ADEP WORKING PAPER SERIES

Measuring All-Cause Mortality with the Census Numident File

Keith Finlay
U.S. Census Bureau

Katie R. Genadek
U.S. Census Bureau

Measuring All-Cause Mortality with the Census Numident File
Keith Finlay, U.S. Census Bureau, email address
Katie R. Genadek, U.S. Census Bureau, email address
ADEP Working Paper 2021-03
July 2021

Abstract

Objectives. To assess the quality of population-level US mortality data in the US Census Bureau Numerical Identification file (Numident) and describe the details of the mortality information as well as the novel person-level linkages available when using the Census Numident.

Methods. We compared all-cause mortality in the Census Numident to published vital statistics from the Centers for Disease Control and Prevention. We provide detailed information on the linkage of the Census Numident to other Census Bureau survey, administrative, and economic data.

Results. Death counts in the Census Numident are similar to those from published mortality vital statistics. Yearly comparisons show that the Census Numident captures more deaths since 1997, and coverage is slightly lower going back in time. Weekly estimates show similar trends from both data sets.

Conclusions. The Census Numident is a high-quality and timely source of data to study all-cause mortality. The Census Bureau makes available a vast and rich set of restricted-use, individual-level data linked to the Census Numident for researchers to use.

Public Health Implications. The Census Numident linked to data available from the Census Bureau provides infrastructure for doing evidence-based public health policy research on mortality. (*Am J Public Health.* 2021;111(S2):e1–e8.
https://doi.org/10.2105/AJPH.2021.306217)

Mortality is a critical outcome in public health surveillance. The COVID-19 pandemic has demonstrated the importance of placing individual death events within their social, economic, and geographic contexts. For example, to identify if Black and Hispanic individuals are overrepresented in COVID-19 mortality rates, high-quality race and ethnicity data linked to mortality records are necessary. To identify why Black and Hispanic individuals are overrepresented, we may need a much richer set of data, potentially going back in time. To study how frontline health workers are impacted by COVID-19 mortality, death records must be linked with occupation or employer data. These kinds of person-level data linkages are often made by public health researchers in smaller settings, but doing so at the population level and in a timely way during a public health emergency is unprecedented.

The Census Bureau's Data Linkage Infrastructure is an ecosystem of survey and administrative records that are linked at the person, address, and business levels. The files are de-identified and made available anonymously to researchers in a restricted-access setting. This infrastructure is an excellent environment to study mortality because it holds full-population death data from the Social Security Administration's Numerical Identification file (SSA Numident). While the SSA Numident has historically suffered from underreporting of deaths and incorrect death dates, the Social Security Administration's (SSA's) efforts to improve death monitoring have greatly enhanced the mortality information in the SSA Numident file.[1–3] Moreover, the Census Bureau has actively analyzed, curated, and documented the death information in the file, working with SSA to disseminate high-quality and complete death data through the Census Bureau's Numerical Identification file (Census Numident).

The linkage of Census Bureau surveys, administrative data, business data, and death data allows for rich analyses of the relationships between all-cause mortality and demographic

characteristics, socioeconomic factors, educational attainment, family structure, residential location, migration, program participation, and early life conditions. While research use of SSA death data is not new,[4] the research possibilities for measuring relationships between social and economic determinants of all-cause mortality with the data available through the Census Bureau are vast and generally underutilized.[5,6] These data are currently accessible to researchers working on approved projects within the Federal Statistical Research Data Center (FSRDC) network.

In this article, we introduce the mortality data in the Census Numident and explain the origin and creation of the data file. To assess the quality of the data, we compared mortality estimates from the Census Numident with the primary population-level published statistics for the United States collected from states by the National Center for Health Statistics and published by the Centers for Disease Control and Prevention (CDC). The availability and linkage of these mortality data to other data held at the Census Bureau for research is also described, as well as how researchers can access these data. We show that the Census Numident is an excellent source to study all-cause mortality, especially for analyzing social determinants, understanding neighborhood context, making evidence-based decisions, and even studying pandemics.

SOCIAL SECURITY ADMINISTRATION AND CENSUS NUMIDENT FILES

SSA uses the SSA Numident to maintain records of Social Security Number (SSN) holders. Although SSNs were created and issued starting in 1936, electronic tracking of SSN information in the Numident began in 1972. The Numident contains all interactions individuals have with SSA related to SSNs, including information on SSN applications, claim records, death information, and requested changes to SSN information. There are now more than 1 billion

transactions within the SSA Numident for approximately 518 million living and deceased SSN holders in the SSA Numident.

The Census Bureau obtains SSA Numident data from SSA to improve Census Bureau survey and decennial census data, perform record linkage, and conduct research and statistical projects. To facilitate the use of the SSA Numident data, the Census Bureau processes quarterly updates from SSA transaction records to create a person-level research file that includes the history of individual-level interactions with the SSA Numident. The Census Bureau calls this processed file the Census Numident. Like the SSA Numident, the Census Numident is a cumulative file. The most recent vintage of the Census Numident is the largest and most up-to-date version, and researchers should use the newest vintage for mortality research.

The Census Bureau assigns a unique, anonymous identifier, called a Protected Identification Key (PIK) to all individuals in the Numident based solely on the SSN. All names and SSNs are removed from the Census Numident file, and the resulting data file, with the PIK added, is then made available to Census Bureau staff and external researchers for approved Census Bureau production and research projects. PIKs are used to link records at the person level over time and across survey and administrative records.

The scope of information in an individual's Census Numident record varies based on when the individual received an SSN, and if the individual has interacted with SSA, such as for a name change. In general, most records include date of birth, place of birth, sex, race/ethnicity, date of SSN application, dates and types of SSA interactions, and the reported date of death (month, day, and year) if deceased (the variables are listed in Table A in the Supplementary Materials, available with the online version of this article at http://www.ajph.org).

DEATH INFORMATION IN THE NUMIDENT

SSA administers the US Old-Age, Survivors, and Disability Insurance program, often referred to as "Social Security." The death information included in the SSA Numident is collected by SSA for the purposes of administering the Old-Age, Survivors, and Disability Insurance program, and the way this information is collected and managed has changed over time. Death information is obtained through several sources including first-party reports of death from family members and representatives and verified third-party reports from friends, state government offices, the Centers for Medicare and Medicaid Services, the Department of Veterans Affairs, and the Internal Revenue Service.

SSA began maintaining death information using electronic methods in 1962[7] and integrated those records into the Numident when it was created.[8] Information on deaths before 1962 is often incomplete or missing.[8] Previous research has shown that because SSA was primarily focused on deaths of claimants, the Numident had greater death coverage for deaths occurring at older ages than deaths occurring at younger ages.[9] However, since 2005, SSA has improved its methods for monitoring deaths by using a new system for electronically registering deaths, the Death Information Processing System. In 2019, SSA undertook the Death Data Improvement Initiative following a report from the Government Accountably Office in 2013 about errors in the death data[1] and 2 reports from the Office of the Inspector General about missing and incorrect deaths in the Numident.[2,3] This resulted in more records with death information and updates to death information for deaths going back to 1960. The SSA Numident is now SSA's single system of record for death information.[10]

The data included in the SSA Numident and, thus, the Census Numident are limited to SSN holders, and their deaths can occur anywhere, including outside of the United States. The SSA

Numident also contains much more complete death records than the oft-used public Death Master File,[11–14] which has always been a subset of the SSA Numident death records (see "Death Data Related to the Census Numident" in the Supplementary Materials for more information). The Death Master File has deteriorated in coverage since a reinterpretation of privacy statutes in 2011, which limited the inclusion of state records.[15–17] While SSA death reports are considered to be measured with some error,[18] they have been used for research on all-cause mortality even before these recent data quality improvements.[19–22]

METHODS

We measured all-cause mortality using the Census Numident by simply counting the deaths based on the recorded year of death. For the primary analyses, we used the Census Numident date of death, which is the most recent death information for a person from the SSA Numident. We benchmarked all-cause mortality estimates using vital statistics mortality estimates from the CDC, the primary source of mortality data for the United States. The CDC data are compiled from death certificates from state vital statistics offices that have been provided to the National Center for Health Statistics. We use the Compressed Mortality Files from CDC WONDER for the years 1980 to 2016. For 2017 to 2018, we used the CDC WONDER public data tool, and for 2019 to 2020 we used counts from the provisional tables.[23] We also compared the Census Numident death counts before 1980 to published mortality estimated from the CDC Compressed Mortality data and the National Vital Statistics System historical tables.[24] The CDC data, including the restricted-use National Death Index, provide date of death, age, sex, race, cause of death, and place of death.
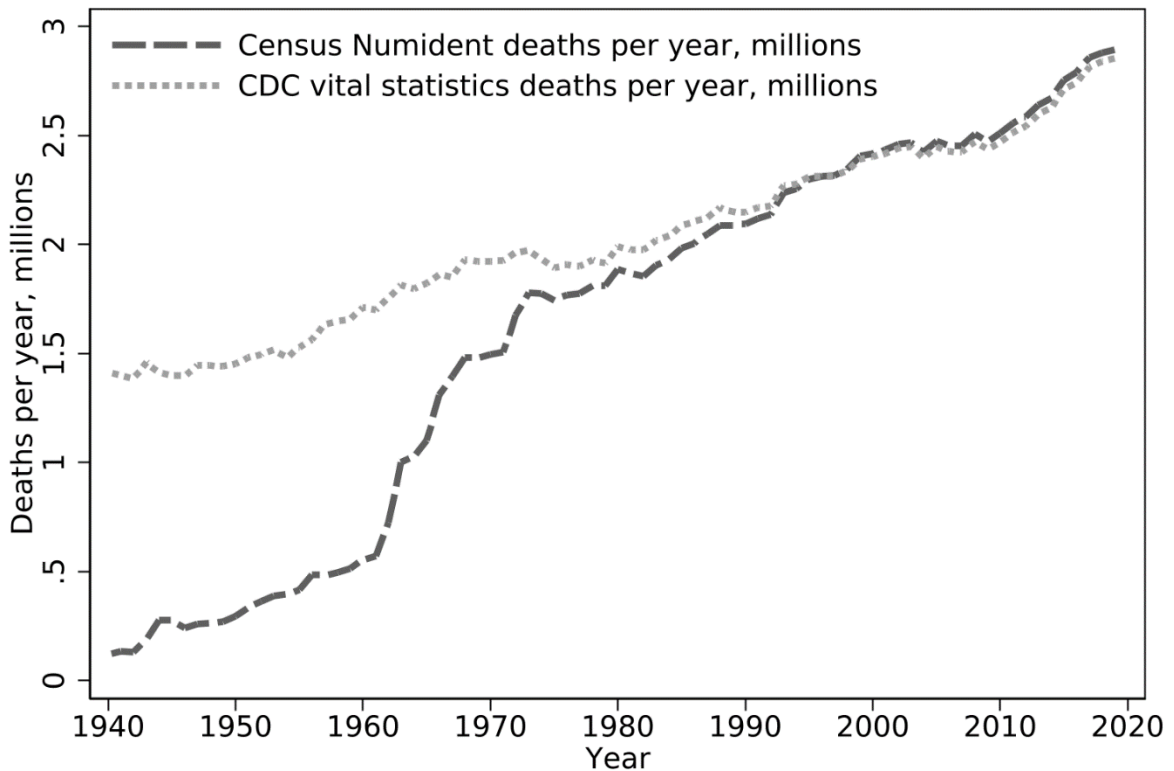
The CDC data have a different universe than the Census Numident, but they provide a useful comparison to assess the coverage of the Numident file for measuring mortality. The CDC-published estimates only include deaths occurring in US states and are not limited to SSN holders, whereas the Census Numident includes deaths of SSN holders dying abroad and in US territories but does not include deaths occurring in the United States for those without an SSN. Thus, almost all the deaths in the United States will appear in each of the files. The difference in counts between the files depends on both error in the data-generating process and the difference between the number of deaths occurring to SSN holders outside of US states and the number of deaths occurring to individuals within the United States without an SSN.

We further benchmarked the Census Numident to the CDC data by performing comparisons of weekly death estimates, age of death, and place of death. Although the Numident does not contain place of death, we were able to proxy for location of death by identifying almost all individuals' most recent residential locations from the Census Bureau's Master Address File - Auxiliary Reference File (MAF-ARF), which is a data file created using information from population-level censuses and administrative records, including annual Medicare enrollment and individual tax filings. Thus, we can approximate death counts by state using records that were assigned a location in the MAF-ARF. Additional analyses of the death counts by race and sex are shown in the Supplementary Materials (Tables D and E). By comparing the Census Numident to the CDC data in more detail, we can assess data quality and identify any shortcomings of mortality records in the Census Numident.

RESULTS

Figure 1 shows yearly mortality estimates from the Census Numident and the CDC from 1940 to the present (the underlying estimates can be found in Table B in the Supplemental Materials). The yearly comparison shows that the Census Numident has more deaths than the CDC estimates in each year from 1997 forward. Going back in time, the Census Numident coverage declines slightly each year until 1985 when the Census Numident contains 95% of the death counts from the CDC and remains around there until 1980. Before 1980, the Census Numident death counts drop steadily until 1967, when they are 75% of the CDC counts. The coverage then drops precipitously, reaching down to 32% of the CDC count in 1960 and continues to decline steadily to under 10% in 1940. This large decrease in deaths captured by the Census Numident occurs directly before the creation of the electronic system for capturing deaths.

FIGURE 1—Coverage of Deaths per Year in the Census Numident vs Centers for Disease Control and Prevention (CDC) Vital Statistics: United States, 1940–2019

The yearly mortality counts suggest that the Census Numident is similar to the vital statistics from CDC on average. Likewise, the weekly mortality estimates from the Census Numident and CDC from 2019 and 2020 are nearly identical. Figure 2 shows the weekly death counts from the Census Numident and the CDC from January 2019 through August 2020. The weekly estimates are nearly the same over the period, both showing the large spike from the COVID-19 pandemic. The weekly coverage of the Census Numident, when compared with the CDC data, ranges from 97% in the final week to 102% in the first week of 2019. These results show that the death data in the Census Numident are accurate and timely. Further analyses show that, in 2019, most

deaths appear in the Census Numident within 7 days after the death occurs. This has improved

since 2000, when more than half of the deaths were added to the SSA Numident 13 or more days

following the death (detailed statistics about reporting delay can be found in Table F in the

Supplemental Materials).

FIGURE 2—Coverage of Deaths per Week in the Census Numident vs Centers for Disease
Control and Prevention (CDC) Vital Statistics, January 2019–September 2020
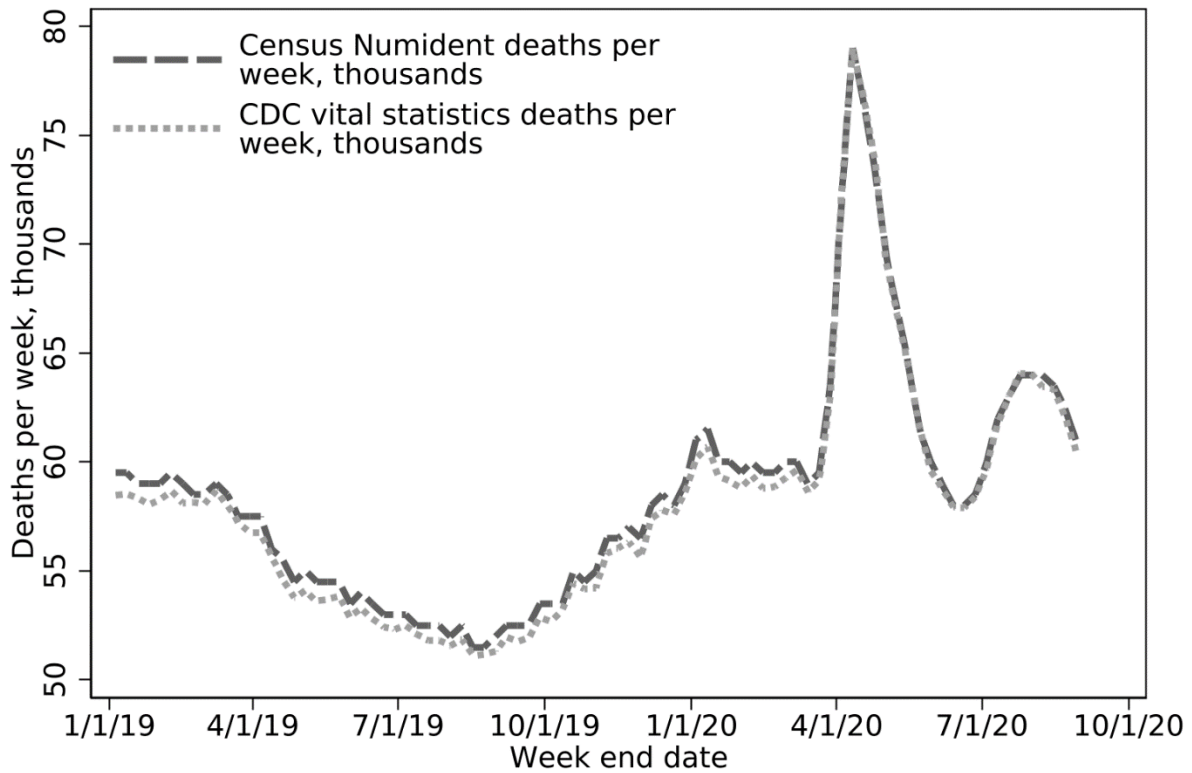

Source Census Numident calculations from the vintage 2020Q4. All Census Numident results were approved for release by the
US Census Bureau, authorization number CBDRB-FY21-ERD002-009. CDC counts were obtained from the CDC WONDER
database.

Table 1 shows death counts that occurred between 2000 and 2018 by age at the time of death

from both the Census Numident and the CDC. Columns 2 and 3 show the percentage of the total

deaths occurring for individuals at those ages within each of the files and Column 4 shows the

ratio of Numident to CDC deaths. The Census Numident has a larger share of deaths for people

of unknown ages (death records missing birth dates or exact date of death) and fewer deaths for

individuals aged younger than 1 year than the CDC data. The infant deaths captured by CDC and

not the Census Numident are likely from live births that do not result in SSN issuance because of

deaths occurring shortly after births.[25] The share of deaths in each of the age categories is

remarkably similar across the files in the other age categories. For ages older than 75 years, the

Census Numident has a slightly larger share of total deaths falling between ages 75 and 84 years

and ages 85 years and older.

Table 1. Mortality Counts by Age at Death, 2000-2018

Age	Census Numident	Numident % of Total	CDC Vital Statistics	CDC % of Total	Numident /CDC
< 1 year	231,000	0.47%	489,447	1.02%	0.47
1-4 years	82,000	0.17%	84,362	0.18%	0.97
5-14 years	114,000	0.23%	114,051	0.24%	1.00
15-24 years	583,000	1.20%	600,236	1.25%	0.97
25-34 years	842,000	1.73%	869,313	1.80%	0.97
35-44 years	1,483,000	3.04%	1,506,617	3.13%	0.98
45-54 years	3,358,000	6.88%	3,365,173	6.99%	1.00
55-64 years	5,899,000	12.09%	5,830,574	12.10%	1.01
65-74 years	8,586,000	17.59%	8,351,622	17.34%	1.03
75-84 years	12,830,000	26.29%	12,514,521	25.98%	1.03
85+ years	14,730,000	30.18%	14,446,308	29.99%	1.02
Missing	65,000	0.13%	4,151	0.01%	15.66
Total	48,803,000		48,176,375		1.01

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 numbers CBDRB-FY21-ERD002-004, CBDRB-FY21-ERD002-009. The Census Numident counts are rounded per Census Bureau Disclosure Review Board guidelines. The CDC vital statistics counts were obtained from the CDC Wonder database.

We were able to use the MAF-ARF to obtain the deceased's state of residence during the

year of death (or location during most recent year in the MAF-ARF before death) for 92% of the

total deaths in the Census Numident from 2010 to 2018. While some of the individuals not

linked to the MAF-ARF because they lived abroad, most were likely not linked because of

incompleteness in the MAF-ARF (though we cannot distinguish between these two groups). The state-level comparison of the Census Numident to the CDC estimates presented in Table 2 support this as the Census Numident death counts are more than 90% of the CDC counts in most states, but there are 11 states with coverage between 85% and 90%, and Hawaii has the lowest coverage rate at 72% of the CDC count. These states in particular suggest that this undercount is not comprised fully of deaths from residents without SSNs, and while it is possible to estimate location of death, these data are incomplete for studying state-level deaths.

Table 2. State-Level Mortality Counts, 2010-2018

States	Census Numident	CDC Vital Statistics	Numident /CDC
Hawaii	69,000	95,857	0.720
Montana	73,500	85,849	0.856
West Virginia	173,000	201,324	0.859
Vermont	44,500	51,428	0.865
New Mexico	137,000	157,207	0.871
Mississippi	243,000	277,152	0.877
Kentucky	363,000	408,180	0.889
Louisiana	348,000	390,890	0.890
Arkansas	248,000	277,887	0.892
Arizona	424,000	474,748	0.893
Idaho	103,000	115,165	0.894
Oregon	278,000	309,821	0.897
Oklahoma	312,000	347,505	0.898
North Dakota	50,000	55,689	0.898
North Carolina	699,000	775,916	0.901
South Carolina	372,000	412,329	0.902
Utah	136,000	150,439	0.904
Maine	112,000	123,730	0.905
Alabama	415,000	458,389	0.905
Texas	1,512,000	1,654,386	0.914
Washington	434,000	474,541	0.915
Georgia	639,000	697,003	0.917
Missouri	486,000	527,724	0.921
New York	1,258,000	1,365,987	0.921
Tennessee	540,000	585,743	0.922
Nevada	185,000	200,165	0.924

Alaska	34,500	37,288	0.925
Ohio	970,000	1,046,075	0.927
Massachusetts	466,000	501,959	0.928
Indiana	514,000	553,409	0.929
Kansas	217,000	233,479	0.929
Virginia	537,000	577,702	0.930
Rhode Island	82,000	88,214	0.930
Colorado	295,000	316,578	0.932
California	2,115,000	2,269,249	0.932
Nebraska	135,000	144,777	0.932
Pennsylvania	1,096,000	1,173,367	0.934
Illinois	885,000	946,599	0.935
New Hampshire	97,000	103,632	0.936
Iowa	246,000	262,491	0.937
Wisconsin	427,000	453,863	0.941
New Jersey	611,000	649,343	0.941
South Dakota	64,000	67,896	0.943
Wyoming	39,500	41,826	0.944
Minnesota	356,000	376,234	0.946
Michigan	797,000	841,628	0.947
District of Columbia	41,000	43,234	0.948
Delaware	72,000	75,720	0.951
Maryland	400,000	419,668	0.953
Connecticut	259,000	270,646	0.957
Florida	1,621,000	1,690,238	0.959
Total	22,031,000	23,860,169	0.923

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.

The results from benchmarking the Census Numident to the CDC vital statistics show that the population-level death counts are similar across the data sources, even though each has a slightly different universe. The timing and quality of the Census Numident data have improved

over time, and the death data in the Census Numident are a high-quality source for measuring all-cause mortality.

DISCUSSION

The results show that the Census Numident accurately estimates all-cause mortality for the United States. While these data do not include cause of death, they become valuable research tools when linked at the person level to other data held at the Census Bureau. Moreover, these data are available to researchers on approved projects through the FSRDC network.

Linking Census Bureau Data to the Census Numident

Data sets available within the Census Bureau's Data Linkage Infrastructure are linkable to the Census Numident and allow researchers to measure the relationship between mortality and demographic characteristics, educational attainment, economic well-being, neighborhoods, migration, public policy, program participation, disability, general health, and many other potential social determinants of mortality. Researchers are already using 2000 and 2010 decennial census data and American Community Survey (ACS) data linked to Census Numident birthplace information to proxy for early life location and exposure.[26] Moreover, researchers can link individuals to the 1940 Decennial Census,[27] and soon to all decennial censuses from 1940 to 2020 , to understand the impacts of early life conditions and place-based exposure on mortality.[28]

As described previously, individual records in the Census Numident are assigned PIKs based solely on SSN. Data from the Census Numident are then included in the Census Bureau's Reference Files, which are used within the Census Bureau's Person Identification Validation System to assign PIKs probabilistically to other Census Bureau data using information such as SSN, name, address, birthdate, and sex.[29] All names and SSNs are removed after PIK application

so that data access by researchers remains confidential. Any data file that has been assigned PIKs can be linked at the individual level to the Census Numident. While the assignment of PIKs to data sets is probabilistic, reviews of the Person Identification Validation System show that PIK assignment has resulted in high-quality linkages with minimal error.[30,31] Because the Census Numident is restricted to SSN holders, linkages to the other data sets are also limited to SSN holders. The data files are restricted-use and available to researchers upon approval for specified projects.

The Census Bureau surveys that can be linked anonymously at the person level to the Census Numident vary in the type of data and population coverage. The 2000 and 2010 Decennial Censuses capture precise location, household structure, and basic demographic information for all residents in the United States, and roughly 90% of the person records in these files have been assigned a PIK. Additional detailed information on educational attainment, federal program participation, migration, employment, income, disability, fertility, veteran status, and dwelling characteristics are available for nearly 20% of Census 2000 (known as the long-form sample). Since 2000, the ACS has been fielded to nearly 3% of the population yearly, and it includes questions similar to the Census 2000 long form. Various Census Bureau surveys also have PIKs assigned to them, and they include more detailed questions on health, well-being, and life experiences for smaller samples. These data files include the Current Population Survey Annual Social and Economic Supplement, the Survey of Income and Program Participation, the National Crime Victimization Survey, and the National Survey of College Graduates.

In addition to survey data, the Census Bureau Data Linkage Infrastructure holds administrative data from federal agencies, state and local governments, and third parties that have had PIKs assigned at the person level. These data include Medicare and Medicaid

enrollment data, the Criminal Justice Administrative Records System, program data from the Department of Housing and Urban Development, and state-level administrative records from the Supplemental Nutrition Assistance Program and Special Supplemental Nutrition Program for Women, Infants, and Children. Researchers can also access the MAF-ARF, which links individuals with PIKs to address-level residential locations from 2000 to the present using comingled survey and administrative data, and the Census Household Composition Key, which links PIKs of parents to PIKs of children born from 1997 to the present.

Finally, the Census Bureau's Longitudinal Employer-Household Dynamics program integrates employer and employee data from state unemployment insurance records with other business and demographic data. Employees in the Longitudinal Employer-Household Dynamics data can be linked by PIK to the Census Numident. And the businesses in the Longitudinal Employer-Household Dynamics data can be linked to the many economic microdata files created by the Census Bureau for research including the Business Register, the Economic Census, and other establishment surveys.

Access and Use of the Census Numident File

The Census Numident file is available to researchers through the FSRDCs, along with all the other data described previously, for use on approved projects. The FSRDC network currently includes 32 physical research centers at universities and research institutions, and many projects are currently approved for virtual access.[32] Researchers can apply to use the Census Numident data through the standard Census Bureau FSRDC application process, which starts by contacting the closest FSRDC.[33]

The Census Numident is updated quarterly with new SSA transactions in March, June, September, and December. As discussed, there are slight delays in death reporting to SSA and inclusion in the Numident updates. At the median, dates of death now appear in the Census Numident a week after death events. About 25% of deaths take at least 2 weeks to appear, and the slowest 5% take 6 weeks to appear.

Public Health Implications

Complete, high-quality mortality data are essential for public health monitoring. Linking mortality data to survey and administrative data allows public health researchers to understand the relationships between mortality and demographic characteristics, social factors, economics, and geographic settings. Large linked data are also essential to evaluate and create evidence-based public health policy. We have shown that the Census Numident is a high-quality, population-wide mortality data source and that the Census Bureau's Data Linkage Infrastructure provides novel linkages to perform groundbreaking research on mortality. The use of these data to measure the relationships between social and economic determinants of all-cause mortality will improve our understanding of public health and health policy in the United States.

About the Authors: Keith F. Finlay is with the U.S. Census Bureau, Suitland, MD. Katie R. Genadek is with the U.S. Census Bureau, Suitland, MD and the Institute of Behavioral Science, University of Colorado – Boulder, Boulder, CO.
Correspondence: Katie R. Genadek, 1440 15th St, Boulder CO 80309, 202-839-0349, katie.r.genadek@census.gov.

References

1. Government Accountability Office. Social Security death data: additional action needed to address data errors and federal agency access United States. 2013. Available at: https://www.gao.gov/products/GAO-14-46. Accessed October 1, 2020.
2. US Social Security Administration Office of the Inspector General. Deceased beneficiaries who had different dates of death on the Social Security Administration's Numident and payment records. 2018. Available at: https://oig.ssa.gov/audits-and-investigations/audit-reports/A-09-17-50259. Accessed October 1, 2020.
3. US Social Security Administration Office of the Inspector General. Master beneficiary record death information that did not appear on the Numident. 2018. Available at: https://oig.ssa.gov/audits-and-investigations/audit-reports/A-06-17-50190. Accessed October 1, 2020.
4. Kestenbaum B. Mortality by nativity. Demography. 1986;23(1):87–90. https://doi.org/10.2307/2061410 PubMed
5. Rendall MS, Weden MM, Favreault MM, Waldron H. The protective effect of marriage for survival: a review and update. Demography. 2011;48(2):481–506. https://doi.org/10.1007/s13524-011-0032-5 PubMed
6. Miller S, Altekruse S, Johnson N, Wherry LR. Medicaid and mortality: new evidence from linked survey and administrative data. National Bureau of Economic Research. 2019. Available at: https://www.nber.org/papers/w26081. Accessed October 1, 2020. https://doi.org/10.3386/w26081
7. The National Archives and Records Administration. Frequently asked questions, Numerical Identification Files. 2018. Available at: https://aad.archives.gov/aad/content/aad_docs/rg047_num_faq_2018Dec.pdf. Accessed October 1, 2020.
8. Puckett C. The story of the Social Security Number. Soc Secur Bull. 2009;69(2):55–74.
9. Hill ME, Rosenwaike I. The Social Security Administration's death master file: the completeness of death reporting at older ages. Soc Secur Bull. 2001–2002;64(1):45–51. PubMed

10. US Social Security Administration. Program Operations Manual System (POMS). 2017. Available at: https://secure.ssa.gov/apps10/poms.nsf/lnx/0202602050. Accessed October 1, 2020.

11. Navar AM, Peterson ED, Steen DL, et al. Evaluation of mortality data from the Social Security Administration Death Master File for clinical research. JAMA Cardiol. 2019;4(4):375–379. https://doi.org/10.1001/jamacardio.2019.0198 PubMed

12. Gavrilov LA, Gavrilova NS. Mortality measurement at advanced ages: a study of the Social Security Administration Death Master File. N Am ActuarJ. 2011;15(3):432–447. https://doi.org/10.1080/10920277.2011.10597629 PubMed

13. Huntington JT, Butterfield M, Fisher J, Torrent D, Bloomston M. The Social Security Death Index (SSDI) most accurately reflects true survival for older oncology patients. Am J Cancer Res. 2013;3(5):518–522. PubMed

14. Hanna DB, Pfeiffer MR, Sackoff JE, Selik RM, Begier EM, Torian LV. Comparing the National Death Index and the Social Security Administration's Death Master File to ascertain death in HIV surveillance. Public Health Rep. 2009;124(6):850–860. https://doi.org/10.1177/003335490912400613 PubMed

15. National Technical Information Service. Important notice. 2011. Available at: https://classic.ntis.gov/assets/pdf/import-change-dmf.pdf. Accessed October 1, 2020.

16. Levin MA, Lin HM, Prabhakar G, McCormick PJ, Egorova NN. Alive or dead: validity of the Social Security Administration Death Master File after 2011. Health Serv Res. 2019;54(1):24–33. https://doi.org/10.1111/1475-6773.13069 PubMed

17. da Graca B, Filardo G, Nicewander D. Consequences for healthcare quality and research of the exclusion of records from the Death Master File. Circ Cardiovasc Qual Outcomes. 2013;6(1):124–128. https://doi.org/10.1161/CIRCOUTCOMES.112.968826 PubMed

18. Cowper DC, Kubal JD, Maynard C, Hynes DM. A primer and comparative review of major US mortality databases. Ann Epidemiol. 2002;12(7):462–468. https://doi.org/10.1016/S1047-2797(01)00285-X PubMed

19. Chetty R, Stepner M, Abraham S, et al. The association between income and life expectancy in the United States, 2001–2014. JAMA. 2016;315(16):1750–1766. https://doi.org/10.1001/jama.2016.4226 PubMed

20. Dupre ME, Gu D, Vaupel JW. Survival differences among native-born and foreign-born older adults in the United States. PLoS One. 2012;7(5):e37177. https://doi.org/10.1371/journal.pone.0037177 PubMed

21. Turra CM, Elo IT. The impact of salmon bias on the Hispanic mortality advantage: new evidence from Social Security data. Popul Res Policy Rev. 2008;27(5):515–530. https://doi.org/10.1007/s11113-008-9087-4 PubMed

22. Elo IT, Turra CM, Kestenbaum B. Mortality among elderly Hispanics in the United States: past evidence and new results. Demography. 2004;41(1):109–128. https://doi.org/10.1353/dem.2004.0001 PubMed

23. Centers for Disease Control and Prevention. CDC WONDER. Available at: https://wonder.cdc.gov. Accessed October 1, 2020.

24. Centers for Disease Control and Prevention. Deaths for selected causes by 10-year age groups, race, and sex: death registration states, 1900–32, and United States, 1933–98. Available at: https://www.cdc.gov/nchs/nvss/mortality/hist290a.htm. Accessed October 1, 2020.

25. Social Security Programs Operations Manual System (POMS). A. Assigning an SSN after death. 2015. Available at: https://secure.ssa.gov/apps10/poms.nsf/lnx/0110225080. Accessed January 1, 2021.

26. Bailey MJ, Hoynes HW, Rossin-Slater M, Walker R. Is the social safety net a long-term investment? Large-scale evidence from the food stamps program. National Bureau of Economic Research. 2020. Available at: https://www.nber.org/papers/w26942. Accessed October 1, 2020.

27. Massey CG, Genadek KR, Alexander JT, Gardner TK, O'Hara A. Linking the 1940 US Census with modern data. Hist Methods. 2018;51(4):246–257. https://doi.org/10.1080/01615440.2018.1507772 PubMed

28. Genadek KR, Alexander JT. The Decennial Census Digitization and Linkage Project. US Census Bureau. 2019. Available at: https://www.census.gov/content/dam/Census/library/working-papers/2019/econ/dcdl-workingpaper.pdf. Accessed October 1, 2020.

29. Wagner D, Lane M. The Person Identification Validation System (PVS): applying the Center for Administrative Records Research and Applications' (CARRA) record linkage software. US Census Bureau. 2014. Available at: https://www.census.gov/content/dam/Census/library/working-papers/2014/adrm/carra-wp-2014-01.pdf. Accessed October 1, 2020.

30. Mulrow E, Mushtaq A, Pramanik S, Fontes A. Assessment of the U.S. Census Bureau's Person Identification Validation System. NORC at the University of Chicago. 2011. Available at: https://www.norc.org/Research/Projects/Pages/census-personal-validation-system-assessment-pvs.aspx. Accessed October 1, 2020.

31. Layne M, Wagner D, Rothhaas C. Estimating record linkage false match rate for the Person Identification Validation System. US Census Bureau. 2014. Available at: https://www.census.gov/content/dam/Census/library/working-papers/2014/adrm/carra-wp-2014-02.pdf. Accessed October 1, 2020.

32. US Census Bureau. Research Data Centers. Available at: https://www.census.gov/content/census/en/about/adrm/fsrdc/locations.html. Accessed January 1, 2021.

33. US Census Bureau. Apply for access. Available at: https://www.census.gov/about/adrm/fsrdc/apply-for-access.html. Accessed January 1, 2021.

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

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 statues 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 Januar 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 Janaury 2021.
[9] Numerical Identification Files (NUMIDENT), 1936-2007: https://catalog.archives.gov/id/12004494 Accessed October 2020.