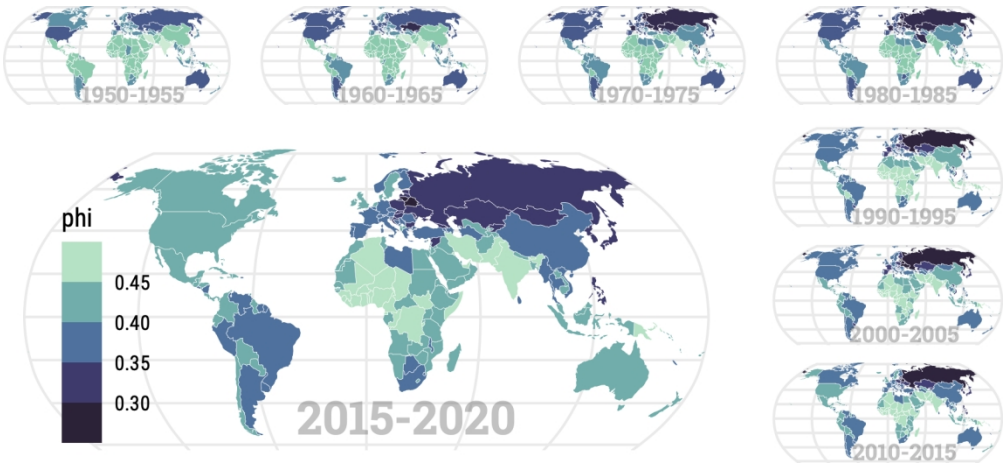

Figure 3 Probability of males outsurviving females across the World, 1950-1954 to 2015-2019. Source: WPP [24] and authors' own calculations using Equation (2).

151x69mm (300 x 300 DPI)

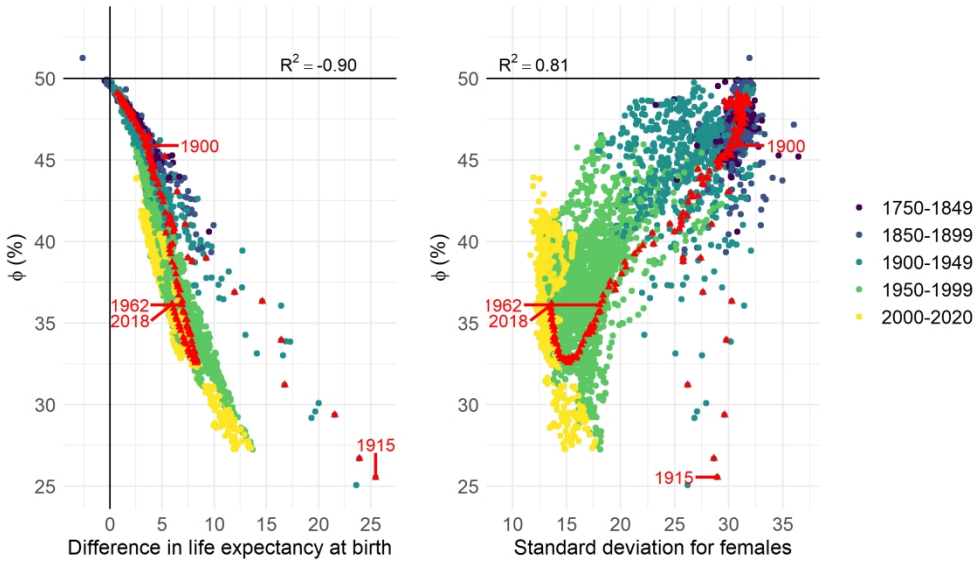

Figure 4 Relation between φ and (a) the sex differences in life expectancy and (b) the standard deviation of lifespans for females for HMD period data since 1751, with France highlighted (red triangles). Source: HMD [23] and authors' own calculations.

1761x1002mm (38 x 38 DPI)

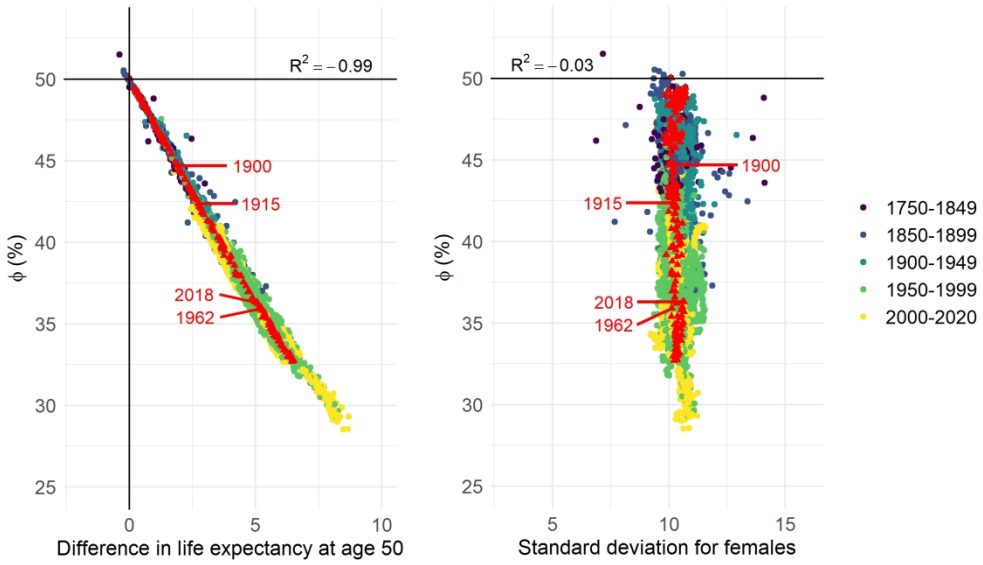

Figure 5 Relation between φ and (a) the sex differences in life expectancy and (b) the standard deviation for females for HMD period data since 1751, conditional to survival to age 50, with France highlighted (red triangles).
Source: HMD [23] and authors' own calculations.

1761x1002mm (38 x 38 DPI)

A. FORMAL RELATION

It can be shown that the outsurvival statistic relates to the joint probability density function of two lifespan distributions, which gives the probability of realizations of two lifespans and thus is related to the overlap of the two distributions. Assume two populations of individuals, with ages at death x and y, respectively. Assume the two populations are independent, meaning that the length of life x does not depend on the length of life y and vice versa. This implies that the joint probability density function, $d_{1,2}(x,y)$, equals the product of the marginal densities so that $d_{1,2}(x,y) = d_1(x)\, d_2(y)$. We are interested in calculating the probability (φ) of individuals in the first population outliving those in the second population. This implies that $0 \leq y < x$ so:

$$\begin{aligned} \varphi \quad &= \int_0^\infty \int_0^x d_{1,2}(x,y)dydx \;\; = \int_0^\infty d_1(x) \int_0^x d_2(y)dydx \\ &= \int_0^\infty d_1(x)D_2(x)dx \\ &= \int_0^\infty d_1(x)[1 - l_2(x)]dx \;\; = 1 - \int_0^\infty d_1(x)l_2(x)dx \\ &= \int_0^\infty d_2(x)\ell_1(x)dx. \end{aligned} \tag{A1}$$

Following the same approach, we can find the complement of φ, labeled φ', which is the probability of individuals in the second population to outlive those in the first:

$$\begin{aligned} \varphi' \quad &= \int_0^\infty \int_0^y d_{1,2}(x,y)dxdy \\ &= \int_0^\infty d_2(y) \int_0^y d_1(x)dxdy \\ &= \int_0^\infty d_2(y)D_1(y)dy \\ &= \int_0^\infty d_2(y)[1 - l_1(y)]\, dy \\ &= 1 - \int_0^\infty d_2(y)l_1(y)\, dy. \end{aligned} \tag{A2}$$

From Equations (B1) and (B2) it can be shown that $\varphi + \varphi' = 1$. Thus, φ is also equal to:

$$\begin{aligned} \varphi \quad &= 1 - \varphi' \\ &= 1 - \left[1 - \int_0^\omega d_2(x)l_1(x)\, dx\right] \\ &= \int_0^\omega d_2(x)l_1(x)\, dx. \end{aligned} \tag{A3}$$

B. SIMULATIONS AND DISCREATE APPROXIMATION

We simulated age at death distributions, using the Gompertz model, using various scale (M) and shape (β) parameters [1]. The distributions were first found using an age width (n) of 0.0001, after

which the data were aggregated within 1-year and 5-years age-groups. The probability that individuals in both population died within the same age-group, $\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2$, was then redistributed between φ and φ' based on two assumptions: equal (equation B1) and proportional redistributions (equation B2). The results are presented in Table B1.

$$\varphi \approx \sum_{x=0}^{\omega} {}_nd_x^1 \; {}_nD_{x-n}^2 + \frac{\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2}{2} \tag{B1}$$

$$\varphi \approx \sum_{x=0}^{\omega} {}_nd_x^1 \; {}_nD_{x-n}^2 + \sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2 \frac{{}_nd_x^1 \; {}_nD_{x-n}^2}{{}_nd_x^1 \; {}_nD_{x-n}^2 + \; {}_nd_x^2 \; {}_nD_{x-n}^1} \tag{B2}$$

The simulations show that equally redistributing $\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2$ between the two other statistics provide very similar results to the continuous data ($n=0.0001$), especially for the 1-year age-group. More differences are found when aggregating by 5-years age-groups, but the difference in φ between the different age-width remains less than 1 percentage point, when equally redistributing $\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2$.

Table B1. Assumptions to redistribute $\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2$ for different mortality scenarios.

	φ	φ'	$\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2$	Eq. B1	Eq. B2
Gompertz: $M_A = 61, M_B = 65, \; \beta_A = 0.12, \; \beta_B = 0.14$					
Continuous	**36.3**	63.7	0.0	-	-
1-year	34.8	62.2	3.0	**36.3**	35.9
5-years	28.2	55.8	15.0	**36.7**	34.3
Gompertz: $M_A = 61, M_B = 70, \; \beta_A = 0.10, \; \beta_B = 0.14$					
Continuous	**23.6**	76.4	0.0	-	-
1-year	22.5	75.2	2.3	**23.6**	23.0
5-years	18.5	70.0	11.3	**24.2**	20.9
Gompertz: $M_A = 68, M_B = 70, \; \beta_A = 0.13, \; \beta_B = 0.14$					
Continuous	**42.8**	57.2	0.0	-	-
1-year	41.2	55.5	3.3	**42.8**	42.6
5-years	34.9	48.8	16.3	**43.0**	41.7
Gompertz: $M_A = 69, M_B = 70, \; \beta_A = 0.10, \; \beta_B = 0.12$					
Continuous	**46.1**	53.9	0.0	-	-
1-year	44.7	52.6	2.7	**46.1**	46.0
5-years	39.4	47.2	13.4	**46.1**	45.5

To further test the model and the redistribution of $\sum_{x=0}^{\omega} {}_nd_x^1 \, {}_nd_x^2$, we simulated 100,000 individual lifespans from an exponential distribution with piece-wise constant rates [2]. We performed this procedure for every population and by sex using as an input empirical death rates retrieved from the HMD [3]. Then we randomly paired males and females and calculated the proportions of males outliving the paired female. Table 3 compares the discrete approach introduced in the main document (eq. B1) and the continuous approach based on simulations. Both approaches provided very similar results.

Table B2. Proportions of males outliving females based on a discrete and continuous approach (simulations).

	% males outliving females		% females outliving males	
	Continuous	Discrete	Continuous	Discrete
Denmark				
1850	46.1	46.4	53.9	53.6
1900	45.8	45.8	54.2	54.2
1950	46.2	46.3	53.8	53.7
2016	40.7	40.6	59.3	59.4
France				
1850	48.6	48.5	51.4	51.5
1915	25.6	25.5	74.4	74.5
1950	39.8	39.8	60.2	60.2
2016	35.9	36.0	64.1	64.0
Japan				
1950	44.1	44.2	55.9	55.8
2016	33.8	33.7	66.2	66.3
Russia				
1960	35.6	35.5	64.4	64.5
2014	30.2	30.0	69.8	70.0

C. DATA

Table C1. Countries/ regions and years with available data in the HMD

Country\region	**Years**	**Country\region**	**Years**
Australia	1921-2018	Japan	1947-2019
Austria	1947-2019	Latvia	1959-2019
Belarus	1959-2018	Lithuania	1959-2019
Belgium	1841-2018	Luxembourg	1960-2019
Bulgaria	1947-2017	Netherlands	1850-2019
Canada	1921-2018	New Zealand	1948-2013
Chile	1992-2017	Norway	1846-2020
Croatia	2001-2019	Poland	1958-2019
Czechia	1950-2019	Portugal	1940-2018
Denmark	1835-2020	Republic of Korea	2003-2018
Estonia	1959-2019	Russia	1959-2014
Finland	1878-2019	Slovakia	1950-2017
France	1816-2018	Slovenia	1983-2017
Germany-East	1956-2017	Spain	1908-2018
Germany-West	1956-2017	Sweden	1751-2019
Greece	1981-2017	Switzerland	1876-2018
Hong Kong	1986-2017	Taiwan	1970-2019
Hungary	1950-2017	UK – England and Wales	1841-2018
Iceland	1838-2018	UK- Scotland	1855-2018
Ireland	1950-2017	UK- Northern Ireland	1922-2018
Israel	1983-2016	USA	1933-2019
Italy	1872-2018	Ukraine	1959-2013

D. EXTRA ANALYSIS: COHORTS, WWP AND MALES' STANDARD DEVIATION

Similar relations as those for period data were also found for cohorts (Figure D1). In the HMD, life table for cohorts were only available for 11 countries: Denmark, England and Wales, Finland, France, Iceland, Italy, Netherlands, Norway, Scotland, Sweden and Switzerland. For cohorts with complete mortality history, the proportions of males outliving females varied between 35% and 49%. Only small changes in φ were observed for cohorts born prior to 1870-1890, with φ varying around 46.5%. For the cohorts born afterwards, φ decreases, reaching a mean of 38.4% for the cohort born in 1925, with values varying between 35.3% (Finland) and 40.4% (Scotland).

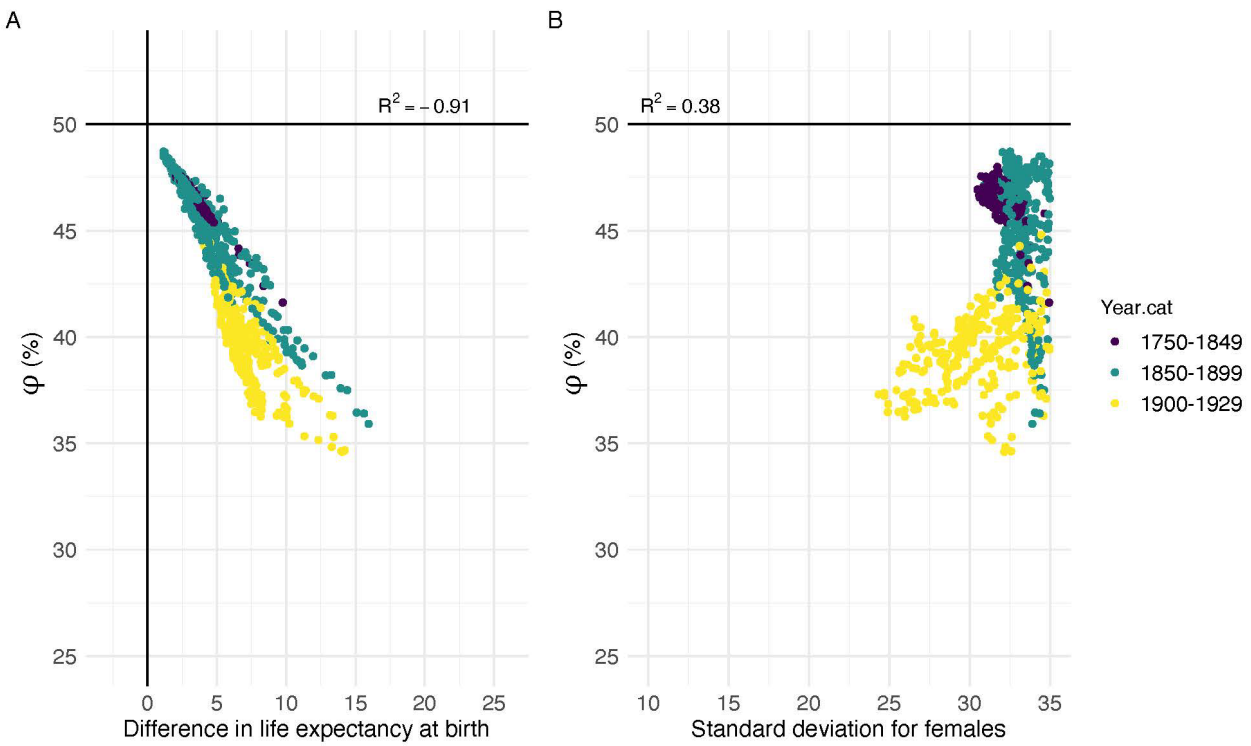

Figure D1 Relation between φ and (a) the sex differences in life expectancy and (b) the standard deviation for females for HMD cohort data.
Source: HMD [3] and authors' own calculations.

Figure D2 also shows the relation between φ and (A) the sex differences in life expectancy at birth and (B) the standard deviation for females for countries in the WWP from 1950-55 to 2015-19. The relation between φ and the two measures is similar to that shown in the main text using the HMD data.

Figure D3 shows the relation between (A) the relation between φ at the standard deviation of the lifespan distribution from birth for males and (B) the same relation, but conditional to survival to age 50. The relation between φ and the SD is similar whether we used the SD for females (as in the main text) or for males (Figure D3).

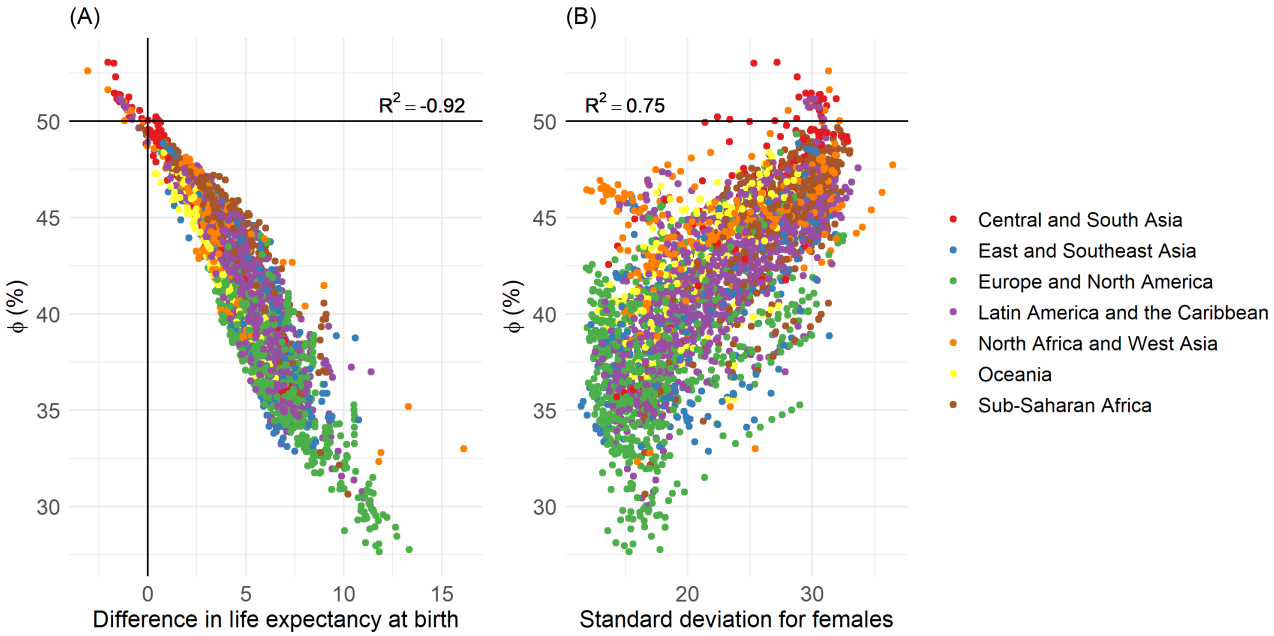

Figure D2 Relation between φ and (a) the sex differences in life expectancy and (b) the standard deviation for females for WPP data.
Source: WPP [4] and authors' own calculations.

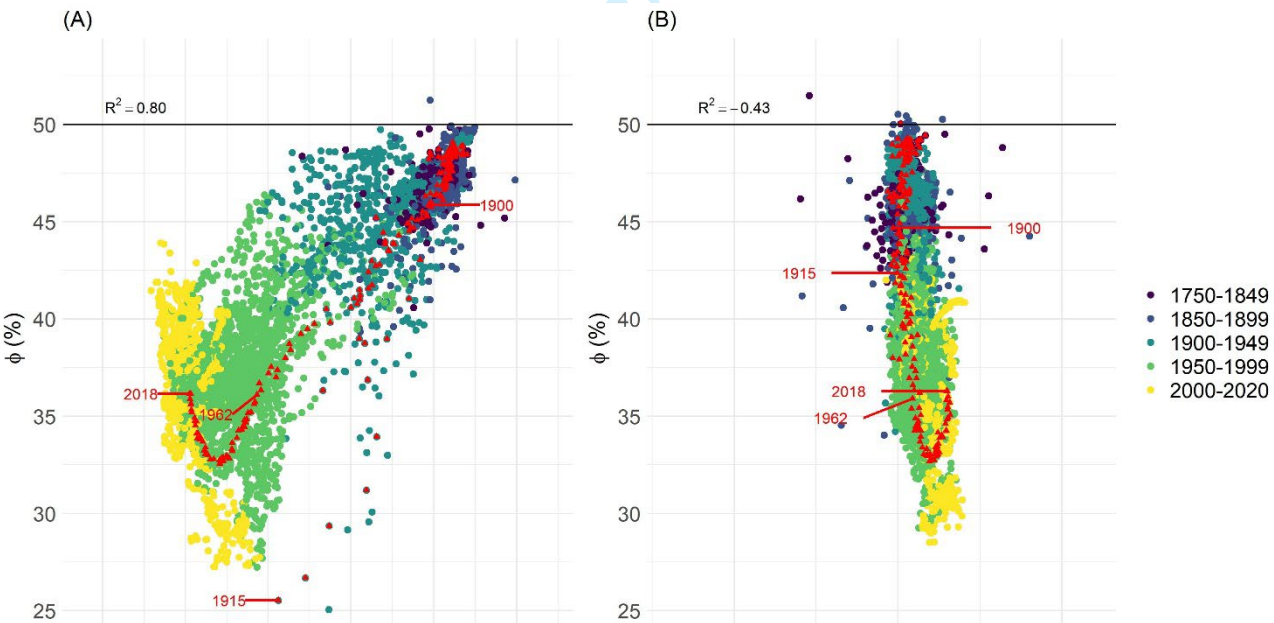

Figure D3 Relation between (a) φ and the standard deviation for males form birth and (b) φ and the standard deviation for males form age 50 for HMD period data.
Source: HMD [3] and authors' own calculations.

E. OTHER MEASURES OF OVERLAP

Figure E1 shows the relation between φ and the stratification index used by Shi et al. [5]. Both indicators are strongly correlated with a correlation coefficient of 0.98. Figure E2 shows a similar relation between φ and the Kullback-Leibler divergence, with a correlation coefficient of -0.93.

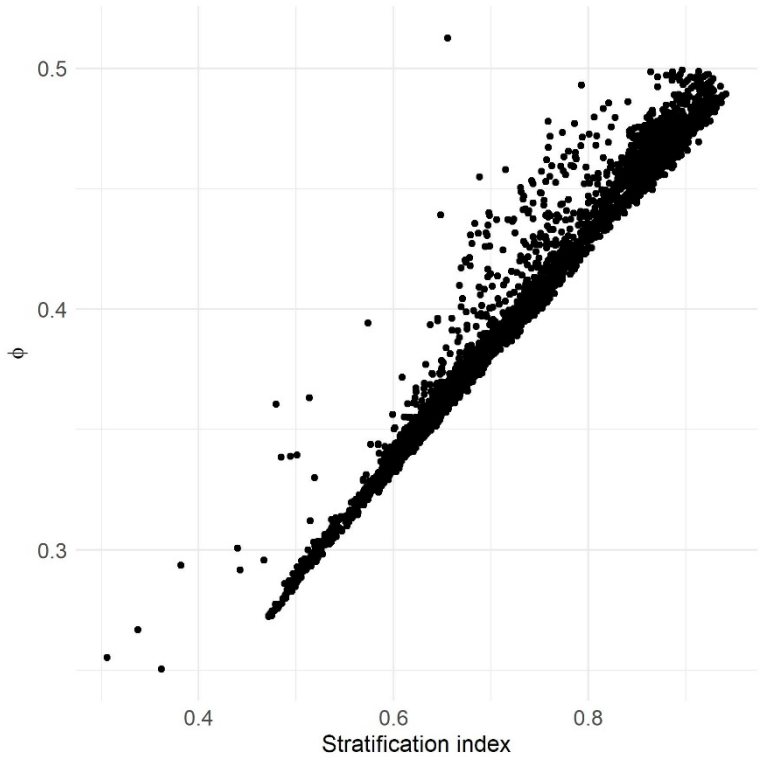

Figure E1. Relation between φ and the stratification index for HMD period data. Source: HMD [3] and authors' own calculations.

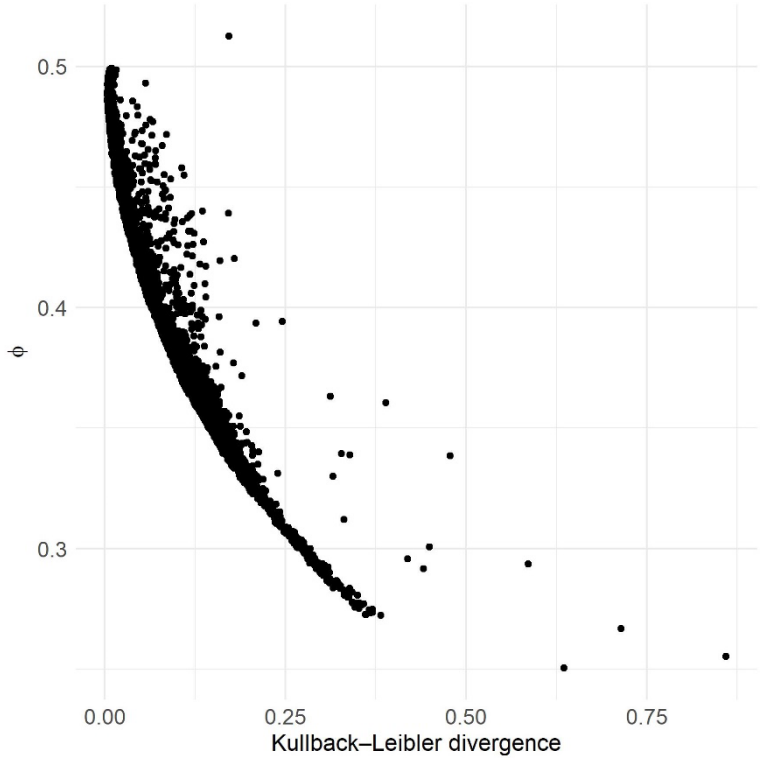

Figure E2. Relation between φ and the Kullback-Leibler divergence for HMD period data. Source: HMD [3] and authors' own calculations.

REFERENCES

1. Missov TI, Lenart A, Nemeth L, Canudas-Romo V, Vaupel JW: **The Gompertz force of mortality in terms of the modal age at death**. *Demographic Research* 2015, **32**:1031-1048.
2. Willekens F: **Continuous-time microsimulation in longitudinal analysis**. In: *New frontiers in microsimulation modelling.* edn. Edited by Harding A, Zaidi A. London: Routledge; 2009: 353-376.
3. HMD: **Human Mortality Database**. In.: University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany), Available at www.mortality.org (Accessed on July 15) . 2021.
4. United Nations: **World Population Prospects 2019, Online Edition. Rev. 1.** In.: Department of Economic and Social Affairs, Population Division, Available at https://population.un.org/wpp/Download/Standard/Mortality/ (Accessed on May 21); 2019.
5. Shi J, Aburto JM, Martikainen P, Tarkiainen L, van Raalte AA: **A distributional approach to measuring lifespan stratification**. *Population Studies* 2022, **(Accepted)**.

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract Reported on p. 1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found p.1
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported p.2-3
Objectives	3	State specific objectives, including any prespecified hypotheses p.3-4
Methods		
Study design	4	Present key elements of study design early in the paper Observational studies. p.4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection Publicly available database. P.6-7
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants Not applicable
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable p. 6-7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group p.6-7
Bias	9	Describe any efforts to address potential sources of bias None
Study size	10	Explain how the study size was arrived at Full population used. P. 6-7
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why p.4-7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding p.4-6
		(*b*) Describe any methods used to examine subgroups and interactions p.4-6 (same methods used for subgroups as for the total population)
		(*c*) Explain how missing data were addressed Handled within the publicly available databases.
		(*d*) If applicable, describe analytical methods taking account of sampling strategy Not applicable
		(*e*) Describe any sensitivity analyses p. 4-5
Results		

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed Not applicable
		(b) Give reasons for non-participation at each stage Not applicable
		(c) Consider use of a flow diagram Not applicable
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders Not applicable
		(b) Indicate number of participants with missing data for each variable of interest Not applicable
Outcome data	15*	Report numbers of outcome events or summary measures Not applicable
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included p. 7-8
		(*b*) Report category boundaries when continuous variables were categorized Not applicable
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period Not applicable
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses p. 8-9
Discussion		
Key results	18	Summarise key results with reference to study objectives p.10
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias p. 10-11
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence p.10-11
Generalisability	21	Discuss the generalisability (external validity) of the study results p.11-12
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based p.13

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely

available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.