

## **The Linkage of National Center for Health Statistics Survey Data to Medicaid Enrollment and Claims Data**

### **Methodology and Analytic Considerations**

**Suggested citation:**

National Center for Health Statistics, Office of Analysis and Epidemiology. The Linkage of National Center for Health Statistics Survey Data to Medicaid Enrollment and Claims Data - Methodology and Analytic Considerations. January 2019. Hyattsville, Maryland.

Available at the following address: <https://www.cdc.gov/nchs/data/datalinkage/nchs-medicaid-linkage-methodology-and-analytic-considerations.pdf>

Last updated: January 28, 2019

## Table of Contents

<b>Introduction</b> .....	1
<b>Data Sources</b> .....	2
<i>National Center for Health Statistics</i> .....	2
<i>Centers for Medicare &amp; Medicaid Services</i> .....	4
<b>Additional Related Data Sources</b> .....	7
<b>Linkage of NCHS Surveys with 1999-2014 Medicaid Data</b> .....	7
<i>Linkage Eligibility</i> .....	7
<i>Child Survey Participants</i> .....	9
<i>CMS Virtual Research Data Center</i> .....	9
<i>Linkage Methods</i> .....	10
<i>Linkage Algorithm for Surveys that collected SSN9</i> .....	10
<i>Linkage Algorithm for Surveys that collected SSN4</i> .....	11
<i>Linkage Rates</i> .....	13
<i>Data Confidentiality</i> .....	13
<b>Analytic Considerations when using the Linked NCHS-CMS Medicaid Files</b> .....	15
<i>General Notices to Users</i> .....	15
<b>General Analytic Guidance</b> .....	15
<i>Merging Linked NCHS-CMS Medicaid Data with NCHS survey data</i> .....	15
<i>Variables to request in RDC applications</i> .....	15
<i>Sample weights</i> .....	17
<i>Temporal alignment of survey and administrative data</i> .....	18
<b>Analysis Using Linked NCHS-CMS Medicaid Data</b> .....	18
<i>NCHS-CMS Medicaid Linked Data Feasibility Files</i> .....	18
<i>PS File Considerations</i> .....	19
<i>Multiple PS records</i> .....	20
<i>Merging the PS File with the claims (IP, LT, OT, and RX) files</i> .....	21
<i>Identifying Medicaid enrollees</i> .....	21
<i>Non-Medicaid/Non-S-CHIP observations</i> .....	22
<i>Days in month</i> .....	22
<i>Dual eligible beneficiaries</i> .....	23

<i>Availability of CHIP enrollment information</i> .....	24
<i>Diagnoses and procedure coding</i> .....	25
<i>State differences in Medicaid</i> .....	25
<i>State-specific data issues</i> .....	26
<i>Managed care vs. fee for service (FFS)</i> .....	26
<i>Waivers</i> .....	27
<i>RX File Considerations</i> .....	28
<i>OT File Considerations</i> .....	29
<i>Duplicate records on the IP, LT, OT and RX Files</i> .....	29
<i>General limitations of MAX</i> .....	29
<b>References</b> .....	31
<b>Acknowledgments</b> .....	33
<b>Additional Resources</b> .....	33
<b>Table 1.</b> Linked NCHS-CMS Medicaid Data - Sample Sizes and Unweighted Percentages, by Survey and Age at Interview: NHIS and LSOA II .....	34
<b>Table 2.</b> Linked NCHS-CMS Medicaid Data - Sample Sizes and Unweighted Percentages, by Survey and Age at Interview: NHANES, NHANES III, and NHEFS.....	38
<b>Table 3.</b> Linked NCHS-CMS Medicaid Data - Sample Sizes and Unweighted Percentages, by Survey and Age at Interview: NHHCS and NNHS .....	40
<b>Table 4.</b> Unweighted Percentage of Linked NCHS-Medicaid Sample Enrolled in a Medicaid Comprehensive Managed Care Program for at Least 1 Month <sup>1</sup> , during 1999–2014 MAX Data Years, by Survey.....	41
<b>Table 5.</b> Unweighted Percentage of Linked NCHS-Medicaid Sample Enrolled in a Medicaid Comprehensive Managed Care Program for the Full Year <sup>1</sup> , during 1999–2014 MAX Data Years, by Survey .....	43
<b>Appendix A:</b> Merging Linked NCHS-CMS Medicaid Files with NCHS Survey Data.....	45

# **The Linkage of National Center for Health Statistics Survey Data to Medicaid Enrollment and Claims Data - Methodology and Analytic Considerations**

## **Introduction**

As the nation's principal health statistics agency, the mission of the National Center for Health Statistics (NCHS) is to provide statistical information that can be used to guide actions and policy to improve the health of the American people. As part of its ongoing efforts to fulfill this mission, NCHS conducts several population-based and establishment surveys that provide rich cross-sectional information on the nation's health such as smoking status, height and weight, health status, and socio-economic circumstances. Although the survey data collected provide information on a wide-range of health related topics, they often lack information on longitudinal outcomes. Through its data linkage program, NCHS has been able to enhance the survey data it collects by augmenting survey information with information from administrative data sources. The linkage of survey information with administrative data provide the unique opportunity to study changes in health status, health care utilization and expenditures in specialized populations, such as low income families with children, the elderly and disabled.

Under an interagency agreement between NCHS and the Centers for Medicare and Medicaid Services (CMS), data from several NCHS surveys have been linked to Medicaid enrollment and claims data. This linkage is the second collaboration in a series of collaborations between the NCHS and CMS. A previous linkage was conducted under an interagency agreement among the NCHS, CMS, the Social Security Administration (SSA), and the Office of the Assistant Secretary for Planning and Evaluation (ASPE) of the Department of Health and Human Services (DHHS).

These linkages support various research initiatives of the participating agencies. The resulting linked data files provide the opportunity to examine the administrative data during the year the survey was conducted, in years following the survey, as well as the years prior to the survey for some NCHS survey participants. The linked NCHS-Medicaid files, in particular, combine health and socio-demographic information from the surveys with enrollment and claims information from the Medicaid and Children's Health Insurance (CHIP) programs, resulting in unique

## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

population-based information that can be used for an array of epidemiological and health services research.

This report describes the most recent linkage conducted between NCHS surveys and CMS administrative records. A brief overview of the data sources, the methods used for linkage, descriptions of the resulting linked data files and analytic guidance are provided. More information about the previous linkage of NCHS survey data and CMS administrative data can be found in *Linkage of NCHS Population Health Surveys to Administrative Records From Social Security Administration and Centers for Medicare & Medicaid Services* [1].

### **Data Sources**

#### *National Center for Health Statistics*

Data from the following NCHS surveys were linked to 1999-2014 Medicaid enrollment and claims data:

- The 1994-2013 National Health Interview Survey (NHIS)
- The Second Longitudinal Study of Aging (LSOA II)
- The 1999-2012 Continuous National Health and Nutrition Examination Survey (NHANES)
- The Third National Health and Nutrition Examination Survey (NHANES III)
- The NHANES I Epidemiologic Follow-Up Study (NHEFS)
- The 2007 National Home and Hospice Care Survey (NHHCS)
- The 2004 National Nursing Home Survey (NNHS)

A brief description of the NCHS surveys follows:

**NHIS** is a nationally representative, cross-sectional household interview survey that serves as an important source of information on the health of the civilian, noninstitutionalized population of the United States. It is a multistage sample survey with primary sampling units of counties or adjacent counties, secondary sampling units of clusters of houses, tertiary sampling units of households, and finally, persons within households. It has been conducted continuously since 1957 and the content of the survey is periodically updated. NHIS has been used as the sampling

## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

frame for other NCHS surveys focusing on specialized populations, including **LSOA II**. Prior to 2007, NHIS traditionally collected full 9-digit Social Security Numbers (SSN) from survey participants. However, in attempt to address respondents' increasing refusal to provide SSN and consent for linkage, NHIS began, in 2007, to collect only the last 4 digits of SSN and added an explicit question about linkage for those who refused to provide SSN. The implications of this procedural change on data linkage activities are discussed later in this report. NHIS is currently planning a content and structure redesign. For detailed information on the NHIS's contents and methods, refer to the NHIS website [2].

**LSOA II** was a prospective study of a nationally representative sample of civilian, non-institutionalized persons 70 years of age and over at the time of their 1994 NHIS interview, which served as the baseline for the study. The LSOA II study design included two follow-up telephone interviews, conducted in 1997-98 and 1999-2000. The LSOA II provides information on changes in disability and functioning, individual health risks and behaviors in the elderly, and use of medical care and services employed for assisted community living. For detailed information on the LSOA II contents and methods, refer to the LSOA II website [3].

**NHANES** is a continuous, nationally representative survey consisting of about 5,000 persons from 15 different counties each year. For a variety of reasons, including disclosure issues, the NHANES data are released on public-use data files in two-year increments. The survey includes a standardized physical examination, laboratory tests, and questionnaires that cover various health-related topics. NHANES includes an interview in the household followed by an examination in a mobile examination center (MEC). NHANES is a nationally representative, cross-sectional sample of the U.S. civilian, noninstitutionalized population that is selected using a complex, multistage probability design.

Prior to becoming a continuous survey in 1999, NHANES was conducted periodically, with the last periodic survey, **NHANES III**, conducted between 1988 and 1994. NHANES III was designed to provide national estimates of health and nutritional status of the civilian, non-institutionalized population of the United States aged 2 months and older. Similar to the continuous survey,

## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

NHANES III included a standardized physical examination, laboratory tests, and questionnaires that covered various health-related topics.

**NHEFS** was a national longitudinal study conducted in collaboration with the National Institutes of Health, National Institute on Aging and other agencies of the Public Health Service. The NHEFS cohort included all persons 25–74 years of age who completed a medical examination as part of NHANES I in 1971–75. The NHEFS study design included four follow-up interviews, conducted in 1982-84, 1986, 1987 and 1992, to investigate the relationships between clinical, nutritional, and behavioral factors assessed at baseline, and subsequent morbidity, mortality, and institutionalization.

For detailed information about the Continuous NHANES, NHANES III and NHEFS contents and methods, refer to the NHANES website [4].

**NHHCS** is one in a series of nationally representative sample surveys of U.S. home health and hospice agencies. It is designed to provide descriptive information on home health and hospice agencies, their staffs, their services, and their patients. NHHCS was first conducted in 1992 and was repeated in 1993, 1994, 1996, 1998, and 2000, and most recently in 2007. Only the most recent year of NHHCS was included in the CMS linkage. For more information on the NHHCS content and methods, refer to the NHHCS website [5].

**NNHS** provides information on nursing homes from two perspectives- that of the provider of services and that of the recipient of care. Data for the surveys were obtained through personal interviews with facility administrators and designated staff who used administrative records to answer questions about the facilities, staff, services and programs, and medical records to answer questions about the residents. NNHS was first conducted in 1973-1974 and repeated in 1977, 1985, 1995, 1997, 1999, and most recently in 2004. Only the 2004 survey was included in the CMS linkage. For more information on the NNHS content and methods, refer to the NNHS website [6].

### *Centers for Medicare & Medicaid Services*

CMS is responsible for administering Medicare, Medicaid, and CHIP, and more recently, the Health Insurance Marketplace. The Medicare program provides health insurance for people aged

## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

65 and over, people under 65 with permanent disabilities who receive SSDI, and people of all ages with end-stage renal disease (ESRD, permanent kidney failure requiring dialysis or a kidney transplant). Medicaid is a means-tested entitlement program that provides health care coverage to some vulnerable populations in the United States, including low-income children, and the aged or disabled poor. CHIP targets low-income uninsured children and pregnant women in families with incomes too high to qualify for most state Medicaid programs. Both Medicaid and CHIP are jointly financed by the federal and state governments, and are administered by the states. For more information about CMS and the programs that it administers, refer to the CMS website [7].

### **Medicaid Analytic eXtract (MAX) Files**

For this linkage, NCHS survey data were linked to the data for the 1999-2014 Medicaid Analytic eXtract (MAX) files. The MAX files are a research-ready data source for Medicaid and CHIP calendar year person-level data on eligibility, service utilization and payment information. MAX data are primarily derived from administrative data provided the states to CMS through the Medicaid Statistical Information System (MSIS). Detailed information about MSIS can be found on the CMS website [8].

For 1999-2012, the MAX files include summary information and claims data for all Medicaid enrollees in each state and the District of Columbia (Puerto Rico and other U.S. territories are excluded). The availability varies by MAX file type and state for 2013-2014. At the time of this linkage, the 2013 MAX files included 28 states and the 2014 included 17 states. The MAX data system consists of a **person summary (PS)** file and four claims files: **inpatient hospital (IP)**, **long-term care (LT)**, **prescription drug (RX)**, and **other services (OT)**. Each of these files is described in greater detail below.

#### *Person Summary (PS) File*

The PS file for each year of MAX data is designed to contain one record for each beneficiary, including persons enrolled in Medicaid, M-CHIP - *a state administered expansion program offering Medicaid benefits to children who were previously ineligible due to their income*, and some (but not all) people enrolled in S-CHIP - *a state administered program distinct from its*



## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

*existing Medicaid program.* In some cases, as described further below, a beneficiary may have more than one record on the PS file during the same year. This might happen, for example, if a person was enrolled in Medicaid in more than one state during the same year. The PS file contains basis of eligibility, monthly enrollment data, type of coverage, demographic information, and summary information regarding expenditures and service use.

### *Inpatient Hospital (IP) File*

The IP hospital file for each year of MAX data contains complete stay records for Medicaid enrollees who used inpatient hospital services and may include more than one record per beneficiary. The file includes admission and discharge dates, diagnosis-related groups (DRG), Medicaid payment for fee-for-service records, third-party payments, Medicaid-paid Medicare copayment and deductible amounts, up to nine ICD–9–CM diagnosis codes, and principal and additional procedure codes.

### *Long-term (LT) Care File*

The LT file for each year of MAX data includes institutional long-term care (LTC) records for services provided by four types of long-term care facilities: mental hospitals for the aged, inpatient psychiatric facilities for persons under age 21, intermediate care facilities for the mentally disabled, and nursing facilities (NF). There may be more than one record per beneficiary on this file. Information in the LT file includes start and end dates of services, patient status at discharge, Medicaid payment amounts for fee-for-service records, third-party payments, Medicaid-paid Medicare copayment and deductible amounts, and up to five ICD–9–CM diagnosis codes. These records do not include procedure codes. Other community-based LTC services (e.g., many home-based and personal care services) are included in the OT file.

### *Prescription Drug (RX) File*

The RX file for each year of MAX data contains prescribed drugs, over-the-counter drugs, and other items dispensed by a freestanding pharmacy (nonhospital-based) and may include more than one record per beneficiary. Information in the RX file includes prescription fill date, new or

## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

refill indicator, National Drug Code, and quantity and day supply. Also included are payment amounts, third-party payments, and Medicaid-paid Medicare copayment and deductible amounts.

### *Other Services (OT) File*

The OT file for each year of MAX data contains two major types of records: 1) records for all non-institutional services delivered that are not reported in other files, and 2) payment records for premiums paid to the following types of Medicaid managed care plans: HMOs, health insurance organizations, prepaid health plans (PHPs), and primary care case management plans (PCCMs). There may be more than one record per beneficiary on this file. The service types in the OT file include physician and professional services, outpatient and clinic services, DME, hospice, home health care, and laboratory and X-ray. Information in the OT files includes dates and types of service, Medicaid payment for fee-for-service enrollees, third-party payments, Medicaid-paid Medicare copayment and deductible amounts, a procedure code, and up to two ICD-9-CM diagnosis codes.

For more information on the MAX files, refer to the CMS website [9].

### **Additional Related Data Sources**

NCHS has also linked to CMS Medicare enrollment and claims data. Linkage of the NCHS survey data with the CMS Medicare data provides the opportunity to study changes in health status, health care utilization and expenditures in the elderly and disabled U.S. populations. For more information about the linked CMS Medicare data, please see the data linkage website [10].

### **Linkage of NCHS Surveys with 1999-2014 Medicaid Data**

#### *Linkage Eligibility*

The linkage of data for NCHS survey participants to Medicaid/CHIP enrollment and claims data was conducted under an interagency agreement between NCHS and CMS. The linkage was performed by NCHS in the CMS Virtual Research Data Center (VRDC) and is not the responsibility of researchers using the data. Approval for the linkage was provided by NCHS' Research Ethics Review Board (ERB) and the linkage was performed only for eligible NCHS

## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

survey participants. The NCHS ERB, also known as an Institutional Review Board or IRB, is an appointed ethics review committee that is established to protect the rights and welfare of human research subjects. Only NCHS survey participants who have provided consent as well as the necessary personally identifiable information (PII), such as date of birth and full or partial SSN or Medicare Health Insurance Claim (HIC) number, are considered linkage-eligible. Linkage-eligibility refers to the potential ability to link data from an NCHS survey participant to administrative data. Due to variability of questions across NCHS surveys, changes to PII collection procedures by the surveys over time, and changes in who is asked specific questions, criteria for NCHS-CMS Medicaid linkage eligibility vary by survey and year.

For many of the surveys initiated prior to and during 2007, including 1994-2006 NHIS, LSOA II, 1999-2008 NHANES, NHANES III, NHEFS, 2007 NHHCS, and 2004 NNHS, a refusal by the survey participant to provide a SSN or HIC number was considered an implicit refusal for data linkage. However, NCHS began to observe an increase in the refusal rate for providing SSN and HIC, particularly for NHIS, which reduced the number of survey participants eligible for linkage [11]. In attempt to address declining linkage eligibility rates, NCHS introduced new procedures for obtaining consent for linkage from NHIS participants. Research was also conducted to assess the accuracy of matching data from NHIS to the National Death Index (NDI) using partial SSN and other PII [12]. The research assessed algorithms using the last four and last six digits of SSN. The results were favorable and provided sufficient data to support changes in how NHIS collected SSN and HIC numbers for linkage [13]. Beginning in 2007, NHIS started requesting only the last four (instead of the full nine) digits of SSN and HIC numbers. In addition, a short introduction before asking for SSN was added and participants who declined to provide SSN or HIC were asked for their explicit permission to link to administrative records without SSN or HIC. Also at this time, the NCHS ERB determined that for 2007 NHIS and all subsequent years, only primary respondents (sample adult and sample children) would be eligible for linkage to administrative records.

The informed consent procedures changed for NHANES as well. NHANES continues to collect full nine digit SSN and HIC numbers. However, beginning with the 2009-2010 NHANES,

## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

participants were explicitly asked for consent to be included in data linkage activities during the informed consent process prior to the interview. Only participants who provided an affirmative response to the linkage question were considered linkage eligible.

### *Child Survey Participants*

NCHS survey participants under 18 years of age at the time of the survey are considered linkage-eligible if the linkage eligibility criteria described above are met and consent is provided by their parent or guardian. However, the consent provided by the parent or guardian does not apply once the child survey participant becomes a legal adult and there is no opportunity for NCHS to obtain consent to link the child participant's survey data to administrative data based on their adult experiences. As a result, in accordance with NCHS ERB guidance, NCHS only includes administrative data that were generated for program participation, claims and other events that occurred prior to the participant's 18<sup>th</sup> birthday on the linked data files provided to researchers.

### *CMS Virtual Research Data Center*

The linkage of NCHS survey data with Medicaid administrative data was performed by authorized NCHS staff within the CMS VRDC. The CMS VRDC is a virtual research environment that allows approved researchers to access Medicare and Medicaid program data from their own workstations. VRDC users are granted direct access to approved CMS data files and are able to conduct analyses for research projects within a secure environment. VRDC users have the ability to download aggregated reports and results to their workstations, following disclosure review. More information about the CMS VRDC can be found on the Research Data Assistance Center (ResDAC) website [14]. ResDAC is a CMS contractor providing free assistance to researchers interested in using CMS data for their research [15].

The data file used for linkage did not contain the NCHS survey public-use identification number, nor did it contain any information that could identify the original survey source. The public-use identification number was replaced with an encrypted linkage identification number used by NCHS staff for data linkage projects.

## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

### *Linkage Methods*

To account for the differences in the way that SSN was collected by the NCHS surveys over time, two different linkage methods were applied for the NCHS-Medicaid linkage. For the NCHS surveys where the full nine digit SSN (SSN9) was collected (1994-2006 NHIS, LSOA II, 1999-2012 NHANES, NHANES III, NHEFS, 2007 NHHCS, and 2004 NNHS), a deterministic or rules-based approach was used. A probabilistic approach based on the framework developed by Fellegi & Sunter [16] was applied for surveys where only the last four digits of SSN (SSN4) were collected (2007-2013 NHIS).

### *Linkage Algorithm for Surveys that collected SSN9*

For surveys that collected SSN9, the same approach used for the previous linkage of NCHS survey data with Medicaid data was followed [1]. To be considered a successful match, agreement was required between the NCHS survey participant's record and the MAX PS record on each of the following items:

- SSN (9-digit)
- Month of birth
- Year of birth
- Sex

## Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations

### *Linkage Algorithm for Surveys that collected SSN4*

For surveys that collected SSN4, a new linkage algorithm was developed to link with limited PII. In summary, a relative likelihood that a pair of records is a true match can be estimated by sum of agreement weights  $A_i$  and disagreement weights  $D_i$ .

$$A_i = \text{Log}_2 \left( \frac{m_i}{u_i} \right)$$

$$D_i = \text{Log}_2 \left( \frac{(1 - m_i)}{(1 - u_i)} \right)$$

for each match variable  $i$

Where  $m_i$  is the probability that the variable agrees given the pair is a true match

$$m_i = \text{P}(\text{variable } i \text{ agrees} \mid \text{true match})$$

and  $u_i$  is the probability that the variable agrees given the pair is not a match

$$u_i = \text{P}(\text{variable } i \text{ agrees} \mid \text{non-match})$$

If the variable agrees,  $A_i$  is calculated; if the variable disagrees,  $D_i$  is calculated.

$$\begin{aligned} w_i &= A_i \text{ if variable } i \text{ agrees} \\ &= D_i \text{ if variable } i \text{ disagrees} \end{aligned}$$

The sum across all the matching variables creates an overall match weight, Match weight  $w = \sum w_i$ . Only pairs with match weight above a cutoff are considered true matches.

The following variables were used for blocking:

- SSN4
- Month of birth
- Year of birth
- Sex

## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

A blocking scheme is designed to reduce the number of pairs generated in the Cartesian product and decrease overall computational time. Match weights were calculated from the agreement/disagreement weights for the other remaining identifier variables:

- Day of birth
- Zip code
- County
- Race/ethnicity
- Predicted probability of being on Medicaid

The predicted probability model of being on Medicaid was developed using results from the 2005 NHIS-Medicaid linkage where the SSN9 was used for linkage. The model was then used to predict the probability that a 2007-2013 NHIS record would link. The dependent variable was an indicator of whether the NHIS 2005 record linked using the SSN9 and the independent variables included the survey report of Medicaid (yes or no), other social services assistance, and income. Income information was based on restricted-use imputed income files. The betas from the 2005 NHIS model were then used to calculate predicted probabilities that a survey record would link to the MAX PS record for the later years of NHIS. An indicator variable was created, equal to 1 for NHIS records with high predicted probability; otherwise it equaled 0. “High” probability was specified by sorting the NHIS records by predicted probability from highest to lowest and summing the sample weights cumulatively until the sum approximated the weighted number of participants linked by SSN-9. With this indicator variable equaling 0 or 1 for the NHIS records, and equaling 1 for all the MAX records, it could be treated like any other match variable. The  $m$  probability was the proportion of true matches (records linked by SSN-9 in 2005) who had the indicator variable equal to 1;  $u$  was the probability that the indicator variable agrees with MAX at random, which, because it equaled 1 for all MAX records, was the proportion of NHIS records with the indicator equal to 1. The agreement and disagreement weights were then incorporated in the overall match weight.

All overall match weights above a threshold of 2.14 were deemed matches and those below were non-matches. The threshold was determined by the minimum sum of Type I and Type II errors.

## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

The algorithm was tested on 2003-2004 and 2011-2012 NHANES data where SSN9 was collected. The 2003-2004 NHANES data were used as the base model (similar to the 2005 NHIS data in the model mentioned above) and the 2011 and 2012 NHANES were used to test to see how well the algorithm performed. The NHANES results were favorable; 96.3% of the SSN9 matches were correctly classified with the SSN4 algorithm.

Upon completion, the results from the SSN9 and SSN4 linkages were combined. A file containing the encrypted NCHS identification number, MSIS identification number, and Medicaid state code for successfully matched survey participants was provided to CMS VRDC staff. MAX data were extracted for successfully matched NCHS survey participants and encrypted data files were shipped to NCHS, where additional quality control checks were performed.

### *Linkage Rates*

The linkage rates for NCHS-CMS Medicaid linkage are provided in [Tables 1 - 3](#). The tables show for each survey, the total survey sample size, the sample size eligible for linkage, the number of eligible survey participants linked to 1999-2014 Medicaid enrollment data and two linkage rates.

The two linkage rates provided in the tables are: a total survey sample linkage rate and an eligible sample linkage rate. The eligible sample for linkage is based upon meeting the linkage eligibility criteria previously described. The linkage rates for each survey were examined overall and by three age groups – less than 18 years, 18 - 64 years, and 65 years and older. Age was defined as the survey participant's age at interview.

### *Data Confidentiality*

The NCHS must provide safeguards for the confidentiality of its survey participants. To ensure confidentiality, all personal identifiers have been removed from the linked NCHS-CMS Medicaid data files. However, there remains the small possibility of re-identification and for this



## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

reason; the linked NCHS-CMS Medicaid data are not available as public-use files. NCHS has provided public-use Feasibility Data Files that include a limited set of variables for researchers to use in determining the feasibility and estimating sample sizes of their proposed research projects. The files can be accessed from the NCHS Data Linkage website [17].

Researchers who want to obtain the linked NCHS-CMS Medicaid data must submit a research proposal to the NCHS Research Data Center (RDC) to obtain permission to access the restricted use files [18].

## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

### **Analytic Considerations when using the Linked NCHS-CMS Medicaid Files**

#### *General Notices to Users*

This section summarizes some key analytic issues for users of the linked NCHS survey data and CMS administrative records. It is not an exhaustive list of the analytic issues that researchers may encounter while using the Linked NCHS-CMS Medicaid Data Files. This document will be updated as additional analytic issues are identified and brought to the attention of the NCHS Data Linkage Team ([datalinkage@cdc.gov](mailto:datalinkage@cdc.gov)). Users of the NCHS-CMS Medicaid linked data are encouraged to visit the ResDAC website [15] for more information on Medicaid data.

### **General Analytic Guidance**

#### *Merging Linked NCHS-CMS Medicaid Data with NCHS survey data*

The Linked NCHS-CMS Medicaid Data Files can only be accessed in a RDC. Within the RDC, the Linked NCHS-CMS Medicaid Data Files can be merged with the NCHS restricted (if needed) and public-use survey data files using unique survey person identification numbers. However, the unique survey identifiers are different across surveys and years. Please refer to [Appendix A](#) for guidance on using and merging the appropriate identification numbers.

#### *Variables to request in RDC applications*

To create analytic files for use in the RDC, a researcher should provide a file containing the variables from the public-use NCHS survey data to the RDC for merging with the requested restricted variables from NCHS surveys and for use with the CMS file variables. The restricted variables from NCHS surveys and the variables from the CMS files that the researcher will use also need to be specifically requested as part of a researcher's application to the RDC. Staff in the RDC verify the full list of variables (restricted and public-use) and check for potential disclosure risk.

Although the complete list of variables used for specific analyses differs, the following variables from NCHS surveys should be considered for inclusion:

## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

- Geography— Geographical information is available on the administrative data for linked participants. However, there may be differences in the information available from the survey and administrative data. It is recommended that users who require information on geography, request this information from the NCHS survey.
  
- Linked mortality data for NCHS surveys—Each of the NCHS surveys that have been linked to the Medicaid data have also been linked to death information obtained from the NDI. The linked NDI mortality files provide date and cause of death for each survey participant who has died. These may be of use to some researchers, but must be specifically requested as part of the researcher’s proposal to RDC. More information about the NCHS-NDI linked mortality files can be found on the NCHS Data Linkage website [19].
  
- NHANES month and year of examination and interview—NHANES is released in 2 year cycles. The exact year (and month) of a survey participant’s interview and examination is not provided on public-use files. However, many researchers will want to know the time elapsed between a given year (or even month) of the Medicaid data and the NHANES interview or examination. The variables that indicate the month and year of NHANES interview or examination must be requested specifically.

It is recommended that researchers include the following variables, available from the public-use NCHS survey files, on their analytic files:

- Sample weights and design variables—these variables are needed to account for the complex design of the NCHS surveys. The names of the weights and design variables differ depending on which NCHS survey is being used. These can be identified using the documentation for each NCHS survey. As discussed below, NCHS recommends adjusting the sample weights to account for linkage eligibility bias.

## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

- Demographic information about survey participants from the NCHS survey— For variables such as race and ethnicity, NCHS demographic information is self- or family respondent-reported and, thus, may be more accurate than demographic data provided in the Medicaid files. Therefore, when possible, the NCHS data should be used for demographic variables.

### *Sample weights*

The sample weights provided in NCHS population health survey data files adjust for oversampling of specific subgroups and differential nonresponse, and are post-stratified to annual population totals for specific population domains to provide nationally representative estimates. The properties of these weights for linked data files with incomplete linkage, due to ineligibility for linkage and non-matches, are unknown. In addition, methods for using the survey weights for some longitudinal analyses require further research. Because this is an important and complex methodological topic, ongoing work is being done at NCHS and elsewhere to examine the use of survey weights for linked data in multiple ways.

One approach is to analyze linked data files using adjusted sample weights. The sample weights available on the NCHS population health survey data files can be adjusted for linkage eligibility, using standard weighting domains to reproduce population counts within these domains: sex, age, and race and ethnicity subgroups. These counts are called “control totals” and are estimated from the full survey sample.

A model-based calibration approach developed within the SUDAAN software package (Procedure WTADJUST or WTADJX) allows auxiliary information to be used to adjust the sample weights. This approach is recommended for adjusting sample weights for the linked files. Because inferences may depend on the approach used to develop weights, within SUDAAN’s WTADJUST or using a different calibration approach, researchers should seek assistance from a statistician for guidance on their particular project. Other approaches or software can be used. NCHS continues to investigate alternate approaches for addressing issues related to missing data, including the use of multiple imputation techniques. More detailed information on adjusting sample weights for linkage eligibility using SUDAAN can be found in Appendix III of *Linkage*

## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

*of NCHS Population Health Surveys to Administrative Records From Social Security Administration and Centers for Medicare & Medicaid Services [1].*

### *Temporal alignment of survey and administrative data*

Each NCHS survey has been linked to multiple years of Medicaid enrollment and claims data. Depending on the survey year, Medicaid data may be available for survey participants at the time of the survey, as well as before or after the survey period. Several factors may influence the alignment of the survey and administrative data, including: age of the survey participant, program eligibility, and discontinuous program coverage.

### **Analysis Using Linked NCHS-CMS Medicaid Data**

#### *NCHS-CMS Medicaid Linked Data Feasibility Files*

To maximize the use of the restricted-use linked NCHS-CMS Medicaid files, NCHS has provided publically available Feasibility Files that include a limited set of variables for researchers to use in determining the feasibility and sample sizes of their proposed research projects. The Feasibility Files include:

- (1) A public-use survey participant identification number so that users can merge variables from NCHS public-use survey data to the Feasibility File.
- (2) An indicator (CMS\_MEDICAID\_MATCH) of whether the NCHS survey participant was eligible for the matching and whether he/she linked to the MAX data.  
CMS\_MEDICAID\_MATCH contains values 1, 2, 3, or 9.
  - a. A “1” indicates that the survey participant is linked; “2” indicates that the survey participant is not linked; “3” indicates a child survey participant with partial administrative data available; and “9” indicates that the survey participant was ineligible for linkage.
  - b. NCHS survey participants were considered ineligible for matching if they are missing key identification data and/or if they refused to provide their SSN or Medicare HIC number at the time of the survey interview (for all surveys) or did not agree to linkage

## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

- without SSN or HIC (for 2007-2013 NHIS) or did not provide an affirmative response to the linkage consent question (for 2009-2012 NHANES). Additional ineligibility criteria included refused, missing, or incomplete information on first name, last name and date of birth.
- c. Ineligible participants should be excluded from all analyses using the linked CMS data.
- (3) Information indicating if a survey participant has a linked data record on any of the MAX files for any of the years of Medicaid/CHIP benefits coverage.

Documentation for the NCHS-CMS Medicaid Feasibility Files can be found on the NCHS Data Linkage website [17].

### *PS File Considerations*

Each Medicaid enrollee is classified by two eligibility groups, a maintenance of assistance status (MAS) group and a basis of eligibility (BOE) group. MAS describes the financial criteria that allow an enrollee to be eligible for Medicaid: receiving cash assistance, being a parent with income below the 1996 Aid to Families with Dependent Children (AFDC) income thresholds, being part of a demonstration project, or other. BOE describes the group to which the enrollee belongs that is categorically eligible for Medicaid (children, elderly, disabled, and other). These measures are combined into the variable EL\_MAX\_ELGBLTY\_CD\_LTST, which concatenates MAS and BOE. MAS is in the first position; BOE is in the second position.

In general, across the years of linked MAX files, variables have changed little, and variable names have been largely standardized by CMS and NCHS. However, occasional additions or changes in variables have been made, so careful examination of the data dictionaries prior to analyses is suggested.

## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

### *Multiple PS records*

Many NCHS survey participants will be linked to multiple MAX file PS records. Most often, this is because a survey participant is linked to several years of MAX data. However, a survey participant may be linked to multiple PS records within the same year. There are multiple explanations for this situation:

- Medicaid enrollees moving between states in a given year
- Eligibility changes resulting in survey participants dis-enrolling and re-enrolling in Medicaid within the same year
- Administrative changes or errors with Medicaid reporting. Some administrative changes and errors can be state-and year-specific. Certain record anomalies in each state have been identified and are provided by CMS [9].
- Most NCHS survey participants with multiple PS records per year had records in multiple states.
- Another potential source of multiple PS records in the same year is false matches due to misreporting of PII or issues with linkage methodology. The validity of multiple records in the same year can be difficult to ascertain. While some records show eligibility in different states in non-overlapping months, others show eligibility in different states in the same months of the same year. A researcher may choose to exclude these records, depending on the research question being explored.

The existence of multiple PS records within 1 year with overlapping months of Medicaid enrollment data between the PS records can complicate analyses. In considering how to assess Medicaid enrollment in the presence of multiple PS records within a year, researchers may consider the use of variables that indicate enrollment by month in each record. By determining whether a person was enrolled in each month across the multiple records within a year, the number of total months of enrollment across records can be obtained. The variables MAX\_ELG\_CD\_MO\_1 through MAX\_ELG\_CD\_MO\_12 indicate whether an enrollee was eligible for Medicaid in a given month and, if so, under what criteria.

## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

To help identify enrollees with multiple PS records within the same year, NCHS has added a set of flag variables to the PS file to identify these observations. FLG\_MULT\_RECS identifies whether enrollees have multiple records in any year. Additional variables identify enrollees who had any multiple PS records in a year and whether they occurred in the same or different state overall and in a given year. FLG\_PRSN\_MULT\_RECS identifies enrollees with multiple records in any year; FLG\_YEAR\_MULT\_RECS identifies enrollees with multiple records in a given year. The response categories for both variables are:

- 0 = No multiple records
- 1 = Multiple records in the same state
- 2 = Multiple records in different states
- 3 = Multiple records in the same and different states

### *Merging the PS File with the claims (IP, LT, OT, and RX) files*

Since the PS file may contain multiple records for a survey participant in a given year, it is important that the following variables are used when merging the PS file with the claims (IP, LT, OT, RX) files to ensure that the appropriate records from each file are merged together:

- SURVEY – Name of survey with data linked to Medicaid/CHIP data
- PUBLICID/SEQN/RESNUM/PATNUM - Unique survey specific participant identification number – (see [Appendix A](#) for survey specific descriptions)
- MSIS\_SEQN – NCHS recoded MSIS identification number
- FILE\_YEAR4 - Calendar year of coverage for Medicaid/CHIP data

The same variables should be used if researchers wish to merge the claims files with each other.

### *Identifying Medicaid enrollees*

The best method for identification of NCHS survey participants who were Medicaid enrollees depends on the exact research question. In general, however, MSNG\_ELG\_DATA on the PS file provides information on enrollment status. The values for MSNG\_ELG\_DATA include:



## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

- . = Enrolled in Medicaid during the year
- 2 = Enrolled in S-CHIP
- 1 = Enrolled in neither Medicaid nor S-CHIP

MSNG\_ELG\_DATA exists on each of the MAX files (PS, RX, IP, LT, OT) and at times, different values are assigned in the different files for the same person (in the same year). However, the value on the PS file is the most valid of these and should be used for all data for that observation.

### *Non-Medicaid/Non-S-CHIP observations*

A small number of observations in the MAX files are coded as having been enrolled in neither Medicaid nor S-CHIP. This generally occurs when claims data are provided to CMS, but no eligibility information is available. Researchers should decide how these observations should be handled in their data analysis based on the specific research question.

### *Days in month*

There are a small number of observations for which the Monthly Days of Eligibility exceeds the total number of days in the month. For example, these variables may suggest that the enrollee was eligible for 31 days in September. These values are caused by administrative errors by the state and should be corrected if an analysis includes these values. An individual can enroll and dis-enroll in Medicaid at any time during a month. As a result, enrollees may have fewer days of Medicaid eligibility in a given month than exist in the month. The number days of Medicaid eligibility per month are captured in EL\_DAYS\_EL\_CNT\_1 through EL\_DAY\_EL\_CNT\_12. The values range from 0 to 31.

## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

### *Dual eligible beneficiaries*

Interest exists among researchers and policy-makers in the group of people who receive benefits from both Medicaid and Medicare, often referred to as “dual eligible.” In the linked NCHS-MAX files, this group can be identified several ways. On the PS file, EL\_MDCR\_DUAL\_ANN identifies dual eligible beneficiaries. Values 50–60 signify that the enrollee was found in the Medicare database. Values 01–10 signify that the enrollee was not found in the Medicare database but was believed to be Medicare-eligible by the state. Values 50–60 should be used to identify dual eligible beneficiaries as the Medicare enrollment database is the preferred indicator of dual enrollment.

To assess whether sample size will be adequate for a particular analysis, as discussed above, using the feasibility files is recommended. While there is no flag for dual eligible beneficiaries on the feasibility files, researchers can use the feasibility files for the Medicare linked data files and the feasibility files for the MAX linked data files to create their own dual eligible flag. This method approximates the number identified as dual eligible on the linked MAX file. More information about the feasibility files for the Medicare linked data files can be found on the NCHS Data Linkage website [20].

For analyses of dual eligible beneficiaries, some researchers may choose to use NCHS surveys linked to both Medicare data and the MAX files. Researchers should note that the methodology used for linking the NCHS surveys to MAX data differs from the methodology used to link NCHS survey data to Medicare data. Detailed information about the methodology used for the NCHS-Medicare linkage can be found in *The Linkage of National Center for Health Statistics*

## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

*Surveys to Medicare Enrollment and Claims Data - Methodology and Analytic Considerations*  
[21].

### *Availability of CHIP enrollment information*

There are two issues related to S-CHIP to consider when using the MAX data. First, states have the option of not reporting information on S-CHIP enrollees to MSIS. Therefore, whereas the data on persons with Medicaid or M-CHIP can be considered universal, the MAX files do not include all S-CHIP enrollees. Variables provide monthly information on CHIP eligibility, as well as whether a person was enrolled in M-CHIP or S-CHIP.

For S-CHIP enrollees in the files, some data elements contain no information. Therefore, variables that are counts of months may not be accurate for persons enrolled in S-CHIP for one or more months since those months are not counted in the total counts. Although S-CHIP enrollees may be a group of particular interest for some researchers, it should be noted that they account for a small percentage of NCHS survey participants linked to the MAX files. For example, among the linked 2013 NHIS participants, less than 1.5% of enrollees in the MAX files are in S-CHIP in any given month. The variables EL\_CHIP\_FLAG\_1 through EL\_CHIP\_FLAG\_12 document CHIP eligibility monthly, and whether an enrollee was in M-CHIP or S-CHIP. The values include:

- 0 = Not eligible for Medicaid or CHIP during this month
- 1 = Enrolled in Medicaid during this month
- 2 = M-CHIP during this month

## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

- 3 = S-CHIP during this month
  
- 9 = CHIP status is unknown

For enrollees with a value of “3” (S-CHIP) for EL\_CHIP\_FLAG\_1 through EL\_CHIP\_FLAG\_12, no information is recorded for other monthly variables for that month.

### *Diagnoses and procedure coding*

Diagnoses are uniformly provided in the MAX files for IP, OT, and LT using *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) codes. However, the coding systems used in the IP and OT files for procedures vary (no procedure codes are provided in the LT files). In the IP files, PRCDR\_CD\_SYS\_1 through PRCDR\_CD\_SYS\_6 describe the type of codes used for procedure codes in PRCDR\_CD\_1 through PRCDR\_CD\_6, respectively. In the OT file, the variable PRCDR\_CD\_SYS describes the type of codes used in the procedure code variable PRCDR\_CD. For each procedure code variable, a separate variable is provided that identifies the coding system used for that code. For the IP file, the majority of procedure codes are ICD-9-CM. However, for the OT file, the majority of codes are either in CPT-4 or HCPCS codes.

### *State differences in Medicaid*

Though Medicaid is administered under general federal guidelines, there is substantial variation in Medicaid at the state level. Medicaid program eligibility, services offered, provider reimbursement and other factors vary greatly from state to state. Consideration of these differences by state may be necessary for many analyses. State identifiers for each NCHS survey need to be specifically requested in these circumstances.

## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

### *State-specific data issues*

Because the data for the MAX files are obtained from each state, data quality differs between states. Prior to conducting analyses, researchers should consult the CMS website on the MAX files [9]. This website provides data dictionaries, data anomalies for the whole MAX file and by state within the MAX files, and summary information by state from the MAX files. For any given analysis, there may be states or variables that present problematic data and careful examination of the resources on the CMS website may reveal these issues before attempting data analysis.

### *Managed care vs. fee for service (FFS)*

Many Medicaid and CHIP enrollees are enrolled in managed care plans, and enrollment in these programs has grown over time. Rates of managed care enrollment also vary markedly across states. The PS file contains variables that identify beneficiaries enrolled in any type of managed care plan and the number of months that they were enrolled in the plan. EL\_PHP\_TYPE\_1\_1 through EL\_PHP\_TYPE\_4\_12 identify up to four types of plans for each month. Types of managed care plans identified in EL\_PHP\_TYPE\_1\_1 through EL\_PHP\_TYPE\_4\_12 are:

- Medical or comprehensive managed care plan
- Dental managed care plan
- Behavioral managed care plan
- Prenatal/delivery managed care plan
- Long-term care managed care plan
- All-inclusive care for the elderly or PACE plan

## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

- Primary care case management or PCCM plan
- Other managed care plan

For enrollees in Medicaid managed care plans, information in MAX is restricted to premium payments and some service-specific utilization information. While records for services delivered (including diagnoses and procedures) are uniformly provided for recipients with FFS coverage, encounter records for those with comprehensive managed care plans are not provided by all states. In some states, only a portion of managed care recipients have encounter data recorded. When included in the files, managed care encounter data list \$0 as the amount paid for the services provided, even when the services are covered by the managed care plan.

A summary of the percentage of NCHS survey participants who were enrolled in a medical or comprehensive managed care plan by year and survey can be found in [Tables 4 - 5](#). Researchers should consider the percentage of survey participants enrolled in a Medicaid managed care plan when determining the feasibility and sample sizes of their proposed research projects.

### *Waivers*

Section 1115 of the Social Security Act provides the Secretary of Health and Human Services broad authority to authorize experimental, pilot, or demonstration projects requested by the states that are likely to assist in promoting the objectives of the Medicaid statute. These projects are intended to test and evaluate a policy or approach that has not been widely used. Some states expand eligibility to persons not otherwise eligible under the Medicaid program, provide services that are not typically covered, or use innovative service delivery systems. Examples include expanding care for children in foster care, providing specialty mental health care, and expanding Medicaid eligibility for family planning services to women of childbearing age not otherwise

## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

eligible for Medicaid. General information about Medicaid waivers can be found on the Medicaid.gov website [22].

Waivers for specific groups make up one of the Maintenance Assistance Status or MAS categories. The variables MAX\_ELG\_CD\_MO\_1 through MAX\_ELG\_CD\_MO\_12 can be used to identify MAS and BOE monthly enrollment information, although the specific type of waiver cannot be identified before 2005. Starting in 2005, MAX files include three elements for each month (MAX\_WAIVER\_TYPE\_1\_MO\_1 through MAX\_WAIVER\_TYPE\_3\_MO\_12) that give detailed information on the type of waivers under which enrollees are eligible for Medicaid.

### *RX File Considerations*

The RX file does not include drugs provided during an inpatient hospital stay. Injectable drugs administered by a health professional are included in the OT file. Beginning in 2006, dual eligible beneficiaries receive receiving the Medicare Part D drug benefit, and their utilization for Part D-covered drugs is in the Medicare Part D Event data. More information on the Medicare Part D prescription drug event (PDE) data can be found on ResDAC's website [15]. Also, NCHS survey data have been linked to Medicare Part D PDE data and more information about these linked data can be found on the NCHS Data Linkage website [10].

As with the other files, occasional additions or changes in variables have been made to the RX files over time; therefore, careful examination of the data dictionaries prior to analyses is suggested.

## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

### *OT File Considerations*

Because of how billing is done for certain types of services, there are multiple claims for the same person and service dates in the OT file. These are not errors or data anomalies. What appear to be duplicate records are not true duplicates, but instead distinct services or portions of a service provided. For example, a patient who has a visit to an outpatient physician and then is sent to an outside laboratory to have a blood test on the same day would have a separate record for each of these services, but both would have the same date of service.

### *Duplicate records on the IP, LT, OT and RX Files*

As with the OT file, we are aware that there are data that appear to be duplicate records on the IP, LT and RX files. Since the MAX claims do not include a record identifier when provided to CMS from the states, it is difficult to determine whether the duplicate records represent data entry errors or realistic scenarios where a person could have many records with matching values. Researchers will need to evaluate each situation to determine the best approach for handling this issue.

### *General limitations of MAX*

There are some general limitations to the information contained in the MAX files. Because these files contain only Medicaid-paid services, they do not capture service use or expenditures during periods of non-enrollment, services paid by other payers, or services provided at no charge. Because MAX files consist only of enrollee-level information, they do not include prescription drug rebates received by Medicaid, Medicaid payments made to disproportionate share hospitals



## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

(DSH)—hospitals that serve a disproportionate share of low-income patients with special needs, payments made through upper payment limit (UPL) programs, and payments to states to cover administrative costs.

In addition, service information in MAX may be missing or incomplete for certain groups of enrollees. This is particularly important for individuals enrolled in both Medicaid and Medicare (dual eligible), persons enrolled in Medicaid managed care plans (either comprehensive or partial plans), and children with S-CHIP coverage. Because Medicare is the first payer for services used by dual enrollees covered by both Medicare and Medicaid, MAX captures such service use only if additional Medicaid payments are made on behalf of the enrollee for Medicare cost sharing or for shared services. Medicare premiums paid by Medicaid on behalf of duals are not included in MAX.

## Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations

### References

1. Golden, C., et al., *Linkage of NCHS Population Health Surveys to Administrative Records From Social Security Administration and Centers for Medicare & Medicaid Services*. National Center for Health Statistics, 2015. Vital Health Stat 1(58).
2. National Center for Health Statistics. National Health Interview Survey (NHIS). Available from: <http://www.cdc.gov/nchs/nhis/index.htm>.
3. National Center for Health Statistics. Second Longitudinal Study on Aging (LSOA II). Available from: <http://www.cdc.gov/nchs/lsao/lsao2.htm>.
4. National Center for Health Statistics. National Health and Nutrition Examination Survey (NHANES). Available from: <http://www.cdc.gov/nchs/nhanes/index.htm>.
5. National Center for Health Statistics. National Home and Hospice Care Survey (NHHCS). Available from: <http://www.cdc.gov/nchs/nhhcs/index.htm>.
6. National Center for Health Statistics. National Nursing Home Survey (NNHS). Available from: <http://www.cdc.gov/nchs/nnhs.htm>.
7. Centers for Medicare & Medicaid Services (CMS). Available from: <http://www.cms.gov/>.
8. Centers for Medicare & Medicaid Services. Medicaid Statistical Information System (MSIS). Available from: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MSIS/>.
9. Centers for Medicare & Medicaid Services. Medicaid Analytic eXtract (MAX) General Information. Available from: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAXGeneralInformation.html>.
10. National Center for Health Statistics. NCHS Data Linked to CMS Medicare Enrollment and Claims Files. Available from: <https://www.cdc.gov/nchs/data-linkage/medicare.htm>.
11. Miller, D., Gindi RM, Parker, JD. *Trends in Record Linkage Refusal Rates: Characteristics of National Health Interview Survey Participants Who Refuse Record Linkage*. in *Joint Statistical Meeting*. 2011. Miami Beach, Florida: American Statistical Association.
12. Sayer, B. and C.S. Cox. *How Many Digits in a Handshake? National Death Index Matching with Less Than Nine Digits of the Social Security Number in Proceedings of the American Statistical Association Joint Statistical Meetings*. 2003.

## Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations

13. Dahlhamer, J.M. and C.S. Cox, *Respondent Consent to Link Survey Data with Administrative Records: Results from a Split-Ballot Field Test with the 2007 National Health Interview Survey*. Presented at the 2007 Federal Committee on Statistical Methodology Research Conference, Arlington, VA, 2007.
14. Research Data Assistance Center. CMS Virtual Research Data Center (VRDC). Available from: <https://www.resdac.org/cms-virtual-research-data-center-vrdc>.
15. Research Data Assistance Center (ResDAC). Available from: <https://www.resdac.org/>.
16. Fellegi, I.P. and A.B. Sunter, *A Theory for Record Linkage*. Journal of the American Statistical Association, 1969. 64(328): p. 1183-1210.
17. National Center for Health Statistics. Public-Use NCHS-CMS Medicaid Feasibility Files. Available from: <https://www.cdc.gov/nchs/data-linkage/medicaid-feasibility.htm>.
18. National Center for Health Statistics. Research Data Center. Available from: <https://www.cdc.gov/rdc/index.htm>.
19. National Center for Health Statistics. NCHS Data Linked to NDI Mortality Files. Available from: <https://www.cdc.gov/nchs/data-linkage/mortality.htm>.
20. National Center for Health Statistics. Public-Use NCHS-CMS Medicare Feasibility Files. Available from: <https://www.cdc.gov/nchs/data-linkage/medicare-feasibility.htm>.
21. National Center for Health Statistics. Office of Analysis and Epidemiology. *The Linkage of National Center for Health Statistics Surveys to Medicare Enrollment and Claims Data - Methodology and Analytic Considerations*. September 2017; Available from: [https://www.cdc.gov/nchs/data-linkage/cms/nchs\\_medicare\\_linkage\\_methodology\\_and\\_analytic\\_considerations.pdf](https://www.cdc.gov/nchs/data-linkage/cms/nchs_medicare_linkage_methodology_and_analytic_considerations.pdf).
22. Medicaid.gov. Section 1115 Demonstrations. Available from: <https://www.medicaid.gov/medicaid/section-1115-demo/index.html>.

## **Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations**

### **Acknowledgments**

Information about the Medicaid/CHIP enrollment and claims files was compiled from the following sources:

Centers for Medicare & Medicaid Services (CMS)  
<http://www.cms.gov/> (accessed January 28, 2019)

Medicaid.gov  
<https://www.medicaid.gov/index.html> (accessed January 28, 2019)

Research Data Assistance Center (ResDAC)  
<http://www.resdac.org/> (accessed January 28, 2019)

### **Additional Resources**

NHANES–CMS Linked Data Tutorial  
<https://www.cdc.gov/nchs/tutorials/NHANES-CMS/index.htm> (accessed January 28, 2019)

**Linked NCHS-CMS Medicaid Data  
Linkage Methodology and Analytic Considerations**

**Table 1. Linked NCHS-CMS Medicaid Data - Sample Sizes and Unweighted Percentages,  
by Survey and Age at Interview: NHIS and LSOA II**

	Sample size			Percent linked	
	Total Sample <sup>1</sup>	Eligible for Linkage <sup>2,3</sup>	Linked to Medicaid Data <sup>4</sup>	Total Sample <sup>5</sup>	Eligible Sample <sup>6</sup>
<b>NHIS 2013</b>	47,417	24,787	7,797	16.44%	31.46%
0 - 17	12,860	4,301	2,458	19.11%	57.15%
18 - 39	12,390	7,251	2,663	21.49%	36.73%
40 - 64	14,435	8,592	1,857	12.86%	21.61%
65 and over	7,732	4,643	819	10.59%	17.64%
<b>NHIS 2012</b>	47,800	26,003	8,300	17.36%	31.92%
0 - 17	13,275	4,760	2,777	20.92%	58.34%
18 - 39	12,434	7,458	2,815	22.64%	37.74%
40 - 64	14,709	9,120	1,906	12.96%	20.90%
65 and over	7,382	4,665	802	10.86%	17.19%
<b>NHIS 2011</b>	45,864	25,539	8,109	17.68%	31.75%
0 - 17	12,850	4,699	2,672	20.79%	56.86%
18 - 39	12,191	7,606	2,805	23.01%	36.88%
40 - 64	13,921	8,737	1,800	12.93%	20.60%
65 and over	6,902	4,497	832	12.05%	18.50%
<b>NHIS 2010</b>	38,434	20,067	6,513	16.95%	32.46%
0 - 17	11,277	4,062	2,337	20.72%	57.53%
18 - 39	10,200	5,927	2,187	21.44%	36.90%
40 - 64	11,507	6,856	1,354	11.77%	19.75%
65 and over	5,450	3,222	635	11.65%	19.71%
<b>NHIS 2009</b>	38,887	21,435	6,875	17.68%	32.07%
0 - 17	11,156	4,131	2,332	20.90%	56.45%
18 - 39	10,277	6,248	2,293	22.31%	36.70%
40 - 64	11,961	7,540	1,578	13.19%	20.93%
65 and over	5,493	3,516	672	12.23%	19.11%
<b>NHIS 2008</b>	30,596	15,189	4,594	15.02%	30.25%
0 - 17	8,815	2,817	1,521	17.25%	53.99%
18 - 39	8,067	4,536	1,515	18.78%	33.40%
40 - 64	9,270	5,332	1,054	11.37%	19.77%
65 and over	4,444	2,504	504	11.34%	20.13%

See footnotes at end of table.

**Linked NCHS-CMS Medicaid Data  
Linkage Methodology and Analytic Considerations**

**Table 1. Linked NCHS-CMS Medicaid Data - Sample Sizes and Unweighted Percentages,  
by Survey and Age at Interview: NHIS and LSOA II – continued**

	Sample size			Percent linked	
	Total Sample <sup>1</sup>	Eligible for Linkage <sup>2,3</sup>	Linked to Medicaid Data <sup>4</sup>	Total Sample <sup>5</sup>	Eligible Sample <sup>6</sup>
<b>NHIS 2007</b>	32,810	14,014	4,175	12.72%	29.79%
0 - 17	9,417	2,580	1,343	14.26%	52.05%
18 - 39	8,864	4,296	1,418	16.00%	33.01%
40 - 64	9,946	4,950	948	9.53%	19.15%
65 and over	4,583	2,188	466	10.17%	21.30%
<b>NHIS 2006</b>	75,716	13,202	4,346	5.74%	32.92%
0 - 17	20,903	1,807	1,027	4.91%	56.83%
18 - 39	22,578	4,302	1,711	7.58%	39.77%
40 - 64	23,845	4,900	1,100	4.61%	22.45%
65 and over	8,390	2,193	508	6.05%	23.16%
<b>NHIS 2005</b>	98,649	45,032	14,965	15.17%	33.23%
0 - 17	26,814	13,097	6,943	25.89%	53.01%
18 - 39	28,982	12,829	4,401	15.19%	34.31%
40 - 64	31,623	14,169	2,429	7.68%	17.14%
65 and over	11,230	4,937	1,192	10.61%	24.14%
<b>NHIS 2004</b>	94,460	46,140	15,245	16.14%	33.04%
0 - 17	26,161	13,541	7,034	26.89%	51.95%
18 - 39	28,015	13,314	4,365	15.58%	32.79%
40 - 64	29,526	13,939	2,508	8.49%	17.99%
65 and over	10,758	5,346	1,338	12.44%	25.03%
<b>NHIS 2003</b>	92,148	49,212	15,709	17.05%	31.92%
0 - 17	25,537	15,832	7,616	29.82%	48.11%
18 - 39	27,809	13,866	4,379	15.75%	31.58%
40 - 64	28,529	14,256	2,442	8.56%	17.13%
65 and over	10,273	5,258	1,272	12.38%	24.19%
<b>NHIS 2002</b>	93,386	53,125	16,575	17.75%	31.20%
0 - 17	26,191	16,888	7,970	30.43%	47.19%
18 - 39	28,298	15,065	4,500	15.90%	29.87%
40 - 64	28,312	15,376	2,702	9.54%	17.57%
65 and over	10,585	5,796	1,403	13.25%	24.21%

See footnotes at end of table.

**Linked NCHS-CMS Medicaid Data  
Linkage Methodology and Analytic Considerations**

**Table 1. Linked NCHS-CMS Medicaid Data - Sample Sizes and Unweighted Percentages,  
by Survey and Age at Interview: NHIS and LSOA II – continued**

	Sample size			Percent linked	
	Total Sample <sup>1</sup>	Eligible for Linkage <sup>2,3</sup>	Linked to Medicaid Data <sup>4</sup>	Total Sample <sup>5</sup>	Eligible Sample <sup>6</sup>
<b>NHIS 2001</b>	100,760	47,772	15,172	15.06%	31.76%
0 - 17	28,572	13,167	6,557	22.95%	49.80%
18 - 39	30,941	14,448	4,401	14.22%	30.46%
40 - 64	30,230	14,661	2,711	8.97%	18.49%
65 and over	11,017	5,496	1,503	13.64%	27.35%
<b>NHIS 2000</b>	100,618	49,623	15,006	14.91%	30.24%
0 - 17	28,495	13,659	6,266	21.99%	45.87%
18 - 39	31,153	15,313	4,347	13.95%	28.39%
40 - 64	29,764	14,925	2,784	9.35%	18.65%
65 and over	11,206	5,726	1,609	14.36%	28.10%
<b>NHIS 1999</b>	97,059	49,586	14,118	14.55%	28.47%
0 - 17	27,271	13,226	5,701	20.90%	43.10%
18 - 39	30,183	15,332	4,148	13.74%	27.05%
40 - 64	28,605	15,131	2,677	9.36%	17.69%
65 and over	11,000	5,897	1,592	14.47%	27.00%
<b>NHIS 1998</b>	98,785	52,786	14,776	14.96%	27.99%
0 - 17	28,122	13,596	5,791	20.59%	42.59%
18 - 39	30,941	16,657	4,303	13.91%	25.83%
40 - 64	28,302	15,975	2,851	10.07%	17.85%
65 and over	11,420	6,558	1,831	16.03%	27.92%
<b>NHIS 1997</b>	103,477	60,989	16,540	15.98%	27.12%
0 - 17	29,792	14,813	6,134	20.59%	41.41%
18 - 39	32,824	19,894	4,929	15.02%	24.78%
40 - 64	28,970	18,220	3,258	11.25%	17.88%
65 and over	11,891	8,062	2,219	18.66%	27.52%
<b>NHIS 1996</b>	63,402	40,592	10,566	16.67%	26.03%
0 - 17	18,128	9,344	3,692	20.37%	39.51%
18 - 39	20,552	13,643	3,215	15.64%	23.57%
40 - 64	17,588	12,203	2,206	12.54%	18.08%
65 and over	7,134	5,402	1,453	20.37%	26.90%

See footnotes at end of table.

**Linked NCHS-CMS Medicaid Data  
Linkage Methodology and Analytic Considerations**

**Table 1. Linked NCHS-CMS Medicaid Data - Sample Sizes and Unweighted Percentages,  
by Survey and Age at Interview: NHIS and LSOA II – continued**

	Sample size			Percent linked	
	Total Sample <sup>1</sup>	Eligible for Linkage <sup>2,3</sup>	Linked to Medicaid Data <sup>4</sup>	Total Sample <sup>5</sup>	Eligible Sample <sup>6</sup>
<b>NHIS 1995</b>	102,467	69,723	17,375	16.96%	24.92%
0 - 17	29,711	15,721	6,072	20.44%	38.62%
18 - 39	32,961	23,793	5,333	16.18%	22.41%
40 - 64	27,840	20,755	3,617	12.99%	17.43%
65 and over	11,955	9,454	2,353	19.68%	24.89%
<b>NHIS 1994</b>	116,179	81,116	18,224	15.69%	22.47%
0 - 17	32,460	16,766	5,820	17.93%	34.71%
18 - 39	37,230	28,290	5,708	15.33%	20.18%
40 - 64	31,918	24,333	3,851	12.07%	15.83%
65 and over	14,571	11,727	2,845	19.53%	24.26%
<b>LSOA II<sup>7</sup></b>	9,447	7,437	1,921	20.33%	25.83%
70 and over	9,447	7,437	1,921	20.33%	25.83%

<sup>1</sup>For 2007-2013 NHIS, only Sample Adult and Sample Child participants are included in the NCHS-CMS Medicaid linkage.

<sup>2</sup>Eligibility for linkage is based on not refusing to provide Social Security (SSN) or Medicare Health Insurance Claim (HIC) numbers and having sufficient personally identifiable information.

<sup>3</sup>For 2007-2013 NHIS, eligibility for linkage is based on providing the last four digits of the SSN or an affirmative response to the follow up question to allow linkage without SSN and having sufficient personally identifiable information.

<sup>4</sup>This group includes linkage-eligible survey participants who linked to Medicaid administrative records at any time during the linkage interval (1999-2014).

<sup>5</sup>This percentage is calculated by dividing the number of linked survey participants by the number of participants in the total sample.

<sup>6</sup>This percentage is calculated by dividing the number of linked survey participants by the total number of linkage-eligible participants.

<sup>7</sup>All LSOA II participants were 70 years or older at time of interview.



**Linked NCHS-CMS Medicaid Data  
Linkage Methodology and Analytic Considerations**

**Table 2. Linked NCHS-CMS Medicaid Data - Sample Sizes and Unweighted Percentages,  
by Survey and Age at Interview: NHANES, NHANES III, and NHEFS**

	Sample size			Percent linked	
	Total Sample	Eligible for Linkage <sup>1,2</sup>	Linked to Medicaid Data <sup>3</sup>	Total Sample <sup>4</sup>	Eligible Sample <sup>5</sup>
<b>NHANES 2011-2012</b>	9,756	5,818	2,789	28.59%	47.94%
0 - 17	3,892	1,907	1,309	33.63%	68.64%
18 - 39	2,261	1,467	725	32.07%	49.42%
40 - 64	2,353	1,593	492	20.91%	30.89%
65 and over	1,250	851	263	21.04%	30.90%
<b>NHANES 2009-2010</b>	10,537	6,550	2,891	27.44%	44.14%
0 - 17	4,010	2,085	1,420	35.41%	68.11%
18 - 39	2,392	1,589	730	30.52%	45.94%
40 - 64	2,612	1,809	497	19.03%	27.47%
65 and over	1,523	1,067	244	16.02%	22.87%
<b>NHANES 2007-2008</b>	10,149	6,678	2,936	28.93%	43.97%
0 - 17	3,921	2,187	1,514	38.61%	69.23%
18 - 39	2,203	1,538	671	30.46%	43.63%
40 - 64	2,469	1,825	485	19.64%	26.58%
65 and over	1,556	1,128	266	17.10%	23.58%
<b>NHANES 2005-2006</b>	10,348	7,290	3,390	32.76%	46.50%
0 - 17	4,785	3,089	2,014	42.09%	65.20%
18 - 39	2,507	1,824	811	32.35%	44.46%
40 - 64	1,867	1,448	326	17.46%	22.51%
65 and over	1,189	929	239	20.10%	25.73%
<b>NHANES 2003-2004</b>	10,122	8,670	4,266	42.15%	49.20%
0 - 17	4,502	3,795	2,517	55.91%	66.32%
18 - 39	2,321	1,987	858	36.97%	43.18%
40 - 64	1,805	1,590	482	26.70%	30.31%
65 and over	1,494	1,298	409	27.38%	31.51%

See footnotes at end of table.

**Linked NCHS-CMS Medicaid Data  
Linkage Methodology and Analytic Considerations**

**Table 2. Linked NCHS-CMS Medicaid Data - Sample Sizes and Unweighted Percentages,  
by Survey and Age at Interview: NHANES, NHANES III, and NHEFS - continued**

	Sample size			Percent linked	
	Total Sample	Eligible for Linkage <sup>1,2</sup>	Linked to Medicaid Data <sup>3</sup>	Total Sample	Eligible for Linkage <sup>12</sup>
<b>NHANES 2001-2002</b>	11,039	9,327	4,117	37.30%	44.14%
0 - 17	5,046	4,253	2,473	49.01%	58.15%
18 - 39	2,507	2,058	820	32.71%	39.84%
40 - 64	2,023	1,773	426	21.06%	24.03%
65 and over	1,463	1,243	398	27.20%	32.02%
<b>NHANES 1999-2000</b>	9,965	7,874	3,605	36.18%	45.78%
0 - 17	4,517	3,471	2,034	45.03%	58.60%
18 - 39	2,263	1,780	674	29.78%	37.87%
40 - 64	1,793	1,505	449	25.04%	29.83%
65 and over	1,392	1,118	448	32.18%	40.07%
<b>NHANES III</b>	33,994	28,129	9,176	26.99%	32.62%
0 - 17	14,377	9,361	4,174	29.03%	44.59%
18 - 39	8,169	7,697	2,066	25.29%	26.84%
40 - 64	6,196	5,976	1,686	27.21%	28.21%
65 and over	5,252	5,095	1,250	23.80%	24.53%
<b>NHEFS<sup>6</sup></b>	14,407	12,960	1,809	12.56%	13.96%
25 - 39	4,965	4,502	776	15.63%	17.24%
40 - 64	5,586	5,025	883	15.81%	17.57%
65 - 74	3,856	3,433	150	3.89%	4.37%

<sup>1</sup>Eligibility for linkage is based on not refusing to provide Social Security (SSN) or Medicare Health Insurance Claim (HIC) numbers and having sufficient personally identifiable information.

<sup>2</sup>For 2009-2012 NHANES, eligibility for linkage is defined as providing an affirmative response to the linkage consent question and having sufficient personally identifiable information.

<sup>3</sup>This group includes linkage-eligible survey participants who linked to Medicaid administrative records at any time during the linkage interval (1999-2014).

<sup>4</sup>This percentage is calculated by dividing the number of linked survey participants by the number of participants in the total sample.

<sup>5</sup>This percentage is calculated by dividing the number of linked survey participants by the total number of linkage-eligible participants.

<sup>6</sup>All NHEFS participants were 25-74 years of age at time of interview.

**Linked NCHS-CMS Medicaid Data  
Linkage Methodology and Analytic Considerations**

**Table 3. Linked NCHS-CMS Medicaid Data - Sample Sizes and Unweighted Percentages,  
by Survey and Age at Interview: NHHCS and NNHS**

	Sample size			Percent linked	
	Total Sample	Eligible for Linkage <sup>1</sup>	Linked to Medicaid Data <sup>2</sup>	Total Sample <sup>3</sup>	Eligible Sample <sup>4</sup>
<b>NHHCS 2007</b>	9,416	9,023	3,877	41.17%	42.97%
0 - 17	258	164	122	47.29%	74.39%
18 - 39	296	278	202	68.24%	72.66%
40 - 64	1,718	1,666	885	51.51%	53.12%
65 and over	7,144	6,915	2,668	37.35%	38.58%
<b>NNHS 2004</b>	13,507	13,371	9,927	73.50%	74.24%
0 - 17	14	14	13	92.86%	92.86%
18 - 39	143	141	128	89.51%	90.78%
40 - 64	1,410	1,390	1,225	86.88%	88.13%
65 and over	11,940	11,826	8,561	71.70%	72.39%

<sup>1</sup>Eligibility for linkage is based on not refusing to provide Social Security (SSN) or Medicare Health Insurance Claim (HIC) numbers and having sufficient personally identifiable information.

<sup>2</sup>This group includes linkage-eligible survey participants who linked to Medicaid administrative records at any time during the linkage interval (1999-2014).

<sup>3</sup>This percentage is calculated by dividing the number of linked survey participants by the number of participants in the total sample.

<sup>4</sup>This percentage is calculated by dividing the number of linked survey participants by the total number of linkage-eligible participants.

**Linked NCHS-CMS Medicaid Data  
Linkage Methodology and Analytic Considerations**

**Table 4. Unweighted Percentage of Linked NCHS-Medicaid Sample Enrolled in a Medicaid Comprehensive Managed Care Program for at Least 1 Month<sup>1</sup>, during 1999–2014 MAX Data Years, by Survey**

<u>Survey</u>	MAX Data Year															
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
<b>NHIS 2013</b>	39.7%	38.2%	37.6%	38.9%	37.7%	37.5%	37.8%	39.3%	41.8%	43.1%	44.3%	44.1%	45.9%	49.3%	54.0%	59.4%
<b>NHIS 2012</b>	40.6%	38.8%	38.3%	39.3%	37.5%	36.2%	36.9%	38.6%	41.4%	41.9%	42.9%	43.1%	46.7%	48.6%	54.3%	59.7%
<b>NHIS 2011</b>	41.5%	39.8%	39.5%	40.3%	39.7%	39.1%	39.6%	43.1%	44.4%	44.1%	46.2%	46.4%	48.7%	52.1%	57.9%	64.0%
<b>NHIS 2010</b>	41.5%	41.1%	40.9%	42.9%	40.3%	40.6%	40.6%	42.8%	45.4%	45.7%	46.6%	47.0%	48.1%	51.7%	58.7%	65.3%
<b>NHIS 2009</b>	42.9%	41.4%	40.7%	42.2%	39.4%	38.9%	39.1%	42.5%	44.6%	44.9%	47.0%	46.1%	48.9%	51.5%	59.5%	64.7%
<b>NHIS 2008</b>	40.9%	38.3%	39.3%	38.9%	36.4%	37.1%	37.9%	40.2%	42.0%	43.2%	44.2%	42.5%	45.0%	47.7%	54.9%	58.5%
<b>NHIS 2007</b>	40.3%	40.5%	40.0%	40.1%	37.7%	37.4%	38.2%	39.4%	41.5%	42.4%	42.7%	40.8%	44.4%	47.9%	55.2%	60.9%
<b>NHIS 2006</b>	38.5%	37.4%	38.5%	38.1%	36.0%	36.4%	37.1%	39.9%	40.1%	40.2%	42.0%	40.9%	43.1%	45.8%	53.5%	56.4%
<b>NHIS 2005</b>	45.0%	43.8%	42.5%	45.1%	41.9%	41.3%	42.2%	44.9%	45.8%	45.4%	47.2%	46.9%	49.5%	53.2%	61.1%	67.0%
<b>NHIS 2004</b>	42.4%	41.5%	42.3%	43.3%	41.6%	41.3%	42.1%	43.6%	44.0%	44.6%	45.8%	44.6%	45.6%	49.7%	57.8%	62.2%
<b>NHIS 2003</b>	42.6%	42.0%	41.5%	43.8%	42.4%	41.9%	42.0%	44.9%	45.8%	45.5%	46.9%	45.7%	46.9%	50.1%	58.0%	63.3%
<b>NHIS 2002</b>	44.6%	43.8%	43.5%	44.6%	42.1%	41.1%	40.7%	42.6%	44.1%	43.8%	45.4%	45.0%	46.4%	48.5%	56.8%	62.4%
<b>NHIS 2001</b>	42.8%	41.8%	43.2%	43.6%	40.0%	39.1%	38.6%	40.2%	42.7%	43.6%	45.5%	44.1%	45.1%	48.9%	57.6%	62.5%
<b>NHIS 2000</b>	42.6%	42.2%	41.6%	43.1%	40.0%	39.9%	39.5%	40.9%	41.1%	42.5%	44.2%	43.7%	44.4%	46.3%	55.7%	59.9%
<b>NHIS 1999</b>	41.7%	40.4%	39.4%	40.8%	37.1%	37.1%	37.4%	38.3%	40.5%	39.8%	41.7%	41.0%	42.0%	44.7%	52.1%	59.4%
<b>NHIS 1998</b>	42.1%	40.5%	39.0%	40.0%	37.4%	36.6%	36.7%	36.7%	38.1%	38.3%	40.2%	39.7%	40.1%	43.4%	49.4%	55.9%
<b>NHIS 1997</b>	40.9%	38.8%	37.6%	38.7%	35.4%	34.9%	35.2%	35.8%	37.0%	36.3%	38.3%	36.8%	37.5%	40.5%	46.0%	51.9%
<b>NHIS 1996</b>	38.5%	37.0%	36.5%	37.3%	34.2%	34.6%	34.8%	35.5%	35.7%	35.9%	37.6%	35.1%	36.5%	38.6%	41.5%	49.5%
<b>NHIS 1995</b>	38.0%	36.8%	36.4%	36.8%	33.6%	33.0%	33.6%	34.0%	34.5%	33.4%	34.8%	33.1%	34.3%	36.4%	41.7%	49.8%
<b>NHIS 1994</b>	37.7%	35.2%	34.3%	35.2%	30.6%	29.9%	30.2%	31.4%	32.4%	31.8%	32.2%	29.6%	30.7%	32.3%	38.6%	46.1%
<b>LSOA II</b>	8.8%	8.1%	7.5%	7.0%	3.5%	4.5%	5.3%	5.0%	6.1%	7.1%	8.5%	7.4%	9.9%	13.3%	15.9%	25.0%

<sup>1</sup>The number of months of enrollment in a Medicaid comprehensive managed care plan (Medicaid Analytic eXtract [MAX] element EL\_PPH\_PLN\_MO\_CNT\_CMCP) was used to compute data shown. Data represent persons having only one MAX record per year.

**Linked NCHS-CMS Medicaid Data  
Linkage Methodology and Analytic Considerations**

**Table 4. Unweighted Percentage of Linked NCHS-Medicaid Sample Enrolled in a Medicaid Comprehensive Managed Care Program for at Least 1 Month<sup>1</sup>, during 1999–2014 MAX Data Years, by Survey - continued**

<u>Survey</u>	MAX Data Year															
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
<b>NHANES 2011-2012</b>	41.5%	40.0%	44.4%	47.1%	49.0%	47.2%	48.3%	51.0%	52.9%	52.7%	52.7%	52.5%	55.7%	58.8%	70.1%	69.3%
<b>NHANES 2009-2010</b>	47.9%	45.7%	44.2%	47.3%	46.7%	46.0%	47.0%	50.7%	52.7%	54.6%	57.2%	57.1%	57.5%	61.9%	68.7%	65.3%
<b>NHANES 2007-2008</b>	45.8%	46.1%	46.5%	46.5%	48.0%	48.7%	49.6%	51.5%	51.9%	51.9%	50.2%	50.4%	52.2%	60.6%	71.7%	74.9%
<b>NHANES 2005-2006</b>	48.2%	46.4%	47.4%	51.3%	46.7%	42.9%	44.5%	45.5%	45.3%	47.7%	50.3%	50.7%	52.0%	51.1%	71.8%	76.6%
<b>NHANES 2003-2004</b>	53.7%	53.1%	52.6%	54.5%	52.7%	50.8%	49.9%	49.8%	53.8%	53.8%	52.0%	51.8%	53.2%	58.3%	64.7%	69.1%
<b>NHANES 2001-2002</b>	45.9%	45.4%	49.0%	49.2%	46.3%	44.7%	43.6%	48.9%	53.6%	53.4%	54.5%	53.8%	56.5%	59.5%	64.5%	65.7%
<b>NHANES 1999-2000</b>	35.3%	37.5%	37.1%	39.3%	39.9%	40.9%	41.3%	43.2%	42.1%	42.9%	43.4%	43.2%	47.0%	51.7%	59.5%	70.1%
<b>NHANES III</b>	36.1%	34.8%	33.0%	33.4%	31.3%	30.4%	29.1%	30.4%	29.0%	28.2%	28.0%	27.5%	30.4%	36.8%	41.3%	47.7%
<b>NHEFS</b>	17.5%	16.1%	16.3%	15.3%	7.0%	7.1%	8.2%	8.8%	10.8%	11.5%	14.8%	12.4%	12.8%	14.5%	17.4%	23.5%
<b>NHCS 2011</b>	29.8%	30.6%	30.1%	31.6%	30.6%	29.8%	29.0%	20.5%	19.6%	25.2%	23.5%	18.2%	25.5%	23.3%	9.0%	12.7%
<b>NHHCS 2007</b>	12.2%	10.5%	9.1%	9.1%	6.7%	6.2%	6.0%	6.7%	7.6%	7.9%	9.2%	9.0%	11.5%	13.0%	18.3%	30.8%
<b>NNHS 2004</b>	9.9%	9.0%	8.3%	7.5%	4.4%	4.2%	4.2%	4.1%	4.9%	6.3%	7.1%	7.5%	7.7%	9.4%	12.9%	22.0%

<sup>1</sup>The number of months of enrollment in a Medicaid comprehensive managed care plan (Medicaid Analytic eXtract [MAX] element EL\_PPH\_PLN\_MO\_CNT\_CMCP) was used to compute data shown. Data represent persons having only one MAX record per year.

**Linked NCHS-CMS Medicaid Data  
Linkage Methodology and Analytic Considerations**

**Table 5. Unweighted Percentage of Linked NCHS-Medicaid Sample Enrolled in a Medicaid Comprehensive Managed Care Program for the Full Year<sup>1</sup>, during 1999–2014 MAX Data Years, by Survey**

<u>Survey</u>	MAX Data Year															
	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
<b>NHIS 2013</b>	17.5%	17.2%	18.0%	16.5%	17.7%	17.2%	18.2%	17.0%	17.3%	18.7%	21.3%	22.7%	23.6%	24.8%	31.4%	37.6%
<b>NHIS 2012</b>	18.2%	18.7%	19.1%	17.2%	17.4%	18.1%	18.3%	17.2%	17.8%	19.5%	22.0%	22.7%	23.8%	26.6%	32.4%	37.8%
<b>NHIS 2011</b>	18.0%	17.8%	19.4%	17.7%	19.1%	19.6%	20.5%	19.5%	19.9%	20.7%	23.2%	25.2%	25.9%	27.5%	35.1%	41.7%
<b>NHIS 2010</b>	18.6%	18.8%	21.1%	18.5%	20.0%	18.9%	19.2%	18.5%	18.7%	22.7%	23.2%	24.3%	26.4%	27.3%	34.5%	43.4%
<b>NHIS 2009</b>	19.8%	18.9%	19.6%	18.5%	18.8%	18.4%	19.9%	18.2%	19.1%	21.2%	23.2%	24.9%	25.7%	28.4%	34.7%	40.5%
<b>NHIS 2008</b>	18.9%	18.1%	18.2%	16.8%	16.6%	17.0%	17.9%	17.5%	18.5%	21.9%	22.9%	24.1%	23.6%	26.2%	33.3%	36.5%
<b>NHIS 2007</b>	17.6%	19.5%	18.4%	17.6%	17.0%	18.7%	19.0%	17.9%	19.5%	20.2%	22.4%	22.1%	23.6%	25.5%	33.1%	37.8%
<b>NHIS 2006</b>	15.9%	16.6%	17.3%	15.0%	15.9%	17.6%	17.6%	17.7%	18.4%	19.3%	19.3%	22.7%	22.9%	25.1%	31.2%	34.1%
<b>NHIS 2005</b>	21.2%	20.5%	20.6%	18.7%	20.9%	21.2%	21.8%	21.7%	21.6%	22.8%	25.6%	26.1%	26.1%	29.4%	37.9%	42.1%
<b>NHIS 2004</b>	18.9%	19.9%	20.6%	19.0%	21.0%	21.0%	21.6%	20.8%	20.7%	22.0%	24.6%	24.4%	25.0%	26.4%	34.2%	39.0%
<b>NHIS 2003</b>	18.6%	18.6%	21.1%	19.6%	21.2%	22.2%	22.5%	21.9%	22.6%	22.1%	24.3%	25.3%	24.7%	27.3%	35.5%	41.4%
<b>NHIS 2002</b>	20.0%	20.3%	21.4%	20.3%	21.7%	21.6%	22.3%	21.1%	20.5%	22.8%	23.8%	25.3%	25.2%	26.8%	34.4%	38.8%
<b>NHIS 2001</b>	20.0%	20.3%	21.5%	20.5%	21.2%	20.5%	20.2%	19.9%	20.2%	22.1%	24.4%	25.2%	25.5%	26.9%	36.2%	39.3%
<b>NHIS 2000</b>	20.7%	20.0%	21.9%	20.1%	21.2%	20.6%	21.2%	20.5%	21.3%	21.3%	23.2%	24.1%	24.2%	25.2%	32.7%	35.9%
<b>NHIS 1999</b>	19.5%	19.3%	21.0%	19.1%	19.9%	19.4%	20.0%	18.7%	19.9%	21.2%	22.7%	23.3%	23.5%	24.0%	30.7%	38.4%
<b>NHIS 1998</b>	21.3%	21.5%	22.1%	19.2%	20.1%	18.9%	19.2%	18.4%	19.1%	20.3%	21.6%	22.8%	22.2%	22.7%	29.0%	33.7%
<b>NHIS 1997</b>	19.8%	19.7%	20.7%	17.8%	18.7%	18.2%	19.0%	18.1%	18.3%	18.9%	20.1%	20.6%	20.4%	22.1%	27.5%	30.8%
<b>NHIS 1996</b>	20.0%	19.5%	19.9%	17.3%	18.9%	18.7%	20.0%	18.8%	17.8%	18.9%	20.8%	20.7%	19.6%	20.8%	25.8%	30.2%
<b>NHIS 1995</b>	19.4%	18.9%	19.7%	17.8%	17.9%	17.6%	18.4%	17.2%	16.8%	17.6%	18.6%	18.7%	17.8%	19.5%	25.7%	29.4%
<b>NHIS 1994</b>	19.0%	18.6%	19.3%	16.6%	16.8%	16.2%	16.2%	16.1%	15.9%	16.7%	18.0%	17.3%	15.7%	17.3%	22.7%	28.3%
<b>LSOA II</b>	5.3%	4.8%	4.6%	2.4%	2.2%	2.9%	3.6%	3.0%	3.5%	4.7%	4.8%	5.2%	5.4%	7.3%	9.8%	15.0%

<sup>1</sup>The number of months of enrollment in a Medicaid comprehensive managed care plan (Medicaid Analytic eXtract [MAX] element EL\_PPH\_PLN\_MO\_CNT\_CMCP) was used to compute data shown. Data represent persons having only one MAX record per year.

**Linked NCHS-CMS Medicaid Data  
Linkage Methodology and Analytic Considerations**

**Table 5. Unweighted Percentage of Linked NCHS-Medicaid Sample Enrolled in a Medicaid Comprehensive Managed Care Program for the Full Year<sup>1</sup>, during 1999–2014 MAX Data Years, by Survey - continued**

<b>Survey</b>	<b>MAX Data Year</b>															
	<b>1999</b>	<b>2000</b>	<b>2001</b>	<b>2002</b>	<b>2003</b>	<b>2004</b>	<b>2005</b>	<b>2006</b>	<b>2007</b>	<b>2008</b>	<b>2009</b>	<b>2010</b>	<b>2011</b>	<b>2012</b>	<b>2013</b>	<b>2014</b>
<b>NHANES 2011-2012</b>	18.9%	20.5%	19.8%	20.7%	21.8%	25.7%	26.0%	24.6%	23.8%	25.1%	25.8%	25.4%	29.5%	31.1%	43.9%	44.4%
<b>NHANES 2009-2010</b>	20.5%	21.5%	23.3%	21.6%	25.6%	24.1%	26.5%	24.2%	25.0%	25.2%	30.7%	34.7%	33.6%	35.3%	43.0%	43.9%
<b>NHANES 2007-2008</b>	21.5%	20.9%	24.7%	24.4%	26.3%	27.4%	27.9%	25.4%	24.6%	26.9%	28.0%	28.0%	28.7%	29.3%	45.4%	50.7%
<b>NHANES 2005-2006</b>	23.1%	23.4%	24.9%	23.1%	23.7%	23.2%	23.0%	22.2%	21.8%	23.0%	30.2%	31.4%	28.8%	30.8%	44.1%	47.2%
<b>NHANES 2003-2004</b>	29.4%	28.0%	28.5%	27.3%	28.5%	27.4%	29.6%	27.5%	27.8%	30.1%	29.5%	31.9%	31.3%	34.0%	39.6%	43.7%
<b>NHANES 2001-2002</b>	23.6%	22.4%	25.0%	22.5%	23.7%	23.4%	23.4%	22.5%	23.2%	29.7%	31.0%	30.6%	31.4%	36.6%	42.1%	42.3%
<b>NHANES 1999-2000</b>	16.3%	17.5%	18.1%	19.2%	21.6%	23.2%	24.9%	22.7%	21.2%	21.9%	23.0%	24.5%	24.0%	28.7%	40.6%	45.8%
<b>NHANES III</b>	19.3%	19.1%	18.1%	16.8%	17.6%	16.6%	17.2%	14.1%	14.6%	14.6%	15.5%	15.5%	16.3%	19.8%	26.4%	30.0%
<b>NHEFS</b>	10.8%	12.1%	11.6%	4.1%	4.3%	4.3%	5.2%	4.9%	5.3%	6.9%	9.2%	8.5%	8.6%	9.7%	12.9%	16.0%
<b>NHCS 2011</b>	13.5%	13.2%	14.1%	14.3%	15.4%	14.7%	14.5%	6.3%	5.9%	5.6%	6.0%	4.7%	3.0%	8.4%	4.8%	7.7%
<b>NHHCS 2007</b>	6.9%	6.3%	5.9%	3.5%	3.2%	3.7%	3.9%	3.8%	2.3%	4.3%	5.5%	6.3%	6.6%	8.2%	11.0%	22.9%
<b>NNHS 2004</b>	6.8%	6.1%	5.8%	2.9%	2.7%	2.6%	2.5%	2.7%	2.5%	3.3%	5.3%	5.3%	5.7%	6.3%	10.2%	14.1%

<sup>1</sup>The number of months of enrollment in a Medicaid comprehensive managed care plan (Medicaid Analytic eXtract [MAX] element EL\_PPH\_PLN\_MO\_CNT\_CMCP) was used to compute data shown. Data represent persons having only one MAX record per year.

**Linked NCHS-CMS Medicaid Data  
Linkage Methodology and Analytic Considerations**

**Appendix A: Merging Linked NCHS-CMS Medicaid Files with NCHS Survey Data**

The data provided on the 1994-2013 NHIS, NHEFS, 1999-2012 NHANES, NHANES III, LSOA II, 2004 NNHS, and the 2007 NHHCS linked CMS Medicaid files can be merged with the NCHS restricted and public-use survey data files using the unique survey specific participant identification number (PUBLICID/SEQN/RESNUM/PATNUM).

Note: At this time the linked CMS Medicaid data files are only available for research use through the NCHS restricted access data center (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 participant identification number on any analytic files sent to the RDC. The RDC will merge data (using PUBLICID, SEQN, RESNUM or PATNUM) from the linked CMS Medicaid files to the analyst's file. The merged file will be housed at the RDC and made available for analysis.

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

**I. National Health Interview Survey (NHIS)**

**1994 NHIS**

<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)`



**Linked NCHS-CMS Medicaid Data  
Linkage Methodology and Analytic Considerations**

**1995-1996 NHIS**

<u>Variable</u>	Public-use <u>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)
```

**1997-2003 NHIS**

<u>Variable</u>	Public-use <u>Location</u>	<u>Length</u>	<u>Description</u>
SRVY_YR	3-6	4	Year of interview
HHX	7-12	6	Household 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).

**Linked NCHS-CMS Medicaid Data  
Linkage Methodology and Analytic Considerations**

**2004 NHIS**

<u>Variable</u>	Public-use <u>Location</u>	<u>Length</u>	<u>Description</u>
SRVY_YR	3-6	4	Year of interview
HHX	7-12	6	Household number
FMX	13-14	2	Family number
FPX	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||FPX));
```

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

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

**2005-2013 NHIS**

<u>Variable</u>	Public-use <u>Location</u>	<u>Length</u>	<u>Description</u>
SRVY_YR	3-6	4	Year of interview
HHX	7-12	6	Household number
FMX	16-17	2	Family number
FPX	18-19	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||FPX));
```

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

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

**Linked NCHS-CMS Medicaid Data  
Linkage Methodology and Analytic Considerations**

**II. National Health and Nutrition Examination Survey 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 linked NHEFS-CMS Medicaid files using this variable ensures that the appropriate information for each survey participant is linked correctly.

**III. 1999-2012 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 linked NHANES-CMS Medicaid 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 linked NHANES III-CMS Medicaid files using this variable ensures that the appropriate information for each survey participant is linked correctly.

## Linked NCHS-CMS Medicaid Data Linkage Methodology and Analytic Considerations

### 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 linked NNHS-CMS Medicaid files using this variable ensures that the appropriate information for each survey participant is linked correctly.

**Linked NCHS-CMS Medicaid Data  
Linkage Methodology and Analytic Considerations**

**VII. 2007 National Home and Hospice Care Survey (NHHCS)**

<u>Item</u>	<u>Length</u>	<u>Description</u>
PATNUM	6	Patient/Discharge Record (Case) Number

All of the 2007 NHHCS public-use data files are linked with the common patient/discharge record (case) number (PATNUM). Merging information from the 2007 NHHCS files to the linked 2007 NHHCS-CMS Medicaid files using this variable ensures that the appropriate information for each survey participant is linked correctly.