Comparison of Lifetable Series by SIS Deciles Using Two Different Methodological Approaches

NOVEMBER | 2024

Mortality and Longevity

Comparison of Lifetable Series by SIS Deciles Using Two Different Methodological Approaches

AUTHOR Magali Barbieri, FSA, MAAA
Co-Director, Human Mortality Database
Director, U.S. Mortality DataBase
University of California, Berkeley

SPONSOR Mortality and Longevity Strategic Research Program

Diversity Equity and Inclusion Strategic Research Program

Financial Reporting Section
Product Development Section
Reinsurance Section
Smaller Insurance Company Section

CONTENTS

Comparison of Lifetable Series by SIS Deciles Using Two Different Methodological Approaches

Two series of lifetables have been created by socioeconomic (SIS) decile for years 1982-2022 using 1) classic demographic techniques and 2) the method protocol of the Human Mortality Database (HMD). This report compares the results and explains differences in the estimation methods. There is a wide overlap between the two approaches since the HMD relies on many of the classic demographic methods, but differences are significant at ages 80 years and over as further explained below.

In either case, the first step in the construction of lifetables is to use basic demographic statistics to calculate mortality rates, from which all other lifetable functions are derived. Both approaches rely on mortality data from the U.S. vital statistics system distributed by the National Center for Health Statistics (NCHS) and population estimates from the Census Bureau. In addition, the HMD methods protocol requires birth data, also distributed by the NCHS. More information on the sources of these data is provided in Appendix 1 of this report.

Before presenting the results of the comparisons between the results of these two different approaches, the next sections describe how the mortality rates are calculated following each approach. In both cases, the demographic data (deaths, population, and, for HMD methods only, births) are first aggregated over all counties in the same SIS deciles as previously defined (see Barbieri, 2022 for additional information). All counties are allocated to one of the 10 socioeconomic deciles based on their socioeconomic characteristics as of the 2000 census (Barbieri, 2022).

Section 1: Classic Demographic Methods

The classic demographic methods (see for instance Shryock, Siegel, and Associates, 2013, or Wachter, 2014) involve the following calculation steps in the construction of a complete period lifetable. The lifetables describe the mortality experience of a fictional cohort of individuals who would have experienced throughout their entire life the mortality conditions of the calendar year of interest. The lifetables include the following functions:

1. Mortality rates (${}_nm_x$): the ratio of the number of deaths to the population counts in each sex and age interval *x* to (*x+n*), where n is the length of the interval (typically one or five years).
2. Probabilities of dying (${}_nq_x$): the proportion of all individuals in the virtual cohort alive at the beginning of each age interval who die during the age interval, i.e., during the interval *x* to (*x+n*).
3. Survivors (l_x): the proportion of the virtual cohort still alive at age *x*. Typically, the root of the lifetable survivorship function is a multiple of 10 (often 100,000).
4. Deaths (${}_nd_x$): the number of individuals in the virtual cohort who died between age x and age (*x+n*).
5. Person-years lived (exposures, ${}_nL_x$): the number of years lived in the interval *x* to (*x+n*) by individuals in the virtual cohort who have survived to age x.
6. Total number of person-years lived after age x(T_x): the sum of remaining years of life lived by all individuals surviving to age x.
7. Remaining life expectancy (e_x): the average number of years lived by individuals in the virtual cohort surviving to age *x*.

Note that lifetables come in two typical flavors: detailed lifetables provide the mortality indicators for each single year of age while abridged lifetables provide the indicators by five-year age groups, up to an open age interval. For the SOA project, lifetables have been constructed by single year of age, by five-year age groups (distinguishing however between the first year of life and the following four years), and by ten-year age groups (again, with the exception of the first age groups, 0 and 1-9 years). Since the functions l_x, T_x, and e_x pertain to exact age x, they do not vary depending on the type of lifetables (i.e., the expectation of life at exact age *x* is the same in the single year, five-year, or ten-year age group lifetables).

The exact formulas for the calculation of each of these lifetable functions are as follows:

$${}_nm_x = {}_nD_x / {}_nP_x \tag{1}$$

where ${}_nD_x$ is the number of deaths in the age interval *x* to (*x+n*) and ${}_nP_x$ is the July 1st population in the same age interval.

$${}_nq_x = (n * {}_nm_x)/(1 + (n- {}_na_x)\, {}_nm_x \tag{2}$$

where ${}_na_x$ represents the average number of years lived by people who died in the age interval *x* to (*x+n*). For the first interval (the first year of life), ${}_na_x$ is typically close to zero in low mortality populations since most children dying before reaching their first birthday, do so in the first few weeks after birth. For all other age groups, these values are typically set at half of the length of the age interval (0.5 for the single year of age lifetables, 2.5 for the five-year age group lifetables—except for the age interval 1-4 years when ${}_1a_4$ is set to 2.0). For the last age interval, ${}_\omega q_x$ is equal to one since everyone dies. For the last (open) age interval, ${}_\omega a_x$ is typically estimated as follows:

$${}_\omega a_x = 1/{}_\omega m_x \tag{3}$$

where ω represents the highest age in the lifetable.

$$l_x = l_{(x-n)} * (1 - {}_nq_x) \quad (4)$$

$$ {}_nd_x = l_{(x-n)} - {}_nl_x \quad (5)$$

$$ {}_nL_x = (n * l_{(x+n)}) + ({}_na_x * {}_nd_x) \quad (6)$$

$$T_x = \sum_x^{\omega} nLx \quad (7)$$

$$e_x = T_x / l_x \quad (8)$$

The above steps have strictly been followed to compute the lifetables for each SIS decile and each calendar year. Note the following points, specific to this SOA project:

- The death tabulations can be prepared up to the maximum age at which individuals have died in any calendar year; however, the population estimates are only provided by the Census Bureau at the county level up to an open age interval of 85 and over. Consequently, the lifetables calculated following classic demographic methods end at 85 and over.
- The population data for the 1980s are only available from the Census Bureau by five-year age groups (i.e., ages 0, 1-4, 5-9, 10-14, etc., up to 85+ years) so the series of lifetables by single year of age only start in 1990. The lifetables by five- and 10-year age groups for years 1990 and after were calculated directly from the single year of age lifetables to be as precise as possible. For the years 1982-1989, the five- and 10-year lifetable series were calculated directly by five-year age group.
- The ${}_0a_1$ were extracted from the HMD U.S. lifetables, thus assuming constancy across all SIS deciles. Note that these values are relatively stable (at 0.13 or 0.14 for the whole period 1982-2022 depending on the calendar year and the sex).

Section 2: HMD Methods

HMD methods fundamentally rely on the same overall concepts as classic demographic methods, but they also include some special techniques designed to overcome the shortcoming of historical demographic data. Such data are often of much lower quality than contemporary data and they are typically less detailed. The HMD protocol thus leverages specific techniques designed by demographers to deal with these issues and to make the HMD lifetable as comparable as possible over time and across countries. In particular:

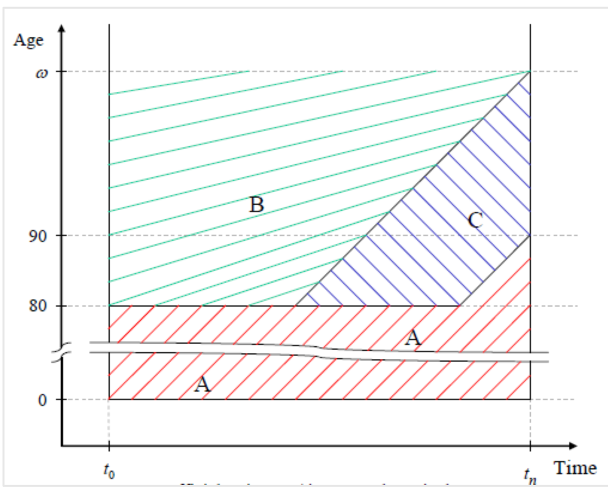
- Before the mortality rates are calculated, deaths are split by Lexis triangle (i.e., cross tabulated by year of birth and age at last birthday) using a couple of different methods, depending on the age considered (see Wilmoth et al., 2021, pp.12-14).
- A special formula has been designed to compute some of the ${}_na_x$ values as demographers have improved on the initial approach as described in the previous section (see Wilmoth et al., 2021, pp.36-38).
- To reduce the impact of the large year-to-year fluctuations in the mortality rates at very high ages (95 years and above), the lifetables are not constructed from the raw rates but from smoothed rates, using the Kannisto method (see Wilmoth et al., 2021, pp.34-36).
- For all ages below 80 years, exposure counts (the denominator of the rates, i.e., ${}_nP_x$ in formula (1) above) are estimated based on the population estimates but also on the births by month and sex to account for the fact that births are not necessarily uniformly distributed over the calendar year, an assumption inherent to the classic method to estimate the mortality rates. This is not so

relevant for the United States, but it is essential to avoid "false cohort effects" in the HMD countries which have experienced the brunt of WWI and WWII, with massive drops in the number of conceptions at the wars beginning and massive increases at the end. See chapter five, pp. 15-28 in Wilmoth et al., 2021, for detailed explanations of the methods implemented.

- For ages 80 years and above, exposure counts are not indirectly estimated to account for the lack of detailed data in many historical populations and the lower quality of the information at these high ages with the common occurrence of age overstatements. At these higher ages, three different methods are implemented, depending on the cohort. For all the cohorts that can presumably be considered as extinct, the extinct cohort method is implemented (see area B in Figure 1). For slightly younger cohorts, the survival ratio method is implemented (area C in Figure 1) and for cohorts for which the highest age in the most recent data year is below 90 years, the cohort-component method is used (area A in Figure 1).

Figure 1

HMD METHODS FOR ESTIMATING POPULATIONS AT AGES 80 YEARS AND ABOVE



Source: Wilmoth et al., 2021, p. 26.

Note: A—Cohort component method
B—Extinct cohort method
C—Survivor ratio method

Of all these methods, the last set, i.e., the indirect estimation of population exposures at ages 80 years and above, creates the largest discrepancies in the lifetables calculated with classic demographic methods.

Section 3: Advantages and Disadvantages of Each Approach

The advantages of using classic demographic methods are the relative simplicity of the calculation steps and the fact that extending the data series when new mortality and population data become available do not modify the lifetable values for calendar years with no new data. For instance, when the SIS lifetable series were extended to include year 2022, the new 2022 mortality file from the NCHS and the revised post-censal population estimates from the Census Bureau for years 2020 to 2023 were used. Consequently, the lifetable values previously calculated for years up to 2019 remained exactly the same as before. By contrast, because of the way population exposures at ages 80 years and above are calculated using HMD methods, each new set of mortality data modifies the mortality values at all ages 80 years and above for all non-extinct cohorts (i.e., for all lifetables in the 25-30 years or so before this last data point).

The disadvantage of using classic demographic methods is the impossibility to calculate detailed lifetables in years when either the mortality or (in the U.S. case for calendar years 1982-1989) population data are not provided in single years of age. In addition, the lifetables have to close in the lowest open age-interval for which mortality or population data are available (85+ years in the U.S. case) while HMD lifetables are presented up to an open age interval of 110 years and above. Finally, the tendency for the oldest cohorts in the U.S. population to over-state their ages (initially documented in Coale and Kisker, 1990, and in Kannisto, 1988), is not accounted for with classic demographic methods, hence the HMD idea to indirectly estimate mortality at ages 80 years and above. However, to be fair, recent studies suggest that the Kannisto model implemented to smooth rates at very high ages in the HMD might not be the most appropriate as it tends to under-estimate mortality by imposing an asymptote at one, which might be too constraining (Camarda, 2022; Dang, 2022; Kannisto, 1988).

Section 4: A Comparison of Lifetable Values in the Two Series

This section presents a comparison of the lifetables in the series calculated using classic (standard) demographic methods and in those calculated using the HMD method protocol, more specifically looking at the expectation of life at birth and at age 80 years (the age at which the HMD methods deviate substantially from classic demographic methods) as well as the proportion surviving to age 80 years. The results are illustrated by Figures 2a-4b below.

Figures 2a and 2b show some small differences in the expectation of life at birth, with lower values overall in the lifetables constructed with HMD methods compared with those constructed with standard methods. The differences are relatively small at the beginning of the time period but, starting around 1990, the gap progressively increases until 2019, after which it reverses and values from standard methods become smaller than those from HMD methods. This reversal is explained by the mix of methods used in the HMD (cohort-component methods for the most recent time period versus survival ratio methods and extinct cohort methods for earlier years) and the way in which the abrupt mortality increase resulting from the COVID-19 pandemic disrupts the application of these methods. This is particularly the case for the survival ratio method which assumes a monotonic and gradual change in mortality.

The pattern is very similar for men and for women though a bit more marked for the later compared to the former: at its maximum, in 2019, the gap reaches 0.4 years for men (for decile 10) and 0.5 years for women (for decile 1). In 2020, when the values are highest using standard methods relative to those using HMD methods, the gap reaches 0.2 years for men and 0.4 years for women. The pattern is also very similar across all socioeconomic deciles, with no clearly discernable socioeconomic gradient.

Figure 2a

EXPECTATION OF LIFE AT BIRTH BY DECILE IN 1982-2022, EACH SEX USING TWO DIFFERENT METHODOLOGICAL APPROACHES (HMD METHODS IN GREY, STANDARD METHODS IN COLOR)

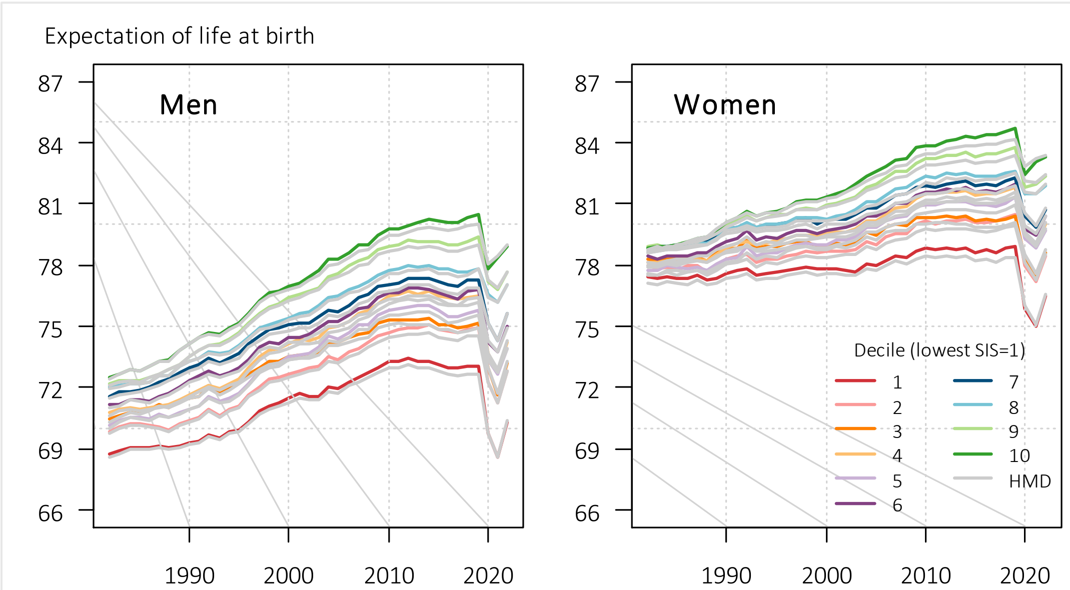
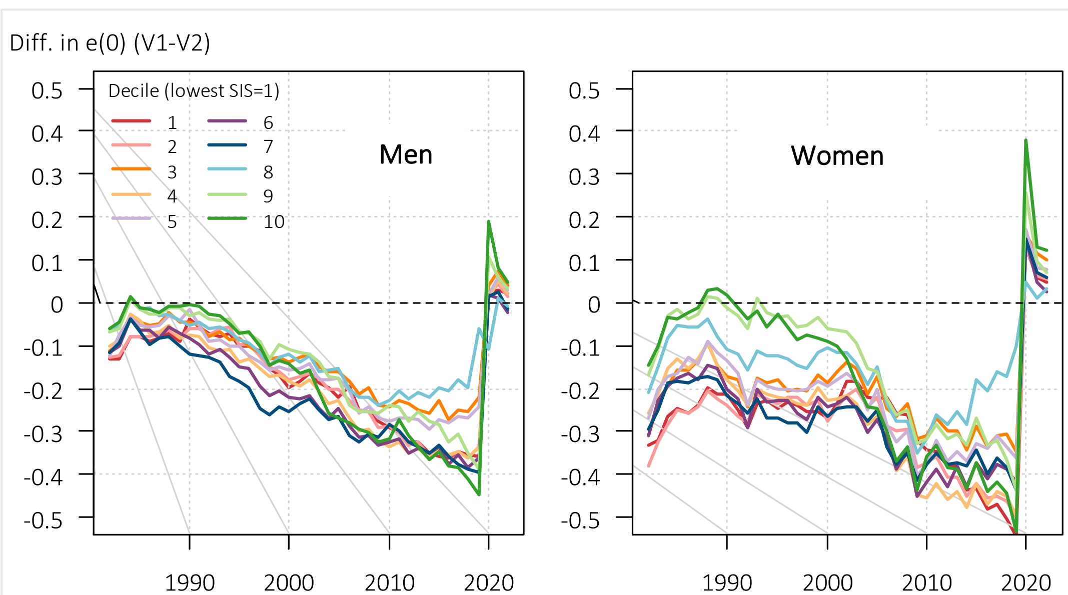


Figure 2b

DIFFERENCES IN THE EXPECTATION OF LIFE AT BIRTH (IN YEARS) BY DECILE IN 1982-2022, EACH SEX BETWEEN TWO DIFFERENT METHODOLOGICAL APPROACHES (V1 = HMD METHODS, V2 = STANDARD METHODS)



The next Figures (3a to 4b) indicate that, as anticipated given the nature of the differences in the two methodological approaches, these differences arise from the way mortality is estimated at ages 80 years and above. Below age 80, the gaps are minuscule: Figure 3a shows a near perfect overlap in the proportions surviving to age 80 and the differences are always below 0.6 per 100,000 survivors (Figure 3b).

Figure 3a

PROPORTION SURVIVING TO AGE 80 YEARS BY DECILE IN 1982-2022, EACH SEX USING TWO DIFFERENT METHODOLOGICAL APPROACHES (HMD METHODS IN GREY, STANDARD METHODS IN COLOR)

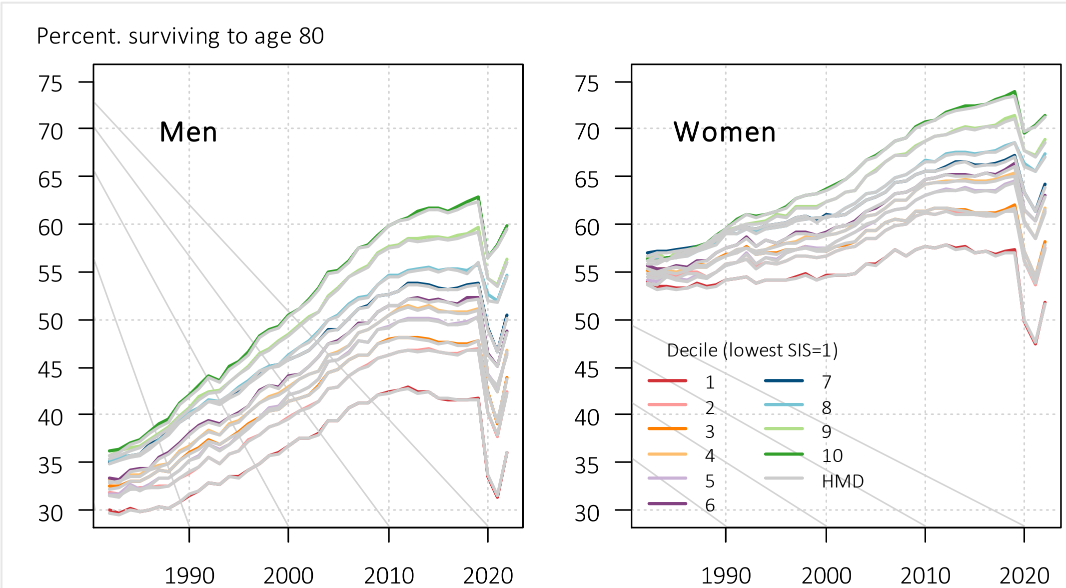
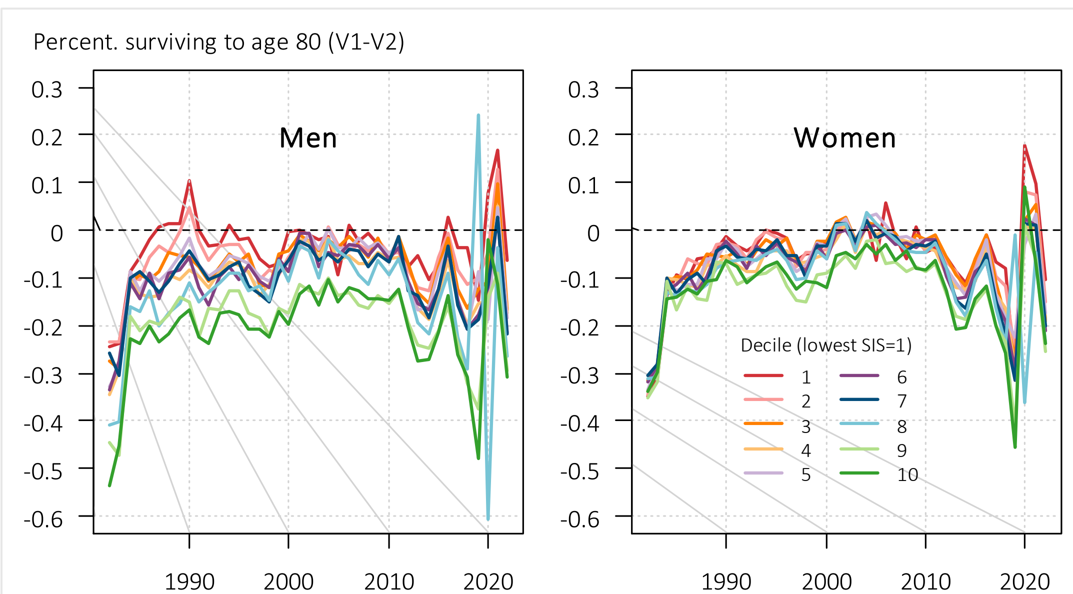


Figure 3b

DIFFERENCES (PER 100,000) IN THE PROPORTION SURVIVING TO AGE 80 BY DECILE IN 1982-2022 BETWEEN TWO DIFFERENT METHODOLOGICAL APPROACHES (V1 = HMD METHODS, V2 = STANDARD METHODS)



By contrast, the gaps in the expectation of life at age 80 are very large (Figure 4a) and the values estimated with standard methods are well above those estimated using HMD techniques. The pattern generally follows that observed with the expectation of life at birth, though with more of a socioeconomic gradient. The differences are largest for the first (most deprived) decile, reaching one year of life in the late 2010s for both men and women (0.3 to 0.7 for deciles 8 to 10). They reverse again starting in 2020 to reach more than 0.2 years for men, 0.4 for women, for the most advantaged decile (decile 10).

Figure 4a

EXPECTATION OF LIFE AT AGE 80 YEARS BY DECILE IN 1982-2022, EACH SEX USING TWO DIFFERENT METHODOLOGICAL APPROACHES (HMD METHODS IN GREY, STANDARD METHODS IN COLOR)

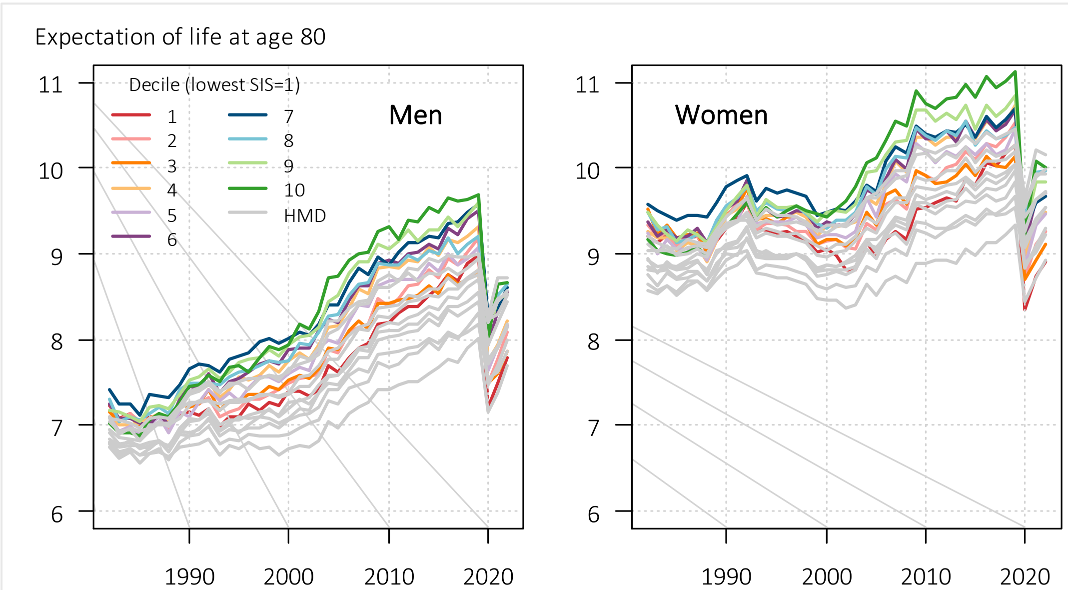


Figure 4b

DIFFERENCES IN THE EXPECTATION OF LIFE AT AGE 80 (IN YEARS) BY DECILE IN 1982-2022, EACH SEX BETWEEN TWO DIFFERENT METHODOLOGICAL APPROACHES (V1 = STANDARD METHODS, V2 = HMD METHODS)

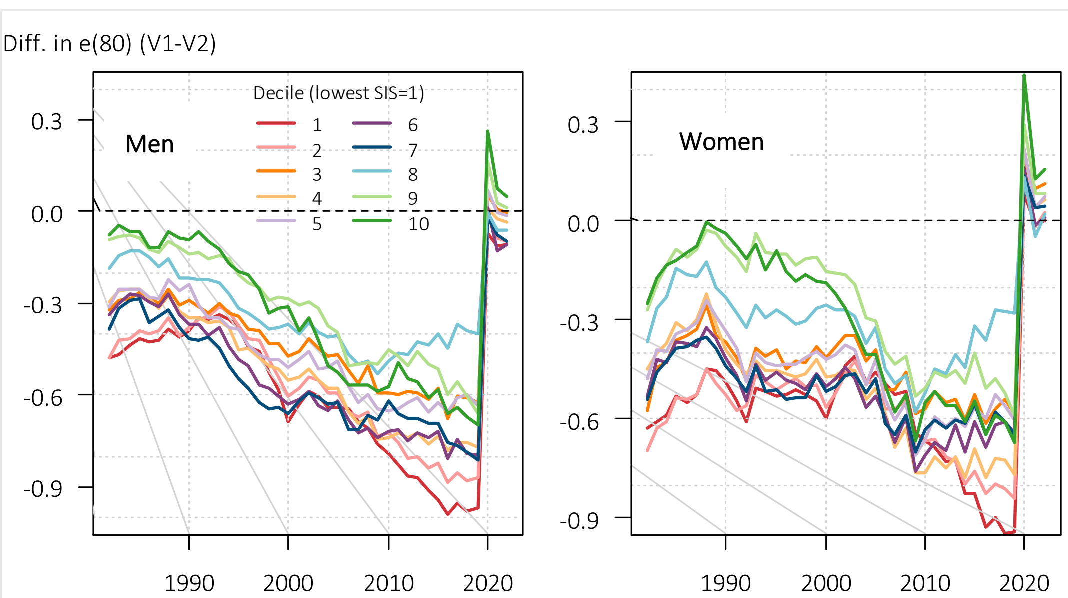
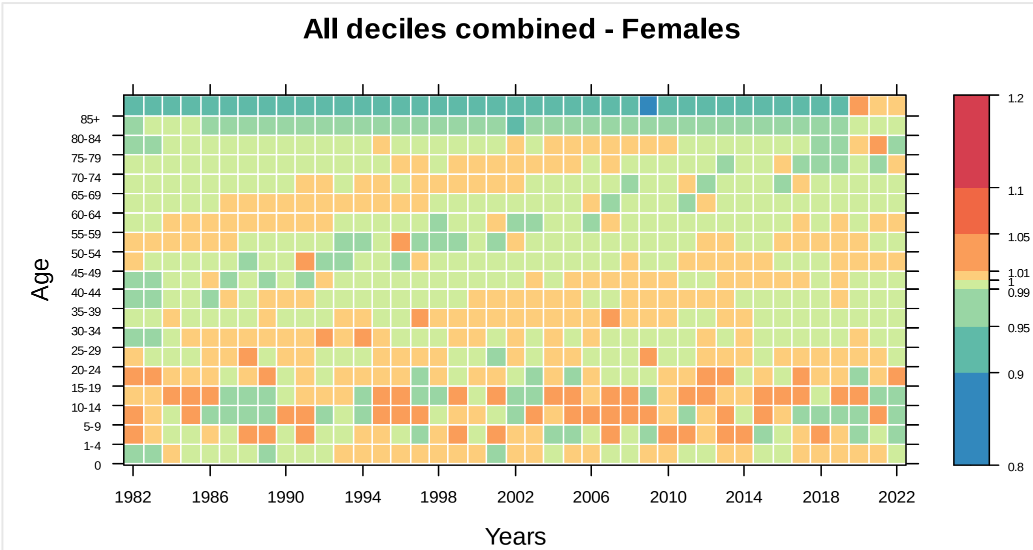
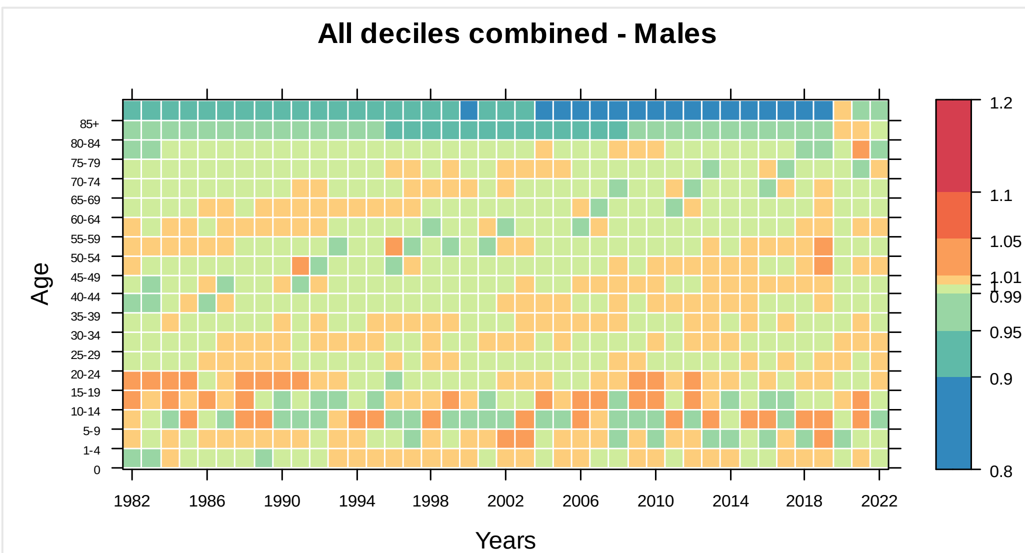


Figure 5 provides more granularity to the comparison. It represents the ratio of the death rates calculated using HMD methods to the death rates calculated using classic demographic technique by single calendar year and by five-year age group for each sex for all deciles combined. It shows that the differences in the rates are typically below 1% except for children above age one and adolescents, where the difference reaches up to 10%, and for people at ages 80 years and above, where it can reach up to 20%. For children and adolescents, the death rates are very low and thus, even a very small absolute difference in the rates will translate into a large relative difference. For older adults, as previously explained, the differences in the methodological approaches are such that large differences were expected. Indeed, instead of the ratio of deaths to the mid-year population used for calculating the rates in standard demographic methods, the HMD implements a combination of extinct cohort and survival ratio methods which only relies on the cohort death counts (and not on the population counts), except for the last few years, when classic methods are used. Indeed, for these last few years, there is no systematic difference in favor of the HMD (Figure 5). The pattern is very similar for men and for women. It is also very consistent across all specific socioeconomic deciles (see Appendix 2), with little differences below age 80 years and a larger gap beyond.

Figure 5

RATIO OF DEATH RATES CONSTRUCTED USING HMD METHODS TO
DEATH RATES CONSTRUCTED WITH STANDARD METHODS, 1982-2022, EACH SEX





In conclusion, though it is difficult to validate either approach, it appears that HMD methods generate higher levels of mortality than standard demographic methods. The special methods implemented in the HMD to construct mortality rates at higher ages have been specifically designed to remedy issues of less detailed and/or poorer-quality data (e.g., deaths and/or populations only available by five-year age groups; deaths and/or populations only available up to an open age-interval; under-reporting of deaths; coverage issues in the census). These issues are known to affect the United States demographic statistics as they do in many other HMD countries (especially for historical periods). However, classic methods have the advantage that once the population estimates from the Census Bureau become final, there will not be any more changes to the estimates even when new mortality data become available. Thus, when the 2010-2019 inter-censal estimates based on the 2020 census have been published, a final series of lifetables for all years 1982-2019 can be generated. Because the Census Bureau revises its post-censal estimates every year, the values in the lifetables for 2020 and after will continue to change each year even when using standard demographic methods. Note however that the detailed lifetables (by single year of age) generated with classic demographic methods can only be calculated for years 1990 and after (when the population estimates by county are also available by single year of age). For the period starting in 1982, only abridged lifetables (by five- or 10-year age group) can be constructed. The choice of the methods to follow in the construction of lifetables thus depends on their use and the user's priorities: if the priority is to obtain the most reliable values, HMD methods should be preferred but if the priority is to obtain a stable, though less detailed, series of lifetables, then classic demographic methods are fine (up to age 80 years at least).

Section 5: Acknowledgments

The researcher's deepest gratitude goes to those without whose efforts this project could not have come to fruition: the Project Oversight Group for their diligent work overseeing, reviewing and editing this report for accuracy and relevance.

Project Oversight Group members:

Jean-Marc Fix, FSA, MAAA

Sam Gutterman, FSA, MAAA, CERA, FCAS, FCA, HonFIA

Larry N. Stern, FSA, MAAA (Chair) At the Name, Credentials

At the Society of Actuaries Research Institute:

Barbara Scott, Senior Research Administrator

Ronora Stryker, ASA, MAAA, Sr. Practice Research Actuary

Patrick Wiese, ASA, Lead Modeling Actuary

Appendix A: Sources of Data

MORTALITY DATA

Restricted-use Mortality Multiple Cause files for years 1982-2022 were provided by the, National Center for Health Statistics (NCHS). They are the same as the publicly available Mortality Multiple Cause files (see https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm) with the essential difference that they include geographic identifiers regarding the deceased's places of residence, down to the county level. These data are available to the HMD team under a Data User Agreement (DUA) with the NCHS. The DUA prevents distributing or publishing any number based on fewer than 10 counts. For the purpose of this SOA project, a random number has been imputed to any cell with less than 10 counts in the tabulations of deaths by county, year of occurrence, sex, and age before any further calculation.

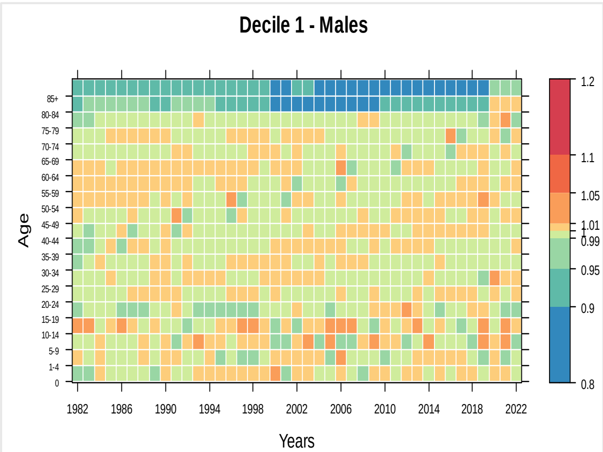
POPULATION DATA

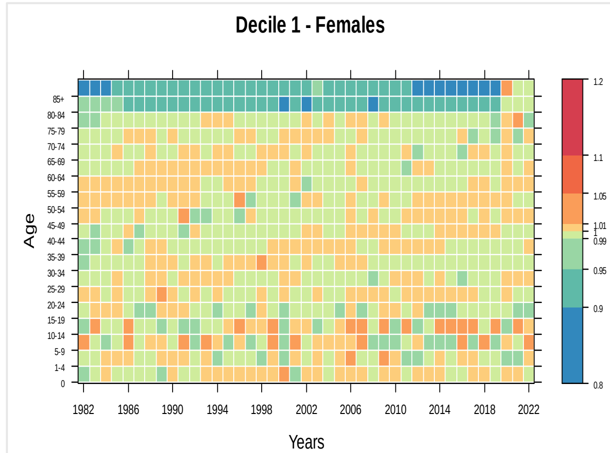
July 1st population estimates by county, calendar year, sex, and age were extracted from the Census Bureau data repository (https://www.census.gov/programs-surveys/popest/data/tables.html). The specific table used from the most recent Vintage of data, i.e., Vintage 2023, is the Annual County and Puerto Rico Municipio Resident Population Estimates by Single Year of Age and Sex: April 1, 2020 to July 1, 2023 (CC-EST2023-SYASEX). Note that, every year, the Census Bureau releases new July 1st population estimates for all years since the most recent census (currently 2020). In addition, after final census results have been published, the final estimates for the most recent inter-censal period are constructed. At the moment, the estimates for the 2010-2019 inter-censal period are still preliminary and the final estimates will not be published until the Fall of 2024 as indicated on the Census Bureau website (https://www.census.gov/programs-surveys/popest/about/schedule.html as of August 7, 2024).

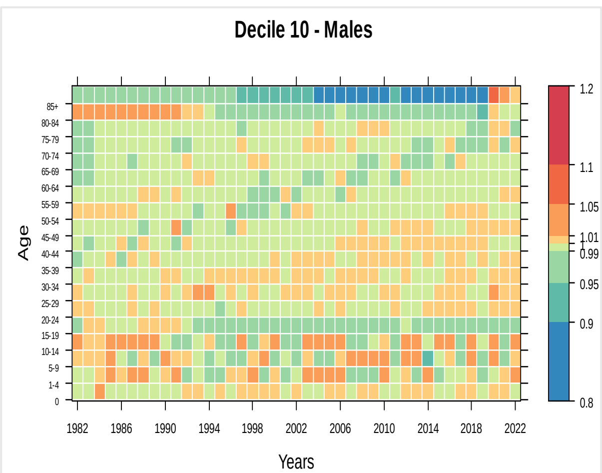
BIRTH DATA

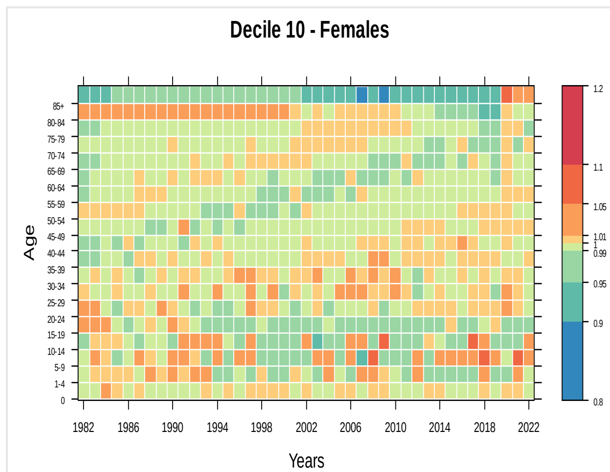
Restricted-use Natality files for years 1982-2022, National Center for Health Statistics (NCHS). Births by year of occurrence and by sex are extracted to better estimate exposure (in combination with the population data described above) following the HMD protocol. They are not used for the estimation of mortality rates using classic demographic methods. Access to the restricted use Natality files is governed by the same DUA as the aforementioned mortality data. The equivalent birth data available at the national level can also be found at https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm.

Appendix B: Ratio of Death Rates Constructed Using Standard Methods to Death Rates Constructed with HMD Methods, 1982-2022, Each Sex









References

Barbieri, M. (2022). Mortality by Socioeconomic Category in the United States, *SOA Research Institute*, Mortality and Longevity Strategic Research Program, https://www.soa.org/4935b3/globalassets/assets/files/resources/research-report/2020/mort-socioeconomic-cat-report.pdf.

Barbieri, M., Wilmoth, J. R., Shkolnikov, V. M., Glei, D., Jasilionis, D., Jdanov, D., ... & Winant, C. (2015). Data resource profile: the human mortality database (HMD). *International journal of epidemiology*, *44*(5), 1549-1556.

Camarda, C. G. (2022). The curse of the plateau. Measuring confidence in human mortality estimates at extreme ages. *Theoretical Population Biology*, *144*, 24-36.

Coale, A.J. & Kisker, E. E. (1990). Defects in data on old-age mortality in the United States: new procedures for calculating mortality schedules and life tables at the highest ages, *Asian and Pacific Population Forum*, 4(1): 1-31.

Dang, L. (2022). Risques de décès aux très grands âges de la vie. Dissertation, Université Paris Nanterre.

Kannisto, V. (1988). On the survival of centenarians and the span of life, Population Studies 42: 389-406.

Shryock, H. S. (2013). The methods and materials of demography. Academic Press.

Wachter, K. W. (2014). Essential demographic methods. Harvard University Press.

Wilmoth, J. R., Andreev, K., Jdanov, D., Glei, D. A., Boe, C., Bubenheim, M., ... & Vachon, P. (2021). Methods Protocol for the Human Mortality Database. Version 6. *University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock.* https://www.mortality.org/File/GetDocument/Public/Docs/MethodsProtocolV6.pdf [version 01/26/2021].

About The Society of Actuaries Research Institute

Serving as the research arm of the Society of Actuaries (SOA), the SOA Research Institute provides objective, data-driven research bringing together tried and true practices and future-focused approaches to address societal challenges and your business needs. The Institute provides trusted knowledge, extensive experience and new technologies to help effectively identify, predict and manage risks.

Representing the thousands of actuaries who help conduct critical research, the SOA Research Institute provides clarity and solutions on risks and societal challenges. The Institute connects actuaries, academics, employers, the insurance industry, regulators, research partners, foundations and research institutions, sponsors and non-governmental organizations, building an effective network which provides support, knowledge and expertise regarding the management of risk to benefit the industry and the public.

Managed by experienced actuaries and research experts from a broad range of industries, the SOA Research Institute creates, funds, develops and distributes research to elevate actuaries as leaders in measuring and managing risk. These efforts include studies, essay collections, webcasts, research papers, survey reports, and original research on topics impacting society.

Harnessing its peer-reviewed research, leading-edge technologies, new data tools and innovative practices, the Institute seeks to understand the underlying causes of risk and the possible outcomes. The Institute develops objective research spanning a variety of topics with its strategic research programs: aging and retirement; actuarial innovation and technology; mortality and longevity; diversity, equity and inclusion; health care cost trends; and catastrophe and climate risk. The Institute has a large volume of topical research available, including an expanding collection of international and market-specific research, experience studies, models and timely research.

Society of Actuaries Research Institute
8770 W Bryn Mawr Ave, Suite 1000
Chicago, IL 60631
www.SOA.org