

Assessing NDI Output

Review Your NDI Edit Results First

Before you begin assessing your NDI matches, it is important to carefully review the file named **EDITS** that is included on your NDI results CD. (Refer to the “[User File Edit Results](#)” section in [Chapter 3](#) for more information on the **EDITS** file.)

Your study subjects’ records are edited before they are searched against the NDI file. The NDI edit program determines whether your data file conforms to NCHS data file specifications, file format requirements, and coding instructions. If your data file cannot be read or does not satisfy the basic edit program requirements, your file will be returned to you for revision and resubmission at no extra charge.

The edit program rejects any user records not containing **at least one** of the following combinations of data items:

1. FIRST and LAST NAME and SOCIAL SECURITY NUMBER
2. FIRST and LAST NAME and MONTH and YEAR OF BIRTH
3. SOCIAL SECURITY NUMBER and DATE OF BIRTH and SEX

These combinations of items are part of the NDI Edit requirements that determine whether a user’s record has sufficient data to qualify for a search. These combinations should not be confused with the following seven NDI matching criteria.

User records rejected by the edit program are not eligible for inclusion in the search of the NDI file.

The edit results provide an error listing of each rejected user record with the reason for the rejection. If you have sufficient data to update the rejected records, you may correct and resubmit such records at no charge.

Furthermore, the edit results also include summary statistics on the accuracy and completeness of each NDI data set item you provided. If you notice that one or more of the data items you submitted are missing, out of range, or have an unexpectedly low completeness percentage, consider correcting those items and resubmitting a corrected file for another NDI search at no charge.

Matching Criteria

The NDI Retrieval Program is used to search the NDI file to determine whether a particular NDI death record qualifies as a possible record match with a particular user record. To qualify as a possible record match, both records must satisfy **at least one** of the following seven conditions or matching criteria, and the specified data items must agree on both records:

1. Social Security number
(Note: For a search to be performed on the nine digits of a Social Security number, the NDI user’s record must also contain EITHER a first and last name OR date of birth and sex; see preceding edit/reject criteria.)
2. Exact month and +/- 1 year of birth, first and last name
3. Exact month and +/- 1 year of birth, first and middle initials, last name
4. Exact month and day of birth, first and last name
5. Exact month and day of birth, first and middle initials, last name
6. Exact month and year of birth, first name, father’s surname
7. **If the subject is female:** Exact month and year of birth, first name, last name (on user’s record), and father’s surname (on NDI record)

FIRST NAME or FIRST and MIDDLE INITIALS must match exactly to satisfy the above matching criteria. In the case of LAST NAME and FATHER’S SURNAME, however, the above matching criteria will be satisfied if there is agreement on names based on either their exact spelling or their New York State Identification and Intelligence System (NYSIIS) codes. The NYSIIS name coding procedure is intended to accommodate common misspellings of names by assigning a specified alpha character(s) to each character (or combination of selected characters) in a name; see examples in [Exhibit 12](#).

With the liberal matching criteria described above, a given user record may generate possible record matches with more than one NDI record, especially when the subject’s name is common (e.g., John Smith or Mary

Jones). The matching criteria are intended to maximize the number of **true** matches that can be found, especially for those NDI users unable to provide Social Security numbers. However, as a byproduct of these liberal criteria, a significant number of **false** matches (false positives) should be expected. The success of the NDI matching process will be determined by the following four factors:

- The effectiveness of the seven matching criteria.
- The quality and completeness of the data on your study subjects.
- The quality and completeness of the death certificate data in the NDI file.
- Your ability to assess the quality of the resulting matches.

Responsibilities of the NDI User

It is your responsibility to review the NDI output files, to assess the quality of each possible NDI record match listed, and to determine which possible matches are true matches. NDI records involved in matches, based on any of the seven matching criteria listed above, should only be considered **possible** matches. The format and content of the NDI output described in [Chapter 3](#) are designed to assist you in determining which matches are **true** matches, which ones are **false** matches, and which are **questionable**, thus requiring further investigation.

If you requested an **NDI Plus** search (to obtain cause of death information), you will receive 10 files of NDI output, compared with eight files of output provided for a “routine” NDI search. **You are encouraged to familiarize yourself with each of these files as described in [Chapter 3](#), especially the NDI Retrieval Report, before beginning to view or process them.** Call the NDI staff (301-458-4444) if you require assistance in processing the files or interpreting the output.

Options for Assessing NDI Output

The suggestions listed below should provide a few ideas on how to handle the wealth of information provided and how to assess the quality of matches that occur. As feedback is received from NDI users, this list of suggestions will be expanded in future revisions of this publication.

NOTE: To determine your **true** matches, refer only to the **REPORT** or **COMBINED** files. **As noted in [Chapter 3](#), do not refer to the CAUSE or PRTCAUSE**

files until you have determined the true matches. The **CAUSE** and **PRTCAUSE** files contain only the cause of death codes for the closer matches; however, many of these matches may also be false matches.

If you requested only a routine NDI search, you will receive only the state and date of death and the death certificate numbers for each possible match. You will not receive the **CAUSE** or **PRTCAUSE** files with your NDI results.

Begin by developing specific criteria or matching conditions that must be satisfied for a match to be deemed a **true** match. Do the same to identify and eliminate **false** matches. Strive to minimize the number of matches that appear to be **questionable**. Some users may prefer to purchase death certificates for the questionable matches before making a final determination, but most users carefully assess the NDI results to decide which matches are the true matches and do not purchase death certificates.

Be prepared to receive a relatively large number of possible NDI record matches, many of which will be **false** matches.

If you submit a large volume of records for an NDI search, be prepared to develop a computer program to isolate **true** matches, to eliminate most **false** matches, and to minimize the number of **questionable** matches. To assist you in developing such a computer program, the NDI output contains the **COMBINED** file, entitled, “Combined File of Matching User and NDI Records.” Refer to [Chapter 3](#) for file format and coding information.

Records that match on a Social Security number will not always be true matches. A few of these matches may involve the record of a spouse. Checking further for agreements on sex, first name, or date of birth is helpful. Furthermore, on rare occasions a false match can occur when the Social Security number is reported or recorded wrong (on the user’s or NDI record), causing a match with another person’s number that contains the same nine digits.

Matches generated solely by those criteria involving **date of birth** may be evaluated initially based on the extent to which the various digits of the **Social Security number** agree or disagree. The “NDI Retrieval Report” uses an “X” to show which digits of the Social Security number agree and a dash (—) to show which digits disagree.

Matches generated solely by those matching criteria involving **month** and **day of birth** may be initially evaluated based on the extent to which the years of birth disagree. The “NDI Retrieval Report” indicates the difference in years between the years of birth on an NDI record and a user record (e.g., +1 year, 1 year, 15 years, and so on). In most instances, matches with greater than 10 years of death probably will be false matches. You will have to determine, however, your own specific criteria for eliminating such false matches based on agreements or disagreements on other data items.

The NDI Retrieval Program will generate and list **multiple** NDI record matches; that is, two or more NDI records may be matched with a given user record. This often occurs in the case of common names. The NDI records are ranked based on the number of NDI data items that are in agreement with those on the user record. Keep in mind that none of the multiple matches is necessarily a **true** match; however, if the criteria you develop select one to be a true match, the remaining NDI record matches can be disregarded or eliminated. Furthermore, several users have commented that matches ranked fourth or greater seldom are true matches. Although this experience has not been analyzed and confirmed to date, you may want to consider either (a) eliminating all matches ranked fourth or greater or (b) applying looser elimination criteria for such matches.

A CD titled “NAMES COUNT!” is available upon request. It is intended as an aid to NDI users with large record volumes who want to develop stricter acceptance criteria for those matches involving common names. This CD contains the frequencies and percentages of last names on the NDI master file from 1979 through 2005. The CD also contains the frequencies and percentages of last names by state and first names by birth decade.

The NDI Retrieval Program incorporates the use of phonetic codes for last names to enhance the marginal effectiveness of the NDI matching criteria. These phonetic codes are based on the New York State Identification and Intelligence System (NYSIIS), available from the website <http://www.dropby.com/NYSIIS.html>. Accepting NYSIIS code agreements on names (rather than just exact agreements) tends to generate a few additional **true** matches; however, accepting NYSIIS code agreements on names can also result in the listing of numerous **false** matches. For example, the last name “**Johnson**” also translates to a NYSIIS code of “**Jansen**.” Similarly, the last names of “**Janzen**” and “**Jeanson**” also translate to the same NYSIIS code of “**Jansen**.” (For more examples of last

names associated with the same NYSIIS code, see Exhibit 12.) Consequently, NDI staff suggest that you apply stricter criteria when assessing matches resulting from NYSIIS agreements. Before determining that a **true** match has occurred, confirm whether sufficient agreements occur on other data items to warrant the designation of a true match.

If you request an NDI file search of only the most recent year of death available (or a few of the more current years of death), do not be surprised if you obtain several possible matches against NDI records having earlier years of death. These NDI records are **BELATED** records that were received from some states after the cutoff date for adding NDI records for a particular year of death. For example, a belated 2003 record will be flagged as being loaded with either the 2004 or 2005 deaths (depending on when it was received) and then will be stored in a special belated file which is **always** searched—regardless of the specific years you request to be searched. Consequently, if you submit records for repeat NDI searches as additional years of death are added to the NDI file, you will always have your records searched against any belated records that were missed in earlier searches. (Note that not all states submit belated records, and the average number of belated records added to the NDI file each year is less than 1,000.)

Probabilistic scores are assigned to each NDI record match. The scoring information appears in positions 172–178 of the **COMBINED** file. Refer to Appendix A for how the probabilistic scores are assigned. The assigned probabilistic scores and associated Status Codes (1 = true match and 0 = false match) are intended only as a guide for the user. The Status Code is also listed in the last column of the **REPORT** file. It is still the user’s responsibility to determine which matches are true or false.

Once you have identified the true matches, the associated coded causes of death can be obtained from the **CAUSE** file or the **PRTCAUSE** file. The **CAUSE** and **PRTCAUSE** files contain ICD–9 and ICD–10 codes for the underlying cause of death, as well as the multiple causes of death (entity axis codes and record axis codes). Complete lists of these codes appear on the CD containing the NDI search results. Explanations of the underlying cause of death and the multiple-cause codes are provided in [Appendix B](#).

Be aware that the coded causes of death may not be available for some of the true matches that you selected. NDI is authorized to release only the coded causes of death for those matching NDI records that are **ranked**

first in the list of possible matches. The only exception permitted is that coded causes can also be released in those few instances when a matching NDI record is assigned a high-enough probabilistic score to be considered a **true** match (regardless of how the record is ranked). (NOTE: If you selected *true* matches for which cause of death codes are not available on the CAUSE or PRTCAUSE files, contact NDI staff before attempting to purchase copies of death certificates. The staff may be able to provide you with the appropriate cause of death codes.)

You may find it necessary to purchase copies of some death certificates whenever the causes of death are not available for some NDI matches (or if you feel you need to verify the **questionable** matches). Visit the NCHS website at <http://www.cdc.gov/nchs/w2w.htm> for a document entitled, "Where to Write for Vital Records."

For more information on NDI, visit the NDI website for a bibliography of "Articles Describing the Performance of the National Death Index," available from: http://www.cdc.gov/nchs/data/ndi/ndi_bibliography_performance.pdf. The website also includes bibliographies of studies that have used NDI.

EXHIBIT 12

Last Name Examples of NYSIIS Codes

For more information, visit: <http://www.dropby.com/NYSIIS.html>.

LAST NAME = JOHNSON	LAST NAME = MORSE	LAST NAME = ANDERSON
NYSIIS = JANSAN*	NYSIIS = MARS*	NYSIIS = ANDARSAN*
JOHNSON	MORSE	ANDERSON
JAENSEN	MAERSCH	AANDERSEN
JAHNSEN	MARSAU	AANDERSON
JAHNSON	MARSAW	AMDERSON
JAHNSSEN	MARSCH	ANDAERSON
JAMSEN	MARSE	ANDARSEN
JAMSION	MARSH	ANODERSON
JANNSOHN	MARSHAE	ANDEERSON
JANSEN	MARSHAUS	ANDERSEN
JANSENIUS	MARSHAW	ANDERSENN
JANSHEN	MARSHIO	ANDERSENS
JANSON	MARZIO	ANDERSEON
JANSONIUS	MARZOA	ANDERSHIN
JANSSEN	MEARSE	ANDERSHONIS
JANSSON	MEHRZAI	ANDERSIN
JANSZEN	MERSI	ANDERSOHN
JANZAN	MEYERSHAW	ANDERSOM
JANZANO	MORSA	ANDERSONS
JANZEN	MORSCH	ANDERSON
JEANSON	MORSEAU	ANDERSSON
JEANSONNE	MORZE	ANDERSSOHN
JEHNSEN	MORZO	ANDERSSON
JEMSON	MOURSI	ANDERSUN
JENSEMA	MUERSCH	ANDERZEN
JENSEN	MURSAU	ANDERZHON
JENSSON	MURSCH	ANDERZON
JENZEN	MURSE	ANDORSON
JOHNSEN	MURSU	ANDIRSON
JOHNSION	MURZI	ANMDERSON
		ANNDERSON

* Not an exhaustive list of all the last names associated with the New York State Identification and Intelligence System (NYSIIS) code.