# NHEFS LINKED MORTALITY FILE
## Detailed Notes for Special Request Variables

Final Mortality Status
NDI submission record type
NDI total possible score
Probability of NDI match and non-match
Potential NDI match results
NDI Score
NDI Class
NDI Status
Data year
Residence of decedent[1]
    State[2]
    County[2, 3] and population size
    City and population size
    Met/Non met county
    Region and division
    SMSA, PMSA/MSA
    FIPS codes for SMSA, PMSA/MSA, CMSA
Autopsy Performed[4]
Day of week[5]
Education (single years, 0-17) coded from death certificate[5, 6]
Hispanic origin or descent coded from death certificate[6]
Hospital (including status of decedent) and other type of place of death[7]
Industry or business coded from death certificate[8]
Injury at work[9]
Marital status coded from death certificate
Occupation (usual) coded from death certificate[8]
Place of accident
Place of birth (State and eight categories outside of the United States) coded from death certificate
Place of death
Occurrence
    State[2, 3]
    County[2, 3] and population size
    Region and division
Race (nine categories) coded from death certificate[10]
Sex coded from death certificate

---

[1]Place of residence for decedents who were nonresidents of the United States has been coded to country of residence.
[2]Includes FIPS codes as well as NCHS code.
[3]Data availability varies by years.
[4]Available for years 1972-1994 only.
[5]Available for years 1989-2000 only.
[6]Applicable for only those States having information on the certificate.
[7]Available for years 1979-2000 only.
[8]Applicable only for those States transmitting information to NCHS. Available for years 1985-1998 only.
[9]Available for years 1993-2000 only.
[10]Beginning 1992, additional categories are available for some States.

# NHEFS LINKED MORTALITY FILE
## Detailed Notes for Special Request Variables

A data file with the complete probabilistic NDI match results is available by request. The special request file differs from the current file in that not every NHEFS 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 in the NHIS matching methodology report.

Final Mortality Status **- *MORTSTAT***

The MORSTAT variable is NCHS's final determination of vital status and should be used as an outcome variable and to calculate survival. All NHEFS participants are assigned a vital status code (0 = assumed alive; 1 = assumed deceased). The ascertainment of vital status for NHEFS participants with matches to NDI records is based upon the NCHS recommended criteria determined by a calibration study.

It is important for users to note that NDI match information, including date and cause of death, may be returned for NHEFS participants who were ultimately determined not to be deceased by NCHS. The NDI match information is provided to allow researchers to redefine vital status based on alternative NDI matching criteria. Users should not determine mortality status through the cause of death information, which is found in the variables *CAUSEAVL* or from the ICD-9 and ICD-10 cause of death codes, rather they should first determine decedents from MORSTAT and the corresponding cause of death information for these individuals.

Users should note that the variable STATUS provided on the file is the vital status assignment returned by the NDI, but does not reflect NCHS's final determination of vital status. Please refer to the matching methodology as well as to the detailed notes for the variables SCORE and CLASS for more information on how NCHS ascertained vital status for NHEFS participants with a match to a NDI record.

## NHEFS LINKED MORTALITY FILE
### Detailed Notes for Special Request Variables

NDI Submission record type—***SUBTYPE***

A value for SUBTYPE is available for NHEFS participants.  Although NHEFS participants could have multiple submission records, the categories of SUBTYPE refer to the submission record with the most complete identifying information.  The categories were developed from the identifying data items that are used in the NDI record retrieval process.  A complete NHEFS submission record has all of the following variables available: social security number (SSN), month of birth, year of birth, first name, middle initial (blank is acceptable), and surname.  For females, complete name information also includes birth surname.

The values for SUBTYPE are:

1 = Male, complete
2 = Male, missing only SSN
3 = Male, all other records
4 = Female, complete
5 = Female, missing only SSN
6 = Female, missing only birth surname
7 = Female, missing both SSN and birth surname
8 = Female, all other records

Users may find the SUBTYPE variable useful in conducting sensitivity analyses and when altering the criteria for ascertaining vital status.  For example, decisions to use alternative cut-off scores to reassign vital status for some NHEFS participants should consider the completeness of the submission record, since missing identifying data items on the submission record may still yield a correct match but generally with a lower score.

NDI Total Possible Score—***POS_TOTAL1***

The score variable reflects the sum of all the individual weights for each of the identification items used in the NHEFS-NDI match.  In addition, a total *possible* score is calculated, which sums the absolute value for all the individual weights for non-missing NDI submission data items, plus the average weight value for missing items.  Since users have the ability to alter the match criteria by choosing a different cut-off score, a total possible score is calculated to assist users in determining the difference between the observed score and the total possible score that a given submission record could achieve, given missing data items and the specific values of the individual submission items.

# NHEFS LINKED MORTALITY FILE
## Detailed Notes for Special Request Variables


Probability of NDI Match and Non-match—**MATCH** and **NONMATCH**


Probabilities of a match or a non-match have been estimated for each of the five NDI classes. The estimated probabilities for NDI classes 2, 3, and 4 are based on logistic regressions on known correct matches from the NHEFS training sample (see matching methodology document for a description of NHEFS training sample). Class 1 matches have an estimated probability of a match equal to 1 and non-match equal to 0, since all class 1 matches are considered exact matches. Class 5 records have an estimated probability of a match equal to 0.002 and non-match equal to 0.998, based on the very small proportion of NHEFS correct matches that are class 5 (1 out of 600 unique class 5 records). Separate models were developed for the probability of a match and the probability of a non-match. The coefficients from the models are used to compute the estimated probabilities. The match and non-match models contain indicator variables corresponding to each submission record variable. For the match model, these variables equal 1 if the NHEFS and NDI records match on that particular variable and 0 otherwise. For the non-match model, the indicator variable equals 1 if the NHEFS and NDI records do not match on that variable. The model for the probability of a match includes variables for first name, month of birth, year of birth, sex, race(white, black, other), state of birth, and class as a categorical variable with class=2 as the reference category. The model for the probability of a non-match includes variables for first name, month of birth, year of birth, sex, race, state of birth, state of residence, and class as a categorical variable with class=2 as the reference category.


The final models are:


$$match = -13.1702 + mm\_fname*3.2717 + mm\_mob*4.8066 + mm\_yob*2.1905 + mm\_sex*5.2475 + mm\_race*2.4247 + mm\_sob*3.2503 + class3*-6.8223 + class4*-9.3514$$


$$nonmatch = -8.0342 + nm\_fname*3.2519 + nm\_mob*4.6921 + nm\_yob*1.9463 + nm\_sex*5.3509 + nm\_race*2.4853 + nm\_sob*2.9460 + nm\_sor*3.1147 + class3*6.1567 + class4*8.9210$$


The resulting predictions are transformed into probabilities using the inverse of the logit transform, $e^{x*b} / 1+e^{x*b}$.


The difference of the two estimated probabilities, p_match – p_nonmatch, is an estimate of the uncertainty of a match with values near zero being the most uncertain.



Potential NDI Match Results – **NDIRECORD**


A variable has been created to indicate whether a potential match record was identified during the NDI match process. Not all records identified as potential matches are ultimately determined to be correct matches. The special request file contains NDI results for all cases where a potential NDI match record exists to allow the researcher the opportunity to alter the criteria for determining final match status. Records coded as NDIRECORD=1 will include potential NDI match results (NDI Score, NDI class, NDI status) for all cases where a potential match record was identified.

Revised NDI Scores - ***SCORE***

The score for each potential NHEFS-NDI match record is developed by summing a set of weights that are assigned to each of the identifying data items used in the NHEFS-NDI match. These weights reflect the degree of agreement between the information on the NHEFS submission record and the NDI death record. The weights correspond to $[\text{Log}_2 (1/p_i)]$, that is base 2 logarithm of the inverse of the probability of occurrence of the value of the identifying data item on the submission record. Weights are either positive or negative. If there is agreement between the NHEFS record and the NDI record for a particular identifying data item, the weight is positive. If there is no agreement, the weight is negative. Data items that are missing on the NHEFS submission record, the NDI record, or both are assigned a weight of zero. The exception is middle initial where a blank middle initial is considered a valid value and receives the appropriate weight. Once weights have been created for each individual data item, they are summed to create a score for each potential match.

$$\textbf{\textit{Score}} = \{\sum W_{SSN1} + \ldots + W_{SSN9}{}^1\} + W_{firstname \times sex \times birthyear} + W_{middleinitial \times sex} + W_{lastname} + W_{race}$$
$$+ \quad W_{sex} + W_{maritalstatus \times sex \times age} + W_{birthdate} + W_{birthmonth} + W_{birthyear} + W_{stateofbirth} + W_{stateof\ residence}$$

This table shows the mean, minimum and maximum values for all weights for each identifying data item, rounded to one decimal place.

| NDI submission record variable | Minimum | Maximum | Mean |
|---|---:|---:|---:|
| Sex | 0.9 | 1.1 | 1.0 |
| Race | 0.3 | 5.0 | 2.7 |
| Day of birth | 4.9 | 4.9 | 4.9 |
| Month of birth | 3.5 | 3.7 | 3.6 |
| Year of birth | 3.0 | 8.3 | 6.4 |
| State of birth | 3.8 | 14.0 | 6.8 |
| State of Residence | 3.4 | 17.5 | 7.2 |
| SSN digits 3,6,7,8,9 | 3.3 | 3.3 | 3.3 |
| SSN digit 1 | 2.1 | 12.8 | 5.7 |
| SSN digit 2 | 3.1 | 3.8 | 3.3 |
| SSN digit 4 | 2.1 | 6.2 | 3.9 |
| SSN digit 5 | 2.6 | 4.4 | 3.5 |
| Last name | 6.1 | 15.9 | 14.4 |
| First name – male | 8.8 | 23.3 | 19.3 |
| First name - female | 7.8 | 23.8 | 19.5 |
| Middle initial – male | 3.0 | 12.4 | 7.0 |
| Middle initial – female | 3.0 | 15.4 | 7.2 |
| Marital status – male | 3.0 | 20.8 | 8.6 |
| Marital status - female | 2.2 | 20.4 | 9.0 |

---

[1] For a record to be assigned the total SSN weight, 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.

Revised NDI Class Codes - *CLASS*

The final five Classes used by NCHS for the NHEFS Linked Mortality files are described below[2].

**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 digits (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 were reassigned to Class 3. SSN is known but 3 or more digits do not agree, but 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, and state of birth, with last name and sex having to be in agreement.

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

**Class 5[3]:** 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 or at least 7 digits of SSN and last name agree, but sex and first name to not agree (indicating that a deceased spouse's SSN is recorded).

*Note:* In the following circumstances, potential matches were switched from a Class 2 to a Class 5 and from Class 5 to Class 3. Potential matches initially in Class 2 are switched to Class 5 when the only matching identifying information is *some* agreement on SSN but
  1. fewer than 5 of these items agree (first name, middle initial, last name, day of birth, year of birth, race, sex, marital status, state of birth) or
  2. last name agrees but sex and first name do not agree (suggesting a spouse's NDI record may have been found).

Since these potential matches are now Class 5, they are considered false matches.

---

[2] 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.
[3] Since potential matches given a class 5 are considered to be false matches, NHEFS respondents with a NDI record match that is class 5 are assumed to be alive.

Potential matches initially in Class 5 are moved to Class 3 if there is the possibility that SSN was either mis-recorded or the possibility that the spouse's SSN was recorded instead of the subject's SSN.  This is possible where last name and sex agree on the NHEFS-NDI record match and at least 6 of these items also agree (first name, middle initial, day of birth, year of birth, race, marital status, state of birth).

Revised NDI status - **STATUS**

All potential NDI match records are assigned a probabilistic score based on agreement of the data items used in NDI match.  After scoring the potential matches, each match is categorized into one of five mutually exclusive classes.  Within each class, matches with a score greater than or equal to a designated cut-off score were considered true matches (STATUS=1), while records with a score less than the cut-off were considered false matches (STATUS=0).  All Class 1 records are considered to be true matches and are assigned a STATUS=1.  All class 5 records are considered to be false matches, are assigned a STATUS=0 and are not provided on the special request file.  The cut-off scores for Class 2 is 47, Class 3 is 45, and Class 4 is 40.   The special request file contains NDI results for all cases where a potential NDI match record exists, regardless of the final assigned vital status to provide the researcher the opportunity to alter the criteria for determining final match status. The standard mortality file only contains NDI match results for cases where the final match status was determined to be deceased.