

NCHS-MAX Feasibility Study File

(Updated 4/2014)

NCHS has completed a data linkage project designed to link administrative data collected from the Centers for Medicare and Medicaid Services (CMS) with data collected in several NCHS population surveys. Medicaid and Children's Health Insurance Program (CHIP) enrollment and claims data covering service from 1999-2009 are available for the following NCHS surveys:

- **1994-2005 National Health Interview Survey (NHIS)**
- **1999-2004 National Health and Nutrition Examination Survey (NHANES)**
- **2004 National Nursing Home Survey (NNHS)**
- **NHANES I Epidemiologic Follow-up Study (NHEFS)**
- **Third National Health and Nutrition Examination Survey (NHANES III)¹**
- **The Second Longitudinal Study of Aging (LSOA II)**

Survey data collected from eligible survey respondents for each of these NCHS surveys were linked to CMS Medicaid Analytic eXtract (MAX) data. MAX contains individual-level information on demographics, monthly enrollment status, eligibility group, and use and costs of services during the year. For each individual linked to the MAX data, information may be available on any of 5 separate MAX files. The MAX files include: the Person Summary (PS) file, the Inpatient Hospital (IP), the Long Term Care (LT) file, the Other Services (OT) file, and the Prescription Drug (RX) file. More information about the types of data on each of these files can be found in: [Documentation and Analytic Guidelines for NCHS surveys linked to Medicaid Analytic eXtract \(MAX\)](#).

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) or Medicare Health Insurance Claim (HIC) number and date of birth. Linkage-eligibility refers to the potential ability to link data from an NCHS survey respondent to MAX data. It is distinct from program-eligibility, which defines whether an individual meets Federal and State-specific Medicaid eligibility criteria.

To maximize public use of these data sets, NCHS has created and released for public use NCHS-MAX Feasibility Study data files. These files provide information about survey respondent eligibility for CMS linkage (linkage-eligibility), final match status (whether the respondent linked to the MAX data), and which specific linked MAX data files are available for each successfully linked survey respondent. These files can be used to determine the maximum available sample size so that the feasibility of conducting analyses utilizing the linked MAX data can be assessed. This should be especially useful to researchers considering whether to initiate a RDC proposal to analyze the restricted use linked MAX files.

Each feasibility file is NCHS survey and survey year specific. A NCHS-MAX Feasibility Study file has been created for all surveys listed above. Each file contains a match status variable: CMS_MEDICAID_MATCH -- to indicate whether the NCHS survey respondent was eligible for CMS Medicaid linkage and if eligible, whether successfully matched to MAX data. The feasibility file also includes a variable to indicate if a survey respondent has a linked data record on each of the five MAX files for each year of Medicaid / CHIP benefit coverage. 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 designed to give researchers a way to gauge whether their analytic samples will have sufficient numbers of sample persons with linked MAX data to warrant an application to the RDC.

¹ In addition to NHANES III, a subset of the Second National Health and Nutrition Examination Survey (NHANES II) sample was linked to data from CMS. Inquiries about this linkage should be directed to the [NCHS Data Linkage Team](#).

Due to confidentiality requirements, the NCHS-MAX linked files are currently restricted for research use and the data are only accessible through the NCHS Research Data Center (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. Due to this change, revisions have been made to the NCHS-MAX feasibility files. More information is available in: [Important information related to NCHS survey participants under 18 years of age at the time of the survey](#).

NCHS-MAX 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: 6

File Position: 1-6

Usage Notes:

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

Researchers linking to 2004 NNHS public-use data should use RESNUM.

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

NCHS-MAX Feasibility Study File DATA CODEBOOK

CMS_MEDICAID_MATCH

CMS Medicaid Match Status

Type: Numeric

Width: 1

File Position: 15

Possible values:

- 1 Linkage-eligible & linked to one or more years of the CMS Medicaid Analytic eXtract (MAX) files
- 2 Linkage-eligible & not linked to any year of the CMS MAX files
- 3 Linkage-eligible & linked to one or more years of the CMS MAX files
 - Child survey participant (less than 18 years of age at time of survey) & turned 18 during the administrative data coverage period - CMS MAX data available prior to the participant's 18th birthday, but information indicating whether child linked after 18th birthday is not available
- 9 Ineligible for CMS Linkage

Usage Notes:

Survey respondents are ineligible for linking to CMS administrative records if they are missing key identification data and/or if they refused to provide their Social Security or Medicare Health Insurance Claim number at the time of the survey interview or did not have a Social Security number verified by the Social Security Administration's Enumeration Verification System.

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_PS_XXXX[†]

On MAX Person Summary (PS) File Indicator (File Year)[†]

Type: Numeric

Width: 1

File Position: 16 - 26

Possible values:

- 0 - Linkage-eligible, linked to one or more years of the CMS MAX files & does not have data on the CMS MAX Person Summary (PS) file for year XXXX[†]
 - Linkage-eligible & not linked to any year of the CMS MAX files
 - Ineligible for CMS Linkage
- 1 Respondent has data on the CMS MAX PS file for year XXXX[†]

Usage Notes:

[†]Represents Medicaid / CHIP coverage years 1999 - 2009

Survey respondents are ineligible for linking to CMS administrative records if they are missing key identification data and/or if they refused to provide their Social Security or Medicare Health Insurance Claim number at the time of the survey interview or did not have a Social Security number verified by the Social Security Administration's Enumeration Verification System.

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-MAX Feasibility Study File DATA CODEBOOK

ON_IP_XXXX[†]

On MAX Inpatient Hospital (IP) File Indicator (File Year)[†]

Type: Numeric

Width: 1

File Position: 27 - 37

Possible values:

- 0** - Linkage-eligible, linked to one or more years of the CMS MAX files & does not have data on the CMS MAX Inpatient Hospital (IP) file for year XXXX[†]
- Linkage-eligible & not linked to any year of the CMS MAX files
- Ineligible for CMS Linkage

- 1** Respondent has data on the CMS MAX IP file for year XXXX[†]

Usage Notes:

[†]**Represents Medicaid / CHIP coverage years 1999 - 2009**

Survey respondents are ineligible for linking to CMS administrative records if they are missing key identification data and/or if they refused to provide their Social Security or Medicare Health Insurance Claim number at the time of the survey interview or did not have a Social Security number verified by the Social Security Administration's Enumeration Verification System.

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_LT_XXXX[†]

On MAX Long Term Care (LT) File Indicator (File Year)[†]

Type: Numeric

Width: 1

File Position: 38 - 48

Possible values:

- 0** - Linkage-eligible, linked to one or more years of the CMS MAX files & does not have data on the CMS MAX Long Term Care (LT) file for year XXXX[†]
- Linkage-eligible & not linked to any year of the CMS MAX files
- Ineligible for CMS Linkage

- 1** Respondent has data on the CMS MAX LT file for year XXXX[†]

Usage Notes:

[†]**Represents Medicaid / CHIP coverage years 1999 - 2009**

Survey respondents are ineligible for linking to CMS administrative records if they are missing key identification data and/or if they refused to provide their Social Security or Medicare Health Insurance Claim number at the time of the survey interview or did not have a Social Security number verified by the Social Security Administration's Enumeration Verification System.

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-MAX Feasibility Study File DATA CODEBOOK

ON_OT_XXXX[†]

On MAX Other Services (OT) File Indicator (File Year)[†]

Type: Numeric

Width: 1

File Position: 49 - 59

Possible values:

- 0** - Linkage-eligible, linked to one or more years of the CMS MAX files & does not have data on the CMS MAX Other Services (OT) file for year XXXX[†]
- Linkage-eligible & not linked to any year of the CMS MAX files
- Ineligible for CMS Linkage

- 1** Respondent has data on the CMS MAX OT file for year XXXX[†]

Usage Notes:

[†]**Represents Medicaid / CHIP coverage years 1999 - 2009**

Survey respondents are ineligible for linking to CMS administrative records if they are missing key identification data and/or if they refused to provide their Social Security or Medicare Health Insurance Claim number at the time of the survey interview or did not have a Social Security number verified by the Social Security Administration's Enumeration Verification System.

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_RX_XXXX[†]

On MAX Prescription Drug (RX) File Indicator (File Year)[†]

Type: Numeric

Width: 1

File Position: 60 - 70

Possible values:

- 0** - Linkage-eligible, linked to one or more years of the CMS MAX files & does not have data on the CMS MAX Prescription Drug (RX) file for year XXXX[†]
- Linkage-eligible & not linked to any year of the CMS MAX files
- Ineligible for CMS Linkage

- 1** Respondent has data on the CMS MAX RX file for year XXXX[†]

Usage Notes:

[†]**Represents Medicaid / CHIP coverage years 1999 - 2009**

Survey respondents are ineligible for linking to CMS administrative records if they are missing key identification data and/or if they refused to provide their Social Security or Medicare Health Insurance Claim number at the time of the survey interview or did not have a Social Security number verified by the Social Security Administration's Enumeration Verification System.

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 2004 NNHS linked MAX 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 MAX 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 MAX 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 PSUNUMR 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-CMS MAX 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-CMS MAX 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-CMS MAX 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 PSUNUMR WEEKCEN SEGNUM HHNUM PNUM)
```

VI. 2004 National Nursing Home Survey (NNHS)

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

All of the 2004 NNHS public-use data files are linked with the common resident record (case) number (RESNUM). Merging information from the 2004 NNHS Files to the 2004 NNHS-CMS MAX linked files using this variable ensures that the appropriate information for each survey participant is linked correctly.