IDENTIFYING THE MOST PROBABLE MATCHES TO THE NATIONAL DEATH INDEX

Robert A. Wright, National Center for Health Statistics
3700 East West Highway, Hyattsville, Maryland 20782

KEY WORDS: linkage, NDI. death index, SOA, LSOA, aging. NHIS

Background

Linkage of survey data to other data bases is one way of improving and expanding the amount of information available, without adding additional burden on the respondent, and, in some circumstances, linkage can provide access to information not readily available to household respondents. One such data base, the National Death Index (NDI), maintains a file of the deaths occurring in the United States. Linkage to this data base allows analysis of causes of death and related information that cannot be known when the person is interviewed. It also allows for confirmation of death as a reason a case is "lost to followup" in longitudinal surveys.

In order to determine how health practices and conditions are related to how long people live and to causes of death, the 1984 National Health Interview Survey Supplement on Aging (SOA) was designed to allow a match to the NDI. All respondents (or their proxies) were told that the National Center for Health Statistics (NCHS) would like to refer to statistical records maintained by NCHS and they were asked for all information (including Social Security Number) that is recommended for linkage with the NDI.

Items Collected

The following information was collected during the SOA interview to facilitate matching:

1. Month, Day, and Year of birth
2. Full name, including First and Last Names and Middle Initial
3. Father's Last Name
4. Social Security Number
5. Sex
6. Race
7. Marital Status
8. State of Residence
9. State (or Country) of birth

Only one other item of information can be used as part of the NDI matching process, the actual or estimated date of death; information which was not available since the survey respondents were living at the time of the SOA interview.

The NDI requires at least one of the following combinations of data items before an NDI match will be attempted:

1. First and Last Name AND Social Security Number (SSN)
2. First and Last Name AND Month and Year of Birth

Records which do not have at least one of these combinations are rejected by a special edit program; 215 such records from the SOA were not submitted to the NDI for matching. Because the records were also matched with the Health Care Financing Administration (HCFA) Medicare records, some of these records were cleaned up and resubmitted.

The NDI Retrieval Program searches the NDI file to determine whether a particular NDI death record qualifies as a possible record match with a particular input record. To qualify as a possible match, both records must satisfy at least one of twelve conditions set by the Retrieval Program.

1. Social Security Number and first name.
2. Social Security Number and last name.
3. Social Security Number and father's surname.
4. If female, Social Security Number and last name on user's record with father's surname on NDI record.
5. Month and exact year of birth and first and last name.
6. Month and exact year of birth and first name and father's surname.
7. If female, month and exact year of birth and first name and last name on user's record with father's surname on NDI record.
8. Month and exact year of birth and first and middle initials and last name.
9. Month within 1 year of birth and first and middle initials and last name.
10. Month within 1 year of birth and first and last names.
11. Month and day of birth and first and last names.
12. Month and day of birth and first and middle initials and last name.

These criteria are designed to maximize the number of true matches identified. Because of this design feature, the Retrieval Program also generates a significant number of false matches (false positives). All matches reported by the Retrieval Program must be examined, therefore, and their number reduced to make efficient use of the Retrieval Program report, either to obtain copies of the death certificates from the States. or to directly link the records to other data sets.

The Retrieval Program Report is sorted so that whenever there is more than one possible match in the NDI file the "first-listed" record for the person is the record determined by the NDI Retrieval Program to be the "best" of the possible matches.

The match between the 16,148 persons in the SOA and the NDI for the years 1984-1986 resulted in 10,302 possible matches. The procedure described below was used to reduce the number of possible matches to a set that more closely represented actual deaths in the SOA population, and to eliminate as many of the false matches as possible.

The 10,302 records of possible matches were screened to remove possible matches where the date of death was before the date of the NHIS/SOA interview in 1984. After screening, 8,599 records remained.

The possible matches were then scored by a computer algorithm. The algorithm was a modified form of two algorithms developed and used by Westat, Inc. Starting with the Westat algorithms, the following criteria were developed by an iterative and somewhat subjective process and were then used to score each case. The criteria relate to five different groups of variables: Social Security Number, date of birth, demographic characteristics, geographic location, and name.

Social Security Number:

- If the Social Security Number (SSN) matched exactly, then the score was increased by 5.
- If 7 or 8 of the digits matched, but not all 9, then the score was only increased by 2.
- If the SSN could not be matched, because it was not available in the NDI file, or because it was not available on the SOA input file, the score was increased by 2.

Date of Birth:
- If the date of birth matched on all three parts, month, day, and year, then the score was increased by 4.
- If only two of the parts matched, then the score was increased by 2.
- If the year of birth did not match exactly, but differed from the NDI year of birth by less than 20 years, or more than 80 years, then the score was increased by 2.

Demographic characteristics:
- If the sex matched, the score was increased by 4.
- If the sex matched and 7 or more of the digits of the SSN matched, the score was increased by 2.
- If the race matched, the score was increased by 2.
- If the marital status matched, the score was increased by 2.

Geographic location:
- If state of residence matched, then the score was increased by 2.
- If state of birth matched, then the score was increased by 2.
- If state of residence matched the state of death, the score was increased by 4.

Name:
- If first name and (if there was a middle initial) the middle initial matched, then the score was increased by 2.
- If the last name matched, then the score was increased by 4.
- If the last name matched only the NYSIIS code, then the score was increased by 2.
- If the last name matched, or the last name matched with the NYSIIS code, the first name matched, and (if there was a middle initial) the middle initial matched, then the score was increased by 2.
- If the sex was female, and the father's surname matched or the last name matched with the father's surname, then the score was increased by 2. The maximum score for the items, assuming an exact match, was:

ITEM	MAXIMUM POSSIBLE SCORE
Social Security Number	7
Date of Birth	4
Sex	4
Race	2
Marital Status	2
State of Residence	2
State of Birth	2
State of Residence with State of Death	4
First Name	4
Last Name	4
Middle Initial	2
Father's Surname, if the sex was female	2

The maximum total score was 37; the minimum score was 4.

Recoding

The scores were then recoded into three groups:

1) Possible matches with scores greater than or equal to 28 were designated good matches.
2) Possible matches with scores of 22, or scores of 24 through 27 were considered likely matches, although they matched on fewer items than the good matches; they were designated fair matches.
3) Possible matches with scores of less than 22 matched on only a few items; they were designated poor matches.
4) Possible matches with scores of 23 matched on only a few items, however, the SSN was an exact match; these were also designated poor matches.
5) A fourth category, non-matches, was created for those SOA records which did not match any record in the NDI.

The SOA records can now be coded as to the likelihood that they have died, date of death, and, after retrieval of the death certificate, cause of death.

For those persons included in the Longitudinal Study of Aging, a longitudinal followup telephone interveiw with a sample of the SOA respondents aged 70 and over, those lost to followup, or reported as deceased can be confirmed.

A year ago, the SOA records were also matched to the NDI; but only to the 1984 and 1985 death records. The results of that match are shown table 3.

Notice that there were fewer matches at all three code levels. As shown in table 4, of the 5,151 person in the LSOA sample, 4,113 were alive at the time of the interview, 605 were reported to be deceased, and the remainder were not interviewed, but were presumed to be alive and eligible for the survey.

For the 5,151 persons recontacted in the LSOA, it is possible to examine some errors in matching. A cross tablulation of the LSOA results with the NDI status code to table 5.

In table 5 there are 11 fair matches with the NDI that indicated that the person in the LSOA was deceased. However, those persons were alive at the time of the 1986 LSOA interview. These probable erroneous matches are readily detectable.

The 46 persons who were poor matches with the NDI and the 169 persons who did not match with the NDI at all may be errors, given that they were reported as deceased at the time of the interview, but they may also be persons who died in 1986, after the latest date of death in available NDI records at that time. There is no way to tell for certain until the next LSOA reinterview is conducted in late 1988, when more information will be available.

Table 1

Table 1 shows the distribution of matches according to score.

Good Matches (Code 1)			Fair Matches (Code 2)			Poor Matches (Code 3)		
Score	Frequency	Percent	Score	Frequency	Percent	Score	Frequency	Percent
28	53	4.6	22	62	2.0	4	17	.5
29	62	5.4	24	44	1.4	6	119	3.8
30	92	8.0	25	13	.4	8	2387	.9
31	179	15.6	26	68	2.2	10	508	16.2
32	41	3.6	27	27	.9	12	837	26.7
33	179	15.6				14	702	22.4
34	12	1.0				16	388	12.5
35	354	30.9				17	1	.0
37	173	15.1				18	185	5.9
						19	2	.1
						20	90	2.9
						21	14	.4
						23	18	.6
Total	1145	100.0		214	6.8		3129	100.0

Table 2

Status Code is Shown Below.

Status	Code	Frequency	Percent
Good Match	1	1,145	7.1
Fair Match	2	214	1.3
Poor Match	3	3,129	19.4
Non-Match	4	11,660	72.2
	Total	6,148	100.0

Table 3

1984-85 Match Results

	SOA		LSOA	
Code	Frequency	Percent	Frequency	Percent
1	1,145	7.1	363	7.0
2	214	1.3	50	1.0
3	3,129	19.4	827	16.1
4	11,660	72.2	3911	75.9
	16,148	100.0	5,151	100.0

Table 4
1986 LSOA Interview Status

	Number	Percent
Completed Interview	4,113	79.8
Deceased	604	11.7
Presumed Alive, No Interview	45	.9
Not Located	389	7.6
Total in the 1986 LSOA	5,151	100.0

Table 5
Number Of Persons In The LSOA 1986 Reinterview By NDI Status and LSOA STATUS

		NDI Status				
LSOA Status	Total	No Rcd	Good	Fair	Poor	None
NO NID Input Record	17	17	0	0	0	0
Alvie	4113	44	0	111	708	3,350
Deceased	604	9	347	33	46	169
Presumed Deceased	16	0	16	0	0	0
Probably Deceased	4	0	0	4	0	0
Probably Not Deceased	71	0	0	0	71	0
Presumed Alive	326	0	0	0	4	322
Total	5,151	70	363	48	829	3,841