MORTALITY DATA FOR THE UNITED STATES MORTALITY DATABASE (USMDB)[1]

By Magali Barbieri and Celeste Winant

INTRODUCTION

The United States Mortality DataBase (USMDB) is developed at the University of California, Berkeley (United States) as part of the larger umbrella project of the Human Mortality Database (HMD), a collaboration between the Max Planck Institute for Demographic Research (MPIDR, Rostock, Germany) and the French Institute for Demographic Studies (INED, Paris, France). The purpose of the USMDB is to provide researchers and others interested in the study of longevity with free and easy access to detailed, up-to-date, and comparable sub-regional mortality data through its website (at usa.mortality.org).

The database contains original period life tables for US geographic entities that are updated every year, when new demographic data become available. Lifetables are provided for years since 1941 for the 4 Census Region, the 9 Census Division, and the 50 states plus the District of Columbia (except for Alaska and Hawaii for which the series start in 1959). The USMDB also provides a period lifetable series at the county level for all US counties for years since 1982. While the methods used to construct the lifetables at the higher regional levels (Census regions, divisions, and states) rely on those of the Human Mortality Database, a different approach had to be followed for the county-level lifetables because county populations are often too small to implement the HMD methods in a meaningful way. This document describes the sources and adjustments to the HMD methods carried out for the construction of the higher-level lifetable series. A separate document [https://usa.mortality.org/uploads/Public/counties/USCountyBayesianEstimationMethodsProtocol20230913web.pdf] describes the sources and statistical modeling approach used to construct the life table series at the county level.

TERRITORIAL COVERAGE

The USMDB includes lifetables series for years since 1941 for the United States as a whole, the 4 Census Regions and Divisions (as defined here: https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf) and the 48 states and the District of Columbia (starting in 1959 for Alaska and Hawaii), in addition to all US counties as described elsewhere (https://usa.mortality.org/uploads/Public/counties/USCountyBayesianEstimationMethodsProtoco

[1] 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.

l20230913web.pdf). The lifetables presented in the USMDB refer to the US resident population. Non-residents were excluded from both the population and the vital statistics data.

Prior to 1959, Alaska and Hawaii were U.S. territories. For the period 1941-1958, we compute life tables for the remaining 48 states and the District of Columbia. 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. Alaska and Hawaii joined Census Division 9 and Census Region 4 when these states became admitted to the Union in 1959. For this current update, lifetables for Census Division 9, Census Region 4, and the United States are separated into two different periods (1941-1958, before the admission of Alaska and Hawaii, and 1959-2020). In the future, we will create single series respectively for each of these geographic aggregates incorporating a territorial adjustment between 1958-1959, following standard HMD methodology.

SOURCES OF DEMOGRAPHIC DATA

The USMDB relies on two distinct sources of statistical information to construct the lifetables: death and birth records from the national vital statistics system and population estimates from the Census Bureau.

Population data

National censuses and population estimates for the United States are collected or constructed by the U.S. Department of Commerce at the US Census Bureau (www.census.gov). The USMDB population data originate from population censuses conducted every 10 years. The USMDB makes use of the censuses conducted from 1940 to 2010. Census counts serve as the basis for producing the 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 1941 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).

For all population count data other than respectively from the 1950 and 1960 Census, no adjustments have been made to the published population estimates.

Vital statistics data

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/). Mortality and natality 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 births and deaths of non-residents and, consequently, to exclude them from tabulations. Therefore, for the years 1941–1969, births and 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.

Historical data In the early years of the 20th century, vital statistics for the United States were based on data from those states admitted to the Birth and Death Registration Areas, the number of which increased over time. To be included in the Birth and Death Registration Areas, 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 Birth and Death Registration Areas. Given the legal requirements for birth and death registration, vital statistics data for the United States are considered to be complete and of acceptable quality since 1933.

Early mortality data

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). However, for issues which have to do with the availability of the detailed information necessary to construct complete lifetables, the USMDB data series only start with calendar year 1941. The present format of the 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.

For years 1941 – 1958, mortality tabulations by age, sex, and state are found in the annual US Census Bureau publications of Vital Statistics of the United States. For years 1941-1949, these tables are published in the volume titled Natality and Mortality Data for the United States Tabulated by Place of Residence. For 1950-1958, these tables are published in the volumes titled Mortality Data. The mortality tabulations for years 1941-1958 were digitized and carefully verified.

Electronic mortality data

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.

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. This includes data 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.

Natality data

Following the HMD Methods Protocol, birth data are only used to adjust cohort population estimates to more accurately estimate Lexis triangle exposure counts when possible. This adjustment is particularly important for years when large month-to-month fluctuations are observed, mostly around the Second World War. Because of the lack of monthly birth tabulations by state, this method is not implemented and by default, the Protocol assumes a uniform distribution of births throughout each calendar year. Births by sex are used to accurately estimate Lexis triangle exposure for the calculation of the infant mortality rate. Births by sex tabulation are available by Census region, division, and state for all years since 1941.

SPECIFIC METHODS

Beyond the Human Mortality Database standard Methods Protocol, we needed to implement a few specific adjustments to the regional US data as explained below.

Deriving Death Counts by Lexis triangles from tabulations

Death counts for years 1941-1958 are available from the Census tabulated by 5-year age groups (ages <1, 1-4, 5-9, 10-14…, 90-94, 95+ (for years 1955-1958) and 95-99, 100+ (for years 1939-1954). For years 1940-1948 and 1955-1958, deaths of unknown age are tabulated separately from deaths of known age. For years 1949-1954, deaths of unknown ages are tabulated with deaths for ages 100 and over. We disaggregate the deaths of unknown age from those with ages 100 and over by applying the mean ratio of deaths from these respective categories over the years 1946-48.

$$D_{UNK}(y = 1949{:}1954) = D_{UNK\ \&\ 100+}(y = 1949{:}1954)\left\langle \frac{D_{UNK}(\tilde{y})}{D_{UNK}(\tilde{y}) + D_{100+}(\tilde{y})} \right\rangle_{\tilde{y}\ in\ 1946:1948}$$

$$D_{100+}(y = 1949{:}1954) = D_{UNK\ \&\ 100+}(y = 1949{:}1954)\left\langle \frac{D_{100+}(\tilde{y})}{D_{UNK}(\tilde{y}) + D_{100+}(\tilde{y})} \right\rangle_{\tilde{y}\ in\ 1946:1948}$$

Where *D* are deaths, *UNK* signifies unknown age, *100+* signifies ages 100 and over, and *UNK & 100+* signifies both deaths of unknown age and those of ages 100 and over.

We next apply HMD methods to first split deaths from five-year age groups and the open age group starting at age 100 for years 1940-1954 and 95 for years 1955-1958 into deaths by single year of age. We do not have information about date of death at the individual record level, so deaths by single year of age for these years are split into Lexis triangles using HMD methods.

Deriving Death Counts by Lexis triangles 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 1959-1988, 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. The specific procedure used to derive the death counts by Lexis triangle for years prior to 1989 is described in Appendix 2.

Constructing annual population estimates for years 1941-1969

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 are available by single year of age since 1959 and by five-year age group for prior decades. To compute death rates by single year of age or by five-year age group, they must be combined with population counts for the corresponding ages, 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 1941-1959 using the classic demographic approach of a cohort component method as described in the 2007 Human Mortality Database Methods Protocol (https://mortality.org/File/GetDocument/Public/Docs/MethodsProtocolV6.pdf). For Alaska and Hawaii, the 1959 estimates were derived by back projecting the 1960 population estimates using the death counts by age within each cohort as well as the birth count for 1959.

For both the 1950 and 1960 Census, tabulations of the population by sex, state, and single year of age are only available for a select sample of the state's population (20% in 1950, 25% in 1960). Tabulations which include enumeration of the entire state population are published with less age detail. Before using these data to construct annual population estimates, we thus imposed the age structure of the single-year-of-age tabulations on the counts from the less-detailed (and more complete) tabulations to create our own population estimates by sex and single year of age that preserves the complete population counts by sex enumerated in each state. For the 1940, 1970, 1980, 1990, 2000 and 2010 Census, tabulations of the population by sex, state, and single year of age are available for the complete sample of the US population.

In addition, for years 1960-1969, we identified two sets of annual population estimates available from the Census Bureau. A first set of estimates 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 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.

Birth data for 1967

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.

Systematic errors introduced by mortality shocks in populations estimates ages 80 and over for the most recently surviving cohorts

The COVID19 epidemic created a spike in death rates over all states, ages, and sexes over the period 2020-2022, with death totals increasing by as much as 30% for some states between years 2019 and 2020. The HMD method used to estimate populations and exposures to risk for the most recently surviving cohorts above age 80 (referred to as the Survivor Ratio Method in the HMD Methods Protocol) assumes that underlying mortality trends vary linearly and monotonically over time in these most recent years. Non-linear mortality shocks such as those introduced by the COVID epidemic violate that assumption. The method distributes the average effect of that shock over the cohorts that survive to age 85 and above, which impact population estimates both before and after the mortality shock, violating causality. While mortality rates in the periods leading up to and including the COVID19 epidemic are mostly influenced by the mortality shock itself, they are also systemically impacted by these non-causal revisions to the exposure at risk. All updates going back to the first that included 2020 mortality data thus have a systematic overestimate of mortality rates and all ensuing life table indicators for the most recently surviving cohorts, ages 80 and above, for years 2019 and prior. The method revises estimates of remaining life expectancy at age 80 (e80) for both females and males in 2019 for the United States as a whole by -0.3 years between the 2020- and 2019-updates, and estimates of e80 for both females and males in 2010 by -0.1 years. Differences in remaining life expectancy at ages 0 (e0), 65 (e65) and 80 (e80) for females and males for the United States, as a whole, between the 2019- and 2020- updates are shown in Figure 1 below.

We expect to see these effects in future updates of life table indicators for surviving cohorts at ages 80 and above, until the time that this shock only impacts extinct cohorts. The systematic effect should diminish with future updates as the shock impacts increasingly aged (and lower-numeracy) cohorts.

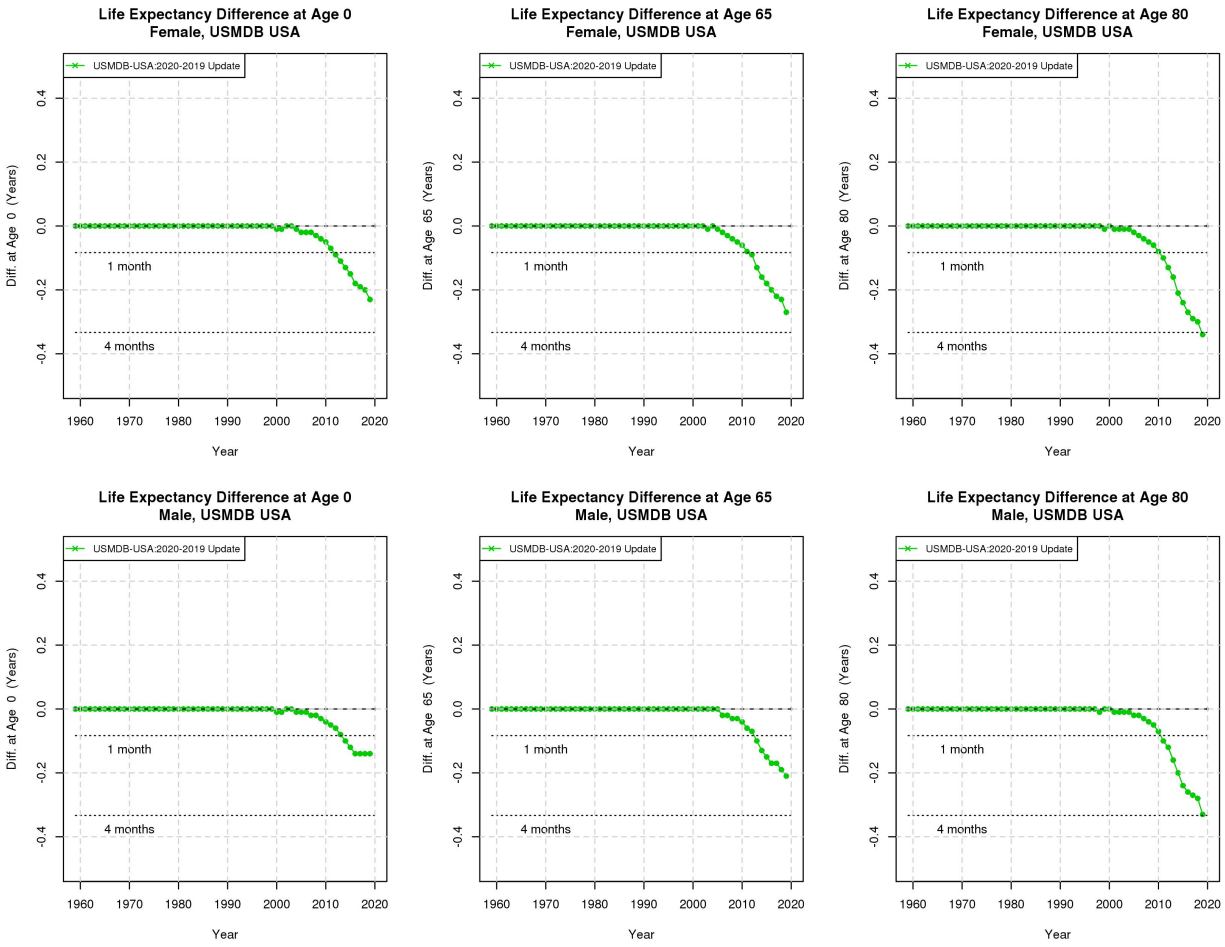


Figure 1: Difference in Remaining Life Expectancy from Ages 0, 65, and 80 between the 2020 update (1959-2020) and previous update (1959-2019) for Females (top) and Males (bottom) for the United States as a whole (USA)

ACKNOWLEDGEMENTS

We are grateful to Elizabeth Arias and Robert Anderson at NCHS for their collaboration to this project. We are also grateful to Nicholas Reynolds for preparing the initial digitization of the mortality data by state for years 1937-1958. 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 USMDB was initially funded by a grant from the National Institute of Aging at the National Institutes of Health (NIA-NIH grant #R01-AG040245, PI: Pr John Wilmoth) and later supported by funding from the University of California, Berkeley Center on the Economics and Demography of Aging (CEDA, NIA-NIH grant #P30-AG012839), another small grant from the NIA (NIA-NIH grant #R03-AG058110, PI: Magali Barbieri), and from the Society of Actuaries. Additional technical and research support has been provided by the French Institute for Demographic Studies (INED). 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 THE USMDB INPUT DATA[2]

POPULATION

Period	Type of Data	Age groups	Comments	RefCode(s)
1940	Official Census estimates	0,1,…,119, 120	Missing Alaska and Hawaii	6
1950	Official Census estimates by single year of age based on a 20% sample	0, 1, ... , 84, 85-89,90-94,95-99,100+	Missing Alaska and Hawaii	7
1950	Official Census estimates by 1, 2, 3, 4, 5 and 10-year age group based on complete sample	0,1-2,3-4,5,6,7-9,10-13,14,15,16-17,18-19,20-24, 25-29,…,65-69,75-84,85+	Missing Alaska and Hawaii	8
1950	Reconstruction of Official Census estimate	Single year of age 0, 1, .., 99, 100+	Missing Alaska and Hawaii	9
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
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

[2] The reference codes indicated in the last column of each of the tables below correspond to the sources listed in Appendix 3.

Period	Type of Data	Age groups	Comments	RefCode(s)
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
2011-2019	Intercensal population estimates for the resident population‡	0, 1 … 85+		120
2020-2023	Intercensal population estimates for the resident population‡	0, 1 … 85+		130

† For details, see this document section titled "Constructing annual population estimates for years 1941-1969."
‡ Data are available on the Census Bureau web site (http://www.census.gov). For the specific URLs and download dates, see Appendix 3 below.

DEATHS

Period	Type of Data	Age groups	Comments	RefCode(s)
1939-1948	Annual death counts for U.S. residents by geographic area, sex, and 5-year age groups	0, 1-4, 5-9, 10-14,…,95-99, 100+, UNK	Excludes Alaska & Hawaii	1939-1948
1949-1954	Annual death counts for U.S. residents by geographic area, sex, and 5-year age groups	0, 1-4, 5-9, 10-14,…,95-99, 100+& UNK	Excludes Alaska & Hawaii, deaths by unknown age (UNK) and ages 100 years and older are tabulated together	1949-1954
1954-1958	Annual death counts for U.S. residents by geographic area,	0, 1-4, 5-9, 10-14,…,90-94, 95+, UNK	Excludes Alaska & Hawaii	1954-1958

Period	Type of Data	Age groups	Comments	RefCode(s)
	sex, and 5-year age groups			
1959-2020	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.

BIRTHS

Period	Type of Data	Comments	RefCode(s)†
1939-1958	Annual births for the *de facto* population by sex	Does not include births for Alaska or Hawaii	2939-2958
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
2007–2023	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

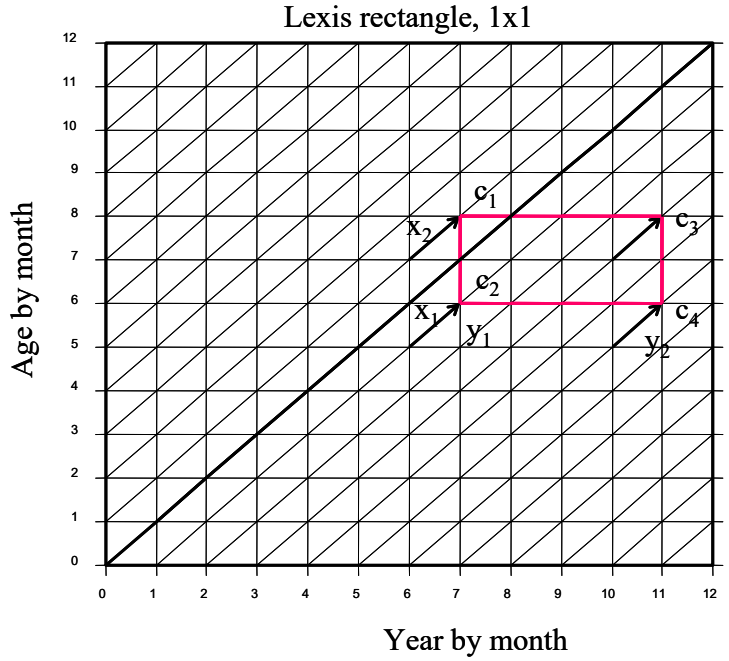
For years of data (1959-1988) when we lack information about the date of birth (and thus about the cohort to which the deceased belonged), we indirectly estimated the Lexis triangle to which deaths should be allocated. As noted earlier, the age variable identifies the age of the decedent at his/her last 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, 1989-2020

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	0.0048
1990	2148463	6023	45	2949	517	9534	0.0044
1991	2169518	5920	29	2476	547	8972	0.0041
1992	2175613	5120	7	1963	467	7557	0.0035
1993	2268553	4891	25	1883	482	7281	0.0032
1994	2278994	4181	29	1578	385	6173	0.0027
1995	2312132	3520	36	1646	427	5629	0.0024
1996	2314690	3584	25	1603	498	5710	0.0025
1997	2314245	3355	17	1697	384	5453	0.0024
1998	2337256	3275	2	1400	409	5086	0.0022
1999	2391399	2980	8	1077	348	4413	0.0018
2000	2403351	2108	2	868	354	3332	0.0014
2001	2416425	2286	4	955	418	3663	0.0015
2002	2443387	2616	2	939	355	3912	0.0016
2003	2448288	2380	2	850	340	3572	0.0015
2004	2397615	2027	3	614	343	2987	0.0012
2005	2448017	2099	8	487	247	2841	0.0012
2006	2426264	1779	1	524	219	2523	0.0010
2007	2423712	1931	8	326	193	2458	0.0010
2008	2471984	1675	0	217	147	2039	0.0008
2009	2437163	1167	1	179	254	1601	0.0007
2010	2648435	891	1	301	125	1318	0.0005
2011	2515458	980	2	144	132	1258	0.0005
2012	2543279	1208	0	170	147	1525	0.0006
2013	2596993	798	0	137	132	1067	0.0004
2014	2626418	661	0	164	163	988	0.0004
2015	2712630	414	0	73	138	625	0.0002
2016	2744248	345	0	52	137	534	0.0002
2017	2813503	285	0	34	129	448	0.0002
2018	2839205	232	0	11	144	387	0.0001
2019	2854838	267	0	21	147	435	0.0002
2020	3383729	279	0	35	116	430	0.0002
2021	3464231	202	0	23	93	318	0.0001
2022	3279857	146	0	21	103	270	0.0001

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.

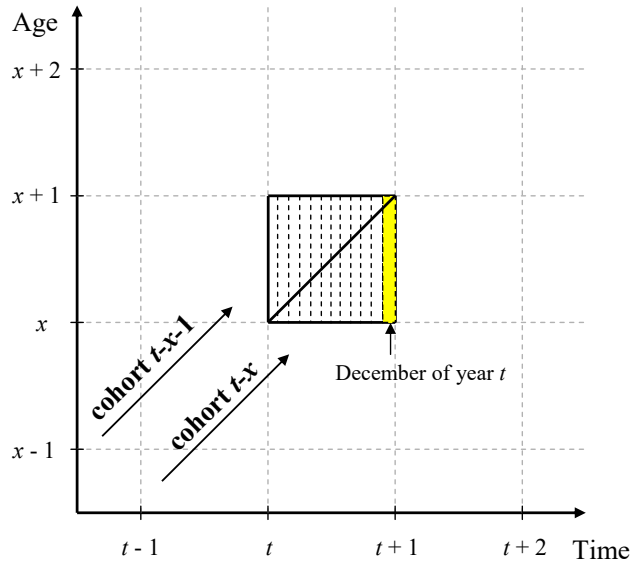
Figure 2. 1x1 Lexis Square by age (in months) and month/year of death



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 3. 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:

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

RefCode 1940
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1940: Part II
Natality and Mortality Data for the United States Tabulated by Place of Residence
Table 9: Deaths by age, sex, race, and population (page 178)

RefCode 1941
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1941: Part II
Natality and Mortality Data for the United States Tabulated by Place of Residence
Table 15: Deaths from selected causes by age, race, and sex: each state, 1941 (page 308)

RefCode 1942
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1942: Part II
Natality and Mortality Data for the United States Tabulated by Place of Residence
Table 21: Deaths from selected causes by age, race, and sex: each state, 1942 (page 219)

RefCode 1943
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1943: Part II
Natality and Mortality Data for the United States Tabulated by Place of Residence
Table 21: Deaths from selected causes by age, race, and sex: each state, 1941 (page 226)

RefCode 1944
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1944: Part II
Natality and Mortality Data for the United States Tabulated by Place of Residence
Table 21: Deaths from selected causes by age, race, and sex: each state, 1944 (page 210)

RefCode 1945
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1945: Part II
Natality and Mortality Data for the United States Tabulated by Place of Residence
Table 19: Deaths from selected causes by age, race, and sex: each state, 1945 (page 181)

RefCode 1946
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1946: Part II
Natality and Mortality Data for the United States Tabulated by Place of Residence
Table 20: Deaths from selected causes by age, race, and sex: United States and each state, 1946 (page 258)

RefCode 1947
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1946: Part II
Natality and Mortality Data for the United States Tabulated by Place of Residence
Table 21: Deaths from selected causes by age, race, and sex: United States and each state, 1947 (page 248)

RefCode 1948
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1948: Part II
Natality and Mortality Data for the United States Tabulated by Place of Residence
Table 19: Deaths from selected causes by age, race, and sex: United States and each state, 1948 (page 248)

RefCode 1949
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1949: Part II
Natality and Mortality Data for the United States Tabulated by Place of Residence
Table 22: Deaths from 64 selected causes by age, race, and sex: United States and each state, 1949 (page 262)

RefCode 1950
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1950: Part III
Mortality Data
Table 56: Deaths from 64 selected causes by age, race, and sex: United States and each state, 1950 (page 256)

RefCode 1951
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1951: Part II
Mortality Data
Table 57: Deaths from 64 selected causes by age, race, and sex: United States and each state, 1950 (page 168)

RefCode 1952
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1952: Part II
Mortality Data

Table 52B: Deaths from 64 selected causes by age, race, and sex: Each state, 1952
(page 174)

RefCode 1953
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1953: Part II
Mortality Data
Table 53B: Deaths from 64 selected causes by age, race, and sex: Each state, 1953
(page 176)

RefCode 1954
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1954: Part II
Mortality Data
Table 53B: Deaths from 64 selected causes by age, race, and sex: Each state, 1954
(page 174)

RefCode 1955
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1955: Part II
Mortality Data
Table 55B: Deaths from 64 selected causes by age, race, and sex: Each state, 1955
(page 180)

RefCode 1956
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1956: Part II
Mortality Data
Table 64B: Deaths from 64 selected causes by age, race, and sex: Each state, 1956
(page 184)

RefCode 1957
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1957: Part II
Mortality Data
Table 66B: Deaths from 64 selected causes by age, race, and sex: Each state, 1957
(page 184)

RefCode 1958
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1958: Part II
Mortality Data
Table 81B: Deaths from 64 selected causes by age, race, and sex: Each state, 1958
(page 322)

RefCode 1
National Center for Health Statistics, United States. Mortality Multiple Cause Restricted Use File, years 1959-2020, Accessed in the Berkeley RDC
Tabulation of deaths by Lexis triangle from individual records by state, single year of age and birth cohort (where available).

RefCode 6
United States. Bureau of the Census. (1942).
Sixteenth census of the United States, 1940, Population
Census of Population: Volume 4. Characteristics by Age, Marital Status, Relationship, Education, and Citizenship.
Parts 2-4
Table 1: Age by single years, by color and sex, for the state, urban and rural: 1940
Washington, D.C.: U.S. Dept. of Commerce, Bureau of the Census.

RefCode 7
A Report of the Seventeenth Decennial Census of the United States
Census of Population: 1950; Volume II – Characteristics of the Populations (1953)
Parts 2-50
Table 15: Age by Color and Sex, for the State, Urban and Rural: 1950 and 1940
Washington, D.C.: U.S. Dept. of Commerce, Bureau of the Census.

RefCode 8
A Report of the Seventeenth Decennial Census of the United States
Census of Population: 1950; Volume II – Characteristics of the Populations (1953)
Parts 2-50
Table 51: Single years of age, by sex, for the State 1950 and 1940, and by color, nativity, and sex, for the State and for cities of 250,000 or more, 1950
Washington, D.C.: U.S. Dept. of Commerce, Bureau of the Census.

RefCode 9
Hybrid series of population data for USA states years 1950 by 1 year age, computed from combined series RefCode 7 and RefCode 8. The primary sources used to construct the hybrid series are:
(Census) Population Data for USA states years 1950, by 1,2,3,4,5,& 10-year age (RefCode 7)
(Census) Population Data for USA states, years 1950 by single year of age (RefCode 8)

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 12)
(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, parts 2-52
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 Summary (Part 1)
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
U.S. Department of Commerce Bureau of the Census

Population and Housing Unit Estimates
Annual Resident Population Estimates for 6 Race Groups (5 Race Alone Groups and Two or More Races) by Age, Sex, and Hispanic Origin for States and the District of Columbia: April 1, 2010 to July 1, 2019; April 1, 2020; and July 1, 2020
(retrieved from
https://www2.census.gov/programs-surveys/popest/datasets/2010-2020/state/asrh/SC-EST2020-ALLDATA6.csv)

RefCode 130
U.S. Department of Commerce Bureau of the Census
Population and Housing Unit Estimates
Annual Estimates of the Civilian Population by Single Year of Age and Sex for the United States and States: April 1, 2020 to July 1, 2023
(retrieved from
https://www2.census.gov/programs-surveys/popest/datasets/2020-2023/state/asrh/)

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 2007-2018
(retrieved from https://wonder.cdc.gov/natality-current.html)

RefCode 2939
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1939: Part II
Natality and Mortality Data for the United States Tabulated by Place of Residence
Table 1: Live births by sex of child, age of mother, race, and parent nativity of white: United States and each State, 1939 (page 23)

RefCode 2940
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1940: Part II
Natality and Mortality Data for the United States Tabulated by Place of Residence
Table 1: Live births by sex of child, age of mother, race, and parent nativity of white: United States and each State, 1940 (page 56)

RefCode 2941
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1941: Part II
Natality and Mortality Data for the United States Tabulated by Place of Residence
Table 1: Live births by sex of child, age of mother, race, and parent nativity of white: United States and each State, 1941 (page 62)

RefCode 2942
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1942: Part II
Natality and Mortality Data for the United States Tabulated by Place of Residence
Table 3: Live births by sex of child, age of mother, race, and parent nativity of white: United States and each State, 1942 (page 68)

RefCode 2943
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1943: Part II
Natality and Mortality Data for the United States Tabulated by Place of Residence
Table 3: Live births by sex of child, age of mother, race, and parent nativity of white: United States and each State, 1943 (page 59)

RefCode 2944
U.S. Department of Commerce Bureau of the Census

Vital Statistics of the United States 1943: Part II
Natality and Mortality Data for the United States Tabulated by Place of Residence
Table 3: Live births by sex of child, age of mother, race, and parent nativity of white: United States and each State, 1944 (page 53)

RefCode 2945
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1945: Part II
Natality and Mortality Data for the United States Tabulated by Place of Residence
Table 3: Live births by sex of child, age of mother, race, and parent nativity of white: United States and each State, 1945 (page 54)

RefCode 2946
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1946: Part II
Natality and Mortality Data for the United States Tabulated by Place of Residence
Table 3: Live births by specified race and sex: United States and each State, 1946 (page 112)

RefCode 2947
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1947: Part II
Natality and Mortality Data for the United States Tabulated by Place of Residence
Table 2: Live births by specified race and sex: United States and each State, 1947 (page 94)

RefCode 2948
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1948: Part II
Natality and Mortality Data for the United States Tabulated by Place of Residence
Table 2: Live births by specified race and sex: United States and each State, 1948 (page 94)

RefCode 2949
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1949: Part II
Natality and Mortality Data for the United States Tabulated by Place of Residence
Table 2: Live births by specified race and sex: United States and each State, 1949 (page 94)

RefCode 2950
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1950: Part II
Marriage, Divorce, Natality, Fetal Mortality and Infant Mortality Data
Table 18: Live births by specified race and sex: United States and each State, 1950 (page 187)

RefCode 2951
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1951: Volume I
Analysis and Summary Tables. Tables for Alaska, Hawaii, Puerto Rico, and Virgin Islands. Marriages, Divorce, Natality, Fetal Mortality and Infant Mortality Data
Table 22: Live births by specified race and sex: United States and each State, 1951 (page 212)

RefCode 2952
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1952: Volume I
Analysis and Summary Tables. Tables for Alaska, Hawaii, Puerto Rico, and Virgin Islands. Marriages, Divorce, Natality, Fetal Mortality and Infant Mortality Data
Table 22: Live births by specified race and sex: United States and each State, 1952 (page 193)

RefCode 2953
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1953: Volume I
Analysis and Summary Tables. Tables for Alaska, Hawaii, Puerto Rico, and Virgin Islands. Marriages, Divorce, Natality, Fetal Mortality and Infant Mortality Data
Table 22: Live births by specified race and sex: United States and each State, 1953 (page 207)

RefCode 2954
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1954: Volume I
Analysis and Summary Tables. Tables for Alaska, Hawaii, Puerto Rico, and Virgin Islands. Marriages, Divorce, Natality, Fetal Mortality and Infant Mortality Data
Table 22: Live births by specified race and sex: United States and each State, 1954 (page 207)

RefCode 2955
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1955: Volume I
Analysis and Summary Tables. Tables for Alaska, Hawaii, Puerto Rico, and Virgin Islands. Marriages, Divorce, Natality, Fetal Mortality and Infant Mortality Data
Table 24: Live births by specified race and sex: United States and each State, 1955 (page 212)

RefCode 2956
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1956: Volume I

Analysis and Summary Tables. Tables for Alaska, Hawaii, Puerto Rico, and Virgin Islands. Marriages, Divorce, Natality, Fetal Mortality and Infant Mortality Data
Table 24: Live births by specified race and sex: United States and each State, 1956 (page 214)

RefCode 2957
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1957: Volume I
Analysis and Summary Tables. Tables for Alaska, Hawaii, Puerto Rico, and Virgin Islands. Marriages, Divorce, Natality, Fetal Mortality and Infant Mortality Data
Table 29: Live births by specified race and sex: United States and each State, 1957 (page 230)

RefCode 2958
U.S. Department of Commerce Bureau of the Census
Vital Statistics of the United States 1958: Volume I
Analysis and Summary Tables. Tables for Alaska, Hawaii, Puerto Rico, and Virgin Islands. Marriages, Divorce, Natality, Fetal Mortality and Infant Mortality Data
Table 30: Live births by specified race and sex: United States and each State, 1958 (page 246)