LIVING
to 100
SOCIETY OF ACTUARIES
INTERNATIONAL SYMPOSIUM

2020 Symposium
Jan. 13–15
Lake Buena Vista, FL

2C - Teaching Session: Data Validation

SOA Antitrust Disclaimer
SOA Presentation Disclaimer

2020 Living to 100 Symposium

EVALUATION OF QUALITY OF MORTALITY ESTIMATES AT ADVANCED AGES
KIRILL ANDREEV

2C: Teaching Session: Data Validation

January 13th, 2019

Draft presentation 5 Jan 2019, not to be cited or copied

SOCIETY OF ACTUARIES®

Disclaimer: The views expressed here do not imply the expression of any opinion on the part of the United Nations Secretariat

Mortality Estimation

Typical steps to derive direct estimates of death rates based on data on deaths from vital registration and data on population from censuses (or population registers):

A) Estimate death counts by year, age, and cohort (by so-called Lexis triangles)

B) Estimate population at 1 January for each calendar year and by single year of age

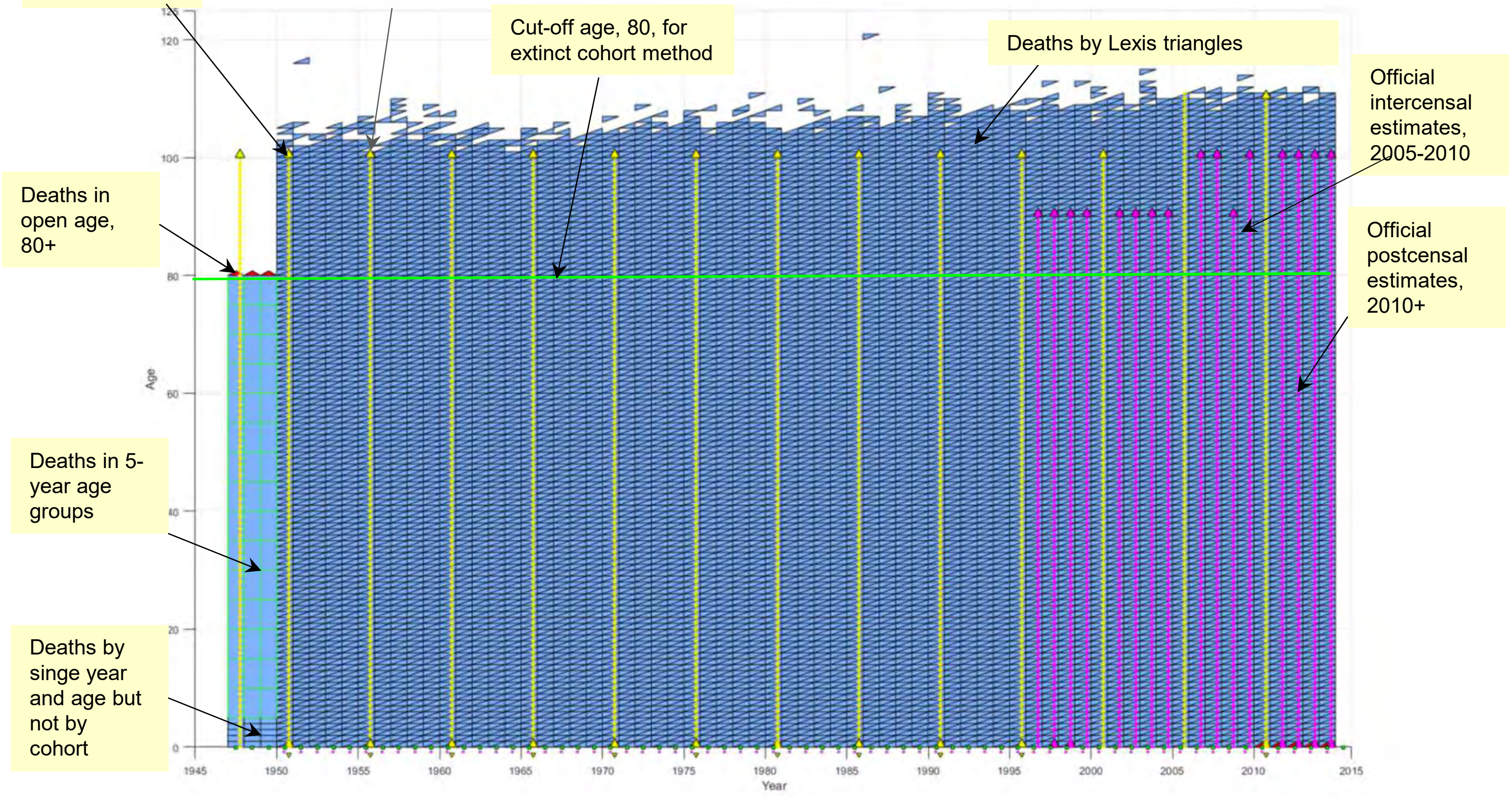
1) for ages below 80 based on census population or/and official population estimates

2) for ages above 80 computed by (almost) extinct cohort method

C) Compute death rates and period/cohort life tables bases on the estimated death counts and population exposure. The data could be further also aggregated by age, period, over countries.

Selected references: Vincent, (1951), Thatcher et al. (2002), Andreev et al. (2003), Wilmoth et al. (2007)

Data on population and deaths available for mortality estimation Japan, Males



1950 census population, ages 0, 1, … 100+

1955 census

Cut-off age, 80, for extinct cohort method

Deaths by Lexis triangles

Official intercensal estimates, 2005-2010

Official postcensal estimates, 2010+

Deaths in open age, 80+

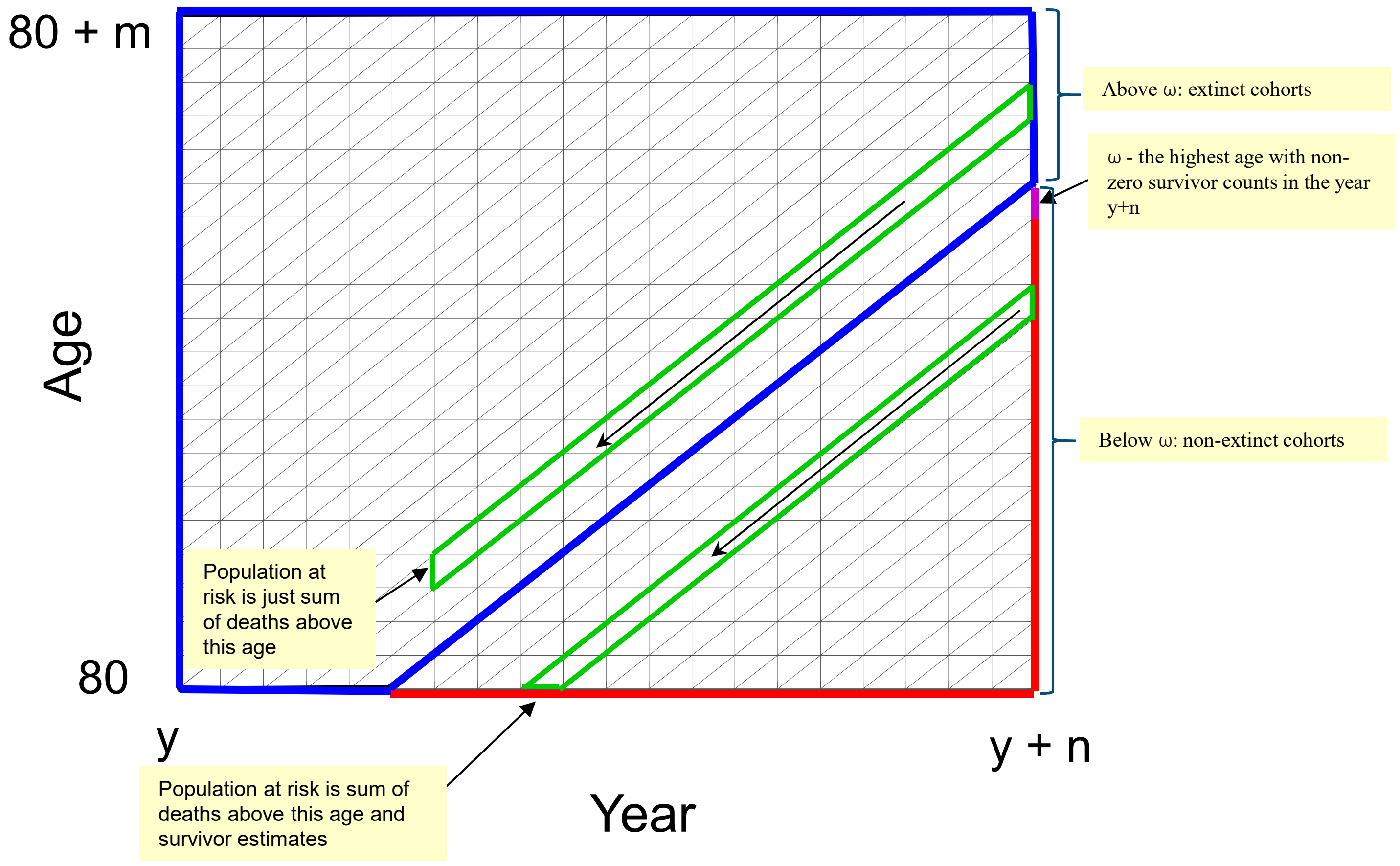
Deaths in 5-year age groups

Deaths by singe year and age but not by cohort

Steps to compute death rates

1) Distribute deaths by Lexis triangles, years 1947-1949 (if not readily available)
2) Compute intercensal estimates, years 1947-1995.Typically based on assumption that residual migration is distributed evenly along the cohorts.
3) For ages 80+, compute population estimates for extinct and almost extinct cohorts by (almost) extinct cohort method
4) Compute death rates by Lexis triangles or rectangles (over single calendar year and age)

Almost Extinct Cohort Method



Above ω: extinct cohorts

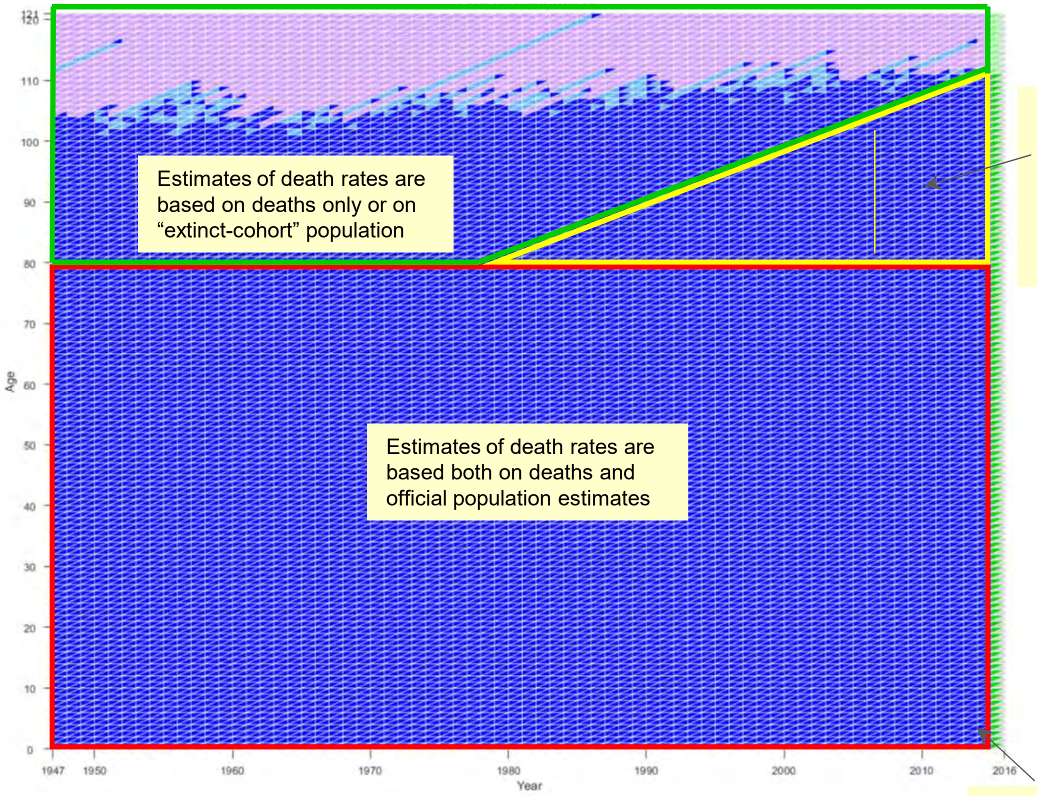
ω - the highest age with non-zero survivor counts in the year y+n

Below ω: non-extinct cohorts

Population at risk is just sum of deaths above this age

Population at risk is sum of deaths above this age and survivor estimates

Resulting estimates of population and deaths*
Japan, Males



Estimates of death rates are based on deaths only or on “extinct-cohort” population

Estimates of death rates are based both on deaths and official population estimates

Estimates of death rates are based both on deaths and survivor estimates

Last year with available data

* This matrix of population, deaths and death rates is also known as Lexis database

Legend

▨ White lines represent population estimates

◢ Blue triangles represent Lexis triangles with population at risk > 0 and death counts > 0 (Mx > 0)

◢ Light blue triangles represent Lexis triangles with population at risk > 0 and death counts equal to 0 (Mx = 0)

◤ Green triangles represent Lexis triangles with population > 0 and missing death counts (survivors of non-extinct cohorts)

◤ Light green triangles represent Lexis triangles with population at risk equal to zero 0 (extinct cohorts)

◢ Purple triangles represent Lexis triangles with both deaths and population at risk equal to zero 0 (Lexis triangles above the highest age at death in a given cohort)

Common Problems in Demographic Data

Age misreporting

- Age heaping (digit preference)
- Age understatement
- Age exaggeration (overstatement)
- Random (symmetric) age misreporting

Completeness and coverage

- death registration
- population censuses
- Mismatch between coverage of death registration and population

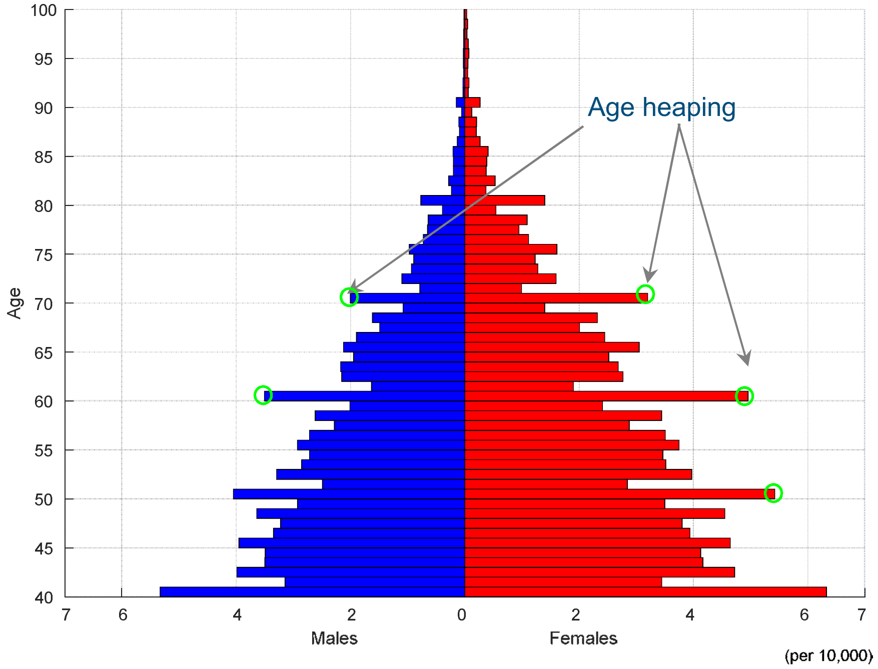
Age Heaping

- Tendency to round the reported age to “convenient” numbers
- Affects both self-reported data (censuses, surveys, administrative data (e.g. Social Security), and not self-reported (death registration), and non-demographic data
- Easy to detect visually and in tables by elevated counts of population and deaths at “round” ages
- In the estimated death rates it manifests itself by significantly elevating (or lowering) death rates by at “round” ages
- If data on date of birth collected, “heaping” could occur at “convenient year of birth” (e.g. 1900), both in death certificate and census data

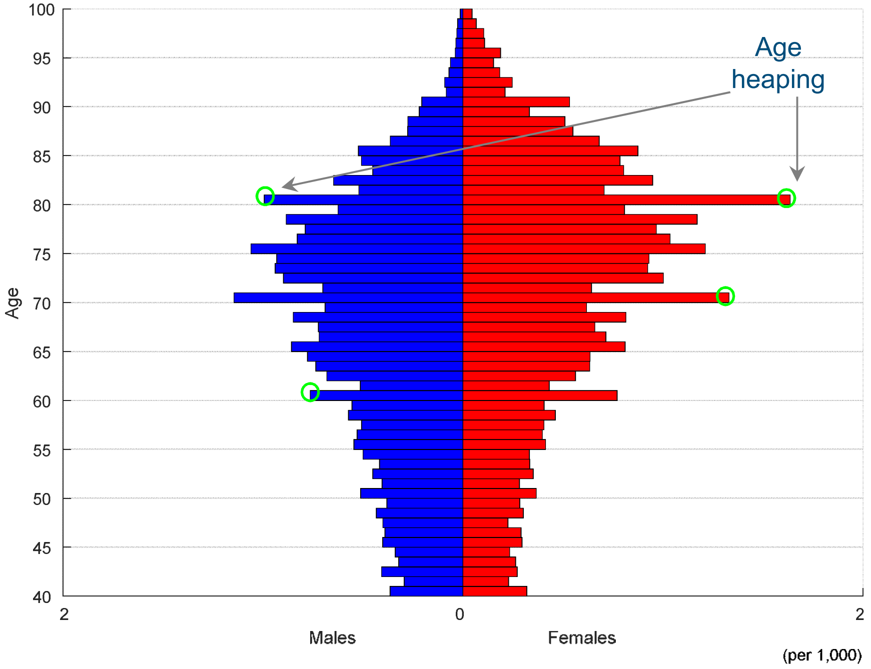
Selected references: Shryock and Siegel (1980)

Example of age heaping in data on population and deaths

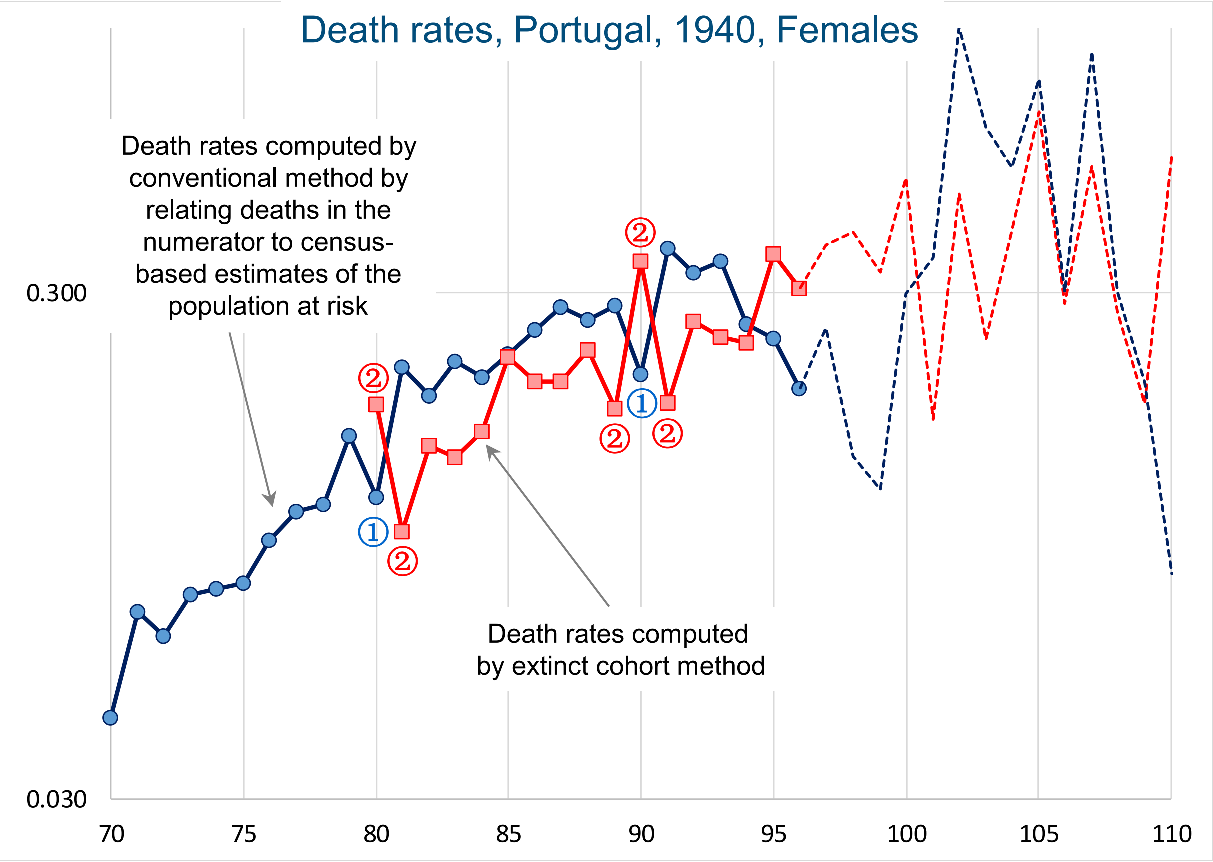
Population by age and sex
Portugal, 1940 Census



Distribution of death counts
Portugal, year 1940



Effect of age heaping on mortality estimates


Death rates, Portugal, 1940, Females

Death rates computed by conventional method by relating deaths in the numerator to census-based estimates of the population at risk

Death rates computed by extinct cohort method

① At round ages (80, 90) death rates computed by conventional method are *lower* than death rates at adjacent ages because of stronger age heaping in census population than in death registration data

② Death rates based on extinct cohort population are *higher* than death rates at adjacent ages due heaping of deaths in death registration data

③ Extinct cohort population estimates are about 10% *higher* than census population for ages 80 or over, about 15% *lower* at ages 85 or over, and about 40% *lower* at ages 90 or over.

Detecting and Quantifying Age Heaping in Demographic Data

- Data visualization

- Comparisons of raw and “smoothed” data and indexes of age heaping based on such comparisons

Indexes of age heaping and related procedures:

- Whipple's index
- Myers' blended method
- Local comparisons e.g. deaths at age 90 vs. log-average of deaths at ages 88-92, ratio of q(80)/q(81) (Kannisto,1999)
- Camarda et al. (2008)– underlining distribution is a smoothed p-spline while deaths/reallocated from adjacent years (possible to estimate reallocation probabilities under constraints and model assumptions)
- Andreev (1998): residual analysis of mortality surfaces smoothed by kernel smoothing (e.g. local Poisson regression) and mortality models methods (suitable for the large bodies of data and for exploring changes in data quality over time)

Adjusting for age heaping

Smoothing of the observed distribution of population or deaths

- Aggregation into age groups purportedly not affected by age heaping and interpolation into single ages again
- Kernel density smoothing (e.g. moving averages or extensions to two dimensions)
- Smoothing splines (e.g. cubic, p-splines)

Effect of age misreporting on death and population distributions and on mortality estimates

- At older ages, age misreporting (both age exaggeration and age understatement) lead to increasingly more population and death counts reported at higher ages as compared with real data. In other words, age misreporting inflates tails of real distributions of death and census counts.

- Death rates based on the data affected by age misreporting are commonly biased downwards

- The bias becomes increasingly severe with age resulting in implausibly low levels of directly computed death rates, low rates of mortality increase over age (slopes of mortality schedules), or even in declines in death rates with age

- Selected population groups in the same country could be affected differently by age misreporting (e.g. males and females, black and white population in the United States, Han Chinese, Māori and non- Māori population of New Zealand)

Selected references: Coale and Kisker (1990), Preston et al. (1999), Andreev et al. (2017)

Detecting Age Misreporting in Demographic Data

Common approach is to explore “plausibility” of the observed data e.g.
a) levels of death rates as compared with mortality estimates for other periods and for countries/populations
b) age patterns of death rates

Matching/linkage studies similar to age validation procedures of extreme-aged individuals (less common approach).

In general, detecting age misreporting is harder than detecting age heaping

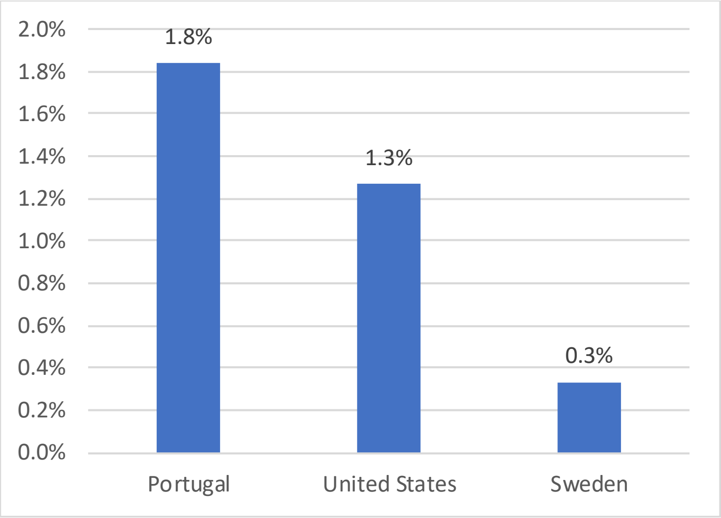
- Not clear-cut conclusions could be derived from the visual inspection of the data as for age heaping
- Often, the data supposedly affected by age misreporting are argued as real. For example, the lower levels of mortality are argued as real and observed mortality crossovers are argued as effect of selection of robust individuals.

Data quality indicators based on tails of the observed distributions

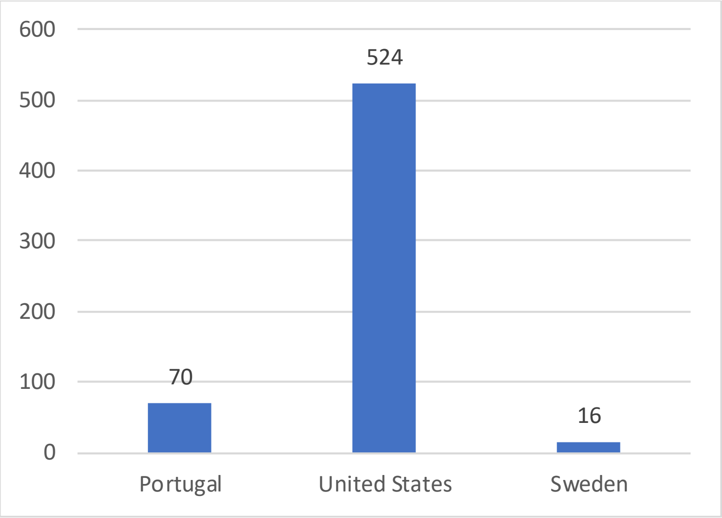
1) Ratio of deaths at ages 105+ to ages 100+, $\frac{D_{105+}}{D_{100+}}$ or ratio of deaths at ages 100+ to ages 85+, $\frac{D_{100+}}{D_{85+}}$ (Kannisto, 1999)
2) The highest age reported (or 2nd, 3d highest etc.) or percentiles (90th, 99th, 99.9th) of reported ages at death based on deaths above age 50 only (Wilmoth and Lundstrom, 1996)
3) Proportions of centenarians $\frac{N_{100+}}{N}$ or supercentenarians $\frac{N_{110+}}{N}$ or $\frac{N_{95+}}{N_{70+}}$ etc. in censuses or population estimates

Example of data quality indicators based on tails of the observed distributions

Ratio of deaths at ages 100+ to deaths at ages 85+, years 1940-49 (percent)



Proportion of centenarians in the total population (per million)



United States and Portugal: 1970 census
Sweden: population estimates based on population register

See Siegel and Passel (1976) for discussion of errors in the 1970 U.S. census and possible was to adjust

Data quality indicators based on plausibility of levels and age pattern of death rates

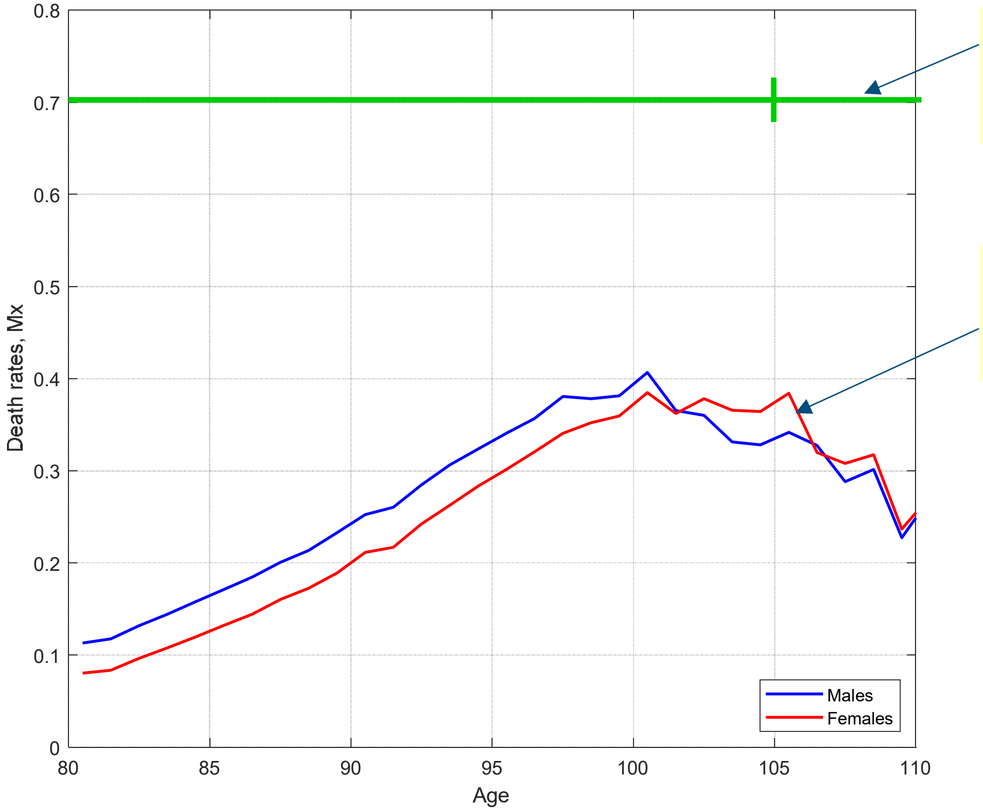
General assumptions:

- General pattern: after age about 30 death rates begin to increase approximately exponentially (Gompertz), level off after age 85, and reach a plateau at ages about 105-110 with maximum $\mu(105)$ = 0.7 or annual probability of dying of q_x = 0.5 or remaining life expectancy $1/\mu(105)$ = 1.4 years
- No firm evidence exists now that death rates decline over age at older ages in any human population

Data quality checks procedures:

- Visual inspection of age patterns of mortality
- Comparing indicators of general mortality and old age mortality between countries (e.g. scatter plots) -- positive correlation between indicators of old age mortality and indicators of overall or adult mortality is expected
- Comparing age patterns of mortality with model age patterns (often based on mortality in selected countries and periods or on “mortality model” e.g. Gompertz, logistic, Kannisto)

Death Rates in the United States, 1959-1969, based on reported deaths and extinct cohort population



Level of “supercentenarians” mortality, μ_x=*0.7*

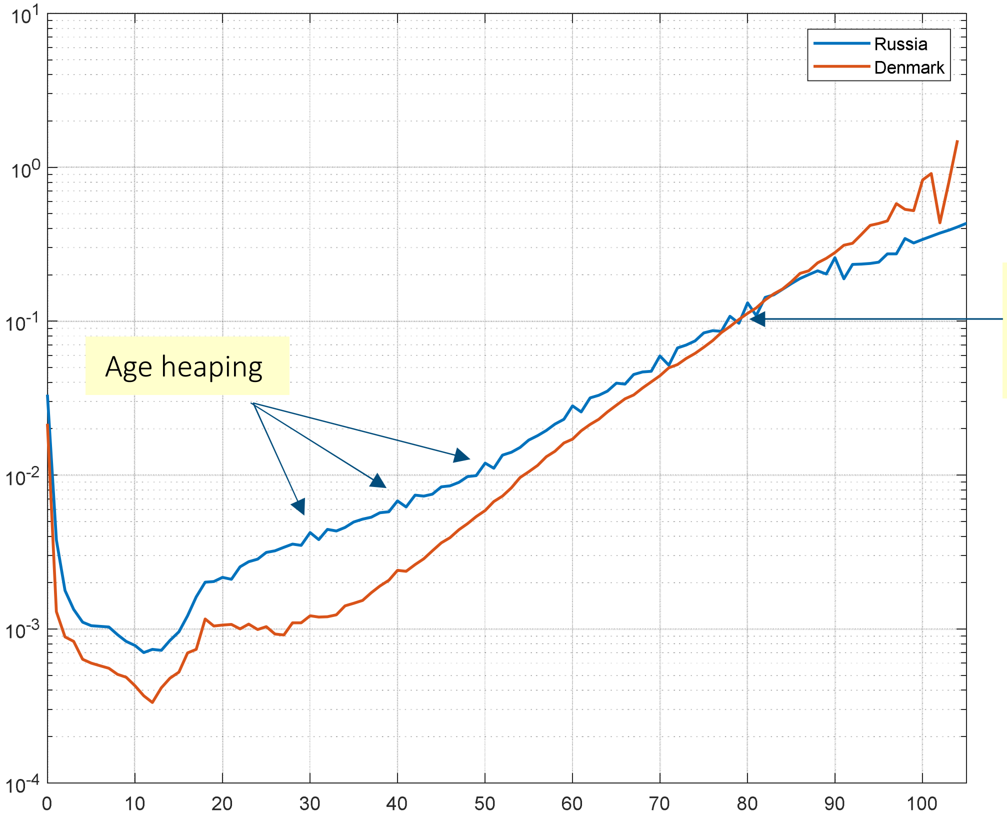
Decline in death rates with age due to age misreporting

Data quality indicators based on plausibility of mortality differences between subpopulations

- Sex ratio of population or deaths over age
- Mortality crossovers e.g. between males and females, black and white population in the United States, between countries
- Mortality differences between general population and subgroups with presumably better age reporting (e.g. Kestenbaum, 1992; Coale and Li, 1991)

Recent review of mortality crossovers: Arias (2019)

Death Rates in Russia and Denmark, Males, years 1960-1969



Age heaping

Mortality crossover due to age misreporting in Russia

Data quality indicators based on trends in death rates

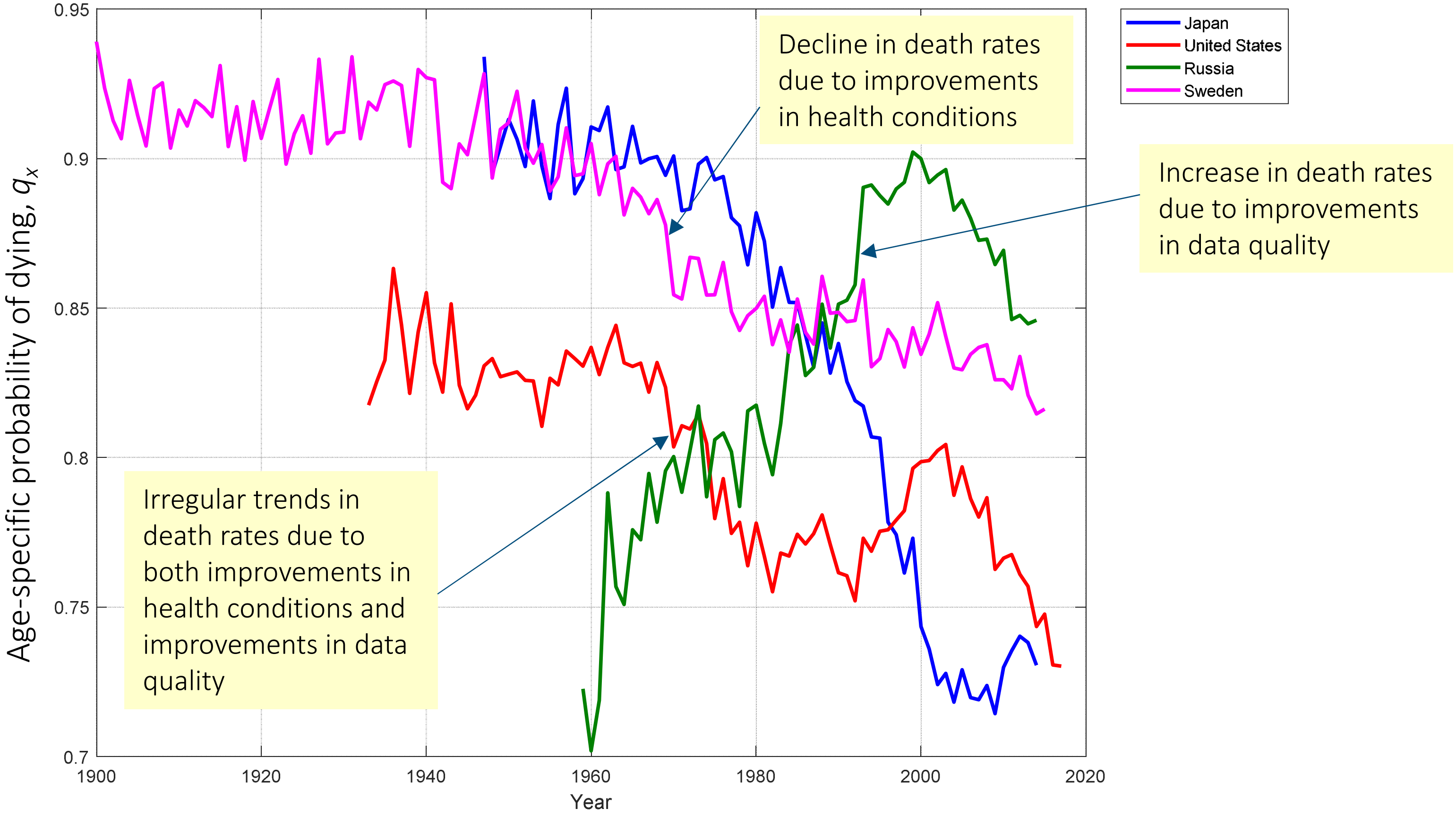
- Quality of data at advanced ages depends primarily on well-functioning civil registration and vital statistics systems and their period of operation in a country

- Data quality tends to improve over time as birth registration covers progressively older segments of population, so-called a ‘‘birth registration effect’’ (Anderson, 1999).

- Combined with age misreporting in earlier periods, improvements in data quality over time could lead to spurious increases or stagnation in death rates while in fact death rates were declining

- Increasing age of mortality crossovers could be an indicator of improvements in data quality

Trends in age-specific probability of dying, 5q95, based on reported deaths and almost extinct cohort population



Decline in death rates due to improvements in health conditions

Increase in death rates due to improvements in data quality

Irregular trends in death rates due to both improvements in health conditions and improvements in data quality

Matching (Record Linkage) Studies

Quality of age reporting on death certificates can be validated by checking age of the same individual reported in independent data sources:

- Birth registration
- Age reported in censuses
- Administrative data sources e.g. Social Security Administration, pension funds etc.

For matching studies in the United States see Rosenwaike and Logue (1983), Preston et al. (1996), Hill et al. (2000).

Adjusting death rates for age misreporting

a) Accept death rates up to a certain age as valid and extrapolate above that age by a mortality model

1) quadratic (Coale and Kisker, 1990)
2) Kannisto model (logistic with asymptote = 1) (Thatcher, 1999)
3) model life tables (often closed by Gompertz (Coale-Demeny) or Kannisto (United Nations) models at the highest ages)

b) Similar to a) but using observed age-specific rates of increase in death rates in a population with similar coverage

1) Medicare population in the United States
2) Life tables from neighboring countries based on data of presumably better quality

c) Combining a) and b) with observed data purportedly unaffected by age misreporting e.g. open age death rates M(x+) or age-specific growth rates (Elo and Preston, 1994)

Adjusting death rates for age misreporting in matching studies (Preston et al., 1996)

Sample size: 5,262 death certificates

Three independent data sources on age of African Americans aged 65+ in 1985:

- Death certificates
- Social security administration records (independent age reporting and strict age validation for people enrolled after 1965 (or born after 1900)
- Age reporting in 1900, 1910, and 1920 censuses (proxy for birth registration)

Three-way linkage was accomplished for 2,657 records, or 50.5%.

"Final" age at death was assigned based on the three-way link, case-by-case basis and used to produce “adjusted” distribution of deaths

The “adjusted” distribution of deaths was converted into age-specific death rates by applying age-specific growth rates (unadjusted) derived from census and population estimates.

The “adjusted” death rates turned out to be lower at ages below 85 and higher above this age, about 30% higher for age group 90-94. The black-white white mortality crossover disappeared as compared with unadjusted white death rates.

Summary

- Existing methods of checking for age misreporting help to highlight possible problems in the data
- No unified or automated approach currently exists for assigning a data quality score or for adjusting death rates at older ages
- Age misreporting generally biases mortality rates downwards and the bias is increasing with age
- Age misreporting and improvements in data quality biases downwards mortality improvements at older ages. As a result, extrapolative mortality projection methods (e.g. Lee-Carter model) applied to such data will underestimate future mortality improvements at older ages and population sizes
- Reliable population data on old age mortality for *extended periods of time* are available for about 13 countries only out of about 200 countries or areas in the world
- Data quality is improving over time and for increasingly more countries direct mortality estimates could be produced based on observed data only, without resorting to adjustments or models

References

Andreev, K. 1999. Demographic Surfaces: Estimation, Assessment and Presentation, with Application to Danish Mortality, 1835–1995: Ph.D. thesis, University of Southern Denmark.

Arias E. (2019) Race Crossover in Longevity. In: Gu D., Dupre M. (eds) Encyclopedia of Gerontology and Population Aging. Springer, Cham

Anderson, R. N. 1999. Method for Constructing Complete Annual U.S. Life Tables. National Center for Health Statistics. Vital and Health Statistics 2(129).

Camarda, C. G., Eilers, P. H. C. and Gampe, J. (2008) Modelling general patterns of digit preference. Statistical Modelling, 8, 385–401.

Coale, A.J., and 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.

Coale, A. and S. Li (1991) The effect of age misreporting in China on the calculation of mortality rates at very high ages. Demography 28(2)

Elo, I. and Preston, S. 1994. Estimating African-American Mortality from Inaccurate Data. Demography, 31(3): 427-458

Hill, M. et al. 2000. Age Reporting among White Americans Aged 85+: Results of a Record Linkage Study. Demography.

Kannisto, V. Assessing the Information on Age at Death of Old Persons in National Vital Statistics. In: Jeune B, Vaupel JW, eds. Validation of Exceptional Longevity. Odense, Denmark: Odense University Press, 1999:173–188. Online at: https://www.demogr.mpg.de/books/odense/6/index.htm

Kestenbaum, B. 1992. A Description of the Extreme Aged Population Based on Improved Medicare Enrollment Data. Demography 29: 565–580.

Preston, S.H., I.T. Elo, I. Rosenwaike, and M.E. Hill. 1996. "African-American Mortality at Older Ages: Results of a Matching Study." Demography 33:193-209.

Rosenwaike, I. and B. Logue. 1983. "Accuracy of Death Certificate Ages for the Extreme Aged." Demography 16:279-88.

Siegel, J. S. and J. S. Passel, 1976, New Estimates of the Number of Centenarians in the United States, Journal of the American Statistical Association, Vol. 71, Vol. 71, No. 355, pp. 559-566.

Shryock H. S., Siegel, J. S., and Associates. (1980). The methods and materials of demography. U.S. Department of Commerce. Bureau of the Census. Fourth Printing (rev.). U S Government Printing Office Washington, D.C

Thatcher, A.R. 1999. The Long-Term Pattern of Adult Mortality and the Highest Attained Age. Journal of the Royal Statistical Society 162 Part 1:5-43.

Wilmoth, J. R. and Lundstrom, H. (1996) Extreme Longevity in Five Countries - Presentation of Trends with Special Attention to Issues of Data Quality. European Journal of Population-Revue Europeenne De Demographie, 12, 63-93.

Acknowledgements

The author acknowledges contribution of the United Nations by making office and IT resources available outside the normal working hours and on weekends

and excellent emotional support of Toby Andreev over this project
(and for keeping me out of that office)

No financial support was received for preparing this presentation

2020 Living to 100 Symposium

NATALIA S. GAVRILOVA, LEONID A. GAVRILOV

Session 2C, Teaching Session: Data Validation

Date: January 13, 2020

SOCIETY OF ACTUARIES®

Presentation Disclaimer

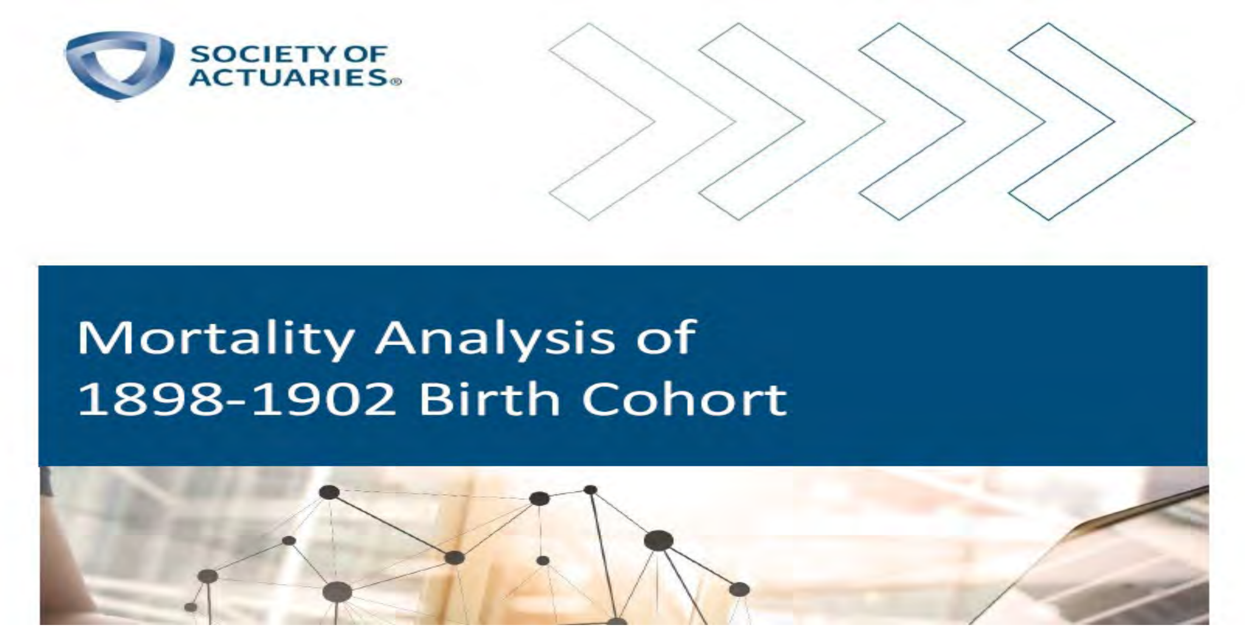
Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the participants individually and, unless expressly stated to the contrary, are not the opinion or position of the Society of Actuaries, its cosponsors or its committees. The Society of Actuaries does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented.

What is the quality of age reporting in the SSA Death Master File across ages and birth cohorts?

A study of data quality for five single-year birth cohorts using direct validation method

Supported by the Society of Actuaries

More details are available in a special report by the Society Of Actuaries



SOCIETY OF ACTUARIES

Mortality Analysis of 1898-1902 Birth Cohort

Supported by the Society of Actuaries

https://www.soa.org/resources/research-reports/2018/birth-cohort/

Study Design

Five single-year birth cohorts:
1898, 1899, 1900, 1901, 1902

Direct age validation of Death Master File (DMF) samples randomly selected at ages 100, 103 and 105 years

Sample sizes: 100 records for ages 100,103,105 years
For age group 109+ years – all available records

Age validation procedure

Age validation was conducted by linkage of DMF records to early historical sources (U.S. censuses, birth and marriage records, draft registration cards). Data linkage was done for 2,711 records.

DMF records were scored according to reliability of age reporting. The scoring system included the following scores:

1 – several early (1950 or earlier) historical sources agree about birth date

2 – one early historical sources agrees about birth date

3 – later sources agree about birth date

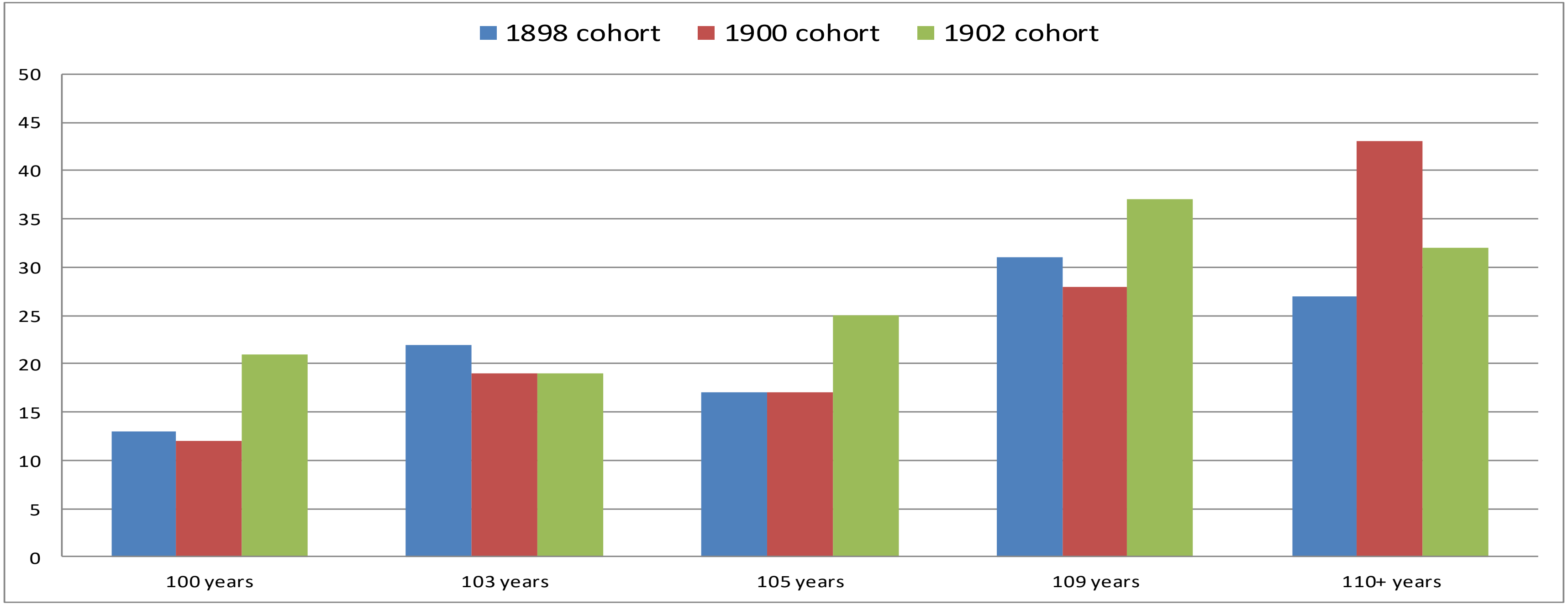
Records with questionable quality:

4 – early sources disagree with DMF record

5 – foreign-born individual arrived in the U.S. later in life

6 – not found in any sources

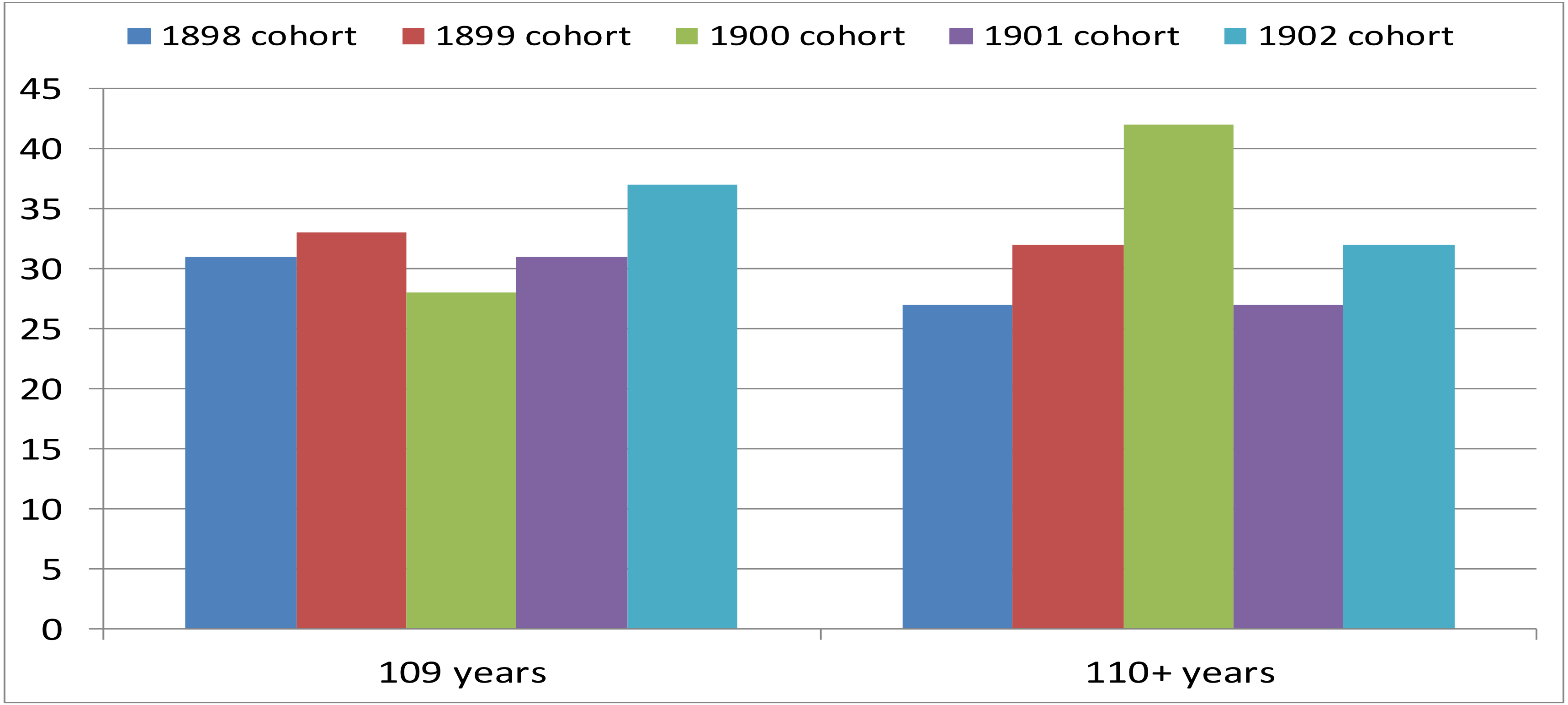
Percent of records with questionable quality as a function of age. 1898, 1900 and 1902 birth cohorts



Results of age validation study for samples of 100 records, by age group.

For ages 109 and 110+ years sample sizes were slightly higher than 100.

Percent of records with questionable quality at extreme old ages. 1898-1902 birth cohorts



Regression model for percentage of poor quality records

Percentage of poor quality records is modeled as a linear function of binary (dummy) variables representing birth cohorts and ages.

$$percent = \text{const} + \square_1 AGE + \square_2 COHORT$$

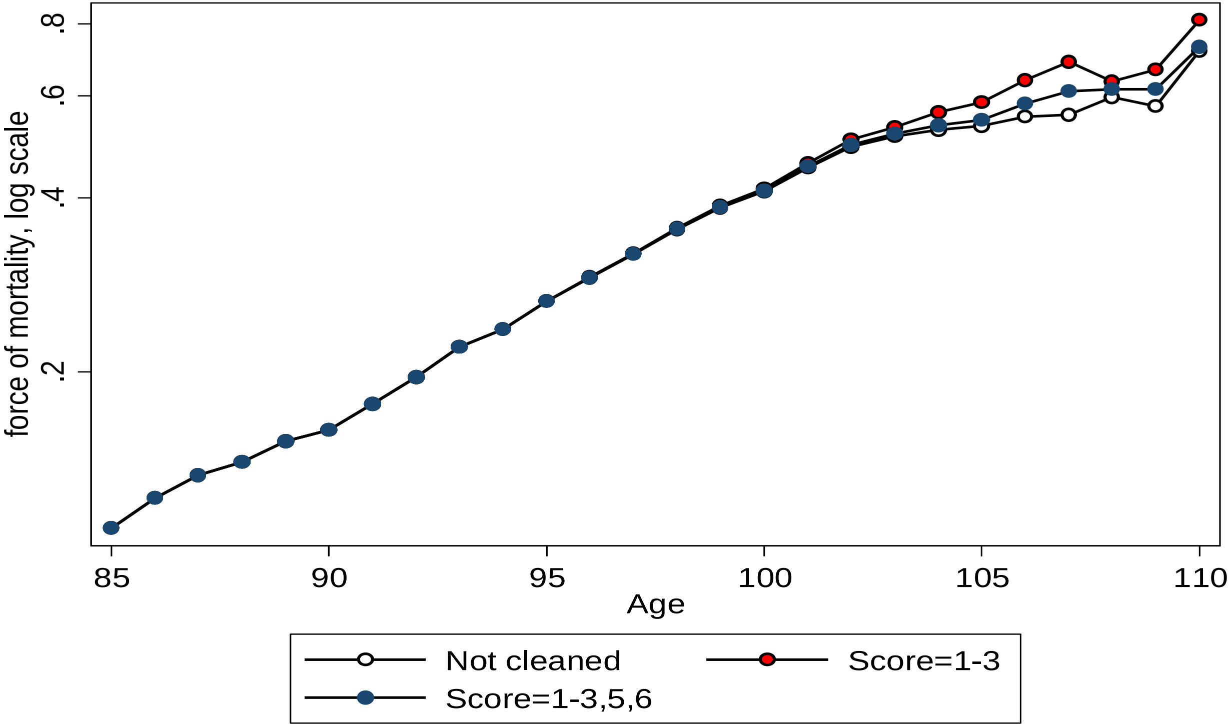
where percent is percentage of poor quality records, AGE and COHORT represent sets of dummy variables (103, 105, 109 for AGE at death with 100 years used as a reference level and 1899, 1900, 1901, 1902 for COHORT birth year with 1898 used as a reference level), β_1 and β_2 are regression coefficients

Regression model for percentage of poor quality data

Variable	Regression coefficients	P-value	95 % confidence intervals
1898 cohort	reference		
1899 cohort	2.00	0.588	-6.55 - 10.55
1900 cohort	-1.75	0.419	-6.69 - 3.19
1901 cohort	-1.56e-15	1.000	-8.55 - 8.55
1902 cohort	4.75	0.057	-0.19 - 9.69
Age 100	reference		
Age 103	4.67	0.092	-1.03 - 10.37
Age 105	4.33	0.112	-1.37 - 10.03
Age 109	**16.67**	**<0.001**	**10.97 - 22.37**
Intercept	**14.33**	**<0.001**	**9.40-19.27**

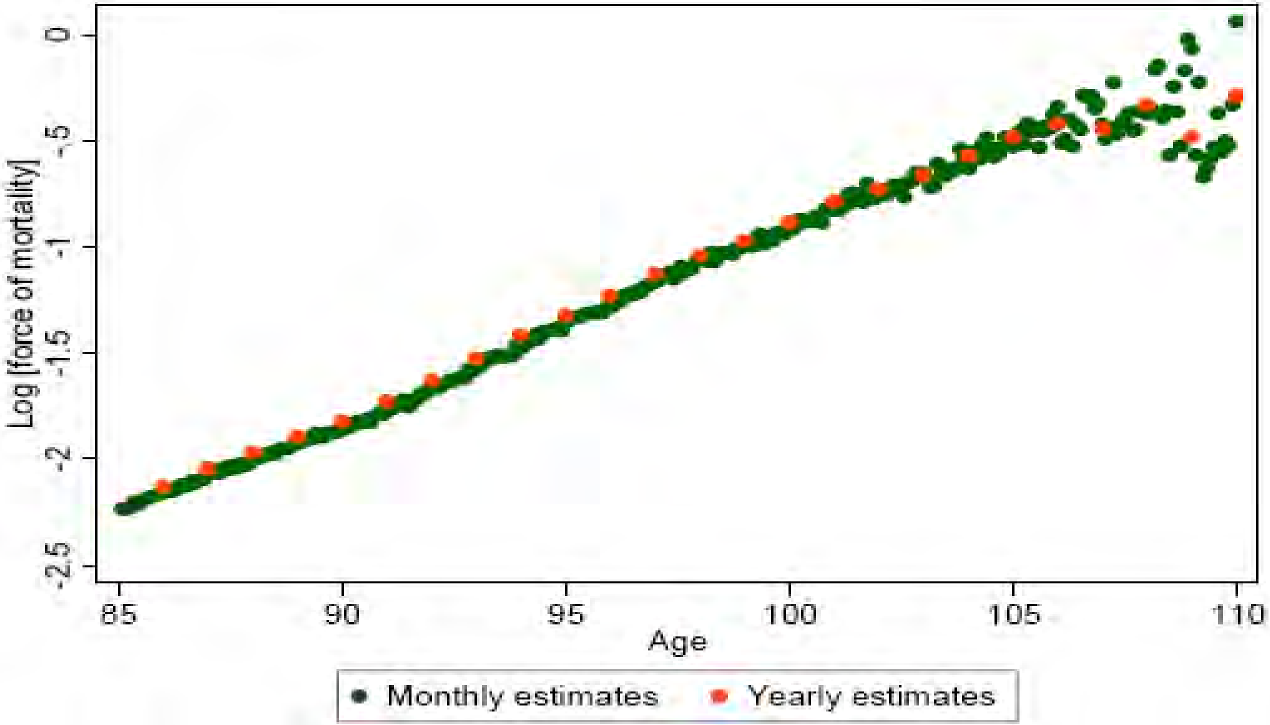
Force of mortality by the data quality score

1900 birth cohort, both sexes



Force of mortality by monthly and yearly estimates after data cleaning

1898-1902 birth cohort, both sexes



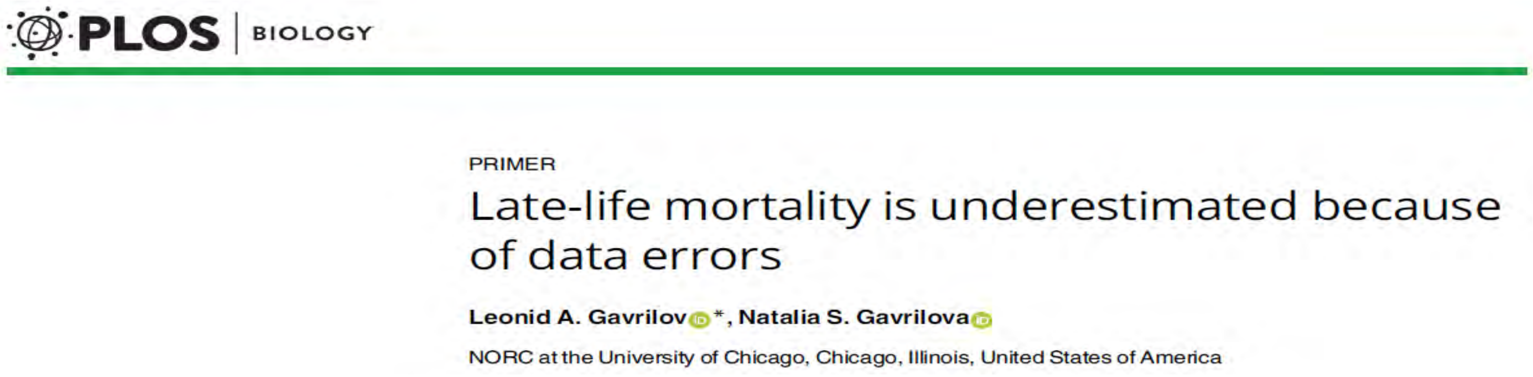
Hypothesis

Mortality deceleration at advanced ages among DMF cohorts may be caused by poor data quality (age exaggeration) at very advanced ages

If this hypothesis is correct then mortality deceleration at advanced ages should be less expressed for data with better quality

Further development

Direct age validation of all records at ages 106 years and over for those born in 1900



PLOS | BIOLOGY

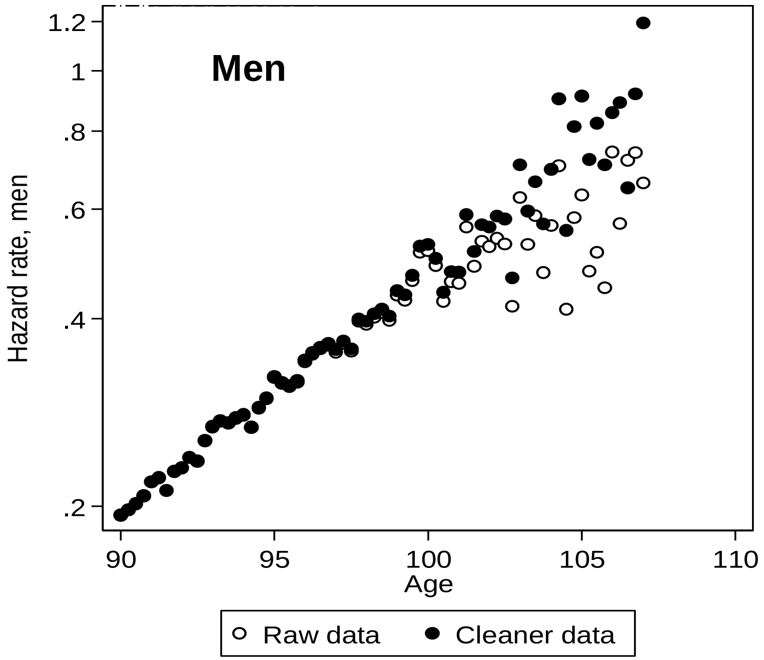
PRIMER

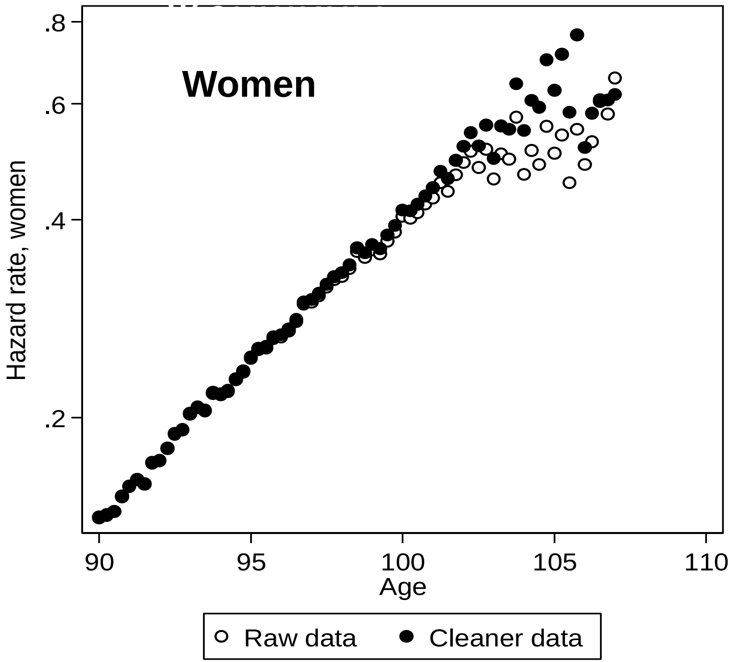
Late-life mortality is underestimated because of data errors

Leonid A. Gavrilov*, Natalia S. Gavrilova

NORC at the University of Chicago, Chicago, Illinois, United States of America

Mortality of U.S. men and women born in 1900 depending on data quality. Age validation conducted for all records aged 106 years and over. Mortality plateau disappears for cleaner data (black circles)





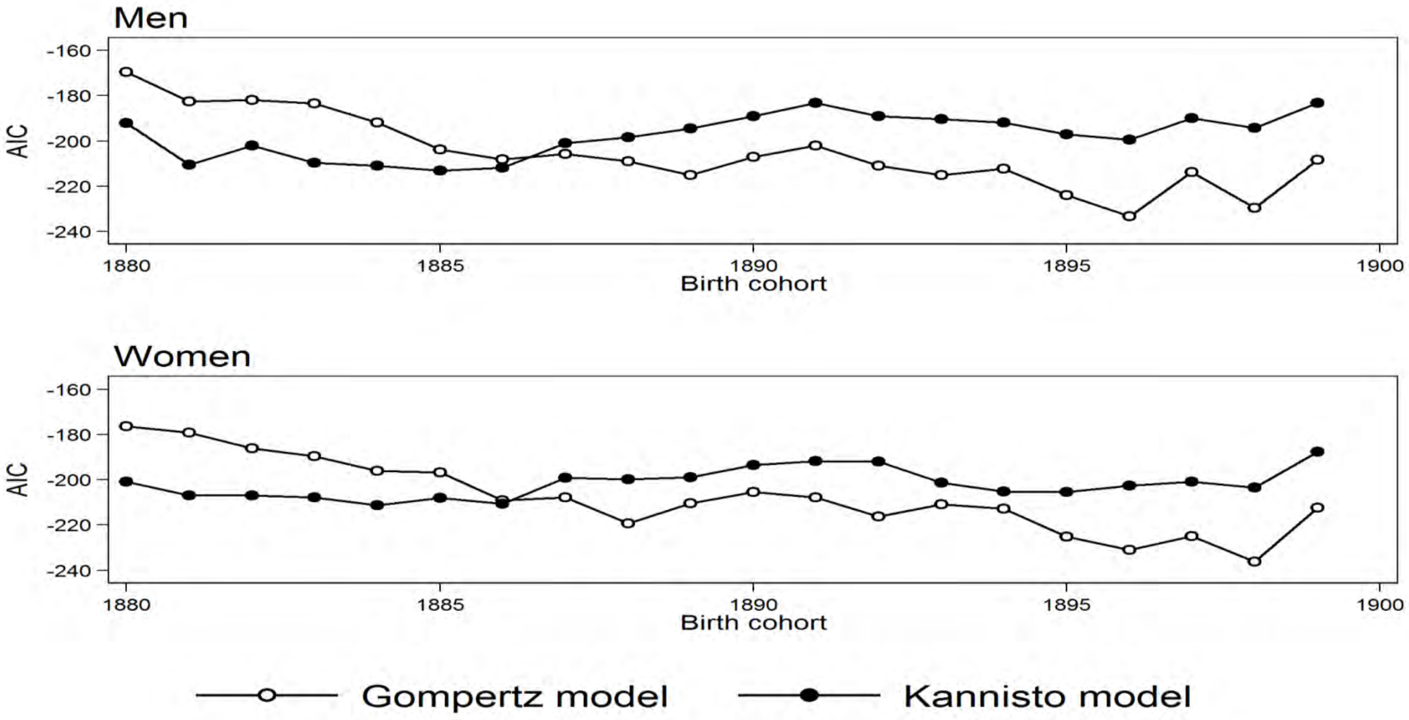
Source: Gavrilov, Gavrilova, *PLOS Biology*, 2019. Data Source: DMF

Is Mortality Deceleration Caused by Age Misreporting?

Age misstatement biases mortality estimates downwards at the oldest ages, which contributes to mortality deceleration (Preston et al., 1999).

If this hypothesis is correct then mortality deceleration should be more prevalent among historically older birth cohorts

Gompertz model outperforms mortality deceleration (Kannisto) model for more recent birth cohorts

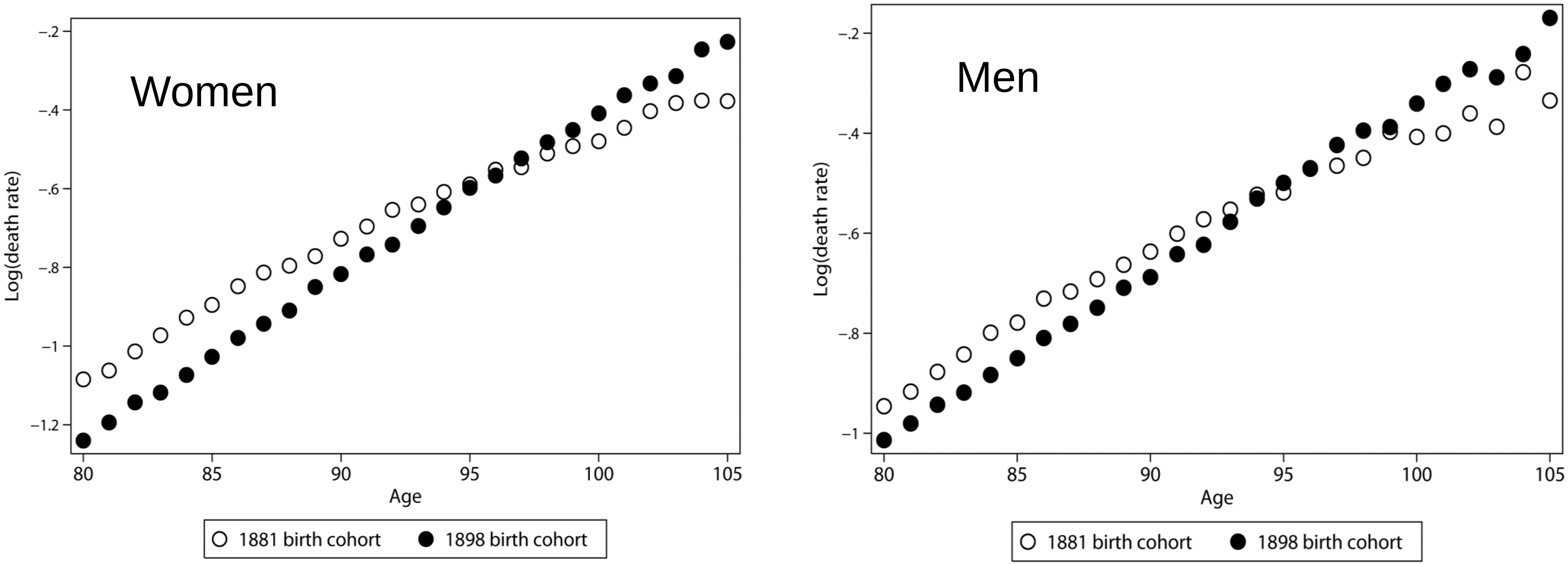


Akaike information criterion (AIC) for Gompertz and Kannisto models. 40 U.S. Birth Cohorts

Source: Gavrilov, Gavrilova, *Gerontology,* 2019. Source of data: Human Mortality Database

Mortality of U.S. men and women in earlier (1881) and later (1898) birth cohorts

Mortality deceleration is observed in early birth cohort only



Source: Gavrilov, Gavrilova, *Gerontology,* 2019. Source of data: Human Mortality Database

Conclusion

Mortality deceleration is more prevalent in historically older birth cohorts when age reporting was less accurate

SOCIETY OF ACTUARIES®

2020 Living to 100 Symposium

2C: Teaching Session: Data Validation

Tom Perls MD, MPH, New England Centenarian Study

BOSTON UNIVERSITY SCHOOL of Medicine

Outline

- Sources of invalid claims
- Red flags of an invalid age claim
- Validating age in the New England Centenarian Study
- Recruiting Issues and Sample Size

SOCIETY OF ACTUARIES®

99% of age claims > 115 years are false

Daily Mail

Mark Porter, Dec. 28, 2019

- “The Island [Dominica] has the highest number of centenarians in the world, and it was from here that the oldest person ever hailed”
- “Born in 1875, Ma Pampo died in 2003, aged 128. She didn’t retire until she was 104. Her next door neighbor [Rose Peter] was 118…”
- “At last count there were 27 centenarians –that’s nearly four per 10,000 of population, 50% higher than the next old-age market leader, Japan, and three times as many as in Britain and the U.S.”

BOSTON UNIVERSITY SCHOOL of Medicine

Types of Myths

- Religious or Patriarch Myth. E.g. Abraham 175 yrs, Moses 120 yrs) ? Time scale
- Village Elder Longevity Myth. Eg. ages 120-160 yrs.
- Fountain of Youth Myth. A substance, eg. Ponce de Leon explored the Florida area in 1513, the glacial milk of the Andes.
- Shangri-La Myth. A place, eg. as described by James Hilton in 1933 in his book *Lost Horizon,* the Caucasus, Vilcabamba, Ecuador, and the Hunza region of Pakistan. Usually invokes many extremely old people. There may be a tourism motive (like Dominica)

"China's Aged and Sick Flock to a Hamlet Known for Longevity" NY Times, 2017

Bama County, China is called China's longevity capital. Baimo cave is said to cure ailments.

Myths…

- Nationalist Longevity Myth
- Spiritual Practice Longevity Myth
- Myth of Familial Longevity
- Individual and Family Notoriety

Mariam Amash was surrounded by her family during all the media attention paid to her while recently claiming to be 120 years old. She would have had to be 66 when she had her son Mohamed who was 54.

Myths....

- Military Age Misreporting
- Administrative Registration Errors (especially unintentional unreported deaths, confusing siblings with each other, etc.)
- Unreported Deaths for pension or other entitlements fraud

Sogen Kato's remains were discovered at his family's home when local officials wanted to honor his age of 111 years in 2010. He had actually died in 1978 at age 79. This sparked a nationwide investigation when the vital status of 234.354 centenarians was unknown and 77,000 would be 120+ years old. One was age 186 years. Many were found to have actually died in WWII.

Red Flags of an invalid age claim

- No mention of the person when they surpassed the record of 122 years
- Maternal age doesn’t make sense
- Lack of early supportive documentation
- Conflicting evidence (e.g. an identity card versus an early census record)

Validating age in the New England Centenarian Study (and our 3 NIH-funded studies)

1. Obtain a birth certificate and when a study participant dies, obtain their death certificate.
2. When a birth certificate is not available, obtain multiple forms of preferably old proof that end up corroborating one another (e.g. military, marriage, school and census records).
3. In the case of people age 110 years old, both #1 and #2 are necessary plus a review of the family pedigree to be sure that the ages of family members make sense relative to the study participant (this is a method called "familial reconstitution").

Hazel, age 111 yrs

Recruitment Challenges

- Rarity of our 103+ year old subjects (from 1/100,000 to 1/5 million in the population) limits our sample size but this may be offset by a likely growing phenotypic homogeneity with older and older ages.
- Still a large frequency of unrecognized deaths on voter registration lists that we use for recruitment mailings (the National Death Index helps with this).
- About 30% of centenarians are in their last year of life.