BMJ Open

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

Marie-Pier Bergeron-Boucher , Jesús-Adrian Alvarez , Ilya Kashnitsky , Virginia Zarulli

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 0, 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.

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. For example, males tend to engage in more risky behaviours, such as smoking and heavy drinking, but oestrogen could also be protective against certain diseases. 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. 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. In 2019, the sex difference in life expectancy was 4.4 years on average worldwide, with large variation across countries. Females have been found to have longer survival and lower death rates than men at all ages and in most modern populations and even under extreme mortality conditions.

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, that is, 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, that is, differences in lifespans within a population, has been receiving increasing attention in the literature. 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. It has been shown that females systematically experience lower lifespan variation than males. 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. 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. 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 low life expectancy outlives an individual from another population with higher life expectancy. 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 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.

METHOD

Outsurvival statistic

Consider two populations with mean and standard deviation (SD) of the age of death (see Ouellette and Bourbeau, and Wilmoth and Horiuchi 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, \( \phi \), 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. 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 \( l_i(x) = 1 - D_i(x) \). The probability that an individual from the first population (males) will outlive an individual from the second population (females) is

\[
\varphi = \int_0^\infty d_2(x) l_1(x) \, dx. \tag{1}
\]

In scenario A of figure 1, \( \varphi \) is 40%.

In the online supplemental 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.
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, that is, 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, \( \varphi \) is higher in scenario C (19%) than in scenario B (14%). Now compare scenario B with 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 \( \varphi \) to 12%. Thus, for the same difference in life expectancy, larger lifespan variation in both populations generally results in larger \( \varphi \). The comparison of scenarios A and C also shows that small differences in life expectancy lead to larger value of \( \varphi \).

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 model. The latter assesses the probability that the stress (population 1) exceeds the strength (population 2) of a material. If the distributions of both populations follow a normal distribution with mean \( \mu_i \) and SD \( s_i \), the probability of failure is \( P(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} \, dz \). This relation formalises what is illustrated in section in figure 1: \( \varphi \) 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 \( \varphi \). To better understand this relation, we analysed the correlation between \( \varphi \) and life expectancy as well as between \( \varphi \) and lifespan variation.

Discrete approximation

Similar equivalences to equation (1) can be developed in a discrete time setting. Let \( d_{i, x} \) be the life table deaths between age \( x \) and \( x+n \) in population \( i \), and \( d_{i, x}^i \) 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:

\[
\varphi \approx \sum_{x=0}^{\infty} \left( \sum_{n=0}^{\infty} d_{i, x+n} d_{i, x}^i \right) + d,
\]

where \( d = \sum_{n=0}^{\infty} d_{i, x} - \sum_{n=0}^{\infty} d_{i, x}^i \) and \( \sum_{x=0}^{\infty} n d_{i, x} d_{i, x}^i \) 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 \( \lim_{n \to 0} \sum_{n=0}^{\infty} n d_{i, x} d_{i, x}^i = 0 \). In the online supplemental 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.

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). The HMD is freely available and 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 UK, 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 online supplemental 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). 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. 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 and population counts from the American Community Service (ACS) from the US Census Bureau. 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 to obtain the population counts from age 90 to 110 by single year of age.
RESULTS

Historical values and trends in $\phi$

Figure 2A shows the outsurvival probability of males over females ($\phi$) since 1850 for all HMD countries and figure 2B for all WPP 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.

For the HMD countries, $\phi$ was slowly decreasing before the First World War, on average from 47.3% in 1850 to 46.0% in 1913. After the war, $\phi$ declined faster. In 1930, the mean $\phi$ across populations was 45.4%, ranging from 42.8% (France) to 48.4% (the Netherlands). By 1985, the mean $\phi$ was 35.3%, ranging from 31.2% (Russia) to 42.8% (Israel). The value of $\phi$ 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 WPP countries, we observed a decrease in $\phi$ 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, $\phi$ was 46.1% on average worldwide, with values ranging between 35.3% (in Kazakhstan) and 52.6% (in Iran). By 2015–2019, $\phi$ declined to 41.2% with values

![Figure 3](image-url)
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.

**Life expectancy and lifespan inequality correlation**

Figure 4 shows that $φ$ is negatively correlated with the differences in life expectancy and positively correlated with females’ SD (as shown in the online supplemental 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 online supplemental materials). This relation empirically demonstrates the formal relation in the Relation to life expectancy and lifespan variation section. The correlation between $φ$ and the SD 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 online supplemental materials).

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 SD 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, 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 online supplemental 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.

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.

**Sex differences by education and marital status**

Tables 1 and 2 show the $φ$ statistic for some subpopulations of males and females in the USA. 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).

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).
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 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 \( \phi \) 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 \( \phi \) 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 \( \phi \).

Other measures of overlap and distance between distributions could have been used. In the online supplemental materials, we compare the outsurvival statistic with a stratification index used by Shi and colleagues20 and the KL divergence. We found that all three indicators are

![Figure 5](relation-between-\( \phi \) and (A) the sex differences in life expectancy and (B) the SD for females for HMD period data since 1751, conditional to survival to age 50 years, with France highlighted (red triangles). Source: HMD26 and authors’ own calculations. HMD, Human Mortality Database.)

Table 1

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Married</td>
<td>Unmarried</td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td>0.39</td>
<td>0.52</td>
</tr>
<tr>
<td>Unmarried</td>
<td></td>
<td>0.26</td>
<td>0.37</td>
</tr>
</tbody>
</table>

Source: MCDD,29 ACS30 and authors’ own calculations using equation 2.

ACS, American Community Service; MCDD, Multiple Cause of Death Dataset.
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: 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, which is consistent with our analysis of φ (see online supplemental materials). The increase and decrease in sex differences in life expectancy were mainly attributed to the smoking epidemic and other behavioural differences between sexes.

The φ values are generally higher in low/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 5 was higher for females than males and remained higher for females in recent years. However, females had a growing mortality advantage above age 15 years 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. 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 dying at a faster pace for females than males below age 50, especially in the first half of the 20th century. 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.

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. Another example is

### Table 2 Outsurvival statistics by sex and education level in the USA, 2015–2019

<table>
<thead>
<tr>
<th></th>
<th>University degree</th>
<th>High school diploma</th>
<th>No high school diploma</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Male</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>University</td>
<td>0.43</td>
<td>0.51</td>
<td>0.53</td>
</tr>
<tr>
<td>High school</td>
<td>0.32</td>
<td>0.39</td>
<td>0.42</td>
</tr>
<tr>
<td>No high school</td>
<td>0.30</td>
<td>0.37</td>
<td>0.39</td>
</tr>
</tbody>
</table>

Source: MCDD, ACS and authors’ own calculations using equation 2. ACS, American Community Service; MCDD, Multiple Cause of Death Dataset.
observed in Latin American populations, where homicides and violent deaths have had an increased burden among males in comparison with females since the 1990s. In Mexico, for example, the increase in homicide mortality, especially among men between 20 and 40 years, contributed to increasing the gap in mortality between females and males. 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, 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.

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.

Twitter Marie-Pier Bergeron-Boucher @bergeron_mp, Jesús-Adrian Alvarez @jssalvry, Ilya Kashnitsky @I_kashnitsky and Virginia Zarulli @VZarulli

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

Contributors M-PB-B designed and conceptualised the study, M-PB-B, J-AA and IK produced the results and did the analysis. M-PB-B, J-AA, IK and VZ contributed to the interpretation of the results; and the drafting, revision and approval of the manuscript. M-PB-B is acting as guarantor.

Funding 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).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. The data are publicly available at http://www.mortality.org and https://population.un.org/wpp/. The R code to replicate the calculations and the figures is openly available in Github https://github.com/Cpop-SDU/outsurvival-in-perspective

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs Marie-Pier Bergeron-Boucher http://orcid.org/0000-0001-7383-3175
Jésus-Adrian Alvarez http://orcid.org/0000-0002-3724-6149
Ilya Kashnitsky http://orcid.org/0000-0003-1835-8667
Virginia Zarulli http://orcid.org/0000-0003-3219-4698

REFERENCES


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 ($\varphi$) of individuals in the first population outliving those in the second population. This implies that $0 \leq y < x$ so:

$$
\varphi = \int_0^\infty \int_0^x d_{1,2}(x,y) dydx = \int_0^\infty d_1(x) \int_0^x d_2(y) dydx
$$

$$
= \int_0^\infty d_1(x) D_2(x) dx
$$

$$
= \int_0^\infty d_1(x)[1 - l_2(x)] dx = 1 - \int_0^\infty d_1(x) l_2(x) dx
$$

$$
= \int_0^\infty d_2(x) l_1(x) dx.
$$

(A1)

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

$$
\varphi' = \int_0^\infty \int_0^y d_{1,2}(x,y) dx dy
$$

$$
= \int_0^\infty d_2(y) \int_0^y d_1(x) dx dy
$$

$$
= \int_0^\infty d_2(y) D_1(y) dy
$$

$$
= \int_0^\infty d_2(y)[1 - l_1(y)] dy
$$

$$
= 1 - \int_0^\infty d_2(y) l_1(y) dy.
$$

(A2)

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

$$
\varphi = 1 - \varphi' = 1 - \left[ 1 - \int_0^\infty d_2(x) l_1(x) dx \right]
$$

$$
= \int_0^\infty d_2(x) l_1(x) dx.
$$

(A3)

B. SIMULATIONS AND DISCRETE APPROXIMATION

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

1
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, \( \sum_{x=0}^{\omega} n d_x^1 n d_x^2 \), was then redistributed between \( \varphi \) and \( \varphi' \) based on two assumptions: equal (equation B1) and proportional redistributions (equation B2). The results are presented in Table B1.

\[
\varphi \approx \sum_{x=0}^{\omega} n d_x^1 n D_x^2-n + \frac{\sum_{x=0}^{\omega} n d_x^1 n d_x^2}{2}
\]

(B1)

\[
\varphi \approx \sum_{x=0}^{\omega} n d_x^1 n D_x^2-n + \sum_{x=0}^{\omega} n d_x^1 n d_x^2 \frac{n d_x^1 n D_x^2-n}{n d_x^1 n D_x^2-n + n d_x^1 n D_x^2-n}
\]

(B2)

The simulations show that equally redistributing \( \sum_{x=0}^{\omega} n d_x^1 n d_x^2 \) 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 \( \varphi \) between the different age-width remains less than 1 percentage point, when equally redistributing \( \sum_{x=0}^{\omega} n d_x^1 n d_x^2 \).

| Table B1. Assumptions to redistribute \( \sum_{x=0}^{\omega} n d_x^1 n d_x^2 \) for different mortality scenarios. |
| ---  | ---  | ---  | ---  | ---  |
| \( \varphi \) | \( \varphi' \) | \( \sum_{x=0}^{\omega} n d_x^1 n d_x^2 \) | 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 \( \sum_{x=0}^{\omega} n d_x^1 n d_x^2 \), 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).

<table>
<thead>
<tr>
<th></th>
<th>% males outliving females</th>
<th>% females outliving males</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Continuous</td>
<td>Discrete</td>
</tr>
<tr>
<td>Denmark</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1850</td>
<td>46.1</td>
<td>46.4</td>
</tr>
<tr>
<td>1900</td>
<td>45.8</td>
<td>45.8</td>
</tr>
<tr>
<td>1950</td>
<td>46.2</td>
<td>46.3</td>
</tr>
<tr>
<td>2016</td>
<td>40.7</td>
<td>40.6</td>
</tr>
<tr>
<td>France</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1850</td>
<td>48.6</td>
<td>48.5</td>
</tr>
<tr>
<td>1915</td>
<td>25.6</td>
<td>25.5</td>
</tr>
<tr>
<td>1950</td>
<td>39.8</td>
<td>39.8</td>
</tr>
<tr>
<td>2016</td>
<td>35.9</td>
<td>36.0</td>
</tr>
<tr>
<td>Japan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1950</td>
<td>44.1</td>
<td>44.2</td>
</tr>
<tr>
<td>2016</td>
<td>33.8</td>
<td>33.7</td>
</tr>
<tr>
<td>Russia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1960</td>
<td>35.6</td>
<td>35.5</td>
</tr>
<tr>
<td>2014</td>
<td>30.2</td>
<td>30.0</td>
</tr>
</tbody>
</table>

C. DATA

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

<table>
<thead>
<tr>
<th>Country/region</th>
<th>Years</th>
<th>Country/region</th>
<th>Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>1921-2018</td>
<td>Japan</td>
<td>1947-2019</td>
</tr>
<tr>
<td>Austria</td>
<td>1947-2019</td>
<td>Latvia</td>
<td>1959-2019</td>
</tr>
<tr>
<td>Belarus</td>
<td>1959-2018</td>
<td>Lithuania</td>
<td>1959-2019</td>
</tr>
<tr>
<td>Belgium</td>
<td>1841-2018</td>
<td>Luxembourg</td>
<td>1960-2019</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>1947-2017</td>
<td>Netherlands</td>
<td>1850-2019</td>
</tr>
<tr>
<td>Canada</td>
<td>1921-2018</td>
<td>New Zealand</td>
<td>1948-2013</td>
</tr>
<tr>
<td>Chile</td>
<td>1992-2017</td>
<td>Norway</td>
<td>1846-2020</td>
</tr>
<tr>
<td>Czechia</td>
<td>1950-2019</td>
<td>Portugal</td>
<td>1940-2018</td>
</tr>
<tr>
<td>Denmark</td>
<td>1835-2020</td>
<td>Republic of Korea</td>
<td>2003-2018</td>
</tr>
<tr>
<td>Estonia</td>
<td>1959-2019</td>
<td>Russia</td>
<td>1959-2014</td>
</tr>
<tr>
<td>Finland</td>
<td>1878-2019</td>
<td>Slovakia</td>
<td>1950-2017</td>
</tr>
<tr>
<td>France</td>
<td>1816-2018</td>
<td>Slovenia</td>
<td>1983-2017</td>
</tr>
<tr>
<td>Germany-East</td>
<td>1956-2017</td>
<td>Spain</td>
<td>1908-2018</td>
</tr>
<tr>
<td>Germany-West</td>
<td>1956-2017</td>
<td>Sweden</td>
<td>1751-2019</td>
</tr>
<tr>
<td>Greece</td>
<td>1981-2017</td>
<td>Switzerland</td>
<td>1876-2018</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>1986-2017</td>
<td>Taiwan</td>
<td>1970-2019</td>
</tr>
<tr>
<td>Iceland</td>
<td>1838-2018</td>
<td>UK- Scotland</td>
<td>1855-2018</td>
</tr>
<tr>
<td>Ireland</td>
<td>1950-2017</td>
<td>UK- Northern Ireland</td>
<td>1922-2018</td>
</tr>
<tr>
<td>Israel</td>
<td>1983-2016</td>
<td>USA</td>
<td>1933-2019</td>
</tr>
<tr>
<td>Italy</td>
<td>1872-2018</td>
<td>Ukraine</td>
<td>1959-2013</td>
</tr>
</tbody>
</table>
**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 $\phi$ were observed for cohorts born prior to 1870-1890, with $\phi$ varying around 46.5%. For the cohorts born afterwards, $\phi$ 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](image)

*Figure D1* Relation between $\phi$ 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 $\phi$ 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 $\phi$ 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 $\phi$ 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 $\phi$ 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 $\varphi$ 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) $\varphi$ and the standard deviation for males form birth and (b) $\varphi$ 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 $\phi$ 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 $\phi$ and the Kullback-Leibler divergence, with a correlation coefficient of -0.93.

![Figure E1](image)

**Figure E1.** Relation between $\phi$ and the stratification index for HMD period data. Source: HMD [3] and authors’ own calculations.
Figure E2. Relation between $\varphi$ and the Kullback-Leibler divergence for HMD period data. Source: HMD [3] and authors’ own calculations.

REFERENCES


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.
