Public-use Linked Mortality Files

Updated May 2022

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). In order to protect the confidentiality of the NCHS survey participants, restricted-use versions of the linked mortality files (LMF) 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 LMF for the 1986-2018 National Health Interview Survey (**NHIS**), 1999-2018 National Health and Nutrition Examination Survey (**NHANES**), and **NHANES III**.

The public-use LMF include a limited set of variables for adult participants only and were subjected to data perturbation techniques to reduce participant disclosure risk. Synthetic data were substituted for follow-up time or underlying cause of death for select records. Information regarding vital status was not perturbed. The public-use LMF provide mortality follow-up data from the date of survey participation through December 31, 2019.

ENHANCED LINKAGE ALGORITHM

The production of the most recent LMF used an enhanced linkage algorithm compared to earlier versions of the files. The enhanced linkage methodology, which relies on the Fellegi-Sunter paradigm, was adopted to account for changes in the survey data collection process and to address the need for an updated linkage algorithm. The enhanced linkage algorithm was designed to improve the quality of matches and to minimize type I (false positive) and type II (false negative) linkage errors. The 2019 LMF supersede all previously released versions of the LMF. Due to the probabilistic nature of the linkage, those that linked to the NDI are assumed deceased and those that did not are assumed alive.

For a small number of records, some values of vital status may have changed from the previous release due to the changes in the linkage algorithm. The vital status remained the same for 98% (or higher depending on the survey year) of the survey participants that were included in both linkages when comparing the same follow up time (e.g., limiting the comparisons to deaths through December 31, 2015). However, there may be some differences in vital status or underlying cause of death data between the previously released public-use LMF and the new 2019 public-use LMF. For example, a small number of participants previously assumed deceased in the 2015 public-use LMF are assumed alive in the 2019 public-use LMF. Similarly, for a small number of participants who were assumed deceased in both linkages, the cause of death code (reported in **UCOD_LEADING**) may have changed. A detailed comparison of the current (2019) and previously released (2015) LMF is available in Appendix II of the Linkage of National Center for Health Statistics Survey Data to the National Death Index—2019 Linked Mortality File (LMF): Methodology and Analytic Considerations.

LINKAGE ELIGIBILITY

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. For NHIS 2015-2018, only sample adult and sample child survey participants were eligible for mortality linkage.

In the public-use LMF, the variable **ELIGSTAT** indicates 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 to conduct data linkage. Because the public-use LMF only include mortality information for survey participants aged 18 and older, children will have a value of 2 for ELIGSTAT and will have missing values for all other mortality or follow-up variables. All NHANES and NHIS participants are included on the LMF, regardless of linkage eligibility.[1] See the Linkage of National Center for Health Statistics Survey Data to the National Death Index—2019 Linked Mortality File (LMF): Methodology and Analytic Considerations for more information.

MORTALITY STATUS

In the public-use LMF, the determination of vital status can be found using the **MORTSTAT** variable. Each survey participant who is linkage eligible for mortality follow-up is assigned a vital status code (0=Assumed alive, 1=Assumed deceased). Vital status may be inferred from other sources for a small percentage of deaths and can be identified by crossing **MORTSTAT** and the leading cause of death variable (**UCOD_LEADING**). These records can be identified by crossing **MORTSTAT** and the leading cause of death variable (**UCOD_LEADING**); however, in a small number of cases UCOD_LEADING was also missing from the death certificate record. For participants who are not linkage eligible, **MORTSTAT** is coded as a numeric missing value.

FOLLOW-UP TIME

For 1986-2018 NHIS participants that are assumed deceased, follow-up time since the interview quarter can be calculated using the quarter of death (**DODQTR**) and the year of death (**DODYEAR**). For those assumed alive, the follow-up time can be calculated using the end of the mortality follow-up period (December 31, 2019).

For 1999-2018 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 follow-up period has been calculated (**PERMTH_EXM**).

[1] For NHIS 2015-2018, only sample adult and sample child survey participants are included on the LMF, since other survey participants in the NHIS Person files were not eligible for mortality linkage.

LEADING CAUSES OF DEATH: RECODES FROM UCOD_113

For 1986-2014 NHIS, 1999-2014 NHANES, and NHANES III, the nine cause-specific death categories included on the public-use LMF include the following groups selected from the NCHS underlying cause-of-death 113 recodes: heart disease (54-68), cancer (malignant neoplasms) (19-43), chronic lower respiratory disease (82-86), unintentional injuries (112-123), cerebrovascular diseases (70), Alzheimer's disease (52), diabetes (46), pneumonia and influenza (76-78), and kidney disease (97-101).

For 2015-2018 NHIS, the cause-specific death categories included on the public-use LMF include the following five groups selected from the NCHS underlying cause-of-death 113 recode: heart disease (54-68), cancer (malignant neoplasms) (19-43), chronic lower respiratory disease (82-86), unintentional injuries (112-123), and cerebrovascular diseases (70). For 2015-2018 NHANES, the cause-specific death categories included on the public-use LMF include the following two groups selected from the NCHS underlying cause-of-death 113 recode: heart disease (54-68) and cancer (malignant neoplasms) (19-43). For these years the cause-specific death categories are limited because of the 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 researchers with analyses that span the entire survey specific mortality period, NCHS created a **UCOD_113** variable to recode all deaths into comparable ICD-10 based underlying cause of death groups. In the public-use LMF, codes from the national leading causes of death (**UCOD_LEADING**) are provided and based on the **UCOD_113** variable available on the restricted-use LMF. The coding for **UCOD_113** is provided in section 5.1 of this document.

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.

CONTRIBUTING CAUSE OF DEATH FLAGS: DIABETES AND HYPERTENSION

Certain contributing causes of death are only coded in the multiple causes of death codes according to ICD coding rules. The current public-use LMF does not include the 20 entity-axis multiple causes of death codes. The multiple causes of death codes are only available on the restricted-use versions of the LMF. However, the current public-use LMF includes two variables, **DIABETES** and **HYPERTEN**, if there was a specific contributing cause of death for these conditions listed as a multiple cause of death (MCOD).

SAMPLE WEIGHTS

The sample weights provided in the NCHS survey data files adjust for oversampling of specific subgroups and differential nonresponse and are post-stratified to annual population totals for specific population domains to provide nationally representative estimates. The use of sampling weights and sample design variables is recommended to account for the complex survey design of the NCHS surveys.

The properties of the survey sample weights for linked data files with incomplete linkage, due to

ineligibility for linkage, are unknown. One approach is to analyze linked data files using adjusted sample weights. The sample weights available on NCHS population health survey data files can be adjusted for linkage eligibility (nonresponse), using standard weighting domains to reproduce population counts within these domains: sex, age, and race and ethnicity subgroups.

Eligibility-adjusted weights are provided for NHIS. Adjusted Sample Adult weights are provided for NHIS 1997-2018, and adjusted person-level weights are provided for NHIS 1987-2014. (There are no person-level adjusted weights for 2015-2018 since only sample adults and sample children were eligible for linkage in those years.)

For analyses using the public-use LMF for other surveys, researchers may consider adjusting the original sampling weight to account for linkage ineligibility due to insufficient identifying data. Depending on the number of ineligible survey participants, ignoring linkage ineligibility may lead to biased mortality estimates. See the Linkage of National Center for Health Statistics Survey Data to the National Death Index—2019 Linked Mortality File (LMF): Methodology and Analytic Considerations for more information.

PUBLIC-USE LINKED MORTALITY FILE SAMPLE PROGRAMS

Please note that all surveys 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 on the NCHS data linkage website. 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** with data input statements to read in the data file and run their analyses using SAS software. 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 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 LMF were 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 LMF data or documentation.

NOTES ON CONSTRUCTING PUBLICID

The NCHS LMF 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 LMF, NHIS has changed its construction of a unique person-level ID. Instructions on how to construct a person-level ID from the NHIS public-use files and merge it with the public-use LMF is available in Appendix III of the Linkage of National Center for Health Statistics Survey Data to the National Death Index—2019 Linked Mortality File (LMF): Methodology and Analytic Considerations. The **PUBLICID** available on the LMF 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 https://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 FILES

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. When the data are used in analyses, they should be cited in the references. The citations should read:

National Center for Health Statistics Division of Analysis and Epidemiology. NHANES III Public-use Linked Mortality Files, 2019. Hyattsville, Maryland. Available from: https://www.cdc.gov/nchs/data-linkage/mortality-public.htm. doi:10.15620/cdc:117141.

National Center for Health Statistics Division of Analysis and Epidemiology. Continuous NHANES Public-use Linked Mortality Files, 2019. Hyattsville, Maryland. Available from: https://www.cdc.gov/nchs/data-linkage/mortality-public.htm. doi:10.15620/cdc:117142.

National Center for Health Statistics Division of Analysis and Epidemiology. NHIS Public-use Linked Mortality Files, 2019. Hyattsville, Maryland. Available from: https://www.cdc.gov/nchs/data-linkage/mortality-public.htm. doi:10.15620/cdc:117143.

The published material should also include a disclaimer that credits any analyses, interpretations, or conclusions reached to the author and not to NCHS, which is responsible only for the initial data. Consumers who wish to publish a technical description of the data should 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.