Public-use Linked Mortality File

Updated March 2020

INTRODUCTION

The National Center for Health Statistics (NCHS) has linked data collected from several NCHS population surveys with death certificate records from the National Death Index (NDI). Due to requirements to protect the confidentiality of the NCHS survey participants, restricted-use versions of the linked mortality files are made available only through the NCHS Research Data Center (RDC).

To complement the restricted-use files and increase data access, NCHS developed public-use versions of the linked mortality files for the 1986-2014 National Health Interview Survey (**NHIS**), 1999-2014 National Health and Nutrition Examination Survey (**NHANES**) and NHANES III. The public-use linked mortality files include a limited set of variables for adult participants only. The public-use versions of the NCHS linked mortality files were subjected to data perturbation techniques to reduce the participant disclosure risk. Synthetic data were substituted for follow-up time and underlying cause of death for select records. Information regarding vital status was not perturbed. The public-use linked mortality file provides mortality follow-up data from the date of survey participation through December 31, 2015. Detailed description of the linkage methodology and analytic guidelines can be found on the NCHS Data Linkage webpage:
https://www.cdc.gov/nchs/data/datalinkage/LMF2015_Methodology_Analytic_Considerations.pdf

LINKAGE ELIGIBILITY STATUS

All participants with sufficient identifying data were eligible for mortality follow-up. Any survey participant record that did not meet the minimum data requirements was ineligible for record linkage. All survey participants are included on the public-use linked mortality files regardless of linkage eligibility.

In the public-use linked mortality files, the variable **ELIGSTAT** notes linkage eligibility. A value of 1 for ELIGSTAT indicates that the survey participant was eligible for the mortality linkage, a value of 2 indicates the survey participant was under age 18 and not eligible for public release and a value of 3 indicates the survey participant was not linkage eligible due to having insufficient identifying data. Please note that all survey participants from the survey files are included on the linked mortality files regardless of linkage eligibility. See the "Linkage of National Center for Health Statistics Survey Data to the National Death Index—2015 Linked Mortality File (LMF): Methodology and Analytic Considerations" for more information (https://www.cdc.gov/nchs/data/datalinkage/LMF2015_Methodology_Analytic_Considerations.pdf).

MORTALITY STATUS

In the public-use linked mortality file the determination of vital status can be found using the **MORTSTAT** variable. Each eligible participant is assigned a vital status code equal to 0 if assumed alive and 1 if assumed deceased. Vital status may have been ascertained from other sources, including linkages with administrative data from the Social Security Administration and Centers for Medicare & Medicaid Services, for a small percentage of deaths. These can be identified by crossing **MORTSTAT** and the leading cause of death variable (**UCOD_LEADING**). Records that have a value of 1 for **MORTSTAT** and a missing cause of death likely came from a non-NDI source. For participants who are not linkage eligible, **MORTSTAT** is indicated as a numeric missing value.

FOLLOW-UP TIME

For 1986-2014 NHIS participants that are assumed deceased, follow-up time can be calculated using the quarter of death (**DODQTR**) and/or the year of death (**DODYEAR**). For those assumed alive, the follow-up time can be calculated using the end of the mortality period (December 31, 2015).

Probabilistic techniques were used in the linkage of the NCHS survey data to the NDI and in the perturbation processes used to create the public-use linked mortality file. While vital status is not perturbed, the perturbation process helps to limit re-identification risks by introducing some noise into the files for date of death and cause of death.

However, these processes may generate some inconsistencies between perturbed date of death and survey interview quarter for a small number of NHIS cases (<0.01%). Researchers have the option to drop these cases or calculate the time-to-event with two potential approaches: (1) set time-to-event to 0.5 years if year of death (**DODYEAR**) equals survey year (**SRVY_YR**), or (2) set time-to-event to 0.1 years if year of death equals survey year and death quarter (**DODQTR**) equals or occurs before survey quarter (**SRVY_QTR**). SAS syntax for both options is provided here:

1. if MORTSTAT = 1 and (DODYEAR = SRVY_YR) then TIME_TO_EVENT = 0.5;
2. if MORTSTAT = 1 and (DODYEAR = SRVY_YR) and (DODQTR <= SRVY_QTR) then TIME_TO_EVENT = 0.1;

For 1999-2014 NHANES and NHANES III follow-up time has been calculated using person months from the date of interview to the date of death or the end of the mortality period (**PERMTH_INT**). In addition, the person months from the mobile examination center date to date of death or the end of the mortality period has been calculated (**PERMTH_EXM**).

LEADING CAUSES OF DEATH: RECODES FROM UCOD_113

For 1986-2014 NHIS, 1999-2006 NHANES, and NHANES III, the nine cause-specific death categories included on the public-use linked mortality files include the following groups from the underlying cause-of-death 113 recodes: heart disease (54-64), cancer (malignant neoplasms) (19-43), chronic lower respiratory disease (82-86), cerebrovascular diseases (70), diabetes (46), pneumonia and influenza (76-78), Alzheimer's disease (52), kidney disease (97-101), unintentional injuries (112-123).

For 2007-2014 NHANES the cause-specific death categories included on the public-use linked mortality files include the following groups from the underlying cause-of-death 113 recode: heart disease (54-64) and cancer (malignant neoplasms) (19-43). For these years of NHANES the

cause-specific death categories are limited because of the short follow-up time and small sample sizes for the other cause-specific death categories.

Cause-of-death coding for all U.S. deaths occurring prior to 1999 follows the 9th revision of the International Statistical Classification of Diseases, Injuries, and Causes of Death (ICD-9) guidelines, while coding for all deaths after 1998 follows the 10th revision of the International Statistical Classification of Diseases, Injuries, and Causes of Death (ICD-10) guidelines. To assist the researcher with analyses that span the entire survey specific mortality period, the **UCOD_113** variable was created as a recode of all deaths occurring prior to 1999 coded under ICD-9 guidelines into comparable ICD-10 based underlying cause of death groups. In the public-use linked mortality file, codes from the leading causes of death (**UCOD_LEADING**) are provided and based on the **UCOD_113** variable available on the restricted-use linked mortality file. The coding for **UCOD_113** can be found in section 5.1 of this document: https://www.cdc.gov/nchs/data/datalinkage/underlying_and_multiple_cause_of_death_codes.pdf

For more on the comparability of ICD-9 and ICD-10 underlying cause of death UCOD 113 recode see:

Anderson RN, Minino AM, Hoyert DL, Rosenberg HM. Comparability of cause of death between ICD-9 and ICD-10: Preliminary estimates. National Vital Statistics Reports; Vol 49 No.2. Hyattsville, MD: National Center for Health Statistics. 2001.

This document is also available on the NCHS website.

PUBLIC-USE LINKED MORTALITY FILE SAMPLE PROGRAMS

Please note that all surveys now use the same file layout and that the data for survey-specific variables will be blank in files for other surveys. A codebook noting all the variables and descriptions is available at https://www.cdc.gov/nchs/data-linkage/mortality-public.htm. There are three sample programs provided to assist researchers when reading in the public-use linked mortality ASCII files.

The first sample program provides researchers using **SAS** data input statements to read in the data file and run their analyses using SAS. By using the ASCII data file (.DAT file) as input to the program, other types of data files (e.g. SAS, SPSS) can be created.

The second program provides researchers using **STATA** with code to read in the data file and run their analyses using the STATA software.

The third program provides researchers using **R** with a program to read in the data file and run their analyses using R software.

Although the public-use linked mortality file was carefully edited, errors may be detected. Please e-mail the NCHS Data Linkage staff at datalinkage@cdc.gov if any errors are detected in the public-use linked mortality file data or documentation.

NOTES ON CONSTRUCTING PUBLICID

The NCHS linked mortality files are person-level files and can be linked to the public-use survey files by matching on the unique person-level **PUBLICID** (for NHIS) and **SEQN** (for NHANES).

Across the NHIS years included in the linked mortality files, NHIS has changed its construction

of a unique person-level ID. Instruction on how to construct a person-level ID from the NHIS public-use files and merge it with the public-use linked mortality files is available in section 5.1 of the "Linkage of National Center for Health Statistics Survey Data to the National Death Index—2015 Linked Mortality File (LMF): Methodology and Analytic Considerations" document: https://www.cdc.gov/nchs/data/datalinkage/LMF2015_Methodology_Analytic_Considerations.pdf

The **PUBLICID** available on the linked mortality files is consistent with the documentation provided for each NHIS year.

STATEMENT OF AUTHENTICITY

This material has been cleared for public distribution by CDC and will be authentic if obtained directly from ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/linked_mortality/. CDC makes every effort to assure the authenticity of electronically distributed documents. However, in all instances where the electronic and official agency record differ, the authenticity of the official agency record is controlling.

GUIDELINES FOR CITATION OF THE PUBLIC-USE LINKED MORTALITY DATA

With the goal of mutual benefit, the NCHS requests that recipients of data files cooperate in certain actions related to their use. Any published material derived from the data should acknowledge NCHS as the original source. The suggested citation to appear at the bottom of all published tables is as follows:

Source: National Center for Health Statistics

When cited in a bibliography, the citation should read:

National Center for Health Statistics. Office of Analysis and Epidemiology, Public-use Linked Mortality File, 2015. Hyattsville, Maryland. (Available at the following address: https://www.cdc.gov/nchs/data-linkage/mortality-public.htm).

The published material should also include a disclaimer that credits any analyses, interpretations, or conclusions reached to the author (recipient of the data file) and not to NCHS, which is responsible only for the initial data. Consumers who wish to publish a technical description of the data should make an effort to ensure that the description is consistent with that published by NCHS.

DATA USE RESTRICTIONS! READ CAREFULLY BEFORE USE.

The Public Health Service Act (Section 308 (d)) provides that the data collected by the National Center for Health Statistics (NCHS), Centers for Disease Control and Prevention (CDC), may be used only for the purpose of health statistical reporting and analysis.

Any effort to determine the identity of any reported case is prohibited by this law.

NCHS does all it can to ensure that the identity of data subjects cannot be disclosed. All direct identifiers, as well as any characteristics that might lead to identification, are either omitted from the data file or perturbed to prevent re-identification. Any intentional identification or disclosure of a person or establishment violates the assurances of confidentiality given to the providers of the information. Therefore, users will:

1. Use the data in these data files for statistical reporting and analysis only.

2. Make no use of the identity of any person or establishment discovered inadvertently and advise the Director, NCHS, of any such discovery (301-458-4500).

3. Not link these data files with individually identifiable data from other NCHS or non-NCHS data files.

By using these data, you signify your agreement to comply with the above-stated statutorily-based requirements.