Investigating the Difference in Mortality Estimates between the Social Security Administration Trustees' Report and the Human Mortality Database

Magali Barbieri

MRRC

Investigating the Difference in Mortality Estimates between the Social Security Administration Trustees' Report and the Human Mortality Database

Magali Barbieri
University of California-Berkeley

September 2018

Michigan Retirement Research Center
University of Michigan
P.O. Box 1248
Ann Arbor, MI 48104
www.mrrc.isr.umich.edu
(734) 615-0422

Investigating the Difference in Mortality Estimates between the Social Security Administration Trustees' Report and the Human Mortality Database

Abstract

This study's goal was to determine whether differences in data or differences in methods explain the divergence between the mortality estimates at ages 65 and older of the Social Security Administration (SSA) and the Human Mortality Database (HMD). These differences, increasing since 1968, are an issue of significant value considering the importance of SSA estimates and projections to determine the long term solvency of the Social Security Trust Funds, as well as of other government programs such as Medicare and Medicaid. The two organizations use different data and different methods to construct their estimates. In particular, the HMD relies on national statistics from the vital registration system and the Census Bureau, while the SSA uses Medicare program enrollment data. Applying the SSA methods to the HMD data showed that differences in the data, rather than in the methods, explain the entire gap in life expectancy at age 65, with the HMD indicator 0.4 years higher for 2014 than the SSA. The study also determined that the gap resulted mostly from lower mortality rates at ages 65 to 84 years (rather than at 85 and older) up to about 2005 to 2006, but that the growing divergence since then is nearly entirely due to increasingly lower mortality at ages above 85. The pattern was found to be similar for men and for women, though the gap is slightly larger for the latter. Additional investigations, with more detailed data, will be necessary to assess whether data reliability or issues of representativeness explain the difference.

Citation

Introduction

The Social Security Administration (SSA) and Human Mortality Database (HMD) life table series for the United States are two of the most widely used sets of mortality indicators at the national level. Life table values differ between the two sources, and the gap has been growing progressively since the early 1980s. The difference in life expectancy at birth has increased from zero in 1980 to over 0.3 years in 2014. Though this difference may not appear as significant, it deserves to be fully understood because of the importance of SSA estimates, which serve as the basis for the SSA annual Trustee's Reports'[1] mortality projections. Any difference in baseline mortality would produce different projection results. Additionally, initial differences may accumulate and get amplified over the projection period with substantial financial implications: The projection is not only crucial to determine the long-term solvency of the Social Security Trust Funds, but also for many other government programs, such as Medicare and Medicaid, as well as for industry strategies in the insurance and retirement sectors in particular (Bennett and Olshansky, 1996; Kashin, King and Soneji 2015). The general goal of the study is to understand the reasons behind the increasing gap in mortality estimation between the SSA and the HMD by carrying out a thorough demographic analysis and, *in fine*, to assess the reliability of these two organizations' underlying data and methods to estimate past and current mortality measures. The results of this demographic research study are expected to contribute to strengthening both the HMD and the SSA mortality projections and models.

The immediate goals of the study are to determine how much of the difference between SSA and HMD in mortality estimation is attributable to differences in the data and how much is

---

[1] Annual Reports of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds.

due to the methods (and, more specifically, which methods really make a difference), and why

the gap between SSA and HMD estimates has been increasing. Though the gap has been noted in

the literature (Kashin, King and Soneji 2015; Goss et al. 2015), the factors behind it have not

been well documented. This analysis is restricted to the period 1968 to 2014. The year 1968 is

the first for which the SSA has used a completely different source of data to construct its life

tables than national statistics and 2014 is the year for which the most recent data are available.

Background

The data and methods used by the Social Security Administration and by the Human

Mortality Database are well documented as described below, but how these contribute to the

difference in mortality estimation remains to be assessed.

Social Security Administration

The Office of the Chief Actuary at the SSA built its life tables from two separate sets of

age-specific death rates using data from 1968 to the present (Bell and Miller 2005). For rates

below age 65, the data used are, for the numerator, the death records by sex and age published by

the National Center for Health Statistics and, for the denominator, population estimates by sex

and age from the Census Bureau. For rates at ages 65 years and above, death and January 1st

population counts by sex and five-year age groups for Medicare beneficiaries are obtained from

the Centers for Medicare and Medicaid Services (CMS) program enrollment. The SSA decided

to use CMS data rather than mortality statistics from the National Center for Health Statistics at

the CDC and population counts from the Estimates and Projections Division at the Census

Bureau because the former are believed to be more reliable than the latter. Though coverage of

the vital registration system and of the U.S. population census is thought to be complete while

Medicare is not, a number of studies have cast doubts on the validity of age reporting in national

statistics, especially for cohorts born before the vital registration system had reached national coverage, in 1933 (Kannisto 1988; Elo and Preston 1994 and 2004; Coale and Kisker 1986 and 1990).

CMS data, which are relevant at ages 65 and older, when most of the population benefits from Medicare, are believed to be fully representative of the Social Security area population (including some groups not covered by the vital statistics system or by the Census, such as the insured foreign-borns who returned abroad at retirement). CMS is also regarded as a more accurate source of statistical information for mortality estimation because a proof of age has to be provided at the time of entitlement to Social Security or Medicare benefits and because both the numerators (death counts) and the denominators (population exposed to the risk of death) are tabulated from the same data collection system rather than from a combination of vital records (for the numerators) and census data (for the denominators) (Goss et al. 2015).

Starting in 1988, instead of using all CMS records, the SSA decided to select only those for Medicare participants who were also eligible for benefits from Social Security or from the Railroad Retirement program. This procedure eliminated about 3 percent of all CMS records when it was initiated (Bell and Miller 2005, p.3). It was justified by lower quality data for the excluded beneficiaries (Kestenbaum and Ferguson 2002). Also excluded are residents of Puerto Rico as well as those beneficiaries who are no longer residents of the U.S. (mostly immigrants who used to work in the U.S. and who have returned home). Since death notifications to CMS are not always timely, enrollees who stop paying a premium are also removed from the calculations, as well as those people who are entitled to a flat benefit not related to prior average earnings (and thus who do not pay any premium), most of them born before 1900 (Kestenbaum and Ferguson 2002).

The methods implemented by the SSA's Office of the Chief Actuary to construct annual complete life tables from these data involve a number of steps and iterations but can be summarized as follows (Bell, Wade, and Goss 2005; Goss et al. 2015). The census population counts are as of midyear, and midyear Medicare counts are estimated by averaging the January 1st counts received from CMS. Rates are computed by dividing the number of deaths for each sex and five-year age groups by the corresponding midyear populations, using these two sets of data depending on the age (younger than 65 or at 65 and older). The rates are converted into probabilities of dying with classic demographic methods. A mathematical formula initially developed by Beers is applied to the probabilities of death in five-year age groups to compute single-year of age probabilities up to 94 years. Because of issue of reliability as regards age reports at higher ages and due to small sample sizes, probabilities of death are assumed to increase at a fixed, sex-specific rate above 94 years. A study by Kestenbaum showed age exaggeration in the CMS data starting at ages slightly younger than 100 years and the resulting under-estimation of death rates in the 1980s (Kestenbaum 1992). Kestenbaum's study confirmed earlier suspicions from Coale and Kisker who determined that some early enrollees in the Medicare program had been excused from providing a definite proof of age if they were already obviously older than 65 at the program inception in 1966 (Coale and Kisker 1986 and 1990).

The Human Mortality Database

The Human Mortality Database (HMD, at www.mortality.org) is a unique open-access collection of detailed mortality and population data for 40 countries with reliable vital registration and censuses, which includes the U.S. for years since 1933 (Barbieri et al. 2015). Its main goal is to document the longevity revolution of the modern era by providing detailed and high-quality mortality data. With more than 50,000 registered users, the HMD is one of the most

widely used data resources in demography. It has been cited in nearly 5,000 publications, including more than 2,800 scientific articles in peer-reviewed journals (http://www.mortality.org/Public/HMD-Publist.pdf). With very few exceptions, a standard set of methods is applied across all time periods and all HMD countries to ensure maximum comparability and to deal with idiosyncrasies and data quality issues in the original data (Wilmoth et al. 2017). The following section describes the methods used to construct the U.S. HMD series more specifically, though it mostly applies to all other HMD mortality series.

The original data used to construct complete life table series are those of the National Center for Health Statistics (death counts by calendar year, sex, single year of age at death and, when available, birth cohort) and the Census Bureau (population estimates by calendar year, sex, single year of age or, for periods when not available, detailed population census counts). These "input" data are used to construct the full set of complete and abridged life tables by year, sex, and age for both periods and cohorts.

Deaths of unknown ages are first redistributed proportionately across all ages for each sex separately. For years when the mortality data do not include the necessary information (before 1989), deaths are further redistributed with a spline function into Lexis triangles (i.e., cross tabulated by both single year of age and birth cohort) using sex-specific regression coefficients estimated from HMD countries and time periods when death counts tabulated by single year of age and birth cohort are available. Deaths in the open age interval are redistributed into Lexis triangles using information on cohort survival for the 20 single years of age below the open age interval (Andreev 2001). For the first year of life, when we know that the vast majority of all deaths occurred within the first few days and weeks after birth, a special method is implemented to allocate deaths to the upper and lower Lexis triangles differently than for other

ages (Andreev and Kingkade 2011). The results of all of these methods is a tabulation of death counts by sex and Lexis triangle for all ages up to 130 years to be used as the numerators of the mortality rates.

The denominator of the rates is represented by the exposures, similarly estimated by Lexis triangle from the Census Bureau annual population estimates. However, only the population counts below age 80 years are used. There are well-known issues with age overstatement among older adults in the U.S. (Coale and Kisker 1990; Elo and Preston 1994; Hill, Preston, and Rosenwaike 2000; Preston and Elo 1999). Furthermore, for early time periods the open age interval is relatively low (85+ years before 1980). To account for these issues, the HMD re-estimates population at higher ages. To do so, the extinct cohort method is combined with the survival ratio method to estimate single year of age population counts from the deaths only, for all ages 80 years and older. This process is expected to produce more robust mortality estimates than using the Census Bureau population estimates because age reports have been showed to be more reliable in vital statistics than in the census and because, in this way, the deaths and exposures are derived from the same source, thus avoiding inconsistencies in the numerators and denominators for the rates (Jdanov et al. 2008). The annual population estimates used for the HMD thus combine the Census Bureau population estimates for ages below 80 and the reconstructed population counts using the extinct cohort and survival ratio methods for ages 80 years and above. These estimates are used to calculate exposure counts by Lexis triangle, further taking into account relative cohort size from the information on the birth counts for the corresponding cohorts.

Deaths and exposure counts by Lexis triangle are then used to compute sex-specific and single year of age mortality rates up to the open age interval 110+ years, with some adjustment at

very high ages to avoid the large year-to-year fluctuations associated with very small counts. The adjustment is carried out using the Kannisto model: A logistic function with an asymptote at one is fitted to the death rates separately for men and for women (Thatcher et al. 1998). The fit is carried out on all ages 80 and above, but the smoothed rates are substituted to the unadjusted rates only for ages 95 years and above. All life table functions are derived from these rates following classic demographic methods.

The data used by the SSA and the HMD are thus the same for all death rates up to age 65 and the differences in methods are small enough to expect relatively similar results. At ages 65 and older, the HMD also uses national statistics but the SSA uses CMS data, and the methods implemented by the two organizations to compute mortality rates and probabilities at higher ages are very different.

Differences in mortality estimates between the SSA and the HMD

Trends in the difference in life expectancy at birth between the HMD and the SSA have followed a relatively linear pattern since 1968. The difference between HMD and SSA estimates fluctuates around zero up to 1988, when the criteria for the CMS extract have been revised by the SSA, as previously described. Starting in 1991, life expectancy at birth in the HMD becomes progressively higher than in the SSA and reaches over 0.3 years in 2014. The overall pattern is similar for both sexes though the difference is typically smaller for women than for men, except for the most recent years when it has become larger for women (i.e., since 2010).

Figure 1. Difference in life expectancy at birth between HMD and SSA (HMD minus SSA), 1968 to 2014
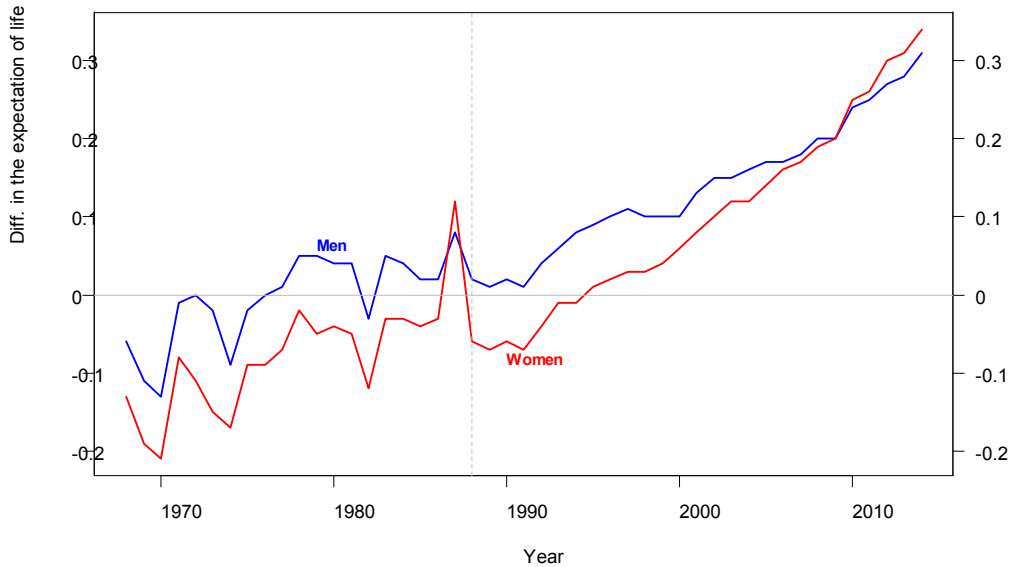
The gap in the life expectancy at birth is entirely attributable to differences in the mortality rates at ages 65 and older. When comparing the proportions surviving to age 65 in the two sets of life tables, the difference is less than one per thousand for all years and each sex (Figure 2). Given the fact that both the HMD and the SSA use the same data sources and very similar demographic methods to estimate life table values younger than 65, this result is not surprising. The rest of this report thus concentrates on mortality estimates at ages 65 and older.

Figure 2. Difference in the percentage of survivors at age 65 (HMD-SSA), 1968 to 2014
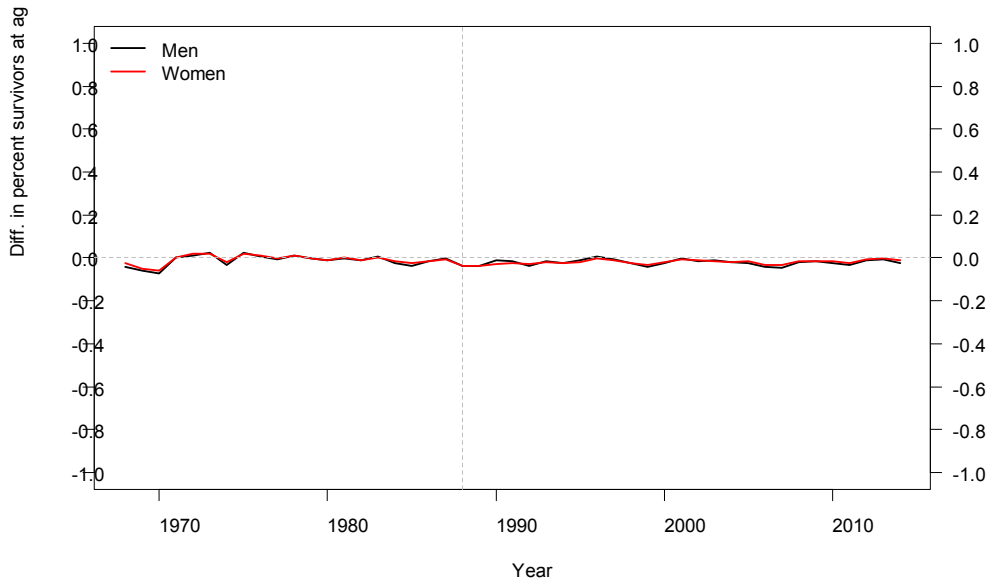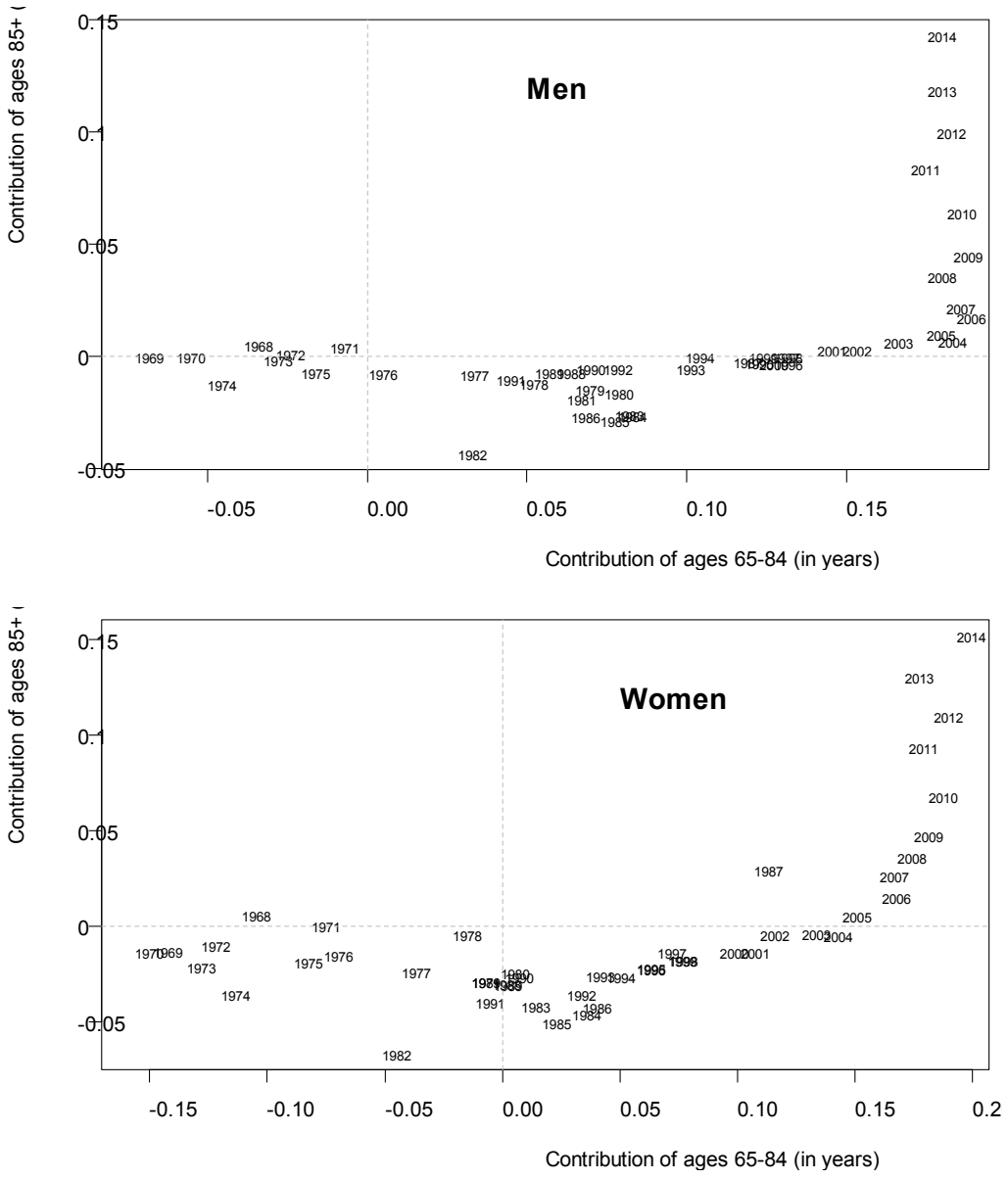

Life expectancy at age 65 is lower in the HMD than in the SSA up to the mid-1970s for men and the early 1990s for women (except for a temporary jump in 1988) and the gap in life expectancy at age 65 has increased to reach 0.4 years for each sex in 2014. Though, as previously explained, the SSA data are expected to be more reliable than the data used by the HMD, the quality of national statistics is expected to have improved over time in the U.S. and the fact that the difference between SSA and HMD estimates of the life expectancy at age 65 has increased since the end of the 1980s is puzzling, providing a strong motivation for further analysis.

Age contributions to the difference in the life expectancy at age 65

We implemented a decomposition method developed by Andreev, Shkolnikov, and Begun to estimate the contribution of each age group to the difference in life expectancy at age 65 between HMD and SSA (Andreev, Shkolnikov, and Begun 2002). In this method, which builds from previous work by Pollack, an algorithm estimates the effects of successively replacing each cell of one matrix by the corresponding cell in the other matrix.

Figures 3a. and 3b. Decomposition of the gap in e65 between HMD and SSA by age group by calendar year, 1968-2014, each sex

Note. These graphs are to be read as follows: In 1968, male life expectancy at age 65 was lower in the HMD than in the SSA lifetable by 0.03 years, and the difference was nearly entirely due to higher death rates at ages 65 to 84 years. By contrast, in 2014, the same indicator is higher in the HMD than in the SSA by 0.32 years, with 0.18 years due to lower death rates at ages 65 to 84 and 0.14 years to lower rates at ages 85+.

The result is striking and very similar for men and for women: While the contribution of the age group 65 to 84 years increases fairly regularly throughout the period (though, for men, it has remained constant since 2006), that of the age group 85 and older, which was close to zero up to 2005, increased quickly and regularly afterwards (Figure 3a and 3b). The share of the age group 65 to 84 years to the absolute difference in the life expectancy at age 65 increased from -0.04 years for men and -0.11 years for women in 1968 (that is, mortality was higher in the HMD than in the SSA life tables at these ages) to 0.18 and 0.16 years, respectively, in 2006, and to 0.18 and 0.19 years in 2014. For the age group 85+ years, the share increased from 0.0 for both sexes in 1968, to 0.01 years also for both sexes in 2005, and to 0.14 for men and 0.15 for women in 2014.

Applying the SSA methods to HMD data

We applied SSA methods to HMD data to measure exactly how much of the gap in mortality at 65 and older is due to differences in methods, and how much is due to differences in data. The purpose is to narrow down the source of the difference between the two sets of estimates. The SSA's Office of the Chief Actuary has provided the necessary data and technical support to help us replicate the SSA process for producing the Trustees Report life tables using the open-source software R.

The main finding of this analysis is that the differences in methods explain none of the gap in the HMD *versus* SSA mortality estimates (Figure 4). If anything, methodological differences actually help to *close* the gap between the two sets of estimates. Indeed, applying SSA methods to HMD data (i.e., data from the vital registration system and the Census Bureau) results in differences in the expectation of life at age 65 even larger than with the original HMD estimates for men: The gap between the blue line and red line on Figure 4 is even bigger than

between the black and red lines (also see Figure 5). For women, the difference in the estimates computed with HMD methods and with SSA methods using the national statistics is negligible (the blue and black lines overlap on Figure 4).

Figure 4. Trends in life expectancy at age 65 with HMD data/HMD methods, HMD data/SSA methods and in the 2017 SSA Trustees' Report by sex, 1968-2014
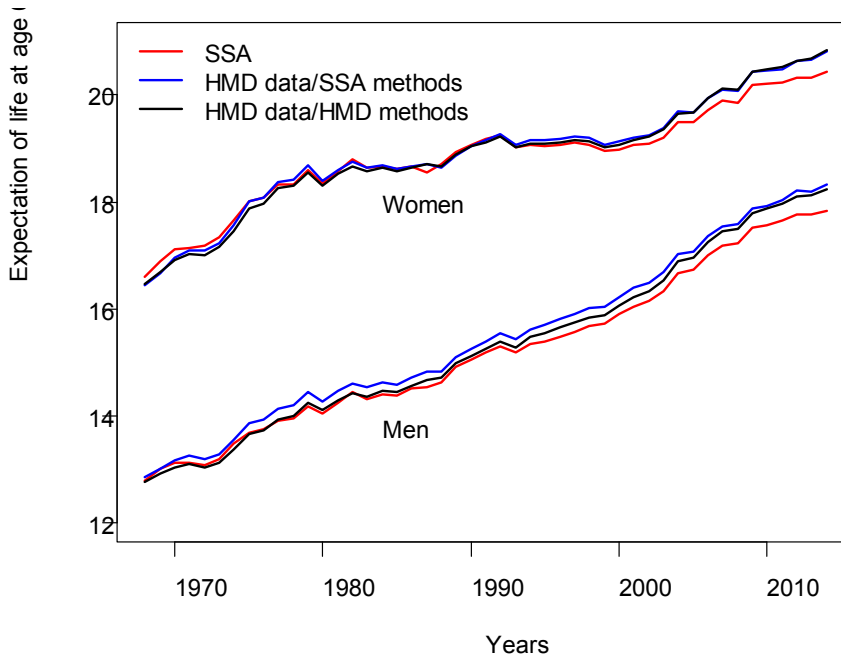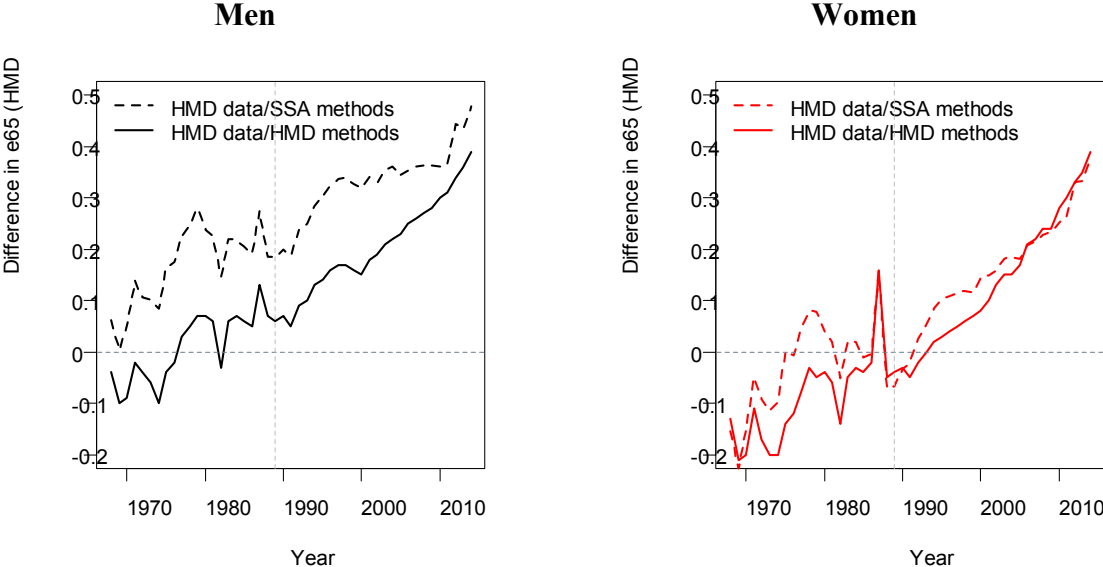
Figure 5. Difference (years) in life expectancy at age 65 (HMD-SSA), 1968 to 2014 using the HMD methods on 1) the HMD data and 2) the SSA data

Men   **Women**

The phenomena of an amplified gap for men when applying SSA methods to HMD data is observed both before and after age 95. For instance, the male survival probability between the ages of 65 and 95 (Figure 6) in 1968 is 1.6 percent in the SSA lifetable, also 1.7 percent in the original HMD lifetable, and 2.2 percent in the hybrid lifetable (constructed by applying SSA methods to HMD data). In 2014, these proportions have increased to 7.3, 8.7, and 9.9 percent, respectively. For women, the proportions are 4.8, 4.8, and 5.3 percent in 1968, and 13.9, 15.6, and 16.3 percent in 2014. When we look at the expectation of life at age 95 (Figure 7), the indicator increases for men from 2.4 to 2.9 years between 1968 and 2014 in the SSA lifetables, from 2.6 to 3.1 years in the HMD lifetables, and from 2.8 to 3.6 in the hybrid lifetable. For women, the values increase from 2.8 to 3.4 years, from 2.8 to 3.5, and from 3.0 to 3.9 years, respectively, in the SSA, HMD, and hybrid lifetables. The increasing gap is thus not the result of differences in the modeling done at very high ages in the two sources.

Figure 6. Trends in the probability of dying between the ages of 65 and 95 (percent) with HMD data/HMD methods, HMD data/SSA methods and in the 2017 SSA Trustees' Report by sex, 1968 to 2014
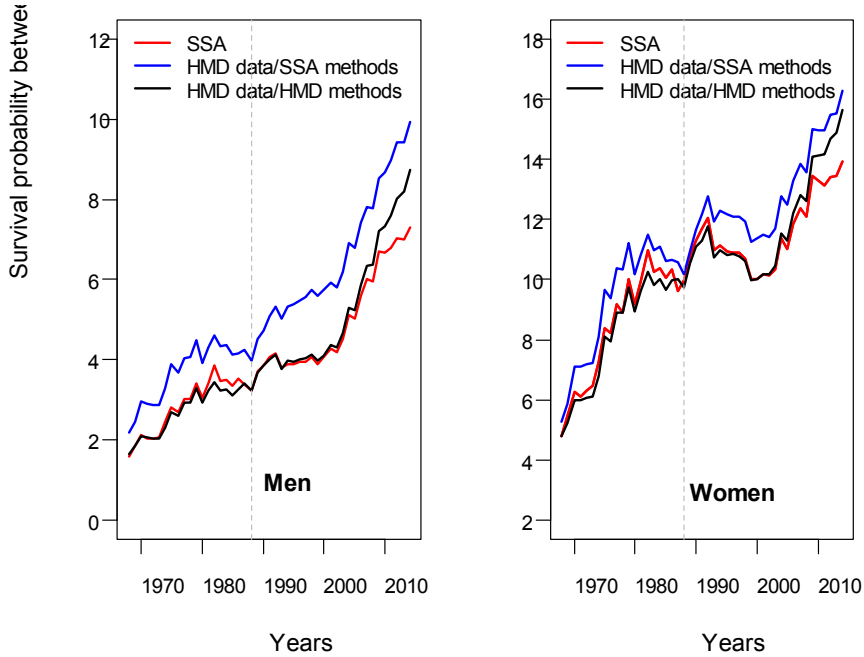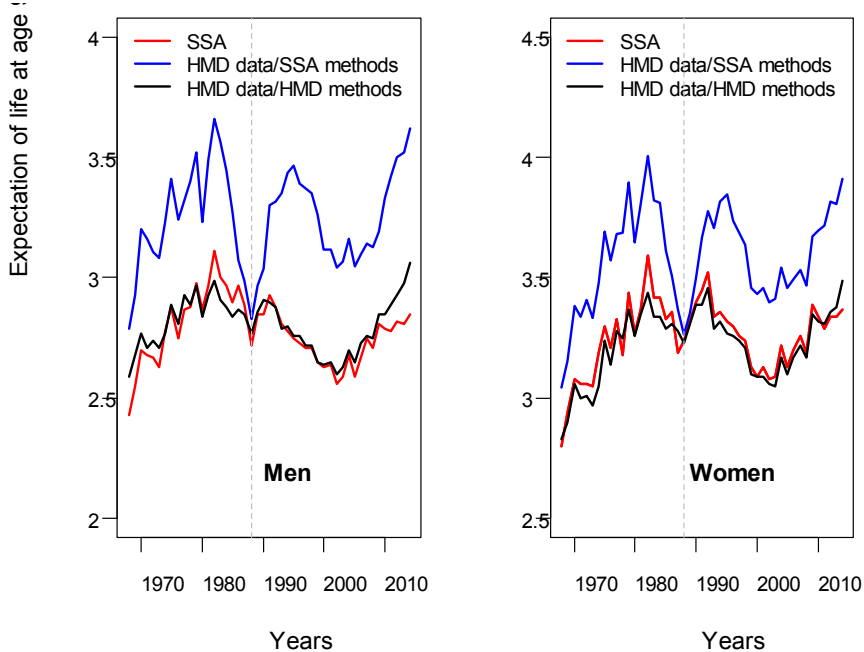

Figure 7. Trends in the expectation of life at age 95 with HMD data/HMD methods, HMD data/SSA methods and in the 2017 SSA Trustees' Report by sex, 1968 to 2014


14

These results clearly indicate that the discrepancy in mortality estimation between the SSA and the HMD lifetables is entirely due to differences in the original data used to construct the lifetables and not to differences in the methods implemented by the two organizations. In fact, with the same set of methods, the gap between the SSA and HMD values would be even larger.

Differences in the SSA and HMD input data

We compared the original (input) data used by the SSA and the HMD to construct their respective lifetables. Note that the HMD input data are national statistics produced by the U.S. National Center for Health Statistics (for the deaths used in the numerator of the mortality rates) and Census Bureau (for the census populations and annual estimates used in the denominator of the rates for ages below 80). We are thus really comparing the data extracted for the SSA by CMS and the national statistics. Discrepancies in the data would indicate either problems with the national statistics, problems with the CMS data, or differences in coverage (i.e., indicating that the CMS data used by the SSA are increasingly less representative of the entire U.S. resident population). Before 1988, the CMS data used by the SSA included the entire Medicare population. Since 1988, however, the CMS data used by the SSA are not representative of the U.S. resident population but they have been restricted to the US-resident population covered by Social Security. Consequently, compared to the total U.S. population, excluded from the CMS database are those individuals residing in the U.S. who do not qualify for Social Security benefits because neither they, nor a spouse (if any), have worked sufficiently in covered employment to become insured for benefits. Note that this is likely to exclude undocumented immigrants from the Social Security estimates while their deaths would be reported to vital statistics.

The comparison indicates two distinct periods: before 1988 and after (Figure 8). At the beginning of the period, in 1968, the death totals and population are very similar in the two sets of data. Then, the number of SSA deaths declines slightly relative to the NCHS estimates and fluctuates around 99 percent up to 1987. During the same time period, the total population count in the SSA declines regularly starting in 1970 up to 1986 and then more suddenly between 1986 and 1987 to reach about 95.3 percent at this later date. Then, in 1988, when the SSA modified its criteria to extract the death and population data used for its mortality estimation from the CMS database, the proportion drops for both the death and the population counts relative to the national statistics: The total number of SSA deaths represent 95.4 percent of the total number of NCHS deaths and the total SSA population represents 93.9 percent of the HMD estimate. Because the criteria used by SSA starting in 1988 to extract its information from the CMS database are more restrictive than those used before 1988, the drop in representation is not surprising. However, the following trends indicate a striking divergence between the SSA death counts and the population counts relative to the national statistics: While the death count remains relatively stable at around 96 percent, the relative population count continues to decline and only reaches 90.6 percent of the HMD estimate in 2014.

Figure 8. Ratio of total population and death count, SSA/HMD*,

1968 to 2014, ages 65+, both sexes
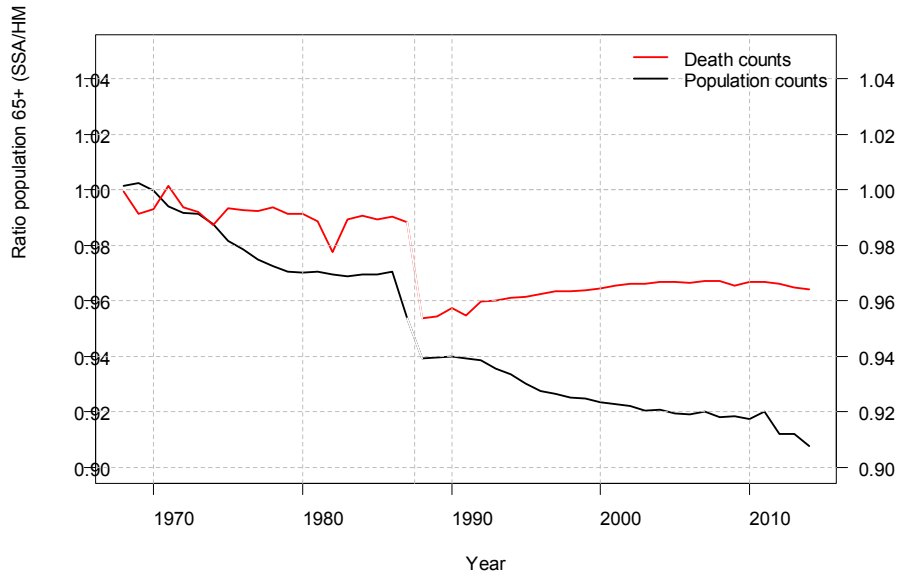
The pattern is very similar for both sexes (Figure 9), though more marked for women as the SSA female death and population counts appear to be a little bit closer to those estimated with national statistics before 1988 while the reverse is found after 1988. For instance, the SSA death total for men represents 99.7 percent of the NCHS total in 1968 and 97.0 in both 1988 and 2014, while the SSA death total for women represents 100 percent of the NCHS total in 1968, 93.9 percent in 1988, and 95.9 percent in 2014. As regards the population totals, the proportions are 100.8 percent for women and 99.2 percent for men in 1968, very similar for both sexes in 1988 (94.3 and 93.7 percent for men and women, respectively), and diverging after 1999 to reach 91.4 and 90.3 percent in 2014.

Figure 9. Ratio of total population and death count, SSA/HMD,

1968 to 2014, ages 65+, each sex
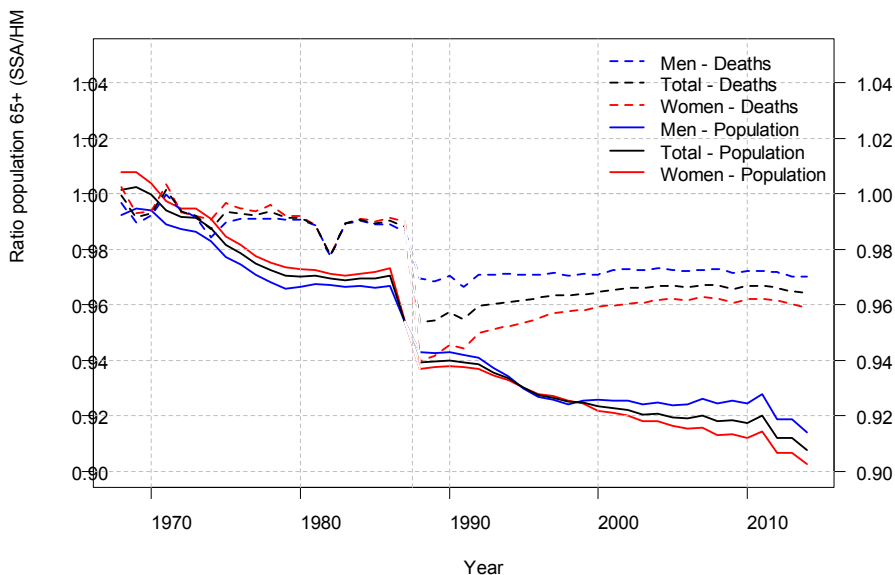
There seems to be a slight age gradient to the difference in population totals between the CMS data and the HMD estimates. Figure 10 represents the ratio of the SSA to the HMD population over all years from 1968 through 2014, for each single year of age up to 85+ years (both sexes combined). It shows that the median ratio ranges from 90.0 percent at age 65 years to 93.4 percent at age 81 years. The lower ratios at ages 65 to 69 years are explained by the fact that some Social Security beneficiaries hold off on filing for benefits to increase their payout. However, the financial advantage of delaying retirement runs out at 70 years and after that age, the ratio of population totals in the SSA to the HMD fluctuates around 93 percent, without any clear pattern.

Figure 10. Distribution of the ratios of total population, SSA/HMD, by age up to 85+ years, over all calendar years 1988 to 2014, both sexes
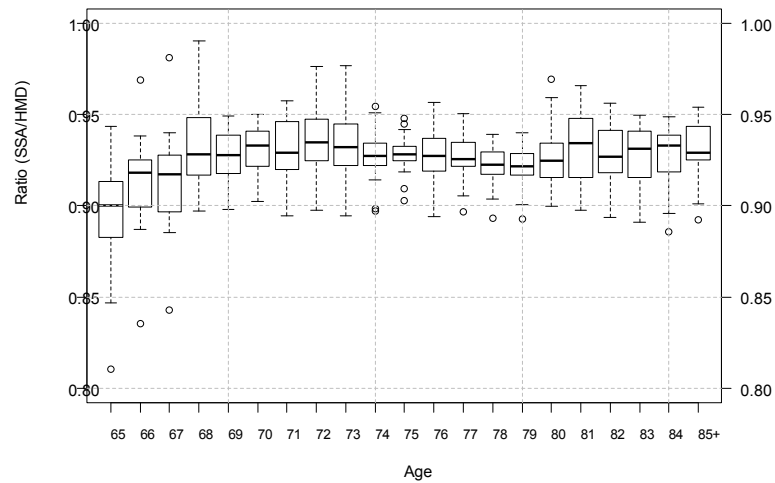
Note: The graphs (boxplots) represent the ratio of the total population counts used by the SSA to estimate mortality to the total population counts in the Human Mortality Database for each single year of age between 65 and 85+ years. The boxplots show the distribution of this ratio over all calendar years 1988 through 2014 for each single year of age. The rectangular box captures half of the observations, i.e., those between the first and third quartiles of the distribution. The thick black line crossing the box toward its middle represents the median. The whiskers located on either side of the rectangle correspond to the lower and upper limits, or "adjacent values", of the normal distribution, except where the extreme values are close to the median, in which case the whiskers end at the extreme value. The adjacent values are calculated from the interquartile range which is multiplied by 1.5 and then subtracted from the first quartile and added to the third.

Plausibility of the HMD population estimates

The increased divergence between the SSA and the HMD population numbers could arise from a progressive overestimation of the population when relying on national statistics. This could result, in particular, from increasing over-count in the census. Indeed, results from the Census Bureau demographic analysis suggests that the population has been over-counted at ages 65 to 84 and under-counted at 85 and older. To examine the plausibility of the HMD estimates at 65 and older, we analyzed the consistency of year to year counts with the annual number of deaths at these ages within each cohort. The idea is to compare the total January 1st population

count at time t at each age x with the total January 1st population count at time (t+1) and age (x+1) after subtracting the deaths recorded at age x to (x+1) during the period t to (t+1). This calculation is only really relevant for population between the ages of 65 and 80 because starting at age 80, the HMD relies on the mortality data from NCHS to estimate population by cumulating the deaths back within each cohort as previously explained and does not use the Census Bureau population data. Since only death data are used to estimate the population 80 and older in the HMD, no difference is expected between the two values. For ages below 80, assuming that there is no reporting problem with the NCHS deaths, any difference between the two values has to be attributed to either migration or error.

If the results showed on Figure 8 arose from problems with the HMD population estimates, we should find a progressive increase in the residual (i.e., the amount of implied migration/error). In fact, our analysis indicates a very plausible pattern which is not consistent with the idea that the HMD population estimates have been increasingly overestimated over time since the end of the 1980s (Figure 11). The residual is negative for the period between 1988 and 1999, which would suggest that more adults 65 and older have left the country than have entered (or, if we attribute all of the residual to errors, that the population has been under-estimated). After a peak at +52,000 in 2001, it fluctuates around +20,000 up to 2010-2011 and increases to +40,000 in 2014. This value is negligible compared to the 4,270,000 difference between the total population count at 65 and older in the HMD and the corresponding population count in the SSA for 2014.

Figure 11. Residuals in the cohort component projection in the HMD in 1988 to 2016, 65 and older
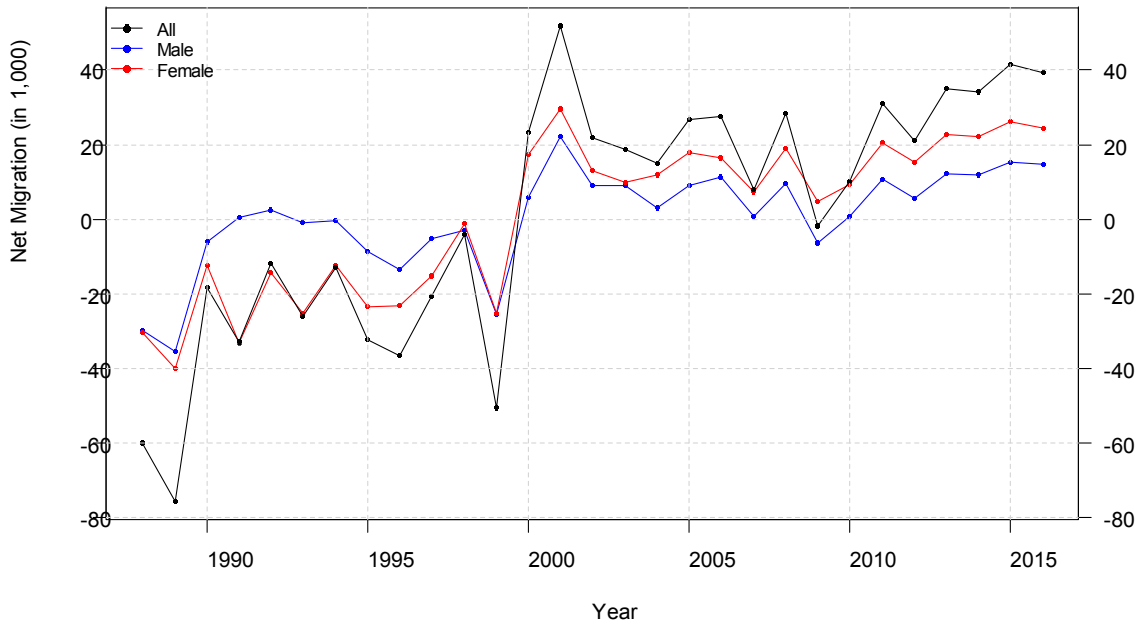
Reconciling the SSA and HMD estimates

If we assume that both SSA and HMD sources of data are reliable, the discrepancy in mortality estimation would have to be explained by a difference in the mortality risks of the SSA population compared with the overall U.S. resident population (as covered by the HMD). As previously explained, the exposure used by the SSA to compute the Trustees' Report mortality rates and associated life table values are derived from the database maintained at the Centers for Medicare and Medicaid Services (CMS). The CMS extract used by the SSA includes a subset of the overall U.S. resident population (since foreign residents and public servants not covered by Social Security in the CMS are excluded from the SSA calculations). We have showed that about 10 percent of the U.S. total resident population (at 65 and older) is excluded from the SSA dataset. We call these excluded individuals the "missing" SSA population.

We have calculated the expectation of life at age 65 for the portion of the U.S. resident population not included in the CMS extract used by SSA (the "missing" population) that would be consistent with the values provided by the SSA and HMD life tables. The result indicates that the "missing" population would need to have a very low, though not unreasonably so, level of mortality to reconcile the two sets of estimates (Appendix Table 1). For instance, in 2014, remaining years of life at age 65 of 17.84 years for men and 20.44 for women in the SSA and of 18.23 years for men and 20.83 years for women in the HMD imply that in the HMD population not covered by SSA, these values would need to be respectively 22.64 and 24.60 years.

We looked at low-mortality populations to gauge the plausibility of these estimates. In Japan, the country with the highest level of life expectancy at age 65 in the world, the remaining length of life at this age was estimated to be 19.28 years for men and 24.16 years for women in 2014. As for the U.S., the state with the highest level of life expectancy at age 65 is Hawaii with values at 19.88 years and 23.11 years, respectively, for men and for women in 2014. It thus appears that a value of 24.60 years for the remaining number of years to live at age 65 for women is a bit high, but not entirely implausible in comparison to other population groups. For men, however, 22.64 seems quite high compared to other low-mortality populations.

Before making a final determination of the results plausibility, we would need to know more about the characteristics of this "missing" SSA population. It is, for instance, reasonable to assume that those people who delay enrolling into Medicare or requesting Social Security benefits until an age older than 65 (and who are thus excluded from the CMS records used by the SSA for its lifetable calculations) have lower mortality risks than those who do enroll into Medicare or request Social Security benefits as soon as they reach 65. Within the current project, we do not have access to the appropriate data to carry out the proper analysis, so this

question remains unanswered. This hypothesis could be tested in future collaborative work with the SSA's Office of the Chief Actuary.

Conclusion

While mortality estimates below age 65 are nearly indistinguishable in the lifetables published by the Social Security Administration and in the Human Mortality Database, there has been a growing gap in the expectation of life at age 65 in the two series of lifetables, which reaches 0.4 years in 2014. Our study has demonstrated that, though both organizations use a different set of methods to construct their mortality estimates, this gap is entirely attributable to differences in the original sources of data (the Centers for Medicare and Medicaid Services database for the SSA and national statistics from the NCHS and the Census Bureau for the HMD). The most puzzling finding in this respect is the fact that, while the total death counts by year for ages 65 and older in the SSA compared to the HMD fluctuates around 96 to 97 percent for all years since 1988, the ratio of the SSA to HMD population counts has declined from about 94 percent in 1988 to less than 91 percent in 2014. This phenomenon could result from an over-estimation of the population in the HMD or an increasingly restricted sample in the CMS data. Our ability to differentiate between these hypotheses is limited by the lack of appropriate data or additional information. Our results warrant further investigation into the data used by the HMD and SSA, both in terms of coverage/representativeness and the quality of age reporting. Further collaboration with the Office of the Chief Actuary of the Social Security Administration is being discussed to explore differences in the data sources in more depth. With this goal in mind, access to the Centers for Medicare and Medicaid Services database and to all of the variables that are used to extract the sub-sample used by the SSA is essential to understand the observed discrepancy in mortality estimation.

References

Andreev, E. M., Shkolnikov, V. M., & Begun, A. Z. (2002). Algorithm for decomposition of differences between aggregate demographic measures and its application to life expectancies, healthy life expectancies, parity-progression ratios and total fertility rates. *Demographic Research*, *7*, 499-522.

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*, dyv105.

Bell, F. C., Wade, A. H., & Goss, S. C. (2005). Life tables for the United States social security area 1900-2100. *Actuarial study*, *120*, Social Security Administration.

Bennett, N. G., & Olshansky, S. J. (1996). Forecasting U.S. age structure and the future of social security: The impact of adjustments to official mortality schedules. *Population and Development Review*, 703-727.

Coale, A. J., & Kisker, E. E. (1986). Mortality crossovers: reality or bad data?. *Population studies*, *40*(3), 389-401.

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.

Elo, I. T., & Preston, S. H. (1994). Estimating African-American mortality from inaccurate data. *Demography*, *31*(3), 427-458.

Elo, I. T., Turra, C. M., Kestenbaum, B., & Ferguson, B. R. (2004). Mortality among elderly Hispanics in the United States: Past evidence and new results. *Demography*, *41*(1), 109-128.

Goss, S., Wade, A., Glenn, K., Morris, M., & Bye, M. (2015). Accuracy of mortality projections in Trustees Reports. *Social Security Administration Actuarial Note* 156, 12 p.

Hill, M. E., Preston, S. H., & Rosenwaike, I. (2000). Age reporting among white Americans aged 85+: results of a record linkage study. *Demography*, *37*(2), 175-186.

Jdanov, D. A., Jasilionis, D., Soroko, E. L., Rau, R., & Vaupel, J. W. (2008). Beyond the Kannisto-Thatcher database on old age mortality: An assessment of data quality at advanced ages. Max Planck Institute for Demographic Research, Rostock, *MPIDR Working Paper 2008-013*.

Kannisto, V. (1988). On the survival of centenarians and the span of life. *Population studies*, *42*(3), 389-406.

Kashin, K., King, G., & Soneji, S. (2015). Systematic bias and nontransparency in U.S. Social Security Administration forecasts. *The Journal of Economic Perspectives*, *29*(2), 239-257.

Kestenbaum, B. (1992). A description of the extreme aged population based on improved Medicare enrollment data. *Demography*, *29*(4), 565-580.

Kestenbaum, B., & Reneé Ferguson, B. (2002). Mortality of the extreme aged in the United States in the 1990s, based on improved Medicare data. *North American Actuarial Journal*, *6*(3), 38-44.

Preston, S. H., Elo, I. T., & Preston, S. H. (1999). Effects of age misreporting on mortality estimates at older ages. *Population studies*, *53*(2), 165-177.

Wilmoth, J. R., & Shkolnikov, V. (2008). Human mortality database. University of California, Berkeley (US), and Max Planck Institute for Demographic Research (Germany).

Wilmoth J. R., Andreev K., Jdanov D., *et al*. (2017) Methods protocol for the Human Mortality Database, Version 6. University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock. Available from*: http://www.mortality.org/Public/Docs/MethodsProtocol.pdf [version 11/27/2017]*

Appendix Table 1. Life expectancy at age 65 in the SSA and HMD lifetables and for the U.S. resident population not included in the CMS extract used by SSA, 1989 to 2014

Year	SSA		HMD		Missing	
	Men	Women	Men	Women	Men	Women
1989	14.92	18.92	14.98	18.88	16.05	18.27
1990	15.06	19.07	15.13	19.04	16.40	18.57
1991	15.20	19.17	15.25	19.12	16.14	18.33
1992	15.31	19.25	15.40	19.23	16.99	18.91
1993	15.19	19.01	15.29	19.01	17.00	19.01
1994	15.34	19.07	15.47	19.09	17.58	19.40
1995	15.40	19.05	15.54	19.08	17.67	19.53
1996	15.49	19.06	15.65	19.10	18.02	19.69
1997	15.58	19.11	15.75	19.16	18.21	19.89
1998	15.68	19.07	15.85	19.13	18.30	20.00
1999	15.72	18.95	15.88	19.02	18.14	20.00
2000	15.91	18.98	16.06	19.06	18.33	20.20
2001	16.05	19.06	16.23	19.16	18.92	20.54
2002	16.15	19.09	16.34	19.22	19.08	20.94
2003	16.33	19.20	16.54	19.35	19.50	21.27
2004	16.67	19.50	16.89	19.65	19.97	21.54
2005	16.73	19.49	16.96	19.66	20.08	21.72
2006	17.00	19.72	17.25	19.93	20.56	22.38
2007	17.19	19.89	17.45	20.11	20.82	22.60
2008	17.22	19.85	17.49	20.09	21.00	22.78
2009	17.51	20.19	17.79	20.43	21.41	23.07
2010	17.57	20.20	17.87	20.48	21.68	23.49
2011	17.66	20.22	17.97	20.52	21.77	23.62
2012	17.76	20.31	18.10	20.64	22.14	23.96
2013	17.76	20.32	18.12	20.67	22.30	24.13
2014	17.84	20.44	18.23	20.83	22.64	24.60