National Health Interview Survey (1986-2004)
Linked Mortality Files

Analytic guidelines

The NHIS Linked Mortality Files are a mortality follow-up study of NHIS participants for the survey years 1986-2004 through December 31, 2006 and represent an update of the mortality experience of eligible NHIS participants. The updated NHIS Linked Mortality Files supersede any previous data releases of a NHIS mortality follow-up study.

These guidelines address the following analytical topics:

I. Eligibility status
II. New sample weights
III. Pooled analyses of NHIS years
IV. NHIS participants with an age last known alive of 100 years or greater
V. Inconsistencies in baseline age and follow-up age
VI. Source of mortality information
VII. Special request data file
VIII. 1992 Hispanic oversample

I. Linkage eligibility status

All NHIS participants with sufficient identifying data to create a NDI submission record were eligible for mortality follow-up. However, due to confidentiality concerns, the public-use version of the NHIS Linked Mortality Files only include mortality follow-up for eligible adults. (See Important information about those 17 years and younger).

Eligibility status for mortality follow-up is indicated by the variable ELIGSTAT and for mortality or survival analyses, analysts should keep only those records with a value of ELIGSTAT = 1. For more information please refer to the File Layout and Detailed Notes for Selected Variables.

II. New NHIS eligibility adjusted sample weights

For the 1986 NHIS Linked Mortality File, there are no eligibility adjusted sample weights. NCHS recommends using the public-use annual final basic weight (WTFA). For the analyst's convenience, the variables WGT_NEW have been populated with the values of WFTA on the 1986 NHIS Linked Mortality File.

For the 1987-2004 Linked Mortality Files, 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. The new eligibility adjusted sample weights provided on the NHIS Linked Mortality Files should be used in place of the original NHIS sample weights to prevent biased mortality estimates.

The Restricted-use Linked Mortality Files include three eligibility adjusted sample weights:
- WGT_NEW refers to a person-level record and is available for the NHIS years 1987-2004
- SA_WGT_NEW refers to a sample adult record and is available for the NHIS years 1997-2004
- SC_WGT_NEW refers to a sample child record and is available for the NHIS years 1997-2004

The Public-use Linked Mortality Files include WGT_NEW and SA_WGT_NEW. Due to confidentiality concerns, the Public-use Linked Mortality Files do not include information on NHIS participants 17 years of age and younger.

Technical note:
Treating the eligible sample from the NHIS Linked Mortality Files 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 the NHIS Linked Mortality Files
Analysts may wish to pool several survey years of the NHIS Linked Mortality Files to implement many types of analyses. This section provides guidance on pooling years of the NHIS.

Just as with the NHIS public-use sampling weights, the linkage eligibility adjusted sampling weights created for the NHIS Linked Mortality Files should be adjusted for pooled data; otherwise, estimates of totals will be too high. For example, the estimated total U.S. civilian noninstitutionalized 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.

NHIS has provided analysts with guidance on creating a conceptual-design structure that allows 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, first paragraph of Introduction and Method 1
- NHIS 1995-1996. See http://www.cdc.gov/nchs/data/nhis/96var.pdf, page 20.
- NHIS 1997-2004 See http://www.cdc.gov/nchs/data/nhis/9705var.pdf, starting on page 10.

The following example has been adapted from the NHIS methods documents referenced above. For an analysis covering the NHIS years 1990-2004, you would create new design variables because the design variables have different names and the stratum identifiers have different lengths across the different design periods. Moreover, since there were changes to the public use design variables for 1995-1996 compared to 1997-2004, it is necessary to treat the data years 1995-1996 as if they were statistically independent from the 1997-2004 years.
The location of the design variables (stratum and PSU) on the person level file for each NHIS design period being treated as statistically independent are:

- NHIS 1990-1994. Stratum location 187-188 and PSU location 189.
- NHIS 1995-1996. Stratum location 337-340 and PSU location 352.
- NHIS 1997-2004. Check the variable layouts as file locations for stratum and PSU change from year to year.

First, all original NHIS design variables need to be named STRATUM and PSU. Suppose the names of the new design variables are NSTRATUM (stratum) and NPSU (PSU). One method to create values for NSTRATUM that are of consistent length and take account of the different sample design periods is to do the following:

For the 1990-1994 data, the STRATUM values are 1, 2, ..., 62. First change the stratum values to 001, 002, ..., 062 in order to create a consistent length for the stratum variables. Next, create NSTRATUM for this design period that are distinct from the other design periods by adding 1000 to the stratum values so that NSTRATUM = 1001, 1002, ..., 1062.

For the 1995-1996 data, additional recoding is required to create compatible structures. SAS code is provided below that has been adapted from documentation found in www.cdc.gov/nchs/data/nhis/96var.pdf.

```
STRATUM = INT(STRATUM/10);

IF STRATUM = 100 THEN STRATUM = 99;
```

As a check the analyst should observe 99 strata and 393 PSUs when using the full 1995 data file, and 99 strata and 295 PSUs when using the full 1996 data file.

Then change STRATUM to 001, 002,..., 099 in order to create a consistent length for the stratum variables. Next, create NSTRATUM for this design period that are distinct from the other design periods by adding 2000 to the stratum values so that NSTRATUM = 2001, 2002, , 2099.

For 1997-2004, the STRATUM values are 1, 2, ..., 339. First change the stratum values to 001, 002, ..., 339 in order to create a consistent length for the stratum variables. Next, create NSTRATUM for this design period that are distinct from the other design periods by adding 3000 to the stratum values so that NSTRATUM = 3001, 3002, ..., 3339.

NPSU can be set equal to the PSU value for each year.

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, particularly if using files other than the NHIS person file, i.e. the Sample Adult or supplements. Analysts of the NHIS Linked Mortality files should use computer software that provides the capability of variance estimation and hypothesis testing for complex sample designs.

IV. NHIS participants with an age last known alive of 100 years or greater

The NHIS (1986-2004) Linked Mortality Files include records where the calculated age presumed alive at the end of mortality follow-up (December 31, 2006) is 100 years or greater. 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, analysts may wish to consider these cases as loss to follow-up and make them ineligible for mortality analyses.

V. Inconsistencies in baseline age and follow-up age

Misreporting or discrepancies in reported age at interview or date of birth can result in values for age at death or age last presumed alive that may be inconsistent with baseline age, resulting in negative follow-up time for survival analyses. When this occurs, the number of cases 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 NHIS participants is based upon matching records to the NDI. However, NCHS collects multiple sources of information to determine the final mortality status of a NHIS survey participant. Examples of other sources of mortality include indication of deceased status from the Social Security Administration, the Centers for Medicare and Medicaid Services, or death certificate review. If a source of mortality, other than a NDI record was available, the NHIS participant was considered deceased. Variables indicating which source or sources were used to determine vital status are available on the linked mortality files. More than one source of mortality may be available. For more information please refer to the File Layout and Detailed Notes for Selected Variables.

VII. Special request data file

A special request data file is available that includes additional death certificate data as well as NDI record match results for potential NDI matches that were considered "false" by the probabilistic matching algorithm, thus providing date and cause of death for those NHIS participants NCHS determined to be alive. To provide the analysts with the opportunity to alter the criteria for determining final vital status, NCHS has provided the SCORE and CLASS for the best NDI record match. The analyst can 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. For more information on the implications of using alternate cut-off scores on vital status ascertainment, please refer to the matching methodology document, Appendix B.

VIII. 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 located at the "1992 NHIS "Readme File – Without Hispanic oversample." Analysts must keep this Hispanic oversample in mind if they wish to combine the 1991 and 1992 data sets.

NCHS has created two mortality linkages for the 1992 NHIS: one containing the Hispanic oversample and one which does not include the Hispanic oversample. When accessing the restricted-use files analysts must indicate which 1992 NHIS Linked Mortality file they wish to analyze in their RDC proposal.