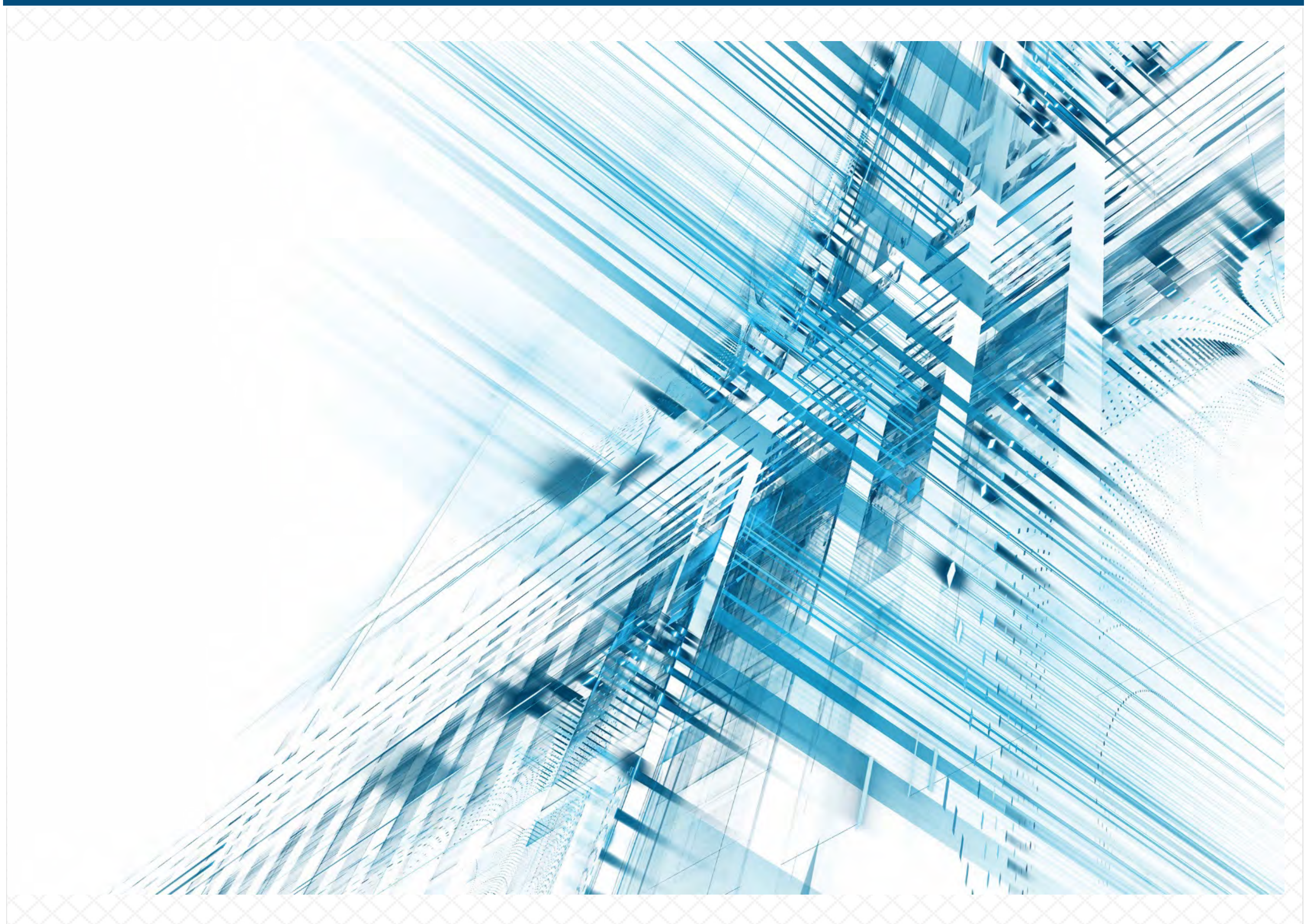
Overview of the United States Mortality DataBase (USMDB)



July 2018

Overview of the United States Mortality DataBase (USMDB)

AUTHOR Dr. Magali Barbieri
Associate Director
Human Mortality Database
University of California, Berkeley

SPONSOR Society of Actuaries

Main goal

The main goal of this project is to construct an open-source and well documented database of detailed period lifetables by sex for each state of the United States (and for the District of Columbia) for all calendar years since 1959 following an approach similar to that of the Human Mortality Database.

Motivation

Currently, life tables by state (the so-called Decennial Life Tables[1]) are available from the National Center for Health Statistics at the Centers for Disease Control for three year-periods centering on each census and updates are published with much delays (for instance, the most recent set of state life tables available are for the period 1999-2001 as the 2009-2011 set has not been published yet). Aggregate mortality estimates (the expectation of life at birth, age-standardized death rates, and the probabilities of dying between selected ages) are available by state from the Institute for Health Metrics and Evaluation[2], produced using a Bayesian method with limited age details. Our project will thus publish information not readily available elsewhere and useful for both actuarial and demographic research. By fostering new studies, it is expected to greatly improve the understanding of geographic variations in mortality in the United States.

Data

The lifetables are constructed using official demographic data, that is vital statistics prepared by the National Center for Health Statistics (NCHS) and census and population estimates data from the U.S. Census Bureau. More specifically, the original data are tabulations of births by state, sex, and calendar year from 1959 to the latest year available (currently 2015); death counts by state, sex, year of occurrence (1959-2015), single year of age, and year of birth; census population by state, sex and single year of age for 1950, 1960 and 1970; and population estimates by state, sex and age (five-year age group for 1970-1979 and single year of age for 1980-2015).

Methodological approach

The data are processed in two steps following the HMD protocol (see http://www.mortality.org/Public/Docs/MethodsProtocol.pdf for details) using computer codes written in the R language: the first step creates tabulations of death and exposure counts by state, calendar year, sex and Lexis triangle (i.e. cross-tabulated by year of occurrence, single year of age and year of birth) (a "Lexis database", or LexisDB), while the second step computes death rates and complete period lifetables by state, calendar year, sex and single year of age up to the open age interval 110 and above (a Lifetable database, or LifetableDB). The set of methods

[1] See https://www.cdc.gov/nchs/products/life_tables.htm#decennial.
[2] http://www.healthdata.org/.

implemented is fairly complex (see the more than 70 pages of explanations in the Methods Protocol referenced above), especially as regards the first step.

Verifications are carried out throughout the process: data quality of the input data is thoroughly investigated and the consistency of the resulting series is carefully checked. Checks are internal, relying on the regularities of the underlying age structure of mortality and gradual changes in the death rates and other mortality indicators over time, and external, using others' estimates (in particular the aforementioned NCHS Decennial Life Tables by state for the periods when available as well as the estimates published by the IHME). Solutions are sought for every identified problem. For instance, a new set of Census Bureau population estimates by state was discovered for years 1970-1979, with data provided by single year of age. We thus attempted to use these instead of the previous five-year age group estimates. However, the output life tables constructed from the more detailed set of estimates yielded an inconsistent pattern in the time series, with what looked like an over-estimation of death rates for the 1970s compared to the previous and later periods (1959-1969 and 1980-2015). Discussions with our contacts at the Census Bureau revealed that the single-year estimates were in fact based on a preliminary sample of the census data while the five-year estimates were based on the full census. After testing various approaches, we finally decided to apply the proportionate distribution of the newly discovered set of estimates by single-year of age to the more reliable five-year death counts.

Prior work

In 2016, birth, death and population data had been collected at the most detailed level possible, and formatted following HMD standards. The two sets of computer codes (LexisDB and LifetableDB codes) had been adjusted to these particular data. Various adjustments to the HMD Methods Protocol have been tested, in particular the age above which exposures were computed using a combination of the extinct cohort method and the survival ratio method (see the HMD Methods Protocol).

In 2017, the computer codes to produce the state life tables were completed and run on the input data. The output is similar to that in the HMD (see www.mortality.org), i.e. lifetables in various age by time formats (single years of age or five year age groups by single calendar year, 5 year- or 10-year period) by sex, for each state and Census Division and Region from 1959 to 2015. We also digitalized all of the Decennial Life tables available from the National Center for Health Statistics (NCHS) at the Centers for Disease Control for years since 1959 (i.e., 1959-61, 1969-71, 1979-81, 1989-91, and 1999-2001). The NCHS is in charge of processing and distributing all vital statistics for the United States. The NCHS life tables were compared with those of the USMDB and were found to be remarkably similar. A prototype for the website to host the publicly available USMDB mortality series was designed.

Current Work

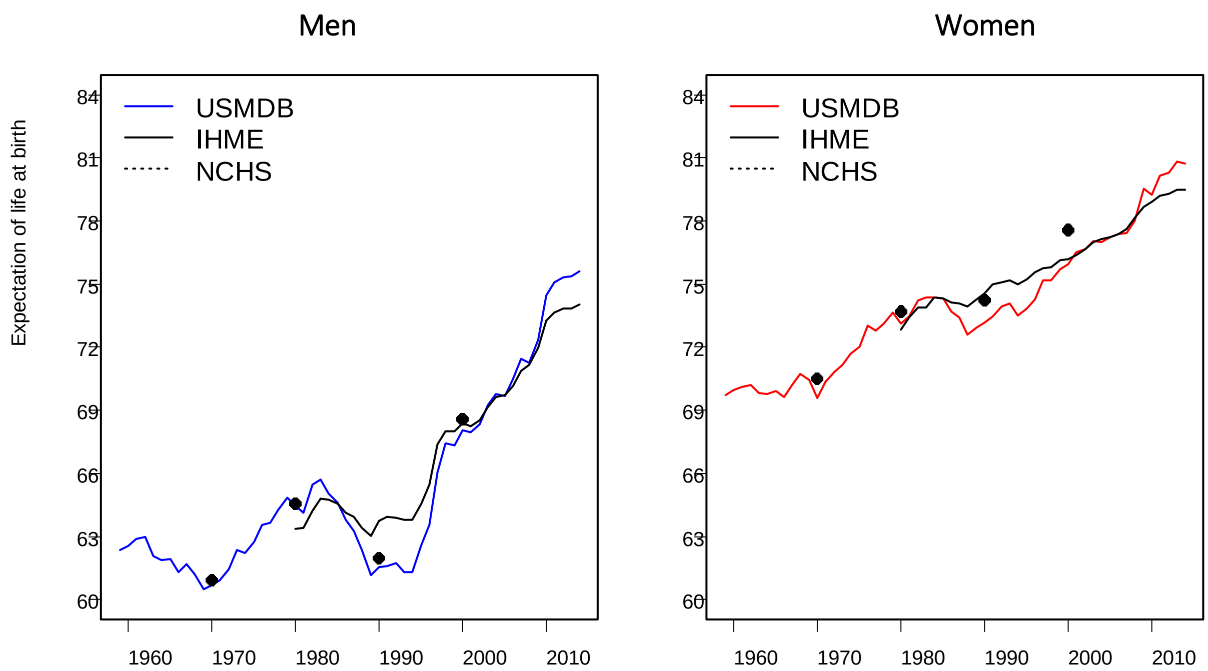
1) Validation of HMD state life tables by comparison with IHME estimates

The USMDB lifetable values by state were compared with those available on the website of the Institute for Health Metrics and Evaluation (IHME, http://ghdx.healthdata.org/us-data), the only other source (with the NCHS Decennial Lifetables) of mortality data at the state level. The IHME has only published a small number of mortality indicators at the state level, namely the life expectation at birth and the probabilities of dying below age five and between age five and 25, 25 and 45, 45 and 65, and 65 and 85 years, for all calendar years from 1908 to 2014.

Again, we found a high consistency between the USMDB and the IHME estimates, except for the smoothing implemented by the IHME, which creates an illusion of stability over time and little year-to-year variations by contrast with the USMDB values in the smaller states (e.g. the District of Columbia as illustrated in Figure 1). The IHME mortality estimates also tend to be higher than the USMDB for the last 4-5 points in the data series. This is due to the fact that the IHME uses a complex statistical model to estimate the state-level age-specific mortality rates by smoothing over the age range, over time (hence the lower estimates of the expectation of life at birth in the IHME compared to the USMDB in the recent period since the mortality trend in the last few years is different from the previous trend) and over space (using information from socio-economically similar populations)[3]. The result of the comparison between HMD and IHME for the whole range of indicators available and for all states is presented in Appendices 2-7.

[3] See more on the IHME model specification in Dwyer-Lindgren, L., Bertozzi-Villa, A., Stubbs, R.W., Morozoff, C., Shirude, S., Unützer, J., Naghavi, M., Mokdad, A.H. and Murray, C.J., 2018. Trends and patterns of geographic variation in mortality from substance use disorders and intentional injuries among US counties, 1980-2014. *Jama*, *319*(10), pp.1013-1023.

Figure 1. A comparison of life expectancy at birth in the USMDB and in the IHME for the District of Columbia, men and women, 1959-2014



2) Construction of life tables at the Census Region level

The Census Bureau divides the United States into 4 Regions (the Northeast, the Midwest, the South and the West), 9 Divisions, and the 50 states, plus the District of Columbia (see Appendix 2 for a map of these administrative units). In addition to the lifetables previously constructed for the 50 US states and D.C. and for each of the 9 Census Divisions, we have added life tables for the 4 Census Regions in a similar format.

3) Database documentation

We have now completed most of the documentation for the website, following largely the structure of the Human Mortality Database. In addition to a specific users' agreement, information about the registration procedure, acknowledgement of technical and financial support, list of all contributors, and preliminary FAQs, we have written a description of the database (Explanatory Notes) and provided detailed descriptions of the methods, background, sources, input data, and other types of useful information for the users in the "Documentation" file available online (and reproduced in Appendix 1 below).

4) Publication of the database and associated website (usa.mortality.org)

We have been working on a website to publish the final life table series. The website is being developed in collaboration and consultation with Arnaud Coche (a computer programmer at the French Institute for Demographic Studies) who has experience developing websites with a MySQL-database[4] architecture via a php-based web (user) interface. The website and database are administered on the back-end with a Xataface[5] interface, allowing non-technical administrators the ability to modify both the structure and the content of the website.

The website uses some features similar to those of the HMD but with notable differences. In particular, because we are envisioning in the longer term the production of life tables by US counties, we had to create a database interface much more powerful than for the current HMD (which includes data for fewer populations than there are states in the US).

In its beta version, the website presents US national, state and regional (Census Divisions and Regions) life tables stored in a static database for all years 1959 to 2015. In this initial version, data are limited to the single year of age and single calendar year (1x1) sex-specific life tables which have all been computed and gone through data-quality checks. Data series are available independently for each geographic unit in two different formats: as comma-separated files and as tabulation-separated files. Comma-separated files are easier to process with modern statistical software but tabulation-separated files are easier to read on screen. In addition, the single-year of age life tables for all geographic units have been bundled within a single zip file for easier download. As this is the very preliminary public version of the USMDB, we are hoping for multiple feedback to improve content and users' experience.

Future work

The main goals for future work are to:

- To further develop the USMDB website based on the feedback received from the initial USMDB users.
- To publish on the website the life tables in alternative formats (single year of age and 5- and 10-year period, as well as abridged life tables – that is by five-year age group rather than single year of age – for similar time periods).
- To update the life table series to 2016 for each state, division and region, when the necessary data become available from the NCHS (deaths) and the Census Bureau (population)

[4] https://en.wikipedia.org/wiki/MySQL
[5] http://xataface.com/wiki/about

- To produce a report on inter-state variations in mortality since 1959 in collaboration with our colleagues at NCHS (Robert Anderson and Elizabeth Arias)

Long term goals

In the longer term, we would like to add cause-of-death data at the state level. We are planning to publish first some age-standardized death rate for a small number of exhaustive and mutually-exclusive cause-of-death categories (maybe 20 or 40). Indeed, because of confidentiality issues, we are not authorized to publish "raw" cause-specific death rates based on fewer than 5 counts without transformation. Furthermore, small death counts (once decomposed by sex, age and cause of death) will produce large year-to-year random fluctuations and render the interpretation of cause-specific mortality trends at the state level difficult. To avoid the risk of disclosing confidential information and meaningless large fluctuations, we will next implement a statistical model to compute more detailed cause-specific mortality indicators. We are planning to adjust a Bayesian model we have developed with two mathematical demographers for this kind of purpose. The model relies on the regularity of mortality over the age range, over time and across space[6]. Using this Bayesian approach, we would also like to offer the same data products at the county level but because of numerous data issues (not the least of which is the difficulty to match mortality and Census population data and to follow changes in county coding in national statistics), this will require additional resources.

Eventually, USMDB users will be able to select life tables from an "A la carte" menu of geographies, that is they will be able to either download all available life tables, or only those for specific states, or only those at the Census Division or Census Region level, or those for all of the states in a particular Division/Region, etc…. For now, users will only have the option to either download all available life tables at once, or to download one set of life tables (i.e. the life tables for a single Division or state) via the data page for that particular administrative unit.

Finally, we would like to improve the usefulness of the USMDB to non-data savvy potential users, such as journalists, policy makers, state health officers, and the general public by creating simple data tables, graphs, and automated reports that could be customized to the particular needs of users through queries. Note that the USMDB website has been designed to accommodate all of these potential developments.

[6] See Alexander, M., Zagheni, E. and Barbieri, M., 2017. A Flexible Bayesian model for estimating subnational mortality. *Demography*, *54*(6), pp.2025-2041.

Additional Background and Documentation file for the USMDB

MORTALITY DATA FOR THE UNITED STATES MORTALITY DATABASE (USMDB)[7]

By Magali Barbieri and Celeste Winant

General

National censuses and population estimates for the United States are produced by the U.S. Department of Commerce at the US Census Bureau (www.census.gov). Vital statistics data are collected and disseminated by the National Center for Health Statistics (NCHS), which is part of the Centers for Disease Control (CDC) (www.cdc.gov/nchs/). Though information on burials was collected routinely in some areas of the United States in the 19th century, it is only in 1933 that the quality of the information collected was deemed good enough (with over 90% of vital events registered) in all of the States and the District of Columbia for the system to cover the whole territory of the United States (National Research Council, 2009). The present format of US death certificates was established in the 1940s and follows the recommended international standards. It has been revised periodically to reflect medical progress and changing public health concerns (Rosenberg, 1999). The last notable revision was implemented in 2003.

Sources of Data

The USMDB population data originate from population censuses conducted every 10 years. The USMDB makes use of the censuses conducted from 1960 to 2010. Census counts serve as the basis for producing annual July 1st population estimates for intercensal and postcensal periods, published by the Census Bureau. Postcensal population estimates are revised every year and intercensal population estimates are substituted to postcensal estimates as new census data become available. Intercensal population estimates are available at the state level for all years since 1970. They have been published on the Census Bureau website (U.S. Census Bureau, Population Estimates Program, https://www.census.gov/programs-surveys/popest.html). Additional (sometimes more detailed) census data are available through the Inter-University Consortium for Political and Social Science Research (ICPSR, www.icpsr.umich.edu) with restricted access for participating organizations. For years between 1959 and 1969, we constructed our own annual intercensal population estimates using census data and birth and death counts by age during the intercensal period. To construct these estimates, we followed strictly the methods of the Human Mortality Database Version 5 (Wilmoth et al., 2007).

[7] This document partly relies on the information included in the Human Mortality Database Background and Documentation File for the United States (http://www.mortality.org/hmd/USA/InputDB/USAcom.pdf), initially prepared by Ludmila Andreeva.

Starting with 1959, data on deaths are available electronically from the National Center for Health Statistics (NCHS) in the form of Mortality Multiple Cause of Death Files (see National Center for Health Statistics, 1959–). These data include individual death records coded from death certificates. Public files are available on the NCHS website, at https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm. However, geographic information (including state of residence and state of occurrence) has been suppressed from the public files for years since 2004 to protect confidentiality. Access to the restricted data is only possible through special arrangement. The USMDB has obtained access to these data through the Berkeley Research Data Center after completing a strict vetting procedure. All of the USMDB mortality data processing has taken place within the RDC (including for years prior to 2004 because more detailed information on the date of birth is available in the restricted files, which allow for a more accurate allocation of deaths to the upper or lower Lexis triangle for each combination of age and year of death).

Territorial Coverage

Since 1959, when both Alaska and Hawaii became US states, national statistics has been publishing information for all 50 states and the District of Columbia. There has been no change to state boundaries since then.

Death Count Data

Coverage and Completeness

In the early years of the 20th century, vital statistics for United States were based on data from those states admitted to the Death Registration Area, the number of which increased over time. To be included in the Death Registration Area, the vital statistics system for a state had to demonstrate coverage of at least 90% of the state population. This process was completed in 1933 for the US as a whole with the admission of Texas into the Death Registration Area. Given the legal requirements for death registration, mortality data for the United States are considered to be complete and of acceptable quality since 1933.

Mortality data for the United States are confined to events registered within the territory of the United States. Vital events to US residents occurring outside of the United States are not included but those to non-US resident occurring within the United States are. Since 1970, it is possible to identify deaths of non-residents and, consequently, to exclude them from tabulations. Therefore, for the years 1959–1969, deaths in the USMDB include both residents and nonresidents (i.e., the *de facto* population), and for the period starting in 1970, only residents are included.

Specific Details

Deriving Death Counts by Lexis triangle from Individual Death Records

Death counts for years since 1989 can be precisely tabulated by Lexis triangle (i.e., by age and birth cohort) because the original data from NCHS include the exact date of birth as well as the exact date of death.

For prior years, the original data identify deaths by single year of age, but not by birth cohort (because the date of birth is not included on the death records). It is possible to estimate deaths by Lexis triangle using the exact date of death, although this approximation (and the resulting mortality estimate) is unlikely to be as accurate as the observed counts for years in which the date of birth is available. Procedures used to derive the death counts by Lexis triangle for years prior to 1989 are described in Appendix 2.

Population Count Data

Coverage and Completeness

Data on population refer to the resident population of the United States. No adjustments have been made to the published population estimates.

Specific Details

Because the Census Bureau annual state-level population estimates by sex and age are only available for years since 1970, we have constructed our own intercensal estimates using a cohort component method described in the 2007 Human Mortality Database Methods Protocol (http://v5.mortality.org/Public/Docs/MethodsProtocol.pdf) for the years 1959-1969. Further details are provided in Appendix 3.

Birth Count Data

Coverage and Completeness

As for the mortality statistics, due to the legal requirements of birth registration, data on births are considered to be virtually complete and of a good quality since 1933.

Birth data for the United States are confined to events registered in the United States. Births to legal residents of the U.S. that occurred in other parts of the world are excluded from published vital statistics. Prior to 1970, births to non-residents (that occurred in the US) were included in the statistics, whereas for 1970 and thereafter, births to non-residents are excluded.

Specific Details

The distribution of births by sex at the state level was missing from the NCHS files for the year 1967. In the USMDB, we thus used the state-specific average of the sex ratio in the two surrounding years, i.e. 1966 and 1968, to split the births by sex within each state.

Acknowledgements

We are grateful to Elizabeth Arias and Robert Anderson at NCHS for their collaboration to this project. We also wish to acknowledge John Wilmoth, current Director of the Population Division at the United Nations, and former PI for this project, who developed the initial idea for the USMD. We also recognize the contributions of Kirill Andreev and of Ludmila Andreeva to the early stages of the USMDB. And of course, this project would not have been possible without prior experience with the Human Mortality Database (HMD), including all the work carried out by both the DataLab team at the Max Planck Institute for Demographic Research and in the Department of Demography at the University of California, Berkeley, to develop the HMD Methods Protocol, of which we have made extensive use in the USMDB.

Financial and logistical support was received from the USMDB two sponsoring institutions, the Department of Demography at the University of California, Berkeley (UCB) and the NIA-funded Center on the Economics and Demography of Aging at UCB (CEDA). The project has been initially funded by a grant (R01 AG040245) awarded to John Wilmoth by the National Institute on Aging (NIA). Additional technical and research support has been provided by the French Institute for Demographic Studies (INED). Additional financial support has been provided to the project on an annual basis since 2016 by the Society of Actuaries-REX Pool Fund as well as through occasional gifts from the following sponsors: the AXA Research Fund, SCOR, Hannover-Re and Reinsurance Group of America (RGA). None of the funders had any role in the data collection, analysis, preparation, review or approval of the final data series.

References

Wilmoth, John R., Kirill Andreev, Dmitri Jdanov, Dana A. Glei, C. Boe, M. Bubenheim, D. Philipov, V. Shkolnikov, and P. Vachon. "*Methods protocol for the Human Mortality Database*." University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock. URL: http://v5.mortality.org/Public/Docs/MethodsProtocol.pdf (2007).

National Center for Health Statistics. (1959–). Mortality Detail Files.

National Center for Health Statistics. (1968–). Multiple Causes of Death Files.

National Research Council. (2009). *Vital Statistics: Summary of a Workshop.* Michael J. Siri and Daniel L. Cork, rapporteurs. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Rosenberg, Harry M. "Cause of death as a contemporary problem." *Journal of the history of medicine and allied sciences* 54, no. 2 (1999): 133-153.

U.S. Census Bureau, U.S. Department of Commerce. (1960 and 1961, 1970, 1980). General Population Characteristics, United States: Summary and each State.

APPENDIX 1: DESCRIPTION OF USMDB INPUT DATA[8]

Deaths

Period	Type of Data	Age groups	Comments	RefCode(s)
1959-2016	Annual death counts for U.S. residents by geographic area, sex, single year of age and date of death (in months)	0, 1,…maximum age attained	Deaths to U.S. residents occurring in outlying territories (e.g., Puerto Rico, U.S. Virgin Islands) or a foreign country (including Canada) are excluded, as are deaths to non-residents for years since 1970. Deaths have been tabulated from individual records. †	1

† For details, see Appendix 2.

Population

Period	Type of Data	Age groups	Comments	RefCode(s)
1960	Official Census estimates	0, 1, … , 84, 85-89,90-94,95-99,100+	Custom computations based on population available from different sources†	10, 11, 12
1970	Official Census estimates	0, 1, … , 84, 85-89,90-94,95-99,100+		20
1971–1979	Intercensal population estimates for the resident population	0, 1-4, 5-9, …, 80-84, 85+		40

[8] The reference codes indicated in the last column of each of the tables below correspond to the sources listed in Appendix 4.

Period	Type of Data	Age groups	Comments	RefCode(s)
1980	Official Census estimates	0, 1, ... , 89, 90-94, 95-99,100+		50
1981–1989	Intercensal population estimates for the resident population‡	0, 1 ... 85+		60
1990	Official Census estimates‡	0, 1, ... , 89, 90-94, 95-99,100+		70
1991–1999	Intercensal population estimates for the resident population‡	0, 1 ... 85+		80
2000	Official Census estimates‡	0, 1, ... , 99, 100-104, 105-109, 110+		90
2001–2009	Intercensal population estimates for the resident population‡	0, 1 ... 85+		100
2010	Official Census estimates‡	0, 1, ... , 89, 90-94, 95-99,100+		110
2011-2016	Intercensal population estimates for the resident population‡	0, 1 ... 85+		120

† For details, see Appendix 3.

‡ Data are available on the Census Bureau web site (http://www.census.gov). For the specific URLs and download dates, see the reference file for the raw data.

Births

Period	Type of Data	Comments	RefCode(s)[†]
1959–1964	Annual births for the *de facto* population by sex	Counts for 1959 have been adjusted to include births that occurred in Hawaii (see section "Births Count Data"). No sex detail for year 1967 (see the Specific Details in the "Birth count data" section above for imputation method).	200
1965–1979	Annual births for the *de facto* population by sex		210
1980–2003	Annual births for the resident population by sex	Births to U.S. residents that occurred abroad are excluded as are births to non-residents.	220
2003–2006	Annual births for the resident population by sex	Births to U.S. residents that occurred abroad are excluded as are births to non-residents.	230
2007–2016	Annual births for the resident population by sex	Births to U.S. residents that occurred abroad are excluded as are births to non-residents.	240

APPENDIX 2:
Tabulation of Deaths from the Mortality Detailed Files by Lexis triangle

The information required to precisely and accurately allocate each death to either the upper or the lower Lexis triangle within each combination of single year of age and calendar year is the birth cohort. This combination of information is available for all years since 1989 in the restricted mortality files to which we have access but information on the date of birth is not available for years 1959-1988. In some instances, there are inconsistencies between the date of birth, the date of death, and the age at death (Appendix Table 1). In addition, some of the information is missing for a small number of records, though the number and proportion of problematic records (with either inconsistencies or missing information) has always been marginal (representing less than 0.5% of all records in 1989) and has been declining consistently over time (representing less than 0.05% of all records for years since 2013). For those records with inconsistencies between the age at death, the date of birth and the date of death, we decided to ignore age and only rely on the dates of birth and death, *except* for those records when the age at death was below one. This is because in such cases, the information provided is extremely detailed (down to the number of minutes lived) which makes it less likely than a coding error would have occurred.

When the birth cohort is unknown

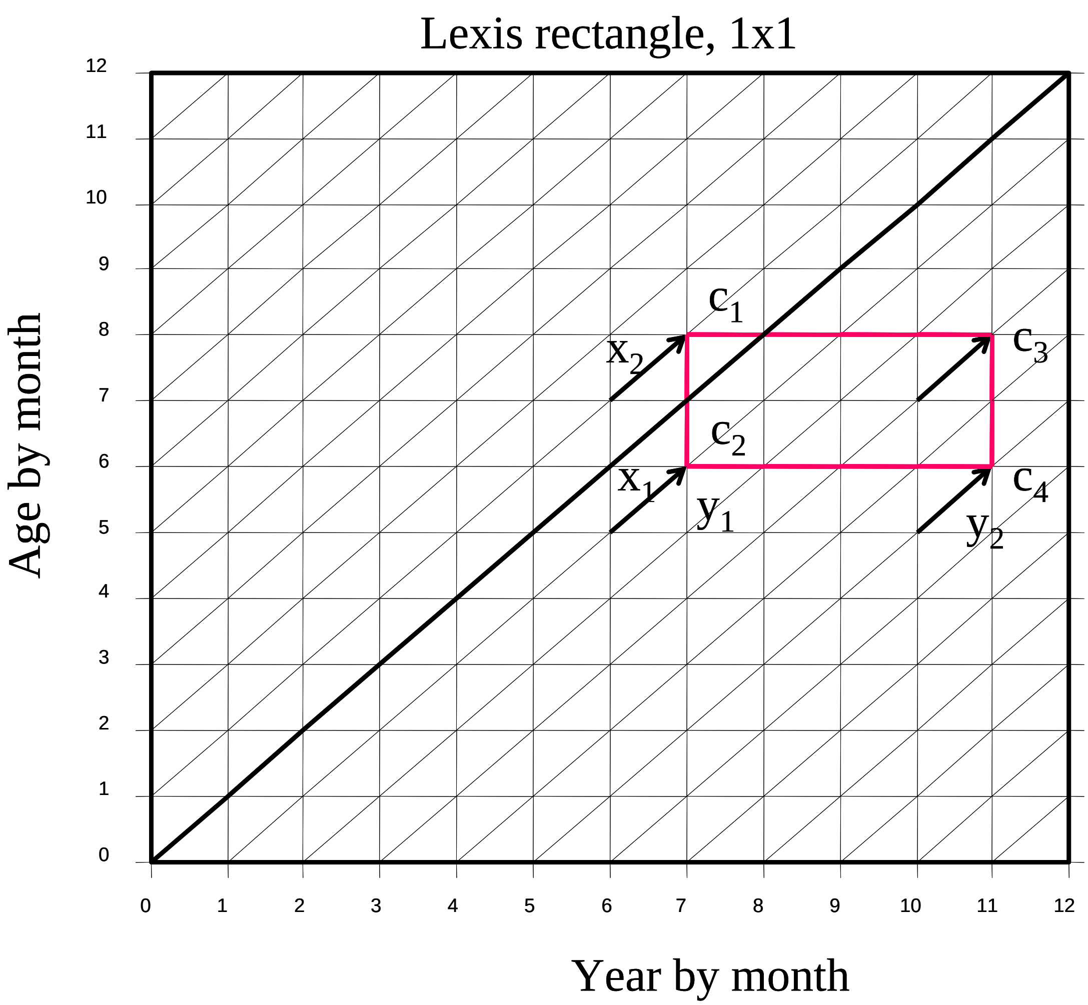
For years of data (1959-1988) when we lack information about the date of birth (and thus about the cohort to which the deceased belonged), we indirectly estimated the Lexis triangle to which deaths should be allocated. As noted earlier, the age variable identifies the age of the decedent at his/her last birthday in single years. The month of death is provided as well as, for several years (1962-1967 and since 1972), the exact date of death. This information was used in the USMDB to allocate the death counts to each Lexis triangle as further explained below.

Table 1. Number of death records with inconsistent or missing information, 1989-2016

Year	All dates/age complete and consistent	Inconsistencies between the age and dates of birth/death	Missing age only	Missing date of birth only	Missing age and date of birth	All incomplete or inconsistent records	In proportion to total records
1989	2150466	6472	45	3183	517	10217	0.48
1990	2148463	6023	45	2949	517	9534	0.44
1991	2169518	5920	29	2476	547	8972	0.41
1992	2175613	5120	7	1963	467	7557	0.35
1993	2268553	4891	25	1883	482	7281	0.32
1994	2278994	4181	29	1578	385	6173	0.27
1995	2312132	3520	36	1646	427	5629	0.24
1996	2314690	3584	25	1603	498	5710	0.25
1997	2314245	3355	17	1697	384	5453	0.24
1998	2337256	3275	2	1400	409	5086	0.22
1999	2391399	2980	8	1077	348	4413	0.18
2000	2403351	2108	2	868	354	3332	0.14
2001	2416425	2286	4	955	418	3663	0.15
2002	2443387	2616	2	939	355	3912	0.16
2003	2448288	2380	2	850	340	3572	0.15
2004	2397615	2027	3	614	343	2987	0.12
2005	2448017	2099	8	487	247	2841	0.12
2006	2426264	1779	1	524	219	2523	0.10
2007	2423712	1931	8	326	193	2458	0.10
2008	2471984	1675	0	217	147	2039	0.08
2009	2437163	1167	1	179	254	1601	0.07
2010	2648435	891	1	301	125	1318	0.05
2011	2515458	980	2	144	132	1258	0.05
2012	2543279	1208	0	170	147	1525	0.06
2013	2596993	798	0	137	132	1067	0.04
2014	2626418	661	0	164	163	988	0.04
2015	2712630	414	0	73	138	625	0.02
2016	2744248	345	0	52	137	534	0.02

Figure 1 illustrates a 1x1 Lexis square (by age and year of death), divided into two Lexis triangles (by age, birth cohort, and year), each of which is further divided into 144 smaller Lexis triangles (by age in months, birth month, and month of death). Suppose that we know a death occurred between July and November (months 7-11) and between age (x + 6 months) and (x + 8 months) (shown by the red rectangle in Figure 1). The red rectangle includes 16 of the smaller Lexis triangles (by month), of which 15 belong to the lower Lexis triangle (by year) and one to the upper triangle. If we assume that the probability of dying is the same in each of the smaller Lexis triangles, then the probability that such deaths occurred to someone from the older cohort is 1/16. Therefore, among all individual death records that fall within this red rectangle, we assign 1/16th of such deaths to the upper triangle and 15/16th to lower triangle. This simple example captures the gist of the method.

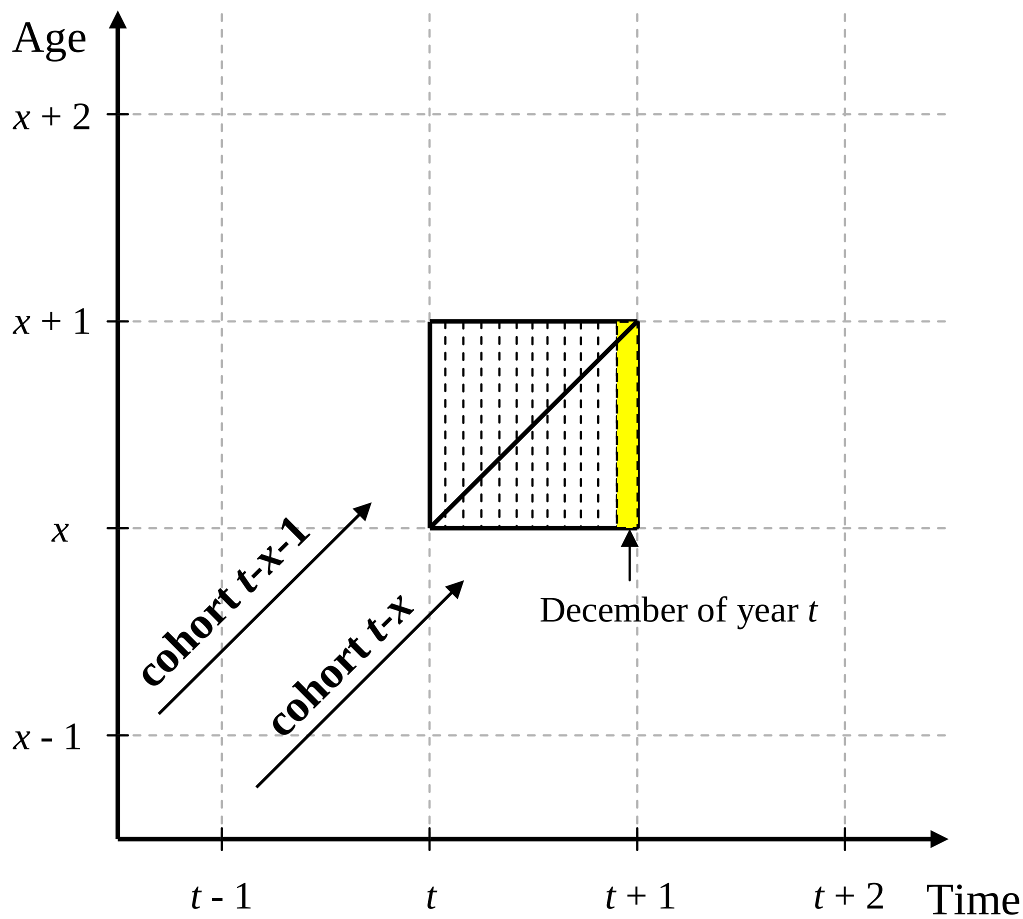
Figure 1. 1x1 Lexis Square by age (in months) and month/year of death



When only the Month of DEATH is Available

For some years (1959–1961 and 1968–1971), only the month and year of death are included in the MDF (not the day of death or the date of birth). Therefore, within a given 1x1 Lexis square (age by calendar year), we can further split the deaths into 12 rectangles representing the month of death (age by month/year of death), as shown on Figure 2. The proportion of deaths falling within the upper and lower triangles of each rectangle can be computed assuming a uniform distribution of deaths. For example, for deaths occurring in December of year *t* (shown in yellow on Figure 2), 23/24th would fall into the lower triangle and 1/24th in the upper triangle.

Figure 2. Illustration of Lexis Triangles and Rectangles



Exact Death of Death Available

For some other years (1962-1967 and 1972-1988), the files include complete information for the date of death but the date of birth is not available. Therefore, a procedure similar to that described in the previous section can be applied. That is, deaths within each 1x1 Lexis square can be split into 365 rectangles representing each possible day of death. Again assuming that deaths are distributed uniformly within each of these rectangles, the proportion of deaths falling within the upper and lower triangles can be computed.

When the Exact Dates of Birth and DEATH are Available

For years since 1989, the exact dates of birth and death (day, month and year) are available in the data files. Thus, it is possible to identify precisely whether the death occurred in the upper triangle or whether it occurred in the lower triangle, without making any assumption.

APPENDIX 3:

Population estimates for 1959-1969

One of the guiding principles of this database is to provide mortality estimates with as much age detail as possible. U.S. data on deaths by single year of age are available starting with 1959. To compute death rates by single year of age, they must be combined with population counts by single year of age, i.e., annual population estimates by state. Such data are not available from the U.S. Census Bureau for years before 1960. We thus had to calculate our own inter-censal estimates for 1959 using the classic demographic approach of a cohort component method. The 1959 estimates by back projecting the 1960 population estimates using the death counts by age within each cohort as well as the birth count for 1959.

In addition, for years 1960-1969, we identified two sets of annual population estimates available from the Census Bureau. A first set of estimates, produced by the Census bureau from a preliminary sample of census data, was available by single year of age but yielded figures which were inconsistent with both the 1970 Census counts and the Census count and estimates for 1980 and beyond. A second set was available by five-year age group, constructed from the full set of census data and highly consistent with prior and succeeding years. We thus decided to use this second set of estimates but redistributed the five-age group deaths to each single year of age using the proportional distribution from the first set of estimates.

APPENDIX 4:

Detailed sources of the data used for the USMDB and corresponding RefCodes (see Appendix 1)

RefCode 1
National Center for Health Statistics, United States. Mortality Multiple Cause Restricted Use File, Accessed in the Berkeley RDC 10-Feb-2017
Tabulation of deaths by Lexis triangle from individual records by state, single year of age and birth cohort (where available).

RefCode 10
Hybrid series of population data for USA states years 1960 by 1 year age, computed from combined series RefCode 11 and RefCode 12. The primary sources used to construct the hybrid series are:
(Census) Population Data for USA states years 1960, by 5 year age (RefCode 11)
(Census) Population Data for USA states, years 1960 (RefCode 12)

RefCode 11
Census of Population: 1960 (Volume I)
U.S. Department of Commerce Bureau of the Census (1961)
Characteristics of the Population
General Population Characteristics
Table 94. Single year of Age by color, nativity, and sex, for the state: 1960
(retrieved from https://www.census.gov/prod/www/decennial.html on 13-Apr-2015)

RefCode 12
Census of Population: 1960
Department of Commerce
Characteristics of the Population
General Population Characteristics, USA Summery
Table 59. Age by color, sex, for the state: 1960
(retrieved from
https://www.census.gov/prod/www/decennial.html
on 5-May-2015)

RefCode 20
Census of Population on
U.S. Department of Commerce Bureau of the Census (1972)
Characteristics of the Population
General Population Characteristics
Table 19. Single year of Age by Race and Sex: 1970
(retrieved from https://www.census.gov/prod/www/decennial.html on 1-Dec-2014)

RefCode 40
United States Department of Commerce. Bureau of the Census.
Intercensal Estimates of the Population of Counties by Age, Sex, and Race [United States]: 1970-1980. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 1992-02-16. https://doi.org/10.3886/ICPSR08384.v1
(Retrieved from https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/8384 on 23-Oct-2015)

RefCode 50
Census of Population on
U.S. Department of Commerce Bureau of the Census
Characteristics of the Population
General Population Characteristics
Table 18. Single year of Age by Race, Spanish Origin, Sex: 1980
(retrieved from
https://www.census.gov/prod/www/decennial.html
on 20-May-2015)

RefCode 60
U.S. Department of Commerce Bureau of the Census
State Intercensal Datasets: 1980-1990
State Population Estimates and Demographic Components of Change: 1981 to 1989, by Age, Sex, Race, and Hispanic Origin
(retrieved from
https://www.census.gov/data/datasets/time-series/demo/popest/1980s-state.html)

RefCode 70
U.S. Department of Commerce Bureau of the Census
Report Number: CP-1 (Volumes 1990 CP-1-2 to 1990 CP-1-52)
1990 Census of Population: Characteristics of the Population
General Population Characteristics
Table 18. Single year of Age by Race, Spanish Origin, Sex: 1990
(retrieved from
https://www.census.gov/prod/www/decennial.html
on 20-May-2015)

RefCode 80
Center for Disease Control and Prevention
National Center for Health Statistics
National Vital Statistics System
Bridged-race intercensal population estimates for July 1, 1990-July 1, 1999
(retrieved from https://www.cdc.gov/nchs/nvss/bridged_race/data_documentation.htm)

RefCode 90
U.S. Department of Commerce Bureau of the Census
Program: Decennial Census – Census United States
Data Set: Census 2000 Summary File 1 (SF 1) 100-Percent Data
Table: PCT012 Sex by age [Total population]
(retrieved from https://factfinder.census.gov)

RefCode 100
Center for Disease Control and Prevention
National Center for Health Statistics
National Vital Statistics System
July 1, 2000-July 1, 2009 Revised bridged-race intercensal population estimates
(retrieved from https://www.cdc.gov/nchs/nvss/bridged_race/data_documentation.htm)

RefCode 110
U.S. Department of Commerce Bureau of the Census
Program: Decennial Census – Census United States
Data Set: Census 2010 Summary File 1
Table: PCT012 Sex by age [Total population]
(retrieved from https://factfinder.census.gov)

RefCode 120
Center for Disease Control and Prevention
National Center for Health Statistics
National Vital Statistics System
Vintage 2016 bridged-race postcensal population estimates
(retrieved from
https://www.cdc.gov/nchs/nvss/bridged_race/data_documentation.htm#vintage2016)

RefCode 200
US Department of Health, Education, and Welfare, Public Health Service, National Center for Health Statistics, United States. Vital Statistics of the United States, Volume 1-Natality. Live Births by Specified Race and Sex: United States and Each State.
(retrieved from https://www.cdc.gov/nchs/products/vsus/vsus_1939_1964.htm on 14 August 2013)

RefCode 210
US Department of Health, Education, and Welfare, Public Health Service, National Center for Health Statistics, United States. Vital Statistics of the United States, Volume 1-Natality. Live Births by Specified Race and Sex: United States and Each State.
(retrieved from https://www.cdc.gov/nchs/products/vsus/vsus_1965_1979.htm on 14 August 2013)

RefCode 220
US Department of Health, Education, and Welfare, Public Health Service, National Center for Health Statistics, United States. Birth Data Files for 1980-2003.
(retrieved from https://www.cdc.gov/nchs/data_access/vitalstatsonline.htm on 22 April 2015)

RefCode 230
United States Department of Health and Human Services (US DHHS),
Centers for Disease Control and Prevention (CDC),
National Center for Health Statistics (NCHS),
Division of Vital Statistics (DVS),
Natality public-use data on CDC WONDER Online Database,
for years 2003-2006 available March 2009
(retrieved from https://wonder.cdc.gov/natality-v2006.html on 30-July-2015)

RefCode 240
United States Department of Health and Human Services (US DHHS),
Centers for Disease Control and Prevention (CDC),
National Center for Health Statistics (NCHS),
Division of Vital Statistics (DVS),
Natality public-use data on CDC WONDER Online Database,
for years 2007-2016 available February 2018
(retrieved from https://wonder.cdc.gov/natality-current.html on 14-September-2017)

About The Society of Actuaries

The Society of Actuaries (SOA), formed in 1949, is one of the largest actuarial professional organizations in the world dedicated to serving more than 30,000 actuarial members and the public in the United States, Canada and worldwide. In line with the SOA Vision Statement, actuaries act as business leaders who develop and use mathematical models to measure and manage risk in support of financial security for individuals, organizations and the public.

The SOA supports actuaries and advances knowledge through research and education. As part of its work, the SOA seeks to inform public policy development and public understanding through research. The SOA aspires to be a trusted source of objective, data-driven research and analysis with an actuarial perspective for its members, industry, policymakers and the public. This distinct perspective comes from the SOA as an association of actuaries, who have a rigorous formal education and direct experience as practitioners as they perform applied research. The SOA also welcomes the opportunity to partner with other organizations in our work where appropriate.

The SOA has a history of working with public policymakers and regulators in developing historical experience studies and projection techniques as well as individual reports on health care, retirement and other topics. The SOA's research is intended to aid the work of policymakers and regulators and follow certain core principles:

Objectivity: The SOA's research informs and provides analysis that can be relied upon by other individuals or organizations involved in public policy discussions. The SOA does not take advocacy positions or lobby specific policy proposals.

Quality: The SOA aspires to the highest ethical and quality standards in all of its research and analysis. Our research process is overseen by experienced actuaries and nonactuaries from a range of industry sectors and organizations. A rigorous peer-review process ensures the quality and integrity of our work.

Relevance: The SOA provides timely research on public policy issues. Our research advances actuarial knowledge while providing critical insights on key policy issues, and thereby provides value to stakeholders and decision makers.

Quantification: The SOA leverages the diverse skill sets of actuaries to provide research and findings that are driven by the best available data and methods. Actuaries use detailed modeling to analyze financial risk and provide distinct insight and quantification. Further, actuarial standards require transparency and the disclosure of the assumptions and analytic approach underlying the work.

Society of Actuaries
475 N. Martingale Road, Suite 600
Schaumburg, Illinois 60173
www.SOA.org