SOCIETY OF ACTUARIES

LIVING to 100
SOCIETY OF ACTUARIES INTERNATIONAL SYMPOSIUM

Mortality and Longevity

Aging and Retirement

Article from

2020 Living to 100 Call for Essays

Recent Trends in Old-Age Mortality and Mortality Improvement

April 2020

Contents

Recent Trends in Old-Age Mortality and Mortality Improvement **3**
Abstract 3
Introduction 3
Methods 3
Conclusions 5
Acknowledgments 6
References 7

Recent Trends in Old-Age Mortality and Mortality Improvement

Natalia S. Gavrilova and Leonid A. Gavrilov

ABSTRACT

This essay analyzes recent trends in old-age mortality and mortality improvement in the United States using data for five single-year cohorts born in 1898–1902. The study was conducted using data from the Social Security Administration Death Master File (DMF). Direct age validation of longevity records with historical resources (such as early censuses) at ages 100, 103, 105 and 109+ years revealed that old-age mortality may be significantly underestimated because of data errors. Age misreporting is particularly strong at extreme old ages of 109+ years, which may result in spurious mortality deceleration. This study has not found any mortality improvement overtime for ages 95+ years and practically no differences in mortality by the region of last residence. The study found that more detailed monthly estimates of the force of mortality demonstrate a more Gompertz-like age trajectory compared to traditional yearly estimates.

INTRODUCTION

Global population aging and rapid growth of older populations in industrialized countries underscores the need for accurate estimates of mortality at advanced ages; these are essential for population and actuarial forecasts. Throughout the actuarial and demographic communities, colleagues debate the shape and level of mortality at the very advanced ages. The goal of this essay, based on a project supported by the Society of Actuaries, is to address that question by identifying the old-age mortality trajectories for the five extinct (or nearly extinct) U.S. cohorts with a birth year of 1898, 1899, 1900, 1901 and 1902 (Gavrilova and Gavrilov 2018). Special attention is devoted to the data-cleaning procedure to identify whether the mortality deceleration at the older ages reported by some could be an artifact of poor data quality.

METHODS

The main source of data in this study is the Social Security Administration Death Master File (DMF) collecting information on deaths in the U.S. In this study, the DMF records are used to evaluate the quality of age reporting at extreme old ages and to estimate the force of mortality for ages 85 years and up, separately by gender, for the U.S. 1898–1902 birth cohorts. We use the last complete version of DMF officially obtained from the U.S. National Technical Information Service (NTIS). Taking into account that the latest deaths in this version were observed in September 2011, we use supplementary data sources for deaths occurring after September 2011 (Social Security Death Index, or SSDI, records on Ancestry.com and records of people age 110 years and over available from the Gerontology Research Group).

DMF records were used also to estimate the force of mortality by month of age for ages 85 years and up, separately by gender, for the U.S. 1898–1902 birth cohorts. Mortality is estimated using the method of extinct generations. Sex has been assigned using existing databases of male and female first names common in 1900, supplemented by additional male and female first names typical for that time. Overall, this procedure provided reliable assignment of sex for approximately 90% of all records. State of Social Security number (SSN) application was assigned based on the first three digits of a person's SSN. State of the last residence was assigned using information available in the

DMF (state code variable and ZIP codes of last residence of the deceased or residence to which the lump-sum death benefit was sent). An earlier study demonstrated that the quality of age reporting in the DMF is acceptable up to ages 106–107 years (Gavrilov and Gavrilova 2011). In this project, we used two approaches for data quality control: direct age validation using linkage to historical resources (such as early censuses) and indirect methods of data quality estimation.

A standard way to improve data quality of records with high longevity is to link death records to historical resources, such as early U.S. censuses, military records and birth certificates (Elo et al. 2013). We had an excellent success rate (over 97%) in linking individual records of centenarians to historical U.S. censuses in an earlier study on exceptional longevity (Gavrilov and Gavrilova 2015). In this study, we conducted a search of individuals in historical records using Ancestry.com, which has a powerful search engine that can find a person in multiple historical sources simultaneously, including all historical U.S. censuses that are open to the public. In this current work, we select records supported by historical resources and give the highest weights to age records from earlier historical sources, for example, 1910 or 1920 U.S. censuses or World War I civil draft registration cards.

We assessed data quality by age group through the direct age validation of samples of records of people at ages 100, 103, 105 and 109+ years taken from the DMF. The assessment involved individual age validation by linkage to historical resources available in the Ancestry.com database. The age of each person in the sample was evaluated by its concordance to a number of historical sources, including but not limited to decennial censuses. Depending on the result of this validation, the value for the age was rated as either good or poor quality.

In the first step of the age validation process, we developed and tested a scoring system based on linkage of DMF records to Ancestry.com resources. Our scoring system included the following six levels of data validity:

1. Several early sources (before 1950) agree about exact birth year with DMF record. Higher weight is given to earlier sources (preferably 1900 and 1910 U.S. censuses or birth certificates).
2. One earlier source agrees.
3. Age is mentioned in at least two independent post-1950 sources, such as obituary index, grave index or state directories, for which dates of birth agree.
4. An earlier source disagrees with DMF regarding age, or different dates of birth appear in other sources (incorrect age).
5. Foreign-born person came to the U.S. after age 20 (unreliable age).
6. Not found in any sources (possible unreliable record).

Records with quality scores 1–3 were considered to be of acceptable data quality, while records with quality scores 4–6 were considered to be of poor quality. We conducted an age validation procedure for samples of records for ages at death at 100, 103 and 105 years for the 1898, 1900 and 1902 birth cohorts that were randomly selected from corresponding age and cohort group. Age validation for the oldest age group (109+) was conducted for all records of five 1898–1902 single-year birth cohorts. In the case of the 1900 birth cohort, the 108+ age group was validated. In total, 1,753 records were validated. Study found that the proportion of records with poor quality of age reporting for the 1900 birth cohort increases from 12% at age 100 years to 32% at age 108+ years and from 13% at age 100 to 29% at age 109+ years in the case of the 1898 birth cohort. Proportion of poor quality records for the 1900 birth cohort at ages 103 years (19%) and 105 years (17%) was only slightly higher than the proportion at age 100 years (13%). Thus, the main drop in quality of age reporting occurs after age 105 years.

We used Pearson's chi-squared test to study differences in the proportion of records with poor quality (scores 4–6) across groups. According to this test, there are no statistically significant differences in data quality between the studied five birth cohorts at ages 100, 103, 105 and 109+ years. This test also did not find significant differences in data quality between men and women at all ages (up to age 108 years) with the exception of age 109+ years. At age 109+ years, men had a significantly higher percent of poor-quality records. Finally, the chi-squared test found statistically significant differences in data quality across age groups mainly due to a higher proportion of records with poor quality at age 109+ years. Our research found that the quality of age reporting did not differ significantly across the five studied historical birth cohorts, despite our initial expectation that more recent cohorts would have better data quality. In addition, a statistically significant increase was noted in the proportion of records with poor quality at age 109 compared to age 105, but no statistically significant difference was found in the proportion of poor-quality records within the ages of 100, 103 and 105 years.

In this study, we used actuarial estimates of the force of mortality (central death rate). Monthly (rather than traditional yearly) empirical estimates of the force of mortality are more accurate at extreme old ages. We found that monthly estimates of the force of mortality are higher after age 105 years compared to traditional yearly estimates (when recalculated in the same units). In addition, monthly estimates of the force of mortality produce more Gompertz-like mortality trajectories that translates to straighter lines on a semilog scale (log of mortality force as a function of age) compared to yearly estimates (up to age 108 years).

We looked at mortality difference by region, but no significant differences were found whether the regions were the region of last residence or the region where the Social Security number was obtained. Thus, regional mortality differences are negligible at extreme old ages. These results again confirm the more general observation that many factors affecting mortality during midlife lose their importance after age 100. Similarly, the year of birth had no effect on mortality for ages over 95 for the studied birth cohorts. That is, there was no expected improvement in mortality over time from the 1898 to the 1902 cohorts. This paradoxical finding corresponds well with earlier results about lack of improvement in mortality of centenarians in the last decades (Gavrilov, Krut'ko and Gavrilova 2017).

Our research confirms that the quality of age reporting after age 105 years in the U.S. is low for the studied cohorts which results in more apparent mortality deceleration at advanced ages. Data quality adjustments based on the results of the age validation procedure we used lead to more Gompertz-like monthly estimates of the mortality trajectories up to age 108 or older (depending on birth cohort and sex). Data cleaning (use of records with confirmed age after age 105 years) results in higher estimates of mortality at advanced ages (Gavrilov and Gavrilova 2019a). A high variability of mortality estimates prevents selection of a specific model for mortality trajectory after age 105 years. Earlier studies as well as this study demonstrated that age misreporting is associated with lower estimates of mortality (Gavrilov and Gavrilova 2019a; Preston, Elo and Stewart 1999). Thus, one may expect an improvement of quality in age reporting over time because of better education in subsequent birth cohorts (Elo et al. 2013). Indeed, study of old-age mortality in 1880–99 single-year U.S. birth cohorts found shifts from mortality deceleration in earlier birth cohorts (born before 1887) to Gompertz-like trajectories for later birth cohorts (Gavrilov and Gavrilova 2019b).

CONCLUSIONS

We can draw the following conclusions from the study of U.S. old-age mortality:

- Applying age validation to DMF records does not reveal sufficient differences in data quality among the studied 1898, 1899, 1900, 1901 and 1902 birth cohorts. No improvement in data quality over time was observed, contrary to our initial expectations.
- The age validation procedure revealed an increase in records with poor quality of age reporting after age 105 years, but not at earlier ages. Problems with data quality can be alleviated by removing observed and estimated (through the interpolation procedure) percentages of records with poor quality for ages 106 years and over.
- Data cleaning at the tail of the survival curve (based on the age validation) results in higher values of mortality estimates around ages 105–108 years compared to raw data. This implies that the observed U.S. mortality at this age interval is underestimated and mortality most likely follows the Gompertz law up to age 108 years.

- Monthly estimates of the force of mortality are higher after age 105 years compared to yearly estimates (when recalculated in the same units).
- Mortality for ages over 95 years does not demonstrate a decline over time in 1898–1902 birth cohorts.
- Mortality estimates by region of last residence reveal small, insignificant differences in old-age mortality without a discernible pattern.

ACKNOWLEDGMENTS

This study was supported by the Society of Actuaries. We would like to thank the members of the Project Oversight Committee for their valuable comments and help in conducting this study.

Natalia S. Gavrilova, Ph.D., is a senior research analyst at Academic Research Centers, NORC at the University of Chicago. She can be reached at *gavrilova@longevity-science.org*.

Leonid A. Gavrilov, Ph.D., is a senior research associate at Academic Research Centers, NORC at the University of Chicago. He can be reached at *gavrilov@longevity-science.org*.

REFERENCES

Elo, Irma T., Laryssa Mykyta, Paola Sebastiani, Kaare Christensen, Nancy W. Glynn and Thomas Perls. 2013. Age Validation in the Long Life Family Study Through a Linkage to Early-Life Census Records. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences* 68, no. 4:580–85.

Gavrilov, Leonid A., and Natalia S. Gavrilova. 2011. Mortality Measurement at Advanced Ages: A Study of the Social Security Administration Death Master File. *North American Actuarial Journal* 15, no. 3:432–47.

———. 2015. Predictors of Exceptional Longevity: Effects of Early-Life and Midlife Conditions, and Familial Longevity. *North American Actuarial Journal* 19, no. 3:174–86.

———. 2019a. Late-Life Mortality is Underestimated Because of Data Errors. *PLoS Biology* 17, no. 2. *https://doi.org/10.1371/journal.pbio.3000148*.

———. 2019b. New Trend in Old-Age Mortality: Gompertzialization of Mortality Trajectory. *Gerontology* 65, no. 5:451–57. *https://doi.org/10.1159/000500141*.

Gavrilov, Leonid A., Vyacheslav Krut'ko and Natalia S. Gavrilova. 2017. The Future of Human Longevity. *Gerontology* 63, no. 6:524–26. *https://doi.org/10.1159/000477965*.

Gavrilova, Natalia S., and Gavrilov, Leonid A. 2018. Mortality Analysis of 1898–1902 Birth Cohort. Society of Actuaries report.

Preston, Samuel H., Irma T. Elo and Quincy Stewart. 1999. Effects of Age Misreporting on Mortality Estimates at Older Ages. *Population Studies* 53, no. 2:165–77.