

Measuring All-Cause Mortality with the Census Numident File: Supplemental Materials

Keith Finlay, PhD, and Katie R. Genadek, PhD

Variables in the Census Numident

The Census Bureau obtains quarterly updates with transactions from the Social Security Administration (SSA) Numident File. The Census Bureau processes these transaction-level quarterly updates to create the Census Numident, a person-level research file that includes the history of individual-level interactions with the SSA Numident. The Census Numident is a cumulative file, and the most recent vintage of the Census Numident is the largest and most up-to-date version.

The Census Bureau makes the Census Numident available to researchers through the restricted Federal Statistical Research Data Center (FSRDC) research environment. Names and Social Security Numbers (SSNs) are not included in the data available for researchers, but the Census Bureau does assign a unique, anonymous identifier, called a Protected Identification Key (PIK) to all individuals in the Numident based on the SSN. The detailed date of death, date of birth, and date of SSA interactions are included in the Census Numident. Variables available in the Census Numident of interest for mortality research are listed in Table A with brief descriptions of the variables.

Table A. Selected Variables in the Census Numident

Birth Date Century	Birthdate
Birth Date Year	Birthdate
Birth Date Month	Birthdate
Birth Date Day	Birthdate
Death Date Century	Date of death
Death Date Year	Date of death
Death Date Month	Date of death
Death Date Day	Date of death
Cycle Date for Date of Death	Date death is posted to Numident
Place of Birth, City	City of birth
Place of Birth, State/Country	State or country of birth
Sex	Most recent male or female indicator
Original Race	First race reported
Best Race	Most recent race with additional edits
Protected Identification Key	Census Bureau unique anonymous identifier

Notes: Variables from the Census Numident used in the analyses presented in this paper or referenced in this paper.

Historical Census Numident and Vital Statistics Comparison

Table B shows the yearly death counts from the Census Numident and the Centers of Disease Control and Prevention (CDC) mortality data back to 1940.¹ The Census Numident counts are rounded according to Census Bureau Disclosure Review Board guidelines. The Census Numident death counts are calculated from the most recent Census Numident vintage (2020Q4). Prior to the SSA's death improvement initiative, the Census Numident contained incomplete death information. Death information was also incomplete due to data transfer errors between SSA and the Census Bureau. We recommend that researchers use the most recent vintage of the Census Numident for death information.

Table B. Yearly Mortality Counts from the Census Numident and CDC Vital Statistics Data

Year	Census Numident	CDC Data	Numident/CDC	Year	Census Numident	CDC Data	Numident/CDC
2019	2,894,000	2,855,000	1.014	1979	1,810,000	1,913,841	0.946
2018	2,880,000	2,839,205	1.014	1978	1,810,000	1,927,788	0.939
2017	2,858,000	2,813,503	1.016	1977	1,776,000	1,899,597	0.935
2016	2,789,000	2,744,248	1.016	1976	1,769,000	1,909,440	0.926
2015	2,754,000	2,712,630	1.015	1975	1,744,000	1,892,879	0.921
2014	2,670,000	2,626,418	1.017	1974	1,775,000	1,934,388	0.918
2013	2,639,000	2,596,993	1.016	1973	1,779,000	1,973,003	0.902
2012	2,587,000	2,543,279	1.017	1972	1,674,000	1,963,944	0.852
2011	2,557,000	2,515,458	1.017	1971	1,505,000	1,927,542	0.781
2010	2,511,000	2,468,435	1.017	1970	1,495,000	1,921,031	0.778
2009	2,472,000	2,437,163	1.014	1969	1,479,000	1,921,990	0.770
2008	2,506,000	2,471,984	1.014	1968	1,481,000	1,930,082	0.767
2007	2,453,000	2,423,712	1.012	1967	1,394,000	1,851,323	0.753
2006	2,451,000	2,426,264	1.010	1966	1,310,000	1,863,149	0.703
2005	2,476,000	2,448,017	1.011	1965	1,101,000	1,823,136	0.604
2004	2,423,000	2,397,615	1.011	1964	1,027,000	1,798,051	0.571
2003	2,467,000	2,448,288	1.008	1963	1,002,000	1,813,549	0.553
2002	2,461,000	2,443,387	1.007	1962	725,000	1,756,720	0.413
2001	2,434,000	2,416,425	1.007	1961	570,000	1,701,522	0.335
2000	2,417,000	2,403,351	1.006	1960	555,000	1,711,982	0.324
1999	2,405,000	2,391,399	1.006	1959	514,000	1,656,814	0.310
1998	2,344,000	2,337,256	1.003	1958	497,000	1,647,886	0.302
1997	2,317,000	2,314,245	1.001	1957	483,000	1,633,128	0.296
1996	2,312,000	2,314,690	0.999	1956	485,000	1,564,476	0.310
1995	2,300,000	2,312,132	0.995	1955	416,000	1,528,717	0.272
1994	2,256,000	2,278,994	0.990	1954	398,000	1,481,091	0.269

1993	2,238,000	2,268,553	0.987	1953	390,000	1,517,541	0.257
1992	2,138,000	2,175,613	0.983	1952	364,000	1,496,838	0.243
1991	2,120,000	2,169,518	0.977	1951	334,000	1,482,099	0.225
1990	2,096,000	2,148,463	0.976	1950	295,000	1,452,454	0.203
1989	2,087,000	2,150,466	0.970	1949	270,000	1,443,607	0.187
1988	2,089,000	2,167,999	0.964	1948	265,000	1,444,337	0.183
1987	2,047,000	2,123,323	0.964	1947	258,000	1,445,370	0.179
1986	2,006,000	2,105,361	0.953	1946	240,000	1,395,617	0.172
1985	1,982,000	2,086,440	0.950	1945	279,000	1,401,719	0.199
1984	1,934,000	2,039,369	0.948	1944	277,000	1,411,338	0.196
1983	1,903,000	2,019,201	0.942	1943	188,000	1,459,544	0.129
1982	1,853,000	1,974,797	0.938	1942	132,000	1,385,187	0.095
1981	1,868,000	1,977,981	0.944	1941	133,000	1,397,642	0.095
1980	1,888,000	1,989,841	0.949	1940	122,000	1,417,269	0.086

Notes: Census Numident calculations from vintage 2020Q4. All Census Numident results were approved for release by the U.S. Census Bureau, authorization numbers CBDRB-FY21-ERD002-004, CBDRB-FY21-ERD002-009. The Census Numident counts are rounded per Census Bureau Disclosure Review Board guidelines. The CDC counts were obtained from the CDC Wonder database.

Universe Comparison of the Census Numident and CDC Data

The universes for the mortality information in the Census Numident and CDC vital statistics are slightly different. Table C shows how the universes overlap.

Table C: Universe comparison for Census Numident and CDC Vital Statistics mortality records

In Census Numident only	In Census Numident and CDC	In CDC only
Deaths of SSN holders outside of U.S. states reported to SSA	Deaths of SSN holders in the U.S. states	Deaths of individuals in the U.S. states who do not have SSNs

We do not know precisely which universe is larger because we do not have exact estimates of the number of individuals in the “Census Numident only” group or the “CDC only” group. Though SSNs are not limited to U.S. citizens, almost all U.S. citizens have been issued an SSN and are in the Numident. The U.S. Department of State’s Bureau of Consular Affairs estimated that nine million U.S. citizens were living abroad in 2016 while the Federal Voting Assistance Program estimated that 4.8 million U.S. citizens lived overseas in 2018.^{2,3} Research at the Census Bureau has estimated that more than four million U.S. residents in 2010 had an Individual Taxpayer Identification Number (ITIN), and thus would reside in the U.S. without an SSN.⁴ Seventeen million respondents in the 2010 Decennial Census could not be linked to the Numident using the information in the 2010 Census.² While some of the 17 million non-links are due to inaccurate identifying information in the Census, some are likely individuals living in the U.S. who do not have an ITIN or SSN. While it is not clear from these estimates which group is expected to be larger, the mortality counts from the Census Numident are consistently greater than the CDC data, peaking at 1.7% more deaths. This suggests that there are more deaths of SSN holders outside the U.S. than deaths occurring in the U.S. to non-SSN holders.

Death Counts by Sex and Race in the Census Numident

Sex is included on applications for SSNs, and individuals can apply to change their sex recorded by SSA. The sex indicator included in the Census Numident has the most recent sex reported for that person. Table D shows sex and age breakdowns for all deaths recorded in the Census Numident from 2000 through 2018, including the percent of deaths occurring by sex within each age group. Columns 5 and 6 show that a very small number of observations are missing sex in

the Census Numident. Missing ages at death were primarily from observations missing the day of the month when the death occurred (97%), with the remaining missing ages caused by incomplete birthdates.

Table D. Mortality Counts by Age and Sex, 2000-2018

Age	Female	% Female	Male	% Male	Missing Sex	% Missing
0-19 years	248,000	38.2%	401,000	61.8%	0	0
20-29 years	203,000	26.9%	551,000	73.1%	0	0
30-39 years	361,000	34.5%	686,000	65.5%	0	0
40-49 years	854,000	37.9%	1,398,000	62.1%	0	0
50-59 years	1,779,000	38.4%	2,855,000	61.6%	30	0.001%
60-69 years	2,925,000	41.0%	4,215,000	59.0%	80	0.001%
70-79 years	4,860,000	46.1%	5,678,000	53.9%	200	0.002%
80-89 years	7,798,000	55.3%	6,315,000	44.7%	500	0.004%
90+ years	5,272,000	69.3%	2,340,000	30.7%	900	0.012%
Missing	25,500	39.4%	39,000	60.3%	150	0.232%
Total	24,325,500	49.8%	24,478,000	50.15%	1860	0.004%

Notes: Census Numident calculations from vintage 2020Q4. Missing Age indicates that age on date of death could not be calculated because the observation was missing the day of the month the death occurred. All Census Numident results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-009. The Census Numident counts are rounded per Census Bureau Disclosure Review Board guidelines.

In addition to having information about sex, many observations in the Census Numident have race/ethnicity indicated. SSA has collected race in SSN applications since the inception of SSNs. The first race options were “White,” “Black,” and “Other.” Additional race/ethnicity options for “Hispanic,” “Asian or Pacific Islander,” and “American Indian or Alaska Native” were added in 1980.⁵ Many Numident observations are missing race because early race codes were not all retained during the computerization of the Numident.⁵ Moreover, starting in 1989, SSNs were issued at birth based on birth certificates sent to SSA by states, and SSA cannot use the race information from the birth certificates.⁵ Like sex, individuals can update their race information when interacting with SSA (for a name change for example), but most observations of those born since 1989 do not have race information. As shown in Table D, roughly 4% of deaths in recent years do not have race indicated in the Census Numident.

The “original race” variable in the Census Numident is the first race reported in the SSA Numident for the given observation. The “best race” variable in the Census Numident is the race code from the most recent SSA Numident transaction, excluding those with unknown or missing race indicators. In addition, if the most recent race was “White” or “Black” or “Other” and a race from an earlier record was “Asian or Pacific Islander” or “American Indian or Alaska Native” or “Hispanic,” then the first race recorded was assigned. If the only race code available was “Unknown” or blank, then these values were retained. Table E shows the breakdown of deaths by year from 2016 to 2020 by race/ethnicity as reported in the “best race” variable in the Census Numident. Death counts are not completely reported for 2020. Researchers could also obtain race/ethnicity information by linking these data to the decennial censuses using the Protected Identification Key (PIK). This would provide race information where the Numident is missing race/ethnicity, and the decennial censuses also provide more detailed race categories than those collected by SSA.

Table E. Mortality Counts by Race and Year of Death, 2016-2020

	2016	%	2017	%	2018	%	2019	%	2020	%
White	2,140,000	76.8%	2,188,000	76.5%	2,195,000	76.2%	2,194,000	75.8%	2,193,000	73.5%
Black	311,000	11.2%	319,000	11.2%	326,000	11.3%	329,000	11.4%	371,000	12.4%
Other	37,500	1.3%	39,500	1.4%	40,500	1.4%	40,500	1.4%	43,500	1.5%
Asian or Pacific Islander	50,000	1.8%	53,000	1.9%	54,500	1.9%	56,000	1.9%	61,000	2.0%
Hispanic	126,000	4.5%	133,000	4.7%	136,000	4.7%	141,000	4.9%	171,000	5.7%
American Indian or Alaska Native	13,000	0.5%	13,500	0.5%	14,000	0.5%	14,000	0.5%	16,500	0.6%
Missing Race	110,200	4.0%	113,200	4.0%	114,150	4.0%	119,100	4.1%	127,040	4.3%

Notes: Census Numident calculations from vintage 2020Q4. The race variable used here is the "best race", which is an edited version of the most recent Numident recording of race. All Census Numident results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-009. The Census Numident counts are rounded per Census Bureau Disclosure Review Board guidelines.

Death Reporting Delays in the Census Numident

Deaths are reported to SSA by family members, family representatives, funeral directors, state agencies, and federal agencies. SSA records the death information in the SSA Numident, which now serves as its single system for death records. As shown in the list of variables in Table A, the Numident has a cycle date of death that identifies when the date of death was updated. Comparing this cycle date of death to the actual date of death, Table F shows the distribution of the death recordation delay in days, by year of death. The promptness of death recording has improved substantially since 2000. From 2000 to 2020, the median number of days between death and death recordation fell from 13 to 6.

Table F. Distribution of Days between Date of Death and Date of Numident Death Posting

Year of Death	Percentile				
	5th	25th	50th	75th	95th
2000	4	8	13	38	208
2001	4	7	12	24	139
2002	4	7	11	21	127
2003	4	7	11	21	122
2004	4	8	13	26	135
2005	4	7	12	23	200
2006	3	7	12	25	166
2007	3	7	12	28	120
2008	3	7	13	30	132
2009	3	6	11	24	102
2010	3	6	11	23	126
2011	2	5	9	19	85
2012	2	5	9	18	74
2013	2	4	8	16	69
2014	2	5	8	16	71
2015	2	5	8	18	96
2016	2	4	8	27	135
2017	2	4	8	19	87
2018	2	4	7	15	72
2019	2	4	7	14	58
2020	2	4	6	12	46

Notes: Census Numident calculations from vintage 2020Q4. All Census Numident results were approved for release by the U.S. Census Bureau, authorization number CBDRB-FY21-ERD002-009.

Death Data Related to the Census Numident

Death Master File

The Death Master File (DMF) is a product produced by SSA and distributed by the National Technical Information Service (NTIS), a federal agency within the Department of Commerce. It includes a subset of deaths from the Numident that SSA is allowed to share based on the source of the death information; death information obtained from state death records is not included.⁶ NTIS shares these data with subscribing companies so they can confirm the accuracy of deaths for their businesses purposes, including the prevention of fraud. Researchers have also used the DMF file to study mortality. DMF coverage has always been incomplete, and it has increasingly deteriorated since a reinterpretation of privacy statutes in 2011.⁷

The Census Bureau obtains the DMF monthly from NTIS because it can provide slightly different death information than is shared in the SSA Numident, which was especially useful before the SSA undertook the improvement of its death records. The Census Bureau continues to obtain the DMF because deaths often appear in the monthly DMF updates before we receive the quarterly updates from SSA. The Census Numident retains the date of death from both sources within the file (date of death variables from a “num_” prefix for SSA Numident of “dmf_” prefix for DMF), and includes a primary date of death which combines data from both sources. The primary death information uses the entry with the most complete death information, and if both are complete then the death information with the most recent cycle date (or date of change) is included.

Medicare Data

Data from the Centers for Medicare & Medicaid Services (CMS) are often used for research and include death information for deceased Medicare recipients. CMS generally receives death information from SSA, and when Medicare enrollees die, families are instructed to tell SSA directly.⁸ The death information in the Medicare data should be the same as in the SSA and Census Numident files, but we have not confirmed this.

Related Public Death Data for Research

The Social Security Death Index (SSDI) also contains death information from the DMF, but it is made available by genealogy groups and companies. Most of the companies provide it in a searchable format on their website. In addition to the DMF, there is a public SSA Numident file available through the National Archives and Records Administration (NARA). Like the DMF, this file also contains a subset of the total deaths in the SSA Numident and researchers can obtain a copy of it from NARA. The most recent version of these data include deaths between 1936 and 2007.⁹

¹ Deaths for Selected Causes by 10-Year Age Groups, Race, and Sex: Death Registration States, 1900-32, and United States, 1933-98. <https://www.cdc.gov/nchs/nvss/mortality/hist290a.htm> Accessed October 2020.

² Consular Affairs by the Numbers.

https://web.archive.org/web/20160616233331/https://travel.state.gov/content/dam/travel/CA_By_the_Numbers.pdf Accessed October 2020.

³ Overseas Citizen Population Analysis Report. <https://www.fvap.gov/info/reports-surveys/overseas-citizen-population-analysis> Accessed January 2021.

⁴ Brown, J. D., Heggeness, M. L., Dorinski, S. M., Warren, L., & Yi, M. Understanding the quality of alternative citizenship data sources for the 2020 census. <https://www2.census.gov/ces/wp/2018/CES-WP-18-38R.pdf> Published 2018. Accessed October 2022.

⁵ Scott S.G. Identifying the race and ethnicity of SSI recipients, *Soc Sec Bull.* 1999;62(4).

⁶ SSA's death information. https://www.ssa.gov/dataexchange/request_dmf.html Accessed January 2021.

⁷ Navar AM, Peterson ED, Steen DL, et al. Evaluation of mortality data from the Social Security Administration Death Master File for clinical research. *JAMA Cardiology.* 2019;4(4):375-379.

⁸ Report a death. <https://www.medicare.gov/sign-up-change-plans/report-a-death> Accessed January 2021.

⁹ Numerical Identification Files (NUMIDENT), 1936-2007: <https://catalog.archives.gov/id/12004494> Accessed October 2020.