1986 – 2000 NHIS Linked Mortality Files: 2004 Restricted Release
Analytic guidelines

The 2004 restricted release of the 1986-2000 NHIS Linked Mortality files represent the *third* linkage of NHIS records to the National Death Index and supersedes the previous NHIS-NDI linkages issued in 1997 and 2000. Due to differences in data editing and NDI matching procedures, there are inconsistencies between these newly created NHIS Linked Mortality files and prior data file releases. No attempt was made to resolve any of these inconsistencies. NCHS recommends that analysts use the new 2004 restricted release of the 1986-2000 NHIS Linked Mortality files.

These analytic guidelines are the most current recommendations from the National Center for Health Statistics (NCHS) and the Data Linkage Program for use with the 2004 restricted release NHIS Linked Mortality Files. These guidelines will be updated on a periodic basis as more information is learned from the analyses of these files.

These guidelines address the following analytical topics:

I. Eligibility status

Only NHIS participants 18 years of age or older at the time of interview were eligible for mortality linkage. Those 17 years or younger at the time of the NHIS interview were ineligible for matching with NDI records as were NHIS participants who had insufficient identifying data to create a NDI submission record. The eligibility status of NHIS participants for mortality follow-up is indicated by the variable *ELIGSTAT*. For mortality or survival analyses, users should keep only the records with a value of *ELIGSTAT = 1* and use the new sample weights, *WGT_NEW*, provided on the file.

II. New adjusted sample weights

For the 2004 NHIS Linked Mortality Files, the NHIS sample weights were adjusted to account for record non-responses for NHIS adults who could not be linked due to insufficient identifying data. The new sample weights (*WGT_NEW)* are provided on the file. Descriptive statistics for the original and adjusted person level sample weights are provided below for NHIS survey years 1986-2000. For more information on the weighting adjustment refer to Guide to Weighting and Variance Estimation in the 1986-2000 NHIS Linked Mortality Files.

Table 1: Description of adjusted Person level survey weights (WGT_NEW) for NHIS respondents (18 years +) considered eligible for mortality follow-up: NHIS survey years 1986-2000

Survey Year	Total Eligible	Minimum	Maximum	Mean	Standard Deviation
1986	43837	349.0	21382.0	3951.4	1105.5
1987	86653	209.0	17838.0	2022.7	718.1
1988	86851	179.0	13765.0	2041.7	628.7
1989	83064	264.0	23704.0	2161.3	790.2
1990	85090	231.0	18803.0	2132.4	751.3
1991	84763	238.0	15152.0	2161.2	722.6
1992	85140	211.0	13024.0	2172.5	711.9
1993	77439	234.0	12038.0	2412.5	919.5
1994	81486	245.0	19638.0	2326.9	806.6
1995	70963	576.0	17086.0	2694.8	1008.4
1996	43892	749.0	23719.0	4394.4	1670.4
1997	67764	0.0	17077.0	2881.6	1069.9
1998	63082	0.0	17003.0	3127.7	1204.3
1999	62137	0.0	17305.0	3212.5	1277.8
2000	64267	0.0	19589.0	3138.4	1274.6

Table 2: Description of original unadjusted Person level sample weight (WTFA) for NHIS respondents (18 years +) considered eligible for mortality follow-up: NHIS survey years 1986-2000

Survey Year	Total Eligible	Minimum	Maximum	Mean	Standard Deviation
1986	43837	345.0	21042.0	3877.8	1084.9
1987	86653	204.0	17540.0	1986.2	705.2
1988	86851	175.0	13486.0	2006.2	617.5
1989	83064	259.0	23388.0	2122.4	775.9
1990	85090	228.0	18688.0	2100.1	740.0
1991	84763	233.0	15067.0	2122.0	710.0
1992	85140	209.0	12690.0	2124.8	696.6
1993	77439	230.0	11694.0	2360.7	900.3
1994	81486	242.0	19244.0	2265.2	785.3
1995	70963	574.0	16636.0	2627.8	983.5
1996	43892	732.0	23220.0	4260.0	1621.9
1997	67764	0.0	16103.0	2667.5	1006.4
1998	63082	0.0	15532.0	2816.5	1105.4
1999	62137	0.0	15454.0	2877.5	1159.4
2000	64267	0.0	18114.0	2812.4	1154.4

III. Variance estimation for the NHIS Linked Mortality files

The data collected in the NHIS are obtained through a complex sample design involving stratification, clustering, and multistage sampling. The use of standard statistical procedures that are based on the assumption that data are generated via simple random sampling (SRS) generally will produce incorrect estimates of variances and standard errors when used to analyze data from the NHIS. For information on the complex NHIS sample design and variance estimation techniques, users should refer to the methods documentation located at http://www.cdc.gov/nchs/about/major/nhis/methods.htm.

Researchers may wish to combine several survey years of the NHIS Linked Mortality files to get reliable estimates for rare health characteristics or mortality outcomes. For the 1986-2000 *NHIS Linked Mortality Files*, the 1986-1994 and 1995-2000 components are based on two independent NHIS designs. Each design has a complex-sampling structure resulting in sampled households being clustered geographically over time, and households/persons being selected with differential

sampling rates. To serve the vast majority of data users of the *NHIS Linked Mortality Files*, a special conceptual-design structure has been created, which is consistent with the true NHIS designs discussed in NCHS (1989) and NCHS (2000). This structure is summarized in the *NHIS Linked Mortality Files* by two new design variables, *STR8600* and *PSU8600*. These new variables represent strata and primary sampling units (PSUs) defined over the 15 years of data.

Users should note that when combining data sets, it is the data users' responsibility to examine the possible changes in the questionnaires as well as variable locations on the data files. For a complete discussion of sample design issues in the NHIS Linked Mortality files and combining multiple files, users should refer to Guide to Weighting and Variance Estimation in the 1986-2000 NHIS Linked Mortality Files.

Users of the NHIS Linked Mortality files should use computer software that provides the capability of variance estimation and hypothesis testing for complex sample designs. Although NCHS, generally, uses Taylor series linearization methods and the SUDAAN software, a summary of available software for the analysis of surveys with complex sample designs can be found at www.fas.harvard.edu/~stats/survey-soft/survey-soft.html. This site includes a selected list of review articles that is not specific to a single software package.

IV. 1992 NHIS Hispanic oversample

For the 1992 NHIS, the Hispanic population was oversampled. The supplemental sample for the 1992 NHIS consists of Hispanic households from the 1991 NHIS, where one or more Hispanic persons 17 years or older were re-contacted in 1992. Hispanic persons from these re-contacted households can be identified in the 1992 NHIS and the 1992 NHIS Linked Mortality file by the code for the processing year variable (the code is "91" for the supplemental Hispanic respondents and "92" for all other respondents). Analysts must keep this Hispanic oversample in mind if they wish to combine the 1991 and 1992 data sets. The Research Data Center (RDC) where the 2004 NHIS Linked Mortality files are available, includes two mortality linkages for the 1992 NHIS - one containing the Hispanic oversample and one with a reduced sample, which does not include the Hispanic oversample. Analysts must indicate which 1992 NHIS Linked Mortality file they wish to analyze in their RDC proposal. Users should note that the appropriate adjusted Person level sample weight (wgt_new) is available for both the original 1992 NHIS and the modified 1992 NHIS without the Hispanic oversample.

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."

V. Altering the criteria for assigning vital status

The 2004 Linked Mortality File includes the NCHS determination for vital status (MORTSTAT) for each eligible NHIS participant. NCHS has also created a special linked mortality file that contains the highest scoring NDI probabilistic match record found for each eligible NHIS participant, including those participants where NCHS determined the NDI potential match record to be false. This file was created to provide researchers with the opportunity to conduct sensitivity analyses by altering the criteria for assignment of vital status. Analysts should refer to the NHIS Linked Mortality files matching methodology, Appendix C for examples of evaluation studies using different criteria to assign vital status for NHIS-NDI match records. Analysts interested in conducting sensitivity analysis should clearly state this request in their RDC application.

Last updated: January 18, 2007