

NCHS-SSA Feasibility Study File

(Updated 4/2014)

NCHS has completed a data linkage project designed to link administrative data collected from the Social Security Administration (SSA) with health related data collected in its population surveys. SSA data on the receipt of Old age, Survivors and Disability Insurance (OASDI) benefits and Supplemental Security Income (SSI) benefits have been linked to the following NCHS surveys:

- 1994-2005 National Health Interview Survey (NHIS)
- NHANES I Epidemiologic Follow-up Study (NHEFS)
- Third National Health and Nutrition Examination Survey (NHANES III)
- 1999-2004 National Health and Nutrition Examination Survey (NHANES)
- The Second Longitudinal Study of Aging (LSOA II)
- 1985, 1995, 1997 & 2004 National Nursing Home Survey (NNHS)

Survey data collected from eligible survey respondents for each of these NCHS surveys were linked to five SSA administrative data files: the Master Beneficiary Record ([MBR](#)) file (which contains OASDI benefit history and payment data; 1962-2007), the Supplemental Security Record ([SSR](#)) file (which contains SSI benefit history and payment data; 1974-2007), the Payment History Update System ([PHUS](#)) file (which contains the actual OASDI benefit payment amount; 1984-2007), the 831 Disability Master ([831 DMF](#)) File (which contains initial disability decision rendered by the Disability Determination Services; 1989-2007) and a special extract of summarized quarters of coverage variables from the Master Earnings file known as the Quarters of Coverage file ([QOC](#); 1953-2007) made specifically for this NCHS linkage.

Linkage of NCHS survey respondents to administrative records is performed for respondents who are “linkage-eligible”. Linkage-eligibility for NCHS survey respondents is determined by whether respondents provided the necessary information: Social Security number (SSN) number and date of birth. Linkage-eligibility refers to the potential ability to link data from an NCHS survey respondent to SSA administrative data. It is distinct from program-eligibility, which defines whether an individual meets Federal Social Security eligibility criteria.

To maximize the use of the linked administrative datasets, NCHS has created public use NCHS-SSA Feasibility Study data files that can be downloaded directly from this website. These files are especially useful to researchers considering whether to initiate a NCHS Research Data Center (RDC) proposal to analyze the restricted-use linked NCHS-SSA files. The feasibility files provide a limited set of variables that can be used to determine the maximum available sample size to assess the feasibility of conducting analyses utilizing the linked NCHS-SSA data.

Each feasibility file is NCHS survey specific and a NCHS-SSA Feasibility Study file has been created for all surveys listed above. Each file contains a match status variable: 1) SSA_MATCH -- to indicate whether the NCHS survey respondent was eligible for SSA linkage and if eligible, whether successfully matched to the SSA master enrollment file and 2) five additional variables to indicate if a survey respondent has a linked data record on each of the five SSA administrative records file. It is important to note that the variables proposed for this data file do not contain any information about benefits or payments; they are simply record status variables designed to give researchers a way to gauge whether their analytic samples will have sufficient numbers of sample persons with linked SSA data to warrant an application to the RDC.

Due to confidentiality requirements, the NCHS-SSA linked files are currently restricted for research use and the data are only accessible through the NCHS RDC, a secure data access portal. Instructions for submitting an RDC proposal to access the data can be found on the [NCHS RDC website](#).

In accordance with NCHS Ethics Review Board (ERB) guidelines, for NCHS survey participants younger than 18 years of age at the time of the survey, NCHS will only provide linked administrative data generated for program participation, claims, and other events that occurred prior to the participant's 18th birthday. Linked administrative data for these survey participants generated on or after their 18th birthday will no longer be released by NCHS. The availability of information in the SSA files for these survey participants depends on the variables requested. Due to this change, revisions have been made to the NCHS-SSA Feasibility Study data files and linkage eligible child survey participants are identified but information about data availability on the specific SSA files has been removed. More information is available in: [Important information related to NCHS survey participants under 18 years of age at the time of the survey](#)

NCHS-SSA Feasibility Study File

DATA CODEBOOK

PUBLICID **NCHS SURVEY IDENTIFIER - PARTICIPANT IDENTIFICATION NUMBER**

Type: Character **Width:** 14 **File Position:** 1-14

Usage Notes:

Public use identifier assigned by NCHS used to link NHIS data and administrative records.

Researchers linking to LSOA II and 1994-2005 NHIS public-use data should use PUBLICID.

See [Appendix A](#) for NCHS survey specific descriptions.

SEQN **NCHS SURVEY IDENTIFIER – SAMPLE SEQUENCE NUMBER**

Type: Numeric **Width:** 5 **File Position:** 1-5

Usage Notes:

Public use identifier assigned by NCHS used to link NHANES data and administrative records.

Researchers linking to NHEFS, NHANES III, and 1999-2004 NHANES public-use data should use SEQN.

See [Appendix A](#) for NCHS survey specific descriptions.

RESNUM **NCHS SURVEY IDENTIFIER – RESIDENT RECORD (CASE) NUMBER**

Type: Numeric **Width:** 7 **File Position:** 1-7

Usage Notes:

Public use identifier assigned by NCHS used to link NNHS data and administrative records.

Researchers linking to NNHS public-use data should use RESNUM. For the 1985 NNHS , the variable RESNUM is constructed using file positions 1-7.

See [Appendix A](#) for NCHS survey specific descriptions.

RESNUM **NCHS SURVEY IDENTIFIER – RESIDENT RECORD (CASE) NUMBER**

Type: Numeric **Width:** 6 **File Position:** 1-6

Usage Notes:

Public use identifier assigned by NCHS used to link NNHS data and administrative records.

Researchers linking to NNHS public-use data should use RESNUM. For the 1995, 1997 & 2004 NNHS, the variable RESNUM is constructed using file positions 1-6.

See [Appendix A](#) for NCHS survey specific descriptions.

NCHS-SSA Feasibility Study File DATA CODEBOOK

SSA_MATCH SSA MATCH STATUS

Type: Numeric **Width:** 1 **File Position:** 15

Possible values:

- 1 LINKAGE-ELIGIBLE & LINKED TO NUMIDENT
- 2 LINKAGE-ELIGIBLE & NOT LINKED TO NUMIDENT
- 3 LINKAGE-ELIGIBLE – CHILD SURVEY PARTICIPANT*
- 9 INELIGIBLE FOR LINKAGE

Usage Notes:

Survey respondents are ineligible for linking to SSA administrative records if they refused to provide their Social Security number at the time of the survey interview or if they are missing key identification data. The Numident file is the SSA master enrollment file of all assigned Social Security Numbers. Not all persons linked to the Numident file will have Social Security benefit information.

*In accordance with NCHS Ethics Review Board (ERB) guidelines, NCHS will no longer release linked administrative data for child survey participants if the administrative data was generated for program participation, claims or other events occurring on or after their 18th birthday.

ON_MBR_FLAG ON MBR FILE INDICATOR

Type: Numeric **Width:** 1 **File Position:** 16

Possible values:

- 0 - LINKAGE-ELIGIBLE & LINKED TO NUMIDENT - RESPONDENT HAS NO DATA ON THE MBR FILE
- LINKAGE-ELIGIBLE & NOT LINKED TO NUMIDENT
- LINKAGE-ELIGIBLE – CHILD SURVEY PARTICIPANT*
- INELIGIBLE FOR LINKAGE

- 1 RESPONDENT HAS DATA ON THE MBR FILE

Usage Notes:

Survey respondents are ineligible for linking to SSA administrative records if they refused to provide their Social Security number at the time of the survey interview or if they are missing key identification data. The Numident file is the SSA master enrollment file of all assigned Social Security Numbers. Not all persons linked to the Numident file will have Social Security benefit information.

*In accordance with NCHS Ethics Review Board (ERB) guidelines, NCHS will no longer release linked administrative data for child survey participants if the administrative data was generated for program participation, claims or other events occurring on or after their 18th birthday.

NCHS-SSA Feasibility Study File

DATA CODEBOOK

ON_SSR_FLAG ON SSR FILE INDICATOR

Type: Numeric **Width:** 1 **File Position:** 17

Possible values:

- 0 - LINKAGE-ELIGIBLE & LINKED TO NUMIDENT - RESPONDENT HAS NO DATA ON THE SSR FILE
- LINKAGE-ELIGIBLE & NOT LINKED TO NUMIDENT
- LINKAGE-ELIGIBLE – CHILD SURVEY PARTICIPANT*
- INELIGIBLE FOR LINKAGE

- 1 RESPONDENT HAS DATA ON THE SSR FILE

Usage Notes:

Survey respondents are ineligible for linking to SSA administrative records if they refused to provide their Social Security number at the time of the survey interview or if they are missing key identification data. The Numident file is the SSA master enrollment file of all assigned Social Security Numbers. Not all persons linked to the Numident file will have Social Security benefit information.

*In accordance with NCHS Ethics Review Board (ERB) guidelines, NCHS will no longer release linked administrative data for child survey participants if the administrative data was generated for program participation, claims or other events occurring on or after their 18th birthday.

ON_PHUS_FLAG ON PHUS FILE INDICATOR

Type: Numeric **Width:** 1 **File Position:** 18

Possible values:

- 0 - LINKAGE-ELIGIBLE & LINKED TO NUMIDENT - RESPONDENT HAS NO DATA ON THE PHUS FILE
- LINKAGE-ELIGIBLE & NOT LINKED TO NUMIDENT
- LINKAGE-ELIGIBLE – CHILD SURVEY PARTICIPANT*
- INELIGIBLE FOR LINKAGE

- 1 RESPONDENT HAS DATA ON THE PHUS FILE

Usage Notes:

Survey respondents are ineligible for linking to SSA administrative records if they refused to provide their Social Security number at the time of the survey interview or if they are missing key identification data. The Numident file is the SSA master enrollment file of all assigned Social Security Numbers. Not all persons linked to the Numident file will have Social Security benefit information.

*In accordance with NCHS Ethics Review Board (ERB) guidelines, NCHS will no longer release linked administrative data for child survey participants if the administrative data was generated for program participation, claims or other events occurring on or after their 18th birthday.

NCHS-SSA Feasibility Study File

DATA CODEBOOK

ON_831_FLAG ON 831 FILE INDICATOR

Type: Numeric **Width:** 1 **File Position:** 19

Possible values:

- 0 - LINKAGE-ELIGIBLE & LINKED TO NUMIDENT - RESPONDENT HAS NO DATA ON THE 831 FILE
- LINKAGE-ELIGIBLE & NOT LINKED TO NUMIDENT
- LINKAGE-ELIGIBLE – CHILD SURVEY PARTICIPANT*
- INELIGIBLE FOR LINKAGE

- 1 RESPONDENT HAS DATA ON THE 831 FILE

Usage Notes:

Survey respondents are ineligible for linking to SSA administrative records if they refused to provide their Social Security number at the time of the survey interview or if they are missing key identification data. The Numident file is the SSA master enrollment file of all assigned Social Security Numbers. Not all persons linked to the Numident file will have Social Security benefit information.

*In accordance with NCHS Ethics Review Board (ERB) guidelines, NCHS will no longer release linked administrative data for child survey participants if the administrative data was generated for program participation, claims or other events occurring on or after their 18th birthday.

ON_QOC_FLAG ON QOC FILE INDICATOR

Type: Numeric **Width:** 1 **File Position:** 20

Possible values:

- 0 - LINKAGE-ELIGIBLE & LINKED TO NUMIDENT - RESPONDENT HAS NO DATA ON THE QOC FILE
- LINKAGE-ELIGIBLE & NOT LINKED TO NUMIDENT
- LINKAGE-ELIGIBLE – CHILD SURVEY PARTICIPANT*
- INELIGIBLE FOR LINKAGE

- 1 RESPONDENT HAS DATA ON THE QOC FILE

Usage Notes:

Survey respondents are ineligible for linking to SSA administrative records if they refused to provide their Social Security number at the time of the survey interview or if they are missing key identification data. The Numident file is the SSA master enrollment file of all assigned Social Security Numbers. Not all persons linked to the Numident file will have Social Security benefit information.

*In accordance with NCHS Ethics Review Board (ERB) guidelines, NCHS will no longer release linked administrative data for child survey participants if the administrative data was generated for program participation, claims or other events occurring on or after their 18th birthday.

Appendix A: Data Usage Issues regarding PUBLICID/SEQN/RESNUM

The data provided on the 1994-2005 NHIS, NHEFS, 1999-2004 NHANES, NHANES III, LSOA II, and 1985, 1995, 1997 & 2004 NNHS linked SSA files can be merged with the NCHS public use survey data files using the unique survey specific Public Identification number (PUBLICID/SEQN/RESNUM).

Note: At this time the linked SSA data files are only available for research use through the NCHS restricted access data center (RDC). Approved RDC researchers may choose to provide their own analytic files created from public use survey files to the RDC. Therefore, it is important for researchers to include survey specific public identification number on any analytic files sent to the RDC. The RDC will merge data (using PUBLICID, SEQN or RESNUM) from the linked SSA files to the analyst's file. The merged file will be held at the RDC and made available for analysis.

Information on how to identify and/or construct the NCHS survey specific PUBLICID, SEQN or RESNUM is provided below.

I. National Health Interview Survey (NHIS)

NHIS 1994

<u>Variable</u>	<u>Public-use Location</u>	<u>Length</u>	<u>Description</u>
YEAR	3-4	2	Year of interview
QUARTER	5	1	Calendar quarter of interview
PSUNUMR	6-8	3	Random recode of PSU #
WEEKCEN	9-10	2	Week of interview within quarter
SEGNUM	11-12	2	Segment number
HHNUM	13-14	2	Household number within quarter
PNUM	15-16	2	Person number within household

Note: Concatenate all variables to get the unique person identifier.

SAS example: `length publicid $14;`

```
PUBLICID = trim(left(YEAR | QUARTER | PSUNUMR | WEEKCEN | SEGNUM | HHNUM | PNUM));
```

Stata example: (note this will convert the variables to string variables)

```
egen PUBLICID = concat(YEAR QUARTER PSU WEEKCEN SEGNUM HHNUM PNUM)
```

NHIS 1995, 1996

<u>Variable</u>	<u>Public-use Location</u>	<u>Length</u>	<u>Description</u>
YEAR	3-4	2	Year of interview
HHID	5-14	10	Household ID number
PNUM	15-16	2	Person number within Household

Note: Concatenate all variables to get the unique person identifier.

SAS example:

```
length publicid $14;  
PUBLICID = trim(left(YEAR || HHID || PNUM));
```

Stata example: (note this will convert the variables to string variables)

```
egen PUBLICID = concat(YEAR HHID PNUM)
```

NHIS 1997-2003

<u>Variable</u>	<u>Public-use Location</u>	<u>Length</u>	<u>Description</u>
SRVY_YR	3-6	4	Year of interview
HHX	7-12	6	Household serial number
FMX	13-14	2	Family number
PX	15-16	2	Person number within Household

Note: Concatenate all variables to get the unique person identifier.

SAS example:

```
length publicid $14;  
PUBLICID = trim(left(SRVY_YR || HHX || FMX || PX));
```

Stata example: (note this will convert the variables to string variables)

```
egen PUBLICID = concat(SRVY_YR HHX FMX PX)
```

*The person identifier was called PX in the 1997-2003 NHIS and FPX in the 2004 (and later) NHIS; users may find it necessary to create an FPX variable in the 2003 and earlier datasets (or PX in later datasets).

NHIS 2004

<u>Variable</u>	<u>Location</u>	<u>Length</u>	<u>Description</u>
SRVY_YR	3-6	4	Year of interview
HHX	7-12	6	Household serial number
FMX	13-14	2	Family number
FPX	15-16	2	Person number

Note: Concatenate all variables to get the unique person identifier.

SAS example:

```
length publicid $14;  
PUBLICID = trim(left(SRVY_YR || HHX || FMX || FPX));
```

Stata example: (note this will convert the variables to string variables)

```
egen PUBLICID = concat(SRVY_YR HHX FMX FPX)
```

NHIS 2005

<u>Variable</u>	<u>Location</u>	<u>Length</u>	<u>Description</u>
SRVY_YR	3-6	4	Year of interview
HHX	7-12	6	Household serial number
FMX	16-17	2	Family number
FPX	18-19	2	Person number

Note: Concatenate all variables to get the unique person identifier.

SAS example:

```
length publicid $14;  
PUBLICID = trim(left(SRVY_YR || HHX || FMX || FPX));
```

Stata example: (note this will convert the variables to string variables)

```
egen PUBLICID = concat(SRVY_YR HHX FMX FPX)
```

II. NHANES I Epidemiologic Follow-up Study NHEFS

<u>Item</u>	<u>Length</u>	<u>Description</u>
SEQN	5	Participant identification number

All of the NHEFS public-use data files are linked with the common survey participant identification number (SEQN). Merging information from multiple NHEFS Files to the NHEFS-SSA linked files using this variable ensures that the appropriate information for each survey participant is linked correctly.

III. National Health and Nutrition Examination Survey (NHANES)

<u>Item</u>	<u>Length</u>	<u>Description</u>
SEQN	5	Participant identification number

All of the NHANES public-use data files are linked with the common survey participant identification number (SEQN). Merging information from multiple NHANES Files to the NHANES-SSA linked files using this variable ensures that the appropriate information for each survey participant is linked correctly.

IV. Third National Health and Nutrition Examination Survey (NHANES III)

<u>Item</u>	<u>Length</u>	<u>Description</u>
SEQN	5	Participant identification number

All of the NHANES III public-use data files are linked with the common survey participant identification number (SEQN). Merging information from multiple NHANES III Files to the NHANES III-SSA linked files using this variable ensures that the appropriate information for each survey participant is linked correctly.

V. The Second Longitudinal Study of Aging (LSOA II)

On the LSOA II survey, researchers need to construct the LSOA II public id from the following variables.

LSOA II

<u>Variable</u>	<u>Public-use Location</u>	<u>Length</u>	<u>Description</u>
YEAR	3-4	2	Year of interview
QUARTER	5	1	Calendar quarter of interview
PSUNUMR	6-8	3	Random recode of PSU #
WEEKCEN	9-10	2	Week of interview within quarter
SEGNUM	11-12	2	Segment number
HHNUM	13-14	2	Household number within quarter
PNUM	15-16	2	Person number within household

Note: Concatenate all variables to get the unique person identifier.

SAS example: `length publicid $14;`

```
PUBLICID = trim(left(YEAR | QUARTER | PSUNUMR | WEEKCEN | SEGNUM | HHNUM | PNUM));
```

Stata example: (note this will convert the variables to string variables)

```
egen PUBLICID = concat(YEAR QUARTER PSU WEEKCEN SEGNUM HHNUM PNUM)
```

VI. National Nursing Home Survey (NNHS)

NNHS 1985

<u>Item</u>	<u>Length</u>	<u>Description</u>
RESNUM	7	Resident Record (Case) Number

NNHS 1995, 1997 & 2004

<u>Item</u>	<u>Length</u>	<u>Description</u>
RESNUM	6	Resident Record (Case) Number

All of the NNHS public-use data files are linked with the common resident record (case) number (RESNUM). Merging information from the NNHS Files to the NNHS-SSA linked files using this variable ensures that the appropriate information for each survey participant is linked correctly.