Documentation and Analytic Guidelines for NCHS surveys linked to Medicaid Analytic eXtract (MAX) files

Revised: 2/20/2014

Suggested Citation: Simon AE, Driscoll AK, Golden C, Tandon R, Duran CR, Miller EA, Schoendorf KC, and Parker JD. Documentation and Analytic Guidelines for NCHS surveys linked to Medicaid Analytic eXtract (MAX) files. Hyattsville, MD: National Center for Health Statistics. Revised February 2014. (Available at the following address: http://www.cdc.gov/nchs/data/datalinkage/documentation_and_analytic_guidelines_nchs_survey_max_linked_data.pdf)

This document provides documentation and analytic guidelines for the NCHS surveys linked to the Medicaid Analytic eXtract (MAX) files. Additional information concerning these linked files is forthcoming.

Table of Contents

		Page Numbers
I)	Section I---Introduction	1
II)	Section II—NCHS Surveys	2
III)	Section III--- Medicaid/CHIP and MAX files	3
	A. Background: Medicaid and CHIP programs	3
	B. Medicaid/CHIP Data (MSIS and MAX)	7
IV)	Section IV--- Linkage Process	12
V)	Section V---Analytic Considerations and Guidelines	15
	A. Analytic Sample Size Estimation – Feasibility Files and Tables	15
	B. Procedures for Merging the NCHS and MAX data	16
	C. Weighting of Linked Data Files	17
	D. Identification of Medicaid Enrollees	18
	E. State Differences in Medicaid	19
	F. Managed Care vs. Fee For Service (FFS)	19
	G. Waivers	20
	H. Data Anomalies	21
	I. Variables to Request in Applications to the Research Data Center (RDC)	22
	J. Multiple PS Records	23
	K. Dual Eligibles	25
	L. CHIP	26
	M. Diagnosis and Procedure Coding	27
	N. Creation of Public ID Numbers for File Linkage	27
VI)	Section VI---References	31
	Table 1. NHIS - 1999-2007[1] Medicaid linked files sample sizes and linkage rates, by NHIS year and age at interview	32
	Table 2. NHANES - 1999-2007[1] Medicaid linked files sample sizes and linkage rates, by NHANES year and age at interview	33

[1] 2008 and 2009 Medicaid linked files are currently available. Revised tables including sample sizes and linkage rates for 1999-2009 are forthcoming.

Section I--Introduction

The National Center for Health Statistics (NCHS) and the Centers for Medicare & Medicaid Services (CMS) have prepared a set of analytic data files linking NCHS survey data to CMS Medicaid utilization data. This dataset linkage provides unique opportunities for research and policy analysis. For example, health status, diagnosis, race/ethnicity, and sociodemographic information from the NCHS surveys can be used to augment the longitudinal utilization data available from Medicaid. Similarly, the Medicaid data can be used to improve identification of Medicaid beneficiaries on the NCHS surveys and to determine Medicaid enrollment during years before or after the NCHS survey year.

These data files complement the previously completed NCHS Survey--Medicare data linkages. Both linkage efforts are part of a larger collaboration between 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) in support of new comparative effectiveness research (CER) initiatives.

This report provides documentation and technical notes for the linked NCHS Survey-Medicaid files. Specifically, research-oriented Medicaid files known as the Medicaid Analytic eXtract (MAX) files from 1999-2009 have been linked to the following NCHS surveys:

1) National Health Interview Survey (NHIS), 1994-2005

2) National Health and Nutrition Examination Survey (NHANES), 1999-2000, 2001-2002, and 2003-2004

3) National Nursing Home Survey (NNHS), 2004

4) NHANES I Epidemiologic Follow-up Study (NHEFS), 1971-1975, 1982-84, 1986, 1987, and 1992

5) Third National Health and Nutrition Examination Survey (NHANES III), 1988-1994

6) The Second Longitudinal Study of Aging (LSOA II), 1994, 1997-1998, and 1999-2000

The MAX files include data for enrollees in Medicaid and in the Children's Health Insurance Program (CHIP) in all 50 states and the District of Columbia. As described in more detail below, they also include data from enrollees who received Medicaid due to expanded Medicaid programs resulting from CHIP legislation as well as incomplete data (discussed below) on children enrolled in CHIP programs administered separately from Medicaid. The MAX files include enrollee level information on enrollment across time, reasons for eligibility for Medicaid

or CHIP, type of insurance plan (fee-for-service or managed care), and for some enrollees, data on utilization and payments to insurance companies.

This data linkage provides the opportunity to examine the Medicaid and CHIP data in the MAX files for NCHS survey respondents during the year the NCHS survey was conducted, in years after the survey was conducted, and/or in years prior to the NCHS survey.

The resulting linked files will be added to a growing list of linked survey-administrative datasets available for use in the NCHS Research Data Center (RDC). Researchers who wish to use the restricted-use data may apply to the NCHS RDC for access. For more information or questions about the linked data, please contact the NCHS Special Projects Branch at datalinkage@cdc.gov.

For instructions about how to apply to the NCHS RDC to use these linked data sets, please contact the RDC at rdca@cdc.gov or view the website for the Research Data Center at http://www.cdc.gov/rdc/.

Terminology and abbreviations related to Medicaid and CHIP programs can present challenges. To the extent possible in this document, we have defined terms when they are used. In addition, glossaries of terminology relating to health insurance programs can be found through the following organizations at the websites below.

1) Centers for Medicare & Medicaid Services:
   https://www.cms.gov/apps/glossary/

2) American Academy of Family Physicians:
   http://www.aafp.org/online/en/home/publications/news/news-now/payment-special-report/20120921paymentsrglossary.html

3) BlueCross BlueShield Association:
   http://www.bcbs.com/glossary/

Section II--NCHS Surveys

NCHS is the nation's principal health statistics agency, charged with compiling statistical information to guide actions and policies to improve the health of the United States. Specifically, NCHS is responsible for compiling and organizing vital statistics for the US, conducting population surveys such as NHANES and NHIS, and conducting establishment surveys of health-care facilities such as the NNHS and the National Ambulatory Medical Care Survey. Data from these surveys and others allow NCHS to:

- document the health status of the population and of important subgroups
- identify disparities in health status and use of health care by race or ethnicity, socioeconomic status, region, and other population characteristics
- describe experiences with the health care system
- monitor trends in health status and health care delivery
- evaluate the impact of health policies and programs

Not all NCHS surveys can be linked to MAX files because some NCHS surveys do not collect the information necessary for linkage. Even within NCHS surveys linked to MAX files, not all respondents are "linkage-eligible," meaning that they did not supply sufficient information to enable the linkage. Thus, Medicaid/CHIP enrollment and claims data are available for those NCHS respondents who were enrolled in Medicaid, provided personal identification data to NCHS, and for whom NCHS was able to match with Medicaid/CHIP administrative records. Data from MAX files for 1999-2009 are available for the following NCHS surveys:

- National Health Interview Survey (NHIS), 1994-2005
- National Health and Nutrition Examination Survey (NHANES), 1999-2004
- National Nursing Home Survey (NNHS), 2004
- NHANES I Epidemiologic Follow-up Study (NHEFS), 1971-1975, 1982-84, 1986, 1987, and 1992
- Third National Health and Nutrition Examination Survey (NHANES III), 1988-1994
- The Second Longitudinal Study of Aging (LSOA II), 1994, 1997-1998, and 1999-2000

Information about each of these NCHS surveys can be obtained at the NCHS website at: http://www.cdc.gov/nchs/index.htm.

Section III---Medicaid/CHIP and MAX Files

A) Background: Medicaid and CHIP programs

Definition/Description

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. The program was enacted in 1965 by Title XIX of the Social Security Act and has since grown to become the third largest source of health care spending in the U.S. after Medicare and employer-provided health insurance. Medicaid accounts for almost a sixth of total national spending on personal health care. Since the 1990s, the number of Medicaid enrollees has exceeded the number enrolled in Medicare.[1]

Medicaid is the main payer of nursing home care and long-term care services overall; it is also the largest source of public funding for mental health care. Health centers and safety-net hospitals that serve low-income and uninsured people rely heavily on Medicaid revenues. Although seniors and people with disabilities make up approximately 25% of all Medicaid enrollees, these two groups account for two-thirds of Medicaid benefit expenditures.[2]

States & Services Covered

Medicaid is administered by states under general guidelines established by the federal government and is financed jointly by federal and state funds. The Federal Medical Assistance Percentage (FMAP), also called the federal match rate, represents the percent of Medicaid

financed by the federal government in each state. The FMAP differs by state and takes into account the average per capita income in a state relative to the national average. For fiscal year 2011, in higher income states, the FMAP is 50 percent; in low-income states, it is as high as 75 percent. The FMAP for 2011 for individual states can be found at http://aspe.hhs.gov/health/fmap11.htm and for earlier years, can be found at http://aspe.hhs.gov/health/fmapearly.htm.

State Medicaid programs must cover mandatory services specified in federal law to receive federal matching funds. Beneficiaries are entitled to receive the mandatory services listed in Box 1.

Box 1.

States are required to provide the following mandatory services:

- Physicians' services
- Hospital services (inpatient and outpatient)
- Laboratory and x-ray services
- Early and periodic screening, diagnostic and treatment (EPSDT) services for persons under 21
- Federally-qualified health center and rural health clinic services
- Family planning services and supplies
- Pediatric and family nurse practitioner services
- Nurse midwife services
- Nursing facility services for persons 21 and older
- Home health care for persons eligible for nursing facility services
- Transportation services
- Medicaid long-term care services

Medicaid long-term care services include comprehensive services provided in nursing homes and intermediate care facilities for the mentally retarded (ICF-MR). Long-term care also includes a wide range of services and supports needed by people to live independently in the community, including home health care, personal care, medical equipment, rehabilitative therapy, adult day care, case management and respite for caregivers.

States are also permitted to cover many services that federal law designates as optional, including dental services, prescription drugs, case management, and hospice services. State variation in Medicaid coverage, with regard to both program eligibility and covered services, results in state differences in enrollment rates and expenditures. Other factors, including the age distribution, the poverty rate, and the Medicaid provider reimbursement rates, also contribute to variation among states in enrollment, service use, and costs. As a result, Medicaid operates as more than 50 distinct programs – one in each state, the District of Columbia, and each of the territories.

Medicaid Eligibility

To qualify for Medicaid, a person must belong to one of the mandatory eligibility groups whose criteria include age and poverty level as well as, in some groups, eligibility for other programs. Federal law requires states to cover certain mandatory eligibility groups in order to receive any federal matching funds, but they have broad flexibility to determine their own methods for assessing income and may also impose asset tests (Box 2).

Box 2.

States are required to cover these mandatory eligibility groups:

- Pregnant women with family income below 133% of the federal poverty level (FPL)
- Infants (up to age 1) born to Medicaid-eligible pregnant women
- Children under 6 with family income below 133% FPL
- Children 6-18 below 100% FPL
- Limited-income families with dependent children who meet states' July 1996 welfare eligibility levels (often below 50% FPL)
- Supplemental Security Income (SSI) recipients

FPL cutoffs for children and parents vary across states and over time. Most elderly and persons with disabilities who receive Supplemental Security Income (SSI) and most low-income aged and disabled Medicare beneficiaries are eligible for Medicaid.[1] Medicare beneficiaries must meet state income and assets criteria.

According to the terms of the Patient Protection and Affordable Care Act of 2010, by 2014 Medicaid will be expanded to include nearly everyone under age 65 with income up to 133% of FPL.

Some Medicaid enrollees may be enrolled in both Medicaid and Medicare, and are referred to as dual eligible beneficiaries, or simply "dual eligibles". In 2007, 8.9 million, or 15.3%, of Medicaid enrollees were also enrolled in Medicare.[3] Several states provide special programs to assist in caring for this group. More information is provided below on how to best identify these enrollees in the linked data, as well as the types of payment data provided for dual eligibles in the MAX files. In addition, more information about dual eligibles can be found through CMS at http://www.cms.gov/Medicare-Medicaid-Coordination/Medicare-MedicaidCoordination.html.

Delivery of Care

Health care through Medicaid is delivered through fee-for-service (FFS) and managed care programs. Medicaid managed care programs are insurance plans whereby a health care organization agrees to provide a defined bundle of health services in return for a fixed monthly fee paid by the state's Medicaid program. States use an array of different types of managed care arrangements in Medicaid. Medicaid managed care plans include comprehensive plans - such as Health Maintenance Organizations (HMO), Health Insurance Organizations (HIO), and Prepaid Health Plans (PHP) - that cover most (but not necessarily all) enrollee health services. Other plans provide more limited services, and coverage varies greatly by plan. A plan may, for example, cover only dental care or behavioral health services. Primary care case management plans (PCCMs) are the least comprehensive managed care type. They involve the payment of a small premium for case management services only.

Since the 1990s, state Medicaid programs have increasingly relied on managed care to organize and deliver services. For example, in 1999, total managed care enrollment accounted for approximately 56% of the Medicaid population.[4] By 2009, about 75% of Medicaid and CHIP enrollees receive some or all of their services through managed care.[5]

Fewer data are available for Medicaid managed care recipients than for those receiving care through fee-for-service plans. The ramifications of this difference are discussed below.

Children's Health Insurance Program (CHIP)

Originally implemented in 1997 via Title XXI of the Social Security Act, CHIP is jointly financed by the federal and state governments, and is administered by the states. It targets low income uninsured children and pregnant women in families with incomes too high to qualify for most state Medicaid programs. Within broad federal guidelines, each state determines its program's design, eligibility groups, benefits, payment levels, and administrative and operating procedures. The federal government provides funding to states for CHIP but caps the amount of funds provided on a matching basis.

Each state has the option of expanding Medicaid eligibility to children who previously had been ineligible due to their income (M-CHIP) and/or creating a program distinct from its existing Medicaid program (S-CHIP). Distinctions are made between these two types of programs in various ways, but this report will refer to expansions of existing Medicaid programs as M-CHIP and programs distinct from existing Medicaid programs as S-CHIP. The federal government matches state spending for CHIP at a higher rate than that for Medicaid. As of 2009, more than 30 states had an S-CHIP program, either by itself or in combination with an M-CHIP program in the same state.[6] S-CHIP programs are allowed to tailor their programs to a greater extent than is permitted for M-CHIP programs. They can, for example, offer different benefit packages, require cost-sharing and premiums, and mandate waiting periods.[6] Data availability is different for M-CHIP and S-CHIP recipients, as discussed below.

B) Medicaid/CHIP Data (MSIS and MAX)

Medicaid data have been collected by each state since 1999 and provided to CMS through the Medical Statistical Information System (MSIS). These data contain enrollee eligibility information and Medicaid claims paid in each quarter of the federal fiscal year (FFY). Data are extracted from the MSIS to create the MAX files which are used to link to the NCHS surveys. The MSIS and MAX files are further described below.

MSIS

The purpose of MSIS is to collect, manage, analyze and disseminate information on enrollees, including utilization and payment for services covered by state Medicaid/CHIP programs. In general, each state file contains one record for each person covered by Medicaid/CHIP for at least one day during the reporting quarter. The files contain specified data elements for: (1) persons covered by Medicaid/CHIP (Eligible files) which consist of demographic and monthly enrollment data; and (2) adjudicated claims (Paid Claims files) which consist of information from medical service-related claims and capitation payments. States submit four types of claims files representing inpatient (IP), long term care (LT), prescription drugs (RX) and other non-institutional services (OT). Claims records contain information on the types of services provided, service providers and dates, costs, and types of reimbursement. The files serve as the historical source of detailed Medicaid eligibility and paid claims data maintained by CMS.

MAX

MAX data are research extracts of MSIS. The MAX data system was developed to provide calendar-year utilization and expenditure information and serve as a research tool for examining Medicaid/CHIP enrollment, service utilization, and expenditures by subgroup and over time. MAX contains individual-level information on demographics, monthly enrollment status, eligibility group, and use and costs of services during the year. It also includes claims-level records for more detailed analysis of patterns of service utilization, diagnoses, and cost of care among Medicaid/CHIP enrollees.

The MAX data systems have been modified from MSIS in a number of ways to enhance their usefulness for research purposes and enable individual enrollee level analysis. While MSIS claims files contain separate records for initial claims, voided claims, and positive or negative adjustments, such records are combined to reflect final service event records in MAX. Changes in program eligibility reported retroactively are incorporated in MAX monthly enrollment measures. MAX data are linked to the Medicare Enrollment Database to help identify people dually enrolled in Medicare and Medicaid.

MAX includes summary information and claims data for all Medicaid enrollees in each state and the District of Columbia but not those in Puerto Rico or other U.S. territories. The MAX data system consists of a person summary (PS) file and four claims files: inpatient (IP), institutional long-term care (LT), prescription drugs (RX) and other services (OT). Each of these files is described in greater detail below.

Person Summary (PS) Files

The Person Summary (PS) file for each year of MAX data contains one record for each person included in the MAX file, including people enrolled in Medicaid, M-CHIP, and some (but not all) people enrolled in S-CHIP. In some cases, as described further below, a beneficiary may have more than one person summary 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.

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 which allows 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, other). These measures are combined into one variable in the PS file (Box 3).

Box 3.

EL_MAX_ELGBLTY_CD_LTST concatenates maintenance of assistance status (MAS) and basis of eligibility (BOE). MAS, which describes the primary financial eligibility criteria met by the enrollee, is in the first position; BOE, which represents the population subgroup through which a person becomes eligible, 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. The data dictionary for the PS file can be found at ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/medicaid/MAX_2009_PS_Dictionary.pdf.

Inpatient Hospital (IP) Files

The inpatient hospital (IP) file contains complete stay records for Medicaid enrollees who used inpatient hospital services. Data include 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. The data dictionary for the IP file can be found at ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/medicaid/MAX_2009_IP_Dictionary.pdf.

Long-term Care (LT) Files

The LT file 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 retarded (ICF/MR), and nursing facilities (NF). Information in the LT file includes beginning 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. However, 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. The data dictionary for the LT file can be found at ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/medicaid/MAX_2009_LT_Dictionary.pdf.

Prescription Drug (RX) Files

The prescription drug (RX) file contains prescribed drugs, over-the-counter drugs, and other items dispensed by a free-standing pharmacy (non-hospital based). Information in the RX file includes prescription fill date, new or refill indicator, National Drug Code (NDC) and quantity and day supply. Also included are payment amounts, third-party payments, and Medicaid-paid Medicare copayment and deductible amounts.

This file does not include drugs provided during an inpatient hospital stay. Injectable drugs administered by a health professional are included in the Other Services (OT file) described below. Beginning in 2006, full-benefit dual eligibles receive 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 at http://www.ccwdata.org. Also, NCHS surveys have been linked to Medicare Part D PDE data and more information about these linked data can be found at http://www.cdc.gov/nchs/data_access/data_linkage/cms_medicare.htm. Drugs provided to patients in LTC facilities are included in this file for most states. However, a small number of states "bundle" these drugs in their LTC facility payment rate; in these cases, no details on drugs are available.

As with the PS and other files, occasional additions or changes in variables have been made to the RX files over time, so careful examination of the data dictionaries prior to analyses is suggested. The data dictionary for the RX file can be found at ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/medicaid/MAX_2009_RX_Dictionary.pdf.

Other Services (OT) Files

The Other Services (OT) file 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, HIOs, PHPs, and PCCMs.

The service types in the OT file include physician and professional services, outpatient and clinic services, durable medical equipment (DME), hospice, home health and lab/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.

Because of the nature of billing 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 rather 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.

The data dictionary for the OT file can be found at ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/medicaid/MAX_2009_OT_Dictionary.pdf.

Limitations of MAX

There are some 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 (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 (Box 4).

Box 4.

MAX files do not include the following:

- Services or expenditures not paid by Medicaid
- Prescription drug rebates
- Medicaid payments to DSH (disproportionate share hospitals)
- Medicaid payments made through UPL (upper payment limit) programs
- Medicaid payments to states for 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 eligibles), 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 that are 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.

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 fee-for-service 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.

In addition, whereas all persons with Medicaid or M-CHIP are included in the MAX files, only some S-CHIP enrollees are included in the MAX files. Further, S-CHIP enrollees that are included in the MAX files may have incomplete data. Further details on these omissions are provided below.

Finally, as with all large data sets, MAX contains some anomalous and possibly incomplete or incorrect data elements. Users should consult MAX anomaly notes, available on the CMS website that describes the MAX files (see Resources for MAX below), for information that may explain unusual patterns in each state's data.

Resources for MAX

MAX files have many intricacies and these must be considered when working with the linked NCHS - MAX data. To assist researchers interested in using the MAX files, we recommend the following resources to learn more about the MAX data, as well as the Medicaid and CHIP programs.

- Centers for Medicare & Medicaid Services
  http://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAXGeneralInformation.html

- The Research Data Assistance Center (ResDAC) at the University of Minnesota
  http://www.resdac.org/

- The Henry J. Kaiser Family Foundation
  http://www.statehealthfacts.org/
  http://medicaidbenefits.kff.org/

- National Pharmaceutical Council
  http://www.npcnow.org/Public/Issues/i_rel_research/Medicaid_Pharmaceutical_Plan_Resources.aspx

Section IV--Linkage Process

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

The process for the collection of this information and the determination of linkage-eligibility has changed over time and differs across surveys. Of note, the proportion of NHIS respondents who provide their SSN varies by survey year and by respondent characteritics[7]. For example, the proportion of respondents 18 years of age and older in the NHIS who provided SSN decreased from 74% in 1997 to a low of 43% in 2006 Therefore, the linkage-eligible records in any given NCHS survey comprise a subset of the respondents to the survey. Of the linkage-eligible, the actual number of records linked depends both on program enrollment and the linkage process. A small number of records that are both linkage eligible and enrolled in Medicaid/CHIP cannot be linked to MAX records due to errors in identifying information either from the survey or MAX data.

NCHS survey participants who were under 18 years of age at the time of the survey are considered linkage-eligible if the listed linkage-eligibility criteria is met and consent is provided by their parent or guardian. However, in accordance with NCHS ERB guidelines, NCHS will only provide linked CMS data generated for program participation, claims and other events that occurred prior to the participant's 18th birthday. Please see the data linkage website for more information on this NCHS ERB guidance:
http://www.cdc.gov/nchs/data/datalinkage/nchs_survey_participants_under_age_18.pdf

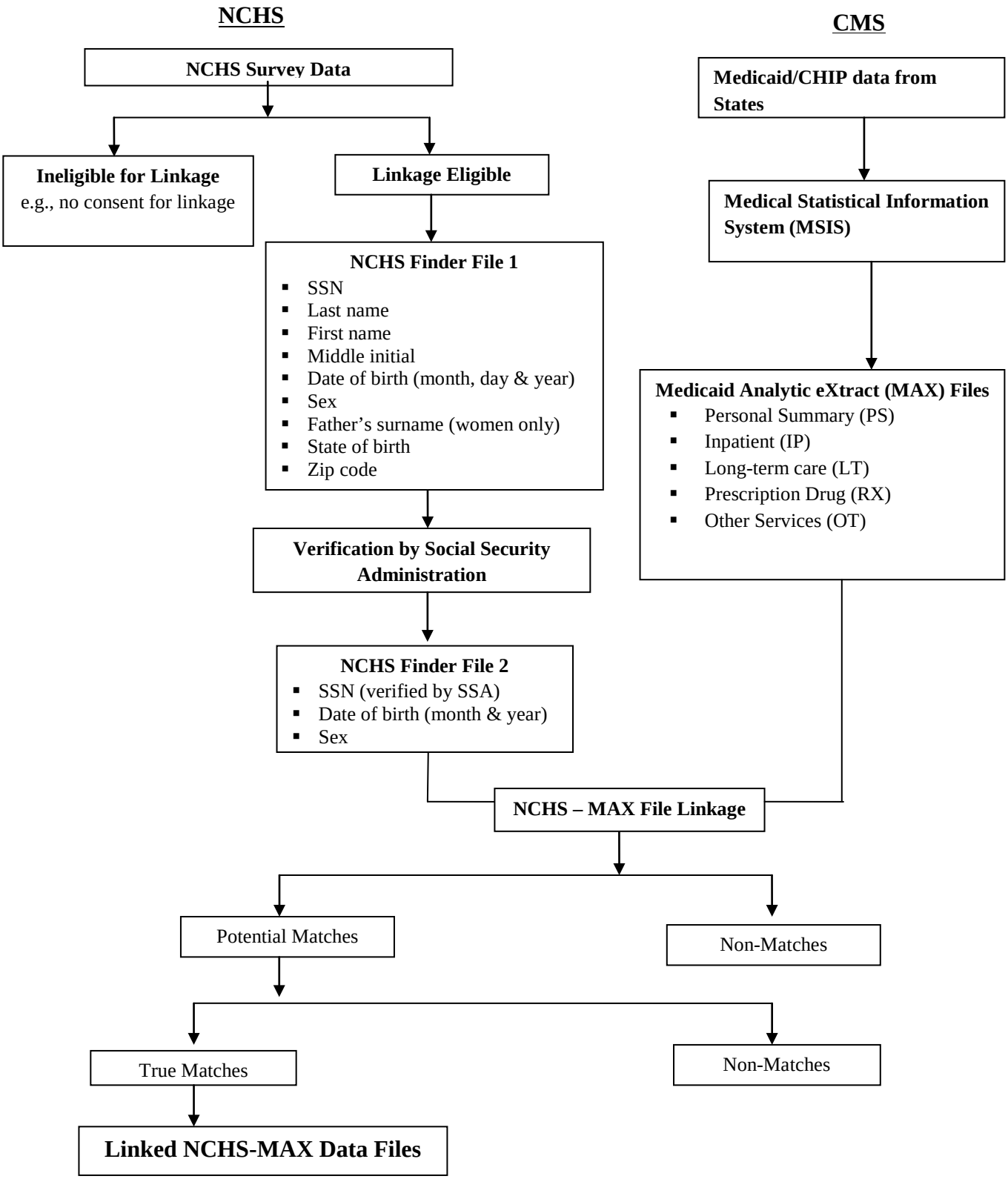
Linking Algorithm

To link the NCHS survey data, all linkage-eligible records from the NCHS surveys were combined into one file, the NCHS Finder File number 1, and sent to the Social Security Administration (SSA) for verification of SSNs. After processing by SSA, most NCHS records in the resulting NCHS Finder File (NCHS Finder File number 2) contained an SSN verified or assigned by SSA, although a small subset of records (0.1%) had a missing SSN. SSA sent the augmented and verified file, NCHS Finder File number 2, to CMS for linkage to the MAX file. The resulting files containing the NCHS records linked to the MAX data were encrypted and sent to NCHS where the processing was completed.

For NCHS records from the verified file to link to the MAX PS file, an exact match on SSN, month and year of birth, and sex was required. From the linked NCHS-PS file, CMS extracted the associated claims from each of the four MAX claims files from the respective state. CMS linked NCHS data separately for each state and then combined the subsets into one national file. The procedure by which linked files were created is displayed in Figure 1 below.

MAX records with missing SSNs are common because many states are unable to collect SSNs from some enrollees. Most of those without an SSN are children, persons who qualify only for family planning benefits, or aliens who qualify only for emergency coverage. CMS removed records without SSNs from the MAX PS file before linking it to the NCHS finder file.

Figure 1. *Linkage algorithm for NCHS survey data linked to CMS MAX files*



To further assist researchers in understanding the linkage eligibility and linkage rates, Tables 1 and 2 provide the full sample numbers for the NHIS and NHANES surveys, the numbers that are linkage eligible in each year and the number actually linked because they are in the MAX file.

Section V—Analytic Considerations and Guidelines

A) Analytic Sample Size Estimation – Feasibility Files and Tables

The NCHS-MAX feasibility study data files allow researchers to obtain approximate sample size estimates for potential analyses and help them determine whether their research idea is feasible with the linked NCHS-MAX data set. The feasibility files indicate which NCHS survey respondents are eligible (and ineligible) for MAX linkage, their final match status, and the specific linked MAX files available for each successfully linked survey respondent. Each feasibility file is NCHS survey- and survey year-specific. For each NCHS survey and survey year, the researcher can use the appropriate feasibility file to determine the number of respondents linked to each of the available MAX files. Also, using these files, researchers may examine the relationships between survey respondent characteristics from the NCHS surveys and linkage eligibility. Each NCHS survey (NHIS, NHANES, NNHS, NHEFS, NHANES III, and LSOA II) uses a different identification variable allowing linkage to the feasibility files and additional variables (Box 5 and section N below). NCHS-MAX feasibility study files can be found at
http://www.cdc.gov/NCHS/data_access/data_linkage/cms/cms_medicaid_feasibility.htm

Researchers should be aware that all NCHS surveys linked to the MAX files have complex survey designs. Therefore, considerations of statistical power should account for the design of the survey, in addition to the unweighted number of observations available for a particular project.

Box 5.

Feasibility files for each NCHS survey include:

- **PUBLICID**: Public ID variable for merging NHIS public use survey records to the feasibility file. The PUBLICID variable must be constructed from other variables provided on the NHIS public use files. Instructions for how to construct PUBLICID can be found in section N below.
- **SEQN**: Public ID variable for merging NHANES public use survey records to the feasibility file
- **RESNUM**: Public ID variable for merging NNHS public use survey records to the feasibility file
- **CMS_MEDICAID_MATCH**: Indicator of the linkage result between NCHS survey participant and MAX data:
  - 1 = Linked to MAX data in any year (1999-2009)
  - 2 = Eligible for MAX linkage but not linked in any year
  - 3 = Child survey participant linked to MAX data - MAX data available prior to the participant's 18th birthday, but information indicating whether child linked after 18th birthday is not available
  - 9 = Not eligible for MAX linkage
- **ON_PS_1999 – ON_PS_2009**: Indicates if NCHS participant linked to PS file in 1999-2009
- **ON_IP_1999 – ON_IP_2009**: Indicates if NCHS participant linked to IP file in 1999-2009
- **ON_LT_1999 – ON_LT_2009**: Indicates if NCHS participant linked to LT file in 1999-2009
- **ON_RX_1999 – ON_RX_2009**: Indicates if NCHS participant linked to RX file in 1999-2009
- **ON_OT_1999 – ON_OT_2009**: Indicates if NCHS participant linked to OT file in 1999-2009

B) Procedures for Merging the NCHS and MAX data

Although MAX and NCHS survey data have been linked, the files are retained separately to allow for the creation of analytic files specific to the needs of researchers. Therefore, the NCHS and MAX data files require merging by the researcher before they can be used for statistical analyses. This section describes how merging of those data should be conducted by the researcher.

The structure of the MAX files has been modified to facilitate the NCHS-MAX data linkage. For each NCHS survey and survey year, a file has been created that includes MAX Personal Summary (PS) file observations for every MAX data year (1999-2009) that link to observations in that NCHS survey file. For example, for NHIS 2000, a PS file exists that has PS file observations from the years 1999-2009 that link to the NHIS 2000. Similar IP, LT, RX, and OT files are available for each NCHS survey and survey year.

As described above and in section N and shown in Box 5, each NCHS survey observation has an identification variable used to link to the Feasibility files. These same variables can also be used to link observations in the NCHS survey to observations within each of the MAX files (PS, IP, LT, RX, and OT). Although the PS linkage file for each NCHS survey year contains data from several MAX years (1999-2009), the identification variable alone is used to link observations in the NCHS surveys to the PS file alone.

Linking observations to data from the MAX claims files (IP, LT, RX, and OT) in addition to the PS file involves a more complicated merging of data sets. Although infrequent, a single observation on an NCHS survey may link to more than one PS file in a single year (see section V(J) below for more details). Observations on the MAX claims files (IP, LT, RX, and OT) relate to a specific PS observation, so observations on the claims files from a particular year must be linked to the correct PS observation when more than one PS file exists within a year.

If multiple PS files within a year link to an NCHS survey observation, and both a PS and an IP file (for example) were linked to the NCHS survey only by the identification variable, it would be unclear which PS observation is related to a specific IP observation. Thus, a researcher trying to merge both data from the PS file and another file to the NCHS survey data must use the following steps to identify which PS observations are related to a specific claims file observation:

1) Merge NCHS survey observations for an individual year to the PS file from that year using the identification variable on the NCHS survey and the PS file, as described above in box 5.

2) Then merge the desired claims file (IP, LT, RX or OT) to the already-merged NCHS-PS file using all three of the following variables:

    a. "file_year4" variable that exists on the claims file (IP,LT, RX, or OT) and the already-merged PS file.

    b. The "msis_seqn" variable on the claims file and the already-merged PS file

    c. The identification variable (as described in box 5) on the claims file and the already merged PS-NCHS survey file.

C) Weighting of Linked Data files

The survey weights provided in NCHS population health survey data files adjust for oversampling of specific subgroups and differential non-response, and are post-stratified to annual population totals for specific population domains to provide nationally representative estimates. The properties of these weights for MAX-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 at NCHS and elsewhere is examining the use of survey weights for linked data in multiple ways.

Until specific recommendations are available, preliminary guidance is to analyze linked data files using adjusted survey weights. The survey weights available on NCHS population health survey data files can be adjusted for incomplete linkage and non-matches, hereafter referred to as non-response, using standard weighting domains to reproduce population counts within these

domains -- gender, age, and race/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) allows auxiliary information to be used to adjust the statistical weights for non-response. This approach is promising and this software is provisionally recommended for adjusting survey weights for the linked files. As 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. Additional information will be posted on the NCHS Data Linkage webpage as it becomes available, including sample SUDAAN code and a detailed literature review of weighting methods.

D) Identification of Medicaid enrollees

Estimates of the number of Medicaid enrollees differ by data source. In general, estimates from household population surveys such as the NHIS, NHANES, or the Current Population Survey (CPS) tend to yield lower estimates than Medicaid enrollment data collected by the states. There are likely several reasons for these differences. Some enrollees may respond incorrectly to population surveys, survey populations may differ from the population included in administrative records, and different reference periods for data from administrative records and the population surveys may account for some of these differences.

A multi-phase research project referred to as the Medicaid Undercount project has been undertaken to explain why discrepancies exist between survey estimates of enrollment in Medicaid and the number of enrollees reported in state and national administrative data. This project, also called the SNACC project (SNACC is an acronym for the federal agencies conducting the project the University of Minnesota's State Health Access Data Assistance Center (SHADAC), the National Center for Health Statistics (NCHS), the Agency for Healthcare Research and Quality (AHRQ), the U.S. Department of Health and Human Services Assistant Secretary for Planning and Evaluation (ASPE), the Centers for Medicare and Medicaid Services (CMS), and the U.S. Census Bureau). More information on this project can be accessed at http://www.census.gov/did/www/snacc/index.html.

The best method for identification of NCHS survey respondents who were Medicaid enrollees depends on the exact research question. In general, however, the variable MSNG_ELG_DATA provides information on their enrollment status (Box 6).

Box 6.

MSNG_ELG_DATA on the PS file indicates enrollment status:

- . = enrolled in Medicaid during the year
- 2 = enrolled in S-CHIP
- 1 = enrolled in neither Medicaid nor S-CHIP. See section H below for further discussion of observations where MSNG_ELG_DATA is coded as "1".

Of note, the variable 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 assigned on the PS file is the most valid of these and should be used for all of the data for that observation, regardless of from which file it originates.

E) State Differences in Medicaid

As described in Section III, 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 can vary greatly from state to state. Consideration of these state-level differences may be necessary for many analyses. State identifiers for each NCHS survey need to be specifically requested in those circumstances (Box 7).

Box 7.

State identifiers are not available in public use data files and must be specifically requested in RDC proposals.

F) Managed Care vs. Fee For Service (FFS)

As described above, 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 Person Summary file contains a variable that can be used to identify beneficiaries enrolled in any type of managed care plan and the number of months that they were enrolled in the plan (Box 8). Additional variables on the PS file identify each of up to 4 different types of managed care plans that the beneficiary was enrolled in during each month of the year (Boxes 8 & 9).

Box 8.

EL_PPH_PLN_MO_CNT_CMCP identifies those enrolled in any type of managed care plan and the number of months enrolled in the plan. EL_PHP_TYPE_1_1 through EL_PHP_TYPE_4_12 identify up to 4 different types of plans for each month.

Box 9.

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 (PACE) plan • Primary care case management (PCCM) plan • Other managed care plan

G) 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 likely to assist in promoting the objectives of the Medicaid statute. These projects are intended to demonstrate and evaluate a policy or approach that has not been widely used. Some states expand eligibility to individuals 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 child-bearing aged women not otherwise eligible for Medicaid. Medicaid enrollees that are eligible for Medicaid as a result of one of these programs are referred to as eligible through a waiver or waiver program. General information about Medicaid waivers can be found at https://www.cms.gov/MedicaidStWaivProgDemoPGI/03_Research&DemonstrationProjects-Section1115.asp.

Before 2005, it is only possible to know the Maintenance Assistance Status (MAS) and the Basis of Eligibility (BOE) of each enrollee by month (Box 10). Waivers for specific groups make up one of the MAS categories, although the specific type of waiver cannot be identified. Starting in 2005, MAX files include three elements for each month that give detailed information on the type of waivers under which enrollees are eligible for Medicaid (Box 10).

Box 10.

MAX_ELG_CD_MO_1 through MAX_ELG_CD_MO_12 provide MAS and BOE monthly enrollment information. MAX_WAIVER_TYPE_1_MO_1 through MAX_WAIVER_TYPE_3_MO_12 give detailed monthly waiver information starting in 2005.

H) Data Anomalies

Some aspects of the data cannot easily be explained. In this section we describe some of these and provide suggestions for how to address these issues during analyses. Potential new anomalies discovered in the course of analysis can be reported to the NCHS Special Projects Branch at datalinkage@cdc.gov.

State-specific data issues

Because the data for the MAX files are obtained from each state, there are differences in data quality between the states. Prior to conducting analyses, researchers should consult the CMS website on the MAX files at http://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAXGeneralInformation.html.

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 in advance of attempting analyses with the data.

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. It is not immediately clear why someone in the MAX files should be included if they were not enrolled in either Medicaid or S-CHIP for any months of the year. One explanation would be that it might reflect a state disenrolling people or reflect claims that arrive before enrollment. However, it is not possible to know with certainty if one of these explanations is valid, if administrative error is an issue, or whether other explanations exist. It is likely most prudent to eliminate these observations from analyses of the data.

These observations represent a small portion of the total data. However, they tend to cluster during the years 1999-2002, although remain <1.5% of those that link to the NHIS files in every year. Additionally, these observations tend to cluster by state, although only 7 states had rates greater than 2%.

For MAX PS files between 1999-2002, some observations for which the enrollee was coded as having neither Medicaid nor S-CHIP during the year have contradictory data when examined at the monthly enrollment level. Specifically, some of these observations are coded as having had

S-CHIP during some individual months, despite the overall yearly variable (MSNG_ELG_DATA) indicating they had neither S-CHIP nor Medicaid during the year. These data have not been changed, as it is not clear which data are correct. Treatment of these records during analysis is left to the researcher.

Days in the 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, but September only has 30 days. These values are due to administrative errors by the state and should be corrected if an analysis includes these values.

An individual can enroll in Medicaid at any time during a month and can also disenroll from Medicaid at any time during a month. As a result, enrollees may have a smaller number of days of Medicaid Eligibility than are in a month (Box 11).

Box 11.

Monthly days of eligibility are captured in the variables EL_DAYS_EL_CNT_1 through EL_DAY_EL_CNT_12; values range from 0 to 31.

I) Variables to Request in Applications to the Research Data Center (RDC)

To create analytic files for use in the RDC, a list of variables from the public-use NCHS survey data is given to the RDC by the researcher for merging with the requested restricted variables from NCHS surveys and for use with the MAX file variables. The restricted variables from NCHS surveys and the exact variables from the MAX files that the researcher will use need to be specifically requested as part of a researcher's application to the RDC. RDC staff will merge the public-use NCHS survey data with the restricted variables from NCHS surveys. In contrast, researchers will be responsible for merging the requested MAX file variables with the NCHS survey data, as described above.

While the complete list of variables used for specific analyses will differ, the variables listed below may be of use to researchers and should be considered for inclusion.

- Person Summary (PS) File variables as needed—few projects can be completed without at least some information from the person summary file. Therefore, these variables should be considered for inclusion.

- Medicaid/S-CHIP identifier on the PS file—MSNG_ELG_DATA

- Geography (State indicator on NCHS surveys where available)

- NHANES month and year of examination and interview (for those using NHANES). The NHANES surveys are analyzed combining at least 2 years (1999-2000, 2001-2002, and 2003-2004). As a result, the exact year (and month) of a respondent's interview and examination is not provided on the public use files for NHANES. However, in analysis of the surveys with the linked MAX files, many researchers will want to know the time elapsed between a given year (or even month) of the MAX data and the NHANES interview or examination. The variables that indicated the month and year of NHANES interview or examination can be obtained in the RDC but must be requested specifically.

- Linked mortality data for the NCHS surveys---all NCHS surveys that have been linked to the MAX data have also been linked to mortality files that provide date and cause of death for each survey respondent who has died. These may be of use to some researchers, but must be specifically requested as part of the researcher's proposal to the RDC.

We also recommend that the following variables, which are available on the public use NCHS survey files, should be provided to the RDC by the researcher for inclusion in the researcher's analytic files:

- NCHS Survey weights and design variables—Survey weights are required to create nationally representative estimates. Likewise, design variables are required to account for the survey design in analyses. 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 in Section V (c) above, these weights should be adjusted.

- Demographic information about survey respondents from the NCHS survey—For some variables (such as race/ethnicity), demographic information from the NHIS and NHANES is self- or family- respondent reported and thus may be more accurate than demographic data provided on the MAX file. Therefore, where possible, the NCHS data should be used for demographic variables.

- NHIS imputed income--- NCHS has created NHIS files that contain imputed income information for those with missing data on exact family income. These files should be used by researchers who are interested in including family income or poverty status in analyses that uses the NHIS linked data files.

J) Multiple PS Records

Many NCHS survey respondents will be linked to multiple MAX file PS records. Most often, this is because a respondent is linked to several years of MAX data. However, less frequently, a survey respondent may be linked to multiple PS records within the same year. This occurs in a

small percentage of records. For example, for NHIS years 1999-2005, 7.9% of NHIS survey respondents who linked to the MAX files had linkages to more than one PS record in at least one year. For NHANES, years 1999-2004, 9.5% of observations that linked to MAX files had linkages to more than one PS record in at least one year. There are multiple explanations for this situation.

- Medicaid enrollees moving between states in a given year
- Eligibility changes resulting in survey respondents disenrolling 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 (https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/downloads/anomalies1.pdf).

NCHS survey respondents with multiple PS records per year were generally due to MAX file records coming from multiple states. Among observations in NHIS years 1999-2005 that linked to multiple PS records in the same year, 86.6% came from different states. For NHANES years 1999-2004, 81.7% came from different states.

Another source of multiple PS records within the same year could be from false matches due to misreporting of personally identifiable information 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 presence of multiple PS records within a year leading to overlapping months of Medicaid enrollment data between the multiple 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 (Box 12). By determining whether a person was enrolled in each month across the multiple records within a year, one can obtain the number of total months of enrollment across records.

To help identify enrollees with multiple PS records within the same year, we have provided a set of variables that act as flags on the data files to identify these observations (Box 13). These variables can help identify enrollees who had any multiple PS records in a year, whether they occurred in the same or different state, and help determine the number of enrollees with multiple PS records per year.

Box 12.

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.

Box 13.

The variable FLG_YEAR_MULT_RECS identifies enrollees with multiple records in the same year:

- 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

The variable FLG_PRSN_MULT_RECS identifies enrollees who had multiple records in any year.

K) Dual Eligibles

There is interest among researchers and policy-makers in the group of people who receive benefits from both Medicaid and Medicare, often referred to as "dual eligibles". In the NCHS – MAX linked files, this group can be identified several ways. However, the suggested method to identify dual eligibles is to use the EL_MDCR_DUAL_ANN variable on the PS file (Box 14).

Box 14.

EL_MDCR_DUAL_ANN identifies enrollees who are dual eligibles. Values 50-59 signify that the enrollee was found in the Medicare database. Values 01-09 signify that the enrollee was not found in the Medicare database but was believed to be Medicare-eligible by the state. We recommend that the values 50-59 be used to identify dual eligibles 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, we recommend using the feasibility files described earlier. While we do not provide a flag for dual eligibles on the feasibility file, researchers can use the feasibility files for the Medicare linked data files (found at http://www.cdc.gov/NCHS/data_access/data_linkage/cms/cms_medicare_feasibility.htm) 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 MAX file.

For analyses of dual eligibles, some researchers will choose to use NCHS surveys linked to both Medicare data and the MAX files. Researchers should be aware that the methodology used for linking the NCHS surveys to MAX data differs from the methodology used to link Medicare data

to the NCHS surveys. The Medicare linkage required that the records match exactly on SSN, gender and day, month and year of birth while the MAX file linkage required matching exactly on SSN, gender, and month and year of birth, but not necessarily on day of birth. We provide a flag in the MAX files that indicates records that did match exactly on month, day and year of birth (Box 15). Using this flag, some researchers may choose to select records from the MAX file that used the same linkage methodology as the Medicare linkage and exclude the other records.

Box 15.

FLG_PRSN_DOB_FLAG identifies observations that linked to the MAX files exactly on SSN, gender, day, month, and year of birth. Similarly, this variable identifies observations that linked only on SSN, gender, month, and year of birth, but not on day.

L) CHIP

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, that is, all persons enrolled in these programs in a given year are included in the MAX files, the MAX files do not include all S-CHIP enrollees in any given year. Variables provide monthly information on CHIP eligibility, as well as whether an enrollee was enrolled in M-CHIP or S-CHIP (Box 16).

For S-CHIP enrollees in the files, some data elements contain no information (Box 16). Therefore, variables that are counts of months are suspect 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 percent of NCHS survey respondents linked to the MAX files. For example, among those linked to the NHIS in 2005, less than 1.5% of enrollees in the MAX files are in S-CHIP in any given month.

Box 16.

Variables EL_CHIP_FLAG_1 through EL_CHIP_FLAG_12 document CHIP eligibility monthly (January – December respectively), and whether an enrollee was in M-CHIP or S-CHIP. • 0 = Not eligible for Medicaid or CHIP during this month • 1 = Enrolled in Medicaid during this month • 2 = M-CHIP during this month • 3 = S-CHIP during this month For enrollees with a value of “3” (S-CHIP) for one of the EL_CHIP_FLAG_1 through EL_CHIP_FLAG_12 variables, no information is recorded for other monthly variables for that month.

M) Diagnosis and Procedure Coding

Diagnoses are uniformly provided in the MAX files for inpatient, outpatient, and long-term care using *International Classification of Diseases, 9th revision, Clinical Modification* (ICD-9-CM) codes. These codes are provided in the IP, LT, and OT MAX files. However, the coding systems used in the IP and OT files for procedures vary (no procedure codes are provided in the LT files). For each procedure code variable, a separate variable is provided that identifies the coding system used for that code (Box 17). For the IP file, the majority, but not all, procedure codes are ICD-9-CM. However, for the OT file, the majority (but again, not all) codes are either in Current Procedural Terminology codes (CPT-4) or Healthcare Common Procedure Coding System codes (HCPCS).

Box 17.

> In the IP files, variables PRCDR_CD_SYS_1 through PRCDR_CD_SYS_6 describe the type of codes used for procedure codes in variables PRCDR_CD_1 through PRCDR_CD_6 respectively.
>
> Similarly, in the OT file, the variable PRCDR_CD_SYS describes the type of codes used in the procedure code variable PRCDR_CD.

N) Creation of Public ID Numbers for File Linkage

The data provided on the feasibility files for the 1994-2005 NHIS, 1999-2004 NHANES, 2004 NNHS, NHEFS, NHANES III and LSOA II can be merged with the NCHS public use survey data files using the unique survey specific Public Identification number, as described in Section V-A. These numbers are also used for linkage of the NCHS surveys to the MAX data files by researchers.

National Health Interview Survey (NHIS)

The items and public-use locations needed to construct the NHIS public ID numbers vary by NHIS survey year.

NHIS 1994

Item	Public-use Location	Length	Description
Year (2 digit)	3-4	2	Year of interview
Quarter	5	1	Calendar quarter of interview
PSU	6-8	3	Random recode of PSU #
Week	9-10	2	Week of interview within quarter
Segment	11-12	2	Segment number
Household	13-14	2	Household number within quarter
Person number	15-16	2	Person number within household

Note: Concatenate all variables to get the unique person identifier. All variables are zero filled.

NHIS 1995, 1996

Item	Public-use Location	Length	Description
Year (2 digit)	3-4	2	Year of interview
Household ID	5-14	10	Household ID number
Person number	15-16	2	Person number within Household

Note: Concatenate all variables to get the unique person identifier. All variables are zero filled.

NHIS 1997-2003

Item	Public-use Location	Length	Description
Year (4 digit)	3-6	4	Year of interview
Household Serial #	7-12	6	Household serial number
Family number	13-14	2	Family number
Person number	15-16	2	Person number within Household

Note: Concatenate all variables to get the unique person identifier. All variables are zero filled.

NHIS 2004

Item	Public-use Location	Length	Description
Year (4 digit)	3-6	4	Year of interview
Household serial #	7-12	6	Household serial number
Family sequence #	13-14	2	Family number
Person sequence #	15-16	2	Person number

Note: Concatenate all variables to get the unique person identifier. All variables are zero filled.

NHIS 2005

Item	Public-use Location	Length	Description
Year (4 digit)	3-6	4	Year of interview
Household serial #	7-12	6	Household serial number
Family sequence #	16-17	2	Family number
Person sequence #	18-19	2	Person number

Note: Concatenate all variables to get the unique person identifier. All variables are zero filled.

National Health and Nutrition Examination Survey (NHANES)

Item	Length	Description
SEQN	5	Participant identification number

All of the NHANES public-use data files are linked with the common survey participant identification number (SEQN).

2004 National Nursing Home Survey (NNHS)

Item	Length	Description
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).

NHANES Epidemiologic Followup Study (NHEFS)

Item	Length	Description
SEQN	5	Participant identification number

All of the NHEFS public-use data files are linked with the common survey participant identification number (SEQN).

Third National Health and Nutrition Examination Survey (NHANES III)

Item	Length	Description
SEQN	5	Participant identification number

All of the NHEFS public-use data files are linked with the common survey participant identification number (SEQN).

The Second Longitudinal Study of Aging (LSOA II)

Item	Public-use Location	Length	Description
Year (2 digit)	3-4	2	Year of interview
Quarter	5	1	Calendar quarter of interview
PSU	6-8	3	Random recode of PSU #
Week	9-10	2	Week of interview within quarter
Segment	11-12	2	Segment number
Household	13-14	2	Household number within quarter
Person number	15-16	2	Person number within household

Note: Concatenate all variables to get the unique person identifier. All variables are zero filled.

Section VI: References

1. Perez V, Schmitz B, Wenzlow AT, Shepperson K, Baugh D, Radke S. The Medicaid Analytic Extract 2004 Chartbook. 2008; http://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAX_Chartbooks.html.
2. Kaiser Commission on Medicaid and the Uninsured. *MEDICAID A Primer: Key Information on Our Nation's Health Coverage Program for Low-Income People.* Washington, DC: The Henry J. Kaiser Family Foundation;2010.
3. Kaiser Commission on Medicaid Facts. *Dual Eligibles: Medicaid's Role for Low-Income Medicare Beneficiaries.* Washington, DC: The Henry J. Kaiser Family Foundation;2011.
4. Kaiser Commission on Medicaid and the Uninsured. Medicaid and Managed Care: Key Data, Trends, and Issues. 2010; http://www.kff.org/medicaid/upload/8046.pdf.
5. Medicaid and CHIP Payment and Access Commission. Report to the Congress on Medicaid and CHIP, March 2011. http://docs.google.com/viewer?a=v&pid=sites&srcid=bWFjcGFjLmdvdnxtYWNwYWN8Z3g6NTZmYjU1ZDcwMTQzMDc0MA.
6. Trenholm C, Mabli J, Wilson A. *SCHIP Children: How Long Do They Stay and Where Do They Go?* Princeton, NJ: Mathematica Policy Research, Inc.;2009.

Table 1. NHIS
1999-2007[1] Medicaid linked files sample sizes and linkage rates, by NHIS year and age at interview

	Sample Size				Linkage Rates			
NHIS Survey Year & Age	Total NHIS	Eligible for Linkage	Linked to MAX PS File		Linkage Rates for Total Sample		Linkage Rates for Eligible Sample	
			In survey year	In any year	In survey year	In any year	In survey year	In any year
1994	**116,179**	**87,079**	**---**	**18,090**	**---**	**15.6**	**---**	**20.8**
<19	33,911	24,326	---	8,609	---	25.4	---	35.4
19-64	67,697	51,061	---	7,082	---	10.5	---	13.9
≥65	14,571	11,692	---	2,399	---	16.5	---	20.5
1995	**102,467**	**73,809**	**---**	**17,076**	**---**	**16.7**	**---**	**23.1**
<19	31,027	21,319	---	8,427	---	27.2	---	39.5
19-64	59,485	43,091	---	6,643	---	11.2	---	15.4
≥65	11,955	9,399	---	2,006	---	16.8	---	21.3
1996	**63,402**	**42,407**	**---**	**10,182**	**---**	**16.1**	**---**	**24.0**
<19	19,011	12,088	---	4,842	---	25.5	---	40.1
19-64	37,257	24,948	---	4,114	---	11.0	---	16.5
≥65	7,134	5,371	---	1,226	---	17.2	---	22.8
1997	**103,477**	**62,689**	**---**	**15,739**	**---**	**15.2**	**---**	**25.1**
<19	31,133	17,794	---	7,550	---	24.3	---	42.4
19-64	60,453	36,881	---	6,333	---	10.5	---	17.2
≥65	11,891	8,014	---	1,856	---	15.6	---	23.2
1998	**98,785**	**53,535**	**---**	**13,727**	**---**	**13.9**	**---**	**25.6**
<19	29,469	15,362	---	6,703	---	22.7	---	43.6
19-64	57,896	31,666	---	5,467	---	9.4	---	17.3
≥65	11,420	6,507	---	1,557	---	13.6	---	23.9
1999	**97,059**	**49,467**	**7,246**	**13,122**	**7.5**	**13.5**	**14.6**	**26.5**
<19	28,704	14,197	3,968	6,389	13.8	22.3	27.9	45.0
19-64	57,355	29,466	2,588	5,372	4.5	9.4	8.8	18.2
≥65	11,000	5,804	690	1,361	6.3	12.4	11.9	23.4
2000	**100,618**	**49,127**	**7,763**	**13,691**	**7.7**	**13.6**	**15.8**	**27.9**
<19	29,933	14,248	4,253	6,745	14.2	22.5	29.8	47.3
19-64	59,479	29,262	2,744	5,596	4.6	9.4	9.4	19.1
≥65	11,206	5,617	766	1,350	6.8	12.0	13.6	24.0
2001	**100,760**	**47,452**	**8,001**	**13,659**	**7.9**	**13.6**	**16.9**	**28.8**
<19	29,988	13,793	4,468	6,836	14.9	22.8	32.4	49.6
19-64	59,755	28,234	2,820	5,573	4.7	9.3	10.0	19.7
≥65	11,017	5,425	713	1,250	6.5	11.3	13.1	23.0
2002	**93,386**	**53,074**	**9,380**	**14,923**	**10.0**	**16.0**	**17.7**	**28.1**
<19	27,477	17,750	5,553	8,138	20.2	29.6	31.3	45.8
19-64	55,324	29,575	3,066	5,624	5.5	10.2	10.4	19.0
≥65	10,585	5,749	761	1,161	7.2	11.0	13.2	20.2
2003	**92,148**	**49,095**	**9,236**	**14,051**	**10.0**	**15.2**	**18.8**	**28.6**
<19	26,844	16,663	5,559	7,712	20.7	28.7	33.4	46.3
19-64	55,031	27,218	2,975	5,312	5.4	9.7	10.9	19.5
≥65	10,273	5,214	702	1,027	6.8	10.0	13.5	19.7
2004	**94,460**	**45,805**	**8,862**	**13,400**	**9.4**	**14.2**	**19.3**	**29.3**
<19	27,526	14,283	5,133	7,017	18.6	25.5	35.9	49.1
19-64	56,176	26,254	2,964	5,322	5.3	9.5	11.3	20.3
≥65	10,758	5,268	765	1,061	7.1	9.9	14.5	20.1
2005	**98,649**	**44,835**	**8,850**	**13,024**	**9.0**	**13.2**	**19.7**	**29.0**
<19	28,168	13,810	5,088	6,772	18.1	24.0	36.8	49.0
19-64	59,251	26,162	3,021	5,316	5.1	9.0	11.5	20.3
≥65	11,230	4,863	741	936	6.6	8.3	15.2	19.2

[1]2008 and 2009 Medicaid linked files are currently available. Revised tables including sample sizes and linkage rates for 1999-2009 are forthcoming.

Table 2. NHANES
1999-2007[1] Medicaid linked files sample sizes and linkage rates, by NHANES year and age at interview

	Sample Size			Linkage Rates	
NHANES Survey Year & Age	**Total NHANES**	**Eligible for Linkage**	**Linked to MAX PS File in any year**	**Linkage Rates for Total Sample in any year**	**Linkage Rates for Eligible Sample in any year**
1999-2000	**9,965**	**7,852**	**3,506**	**35.2**	**44.6**
<19	4,804	3,726	2,270	47.2	60.9
19-64	3,769	3,011	839	22.3	27.9
≥65	1,392	1,115	397	28.5	35.6
2001-2002	**11,039**	**9,274**	**3,874**	**35.1**	**41.8**
<19	5,347	4,489	2,596	48.6	57.8
19-64	4,229	3,544	947	22.4	26.7
≥65	1,463	1,241	331	22.6	26.7
2003-2004	**10,122**	**8,624**	**3,932**	**38.8**	**45.6**
<19	4,792	4,032	2,577	53.8	63.9
19-64	3,836	3,296	1,010	26.3	30.6
≥65	1,494	1,296	345	23.1	26.6

[1]2008 and 2009 Medicaid linked files are currently available. Revised tables including sample sizes and linkage rates for 1999-2009 are forthcoming.