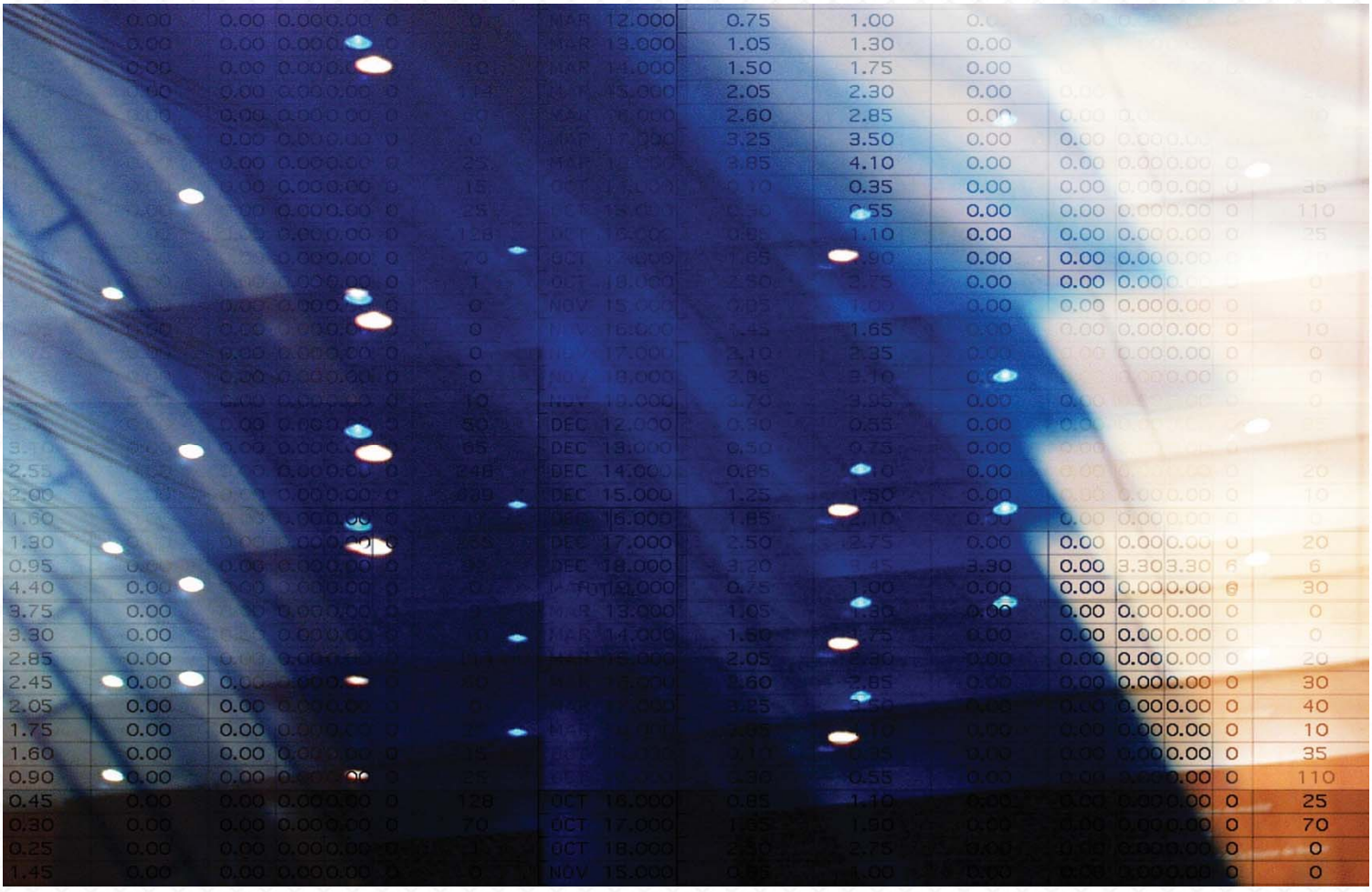
SOCIETY OF ACTUARIES

Expanding the Human Mortality Database to include Cause-of-Death Information



February 2017

Expanding the Human Mortality Database to include Cause-of-Death Information

SPONSOR Society of Actuaries
Research Expanding Boundaries Pool

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

Caveat and Disclaimer

The opinions expressed and conclusions reached by the authors are their own and do not represent any official position or opinion of the Society of Actuaries or its members. The Society of Actuaries makes no representation or warranty to the accuracy of the information.

Copyright ©2017 All rights reserved by the Society of Actuaries

TABLE OF CONTENTS

© 2017 Society of Actuaries

Expanding the Human Mortality Database to include cause-of-death information

Section 1: Background

The Human Mortality Database is a unique open-access collection of detailed mortality and population data for 38 countries with complete and reliable vital registration and census data. The HMD currently covers the United States, almost all of Europe, including countries of the former U.S.S.R., Japan, Australia and other mostly high-income countries with rapidly aging populations. First published in 2002, the project is the result of an initial collaboration between two teams of researchers, at the University of California, Berkeley in the United States, and at the Max Planck Institute for Demographic Research (MPIDR) in Rostock, Germany, which have been more recently joined by the French Institute for Demographic Studies (INED) in Paris.

The Human Mortality Database (HMD) contains calculations of death rates and life tables for national populations (countries or areas), as well as the original input data used to construct these tables and an extensive documentation. The input data consist of death counts and birth counts from vital statistics, plus census counts and population estimates from various official sources. A rigorous methods protocol (Wilmoth *et al.* 2007) is systematically implemented to process the input data and construct all HMD mortality series in a standardized manner, accounting for all kinds of data idiosyncrasies and demographic disruptions. The data series for each country are updated periodically. The constructed and updated series are carefully checked before publication.

The HMD data are used by scientists from many different fields to study mortality trends, ageing and population change or to develop simulations of various risks on the population structure; by international organizations, such as the United Nations or the World Health Organization, to strengthen their own databases; and by private-sector organizations (insurance companies and actuaries in particular) who need reliable and harmonized life table estimates.

Over 40,000 users have accessed the HMD at various times since 2002. As of December 2015, a non-exhaustive list of known publications that cite the HMD (accessible at http://www.mortality.org/Public/HMD-Publist.pdf) includes about 3,000 journal articles, 650 books or book chapters, and more than 150 dissertations and theses. The HMD has been used to document a wide variety of topics and fostered studies directed towards better understanding mortality change. As a growing course of study in mortality improvement focuses on specific trends by causes of death, it has naturally led to repeated requests from the HMD user community to make available such type of information. Combined with the prior and long-term support from the U.S. National Institutes of Health, the more recent support provided by the Society of Actuaries has allowed the HMD project to start offering access to cause-of-death information for countries already included in the database.

In almost all countries (though notably not in the former Soviet Republics until the 1990s at the earliest), causes of death have been categorized according to the World Health Organization International Classification of Diseases (ICD), which has been revised ten times since the late 19th century. Such data

© 2017 Society of Actuaries

contain rich information about historical changes in patterns of disease and other fatal risks. Death rates by sex, age and for a selected number of causes are available for various periods from national statistics offices. However, the data are not comparable across countries and over time because the methods implemented are typically not well documented. Furthermore,these methods, as well as the cause-of-death category shortlists used to disseminate the data, vary from country to country and, sometimes, from year to year in the same country. Death counts by cause are also available for a large number of countries in the World Health Organization (WHO) and in the Eurostat databases. The usefulness of these databases is however limited because of the very short time series available (at most 15-16 years in the Eurostat database) or because the cause-of-death categories used to classify the deaths vary widely from one period to the next and it is nearly impossible to reconstruct a consistent series of cause-specific death counts (in the WHO database). The proposed work was thus developed to produce a unique collection of historical standardized cause-specific mortality series which, when combined with the existing HMD mortality series, would further increase the usefulness of the database and enable a deeper understanding of mortality change. The new cause-of-death series were to be developed for the following eight countries: Canada, the Czech Republic, England and Wales, France, Japan, Norway, Sweden, and the United States for periods going back as close as possible to 1950.

Section 2: Scope of work

In 2016, the researcher and her team focused on the development and completion of the HMD data series down to a cause of death level for an initial set of eight populations. The aim was to create data series on causes of death from 1950 or the earliest date available up to the present, using a list of 92 categories and 20 broader groupings chosen to minimize the disruptions associated with revisions of the international classification system (ICD-6 up through ICD-10), without adjusting the underlying data. Though cause-of-death data are available in some countries for periods before 1950, it is impossible to construct consistent cause-of-death series going back further as the ICD Revision (ICD-5) in use before 1950 is structurally different from the one published in 1950 (ICD-6). In 2017, this work will be completed by the publication of the eight country series on dedicated web pages within the Human Mortality Database website. In addition, it is expected that similar cause-of-death data will be added for another 2-3 countries. Initial contacts with national statistics offices suggest that the most promising candidates at this point are Italy, Switzerland, Belgium, Finland and, possibly, Scotland and Ireland.

Section 3: Methodological approach

The construction of cause-of-death series for the Human Mortality Database involves five basic steps which are further described below in detail: 1) collection of input data (death counts by year, sex, age and cause of death); 2) input data formatting; 3) input data quality assessment; 4) construction of intermediary files (cause-of-death input data and HMD series); 5) production of the final cause-of-death data series.

3.1. COLLECTION OF INPUT DATA

© 2017 Society of Actuaries

Cause-of-death data have been collected over the years by the HMD project with help from in-country experts (statisticians working for national statistics offices and academic scholars). The cause of death of interest for this project is the underlying cause as defined by the World Health Organization. The underlying cause of death is "the disease or injury that initiated the train of events leading directly to death, or the circumstances of the accident or violence which produced the fatal injury". It is selected following a number of standardized rules developed by WHO among all the causes listed by the certifying physician on the death certificate. Only for the most recent years are data typically available in electronic format. Electronic data are not always available publicly and special tabulations have to be requested for a number of countries. For earlier years (especially the 1950s and 1960s), data have to be located and scanned from publications or from country archives, and they have to be digitalized in a careful manner to avoid errors (with systematic checks across sexes, ages and causes).

The two main barriers to access available data have been the issues of confidentiality and cost. Regarding confidentiality, many countries have been increasingly reluctant to give access to the most detailed data, especially those with small populations where the indirect identification of single individuals is a possibility. Two accommodations had to be made to protect privacy rights. First, it was decided that only 3-digit codes would be used in the HMD COD categories up to ICD-9, which also facilitated access to data only available from paper publications (which rarely include the full 4-digit detail). However, because of the complicated mapping of ICD-9 codes into ICD-10 codes for some of the external cause categories (most notably traffic accidents), the 4-digit detail is still necessary for ICD-10 in order to guarantee a minimum degree of consistency for the 92 HMD COD categories over time. Second, the research team had to agree not to provide access to the COD input data to HMD users. Specifically, the original cause-of-death data with all 3-digit (for ICD-6 through ICD-9) and 4-digit (for ICD-10) codes are not available to users on the public HMD website and the COD fractions, death counts and death rates are provided only for the 92 COD aggregates defined for the HMD.

3.2. INPUT DATA FORMATTING

Input data are provided in a variety of formats, even when electronic, and preparing all death counts by sex, age, and cause in a uniform format is another time consuming task. A document with instructions has been developed to ensure that the HMD staff follows a standard process guaranteeing homogeneity in the way the input data is prepared for processing (Appendix C). The input data file is a text file with six columns corresponding to the year of occurrence/registration, the sex and the age of the deceased, the ICD-code and the ICD-revision implemented that year, and the total death count for the corresponding year, sex, age and ICD-code.

3.3. DATA QUALITY ASSESSMENT

Beyond checking for internal consistency of aggregate counts over each sex and age group at the data entry stage, the quality of the COD statistics is verified in three different ways.

First, death counts aggregated by sex and age over all causes are compared with those already in the HMD (which often come from a different source). Discrepancies are expected to occur because of known delays in registering deaths from certain causes which trigger legal procedures (homicides for instance) but the differences are discussed with the National Statistics Offices to make sure that they do not result from errors.

© 2017 Society of Actuaries

Second, some countries have been adjusting slightly the WHO ICD scheme to fit their own needsThese typically consist in disaggregating existing categories. For instance, the United States uses an adapted classification for years covered by ICD-8, with many additional sub-categories. This country also added two new COD categories (U011 – Terrorism, Suicide – and U030 – Terrorism, Homicide) to account for the deaths of the 2001 terrorist attacks. Since the added or modified codes are rarely documented, the National Statistics Offices have been contacted to make sure, again, that these were not errors (which did happen in a few cases) and to clarify which diseases or conditions correspond to the added or modified codes. This information is included in each of the country-specific COD documentation.

Third, any severe disruptions in cause-specific death rate series not associated with a change in the standard ICD is investigated. In many cases, these result from non-ICD related changes in coding, but sometimes, errors in the coding or in the publication of COD data have been caught and corrected. Again, this is the type of information that will find its way into the country-specific documentation.

3.4. TABULATION OF DEATHS BY HMD CAUSE-OF-DEATH CATEGORIES

The following step is relatively straightforward and easy to automate. It is to convert the ICD codes into HMD COD codes. The list of the 92 cause-of-death categories with corresponding ICD codes for each ICD revision is provided in Appendix A. Death counts are aggregated by calendar year, sex, and age over the ICD codes making up each HMD COD category.

3.5. PRODUCTION OF FINAL CAUSE-OF-DEATH SERIES

COD fractions are next computed by simply dividing the number of deaths (in each calendar year, sex, and age group) for each of the 92 COD category by the total number of deaths for all causes combined (for the corresponding calendar year, sex, and age group). By design, the sum of the COD fractions over the 92 COD categories (for each calendar year, sex, and age group) adds up to 1.

All-cause death counts and death rates are then "grabbed" from the HMD website for each country (calendar year, sex, and age) and multiplied by the COD fractions for the corresponding year, sex, and age group to produce the cause-specific death counts and death rates to be published in the final series (as in Appendix B). The all-cause HMD files used for these calculations are Deaths_5x1.txt and Mx_5x1.txt when the cause-of-death data is only available by five-year age groups and Deaths_1x1.txt and Mx_1x1.txt when they are available by single-year of age.

Table 1 below lists the 8 countries for which cause-of-death series have been constructed so far and indicates which periods are covered in each country.

Table 1.
HMD CAUSE-OF-DEATH SERIES AVAILABLE BY COUNTRY AND TIME PERIOD

Country	HMD series
Canada	1950-2009
Czech Republic	1968-2013
England and Wales	1950-2014
France	1958-2012
Japan	1950-2013
Norway	1951-2012

© 2017 Society of Actuaries

Sweden	1952-2012
United States	1959-2013

The HMD cause-of-death series provides three types of indicators: cause-specific fractions, cause-specific death counts and cause-specific death rates. Each indicator is broken down by country, calendar year, sex, age group, and each of the 92 cause-of-death categories.

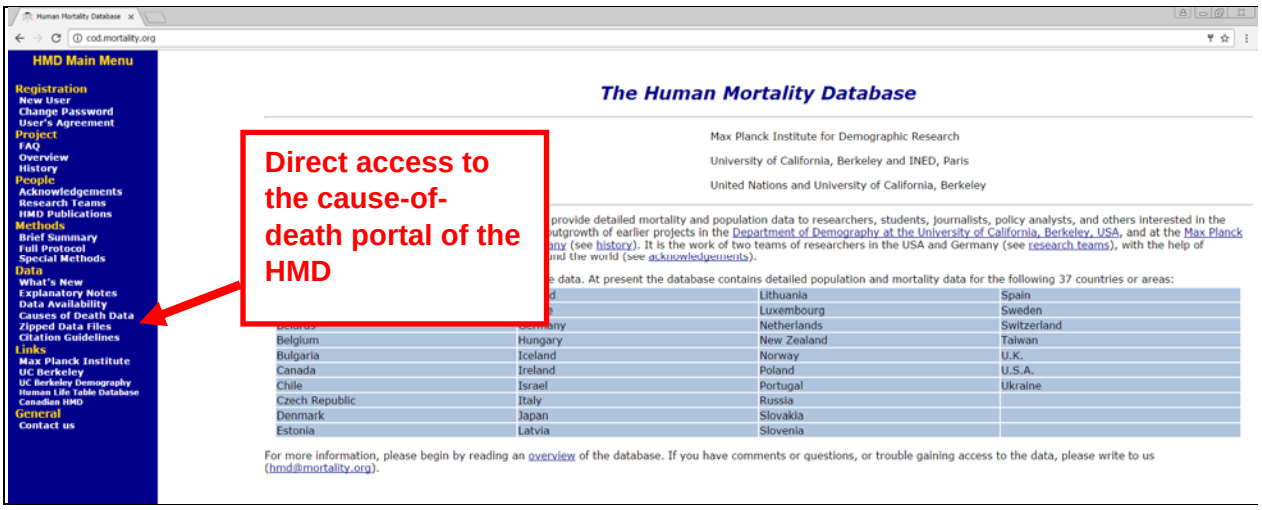
Section 4: Accessing cause of death information in the HMD

A prototype for the cause-of-death web pages to be added to the regular HMD website is ready and available for review by the SOA research staff and POG members. The cause-of-death series will be made available on the HMD public website as soon as they have been finalized. A link to the HMD cause-of-death portal will be included on the SOA web pages for this project.

The home page of the prototype is a copy of the regular HMD website homepage (Figure 1). The only difference is that an additional item has been included in the left-hand side menu to provide direct access to the cause-of-death portal of the HMD.

Figure 1.

PROTOTYPE HOME PAGE

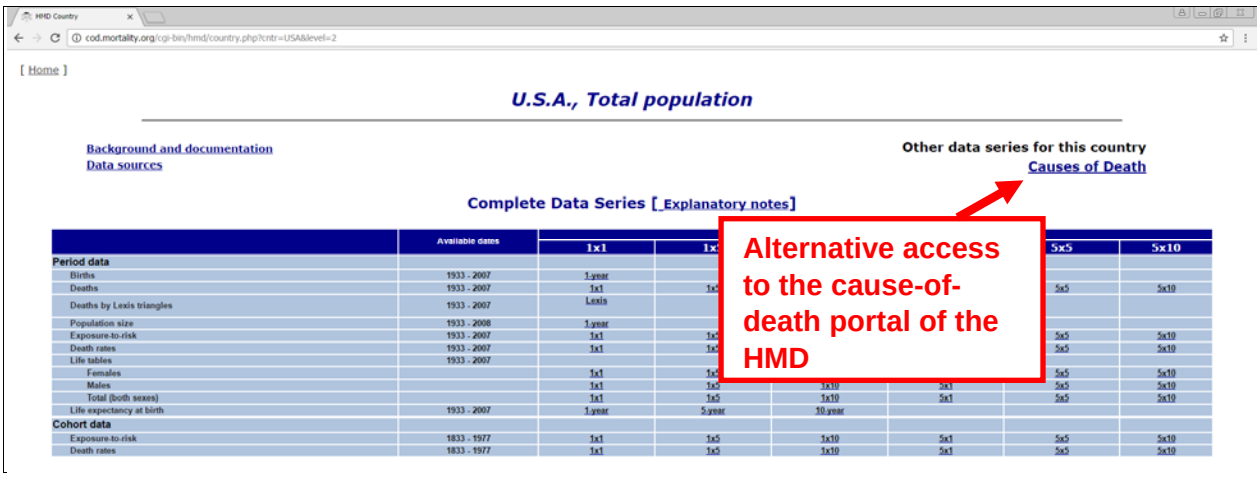


Another way to access the cause-of-death portal is to click on a country name, for instance U.S.A., where a link to the cause-of-death data has been added to the list of other available series for this country (Figure 2).

Figure 2.

© 2017 Society of Actuaries

ACCESS TO CAUSE-OF-DEATH DATA THROUGH A REGULAR HMD COUNTRY PAGE

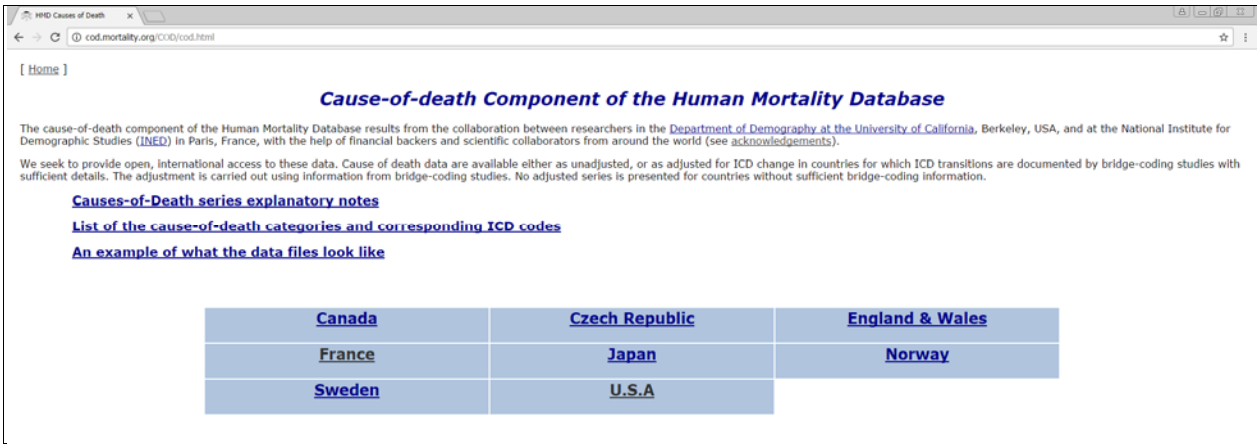


The HMD cause-of-death portal (Figure 3) gives access to:

- the cause-of-death series for each of the eight countries included in the first phase of this project
- a description of the series and some information regarding the process followed to construct the series (see the link to "Cause-of-Death series explanatory notes")
- the shortlist of the 92 cause-of-death categories and 20 cause-of-death chapters built to summarize the data ("List of the cause-of-death categories and corresponding ICD codes")
- an example of what the cause-of-death series look like since the files are too large for users to be able to see them on screen (by contrast with the all-cause HMD series).

Figure 3.

HMD CAUSE-OF-DEATH PORTAL



By clicking on one of the country names, the user is taken to a standard country page for the cause-of-death series (see an example for the U.S.A. on Figure 4). In the example provided here, users can access both the original input data (see table at the bottom of the country page) and the complete data series (first table on the country page). However, the United States is one of the few countries where there is no restriction on HMD users' access to the original data. Because of the aforementioned concerns with

© 2017 Society of Actuaries

privacy rights, the statistical offices in most countries have requested that the HMD does not provide access to the original (input) data but only to the reconstructed cause-of-death series for a limited number of cause-of-death categories. In most countries, the Input Data table is thus empty.

Figure 4.

EXAMPLE OF A COUNTRY PAGE FOR THE HMD CAUSE-OF-DEATH SERIES

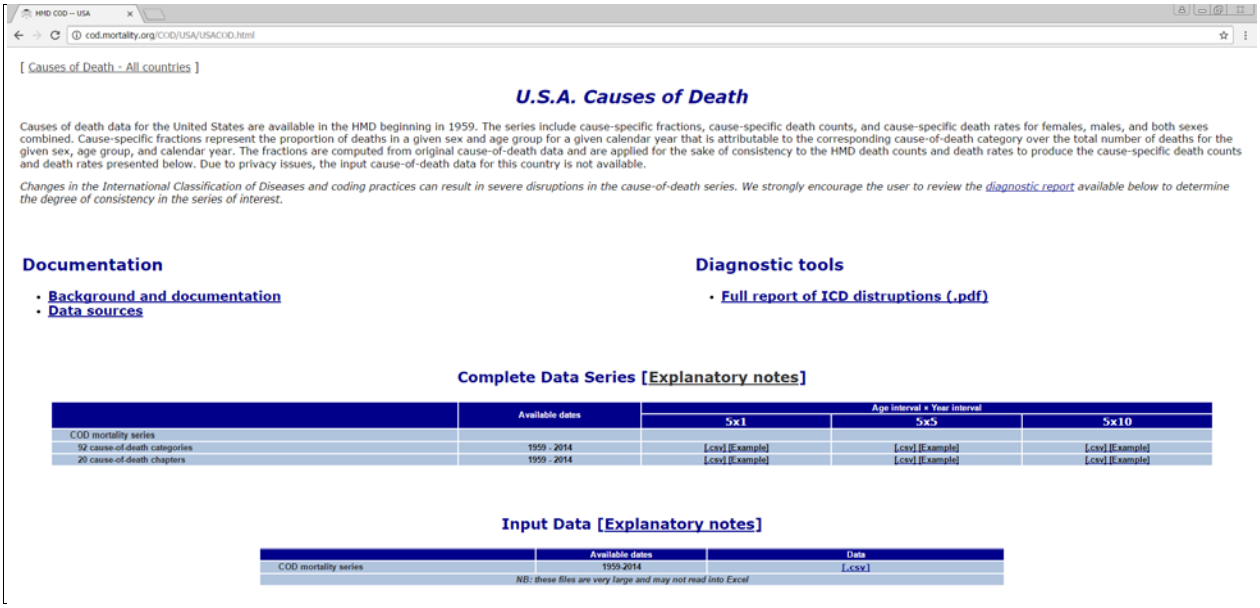


Figure 5 represents the first 42 lines of a cause-of-death file, here for the United States. The first column indicates the year to which the COD fractions, death counts and death rates pertain; the second column, the age or age group; the third column the HMD COD category (one of 92) or the total over all causes ("All"); the next three sets of three columns show the COD fractions, the death counts and the death rates (for females, males, and both sexes combined); and the last column indicates which Revision of the International Classification of Diseases was in used in this particular country and year (in this example, the 7th Revision in 1959). The last six columns (death counts and death rates for each specific age group, sex, and cause-of-death category) are derived from the first three, i.e. cause-of-death fractions are first computed from national cause-of-death data files and applied to the HMD all-cause (existing) death count and death rate series for the sake of consistency.

© 2017 Society of Actuaries

Figure 5.

EXTRACT (FIRST 42 LINES) OF THE UNITED STATES HMD CAUSE-OF-DEATH SERIES

The United States of America, Deaths by HMD Cause-of-Death (5x5) 91 HMD causes Last modified: Wed Sep 21 13:24:19 2016

Year	Age	COD.cat	COD.Frac.F	COD.Frac.M	COD.Frac.T	Deaths.F	Deaths.M	Deaths.T	Rates.F	Rates.M	Rates.T
1959	0	01	0.000355	0.000512	0.000445	17	33.04	50.04	0.8429	1.5829	1.2192
1959	0	02	0.000188	0.000186	0.000187	9	12.02	21.02	0.4462	0.5759	0.5121
1959	0	03	0.028146	0.025601	0.026685	1348.27	1652.96	3001.22	66.8339	79.1966	73.1204
1959	0	04	0.001754	0.001363	0.00153	84.02	88.02	172.03	4.1647	4.2171	4.1914
1959	0	05	0.001927	0.001755	0.001828	92.31	113.32	205.63	4.5759	5.4292	5.0098
1959	0	06	0.003905	0.003691	0.003782	187.04	238.29	425.33	9.2715	11.4169	10.3624
1959	0	07	0.000167	0.000263	0.000222	8.02	17	25.02	0.3975	0.8147	0.6096
1959	0	08	0	0	0	0	0	0	0	0	0
1959	0	09	0.007496	0.005685	0.006456	359.07	367.07	726.14	17.7992	17.5872	17.6914
1959	0	10	0.000021	0.000031	0.000027	1	2	3	0.0496	0.0959	0.0731
1959	0	11	0	0	0	0	0	0	0	0	0
1959	0	12	0.000042	0	0.000018	2	0	2	0.0992	0	0.0487
1959	0	13	0	0.000031	0.000018	0	2	2	0	0.0959	0.0487
1959	0	14	0.000021	0	0.000009	1	0	1	0.0496	0	0.0244
1959	0	15	0.000167	0.000047	0.000098	8	3	11.01	0.3967	0.1439	0.2681
1959	0	16	0	0.000046	0.000027	0	3	3	0	0.1438	0.0731
1959	0	17	0	0	0	0	0	0	0	0	0
1959	0	18	0.000021	0.000046	0.000036	1	3	4	0.0496	0.1438	0.0975
1959	0	19	0.000021	0	0.000009	1	0	1	0.0496	0	0.0244
1959	0	20	0.000021	0	0.000009	1	0	1	0.0496	0	0.0244
1959	0	21	0	0	0	0	0	0	0	0	0
1959	0	22	0	0	0	0	0	0	0	0	0
1959	0	23	0.000021	0	0.000009	1	0	1	0.0496	0	0.0244
1959	0	24	0	0.000015	0.000009	0	1	1	0	0.0479	0.0244
1959	0	25	0.000125	0.000155	0.000142	6	10	16	0.2975	0.4792	0.3899
1959	0	26	0.000021	0.000015	0.000018	1	1	2	0.0496	0.0479	0.0487
1959	0	27	0.000021	0	0.000009	1	0	1	0.0496	0	0.0244
1959	0	28	0.000167	0.000201	0.000187	8	13.01	21.01	0.3966	0.6233	0.5119
1959	0	29	0	0	0	0	0	0	0	0	0
1959	0	30	0.001044	0.001054	0.00105	50.03	68.03	118.06	2.48	3.2596	2.8764
1959	0	31	0.000982	0.000697	0.000818	47.02	45.02	92.04	2.3308	2.157	2.2424
1959	0	32	0.001337	0.000991	0.001139	64.04	64.01	128.05	3.1743	3.067	3.1197
1959	0	33	0.001964	0.001983	0.001975	94.06	128.03	222.08	4.6624	6.1339	5.4107
1959	0	34	0.000188	0.000077	0.000125	9	5	14.01	0.4463	0.2397	0.3412
1959	0	35	0	0	0	0	0	0	0	0	0
1959	0	36	0.002717	0.003269	0.003033	130.13	211.04	341.17	6.4507	10.1114	8.3122
1959	0	37	0	0	0	0	0	0	0	0	0
1959	0	38	0	0	0	0	0	0	0	0	0
1959	0	39	0.001837	0.001691	0.001753	88.02	109.16	197.18	4.363	5.2302	4.804
1959	0	40	0.00717	0.007684	0.007465	343.45	496.1	839.55	17.0248	23.7691	20.4543
1959	0	41	0	0	0	0	0	0	0	0	0

As for the regular HMD mortality series, the country pages for the cause-of-death data provide access to country-specific documentation as well as to the statistical series. In addition, in order to increase users' awareness of cause-of-death classification issues, diagnostic tools to evaluate the consistency of cause-of-death series are provided within a country-specific report that is automatically generated each time the cause-of-death series are revised.

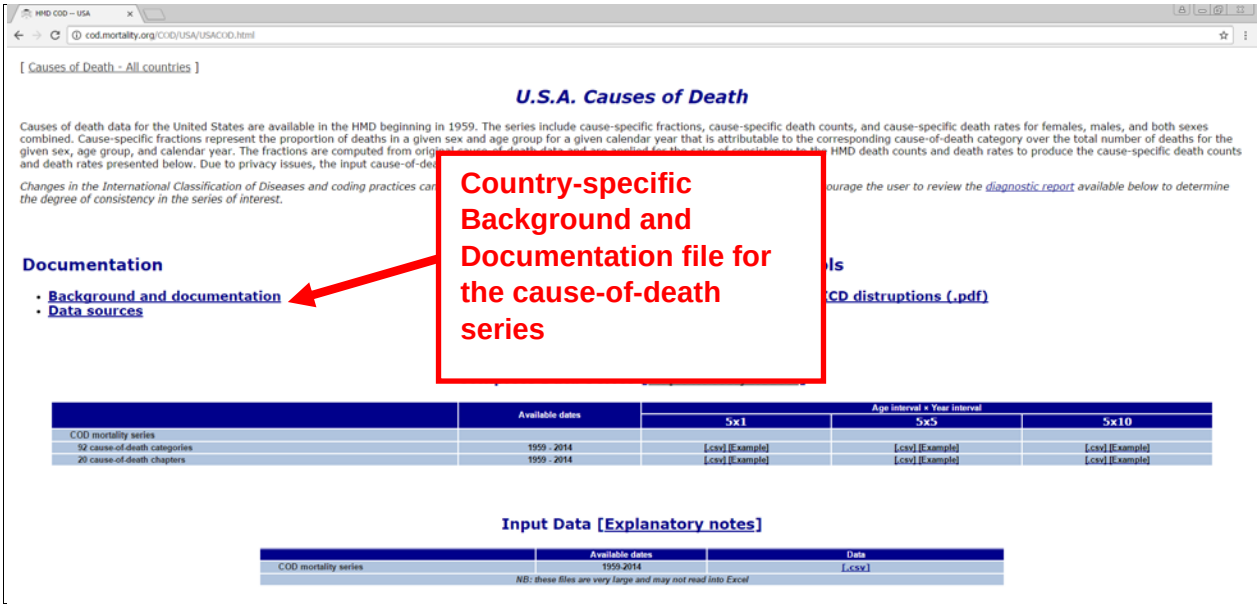
The "Background and documentation" file accessible from the country-specific cause-of-death page under the "Documentation" heading (Figure 6) presents a history of the collection of cause-of-death data (including the time periods when each ICD Revision was implemented), information about any

© 2017 Society of Actuaries

specificities regarding the original cause-of-death data provided by the country national statistics office, a discussion of data sources (also listed in great details in the "Data sources" file available under the same heading), and a table comparing the death counts by year as calculated from the cause-of-death data. Indeed, since sources for cause-specific and all-cause data might differ, depending on the country, and because of delays in cause-of-death reporting due to legal constraints in all countries, there are typically some discrepancies between the two series. For the sake of consistency between the already existing all-cause series presented in the HMD and the new cause-of-death series, adjustments are carried out to force agreement in the death counts and death rates by calendar year, sex, and age between the two sets of mortality series as described in the explanatory notes.

Figure 6.

COUNTRY-SPECIFIC DOCUMENTATION FILE FOR THE CAUSE-OF-DEATH SERIES

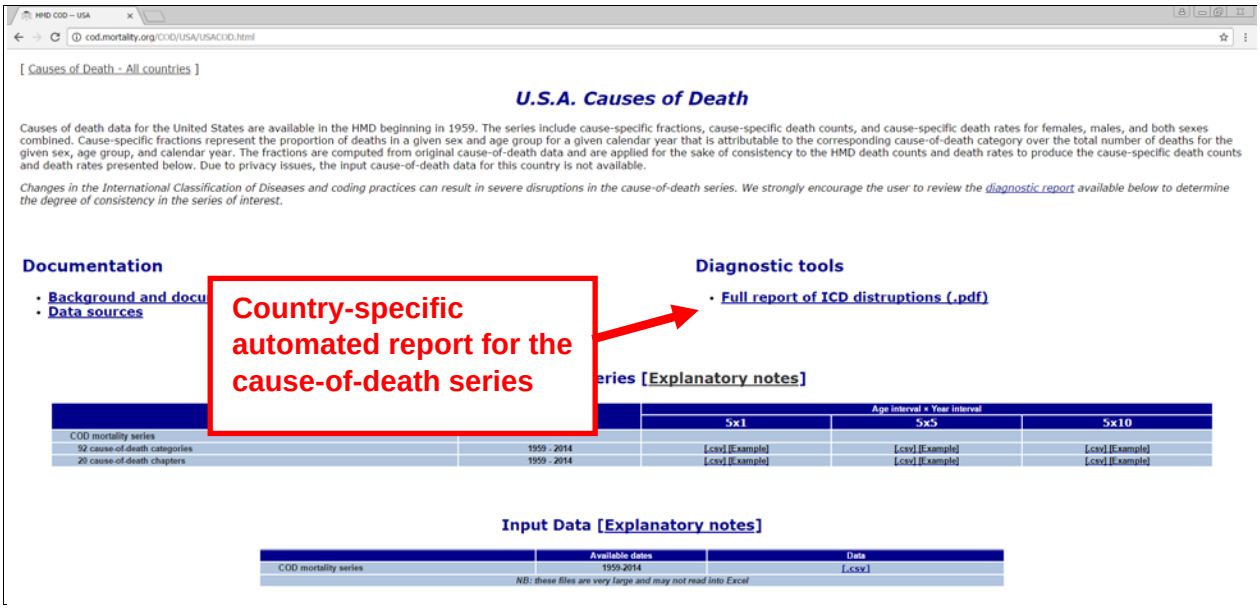


The country-specific diagnostic report includes a title page with the exact date of production and the years for which the HMD cause-of-death series are available for that country and one page each for the 92 cause-of-death categories (Figure 7).

© 2017 Society of Actuaries

Figure 7.

AUTOMATED COUNTRY-SPECIFIC REPORT FOR THE CAUSE-OF-DEATH SERIES



Each page lists the cause-of-death category number and disease name, presents a table showing the International Classification of Disease (ICD) codes under each Revision, a figure of the death counts per year with a red marker at the time when a new ICD Revision was introduced (the upper and lower edges of the red marker roughly corresponds to a confidence interval), and another table presenting the statistical score of the disruptions introduced by each ICD change (the higher the score, the more severe the disruption in trends associated with the transition from one ICD-Revision to another). A technical description of how the red markers on the figure and the scores in the bottom table have been constructed is provided in Appendix A, which is also reproduced at the end of each country report). Figure 8 provides an example of a typical report page.

© 2017 Society of Actuaries

Figure 8.

EXAMPLE OF A TYPICAL PAGE IN A COUNTRY-SPECIFIC DIAGNOSTIC REPORT

47 : Arteriosclerotic/ischaemic and degenerative

Table 94: Code ranges by ICD

	ICD	Codes
1	10	I200-I259
2	6	4200-4209
3	7	4200-4209
4	8	4100-4149
5	9	4100-4149

Figure 47: Death counts (all ages, both sexes) by year, HMD COD n° 47, The United States of America, 1959-2014

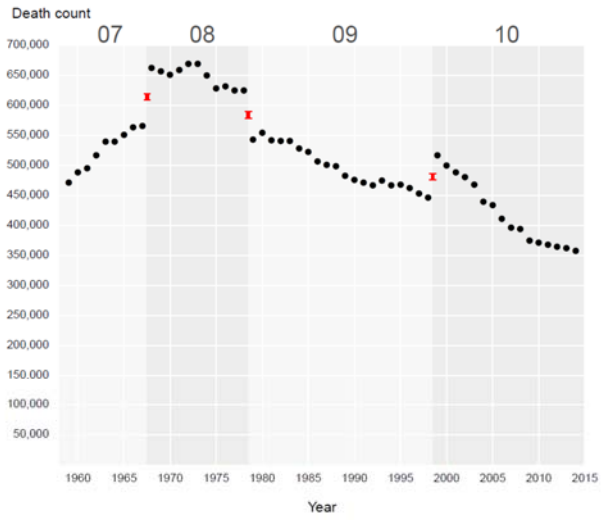


Table 95: Disruption scores by ICD transition

	Transition	Year	Score
1	ICD-7/ICD-8	1968	17.95
2	ICD-8/ICD-9	1979	14.62
3	ICD-9/ICD-10	1999	13.74

© 2017 Society of Actuaries

Appendix A—List of 92 disease categories and 20 broad groupings with corresponding ICD codes

(Sixth through Tenth Revisions) and consistency with Eurostat and NCHS groupings

Group	Cause	Disease(s)	ICD-6 code(s)	ICD-7 code(s)	ICD-8 code(s)	ICD-9 code(s)	ICD-10 code(s)	Eurostat category	NCHS category
I		Certain infectious diseases	001-138, 571, 764	001-138, 571, 764	000-136	001-139	A00-B99		
	1	Tuberculosis	001-008, 010-019	001-008, 010-019	010-019	010-018	A15-A19, B90	x	x
	2	Syphilis	020-029	020-029	090-097	090-097	A50-A53		x
	3	Infectious Gastro-enteritis	571, 764	571, 764	008, 009	008, 009	A04, A08-A09		
	4	Dysentery	045-048	045-048	004, 006	004, 006	A03, A06		
	5	Meningococcal infection	057	057	036	036	A39	x	
	6	Septicemia	053	053	038	038	A40-A41		
	7	Viral hepatitis	092	092	070	070	B15-B19	x	
	8	HIV-AIDS	--	--	--	042-044	B20-B24	x	x
	9	Other infectious diseases	030-044, 049-052, 054-056, 058-091, 093-138	030-044, 049-052, 054-056, 058-091, 093-138	000-003, 005, 007, 020-035, 037, 039-068, 071-089, 098-136	001-003, 005, 007, 020-035, 037, 039-041, 045-066, 071-088, 098-139	A00-A02, A05, A07, A20-A38, A42-A49, A54-B09, B25-B89, B91-B99		
II		Malignant neoplasm	140-205	140-205	140-209	140-208	C00-C97		x
	10	Lip/oral cavity/pharynx	140-148	140-148	140-149	140-149	C00-C14	x	
	11	Esophagus	150	150	150	150	C15	x	
	12	Stomach	151	151	151	151	C16	x	x
	13	Colon	153	153	153	153	C18	x	x
	14	Rectum	154	154	154	154	C19-C21	x	
	15	Liver, gallbladder and bile ducts	155, 156	155, 156	155, 156	155-156	C22-C24	x	
	16	Pancreas	157	157	157	157	C25	x	x
	17	Larynx	161	161	161	161	C32	x	
	18	Trachea, bronchus, and lung	162, 163	162, 163	162	162	C33-C34		x
	19	Skin	190, 191	190, 191	172, 173	172, 173	C43-C44	x	
	20	Breast	170	170	174	174, 175	C50	x	x
	21	Cervix uteri	171	171	180	180	C53	x	x
	22	Other parts of uterus	172-174	172-174	182	179, 182	C54-C55	x	
	23	Ovary	175	175	183	183	C56	x	
	24	Prostate	177	177	185	185	C61	x	x
	25	Kidney	180	180	189	189	C64-C66, C68	x	x
	26	Bladder	181	181	188	188	C67	x	

© 2017 Society of Actuaries

Group	Cause	Disease(s)	ICD-6 code(s)	ICD-7 code(s)	ICD-8 code(s)	ICD-9 code(s)	ICD-10 code(s)	Eurostat category	NCHS category
	27	Hodgkin lymphoma	201	201	201	201	C81		
	28	Non-Hodgkin lymphoma	200, 202, 205	200, 202, 205	200, 202	200, 202	C83-C85		x
	29	Myeloma	203	203	203	203	C90		
	30	Leukemia	204	204	204-207	204-208	C91-C95	x	x
	31	Other malignant neoplasms	152, 158-160, 164-165, 176, 178-179, 192-199	152, 158-160, 164-165, 176, 178-179, 192-199	152, 158-160, 163-171, 181, 184, 186-187, 190-199, 208-209	152, 158-160, 163-171, 176, 181, 184, 186-187, 190-199	C17, C26-C31, C37-C41, C45-C49, C51, C52, C57-C60, C62-C63, C69-C80,C82, C86-C88, C96-C97		x
III	32	Other neoplasms	210-239	210-239	210-239	210-239	D00-D48		
IV	33	Diseases of the blood and blood-forming organs	290-299	290-299	280-289	279-289	D50-D89	x	
V		Endocrine, nutritional and metabolic diseases	250-289	250-289	240-279	240-278	E00-E90	x	
	34	Diabetes Mellitus	260	260	250	250	E10-E14	x	x
	35	Overweight, obesity, and other hyperalimentation	287	287	277	278	E65-E68		
	36	Other endocrine, nutritional and metabolic diseases	250-254, 270-286, 288-289	250-254, 270-286, 288-289	240-249, 251-276, 278-279	240-246, 251-277	E00-E07, E15-E64, E70-E90		
VI		Mental and behavioural disorders	300-326	300-326	290-315	290-319	F00-F99	x	
	37	Alcohol abuse (including alcoholic psychosis)	307, 322	307, 322	291, 303	291, 303	F10	x	
	38	Drug dependence, toxicomania	323	323	304-305	304-305	F11-F19	x	
	39	Other mental and behavioural disorders	300-306, 308-321, 324-326	300-306, 308-321, 324-326	290, 292-302, 306-315	290, 292-302, 306-319	F00-F09, F20-F99		
VII		Diseases of the nervous system and the sense organs	340-398	340-398	320-389	320-389	G00-G44, G46-H95	x	
	40	Meningitis (other than meningoccocal and tuberculous)	340	340	320	320-322	G00-G03	x	
	41	Parkinson's disease	350	350	342	332	G20-G21		
	42	Alzheimer's disease	--	--	--	331	G30		x
	43	Multiple sclerosis	345	345	340	340	G35		
	44	Other diseases of the nervous system and the sense organs	341-344, 351-398	341-344, 351-398	321-333, 341, 343-389	323-330, 333-337, 341-389	G04-G13, G23-G26, G31, G36-G44, G46-H95		

© 2017 Society of Actuaries

Group	Cause	Disease(s)	ICD-6 code(s)	ICD-7 code(s)	ICD-8 code(s)	ICD-9 code(s)	ICD-10 code(s)	Eurostat category	NCHS category
VIII		Heart disease	400-447, 465	400-447, 465	390-429, 450	390-429	I00-I52	x	x
	45	Chronic rheumatic heart diseases	410-416	410-416	393-398	393-398	I05-I09		
	46	Acute rheumatic heart diseases	400-402	400-402	390-392	390-392	I00-I02		
	47	Arteriosclerotic/ischaemic and degenerative	420	420	410-414	410-414	I20-I25	x	x
	48	Hypertensive heart disease	440-447	440-447	400-404	401-405	I10-I15		
	49	Pulmonary heart disease and diseases of pulmonary circulation	465	465	426, 450	415-417	I26-I28		
	50	Other forms of heart disease	421-434	421-434	420-425, 427-429	420-429	I30-I52		
IX	51	Cerebrovascular disease	330-334	330-334	430-438	430-438	I60-I69, G45	x	x
X		Other and unspecified disorders of the circulatory system	450-464, 466-468	450-464, 466-468	440-448, 451-458	440-459	I70-I99		
	52	Artherosclerosis	450	450	440	440	I70		x
	53	Aortic aneurysm	451	451	441	441	I71		
	54	Other diseases of arteries, arterioles and capillaries	452-456	452-456	442-448	442-449	I72-I78		
	55	Other disorder of the circulatory system	460-464, 466-468	460-464, 466-468	451-458	451-459	I80-I99		
XI		Respiratory diseases	240-241, 245, 470-527, 763	240-241,245, 470-527, 763	460-474, 480-486, 490-493, 500-519	460-519	J00-J47, J60-J98, U04	x	
	56	Influenza	480-483	480-483	470-474	487	J09-J11	x	x
	57	Other acute respiratory infections	470-475, 500	470-475, 500	460-466	460-466	J00-J06, J20-J22, U04		
	58	Pneumonia	490-493, 763	490-493, 763	480-486	480-486	J12-J18	x	x
	59	Chronic bronchitis	501-502	501-502	490, 491	490, 491	J40-J42		
	60	Asthma	241	241	493	493	J45-J46	x	
	61	Other obstructive pulmonary diseases	526-527	526-527	492, 518	492, 494, 496	J43-J44, J47	x	x
	62	Other respiratory diseases	240, 245, 510-525	240, 245,510-525	500-517, 519	470-478, 488, 495, 500-519	J30-J39, J60-J98		

© 2017 Society of Actuaries

Group	Cause	Disease(s)	ICD-6 code(s)	ICD-7 code(s)	ICD-8 code(s)	ICD-9 code(s)	ICD-10 code(s)	Eurostat category	NCHS category
XII		Diseases of the digestive system	530-570, 572-587	530-570, 572-587	520-577	520-579	K00-K93	x	
	63	Gastric and duodenal ulcer	540-542	540-542	531-534	531-534	K25-K28	x	x
	64	Gastro-enteritis (non-infectious)	543, 572	543, 572	535, 561-563	535, 555, 556, 558, 562	K29, K50-K52, K57		
	65	Chronic liver diseases and cirrhosis	581	581	571	571	K70, K73-K74	x	x
	66	Other diseases of the digestive system	530-539, 544-570, 573-580, 582-587	530-539, 544-570, 573-580, 582-587	520-530, 536-560, 564-570, 572-577	520-530, 536-553, 557, 560-561, 563-570, 572-579	K00-K22, K30-K46, K53-K56, K58-K67, K71-K72, K75-K93		
XIII		Diseases of the skin and subcutaneous tissue	243-244, 690-716	243-244, 690-716	680-709	680-709	L00-L99	x	
	67	Infections of skin and subcutaneous tissue	690-698	690-698	680-686	680-686	L00-L08		
	68	Non-infectious diseases of the skin and subcutaneous tissue	243-244, 700-716	243-244, 700-716	690-709	690-709	L09-L99		
XIV		Diseases of the musculoskeletal system/connective tissue	720-749	720-749	710-738	710-739	M00-M99	x	
	69	Rheumatoid arthritis and osteoarthrosis	722-723	722-723	712-713	714-715	M05-M06, M15-M19	x	
	70	Other diseases of the musculoskeletal system/connective tissue	720-721, 724-749	720-721, 724-749	710-711, 714-738	710-713, 716-739	M00-M03, M08-M13, M20-M99		
XV		Diseases of the genitourinary system	590-637, 792	590-637, 792	580-629, 792	580-629	N00- N99	x	
	71	Nephritis, nephrosis and renal failure	590-594, 792	590-594, 792	580-584, 792	580-586	N00-N05, N17-N19	x	x
	72	Infections of kidney	600	600	590	590	N10-N12, N13.6, N15		
	73	Other diseases of kidney and ureter	601-604	601-604	591-594	587-589, 591-594	N06-N08, N13.0-N13.5, N13.7-N14, N16, N20-N29		
	74	Other diseases of the genitourinary system	605-637	605-637	595-629	595-629	N30-N99		
XVI	75	Complications of pregnancy, childbirth and puerperium	640-689	640-689	630-678	630-676	O00-O99	x	x
XVII	76	Certain conditions originating in the perinatal period	760-762, 765-776	760-762, 765-776	760-779	760-779	P00-P99	x	x

© 2017 Society of Actuaries

Group	Cause	Disease(s)	ICD-6 code(s)	ICD-7 code(s)	ICD-8 code(s)	ICD-9 code(s)	ICD-10 code(s)	Eurostat category	NCHS category
XVIII		Congenital malformations/anomalies	750-759	750-759	740-759	740-759	Q00-Q99	x	x
	77	Congenital malformations of the nervous system	750-753	750-753	740-743	740-742	Q00-Q07	x	
	78	Congenital malformations of the circulatory system	754	754	746-747	745-747	Q20-Q28	x	
	79	Other congenital malformations/anomalies	755-759	755-759	744-745, 748-759	743-744, 748-759	Q10-Q18, Q30-Q99		
XIX		Ill-defined or unknown	242, 780-791, 793-795	242, 780-791, 793-795	780-791, 793-796	780-799	R00-R99	x	
	80	Senility without psychosis	794	794	794	797	R54		
	81	Sudden death	--	--	795	798	R95-R96	x	x
	82	Unknown and unspecified causes	795	795	796	799	R97-R99	x	
	83	Other ill-defined or unknown	242, 780-791, 793	242, 780-791, 793	780-791, 793	780-796	R00-R53, R55-R94		x
XX		External causes	E800-E999	E800-E999	E800-E999	E800-E999	V01-Y98	x	
	84	Motor vehicle accidents	E810-E835	E810-E835	E810-E823	E810-E819, E820-E825	V02-V04, V09.0, V09.2, V12-V14, V19.0-V19.2, V19.4-V19.6, V20-V79, V80.3-V80.5, V81.0-V81.1, V82.0-V82.1, V83-V86, V87.0-V87.8, V88.0-V88.8, V89.0, V89.2	x	x
	85	Accidental falls	E900-E904	E900-E904	E880-E887	E880-E888	W00-W19	x	
	86	Accidental poisoning by alcohol	E880	E880	E860	E860	X45		
	87	Other accidental poisoning	E870-E879, E881-E895	E870-E879, E881-E895	E850-E859, E861-E877	E850-E859, E861-E869	X40-X44, X46-X49	x	

© 2017 Society of Actuaries

Group	Cause	Disease(s)	ICD-6 code(s)	ICD-7 code(s)	ICD-8 code(s)	ICD-9 code(s)	ICD-10 code(s)	Eurostat category	NCHS category
	88	Other accidents	E800-E802, E840-E866, E910-E965	E800-E802, E840-E866, E910-E965	E800-E807, E825-E845, E890-E949	E800-E807, E826-E848, E870-E879, E890-E949	V01, V05-V06, V09.1, V09.3-V11, V15-V18, V19.3, V19.8-V19.9, V80.0-V80.2, V80.6-V80.9, V81.2-V81.9, V82.2-V82.9, V87.9, V88.9, V89.1, V89.3-V89.9, V90-V99, W20-X39, X50-X59	x	x
	89	Suicide	E970-E979	E970-E979	E950-E959	E950-E959	X60-X84	x	x
	90	Homicide	E980-E985	E980-E985	E960-E969	E960-E969	X85-Y09	x	x
	91	Events of undetermined intent	--	--	E980-E989	E980-E989	Y10-Y34	x	
	92	Other external causes	E990-E999	E990-E999	E970-E979 E990-E999	E970-E979 E990-E999	Y35-Y98		x

© 2017 Society of Actuaries

Appendix B— Diagnostic tools to assess disruptions due to changes in the ICD[1]

Two diagnostic tools have been developed to help users assess the impact of ICD change on cause-of-death series: a statistical marker, hereby called disruption score, and a graphical marker of the degree of disruption for each specific cause.

1. COMPUTATION OF THE DISRUPTION SCORE

The disruption scores show the degree of discontinuity in a series of death counts at the time when a new ICD scheme is introduced. One disruption score is computed separately for each ICD transition and cause-of-death category (for both sexes and all ages combined) in each country. The score itself is similar to a t-statistics in that it is computed as the ratio of the difference between the estimated number of deaths in the years just before and just after the ICD change to the root mean square deviation of expected (fitted) to observed deaths for the period before and after the change in ICD. The expected numbers of deaths are estimated from a robust smoothing of the observed deaths carried out once for the period before the ICD change and once for the period after the change.

The disruption score $S_i(t)$ for the death counts from cause i during a year t when a new ICD revision was introduced is defined as

$$S_i(t) = \frac{| \hat{d}_i^*(t-1) - \hat{d}_i(t)}{RSME(t)}$$

with $\hat{d}_i^*$ the expected number of deaths from cause i for the year just before the ICD change and the expected number of deaths from cause i for the year when the revised ICD was implemented and $RSME(t)$ the root mean square deviation of expected to observed deaths for the period before and after the change in ICD. For a dynamic explanation of the mathematics behind disruption scores, see http://www.youtube.com/watch?v=wP_7RXjV36A .

Alternative robust smoothing models have been examined but we found that the disruption score is sensitive enough to small changes in smoothing forms or parameters (whether using parametric or non-parametric models). Indeed, the disruption score is much more sensitive to the value of the root mean square deviation of expected to observed deaths which varies little from a model to another. Furthermore, the disruption score is not intended to be used as a diagnostic tool but as a screening tool: no decision is made based on the score itself; it used simply used to quickly separate out those cause-specific death count series that deserve a second look from those which are highly consistent over time.

2. GRAPHICS

Each plot shows the numbers of deaths for a specific cause for the years before and after a change in ICD scheme. A vertical red bar marks the year in which the ICD system changed. The length of the bar is related to a "confidence interval" based on the variability of the death count series for the periods before and after the ICD change. To construct the red bar, the root mean square error was computed from a robust smooth of the numbers of deaths before and after the change in ICD system; this is

[1] The disruption score and the graphics have been designed by Robert Chung (University of California, Berkeley).

analogous to a standard deviation from a mean. The length of the red bar is twice the root mean square error and it is centered on the plot halfway between the ends of the robust smooth lines.

Figure B1. provides an example of a cause-of-death series, namely pneumonia, in the United States, with various degrees of disruptions depending on the ICD Revision. The dots represent the observed number of deaths in each calendar year while the grey lines correspond to the smooth in the observed deaths for each ICD period. Disruption scores are represented by the red markers. For this series, the scores are: 2.53 for the transition from ICD-7 to ICD-8 (in 1968), 3.95 for the transition from ICD-8 to ICD-9 (in 1979), and 11.36 for the transition from ICD-9 to ICD-10 (in 1998). As indicated by the red markers on the graph, the series might have been disrupted by the change in classification for all three transitions but the break is particularly large for the last transition (from ICD-9 to ICD-10), requiring caution when interpreting the overall trend.

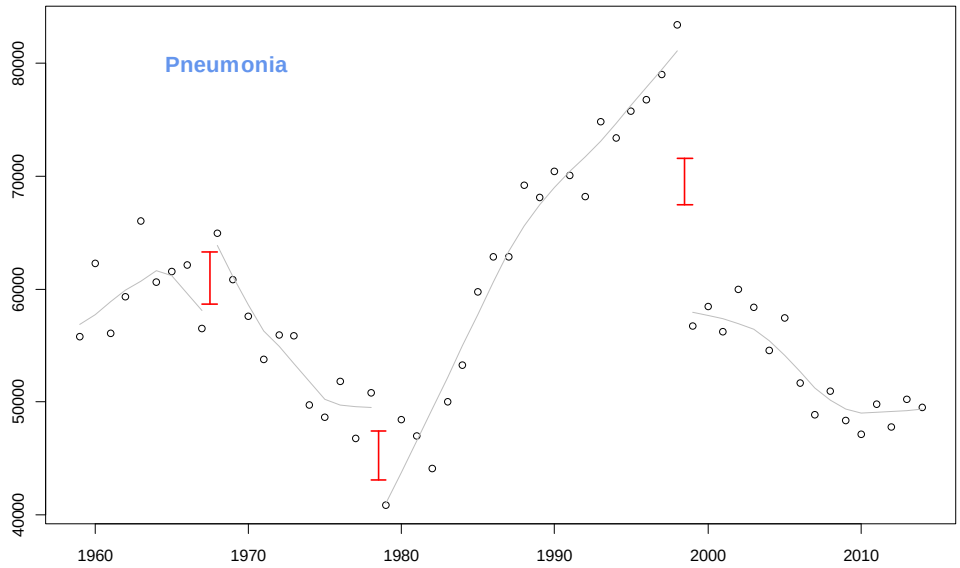


Figure B1. Number of deaths from pneumonia (all ages, both sexes) in the United States by calendar year, 1959-2014

The reason for the large disruption in the number of deaths from pneumonia between 1998 and 1999 results from some clarification by the World Health Organisation in the rule to select the underlying cause of death on the death certificate. Up to 1998, imprecision in the rule lead to the frequent selection of pneumonia, a common complication for many different diseases, as the underlying cause. With the new formulation of the selection rule in ICD-10 manuals, pneumonia is never supposed to be selected as the underlying cause when it is listed on the death certificate as a complication from other diseases, thus the significant drop in the number of deaths attributed to pneumonia between the two ICD periods.

Appendix C — Structure of the cause-of-death input database

[Note: This Appendix reproduces the instructions to country specialists to prepare the input data in a standardized format]

The Cause-of-death Input Database (COD-InputDB) closely follows the format of the HMD all-cause Input Database though it only contains one type of data: deaths by cause. For each population, we have 8 different files (9 where births-by-month data are available):

- XXXcod.txt (for death data),
- XXXcodref.txt (references),
- XXXcodnote.txt (for specific notes),
- XXXcodcom.pdf (for background and documentation),
- XXXreadme.txt (with a descriptive identifier for this population),

where XXX is the same population code as for the all-cause HMD series (3-7-digits, all character values in uppercase). The first three digits represent a 3-letter country code (e.g., FRA, USA). An optional fourth digit identifies the national population, where "T" denotes the total population (i.e., including both military and civilian) and "C" denotes the civilian population. In cases where this information is known or unclear (e.g., not consistent across deaths and population estimates), the fourth digit may be omitted (if the data represent the national population as for all cause-specific files for now) or indicated by "_" (if the data represent a subpopulation identified by the 5^{th}-7^{th} digits). The fifth, sixth, and seventh digits may be used to identify a subpopulation (e.g., region/province, racial/ethnic group) as needed. This code may be any combination of up to three numbers of uppercase characters. For example,

NOR	identifies the national population of Norway (unknown/uncertain whether military are included)
FRAT	identifies the total national population of France
FRAC	identifies the civilian national population of France
NZL_NP	identifies the national population of New Zealand
NZL_MA	identifies the Maori population of New Zealand

The cause-of-death Background and Documentation file (XXXcodcom.pdf) is in portable document format (.pdf). All other files are in ASCII format. Each data file contains a standard heading (first line), which represents a field identifier; the headings are case-sensitive. In this file, each line (record) is defined independently from other records (i.e., each record contains all necessary information, and the sequence of the records has no significance). Each record is unique (i.e., there are no duplicate records). We use CRLF ("\r\n") combination of characters as a record delimiter and a comma (",") as the field delimiter. Missing values are coded as a single dot (".") . Optionally, the data file may contain a number of spaces to improve text file readability. The spaces have no other function. [The order of the columns really does not matter to the computer, but it is best to keep to the order given in the examples.]

In addition to the input data file, four text files must be systematically provided and updated: the short description file (XXXreadme), which is the exact same one as for the all-cause HMD series, the cause-of-death Background and Documentation file (XXXcodcom), the cause-of-death Reference file (XXXcodref) and the cause-of-death Note file (XXXcodnote).

Description of formats:

1. Deaths by cause

File name: XXXcod.txt

Heading:

PopName, Area, Year, YearReg, YearInterval, Sex, Age, AgeInterval, ICD-code, ICDRevision, RefCode, Access, Deaths, NoteCode1, *NoteCode2*, *NoteCode3, LDB*

For each death count, we have one record. Each death count (field "*Deaths*") is determined by population code (field "*PopName*"), geographical coverage (field "*Area*"), calendar year (field "*Year*"), year of registration ("*YearReg*"), the length of year interval of the Lexis element (field "*YearInterval*"), sex (field "*Sex*"), age (field "*Age*"), the length of age interval (field "*AgeInterval*"), the shape of Lexis element (field "*Lexis*"), the ICD Revision in use during the year in the country (field "*ICDNumber*"), the cause of death, as coded in the International Classification of Diseases (field "*ICDCode*"), the reference code (field "*RefCode*"), and the type of access (field "*Access*"). Each record also contains three fields that link to specific comments ("*NoteCode1*", "*NoteCode2*", and "*NoteCode3*"). The final field ("*LDB*") indicates whether the data are used to create the Lexis database. Note: If this file includes death counts for a particular calendar year, but an observation is omitted for a particular sex, age, and cause-of-death combination, then for the purposes of calculating cause-specific mortality estimates, we assume there were no deaths for this group.

The format of the fields is as follows:

1.1) Population code (3-7-digits). Same code ("XXX") as in the file name.

1.2) Area (2-digit). For example: 01, 10, 20, …. This field serves to reflect territorial (or population) coverage.

1.3) Calendar year (4-digit). Year in which the deaths occurred.

1.4) Year of registration (4-digit). In recent years, some countries (e.g., Finland) have reported deaths registered in current year, but which actually occurred earlier in time. In such cases, *Year*, *Age,* etc. are coded to refer to the date of occurrence, while *YearReg* is set to the registration year. If no distinction is made between year of occurrence and year of registration[2], then *Year = YearReg*.

1.5) The length of the year interval (1- or 2-digit). For example: 1, 2, 3, 5, 10.

1.6) Sex (1-char). The character 'm' denotes males, 'f' denotes females, and 'b' indicates both.

1.7) Age (1-, 2- or 3-digit or 3-char). For age groups, the value is always equal to the lower age limit, and TOT[3] and UNK stand for total and unknown ages, respectively.

1.8) The length of the age interval (1 or 2 digits or '+' for open age interval). For example: 1, 2, 3, 5, 10, +.

1.9) Lexis element (2-char). This field denotes the shape of the Lexis element, where: TL=lower triangle, TU=upper triangle, RR=rectangle, VV=parallelogram with vertical left and right sides (i.e., period-cohort), VH=parallelogram with horizontal upper and lower sides (i.e., age-cohort); RV=Same as VV except also includes TL for the first age in the interval (e.g., cohorts aged 0-4 on Dec 31st—includes those born in the current calendar year).

1.10) ICD Revision in use in the given country and time period (1 digits for ICD-6 through ICD-9 and 2 digits for ICD-9).

---

[2] Recording deaths by year of occurrence is a new trend in statistical publications. In earlier years, deaths were recorded only by year of registration.

[3] The total is always taken from the original data. It may be different from the total yielded by summing up age specific counts where there have been errors (frequent in paper publications; exceedingly rare when the data are provided in electronic format).

1.11) The cause-of-death category as coded in the International Classification of Disease (3 or 4 digits for Revisions 6-9 of the ICD; 4 digits for Revision 10).

1.12) Reference code (numeric, no fixed-length). A numeric code that identifies a data source provided in the file XXXcodref.txt (see description for XXXcodref.txt).

1.13) Access (1-char). This field indicates the confidentiality/accessibilityof the data ('C' - confidential, 'O' – publicly accessible, 'U' – unknown).

1.14) Deaths (numerical field, no fixed length). Number of deaths.

1.15) Note codes (numeric, no fixed length). These three fields ("*NoteCode1*", "*NoteCode2*", "*NoteCode3*") link to specific notes contained in the file "XXXcodnote.txt". These fields may be empty (denoted by a single dot '.'). [NB: these codes should be of text type.]

1.16) LDB (numeric, 0/1). This field ("*LDB*") is coded "1" if the data are used to create the Lexis database and coded "0" if not. [This should be the last column of data in the file.]

Example: Total National Population of The United States

PopName	*Area*	*Year*	*YearReg*	*YearInterval*	*Sex*	*Age*	*AgeInterval*	*Lexis*	*ICDnumber*	*ICDcode*	*RefCode*	*Access*	*Deaths*	*NoteCode1*	*NoteCode2*	*NoteCode3*	*LDB*
USA	01	1959	1959	1	m	1	1	RR	7	0020	30	O	7	.	.	.	1
USA	01	1959	1959	1	m	1	1	RR	7	0100	30	O	17	.	.	.	1
USA	01	1959	1959	1	m	1	1	RR	7	0192	30	O	3	.	.	.	1
USA	01	1959	1959	1	m	1	1	RR	7	0420	30	O	3	.	.	.	1
USA	01	1959	1959	1	m	1	1	RR	7	0450	30	O	1	.	.	.	1
USA	01	1959	1959	1	m	1	1	RR	7	0454	30	O	3	.	.	.	1
USA	01	1959	1959	1	m	1	1	RR	7	0480	30	O	9	.	.	.	1
USA	01	1959	1959	1	m	1	1	RR	7	0491	30	O	1	.	.	.	1
USA	01	1959	1959	1	m	1	1	RR	7	0510	30	O	4	.	.	.	1
USA	01	1959	1959	1	m	1	1	RR	7	0530	30	O	1	.	.	.	1
USA	01	1959	1959	1	m	1	1	RR	7	0533	30	O	1	.	.	.	1
USA	01	1959	1959	1	m	1	1	RR	7	0534	30	O	36	.	.	.	1
USA	01	1959	1959	1	m	1	1	RR	7	0550	30	O	1	.	.	.	1
USA	01	1959	1959	1	m	1	1	RR	7	0560	30	O	9	.	.	.	1
USA	01	1959	1959	1	m	1	1	RR	7	0561	30	O	10	.	.	.	1

2. Cause-of-death notes

File name: XXXcodnote.txt

This file contains specific notes pertaining to individual cause-of-death data points. These notes are identified by a note code that links to specific data points in the data file. The first line of each record contains a code (the same as *NoteCode1, NoteCode2*, and *NoteCode3* in the death file). Lines following the note code contain specific notes (which may be written in a free form, but shouldn't include blank lines). Blank lines are used for separating the records. Each subpopulation within a country (eg. England and Wales, Scotland, and Ireland within the United Kingdom) must have its own XXXcodnote.txt file, but it is left to the discretion of the country specialist whether those files are unique (containing only information for that particular subpopulation) or whether they are simply copies of one master file (containing information for all subpopulations within that country) with different filenames (e.g., GBRTENWcodnote.txt, GBR_NIRcodnote.txt, and GBR_SCOcodnote.txt).

Example:

1
This number is probably not correct, but we don't have other data

2
For this year and age we need additional information

……

3. References

File name: XXXcodref.txt

This file contains the description of sources. Each data source has a corresponding record in this file. Each record in this file is separated by a blank line. Each subpopulation within a country must have a separate XXXref.txt file, but it is left to the discretion of the country specialist whether those files are unique (containing only information for that particular subpopulation) or whether they are simply copies of one master file (containing information for all subpopulations within that country) with different filenames (e.g., GBRTENWcodref.txt, GBR_NIRcodref.txt, and GBR_SCOcodref.txt). Each record contains the following fields:

- *RefCode:* Reference code (the same as in the data files)
- *Source:* Full (or as full as possible) reference to the source
- *Comments:* List the calendar year(s) to which each reference code corresponds. This field may also contain comments from the country specialist about this submission.
- *Date:* Date of incorporation into the Database (format: dd.mm.yyyy)
- *Reference Person:* Person who added this entry to the reference file

The name of each field is included in the record on a separate line.

Example:

RefCode
30
Source
National Center for Health Statistics, United States. Mortality Data – Multiple Cause of Death Public Use Files, 1959-2007, downloaded from http://www.nber.org/data/vital-statistics-mortality-data-mulitiple-cause-of-death.html.
Comments
(Deaths by cause) Data for Deaths 1959-2007. Tabulation of death counts by cause, year of occurrence, sex and age carried out by Magali Barbieri from individual records in electronic format.
Date
12-Sept-2012
Reference Person
Magali Barbieri (magali@demog.berkeley.edu)

4. Cause-of-death Background and Documentation

File name: XXXcodcom.pdf

This file is constructed by the country specialist and made available to users. For a given population, the Background and Documentation for the primary population (e.g., Total National Population) should contain the information that pertains to the population as a whole (e.g., completeness of coverage, period when each ICD Revision was in use, data quality, historical information, country-specific ICD codes, etc.). The Background and Documentation for *subpopulations* should refer the user to the "main" Background and Documentation, but provide any additional information that is specific to that subpopulation. Information related to specific data points should be included in the Notes file rather than the Background and Documentation file. For example, comments regarding the source of data from country C for calendar year Y are given in the "*Comments*" field of the reference file. However, comments regarding comparative characteristics of different data sources for year Y of country C should be stored in XXXcodcom.pdf.

This file contains all (important) information that is known to country specialist. Nevertheless, it cannot be used as a full description of statistical system or history of the country.

5. Descriptive Identifier

File name: XXXreadme.txt

This file is borrowed from the all-cause HMD Input database. The first line of this file contains a descriptive identifier for the population (i.e., name of country or area, type of population, and population subgroup) that will be used in the header for all data files displayed on the HMD website. For populations where the population classifier (e.g., total or civilian) is indeterminate, it is particularly important that the population is described to the best of your knowledge. At the discretion of the country specialist, other unspecified information (e.g. history of updates) can be included on subsequent lines.

For example:

FRATNPreadme.txt reads:
France, National Total Population

ITA_NPreadme.txt reads:
Italy, National Population (excludes military deaths, but population includes military)

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 24,000 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 non-actuaries 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