The 1986-2000 National Health Interview Survey Linked Mortality Files: Matching Methodology

Suggested citation: National Center for Health Statistics. Office of Analysis and Epidemiology, The 1986-2000 National Health Interview Survey Linked Mortality Files: Matching Methodology, October 2005. Hyattsville, Maryland. (Available at the following address: http://www.cdc.gov/nchs/data/datalinkage/matching_methodology_nhis_final.pdf)

Introduction

Record linkage refers to the bringing together, from two or more independent files, records believed to relate to the same "entity", e.g. people or families (Fair and Whitridge 1997 – Record Linkage Techniques: Proceedings of an International Workshop and Exposition). The NHIS Linked Mortality files link 15 years of records of adult participants (age 18 years and older) in the National Health Interview Survey (NHIS) with death records from the National Death Index (NDI). The NHIS is the principal source of information on the health of the civilian, non-institutionalized population of the United States and has been conducted continuously since 1957. For detailed information on the NHIS's contents and methods, refer to the NHIS web page. The NDI is a central computerized data base of all certified deaths in the United States since 1979. For detailed information on the NDI's contents and methods, refer to the National Death Index (NDI).

The 2004 restricted release of the NHIS Linked Mortality files represents the *third* linkage of NHIS records linked to the NDI and supersedes the previous NHIS-NDI linkages issued in 1997 and 2000. The NHIS Linked Mortality files are available for each NHIS survey year from 1986 through 2000, with mortality follow-up from the NHIS adult participant's date of interview through December 31, 2002. This new resource allows researchers the opportunity to use the richness of the NHIS questionnaires to examine the association of a variety of health and health care factors with mortality.

Overview

The linking of NHIS and NDI records was conducted entirely by probabilistic matching, meaning that no death certificate verification is undertaken for NHIS records that return NDI record matches. A matching methodology was employed for the NHIS Linked Mortality files that was similar, but not identical, to the standard methodology offered by the NDI. Specifically, NCHS developed new weights associated with the specific value of each identifying element on the submission record to create scores for potential matches and implemented a more restrictive criteria for identifying potential matches than the NDI standard approach (see Step 3 below). Also, NCHS conducted a new calibration study to establish the cut-off scores for determining whether a NDI match is considered a true match or a false match (Step 4 below).

This document explains the matching methodology NCHS employed to link NHIS records to death records in the NDI. In this document, data users will find detailed information on the following steps involved in the NHIS-NDI match process.

1. Creating NDI submission records from NHIS respondent records
2. Selecting potential matches between NHIS and NDI records
   - Selection is based upon 7 different criteria
   - Selection creates a pool of potential matches
3. Scoring and classifying potential matches
   - Scores are based upon weights associated with the *values* of each identifying data item
   - Classes are based upon *which* identifying data items match
4. Determining final match status and assigning vital status

Figure 1 depicts the NDI matching process for the NHIS surveys. Users interested in a detailed description of the standard NDI matching methodology should refer to the National Death Index (NDI).

NHIS-NDI record linkage

Only NHIS participants 18 years of age or older at the time of interview, with sufficient identifying information, were eligible for matching with NDI records. A second group of NHIS participants was ineligible because of insufficient identifying data to create a submission record, resulting in unknown (missing) vital status for these adult NHIS respondents. As ignoring these ineligibles might lead to biased mortality analysis, NCHS developed new sample weights that are adjusted to account for these non-responses. These weights are available on the NHIS Linked Mortality files. Users should note the percentage of NHIS respondents 18 years and older who are ineligible due to insufficient data markedly increased in 1997 compared with prior years (see Tables 1 and 3 in tabular data). For a description of how the new sample weights were created, please refer to Guide to Weighting and Variance Estimation in the 1986-2000 NHIS Linked Mortality Files.

1. Submission Records

NCHS prepared a base submission record for the NDI for each eligible NHIS participant for the years 1986 – 2000[1]. Each NHIS submission record contained up to 12 identifying data items (see below). The NHIS routinely collects all of the 12 data items used by the NDI for matching and has essentially 100 percent complete reporting of these items except for social security number (SSN) and middle initial (see Table 3 in Tabular data). In general, there is no attempt to recontact NHIS participants, but NCHS does update and edit identifying information[2]. Names collected during the NHIS interview such as "John Doe", "Female #1", and "Person" are deleted as non-valid names. Names containing obvious keystroke errors such as "Jo9hn" are corrected.

Data items on the NHIS submission record

1. Social Security Number
2. First name
3. Middle initial
4. Last name (or birth surname)
5. Month of birth
6. Day of birth
7. Year of birth
8. Sex
9. State of birth
10. Race
11. State of residence
12. Marital status

[1] As previously noted, there have been two prior NHIS-NDI linkages. For the most recent NHIS-NDI linkage, all eligible NHIS survey participant records were submitted to the NDI, not just those that previously had not matched.
[2] For example, selected NHIS years have been linked to data from the Center for Medicare and Medicaid Services that can result in updated or corrected data for name, SSN, and date of birth.

In addition to the base submission record, the NDI allows multiple alternate submission records for each eligible NHIS participant. In order to increase the chances for selection of the correct death record, NCHS generated alternate submission records, e.g. when identification data was questionable or when the NHIS participant had a multi-part name. For a detailed description of the rules NCHS used to generate alternate NHIS submission records, refer to Appendix A.

Before the NDI processes any submission record, each record is screened to determine if it contains at least one of the following combinations of identifying data elements:

1. Social Security number, sex, full date of birth present
2. Last name, first initial, month of birth, year of birth present
3. Last name, first initial, Social Security number present

Any submission record that did not meet these minimum data requirements was ineligible for record linkage. Table 1 in Tabular Data (column 5) shows the number of NHIS participants who were ineligible because they did not meet the minimum NDI data requirements. All accepted NHIS submission records are further edited by the NDI system to provide a consistent format for identifying data elements before performing the NDI record search and retrieval process. For example, the NDI editing process converts text to all upper case and removes suffixes from last names. Also, since spelling variants of names are common, NDI codes last names based on the way a name sounds rather than how it is spelled[3]. For example, records with last names Smith and Smyth receive equivalent NYSIIS codes and both would be selected as a potential match for a NHIS submission with Smith (or Smyth) as a last name.

2. Selecting NHIS-NDI potential match records

The NDI system selects death record matches based on a set of established match criteria. The seven criteria listed below were the criteria in use at the time of the most recent NHIS-NDI match.

1. Social Security Number
2. First and last name, exact month of birth, year of birth within 1 year
3. Last name, first initial and middle initial, exact month of birth, year of birth within 1 year
4. First and last name, exact month of birth, exact day of birth
5. Last name, first initial and middle initial, exact month of birth, exact day of birth
6. First name, father's surname, exact month of birth, exact year of birth
7. For females only, first name, exact month and year of birth, and last name from the user's record matching birth surname on the NDI record (for females who change their name after marriage, but don't supply a birth surname)

Any NDI record that matches a NHIS submission record on any one of these seven criteria is selected. As one or more NDI records may be matched to a given NHIS record, the NDI record selection process can return several hundred *potential* matches for each NHIS person, many of which will be non-matches or duplicate records.

[3] The sound alike system is a variation of the New York State Identification Intelligence System or NYSIIS, which converts a name to a phonetic coding.

3. Scoring and classifying potential match records

Assessing the quality of potential matches and determining the best match for each NHIS participant requires a consistent approach. The matching methodology begins by assigning probabilistic scores for each potential match. The score is the sum of the weights assigned to each of the identifying data items used in the NHIS-NDI record match, where the weights reflect the degree of agreement between the information on the NHIS submission record and the NDI death record. NCHS developed the weights, known as binit weights, based upon the frequency of occurrence of the 12 data items in the NDI files for years 1979 to 2000, which represents about 49 million persons. The weights correspond to $[\text{Log}_2 (1/p_i)]$: the base 2 logarithm of the inverse of the probability of occurrence of the value of the identifying data item on the submission record.

- Social Security number – each digit in each position of the SSN (1 to 9) has a corresponding binit weight, with the total SSN weight being the sum of the weights for each of the nine digits. For a record to receive the total SSN binit weight at least 8 digits need to agree. If seven digits agree, then 7/9 of the total weight is assigned. If fewer than seven digits agree then the total SSN weight becomes negative.
- Name – a common first name, such as "John", that has a higher probability of occurrence has a lower binit weight than an uncommon name, such as "Jonas". First name weights are stratified by both sex and year of birth since first names are sex specific and the popularity of first names varies over time. Weights for first and last names are limited to a finite set of values and any name not appearing in the set (meaning it is less common) receives the maximum value of the weight.

Weights are either positive or negative. If there is agreement between the NHIS record and the NDI record for a particular identifying data item, the weight is positive. If there is no agreement, the weight is negative. Some items, such as year of birth, allow a tolerance (+/- 3 years) and are still considered to agree. With the exception of middle initial, data items that are missing on the NHIS submission record, the NDI record, or both receive a weight of zero. A blank middle initial is considered a valid value and receives the appropriate weight. The score for each potential match is the sum of the weights for each individual data item.

$$\textbf{\textit{Score}} = \{\sum W_{SSN1} + \ldots + W_{SSN9}{}^{4}\} + W_{firstname\ x\ sex\ x\ birthyear} + W_{middleinitial\ x\ sex} + W_{lastname} + W_{race} + W_{sex} + W_{maritalstatus\ x\ sex\ x\ age} + W_{birthdate} + W_{birthmonth} + W_{birthyear} + W_{stateofbirth} + W_{stateof\ residence}$$

After scoring the potential matches, each is categorized into one of five mutually exclusive classes[5]. Whereas weighting and scoring take into account the probability that the NHIS record and the NDI record share a particular value for the identifying items, the classes take into account which identifying items agree. They reflect the fact that some of the 12 NDI identifying

[4] For a record to be assigned the maximum weight for SSN, there needs to be agreement on at least 8 digits. If seven digits agree, then 7/9 of the total weight is assigned. If fewer than seven digits agree then the total SSN weight becomes negative.

[5] All total scores were adjusted to reflect the final class code for the potential matches. For example, any record that was switched from Class 5 to Class 3 had its score adjusted to reflect that SSN is missing, with the value of 0 assigned to SSN.

items are more important for determining true matches than others. For example, as SSN is a key identifier in the matching process, each NHIS-NDI record match is initially classified according to whether SSN is present and agrees (Class 1 or 2), is present but disagrees (Class 5) or is unknown (Class 3 or 4). Additionally, non-changing identifying information is more important than information that can change over time. Many women, for example, assume their spouse's name at marriage, a common example of legitimate change over time. Birth surname, however, does not change and is thus an important matching variable for women. By contrast, state of residence and marital status may change between the NHIS interview date and the date of death and are, therefore, less important as matching variables.

The final five classes used by NCHS for the NHIS Linked Mortality files are as follows:

Class 1: Agrees on at least 8 (of 9) digits of SSN, first name (including NYSIIS match), middle initial (including blank), last name (including NYSIIS match), birth year (+/- 3 years), birth month, sex, and state of birth.

Class 2: Agrees on at least 7 (of 9) digits of SSN and at least 5 more of the following items: first name (including NYSIIS match), middle initial (including blank), last name (including NYSIIS match), birth year (+/- 3 years), birth month, sex, and state of birth.

Class 3: There are two types of Class 3 matches:

Type A: SSN is unknown, but last name matches (including NYSIIS match) and at least 7 of the following items agree: first name (including NYSIIS match), middle initial (including blank), last name (including NYSIIS match), birth year (+/- 3 years), birth month, sex, and state of birth.

Type B: Records in this category were initially put in Class 5 but switched to Class 3 because SSN is known but 3 or more digits do not agree[6]. Additionally, at least 8 of the following items agree: first name (including NYSIIS match), middle initial (including blank), last name (including NYSIIS match), birth year (+/- 3 years), birth month, sex, race, marital status, and state of birth. Last name and sex must agree.

Class 4: SSN is unknown on either the NHIS submission record or the NDI record and fewer than 8 of the items listed in Class 3 match.

Class 5: SSN is present but fewer than 7 (of 9) digits on SSN agree or at least 7 digits on SSN agree but fewer than 5 of the following items agree: first name (including NYSIIS match), middle initial (including blank), last name (including NYSIIS match), birth year (+/- 3 years), birth month, sex, and state of birth.

4: Selecting matches and assigning vital status

[6] This class switch occurs if after review, there is the possibility that SSN was either recorded incorrectly or that the spouse's SSN was recorded instead of the subject's SSN.

As already described in Section 2, each eligible NHIS participant may have multiple submission records and each submission record may return one or more matches to a NDI record. The NHIS Linked Mortality files do NOT include all of the potential matching NDI records. Rather, for those NHIS participants with a potential match to the NDI, NCHS employed a strategy to provide the single best NDI match record for inclusion on the linked NHIS mortality file.

First, NHIS-NDI potential match records that had a date of death prior to the date of interview, a score of zero or less, or final categorization of Class 5 were considered false matches and eliminated from the pool of potential matches. Next, among the remaining pool of potential matches, match records that referred to the same death certificate (duplicates) were eliminated. This process reduced the pool of potential matches by 85% to 90%. Many participants, however, still had more than one NDI record as a potential match. The remaining potential matches for each eligible NHIS participant were ranked first on class (from 1 to 4) and then within class by highest score. The NDI match with the highest score within the best class was selected as the one match for inclusion on the mortality file. In the event of a tie among NDI record matches for a particular NHIS record, the tiebreaker reflected the importance of matching items[7]. NCHS then determined whether each best record match was true or false.

A true match reflects *both* the correct vital status of the survey participant and a match to the correct death certificate data. All Class 1 match records were considered true matches. For match records with Classes 2, 3, and 4, NCHS determined whether the match was true or false using cut-off scores developed from the NHANES I Epidemiologic Follow-Up Study (NHEFS) calibration sample, which has verified mortality outcomes for its sample. Within each class, matches with a score *greater than or equal* to the cut-off score were considered true matches, while records with a score less than the cut-off were considered false matches. *The cut-off scores for Classes 2, 3, and 4 were 47, 45, and 40, respectively.* In general, the process was to select the cut-off scores within Classes 2, 3, and 4 that simultaneously maximized the proportion of people correctly classified and minimized the number of people incorrectly classified, with particular attention given to minimizing the number of false positives. Users should refer to Appendix B for a description of the results of the calibration study.

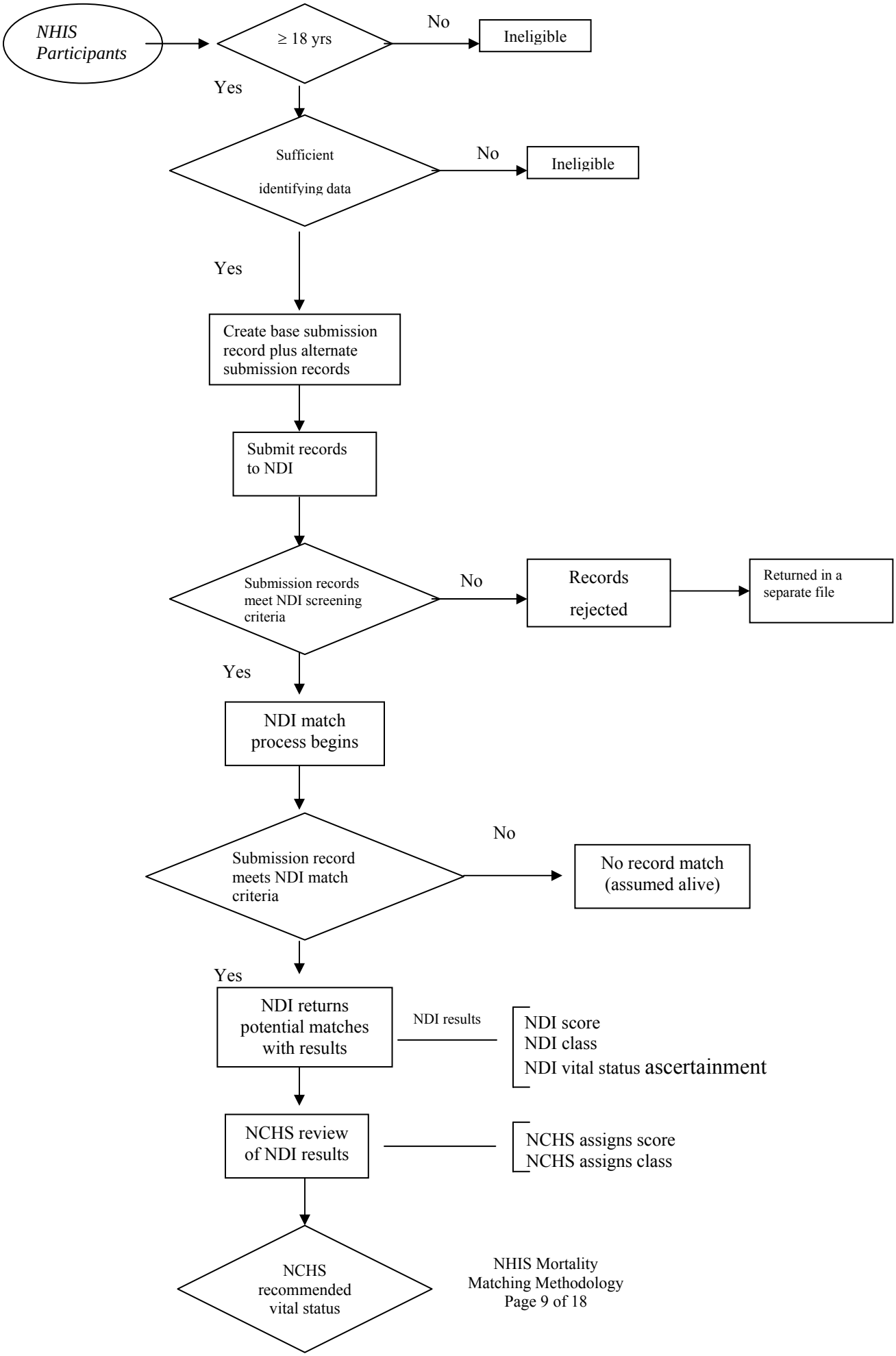
[7] The order is: number of digits of SSN; sex; last name; first name; state of birth; year of birth; month of birth; day of birth. If all of these are the same, then a random number is used.

Final Notice to Users

Eligible NHIS participants with a "true" NDI record match are assumed to be dead. Eligible NHIS participants with no potential NDI record match as well as those with a NDI record match that is considered a "false" match are assumed to be alive. Ineligible NHIS respondents should be excluded from mortality analyses. Analysts should use the variable MORTSTAT to determine vital status.

A data file with the additional probabilistic NHIS-NDI match results is available by request. This special request file differs from the current file in that not every NHIS participant with a NDI record is considered deceased. The special request file includes NDI record match results for potential NDI matches that were considered "false" by the probabilistic matching algorithm. NCHS has provided the SCORE and CLASS for the best NDI record match, regardless of the final assigned vital status, to provide the user with the opportunity to alter the criteria for determining final match status. The user 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 see Appendix C.

Figure 1: NHIS-NDI Matching Process



Appendix A

Creating Alternate Submission Records

The primary purpose of using alternate submission records is to increase the chances of returning a correct death record for those NHIS participants who are, in fact, deceased. The NDI allows multiple alternate submission records for each survey person. Rules for creating alternate NDI submission records were based upon a calibration study using the NHANES I Epidemiologic Follow-up Study (NHEFS). The NHEFS calibration study has a sample of 12,699 people whose vital status is known for a definite time period beginning January 1979 through either the date of death for decedents or a final interview date for non-decedents. NCHS created base submission records for this sample and submitted them to the NDI record retrieval process. For those known to be deceased but who did not return an NDI record match, NCHS compared the identifying information on the submission record to the information on the death certificate. The process revealed the most common reasons a NDI record was not returned.

Name inaccuracies are the most common type of mismatch error encountered when matching to the NDI system. Since death certificates are official records, they will list the full proper name of the decedent. However, survey respondents may provide nicknames or middle names as their first names. To account for nicknames being listed as the first name, NCHS used a nickname to proper name conversion process that created alternate submission records with the most popular formal name associated with that nickname. For example, if a NHIS record listed the respondent name as Beth, two submission records were created. The base submission record included Beth as the first name and the alternate submission record included Elizabeth as the first name.

Multipart first or last names also increase the chances of a NHIS and NDI record not matching. Such differences in name reporting are particularly common for the U.S. Hispanic population. For example, mother's and/or father's surname may both be reported as two last names in a particular order during the survey contact but may be reversed on the death record. To take into account potential recording discrepancies caused by multi-part names, alternate records were created using all of the components of multi-part names both separately and together. Only names with either a space or hyphen are treated as multipart names.

Middle initial plays an important role in NDI matching. Since the NDI allows a blank as a valid value for middle initial, an alternate record is created by dropping the middle initial from any base submission record where it is non-blank.[1] NCHS believes converting non-blank middle initials to blank to be a useful alternate submission strategy because a comparison of where middle initial is missing on either the survey data record or the correct matching NDI record showed that in only 50% of the cases middle initial was missing on both records.

In summary, for the NHIS-NDI linkage, the following rules were used for generating alternate submission records:

1. Use proper name in place of nickname for first name

[1] Preliminary research performed at NCHS has found that many survey data files include a blank middle initial about 25% of the time, making blank the single most commonly reported middle initial.

2. Multipart first and last names are submitted as is, and alternately each part of the name is submitted as the first or last name
3. Switch first name and middle name
4. Blank out middle name
5. Add alternate surnames when evidence of legal name change is available
6. When Medicare number is available, substitute Medicare number for Social Security Number (SSN) (note: some subjects have a Medicare number but no SSN)
7. Use alternate birth date or SSN data, if collected
8. If month of birth is missing, submit twelve records, one with each month

The rules for alternate submission record creation are multiplicative in nature. For example, a participant may have both an imputed month of birth (12 separate records) and two-part first name (3 separate records) resulting in 36 NDI submission records.

Appendix B

Evaluation of vital status ascertainment for the NDI record selection and match process

1. NHANES I Epidemiologic Follow-up Calibration Sample

Since the NDI record selection and match processes do not have an independent means of assessing whether a NHIS-NDI match is true or false, NCHS undertook a calibration study to determine the adequacy of the probabilistic approach utilized to match NHIS participants to NDI records. Such a study is necessary in order to assess the number of false negatives and false positives.

With regard to false negatives, there are several ways the death of a NHIS participant could be missed. Some of these ways are due to the universe of deaths in the NDI, some to the NDI selection process and some to the ranking, scoring and classification of matches employed by NCHS in the NHIS-NDI linkage (see sections 3 and 4 of the main document). Specifically, there are five ways a death could be missed in the NHIS mortality files:

- Deaths outside the United States are not included in the NDI database;
- A small number of deaths occurring in the U.S. are not part of the NDI database;
- Deaths not retrieved in the NDI record selection process due to lack of agreement of identification data;
- True deaths retrieved in the NDI record selection process are dropped from the pool of potential matches because they are not the top ranked death record by NCHS;
- True deaths retrieved in the NDI record selection process are assigned a score below the threshold for determining a match a true match.

False positives often arise by finding a match for a relative or someone with a common name. A small number of false positives also occur when true decedents are matched to the wrong NDI record. Although these individuals are assigned the correct vital status, as the death record is wrong, the date and cause of death are unlikely to be correct.

The calibration study used the NHANES I Epidemiologic Follow-up survey (also known as NHEFS), which was conducted from 1971-1992. NHEFS provides a unique opportunity to assess the quality of the NHIS-NDI matching process because it is a longitudinal study with a high participation rate and highly complete and verified identification data. In the NHEFS sample, there are 12,699 people for whom active follow-up was conducted so that their vital status was known beginning January 1979[1] through either the date of death or a final interview date (for non-decedents). In this sample, four deaths occurred outside the United States, leaving 3,454 deaths that were available to be included in the NDI database and for which a match to a NHEFS participant was possible. NCHS applied the same approach for creating submission records, selecting NDI records, and ranking, scoring, and classifying matches to the NHEFS sample as for the NHIS-NDI linkage to determine how many of the 3,454 deaths could be found.

[1] The NDI was established in 1979. Persons in the NHEFS sample who died before 1979 were not considered in this study.

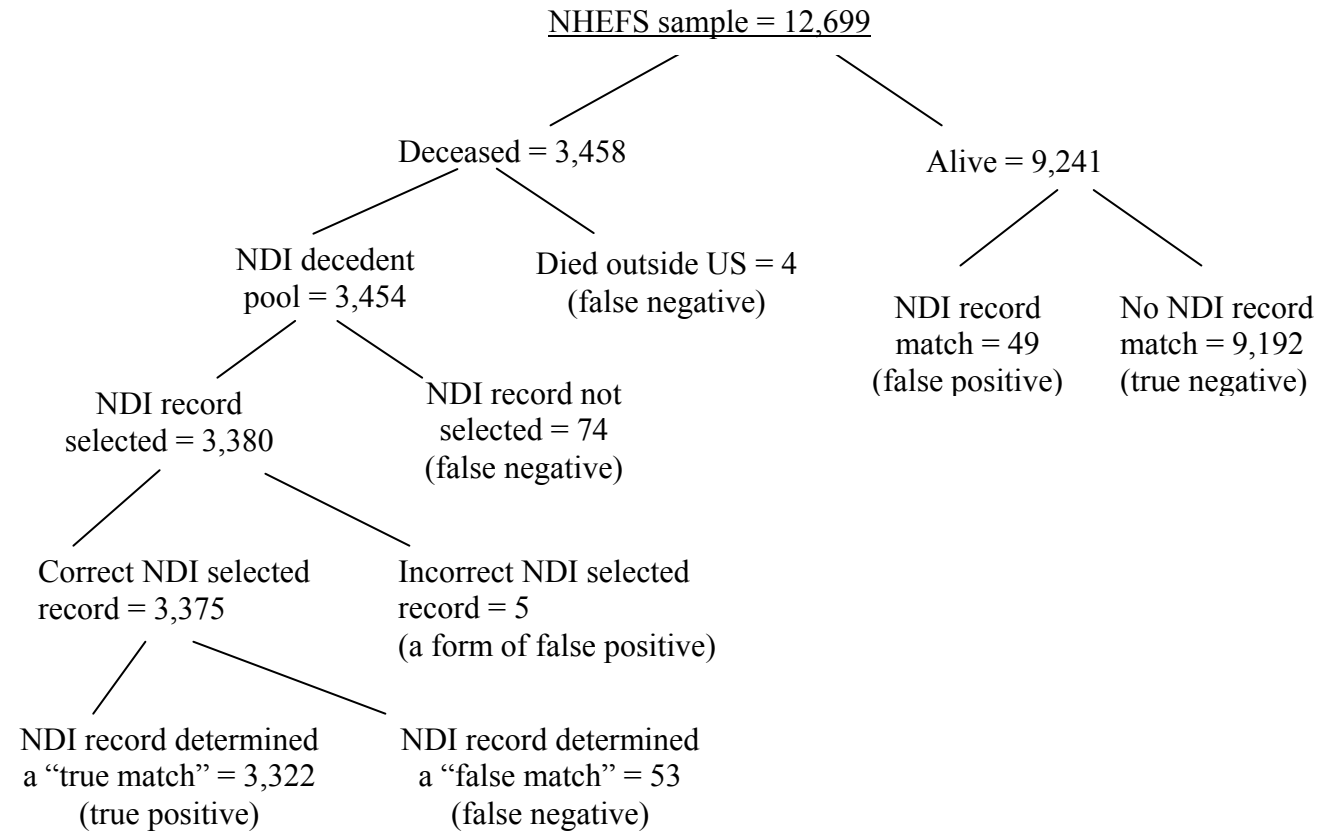
Figure 2 depicts the selection process and match status determination of the NHEFS sample. Among the 3,454 NHEFS decedents, 3,380 had a NDI record selected as a potential match and 74 did not. Among the NDI potential matches for the NHEFS decedents, 3,375 had the *correct* NDI record selected. Using the cut-off scores for Classes 2, 3, and 4 as described in section 4 of the documentation, resulted in 3,322 being considered true matches and correctly assigned as deceased, whereas 53 were considered false matches and incorrectly assigned a vital status as alive. Additionally, there were 79 NHEFS decedents who did not have a true match to a NDI record - 5 were decedents who were assigned as dead, but because their NDI record match is to the wrong person, the date and cause of death will not be correct (a form of false positive) and 74 were decedents who did not return a NDI record and were incorrectly assigned a vital status of alive. Among NHEFS non-decedents, 49 returned a NDI record that was selected as a true match and were incorrectly assigned a vital status of deceased.

Table 1 shows the cut-off scores for Classes 2, 3, and 4 employed to determine the match status of NDI potential matches to NHEFS records and the proportion correctly classified. Based upon the matching methodology described for the NHIS Linked Mortality files, across all four classes, 96.1% of NHEFS decedents were correctly classified as deceased and matched to the correct death certificate, 99.4% of non-decedents were correctly classified as alive, with an overall 98.5% of NHEFS respondents correctly classified.

Table 1:Cut-off scores and proportion of NHEFS subjects correctly classified.

Within Class	Cut-off Score	Correctly classified overall (%)	Correctly classified as dead (%)	Correctly classified as alive (%)
2	≥ 47	98.0	98.5	40.0
3	≥ 45	89.7	94.7	67.5
4	≥ 40	98.6	60.5	99.4
Total across classes	**-----**	**98.5**	**96.1**	**99.4**

Figure 2.



---

2. Comparison of Vital Status in the NHIS Linked Mortality Files with the Medical Expenditure Panel Study (MEPS)

The probabilistic approach utilized to match NHIS participants to NDI records was validated using the NHANES I Epidemiologic Follow-up Survey (see section 1). In addition, NCHS evaluated the completeness and accuracy of NHIS-NDI matched mortality data using the Medical Expenditure Panel Study (MEPS). The MEPS is a sample of families and individuals drawn from a sub-sample of households that participated in the NHIS, and thus also includes some participants in the NHIS-NDI linkage. The MEPS is designed to collect information on medical expenditures, but to the extent possible, interviewers identify decedents and record their date of death based on reports by a household member or a proxy respondent (e.g., neighbor). However, unlike the NHEFS cohort where vital status is confirmed by a proxy interview or death certificate, the MEPS cohort used for this analysis does not have a confirmed vital status. In this regard, this analysis is not a validation study but rather a comparison of mortality ascertainment between the MEPS and the NHIS-NDI linkage process.

The MEPS features a panel design with five rounds of interviewing covering 2 full calendar years. For the mortality comparison, NCHS used data from all panels for the six years of the MEPS data from 1996 to 2001, which correspond to a subset of NHIS respondents from the years 1995 to 2000. Only NHIS adults (18 years of age and older) who met the NDI criteria of sufficient data for

matching were eligible for the mortality comparison, leaving 43,700 eligible NHIS adults for whom mortality status could be compared between MEPS and the NHIS-NDI matched mortality data. To conduct the mortality comparison, only NDI deaths occurring within the two calendar year period of the MEPS surveillance period were included. Date of death was used to determine if the correct death certificate was identified by the NDI match as it is the only mortality information available for the MEPS respondents. .

The table indicates the vital status ascertainment for eligible NHIS participants in both the MEPS survey and the NHIS-NDI match.

Comparison of vital status ascertainment for eligible NHIS participants:

	MEPS		
IS-NDI	Alive	Deceased	
live	42,739	72	42,711
eceased	142	747	889
al	42,881	819	43,700

Of the 43,700 eligible NHIS adults, the NHIS-NDI match indicated 889 deaths. For 747 of the 889 deaths, the MEPS data also indicated that the survey participant was a decedent (84% agreement). Agreement was based upon the comparison of the date of death listed by the MEPS to the date of death on the death certificate selected by the NHIS-NDI probabilistic match (606 matched exactly on month, day and year of death; 77 matched on month and year, and 63 matched on year alone; one matched on month alone).

The MEPS data indicated 819 deaths among its participants, of which 72 were not identified in the NHIS-NDI match. Further analysis of these 72 MEPS deaths was undertaken to determine whether they represented probable false negatives in the NHIS-NDI match, meaning that the NHIS-NDI failed to identify a decedent. For 25 of the cases, it was found that the NHIS-NDI does, in fact, include a death record, but the date of death is outside the MEPS observation period used for this analysis. The Social Security Administration's Death Master File (DMF) confirmed the NHIS-NDI date of death in seven cases and confirmed the MEPS death date in three cases. For the remaining 47 cases, 28 had no potential NDI match record selected, meaning that they would be assumed alive based upon the NHIS-NDI matching methodology. The other 19 had a potential NDI match record that was ultimately considered to be a false match based on the probabilistic match score. To assess whether these 19 cases were false negatives, the date of death from the potential NDI match record was compared to the MEPS date of death. For 12 of the 19, the dates of death matched, indicating that the NHIS-NDI incorrectly determined the potential NDI match record to be false. In the remaining 7 discrepant cases, the date of deaths did not match leaving the final vital status assessment uncertain.

The NHIS-NDI probabilistic match procedure is considered conservative, in that it attempts to minimize the potential for wrongly classifying an alive person as dead. For this reason, further analysis was undertaken of the 142 deaths that the NHIS-NDI match identified, but that were not present in MEPS. The Social Security Administration's Death Master File (DMF) confirmed 28 of these cases as true deaths. For the remaining 114 discordant cases, 59 are considered excellent or very strong matches in the NHIS-NDI matching system. Therefore, the remaining 55 cases are probable NHIS-NDI false positives, which is approximately 6% of the deaths identified by the NHIS-

NDI matching methodology. This estimated false positive rate is higher than the approximately 1.5% rate from the NHEFS calibration study (Section 1). However, unlike the NHEFS calibration study where the vital status of the NHEFS cohort was known, the MEPS study does not have as its primary purpose to collect vital status information on its participants. For this reason, it is quite possible that in cases where the NHIS-NDI match process identified a death but the MEPS did not, the MEPS may have missed or underestimated mortality of participants, particularly those living alone, rather than the NHIS-NDI matching process overestimating mortality. However, no independent source exists to confirm vital status for most of the cases.

In summary, the NHIS-NDI matching methodology identified 889 deaths from the MEPS sample of which 747 (84%) were also identified by the MEPS. An additional 28 deaths found by the NHIS-NDI that were not indicated in MEPS were in fact true deaths, while another 59 NHIS-NDI deaths were considered such good matches that we are confident that they are true deaths. Thus, among the 889 NHIS-NDI deaths in the MEPS sample, the percentage of "true" deaths ranges from 84 to 94%. The most conservative estimate of missed mortality in the NHIS-NDI is 65 deaths (72-7); however, only 15 (12+3) of these can be verified from another source.

The results of the comparison of the MEPS ascertainment of vital status to the probabilistic matching methodology provides further support of the validity of the probabilistic matching methodology of the NHIS-National Death Index linkage.

Appendix C

Altering the criteria to assign vital status

The NHIS Linked Mortality File (Standard and Special Request versions) includes the NCHS recommended vital status ascertainment (MORSTAT) for each eligible NHIS participant. For NHIS participants with a NDI record match, the special request file also includes the NCHS class and score for that record from which the determination of vital status was made. The CLASS and SCORE variables are included on the special request file so that users can alter the criteria for determining vital status and conduct their own sensitivity analyses. The choice of cut-off scores involves decisions regarding sensitivity and specificity. Alternative criteria for determining vital status must balance the trade-off of false positives versus false negatives. Below are two examples of studies evaluating different criteria to ascertain vital status for NHIS-NDI match records.

Using the 1986-1990 NHIS linked to the NDI with mortality follow-up through 1991, Liao et al. (1998) evaluated death rates using three different criteria to identify deaths. Criterion 1 was the most conservative, requiring an exact match on SSN. Only NHIS participants with a NDI record with a Class 1 or 2 match were considered true matches and assumed deceased; Criterion 2 was the NCHS recommended ascertainment of vital status using NCHS's cut-off scores[1]; and Criterion 3 was the least stringent with all of Class 1, 2, and 3 matches plus Class 4 matches with scores higher than the recommended cut-off considered true matches. Mortality estimates were lowest based upon criterion 1 and highest with criterion 3. Furthermore, the use of different criteria to determine vital status had differential effects upon death rates for sex and race/ethnic groups.

Using the 2004 release of the NHIS Linked Mortality files for the NHIS years 1988-1994, NCHS conducted its own evaluation of alternative criteria to assign vital status and its impact on death rates. The analysis evaluated four criteria that differed in the established cut-off scores for determining which Class 2, 3, or 4 matches were true matches. For all four criteria, Class 1 match records were considered true matches. Criterion 1 reflected the NCHS recommended cut-off scores; criterion 2 made all Class 2 matches true matches and all Class 4 matches non-matches; criterion 3 increased by four points the cut-off scores for Classes 2, 3, and 4 from the NCHS recommended scores; and criterion 4 lowered by four points the cut-off scores for Classes 2, 3, and 4 from the NCHS recommended scores. Table 1 displays the results.

Across the four criteria, the number of Class 1 matches (assumed deceased) was 36,345. Altering the cut-off scores for the other classes, Criterion 3 produced the most conservative results with 58,934 deaths overall and criterion 4 the least conservative with 64,142 deaths. The NCHS recommended cut-off scores produced 61,021 deaths overall. Not surprisingly, altering the matching criteria differentially affected the mortality rates by race/ethnicity, sex, and age, with Hispanics being the most affected (data not shown).

[1] Users should note that since this analysis is based upon a previous NHIS-NDI record linkage, the NCHS recommended classifications and cut-off scores differ from the recommended classifications and cut-off scores in current NHIS Linked Mortality files.

Inferences from studies that use the NHIS Linked mortality files could be affected by the matching criteria chosen and the ascertainment of vital status. Researchers interested in examining the robustness of their findings can perform sensitivity analyses by altering the criteria for matching. This may be particularly important for studies examining mortality patterns for specific race/ethnic groups.

Table 1: Number of deaths assigned to NHIS participants based upon four criteria for determining NHIS-NDI record linkages to be matches or non-matches

	Match Record Class	Match Record Score	No. Assumed Dead
Criterion 1: NCHS recommendation[2]	2	≥ 47	10,273
	3	≥ 45	11,855
	4	≥ 40	2,548
Criterion 2	2	All	10,734
	3	≥ 45	11,855
	4	None	--------
Criterion 3 (4 points above NCHS recommended cut-off scores)	2	≥ 51	10,038
	3	≥ 49	10,817
	4	≥ 44	1,291
Criterion 4 (4 points below NCHS recommended cut-off scores)	2	≥ 43	10,415
	3	≥ 41	12,595
	4	≥ 36	4,787

[2] The values for MORSTAT on the NHIS Linked Mortality files are based upon this criterion.