1986 – 2000 NHIS Linked Mortality Files: 2007 Public Release
Analytic guidelines

The 2007 public-use 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 public data file releases. No attempt was made to resolve any of these inconsistencies. NCHS recommends that analysts use the new 2007 public 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). 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. (See Detailed Notes for Selected Variables)

II. New adjusted sample weights

For the 2007 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. 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.

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. To serve the vast majority of data users of the NHIS Linked Mortality Files, two variance estimation variables have been created in the NHIS Linked Mortality Files. These variables *STR_POOL* and *PSU_POOL*, are consistent with the existing NHIS public-use file variance estimation variables and represent strata and primary sampling

units (PSUs) defined over the 15 years of data. 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. In addition, 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.

To assist data users who wish to combine multiple years of NHIS survey data for mortality analysis, two 1992 NHIS public release linked mortality files have been made available for the 1992 NHIS - one containing the Hispanic oversample and one with a reduced sample, which does not include the Hispanic oversample. Researchers must be careful to download the appropriate 1992 NHIS Linked Mortality public-use file they wish to analyze. 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. Calculating correct Age at Interview & Age at Death

The Public-use NHIS files provide a top-coded age at interview and a corresponding bottom coded date of birth. Analysts need to be aware that an accurate age at death cannot be calculated using public-use age at interview or date of birth for survey respondents ages 90 and over in 1996 NHIS and ages 85 or over for NHIS 1997-2000.

NCHS recommends that researchers who require a more precise measure of age at death apply and use the restricted mortality files in the NCHS Research Data Center. These restricted use files contain the non-altered age at death and age at interview for all survey respondents.

VI. ICD coding comparability

The public-use 1986-2000 NHIS Linked Mortality Files contain information on date and cause-of-death that is obtained through the linkage of NHIS participants' information to death certificate records maintained by the National Death Index (NDI). Death certificate cause-of-death information is coded under the Ninth Revision of the *International Classification of Diseases* (ICD-9), for deaths occurring before 1999, and the Tenth Revision (ICD-10), for deaths occurring in 1999 and later. The classification and rule changes between ICD–10 and ICD–9 have resulted in the shifting of deaths away from some underlying cause-of-death categories into others. This has had a considerable impact on the frequency of deaths for some of the major underlying causes and creates discontinuities in cause-of-death trends from 1998 to 1999. Analysts should consider the potential analytic impact of ICD coding differences when conducting cause-specific analyses using the NHIS Linked Mortality files.

NCHS routinely develops grouped recodes for underlying cause-of-death to assist users with the tabulation of underlying causes of death. Since most researchers using the NHIS Linked Mortality Files will conduct analyses that span the entire mortality follow-up period (1986-2002), we have recoded all deaths into the 113 underlying cause-of-death groups used under ICD-10. The ICD-9 and ICD-10 recodes for this list can be found below. This recode provides underlying cause-of-death information, in the same format, for the entire NHIS mortality follow-up period. However, this recode does not control for the transition in death certificate cause-of-death coding rules between ICD-9 and ICD-10.

NCHS has conducted a comparability study[1] to measure the extent of discontinuities in cause-of death trends from 1998 through 1999. The study found that the discontinuity in trend is substantial, particularly for causes of death such as septicemia, influenza and pneumonia, and Alzheimer's disease. Thus, failure to take into account the effects of the implementation of ICD–10 can result in a greatly distorted view of changes in mortality risk for some causes of death. The report provides comparability ratios (CR) for each underlying cause-of-death, which can be applied to adjust the calculation of population mortality rates. However, the CR cannot be applied to individual level mortality risk estimated from regression equations.

We strongly encourage all users of the NHIS Linked Mortality Files, who are examining cause-specific mortality, to refer to the report on the effect of the transition from ICD-9 to

[1] Comparability of Cause of Death Between ICD-9 and ICD-10: Preliminary Estimates

ICD-10 to determine whether a cause-of-death has had a significant discontinuity in classification. Researchers who are examining a cause-of-death, which has a CR that differs substantially from 1.0, should consider confirming results obtained using the underlying cause-of-death 113-group recode for the entire follow-up period with results obtained using mortality data through 1998, prior to the ICD classification and rule changes. In addition, researchers may conduct comparative analyses using the original ICD-9 underlying cause-of-death codes for deaths occurring before 1999. The individual ICD-9 underlying cause-of-death codes as well as the grouped recodes are available on the Restricted-use NHIS Linked Mortality Files available through the NCHS Research Data Center

Last updated: September 5, 2007

Table C. Comparable category numbers for 113 selected causes of death according to the Ninth and Tenth Revisions, *International Classification of Diseases*

List number	Cause of death	Category codes according to ICD–10[1]	Category codes according to ICD–9[1]
001	Salmonella infections	A01–A02	002–003
002	Shigellosis and amebiasis	A03,A06	004,006
003	Certain other intestinal infections	A04,A07–A09	007–009
004	Tuberculosis	A16–A19	010–018
005	Respiratory tuberculosis	A16	010–012
006	Other tuberculosis	A17–A19	013–018
007	Whooping cough	A37	033
008	Scarlet fever and erysipelas	A38,A46	034.1–035
009	Meningococcal infection	A39	036
010	Septicemia	A40–A41	038
011	Syphilis	A50–A53	090–097
012	Acute poliomyelitis	A80	045
013	Arthropod–borne viral encephalitis	A83–A84,A85.2	062–064
014	Measles	B05	055
015	Viral hepatitis	B15–B19	070
016	Human immunodeficiency virus (HIV) disease	B20–B24	042–044
017	Malaria	B50–B54	084
018	Other and unspecified infectious and parasitic diseases and their sequelae	A00,A05,A20–A36,A42–A44,A48–A49,A54–A79,A81–A82,A85.0–A85.1,A85.8,A86–B04,B06–B09,B25–B49,B55–B99	001,005,020–032,037,039–041,046–054,056–061,065–066,071–083,085–088,098–134,136,139,771.3
019	Malignant neoplasms	C00–C97	140–208
020	Malignant neoplasms of lip, oral cavity and pharynx	C00–C14	140–149
021	Malignant neoplasm of esophagus	C15	150
022	Malignant neoplasm of stomach	C16	151
023	Malignant neoplasms of colon, rectum and anus	C18–C21	153–154
024	Malignant neoplasms of liver and intrahepatic bile ducts	C22	155
025	Malignant neoplasm of pancreas	C25	157
026	Malignant neoplasm of larynx	C32	161
027	Malignant neoplasms of trachea, bronchus and lung	C33–C34	162
028	Malignant melanoma of skin	C43	172
029	Malignant neoplasm of breast	C50	174–175
030	Malignant neoplasm of cervix uteri	C53	180
031	Malignant neoplasms of corpus uteri and uterus, part unspecified	C54–C55	179,182
032	Malignant neoplasm of ovary	C56	183.0
033	Malignant neoplasm of prostate	C61	185
034	Malignant neoplasms of kidney and renal pelvis	C64–C65	189.0,189.1
035	Malignant neoplasm of bladder	C67	188
036	Malignant neoplasms of meninges, brain and other parts of central nervous system	C70–C72	191–192
037	Malignant neoplasms of lymphoid, hematopoietic and related tissue	C81–C96	200–208
038	Hodgkin's disease	C81	201
039	Non-Hodgkin's lymphoma	C82–C85	200,202
040	Leukemia	C91–C95	204–208
041	Multiple myeloma and immunoproliferative neoplasms	C88,C90	203
042	Other and unspecified malignant neoplasms of lymphoid, hematopoietic and related tissue	C96	- - -

See footnotes at end of table.

Table C. Comparable category numbers for 113 selected causes of death according to the Ninth and Tenth Revisions, *International Classification of Diseases*—Con.

List number	Cause of death	Category codes according to ICD–10[1]	Category codes according to ICD–9[1]
043	All other and unspecified malignant neoplasms	C17,C23–C24,C26–C31,C37–C41,C44–C49, C51–C52,C57–C60,C62–C63,C66,C68–C69,C73–C80,C97	152,156,158–160,163–171,173,181,183.2–184,186–187,189.2–190,193–199
044	In situ neoplasms, benign neoplasms and neoplasms of uncertain or unknown behavior	D00–D48	210–239
045	Anemias	D50–D64	280–285
046	Diabetes mellitus	E10–E14	250
047	Nutritional deficiencies	E40–E64	260–269
048	Malnutrition	E40–E46	260–263
049	Other nutritional deficiencies	E50–E64	264–269
050	Meningitis	G00,G03	320–322
051	Parkinson's disease	G20–G21	332
052	Alzheimer's disease	G30	331.0
053	Major cardiovascular diseases	I00–I78	390–434,436–448
054	Diseases of heart	I00–I09,I11,I13,I20–I51	390–398,402,404,410–429
055	Acute rheumatic fever and chronic rheumatic heart diseases	I00–I09	390–398
056	Hypertensive heart disease	I11	402
057	Hypertensive heart and renal disease	I13	404
058	Ischemic heart diseases	I20–I25	410–414,429.2
059	Acute myocardial infarction	I21–I22	410
060	Other acute ischemic heart diseases	I24	411
061	Other forms of chronic ischemic heart disease	I20,I25	412–414,429.2
062	Atherosclerotic cardiovascular disease, so described	I25.0	429.2
063	All other forms of chronic ischemic heart disease	I20,I25.1–I25.9	412–414
064	Other heart diseases	I26–I51	415–429.1,429.3–429.9
065	Acute and subacute endocarditis	I33	421
066	Diseases of pericardium and acute myocarditis	I30–I31,I40	420,422–423
067	Heart failure	I50	428
068	All other forms of heart disease	I26–I28,I34–I38,I42–I49,I51	415–417,424–427,429.0–429.1,429.3–429.9
069	Essential (primary) hypertension and hypertensive renal disease	I10,I12	401,403
070	Cerebrovascular diseases	I60–I69	430–434,436–438
071	Atherosclerosis	I70	440
072	Other diseases of circulatory system	I71–I78	441–448
073	Aortic aneurysm and dissection	I71	441
074	Other diseases of arteries, arterioles and capillaries	I72–I78	442–448
075	Other disorders of circulatory system	I80–I99	451–459
076	Influenza and pneumonia	J10–J18	480–487
077	Influenza	J10–J11	487
078	Pneumonia	J12–J18	480–486
079	Other acute lower respiratory infections	J20–J22	466
080	Acute bronchitis and bronchiolitis	J20–J21	466
081	Unspecified acute lower respiratory infection	J22	- - -
082	Chronic lower respiratory diseases	J40–J47	490–494,496
083	Bronchitis, chronic and unspecified	J40–J42	490–491
084	Emphysema	J43	492
085	Asthma	J45–J46	493
086	Other chronic lower respiratory diseases	J44,J47	494,496
087	Pneumoconioses and chemical effects	J60–J66,J68	500–506

See footnotes at end of table.

Table C. Comparable category numbers for 113 selected causes of death according to the Ninth and Tenth Revisions, *International Classification of Diseases*—Con.

List number	Cause of death	Category codes according to ICD–10[1]	Category codes according to ICD–9[1]
088	Pneumonitis due to solids and liquids	J69	507
089	Other diseases of respiratory system	J00–J06,J30–J39,J67,J70–J98	034.0,460–465,470–478,495,508–519
090	Peptic ulcer	K25–K28	531–534
091	Diseases of appendix	K35–K38	540–543
092	Hernia	K40–K46	550–553
093	Chronic liver disease and cirrhosis	K70,K73–K74	571
094	Alcoholic liver disease	K70	571.0–571.3
095	Other chronic liver disease and cirrhosis	K73–K74	571.4–571.9
096	Cholelithiasis and other disorders of gallbladder	K80–K82	574–575
097	Nephritis, nephrotic syndrome and nephrosis	N00–N07,N17–N19,N25–N27	580–589
098	Acute and rapidly progressive nephritic and nephrotic syndrome	N00–N01,N04	580–581
099	Chronic glomerulonephritis, nephritis and nephropathy not specified as acute or chronic, and renal sclerosis unspecified	N02–N03,N05–N07,N26	582–583,587
100	Renal failure	N17–N19	584–586
101	Other disorders of kidney	N25,N27	588–589
102	Infections of kidney	N10–N12,N13.6,N15.1	590
103	Hyperplasia of prostate	N40	600
104	Inflammatory diseases of female pelvic organs	N70–N76	614–616
105	Pregnancy, childbirth and the puerperium	O00–099	630–676
106	Pregnancy with abortive outcome	O00–O07	630–639
107	Other complications of pregnancy, childbirth and the puerperium	O10–O99	640–676
108	Certain conditions originating in the perinatal period	P00–P96	760–771.2,771.4–779
109	Congenital malformations, deformations, and chromosomal abnormalities	Q00–Q99	740–759
110	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified	R00–R99	780–799
111	All other diseases (Residual)	Residual	Residual
112	Accidents (unintentional injuries)	V01–X59,Y85–Y86	E800–E869,E880–E929
113	Transport accidents	V01–V99,Y85	E800–E848,E929.0,E929.1
114	Motor vehicle accidents	V02–V04,V09.0,V09.2,,V12–V14,V19.0–V19.2,V19.4–V19.6,V20–V79,V80.3–V80.5,V81.0–V81.1,V82.0–V82.1,V83–V86,V87.0–V87.8,V88.0–V88.8,V89.0,V89.2	E810–E825
115	Other land transport accidents	V01,V05–V06,V09.1,V09.3–V09.9,V10–V11,V15–V18,V19.3,V19.8–V19.9,V80.0–V80.2,V80.6–V80.9,V81.2–V81.9,V82.2–V82.9,V87.9,V88.9,V89.1,V89.3,V89.9	E800–E807,E826–E829
116	Water, air and space, and other and unspecified transport accidents and their sequelae	V90–V99,Y85	E830–E848,E929.0,E929.1
117	Nontransport accidents	W00–X59,Y86	E850–E869,E880–E928,E929.2–E929.9
118	Falls	W00–W19	E880–E888
119	Accidental discharge of firearms	W32–W34	E922
120	Accidental drowning and submersion	W65–W74	E910
121	Accidental exposure to smoke, fire and flames	X00–X09	E890–E899
122	Accidental poisoning and exposure to noxious substances	X40–X49	E850–E869,E924.1
123	Other and unspecified nontransport accidents and their sequelae	W20–W31,W35–W64,W75–W99,X10–X39,X50–X59,Y86	E900–E909,E911–E921,E923–E924.0,E924.8–E928,E929.2–E929.9
124	Intentional self-harm (suicide)	X60–X84,Y87.0	E950–E959
125	Intentional self-harm (suicide) by discharge of firearms	X72–X74	E955.0–E955.4

See footnotes at end of table.

Table C. Comparable category numbers for 113 selected causes of death according to the Ninth and Tenth Revisions, *International Classification of Diseases*—Con.

List number	Cause of death	Category codes according to ICD–10[1]	Category codes according to ICD–9[1]
126	Intentional self-harm (suicide) by other and unspecified means and their sequelae	X60–X71,X75–X84,Y87.0	E950–E954,E955.5–E959
127	Assault (homicide)	X85–Y09,Y87.1	E960–E969
128	Assault (homicide) by discharge of firearms	X93–X95	E965.0–E965.4
129	Assault (homicide) by other and unspecified means and their sequelae	X85–X92,X96–Y09,Y87.1	E960–E964,E965.5–E969
130	Legal intervention	Y35,Y89.0	E970–E978
131	Events of undetermined intent	Y10–Y34,Y87.2,Y89.9	E980–E989
132	Discharge of firearms, undetermined intent	Y22–Y24	E985.0–E985.4
133	Other and unspecified events of undetermined intent and their sequelae	Y10–Y21,Y25–Y34,Y87.2,Y89.9	E980–E984,E985.5–E989
134	Operations of war and their sequelae	Y36,Y89.1	E990–E999
135	Complications of medical and surgical care	Y40–Y84,Y88	E870–E879,E930–E949

- - - No comparable category classified by ICD–9 exists.

[1]ICD–10 is *International Classification Diseases*, Tenth Revision, and ICD–9 is *International Classification of Diseases*, Ninth Revision.