

Chronic Condition Data Warehouse

Getting Started with Medicaid Analytic
eXtract (MAX) Data Files-
A Technical Guidance Paper

Version 1.1
October 2014

Table of Contents

Chapter 1: Overview	1
Chapter 2: File Types.....	2
Chapter 3: Methods.....	3
A. Understanding Eligibility Data	3
1. Eligibility for Benefits	3
a. Categories of Eligibility	3
b. Children’s Health Insurance Program (CHIP)	4
c. Scope of Benefits	5
Table 1. Mandatory and Optional Benefits.....	5
Table 2. Data Values for the Different Levels of Benefits	6
d. Managed Care	7
Table 3. Examples and Data Values of Types of PHPs*	7
e. Cost Sharing.....	8
Code example 1: Describe a Sample of Medicaid Enrollees using Eligibility Variables.....	8
2. Demographic Information.....	11
3. Consideration of Vital Status (death date)	11
B. Understanding Utilization Data	12
1. Contents of Data Files.....	12
2. Categorizing Services	12
Code example 2: Creating Unique Person Identifier.	13
Table 4. MAX TOS Values.....	14
3. Medicaid Programs	15
4. How Services are Billed.....	15
5. Waivers	16
Code example 3: Identifying Wavier and Non-wavier Long Term Care Services.	16
Table 5. Service Categories and Associated Values	17
C. Describing and Summarizing Utilization.....	17
1. Totals Variables	17
2. How to Link PS to Claims	18
Table 6. Recommended Unduplication Methodology	18

3. Identifying Care for a Particular Diagnosis 19
 Code example 4: Using Various Merges for Linking PS and Utilization Files. 19
 Code example 5: Identifying Claims for Specific Services. 20
4. Identifying Services for Particular Enrollees 21
 Code example 6: Using a Population Subset – Identify Specific Types of Care Received. 21
 Code example 7: Classifying Institutional Long Term Care Services. 22
5. Summary Use and Payment 25
 Code example 8: Calculating Payments Per Capita and Per Member Month. 25
Chapter 4: Receiving CCW Data 27
Chapter 5: Further Assistance with MAX Data 27
Appendix A: List of Acronyms and Abbreviations 28

Getting Started with CMS Medicaid Analytic eXtract (MAX) Data Files
A Technical Guidance Paper

Chapter 1: Overview

More than 60 million people in the U.S. are covered by Medicaid or the Children's Health Insurance Program (CHIP).¹ Medicaid is a state-administered benefit with guidance/requirements and shared funding from the federal government and individual state governments (Title XIX of Social Security Act from 1965). Each state must provide the minimum federally mandated services and coverage for federally mandated eligibility groups; however, benefits vary from state to state. Many groups of people are covered by Medicaid, depending on the state's requirements (e.g., age, whether pregnant, disabled, blind, or aged, income level and resources, U.S. citizenship or lawful immigration status). There are also special rules for those who live in nursing homes and for disabled children living at home. For more information on the Medicaid program, see the Centers for Medicare & Medicaid Services (CMS) website (<http://www.medicaid.gov/>).

Each state compiles information regarding enrollment, service utilization, and payment into their Medical Statistical Information System (MSIS). The resulting files are sent to CMS. CMS and its contractors compile the various state MSIS data files into a uniform data structure for each calendar year and make these annual state segment files available to researchers as the Medicaid Analytic eXtract (MAX) data files. Additional details regarding the construction of the MAX files is available on the CMS website (https://www.cms.gov/MedicaidDataSourcesGenInfo/07_MAXGeneralInformation.asp).

Researchers will find useful information about file content by state and year by reviewing the MAX Data Validation Tables and Data Anomalies Reports under the MAX General Information on Data section of the CMS website. Information regarding some state-specific information is also available, such as managed care penetration. MAX file layouts and other helpful information regarding use of MAX files can be referenced on the Research Data Assistance Center (ResDAC) web site at (<http://www.resdac.org/cms-data/file-family/Medicaid-Analytic-Extracts-MAX>).

The MAX data are contained in the CMS Chronic Condition Data Warehouse (CCW) and are provided to academic researchers and certain government agencies, which have been approved under a Data Use Agreement (DUA) to obtain MAX data for research purposes. The CCW MAX data contain identifiable information, and are subject to the Privacy Act and other Federal government rules and regulations (see ResDAC web site for information on requesting MAX data). The data dictionary for the CCW MAX files is available on the CCW website (<https://www.ccwdata.org/web/guest/data-dictionaries>).

¹ <http://www.medicaid.gov/>

The CCW MAX files are linked by a unique CCW identifier, allowing researchers to analyze information across the continuum of care, within and across years, as well as across states, without using actual identifying information (e.g., MSIS_ID, Social Security number, or Medicare Health Insurance Claim number for dual eligibles). While the MAX files were designed from the ease-of-use perspective, learning the nuances of any new data source can be challenging. Creating the appropriate analytic dataset can be the key to simplifying analysis of the data.

This document provides technical guidance to approaching analysis of MAX data and provides a framework to creating customized, analytic data files. In addition to narrative technical information, examples of SAS[®] code are shared.

Chapter 2: File Types

There are six types of data files available from the CCW MAX database, consisting of the following:

- **Person Summary (PS)** - Person-level information regarding Medicaid-eligibles who have enrolled in a state Medicaid program or a Medicaid-expansion CHIP in the year, whether or not they used any services
- **Inpatient Hospital (IP)** - Contains complete stay records for enrollees who used inpatient services. Records include diagnoses, procedures, discharge status, length of stay, and payment amounts.
- **Long Term Care (LT)** - Includes service data from four types of long-term care facilities that serve Medicaid enrollees.
- **Prescription Drug (RX)** - Contains records of prescription and over-the-counter drugs paid by Medicaid. Claims records include a National Drug Code (NDC) to indicate what drug product was used, service date and payment information.
- **Other Services (OT)** - Includes outpatient, physician and professional services, hospice, home health, lab/X-ray, durable medical equipment (DME), premium payments etc., and all Medicaid records not reported in any other MAX service files
- **Medicaid Enrollee Supplemental File (MESF)** - Historically, the MAX PS File was developed prior to the CCW; therefore it lacked some of the key features of CCW data files that are designed to aid in the study of quality of care and health of the population. This was remedied when CCW began disseminating an additional file designed to supplement the information in the MAX PS. The Medicaid Enrollee Supplemental File consists of two optional components called “segments”. These became available in 2014. The two file segments are the:
 - Conditions segment, and
 - NDI segment.

Data dictionaries for the CCW MAX files are available on the CCW website (<https://www.ccwdata.org/web/guest/data-dictionaries>). Additional information regarding the source data is available on the CMS website (see <http://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAXGeneralInformation.html>). Additional information regarding MAX files, and educational opportunities involving MAX data are available through ResDAC (see <http://www.resdac.org/Index.asp>).

Chapter 3: Methods

A. Understanding Eligibility Data

1. Eligibility for Benefits

Determinations regarding enrollment in Medicaid are made each month. Some enrollees may be deemed eligible for a portion of a year or an entire year, whereas others may be eligible for benefits for a single month. A person may be eligible for a different level of benefits at different points in time. To accommodate this variability, the MAX PS file has monthly variables to indicate whether the person was Medicaid covered (MAX_ELG_CD_MO_XX). Some investigators may not need to know exactly which month Medicaid enrollment occurred – in which case there is a summary variable in the PS file which simply counts months enrolled in Medicaid during the year (EL_ELGLTY_MO_CNT).

Benefits vary among those with Medicaid coverage. Many people may be approved to receive full Medicaid benefits, whereas others have only restricted benefits which cover specific types of care.

a. Categories of Eligibility

Federal law identifies over 25 different eligibility categories. CMS has classified these into five broad groups²:

- Children
- Pregnant women
- Adults in families with dependent children
- Individuals with disabilities
- Adults 65 years of age or older

Potential enrollees must meet financial tests, which vary by category (see the www.cms.gov website for more details). For example, some of the mandatory Medicaid eligibility groups include: children 0-6 years of age whose family income is up to 133% Federal Poverty level

² CMS.gov website has many resources to describe Medicaid eligibility. See, for example, <http://www.cms.gov/Outreach-and-Education/Training/CMSNationalTrainingProgram/Downloads/2013-Medicaid-and-CHIP-Workbook.pdf>

(FPL), children 6-10 years old from families up to 100% FPL, and pregnant women up to 133% FPL. States may offer benefits at varying levels. The Kaiser Family Foundation website offers some information regarding state Medicaid benefits³ (see: <http://www.kff.org/medicaid/index.cfm>).

States may optionally cover certain groups, including individuals with incomes higher than the minimum federal thresholds. For example, some states provide coverage for impoverished adults without dependent children or those who are medically needy. There are also some conditions or diseases which may result in Medicaid eligibility – such as patients served through the National Breast and Cervical Cancer Early Detection Program (NCCEDP), or tuberculosis-related services (TB).

Investigators may use the MAX eligibility codes (e.g., monthly MAX_ELG_CD_MO_XX or EL_MAX_ELGLTY_CD_LTST) to classify Medicaid enrollees into categories. Two common schemes are the Medicaid Assistance Status (MAS) and the Basis of Eligibility (BOE). Both of these variables are derived from the two-digit MAX eligibility code variables, referred to earlier in this paragraph. The first digit of the eligibility code is used to define MAS, which subdivides the population into: 1) cash; 2) medically needy [MN]; 3) poverty; 4) other; and, 5) W1115 (1115 waiver). The second digit of the eligibility code is used to define BOE, which subdivides the population into: 1) aged; 2) blind or disabled; 3) child; 4) adult; and, 5) unknown. A SAS[®] code example, which demonstrates how to construct the MAS and BOE is presented later in this chapter (see Code example 1).

b. Children's Health Insurance Program (CHIP)

The Children's Health Insurance Program (CHIP; also referred to as the State Children's Health Insurance Program [SCHIP]) was authorized through the Balanced Budget Act of 1997, and funded as Title XXI of the Social Security Act. The objective of CHIP is to enable states to extend insurance coverage to low-income children who are not eligible for Medicaid and who do not have private insurance. The CHIP benefit is structured similarly to Medicaid in that it is a state program with federal matching funds available for children based on income limits relative to the FPL. States may elect to extend coverage to children beyond the federal matching income level. Each state is allotted a certain level of matching funds each year – and federal funds are capped at this amount. Additional details regarding state CHIP programs can be found on the Medicare.gov website⁴.

States can operate CHIP one of three ways: 1) as an expansion of Medicaid [referred to as

³ Kaiser Family Foundation website. See, for example, <http://medicaidbenefits.kff.org/index.jsp?CFID=105676655&CFTOKEN=34377891&jsessionid=60304b118c30fb6a2fa8521d184d47114933>

⁴ Medicare.gov website has many resources regarding CHIP. See, for example, <http://www.medicare.gov/Medicaid-CHIP-Program-Information/By-Topics/Childrens-Health-Insurance-Program-CHIP/Childrens-Health-Insurance-Program-CHIP.html>

M-CHIP]; 2) as a separate child health insurance program [referred to as SCHIP]; or, 3) a combination of these two approaches. Children enrolled in SCHIP are not considered to be enrolled in Medicaid (note: exception is Medicaid-expansion CHIP, which is equivalent to Medicaid coverage). There are some data in the MAX PS file regarding children enrolled in CHIP when used as a Medicaid expansion (M-CHIP). However, S-CHIP enrollment information is not a required part of the MAX files, and no S-CHIP utilization information is contained in the MAX database.

Investigators are able to determine which children are covered by CHIP by using the monthly CHIP eligibility flags (EL_CHIP_FLAG_X). Medical care utilization (claims or encounter) and payment information is not available for those enrolled in CHIP. Note that within a year, a person may have a combination of Medicaid and CHIP coverage; therefore it is possible that people with CHIP enrollment may have some claims in the MAX files.

c. Scope of Benefits

States are required to cover certain benefits, whereas others are optional. Similarly, not all enrollees are eligible to receive comprehensive benefits and may only be eligible for restricted benefits.

- 1) Mandatory and optional benefits. All children in Medicaid are entitled to comprehensive services through the Early, Periodic Screening, Diagnosis and Treatment (EPSDT) program. Other optional benefits provide coverage only for specific types of care. For example, the National Breast and Cervical Cancer Early Detection Program (NCCEDP) and tuberculosis (TB) programs offer only disease-specific services. If you are examining variation between states with regard to volume or cost of services, it is important to consider the nature of the services (i.e., mandatory versus optional), and to understand the nuances of how the state(s) structures the benefit. Table 1 below lists the different types of mandatory and optional benefits typically offered by states. Refer to the CMS website for more details (<http://www.medicaid.gov/Medicaid-CHIP-Program-Information/By-Topics/Benefits/Medicaid-Benefits.html>).

Table 1. Mandatory and Optional Benefits

Mandatory Benefits	Optional Benefits
<ul style="list-style-type: none"> • Inpatient hospital services • Outpatient hospital services • Early and Periodic Screening, Diagnostic, and Treatment (EPSDT) services • Nursing facility services • Home health services • Physician services • Rural health clinic services 	<ul style="list-style-type: none"> • Prescription drugs • Clinic services • Physical therapy • Occupational therapy • Speech, hearing and language disorder services • Respiratory care services • Other diagnostic, screening, preventive and rehabilitative services • Podiatry services

<ul style="list-style-type: none"> Federally qualified health center services Laboratory and X-ray services Family planning services Nurse Midwife services Certified Pediatric and Family Nurse Practitioner services Freestanding Birth Center services (when licensed or otherwise recognized by the state) Transportation to medical care Tobacco cessation counseling for pregnant women 	<ul style="list-style-type: none"> Optometry services Dental services Dentures Prosthetics Eyeglasses Chiropractic services Other practitioner services Private duty nursing services Personal Care Hospice Case management Services for Individuals Age 65 or Older in an Institution for Mental Disease (IMD) Services in an intermediate care facility for the mentally retarded State Plan Home and Community Based Services- 1915(i) Self-Directed Personal Assistance Services- 1915(j) Community First Choice Option- 1915(k) Other services approved by the Secretary*
---	---

2) Full benefits or restricted benefits. Some Medicaid enrollees may be eligible only for partial/restricted benefits rather than comprehensive services (e.g., people may receive Medicare premium assistance or prescription drug coverage only). The monthly EL_RSTRCT_BNFT_FLG_XX variables describe the scope of benefits for the enrollee. One would expect to see very different utilization and cost for persons eligible for different levels of benefits. Code example 1, below, illustrates an option for using information regarding full versus restricted benefits. Table 2 below documents the data values for the EL_RSTRCT_BNFT_FLG_XX variable. Full data dictionaries are available through the CMS or ResDAC web sites, as mentioned previously.

Table 2. Data Values for the Different Levels of Benefits

0 Not Medicaid eligible
1 Full Medicaid
2 Restricted Medicaid - alien
3 Restricted Dual (QMB/SLMB only)
4 Restricted - pregnancy
5 Restricted - other
6 Medicaid – FP only
7 Medicaid – alternate coverage
9 Medicaid – unknown restrictions
X Medicaid– Prescription (RX) drug only
Y Dual – RX only (QMB/SLBM only)
Z Dual – RX only (EDB info = no state dual \$)

d. Managed Care

Many states use a managed care delivery system for their Medicaid benefits (see <http://www.medicaid.gov/>). These are capitated plans or plans through which health care providers (e.g., physicians, nurse practitioners) are paid a set amount for each enrolled person assigned to them for a prescribed period of time whether or not the enrollee seeks care. These managed care plans cover specified services. Some managed care plans are designed to offer comprehensive medical benefits, such as those offered by Health Maintenance Organizations (HMOs); whereas other plans offer only specific benefits or “carve outs”, such as mental health or dental care. States have flexibility with regard to administering the Medicaid benefit. They also may apply for waivers from CMS to operate managed care plans or other demonstration projects within the Medicaid program. These waivers include section 1915(b) or Section 1115. Although not technically managed care, some Medicaid programs have a Primary Care Case Management (PCCM) program, which involves a small monthly stipend to primary care providers who agree to coordinate care for FFS recipients.

All of these managed care arrangements are referred to as pre-paid health plans (PHP). There are monthly indicators in the PS file to capture up to four PHPs, if any, for each person enrolled (EL_PHP_ID_X_XX). Information regarding the type of PHP for each month, for up to four plans, is contained in (EL_PHP_TYPE_X_XX) and illustrated in Table 3. Investigators may also find it helpful to use the seven summary variables which count the total number of prepaid health plans during the year. The seven types of plans include: comprehensive managed care plans (EL_PPH_PLN_MO_CNT_CMCP), all-inclusive care for the elderly (EL_PPH_PLN_MO_CNT_AICE), behavioral managed care, dental managed care, long-term care managed care, prenatal delivery managed care, and primary care case management.

Table 3. Examples and Data Values of Types of PHPs*

00 = Individual Was Not Eligible For Medicaid This Month
01 = Enrolled In A Medical Or Comprehensive Managed Care Plan (e.g., HMO).
02 = Enrolled In A Dental Managed Care Plan
03 = Enrolled In A Behavioral Managed Care Plan
04 = Enrolled In A Prenatal/Delivery Managed Care Plan
05 = Enrolled In A Long-Term Care Managed Care Plan
06 = Enrolled In A Program For All-Inclusive Care For The Elderly (PACE)
07 = Enrolled In A Primary Care Case Management Managed Care Plan (PCCM)
08 = Enrolled In An Other Managed Care Plan
88 = Not Applicable, This Record Is Not An Encounter Record
99 = Eligible's Managed Care Plan Status Is Unknown.

* shaded rows are often used to indicate enrollment in comprehensive managed care services where there is not much if any FFS claim activity. Note that some states also use the value ‘08’ (caution: be mindful of which states operate managed care in particular ways).

e. Cost Sharing

Some enrollees have other health insurance coverage (e.g., dual Medicare/Medicaid or third party private insurance coverage). Medicaid is the payer of last resort – meaning that all other forms of insurance are primary. (Note: exceptions include other federal health systems – such as the Veterans Affairs [VA] and Indian Health Service [IHS]). These other third party payments reduce the cost liability for Medicaid.

For people dually enrolled in Medicare and Medicaid, states are subject to cost-sharing for certain services (Note: for Medicare enrollees who are not dually enrolled, the consumer would be responsible for cost-sharing). These costs may include copayments or coinsurance (MDCR_COINSUR_PYMT_AMT) and deductibles (MDCR_DED_PYMT_AMT).

Code example 1: Describe a Sample of Medicaid Enrollees using Eligibility Variables

Define a sample of Medicaid enrollees with a specified duration and scope of Medicaid coverage. This code includes variables which can be used to determine whether the person was entitled to the full scope of Medicaid benefits or only partial benefits. It also includes variables to determine whether the person was enrolled in FFS or managed care coverage, and the eligibility category.

The first portion of the code uses the monthly eligibility variables (MAX_ELG_CD_MO_XX) to verify Medicaid enrollment each month. It also uses the monthly benefit variables to determine whether the person was enrolled for full or restricted benefits (EL_RSTRCT_BNFT_FLG_XX). The arrays are designed to the number of covered months.

The second portion of the code looks at fee-for-service (FFS) coverage for Medicaid (rather than managed care), for comprehensive medical benefits, prenatal care/delivery and long-term care, using the monthly EL_PHP_TYPE_X_XX variables for all four possible types of PHPs (a total of 48 variables considered). Importantly, this example shows how some values indicating managed care coverage may vary by state. Investigators are encouraged to understand the payment structures and waivers used by states included in the analysis. Note that investigators interested in topics such as behavioral and mental health or dental health may wish to modify the code example to capture the population with these types of benefits offered through PHPs.

The third portion of the code creates the MAS and BOE from the latest uniform eligibility code variable (EL_MAX_ELGLTY_CD_LTST).

The input data source is the MAX PS File from 2007 (note: the variables and variable names are generally the same over time, however modifications occur periodically).

The following SAS[®] code assumes that you are developing a new data file (called coverage) by using MAX coverage specifications. Then you may use these variables to subset your population (e.g., FFS only or those entitled to full, rather than restricted benefits) and/or to determine member months of coverage in Medicaid, FFS, or managed care plans.

```

data coverage;
  set a.maxdata_ps_2007;

  *determine number of months with Medicaid and scope of benefits*/;
  array max_elg(12) MAX_ELG_CD_MO_1-MAX_ELG_CD_MO_12;
  array Mdc_d_mo (12) MDCD_MO_1-MDCD_MO_12;
  array rstr(12) EL_RSTRCT_BNFT_FLG_1-EL_RSTRCT_BNFT_FLG_12;
  array cov_mo(12) COV_MO_1-COV_MO_12;

  do i=1 to 12;
    if max_elg(i) not in ('00','99','') then
      Mdc_d_mo(i)=1;
    else if max_elg(i) in ('00','99','') then
      Mdc_d_mo(i)=0;
    MAX_MOS=sum(of Mdc_d_mo:);

    if Mdc_d_mo(i)=1 and rstr(i) in ('1') then
      cov_mo(i)=1;
    else if Mdc_d_mo(i)=1 and rstr(i) in
('2','4','5','6','7','8','9','3','X','Y','Z') then
      cov_mo(i)=0;
    FULL_MOS=sum(of cov_mo:);
  end;

  ***Define FFS Status;
  array php1(*) EL_PHP_TYPE_1_1-EL_PHP_TYPE_1_12;
  array php2(*) EL_PHP_TYPE_2_1-EL_PHP_TYPE_2_12;
  array php3(*) EL_PHP_TYPE_3_1-EL_PHP_TYPE_3_12;
  array php4(*) EL_PHP_TYPE_4_1-EL_PHP_TYPE_4_12;
  array mx_f(*) MDCDFFS1-MDCDFFS12;

  do i= 1 to 12;
    /*consideration of state-specific managed care coverage*/
    if STATE_CD in ('AL','CA','FL','WI') then
      do;
        if (php1(i) not in ('01','05','06','08','') and
          php2(i) not in ('01','05','06','08','') and
          php3(i) not in ('01','05','06','08','') and
          php4(i) not in ('01','05','06','08','')) then
          mx_f(i)=1;
        else if (php1(i) in ('01','05','06','08') or
          php2(i) in ('01','05','06','08') or
          php3(i) in ('01','05','06','08') or
          php4(i) in ('01','05','06','08')) then
          mx_f(i)=0;
        end;
      else
        do;
          if (php1(i) not in ('01','05','06','') and
            php2(i) not in ('01','05','06','') and
            php3(i) not in ('01','05','06','') and
            php4(i) not in ('01','05','06','')) then
            mx_f(i)=1;
          else if (php1(i) in ('01','05','06') or
            php2(i) in ('01','05','06') or
            php3(i) in ('01','05','06'))
  
```

```

php4(i) in ('01','05','06')) then
  mxf(i)=0;
end;

/*calculation of summary FFS and managed care variables to use in later
analyses*/

if mxf(i)=1 and mdc_d_mo(i)~1 then mxf(i)=0;

if 0<sum(of mxf(*))=Max_mos then
  MDCDFFS=1;
else if 0<=sum(of mxf(*))< Max_mos then
  MDCDFFS=0 ;
end;

if MAX_MOS > 0 then ffs_mos=sum(of MDCDFFS1-MDCDFFS12);
if MAX_MOS > 0 then MC_MOS=Max_mos - ffs_mos;
if MAX_MOS > 0 and MC_MOS>0 then MC=1; else if MAX_MOS>0 then MC=0;

/** describing eligibility **/

length MAS BOE $10;
if substr(EL_MAX_ELGBLTY_CD_LTST,1,1)='1' then MAS='1 CASH';
else if substr(EL_MAX_ELGBLTY_CD_LTST,1,1)='2' then MAS='2 MN';
else if substr(EL_MAX_ELGBLTY_CD_LTST,1,1)='3' then MAS='3 POVERTY';
else if substr(EL_MAX_ELGBLTY_CD_LTST,1,1)='4' then MAS='4 OTHER';
else if substr(EL_MAX_ELGBLTY_CD_LTST,1,1)='5' then MAS='5 W1115';
else MAS='9 UNKNOWN';
if substr(EL_MAX_ELGBLTY_CD_LTST,2,1)='1' then BOE='1 AGED';
else if substr(EL_MAX_ELGBLTY_CD_LTST,2,1)='2' then BOE='2 BLIND';
else if substr(EL_MAX_ELGBLTY_CD_LTST,2,1) in ('4','6','8') then BOE='4
CHILD';
else if substr(EL_MAX_ELGBLTY_CD_LTST,2,1) in ('5','7','A') then BOE='5
ADULT';
else BOE='9 UNKNOWN';

label

MAX_MOS='Number of months enrolled in Medicaid'
FULL_MOS='Number of months with full, rather than restricted Medicaid
coverage'
FFS_MOS='Number of months of FFS coverage, not HMO'
MC_MOS='Number of months of Managed Care coverage'
MC='at least one month of managed care coverage'
MDCDFFS='Person had FFS Medicaid all covered months, or not'
MAS='Maintenance Assistance Status'
BOE='Basis of Eligibility';

run;

proc freq data=coverage;
tables MAX_MOS*MAS*BOE/list missing;
run;

```

Investigators may modify this SAS[®] code example as desired to meet the inclusion criteria for their study or to describe the population. Some of the variables created in this code example are used in subsequent code examples.

2. Demographic Information

Several demographic variables are available in MAX files. The MAX PS has an age group variable (EL_AGE_GRP_CD), also date of birth (EL_DOB). It has a variable to indicate the gender of the enrollee (EL_SEX_CD), and this variable is also combined with a dichotomous race categorization (white or non-white, by sex) (EL_SEX_RACE_CD).

There are five separate dichotomous variables for race, with each variable having a specific meaning. The five possibilities for race are: 1) RACE_CODE_1 (white – yes/no); 2) RACE_CODE_2 (black – yes/no); 3) RACE_CODE_3 (American Indian/Alaskan – yes/no); 4) RACE_CODE_4 (Asian – yes/no); and, 5) RACE_CODE_5 (Hawaiian/Pacific Islands – yes/no). There is also an ethnicity code (ETHNICITY_CODE), which indicates whether the person is Hispanic/Latino or not. Finally, there is a variable which combines the various race and ethnicity variables into a single variable, called EL_RACE_ETHNCY_CD (e.g., white - not Hispanic, black – not Hispanic).

If the beneficiary is also enrolled in Medicare, additional variables which may be helpful in describing race or ethnicity are available. For example, the Medicare language code designates the preferred language for receiving Medicare materials (MDCR_LANG_CD). There is also a Medicare race/ethnicity code from the Medicare Enrollment Database (MDCR_RACE_ETHNCY_CD).

The MAX PS file includes some variables helpful in understanding the geographic location of the enrollee. Variables include county (EL_RSDNC_CNTY_CD_LTST), zip code (EL_RSDNC_ZIP_CD_LTST), and state (STATE_CD).

3. Consideration of Vital Status (death date)

Within the PS file there is a variable which may be used to indicate a person who was enrolled in Medicaid for some portion of the year has died. The EL_DOD – is the date of death from MSIS. The EL_DOD has been known to be unreliably reported in MAX; please refer to the CMS website for MAX FAQs on this topic (<https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAXGeneralInformation.html>). If the person is also dually enrolled in Medicare, there is a Medicare date of death (MDCR_DOD).

CW delivers a supplemental file related to date of death that you may request. It is the Medicaid Enrollee Supplemental File – National Death Index (MESF – NDI) file, which contains the cause of death, according to official death certificate data purchased from the National Death Index (NDI) through the Centers for Disease Control and Prevention (CDC). This segment contains only the subset of enrollees who had a date of death in the MAX PS

file, or if the person was dually enrolled in Medicare, a date of death in Medicare records. Record layouts for the MESF NDI file are on the CCW website. Additional details regarding these files, including DUA requirements are available through ResDAC.

B. Understanding Utilization Data

The MAX claims files may be very large. It will be important to determine whether you need to include all claims for your sample, regardless of the reason for the medical care, or whether you may be able to satisfy your analytic objectives by querying the data files and extracting only the claims related to receipt of specific types of care, or care for certain conditions.

Some enrollees will not have any utilization (claims) data, whereas others will have a very large volume of services. There are several steps which must be taken to accurately organize utilization: 1) Verify you are examining utilization for the appropriate claim type (IP, LT, OT, RX, or a combination of files); 2) Determine the care setting(s) you need to examine within the files (e.g., only Hospice services within the OT file); and, 3) Determine whether you need to explore all care for your sample or look only for certain types of care (i.e., diagnoses, procedures, or any care within a particular setting/level of care).

1. Contents of Data Files

The record layout for the four claims files is similar. All files contain person identifiers (e.g., BENE_ID, MSIS_ID, STATE_CD, EL_DOB, etc.). In addition, there are variables to indicate the dates of service (SRVC_BGN_DT and SRVC_END_DT) and the specific services received, which may vary by claim type (e.g., DIAG_CD_X, PRCDR_CD, NDC). Each claim also has information regarding charges (e.g., UB_92_REV_CD_CHGS_X, CHRГ_AMT) and payments (e.g., MDCD_PYMT_AMT, MDCD_COINSUR_PYMT_AMT).

2. Categorizing Services

There are several options with regard to categorizing the Medicaid services used. Investigators are reminded that all claims are processed by states using the MSIS_ID. Many users will find it helpful to be able to identify services for a unique person within a year, within a state. There are three variables which must be considered when identifying a unique person:

- BENE_ID - goes across MSIS_IDs over time and across states
- MSIS_ID - state assigned value at a point in time (same ID may appear in other states or within the state at a different point in time; variable is encrypted), and
- STATE_CD – state code.

Using the MSIS_ID and STATE_CD alone, even for a single year, might result in duplicate records for a person; it is possible for someone eligible for Medicaid to be ineligible at a later point in time. It is also possible that the person becomes eligible again within the same year. This discontinuous enrollment may result in the person having more than one MSIS_ID. If BENE_ID were completely populated, this variable, in combination with STATE_CD, would

precisely identify a unique person; however, the BENE_ID is fairly new for MAX data, and is not populated for approximately 7% of MSIS_IDs. To be able to identify the eligibility information for unique people within a state, it is recommended that investigators aggregate information using: BENE_ID (when populated), MSIS_ID, and STATE_CD.

Code example 2: Creating Unique Person Identifier.

This code is an example of how to create of a new “ID” variable, which is useful in uniquely identifying a person – and all service use associated with the person.

The following SAS® code demonstrates creation of a new variable (called ID), which considers the BENE_ID and the MSIS_ID to uniquely identify a person in the MAX files. Here we show this variable for the PS file. Later, you may use these variables to calculate per capita costs or service use.

```
data person;
  set a.maxdata_ps_2007;
  length id $32;
  if BENE_ID~=' ' then ID=BENE_ID;
  else if BENE_ID=' ' then ID=MSIS_ID;
run;
```

- a. The most basic form of service categorization is the MAX file layout. The claims are partitioned into four services based on the type of service provided (IP, OT, RX, LT).
- b. Medicaid payments may occur for premiums or for medical services. To identify the payments associated with premiums, the PS file has a variable called TOT_MDCD_PREM_CLM_CNT – the count of premium payment claims per person. There are also summary payment variables for premium amounts (e.g., for HMO or PCCM; PREM_MDCD_PYMT_AMT_HMO and PREM_MDCD_PYMT_AMT_PCCM). For people enrolled in managed care or prepaid health plans, FFS claims for the types of services covered in the PHP for the months in managed care will not appear in the MAX files. For people enrolled in PCCM, this variable indicates the payment for the PCCM service – yet there will still be claims for medical care in the MAX files. Also, for people enrolled in dental or behavioral capitated plans, the medical claims will still appear in the MAX files.
- c. A key variable on the claim record is used to indicate the type of claim (TYPE_CLM_CD). This variable indicates whether the claim is for: 1) a FFS medical claim; 2) a capitated payment; 3) an encounter [or “dummy” record]; 4) a service tracking claim for a lump sum payment; 5) a supplemental payment – above capitation; or, 9) unknown. Payments should not appear for encounter records.
- d. Every claim is classified according to the type of service. There are two variables which classify the type of service: the MSIS_TOS and the MAX_TOS. The concept behind these

two variables is the same; however there is variability between states with regard to algorithms used to classify services, making it inadvisable to compare MSIS_TOS across states. The MAX_TOS variable re-groups services using a uniform coding scheme across states (e.g., by considering procedure codes), and adds four TOS codes which are not part of MSIS. A difference in these two variables, for example, is that some states may classify services provided during a physician visit (e.g., labs or drug) as being part of the visit (i.e., a physician service), whereas other states would classify the service as a lab or drug service.

Some types of services may be found in more than one file. Table 4 below illustrates which MAX_TOS values were found in which MAX files in 2007.

Table 4. MAX TOS Values

MAX Type of Service (TOS)					
value	Label	IP	LT	RX	OT
01	Inpatient hospital	X			
02	Mental hosp for aged		X		
04	IP psych – youth		X		
05	ICF – MR		X		
07	NF		X		
08	Physician				X
09	Dental				X
10	Other Provider				X
11	OP hospital				X
12	Clinic				X
13	Home health				X
15	Lab and x-ray				X
16	Drugs			X	X
19	Other services				X
24	Sterilizations	X			X
25	Abortions	X			X
26	Transport				X
30	Personal care				X
31	Case Management				X
33	Rehab				X
34	PT, OT, Speech				X
35	Hospice				X
36	Nurse Midwife				X
37	Nurse Practitioner				X
38	Private nurse				X
51	DME			X	X
52	Residential care				X
53	Psychiatric				X

MAX Type of Service (TOS)					
value	Label	IP	LT	RX	OT
54	Adult day care				X
99	Unknown	X	X	X	X

The PS file has summaries of record/claim counts by these TOS categories. The TOS codes appear as the suffix for some variables. For example, FEE_FOR_SRVC_IND_16 is for drugs and have National Drug Codes (NDCs) for prescription and over the counter (OTC) drug (TOS =16); ENCTR_REC_CNT_51 is for NDCs that represent durable medical equipment and supplies (DME) (TOS=51).

Medicaid may cover services often not traditionally paid by some private health insurance plans – e.g., dental, care management, over the counter (OTC) drugs, and long-term care or nursing facility (LTC/NF) stays. Investigators wishing to compare utilization and payments between Medicaid and other types of health plans may need to limit the Medicaid types of services to allow for more of an apples-to-apples comparison (e.g., with private health insurance plans or Medicare).

e. Service place. The OT file contains a PLC_OF_SRVC_CD to describe the care setting (e.g., school, home, ambulance, mass immunization center).

Additional examples of various methods for classifying the location/setting of care can be found in a CCW guidance document (see CCW Technical Guidance: Getting Started with CMS Medicare Administrative Research Files at <https://www.ccwdata.org/web/guest/technical-guidance-documentation>). Examples include identification of services using revenue center codes (e.g., IP variable UB_92_REV_CD_GP_X or OT variable UB_92_REV_CD) or procedure codes (e.g., OT variable PRCDR_CD). The UB revenue center codes are not widely reported in MAX.

3. Medicaid Programs

Some Medicaid services may be from various funding programs, all of which have specific eligibility requirements and payment structures. The MSIS type of program code (MSIS_TOP) is another way to look at services (i.e., ESPDT, Family Practice [FP], Rural Health Clinic [RHC], Federally-Qualified Health Clinic [FQHC], IHS, Home and Community Based [HCB]-Aged, Home and Community Based Waiver [HCBW]). This variable is useful for determining which enrollees were receiving services under waiver programs (i.e., values of 6 or 7 in this field = home and community-based waivers).

4. How Services are Billed

Providers use the UB-04 claim form (also known as CMS 1450) to submit claims for inpatient services. The previous version of the claim form was called a UB_92. In the MAX IP file, there are 23 variables which provide details regarding the particular revenue centers (i.e.,

hospital cost centers) on what used to be known as the UB 92 form (e.g., UB_92_REV_CD_GP_1). The values and meaning of these revenue centers can be found in the Medicare Institutional claims file data dictionary on the ResDAC site (see:<http://www.resdac.org/cms-data/variables/revenue-center-code>). For these 23 revenue centers, there is also information regarding the associated units (e.g., volume of services or UB_92_REV_CD_UNITS_1) and charges (e.g., UB_92_REV_CD_CHGS_1).

Some services may be “bundled” and not billed as separate, discrete services. For example, in acute hospital setting care is often paid based using a Diagnosis Related Group (DRG). Similarly, bundling of services is fairly common in IP and LT claims, whereby drugs and certain services or therapies are included in per diem payments. When payment is bundled, specific procedure codes may not appear on the claim.

Procedure codes also appear in the data, and may affect payment. See, for example, the OT variable PRCDR_CD.

5. Waivers

For persons receiving long term care, a variety of services may be provided. These services may appear in different MAX files. Long term care may be community or facility-based (i.e., investigators should use information from both the LT and OT files). Both types of services should be considered when examining utilization and costs (for people receiving these services). Code example 3, below, illustrates a method for identifying waiver and non-waiver services. Note that there is additional information regarding long-term care services in the OT file, such as the Community-Based Long-Term Care services flag (CLTC_FLAG).

Code example 3: Identifying Wavier and Non-wavier Long Term Care Services.

This code is an example of how to identify waiver services in the OT file. This code uses two key service classification variables, the MSIS_TOP and MAX_TOS, to cluster similar services into a service type category (variable called SRVC). Two values from the MSIS_TOP indicate wavier services (6,7). The values of SRVC derived from the OT file all begin with the “OT”.

Table 5, below displays the service categories, and the associated values created when using Code example 3.

Table 5. Service Categories and Associated Values

Value	Description
OT_HH	Other Therapy file, home health
OT_PCS	Other Therapy file, personal care services
OT_DME	Other Therapy file, durable medical equipment
OT_OTH	Other Therapy file, other type of service
OT_DRUG	Other Therapy file, drug

The following SAS[®] code demonstrates creation of service categories (variable called SRVC) which use the MSIS_TOP and MAX_TOS to aggregate and describe services.

```

data ot;
  set a.maxdata_ot_2007;
  length srvc $8;

  if MAX_TOS=13 then SRVC='OT_HH';
  else if MAX_TOS=30 then SRVC='OT_PCS';
  else if MAX_TOS=51 then SRVC='OT_DME';
  else if MAX_TOS=16 then SRVC='OT_DRUG';
  else SRVC='OT_OTH';

  if MSIS_TOP in (6,7) then WAIVER=1; else WAIVER=0;

run;

proc freq data=ot;
  tables SRVC * WAIVER;
run;

```

C. Describing and Summarizing Utilization

You have many options for exploring and describing utilization. Numerous variables in the claims files make it possible to classify the types of services received (i.e., care setting and procedures), and the reason (i.e., diagnosis) for the care. There are also many options for the unit of analysis (e.g., number of visits, total hospitalized days, per capita utilization). Below some common methods for counting and summarizing utilization are presented.

1. Totals Variables

Within the PS file there are several summary variables which count the per person number of claims and encounters, as well as variables which aggregate the per person payments (e.g., TOT_MDCD_PYMT_AMT). These totals are also provided for IP and LTC settings (e.g., TOT_IP_DSCHRG_CNT). There are totals and for several of the program types such as,

FP, FQHC, HCBCA, HCBCS, IHS, RHC (e.g., TOT_PYMT_AMT_FQHC), and totals for these programs by setting for each of the four types of claims files (e.g., OT_PYMT_AMT_FQHC).

2. How to Link PS to Claims

If you want to join a person-level file, such as the PS file or a subset of this file, there are several options for performing the merge, depending on the objectives.

a. Linkage keys - It is recommended that investigators use three variables as keys when linking files, to ensure integrity of each record. This is important because an individual person may have more than one MSIS_ID – and may have utilization records using more than one MSIS_ID. Then, for accurately tabulating per person statistics, it is helpful to create a unique ID variable (described above).

- BENE_ID
- STATE_CD
- MSIS_ID

Table 6 below illustrates some data integrity problems which may arise if the record for individual persons are handled improperly. A sample data file from 2007 is used for illustrative purpose.

- The sample PS file has 616,730 records (i.e., MSIS_ID and state combinations)
- After unduplicating, by consideration of the BENE_ID, MSIS_ID and STATE_CD, we have 616,582 unique people in the sample.
- If we only use the BENE_ID or, if BENE_ID is not populated, use the MSIS_ID. If the state is not also used as a key variable, some people are erroneously considered the same person due to a duplicate MSIS_ID assignment between states.

Table 6. Recommended Unduplication Methodology

PS File (which is at the MSIS_ID by STATE_CD level)	(BENE_ID or MSIS_ID) by STATE_CD	(BENE_ID or MSIS_ID) without STATE_CD
616,730	616,582	614,918

Shaded cell = recommended method.

All files must be sorted (by BENE_ID, STATE_CD, and MSIS_ID) before merging (refer to Code example 4, below).

b. Type of merge - Once the linkage keys have been established, some additional decisions regarding how to join the files must be addressed. A couple of merge options exist: 1) merge the utilization file to PS (or other subpopulation defined by the investigator) only if there is a record in both files (Note: within a utilization file, it is possible that claims exist for persons not appearing in the PS file); 2) merge the utilization file to PS so that there is a record for all

PS records. The first option allows investigators to examine utilization and calculate statistics for service users, whereas the second option enables per capita utilization statistics. Use caution since you may need only a subset of variables. It will be important to keep the file as trim as possible as it can get very wide.

Code example 4: Using Various Merges for Linking PS and Utilization Files.

The SAS® code example demonstrates a method for joining the MAX PS file with any of the claims files, after determining whether to keep all records in all files – or only specific ones.

The following SAS® code uses the three key linkage variables to merge the relevant data files. The input data files must first be sorted before the merge. In this example, we merge only the IP records that have a corresponding PS record for the person.

```
proc sort data= a.maxdata_ps_2007; out=ps;
by BENE_ID STATE_CD MSIS_ID;
run;

proc sort data= a.maxdata_ip_2007 out=ip;
by BENE_ID STATE_CD MSIS_ID;
run;

data b.inpt;
merge ps (in=x) ip (in=y);
if x ;
by BENE_ID STATE_CD MSIS_ID;
run;
```

As mentioned above, the merge can be performed in different ways. By substituting the “if x;” statement in the data step for “if x and y;” there will only be records for people with both an IP claim and a PS record.

3. Identifying Care for a Particular Diagnosis

As an analyst, you may be interested only in certain types of care. For example, perhaps you only want to examine hospitalizations related to hip fracture or heart failure. Most state Medicaid agencies use DRGs to classify hospital discharges. Note that the DRGs used may be the same as the Medicare DRGs, or states may use state-specific codes. The DRG indicator variable in the IP file (DRG_REL_GROUP_IND) should be examined to determine whether the CMS coding scheme was used, or if the state should be contacted to learn the meaning of the DRGs. Each DRG has a payment weight assigned to it, based on the average resources used to treat patients in that DRG. Researchers exploring reasons for inpatient service use may find it helpful to use the DRG_REL_GROUP variable in the IP data file.

Similarly, a common way to examine reasons for care is to look at the diagnoses on the claim (e.g., DIAG_CD_1). This is the medical or surgical reason that hospitalization required. The diagnosis code is a more granular system than DRG (i.e., many diagnosis codes may fall into a single DRG).

Additional diagnoses are present if there are complicating factors (e.g., may not be the major reason for the stay/visit, but have a bearing on the condition of the patient). These are called secondary diagnosis codes (i.e., DIAG_CD_2 - DIAG_CD_9). It is common to see a condition such as diabetes listed as a secondary diagnosis – a co-occurring condition which must be managed, rather than the major cause of the hospitalization.

Code example 5: Identifying Claims for Specific Services.

The SAS® code example demonstrates a process for identifying all claims for a particular condition – asthma. Here we use the diagnosis codes in any of the diagnosis fields within the IP file (i.e., DIAG_CD_1 - DIAG_CD_9). Investigators may decide to examine only a particular subset of IP claims.

The following SAS® code uses the IP file to identify claims for diagnoses of interest, using any of the 9 positions for the diagnosis code. It outputs two files: a file with all IP claims (called asthma) that contains a claim-level asthma indicator (ASTH) and a finder file of people with IP asthma treatment (called ipfinder).

```

data asthma ipfinder (keep=bene_id msis_id state_cd);
  set b.inpt;

  /*looking for asthma diagnosis codes in any position on the
  claim*/

  array diag (9) DIAG_CD_1 - DIAG_CD_9;

  ASTHM=0;
  do i= 1 to 9;
    if DIAG (i) in
    ('49300', '49301', '49302', '49310', '49311', '49312', '49320', '49321',
    '49322',
    '49381', '49382', '49390', '49391', '49392') then ASTHM=1;
  end;

  output asthma;
  if ASTHM=1 then output ipfinder;

  run;
  
```

Additional examples of various methods for examining utilization – including care for a particular diagnosis, care for individuals with particular conditions, or people who received

particular procedures – can be found in a CCW guidance document called “CCW Technical Guidance: Getting Started with CMS Medicare Administrative Research Files” (<http://www.ccwdata.org/analytic-guidance/index.htm>).

4. Identifying Services for Particular Enrollees

As discussed above, to examine patterns of medical utilization, it is sometimes necessary to combine information from different claim types so that a complete assessment of the care received by individuals is obtained.

a. Prescription drugs - In general, prescription drugs (and possibly a small number of over-the-counter drugs) are found in the RX file. Some specific types of drugs, particularly infused drugs which are administered in a medical setting (e.g., chemotherapy) are found in the OT file. A MAX_TOS value of 16 is used to identify drug services. As mentioned previously, prescription drugs are often bundled into a per diem facility payment – and these drugs will not appear in the RX file.

The RX file contains some variables which are helpful for determining the type of drug received by the patient. The major classification system is the National Drug Code (variable is called NDC). Although there are variables in the data file which appear to be drug classification schemes (e.g., American Hospital Formulary code [AMER_HSPTL_FRMLRY_SYS_CD], First DataBank® [GCN_SEQNO, HIC3, THRTPC_CLASS_CD_SPECIFIC] and Medispan® [MEDISPAN_DRG_CTGRY]), these data are proprietary and only CMS data users covered under the Medispan license have access to these fields. These variables are blank for all other researchers.

It is important to note that Medicare began offering prescription drug coverage, called Medicare Part D, in 2006. For Medicaid enrollees who were also dually enrolled in Medicare, prescription drugs would primarily appear in the Medicare Part D data files (i.e., Medicare is the primary payer for Medicare covered drugs) – although some drugs for dual eligibles appear in the RX files (e.g., to represent the cost-sharing portion of the drug claim; also drugs not covered by Medicare – such as OTC drugs and certain prescription drugs which are not part of the Medicare Part D standard benefit).

Code example 6: Using a Population Subset – Identify Specific Types of Care Received.

For this example, we start with the subset of people who had an inpatient claim for asthma (in Code example 5, above). Using the unique person identifiers for this population (BENE_ID, STATE_CD and MSIS_ID) we are seeking to identify all RX services received by these patients.

```
proc sort data=ipfinder nodupkey;  
  by BENE_ID STATE_CD MSIS_ID;  
run;
```



```
proc sort data=a.maxdata_rx_2007 out=rx;
      by BENE_ID STATE_CD MSIS_ID;
run;

data asthma_rx;
merge ipfinder (in=x) rx(in=y);
by BENE_ID STATE_CD MSIS_ID;
if x;
run;

/*examine the NDCs used by asthma patients*/

proc freq data=asthma_rx;
table NDC;
run;
```

Using our test data file, the ipfinder file yields 5,387 IP claims with an asthma diagnosis code. After unduplicating the person identifiers, we have a finder file with 3,963 patients who had IP care with an asthma diagnosis. We obtain 122,627 RX records with a populated NDC code for these patients.

b. Long-term care services may take place in either institutional or non-institutional settings. The LT file contains claims for facility-based services (e.g., nursing facilities, mental hospitals, inpatient psychiatric and Intermediate Care Facilities for the Mentally Retarded [ICF/MR]). As mentioned previously, supportive services for long-term care may also be home-based (e.g., home health or personal care services). These non-institutional care claims appear in the OT file (see Code example 3, above).

Claims in the LT file may span various lengths of time – and often consist of weekly or monthly claims for persons who are continuously enrolled for a long-term stay. There may be multiple records for the person for the same service dates (i.e., same SRVC_BGN_DT and SRVC_END_DT) at the same facility (PRVDR_ID_NUM). Each of these records should be considered when examining costs of long-term care, although it is not possible to tell which records are for per-diem payments and which are for various ancillary services (e.g., therapy). If examining the duration of the stay, or the number of stays per person, it will be necessary to unduplicate records so that only one record per patient per facility per time frame is considered.

Code example 7: Classifying Institutional Long Term Care Services.

This code is an example of how to classify services in the LT file using the MAX_TOS variable.

The next portion of the code examines the dates on the claims, to determine a total number of days in LT facilities, for each service type. A means statement is used to summarize the average per person number of days by setting. The final portion of code examines total payments from the LT file, and uses SQL code to illustrate an option for efficiently summarizing the payment data.

The following SAS® code assumes that you are developing a new data file (called long) from the MAX LT file. Then you may use these variables to calculate the number of days per person in each setting of care, and create a summarized data file (called long2). This data file is further summarized into per person averages in a file called ind_summ.

Finally, a data file is created to summarize per person Medicaid payments for the LT file (called LExample). Tabulations are performed using this table.

```

data Long;
  set a.maxdata_lt_2007;
  length ID $32;
  length SRVC $10;

  if BENE_ID~= ' ' then ID=BENE_ID;
  else if BENE_ID=' ' then ID=MSIS_ID;

  if MAX_TOS=7 then SRVC='LT_I_NFS';
  else if MAX_TOS in (2,4,5) then SRVC='LT_I_OTH';
  else SRVC='LT_OTH';

run;

/*calculate facility days - need to unduplicate the redundant dates on
the records*/

proc sort data=long;
  by BENE_ID MSIS_ID STATE_CD SRVC SRVC_BGN_DT SRVC_END_DT;
run;

data long;
  set long;
  format END_DT START_DT yymmddn8.;
  by BENE_ID MSIS_ID STATE_CD SRVC SRVC_BGN_DT SRVC_END_DT;
  if first.STATE_CD then do;
    END_DT=SRVC_END_DT;
    NEWSTAY=1;
    START_DT=SRVC_BGN_DT;
  end;
  else do;
    if intck('day',END_DT,SRVC_BGN_DT) > 1 then do;
      NEWSTAY=1;
      START_DT=SRVC_BGN_DT;
    end;
    else NEWSTAY=0;
    if SRVC_END_DT > END_DT then END_DT=SRVC_END_DT;
  end;
  retain END_DT START_DT;
run;

proc sort data=long;
  by BENE_ID MSIS_ID STATE_CD SRVC START_DT END_DT;
run;

```

```

data long2;
  set long;
  by BENE_ID MSIS_ID STATE_CD SRVC START_DT END_DT;
  if last.START_DT then DAYS=intck('day',START_DT,END_DT)+1;
run;

/*individual summary file - one record per person per LT service type,
which sums the number of days for all claims*/

proc means data=long2 noprint;
  by BENE_ID MSIS_ID STATE_CD SRVC;
  output out=ind_summ(drop= _freq_ _type_) sum(DAYS)=;
run;

/*Average number of days and number of individuals by service type*/

proc sort data=ind_summ;
  by SRVC;
run;

proc means data=ind_summ N mean median;
  by SRVC;
  var DAYS;
run;

```

Note: this outputs summary tables for the 3 LT service categories.

```

/* use SQL code - calculate total payments per person
All LT claims are used*/

proc sql;
create table b.ltexample as
select
  ID,
  count(MSIS_ID) as CLAIM_COUNT,
  count (distinct ID) as PERSON_COUNT,
  sum(MDCD_PYMT_AMT) as MEDICAID_PAY_SUM
from long
group by ID, SRVC;
quit;

proc tabulate data=b.ltexample;
  class PERSON_COUNT;
  var MEDICAID_PAY_SUM ;
  table (MEDICAID_PAY_SUM * sum={label="total $"} PERSON_COUNT *
n={label="# people"} ) ;
  table mean*(MEDICAID_PAY_SUM * (PERSON_COUNT));
run;

```

Investigators interested in more information regarding persons residing in a long-term care facility (i.e., those with records in the MAX LT files) may wish to consider obtaining CCW assessment data files. For persons in the LT file who have a BENE_ID, the corresponding

Minimum Data Set (MDS) assessment data could be obtained; refer to the ResDAC website (<http://www.resdac.org/cms-data/file-family/assessments>). The MDS provides detailed information regarding the health and functional status of the patient. Researchers may wish to unduplicate person records in the MAX files before they join to the MDS using the BENE_ID (i.e., to avoid a potential many-to-one MAX to MDS linkage). Refer to Code example 2 for more details.

5. Summary Use and Payment

There are many variables within the PS file which summarize Medicaid use and payments. Investigators may find that many of the aggregate utilization and payment statistics can be easily performed with the PS file without requiring tabulations from the claims files. It is common to explain insurance statistics in terms of member months of enrollment. It is also useful to understand the total number of enrollees and the per capita statistics. Code example 8 below demonstrates the use of variables in the PS file for examination of member months of enrollment, per member per month payments, and per capita payments.

Code example 8: Calculating Payments Per Capita and Per Member Month.

For this example, we use the PS file, and create several summary member month and payment variables, looking at total Medicaid payments (TOT_MDCD_PYMT_AMT), payment for FFS use (TOT_MDCD_FFS_PYMT_AMT), and payments for managed care enrollees (the difference between these two variables). We also use summary enrollment counts which were created in an earlier example: the count of months enrolled in Medicaid (MAX_MOS), the count of months in FFS Medicaid (FFS_MOS) and the count of months enrolled in managed care plans (MC_MOS).

For the last portion of this example, we show that these summary statistics can be created at the state level, in case between state comparisons are desirable.

Note: investigators may wish to count only a subset of the people (e.g., those enrolled in full rather than restricted benefits).

The following SAS[®] code uses the coverage file (from Code example 1) to summarize months of Medicaid coverage overall, and for FFS and managed care. Total payments are summarized similarly, to allow for per member per month payment determinations. Per capita payments are also determined, and the example demonstrates how to calculate this by state.

```
data b.payment;
set coverage;

if FFS_MOS > 0 then MDCD_FFS_PM_PMT =TOT_MDCD_FFS_PYMT_AMT/FFS_MOS;
if MAX_MOS > 0 then MDCD_PM_PMT=TOT_MDCD_PYMT_AMT/MAX_MOS;
TOT_MDCD_MC_PMT_AMT=TOT_MDCD_PYMT_AMT-TOT_MDCD_FFS_PYMT_AMT;
if MC_MOS > 0 then MDCD_MC_PM_PMT=TOT_MDCD_MC_PMT_AMT/MC_MOS;
```

```

label
  MDCD_FFS_PM_PMT='FFS per member per month payment'
  MDCD_PM_PMT='Total Medicaid per member per month payment'
  MDCD_MC_PM_PMT='Managed Care per member per month payment'
  TOT_MDCD_MC_PMT_AMT = 'Total managed care payment amount';

run;

/*calculation of per member per month average payments*/

proc means data=b.payment;
  var MDCD_PM_PMT MDCD_FFS_PM_PMT MDCD_MC_PM_PMT;
  title 'PMPM payments';
run;

/*this proc means outputs a data file with only 2 variables - the
total Medicaid payments and the total enrolled in Medicaid*/

proc means data=b.payment noprint;
  output out=MDCD_summary(drop=_freq_ _type_)
    sum(TOT_MDCD_PYMT_AMT) = n(MAX_MOS)=PERSON_CNT;
  title'Total Per capita payments';
run;

/*the next procs sort data into managed care vs. FFS by state- then
computes total payments and people enrolled and outputs this as a dat
file (called state_summary)*/

proc sort data=b.payment out=State_pay;
by MC STATE_CD;
run;

proc means data=state_pay noprint;
by mc state_cd;
output out=state_summary (drop=_freq_ _type_)
  sum(TOT_MDCD_FFS_PYMT_AMT TOT_MDCD_MC_PMT_AMT) = N(FFS_MOS
)=PERSON_CNT;
  title'FFS or MC Per capita payments by state';
run;

```

Researchers may wish to calculate per capita payments for FFS or managed care, however it is important to remember that enrollees may have a combination of FFS and managed care during the year. If using the per capita calculations (as opposed to member months calculations), enrollees may appear in both FFS and managed care categories.

Chapter 4: Receiving CCW Data

Data files are typically shipped to you on either a Universal Serial Bus (USB) external hard drive or a Digital Video Disc (DVD)/Compact Disc (CD). These data files are packaged as encrypted self-decrypting archive (SDA) files (see the CCW User Guide at <https://www.ccwdata.org/web/guest/user-documentation> for additional information on encryption). The decryption password is sent to you electronically via email. When you receive the data package (via hard drive or DVD/CD), copy them from the shipping media to your local workspace. Note some data shipped on a hard drive can be decrypted on that hard drive, depending on the size of the data files. Using the password provided to you via email, follow the Decryption Instructions enclosed in the data package. Each SDA contains the data file(s), SAS® code and a file transfer summary (.fts) file which can be used to verify the data was read in correctly.

Chapter 5: Further Assistance with MAX Data

ResDAC offers free assistance to those using MAX data for research. The ResDAC website provides links to descriptions of the CMS data available, request procedures, supporting documentation, workshops on how to use MAX data and other helpful resources. Visit the ResDAC web site at (<http://www.resdac.org>) for additional information. Researchers may also e-mail questions to resdac@umn.edu or call the Assistance Desk at 1-888-973-7322.

In the event that a ResDAC technical advisor is not able to answer the question, the technical advisor will direct you to the appropriate person.

www.ccwdata.org

Email: CMSdata@gdit.com

Phone: 1-866-766-1915

Appendix A: List of Acronyms and Abbreviations

Acronym	Definition
BETOS	Berenson-Eggers Type of Service
BOE	Basis of Eligibility
CD	Compact Disc
CHIP	Children's Health Insurance Program
CMS	Centers for Medicare & Medicaid Services
CCW	Chronic Condition Data Warehouse
DVD	Digital Video Disc
DME	Durable Medical Equipment
DRG	Diagnosis Related Group
DUA	Data Use Agreement
ED	Emergency Department
EPSDT	Early, Periodic Screening, Diagnosis and Treatment
FQHC	Federally-Qualified Health Clinic
FP	Family Practice
FFS	Fee-for-Service
FPL	Federal Poverty Level
HCPCS	Healthcare Common Procedure Coding System
HMO	Health Maintenance Organization
HCB	Home and Community Based
HCBW	Home and Community Based Waiver
ICD-9	International Statistical Classification of Diseases and Related Health Problems
IHS	Indian Health Service

ICF/MR	Intermediate Care Facilities for the Mentally Retarded
IP	Inpatient Hospital
LT or LTC	Long Term Care
MAS	Medicaid Assistance Status
MAX	Medicaid eXtract
M-CHIP	Medicaid Expansion of CHIP
MDS	Minimum Data Set
MMIS	Medicaid Management Information Systems
MN	Minnesota
NBCCEDP	National Breast and Cervical Cancer Early Detection Program
NDC	National Drug Code
NF	Nursing Facility
OT	Other Services
OTC	Over the Counter
PCCM	Primary Care Case Management
PHP	Pre-Paid Health Plans
PS	Person Summary
ResDAC	Research Data Assistance Center
RHC	Rural Health Clinic
RX	Prescription Drug
SCHIP	State Children's Health Insurance Program
SDA	Self-Decrypting Archive
SNF	Skilled Nursing Facility
SSA	Social Security Administration

TB	Tuberculosis-related Services
USB	Universal Serial Bus
VA	Veterans Affairs