Analytic Guidelines for NCHS 2011 Linked Mortality Files

These guidelines address the following analytical topics:

I. Linkage eligibility status

All participants with sufficient identifying data were eligible for mortality follow-up. Each record was screened to determine if it contained at least one of the following combinations of identifying data elements:

1. SSN (nine digits or last four digits), sex, full date of birth present
2. Last name, first name, month of birth, year of birth present
3. Last name, first name, SSN (nine digits or last four digits) present

Any survey participant record that did not meet these minimum data requirements was ineligible for record linkage.

Eligibility status for mortality follow-up is indicated by the variable ELIGSTAT. For mortality or survival analyses, analysts should limit their analysis to those survey records with a value of ELIGSTAT = 1.

II. Use of survey weights

Survey sampling weights

The use of sampling weights and sample design variables is recommended for NCHS survey analyses to account for the complex survey design of NCHS studies. Failing to account for the complex survey design may produce biased estimates and overstated significance levels.

Eligibility adjusted sampling weights

For analyses using the Linked Mortality Files, it is recommended that survey participants classified as eligible for mortality follow-up have their original sampling weight adjusted to account for those ineligible for linkage to the National Death Index (NDI) due to insufficient identifying data. Ignoring those ineligible for linkage in the 2011 Linked Mortality Files may lead to biased mortality estimates. NCHS has provided guidance on methods available to adjust sampling weights: Use of Survey Weights for Linked Data Files – Preliminary Guidance.

NHIS eligibility adjusted sample weights

The NCHS Special Projects Branch has provided eligibility adjusted weights for the NHIS from 1987-2009. For the 2011 Linked Mortality Files, there are no eligibility adjusted sample weights for the 1985 and 1986 NHIS. NCHS recommends using the public-use annual final basic weight (WTFA) for those survey years.

For the 1987-2009 NHIS, NHIS participants classified as eligible for mortality follow-up had their original NHIS sampling weight adjusted to account for those ineligible for linkage to the NDI due to insufficient identifying data in the 2011 Linked Mortality Files. The new eligibility adjusted sample weights provided on the 2011 Linked Mortality Files should be used in place of the original NHIS sample weights to prevent biased mortality estimates.

The 2011 Linked Mortality Files include three eligibility adjusted sample weights for the NHIS:

- WGT_NEW refers to a person-level record and is available for the NHIS years 1987-2009

- SA_WGT_NEW refers to a sample adult record and is available for the NHIS years 1997-2009

- SC_WGT_NEW refers to a sample child record and is available for the NHIS years 1997-2009

The NHIS from 1987 to 1996 did not include sample adult or sample child files.

Technical note:
Treating the eligible sample from the NHIS as a subsample of the original NHIS sample allows for the original post-stratification adjustment method to be used to inflate the sampling weights. The tacit assumption is the adjustment cells used will mitigate estimation bias due to using only the eligible sample.

III. Pooled analyses of NCHS Linked Mortality Files

Analysts may wish to pool several survey years of the 2011 Linked Mortality Files to increase sample size for many types of analyses. The eligibility adjusted sampling weights should be adjusted for pooled data; otherwise, estimates of totals will be too high. For example, the estimated total U.S. civilian non institutionalized population from two years of pooled data, using unadjusted weights, would be about twice as large as it should be. A simple, valid weight

adjustment procedure that NCHS recommends is to divide each sample weight in the pooled dataset by the number of years that are being pooled; e.g., divide by 2 when two years of data are combined, divide by 3 when three years of data are combined, etc.

Pooled analysis methods for estimating variance

Analysts should note that when combining data sets, it is the data users' responsibility to examine possible changes in variable names and/or locations on the data files (e.g. changes in variable location or name). Differences in the study design variables may also be an issue when pooling survey years within a specific survey.

NHIS has provided analysts with guidance for variance estimation for pooled analyses of NHIS years. Please refer to the following NHIS file documentation for additional information:

- NHIS 1986-1994. See http://www.cdc.gov/nchs/data/nhis/8594var.pdf
- NHIS 1995-1996. See http://www.cdc.gov/nchs/data/nhis/96var.pdf
- NHIS 1997-2005 See http://www.cdc.gov/nchs/data/nhis/9705var.pdf
- NHIS 2006-2011 See http://www.cdc.gov/nchs/data/nhis/2006var.pdf

The National Health and Nutrition Examination Survey provides tutorials on pooling years of NHANES data, including construction of appropriate pooled sample weights. Links to the NHANES tutorials can be found here: http://www.cdc.gov/nchs/tutorials/nhanes/.

IV. Linkage survey participants with improbable ages

The NCHS 2011 Linked Mortality Files include records where the calculated age presumed alive at the end of mortality follow-up is 100 years or more. For these cases there was no valid NDI record match or any other source of mortality information. Yet, given the probabilistic nature of the mortality ascertainment and the lower likelihood of being alive at 100 years or older, analysts may wish to consider these cases as loss to follow-up and make them ineligible for mortality analyses.

One possible method to determine an age cutoff at which participants should be considered lost to follow-up is to use the probability that survey participant was alive as of December 31, 2011. The NCHS Special Projects Branch has calculated probabilities based on cohort life tables from: Life Tables for the United States Social Security Area 1900-2100. SSA Pub. No. 11-11536 available at: http://www.ssa.gov/OACT/NOTES/pdf_studies/study120.pdf. Probabilities were conditional only on year of birth and sex of the survey respondent; they were not adjusted for last known alive year (typically the year of survey response). These probabilities will be made available as a separate table that can be merged with the 2011 Linked Mortality Files and can be used as an additional tool to help researchers.

V. Inconsistencies in baseline age and follow-up age

Misreporting or discrepancies between reported age at interview and the date of birth may result in values for age at death or age last presumed alive that are inconsistent with baseline age, resulting in negative follow-up time for survival analyses. The number of cases where this occurs is small but analysts should be aware and make appropriate adjustments to the data.

VI. Source of mortality information

The primary determination of mortality for eligible participants is based upon matching survey records to the NDI. However, NCHS also relies on additional sources of information to determine the mortality status of a survey participant. Other sources include the Social Security Administration, the Centers for Medicare and Medicaid Services, and death certificates. If a source of mortality, other than a NDI record was available, the participant was considered deceased. Variables indicating which source or sources were used to determine vital status are available on the 2011 Linked Mortality Files. For more information please refer to the Data Dictionary.

VII. 2011 Linked Mortality Files match result variables

To provide the analysts with the opportunity to alter the criteria for determining final vital status, NCHS includes the SCORE and CLASS variables for the best NDI record match on the 2011 Linked Mortality Files. The analyst can conduct sensitivity analyses or take either a more or less conservative approach to vital status ascertainment by setting a different cut-off score within each class and/or determining which classes contain true matches. Researchers interested in how changing score cut-offs can affect the determination of vital status in linked mortality files can refer to the following article: Lariscy JT. Differential record linkage by Hispanic ethnicity and age in linked mortality studies: implications for the epidemiologic paradox. J Aging Health. 2011 23(8):1263-84. More information on the match result variables can be found in the Data Dictionary.

VIII. Statistical issues with linked mortality data

NCHS Linked Mortality Files present unique statistical issues due to the potential for differential follow-up times and potential censoring bias. The paper below may be of interest to analysts as it provides an overview of some of these issues.

> Statistical Issues in Analyzing the NHANES I Epidemiologic Follow-up Study (NHEFS) [PDF - 3 MB]

IX. 1992 NHIS Hispanic oversample

For the 1992 NHIS, the Hispanic population was oversampled. For more information on the 1992 NHIS without the Hispanic oversample, please refer to the supporting documentation from the NHIS: 1992 NHIS "Readme File – Without Hispanic oversample. Analysts must keep this Hispanic oversample in mind if they wish to combine data sets.

X. Important information on merging restricted and public-use NCHS survey data

The data provided on the NCHS 2011 Linked Mortality Files can be merged with the NCHS public-use survey data files using unique survey person identification numbers. However, the identifying variables are different across surveys and years. For guidance on using and merging the appropriate identification numbers please refer to the Data Dictionary.

Suggested citation: National Center for Health Statistics. Office of Analysis and Epidemiology. Analytic Guidelines for NCHS 2011 Linked Mortality Files, *August, 2013*. Hyattsville, Maryland. Available at the following address: http://www.cdc.gov/nchs/data_access/data_linkage/mortality/linkage_methods_analytical_support.htm