Compressed Mortality File 1968-1988

Description of the Mortality File

Mortality data on the CMF 1968-1988 are based on information from all death certificates filed in the 50 States and the District of Columbia during 1968-1988, excluding deaths of nonresidents (e.g. deaths of nonresident aliens, nationals residing abroad, and residents of Puerto Rico, the Virgin Islands, Guam, and other territories of the United States) and fetal deaths. Mortality data from the death certificates are coded by the States and provided to NCHS through the Vital Statistics Cooperative Program or coded by NCHS from copies of the original death certificates provided to NCHS by the State registration offices. Descriptions of the vital statistics reporting system maintained by NCHS and of the technical details of the mortality data are available in the Technical Notes in the annual National Vital Statistics Reports *Deaths: Final Data*

Only a select set of variables are extracted from the death records for inclusion on the CMF mortality record: 1) State and county of residence, 2) year of death (rather than the full date of death), 3) race (White, Black, other races), 4) sex, 5) age group at death, 7) underlying cause-of-death (4-digit ICD code, for 1968-78, the 8 8th revision; for 1979-88, the 9th revision), and 7) 69- or 72 cause-of-death recode. Including only these seven variables on the file and recoding some of them into a limited number of categories results in numerous records having identical values on all of the variables. Aggregating the records with identical values on all of the variables into one record and adding a count to that record indicating the number of records that have been aggregated substantially reduces the number of records on the file. . For example, two white male residents of Clay County, Alabama, with ages between 35 and 44 years, died from "bronchus and lung, unspecified" (ICD 162.9) in 1979. Their records were combined into one, with a 2 in the count field. Note that there are no records on the file with zero in the count field. If no deaths occurred for a particular combination of variable values, no record appears.

Specific details

1. Cause-of-death on the CMF is the underlying cause-of-death, which is defined by WHO as "the disease or injury which initiated the train of events leading directly to death, or the circumstances of the accident or violence which produced the fatal injury". When more than one cause or condition is listed on the death certificate, the underlying cause is determined by the sequence of conditions and associated selection rules. Underlying cause of death for the years 1968-78 is classified in accordance with the *Eighth Revision International Classification of Diseases, Adapted for Use in the United States* (ICDA-8) codes. Cause of death for the years 1979-88 is classified in accordance with the *International Classification of Disease, Ninth Revision* (ICD-9) codes.

Changes in classification of causes of death due to the ICD revisions result in discontinuities in cause-of-death trends. Comparability ratios for the 8th and 9th revisions are available for selected causes of death.

2. The fourth digit of the ICD code can assume the values 0-9 and blank. If the fourth digit of the ICD code is a "blank" on the source file, it is a blank on this file. Care must be taken when reading in the ICD code to distinguish between blanks and zeros.

3. For injuries and poisonings, the external cause is coded (E800-E999) rather than the Nature of Injury (800-999). The letter "E" is not included in the code.

4. In 1988, if there were three or fewer deaths for a given Georgia county of residence (of deaths occurring in Georgia) with HIV infection (ICD codes *042-*044, 795.8) cited as a cause-of-death (underlying or non-underlying cause), these records were assigned a "missing" place of residence code (FIPS code = 13999).

5. Race and Hispanic origin are reported separately on the death certificate in accordance with standards set forth by the Office of Management and Budget (OMB). Thus, persons of a given race may be Hispanic or non-Hispanic.

6. Death rates by race are based on information from death certificates (numerators of the rates) and on population estimates from the Census Bureau (denominators). Information included on the death certificate about the race of the decedent is reported by the funeral director as provided by an informant, often the surviving next of kin, or, in the absence of an informant, on the basis of observation. Race information from the census is by self-report. To the extent that race is inconsistent between these two data sources, death rates will be biased. Studies have shown that persons self-reported as American Indian or Asian on census and survey records may sometimes be reported as white on the death certificate, resulting in an underestimation of deaths and death rates for the American Indian and Asian groups. Bias also results from undercounts of some population groups in the census, particularly young black males, young white males, and elderly persons, resulting in an overestimation of death rates. Rosenberg estimated that the misclassification and undercoverage result in overstated death rates for the white and black populations (1% and 5%, respectively) and understated death rates for other population groups (American Indians, 21%; Asian or Pacific Islanders, 11%).

7. The State and county FIPS codes contain leading zeros in both the 2-byte State code and the 3-byte county code.

Mortality File Layout

File name	Years	Number of records	Record t length	File size	Format
MORT6878.txt	1968-1978	8,774,864	23	215MB	ASCII
MORT7988.txt	1979-1988	8,776,385	23	215MB	ASCII

The file is sorted by locations 6-9, 1-5, 10, 11-12, 13-16.

Location	Field size	Item and code outline	Format
		FIPS Codes	
1-2	2	FIPS State code	Numeric
3-5	3	FIPS county code	Numeric
6-9	4	<u>Year of death</u> 1968-1978 1979-1988	Numeric
10	1	<u>Race-sex</u> 1　White male 2　White female 3　Black male 4　Black female 5　Other races male 6　Other races female	Numeric
11-12	2	<u>Age at death</u> 01　under 1 day 02　1-6 days 03　7-27 days 04　28-364 days 05　1-4 years 06　5-9 years 07　10-14 years 08　15-19 years 09　20-24 years 10　25-34 years 11　35-44 years 12　45-54 years 13　55-64 years 14　65-74 years 15　75-84 years 16　85+ years 99　Unknown	Numeric
13-16	4	<u>ICD code for underlying cause-of-death</u> 1968-78: ICDA-8 1979-88: ICD-9	3 digits: Numeric 4-digits: Character

Location	Field size	Item and code outline	Format
17-19	3	<u>Cause-of-Death Recode</u> 1968-78: 69 Cause Recode 1979-88: 72 Cause Recode	Numeric
20-23	4	<u>Number of deaths</u>	Numeric

Description of the Population File

The national, State, and county population estimates on the CMF are estimates of the resident population of the United States produced by the U.S. Census Bureau. The estimates for 1971-79 and 1981-88 are intercensal estimates of July 1 resident populations. The 1970 and 1980 population estimates are April 1 modified census counts. The 1968 and 1969 state and county population estimates were calculated by NCHS using linear extrapolation.

Specific details

1. **National, State, and county estimates** - There are records on the file for each geographic level (total U.S., state, county. National, State, and county population estimates can be identified by using the FIPS code in location 1-5 or the record type variable in location 140. National population records have a State FIPS code of "00" and a county FIPS code of "000" and a record type code of "1". State population records have a nonzero 2-digit State FIPS code and a county FIPS code of "000" and a record type code of "2". County population records have nonzero 2-digit State and 3-digit county FIPS codes. The record type value for county records is "3".

   The sum of the population estimates of counties within a state may not equal the state population estimate, and the sum of all state population estimates or all county population estimates may not equal the national population estimate. This is due to rounding error that results because fractional estimates are not allowed. In general, the national population estimates should be used when calculating national death rates and the state population estimates should be used when calculating state death rates.

2. **Rounding of national estimates -** For non-censal years prior to 1992, the NCHS Division of Vital Statistics used national population estimates rounded to the nearest 1,000 to calculate published death rates. On the CMF, the national population estimates for 1968-69 and 1971-79 are rounded to the nearest 1,000 in accordance with this practice. However, this means that calculation of rates for aggregated age, race, and/or sex groups involves using population estimates that were rounded before aggregation rather than after aggregation. As a result, rates for aggregated groups calculated using the rounded estimates on the CMF may differ slightly from those published by NCHS. The national population estimates for 1981-88 on this file are not rounded so that the user can round them after aggregating across subgroups and avoid the rounding error problem.

3. When the age group 0-4 years rather than the age groups Under 1 year and 1-4 years appeared on the Census file, the age group 0-4 years was multiplied by 0.8 to obtain an estimate of the population 1-4 years.

4. The State and county FIPS codes contain leading zeros in both the 2-byte State code and the 3-byte county code.

5. **Live births and population under one year of age -** To permit the calculation of infant mortality rates and maternal mortality rates, NCHS live-birth data were substituted for the estimates of the population under one year of age. The race code for live births is "race of mother" as stated on the birth certificate.

6. In 1988, there was an additional county in Georgia with a "missing" county code of "999". The six records for this county have population counts of zero.

7. **Limitations of State and county population estimates -** The State and county population estimates have been provided for age-race-sex groups for the user's convenience in aggregating to various groups. However, the limitations of the methodology used to derive State and county estimates are such that the U.S. Census Bureau does not consider the estimates to be accurate for each age-race-sex cell. The Census Bureau believes that aggregating the individual cells to larger groups will reduce the level of error. Further, although most of the estimates are not rounded, the U.S. Census Bureau does not consider the estimates to be accurate to the last digit.

8. **Source of population estimates**

    The table summarizes the sources of the population data on the CMF.

Sources of population data for 1968-1988, by year and type of estimate			
Date of estimate	Geographic type		
	National	State	County
July 1, 1968-69	Intercensal series	Derived by NCHS from 1970 and 1971 State estimates	Derived by NCHS from 1970 and 1971 county estimates
April 1, 1970	Modified –race census counts	Modified–race census counts	Modified-race census counts
July 1, 1971-79	Intercensal estimates	Sum of county intercensal estimates	Intercensal estimates
April 1, 1980	Modified –race census counts	Modified–race census counts	Modified-race census counts
July 1, 1981-88	Intercensal estimates	Intercensal estimates	Intercensal estimates

1968-69 population estimates - National population estimates are U.S. Bureau of the Census intercensal estimates of the July 1 resident population by race, sex, and age group. State and county population estimates were calculated by NCHS using linear extrapolation from the corresponding July 1, 1970 and July 1, 1971 state and county estimates.

1970 population estimates - National, State, and county population estimates are from a modified version of the April 1, 1970 census. The original census counts were modified by the U.S. Bureau of the Census to correct: 1) errors discovered in the data, 2) race misclassification - persons of Hispanic origin who reported their race as "other" were recoded as "white".

1971-79 population estimates - National and county estimates are U.S. Bureau of the Census intercensal estimates of the July 1 resident population. The Bureau of the Census did not produce state population estimates by age, race, and sex for the 70's. Therefore, the state population estimates for 1971-79 on this file are simply the sum of the population estimates for the counties in each state.

Three Virginia independent cities (Manassas, Manassas Park, and Poquoson) did not appear on the Census file. While these independent cities are not on the mortality file for 1968-78, they are on the file for 1979 onwards. Therefore, 979 populations for these three cities were estimated from the July 1, 1980 and July 1, 1981 estimates of these cities. The 1979 population estimates for the counties containing the cities were reduced by the estimated city populations.

1980 population estimates - National, state, and county population estimates are from a modified version of the April 1, 1980 census. The original census counts were modified by the U.S. Bureau of the Census: 1) persons who reported their race as "other" (the majority being of Hispanic origin) were reassigned to one of the official race groups, 2) an adjustment was made for the overcount of centenarians

April 1, 1980 population estimates for three Virginia independent cities, (Manassas, Manassas Park, and Poquoson) had to be extrapolated from July 1, 1980 estimates. The April 1 populations for the three cities were calculated as a proportion of the April 1 county population, where the proportion was calculated from the July 1, 1980 estimates. The April 1 population estimates for the counties containing the three cities were reduced by the cities estimated April 1 populations.

1981-88 population estimates - National, State, and county estimates are U.S. Bureau of the Census intercensal estimates of the July 1 resident population.

Population File Layout

File name	Years	Number of records	Record length	File size	Format
POP6878.txt	1968-1978	206,712	140	29MB	ASCII
POP7988.txt	1979-1988	189,966	140	27MB	ASCII

The file is sorted by locations 6-9, 1-5, 10, 11.

Location	Field size	Item and code outline	Format
		FIPS codes	
1-2	2	FIPS State code	Numeric
3-5	3	FIPS county code	Numeric
6-9	4	Year 1968-1978 1979-1988	Numeric
10	1	Race-sex 1 White male 2 White female 3 Black male 4 Black female 5 Other races male 6 Other races female	Numeric
11-18	8	Number of live births	Numeric
19-26	8	Population in age group: 1-4 years	Numeric
27-34	8	Population in age group: 5-9 years	Numeric
35-42	8	Population in age group: 10-14 years	Numeric
43-50	8	Population in age group: 15-19 years	Numeric
51-58	8	Population in age group: 20-24 years	Numeric
59-66	8	Population in age group: 25-34 years	Numeric
67-74	8	Population in age group: 35-44 years	Numeric
75-82	8	Population in age group: 45-54 years	Numeric
83-90	8	Population in age group: 55-64 years	Numeric
91-98	8	Population in age group: 65-74 years	Numeric
99-106	8	Population in age group: 75-84 years	Numeric
107-114	8	Population in age group: 85+ years	Numeric

Location	Field size	Item and code outline	Format
115-139	25	<u>County name</u>	Character
140	1	<u>Record type</u> 1=National population 2=State population 3=County population	Numeric