Chronic Conditions Warehouse

Your source for national CMS Medicare and Medicaid research data

Chronic Conditions Warehouse Virtual Research Data Center

T-MSIS Analytic Files (TAF) Research Identifiable Files (RIFs) User Guide

APRIL 2025 | VERSION 1.12

Revision Log

Date	Changed by	Revisions	Version
April 2025	K. Schneider	Edits made to comply with Executive Order 14168 and NUBC™ licensing	1.12
December 2024	K. Schneider	Updated Appendix B — State Cutover to T-MSIS to add Guam and reflect cessation of MT TPA and consolidation of WY CHIP into WY data file; updated Appendix D — TAF RIF Releases	1.11
December 2023	K. Schneider	Updated Appendix D — TAF RIF releases; described upstream beneficiary identifier change and potential for clusters of BENE_IDs with different MSIS_ID/SUBMTG_STATE_CD values. Updated cms.gov URLs that have changed	1.10
December 2022	K. Schneider	Updated Appendix D — TAF RIF Releases	1.9
October 2022	K. Schneider	Updated document for the 2020 data release; described new values pertinent to the COVID public health emergency (PHE) in the DE base file ELGBLTY_GRP_CD and RSTRCTD_BNFTS_CD_LTST; added reference to some (not all) new fields, including DE base file Monthly Reason For Change in Eligibility Status (ELGBLTY_CHG_RSN_CD_01–12); claims — a new federally-derived service category (FED_SRVC_CTGRY_CD), AHRQ Clinical Classifications Software Refined (CCSR) Diagnosis Category Code variable (DGNS_1_CCSR_CTGRY_CD) and a taxonomy code for the billing provider (BLG_PRVDR_NPPES_TXNMY_CD) derived from the NPPES files; described upstream beneficiary identifier change and potential utility for a 2020 BENE_ID bridge file	1.8
November 2021	A. Walsh	Deleted Appendix D — Type of service codes by file, update previous Appendix E — TAF RIF Releases (which becomes Appendix D)	1.7
September 2021	K. Schneider	Updated Appendix E — TAF RIF Releases	1.6
September 2021	K. Schneider	Adjusted language to reflect inclusion of all claims in the RIF, including service tracking and non-person claims.	1.5
June 2021	K. Schneider R Van Gilder A. Walsh	Updated document template and inserted Chapter 4 to describe new data files — APL and APR	1.4
December 2020	K. Schneider R. Van Gilder	Updated paper to remove the 2018 column in Table 1 and Appendix B (since it was identical to 2017)	1.3
November 2020	K. Schneider R. Van Gilder	Updated paper for the 2014–2016 data “Release 2,” updated DQ references	1.2
August 2020	K. Schneider R. Van Gilder	Updated paper for the 2017–2018 data release, including Table 1 and Appendix B. Introduced the Data Quality (DQ) Atlas and added hyperlinks	1.1
November 2019	R. Van Gilder K. Schneider	Initial release	1.0

Table of Contents

List of Tables

List of Figures

List of Data Tips

Overview

The Centers for Medicare & Medicaid Servicers (CMS) uses the Chronic Conditions Warehouse (CCW) to develop and manage CMS research data sources. Academic researchers and certain government agencies with approval under a Data Use Agreement (DUA) to obtain CMS data for research purposes may request Research Identifiable Files, or RIFs. CMS has authorized the release of a new format of Medicaid and Children's Health Insurance Program (CHIP) data to the research community called the Transformed Medicaid Statistical Information System (T-MSIS). This user guide describes the new T-MSIS research files, their strengths and weaknesses compared to previous Medicaid data formats, and how to obtain them.

Medicaid is a state-administered benefit with guidance/requirements and shared funding from the federal government and individual state governments (aka Title XIX of the Social Security Act from 1965). Each state must provide the minimum federally mandated services and coverage for federally mandated eligibility groups; however, state benefits may vary from state to state. Medicaid covers many groups of people, depending on the state's requirements (e.g., age, pregnancy status, disabled, blind, 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 children with disabilities living at home. CMS administers the Medicaid program jointly with states. Reference the CMS Medicaid.gov website for additional information regarding Medicaid.

The Balanced Budget Act of 1997 amends the Social Security Act to add Title XXI, the State Children's Health Insurance Program (aka S-CHIP, or simply CHIP). CHIP enables states to extend insurance coverage to low-income children who are not eligible for Medicaid and do not have private insurance. States have the flexibility to design their program within federal guidelines. Benefits vary by state and the type of CHIP program (e.g., whether the CHIP program operates as an expansion of Medicaid [sometimes referred to as M-CHIP], a separate child health insurance program [referred to as S-CHIP], or a combination of the two). For more information on the CHIP program, reference the CMS website.

The Patient Protection and Affordable Care Act of 2010 (ACA) expanded insurance coverage by offering private individual policies via a centralized "marketplace" and state Medicaid expansions. The ACA expands Medicaid coverage for most low-income adults to 138% of the federal poverty level (FPL). As of 2014, states may decide whether to adopt Medicaid expansion; not all states have expanded their Medicaid programs.[1]

Each state compiles information regarding their Medicaid and CHIP enrollment, service utilization, and payments in the recently implemented T-MSIS format and provides T-MSIS data files to CMS. Using the T-MSIS data files from states, CMS creates the T-MSIS Analytic Files (TAFs). The CCW team obtains the TAF files, loads them to a database, and creates claims and enrollment Research Identifiable Files (RIFs). These files are available to academic researchers and certain government agencies with an approved DUA for research purposes. The TAF RIFs contain personally identifiable information (PII) and are subject to the Privacy Act and other federal government rules and regulations (reference the ResDAC website for details on requesting TAF RIF data).

For Medicaid data from 1999 through 2014,[2] CMS produced annual MAX (Medicaid Analytic eXtract) files using data derived from the Medicaid Statistical Information System (MSIS). CMS has phased out MSIS, and therefore MAX, as states converted their Medicaid and CHIP data submissions to CMS into the new T-MSIS format.

---

[1] HealthCare.gov. Medicaid & CHIP. "Medicaid expansion & what it means for you." https://www.healthcare.gov/medicaid-chip/medicaid-expansion-and-you// (Accessed 12/09/2024)

[2] Not all states have MAX files for 2014; 2013 MAX may be the last year CMS has included some states with this data format. Refer to Appendix B — State Cutover to T-MSIS and Availability of MAX and TAF Research Files.

T-MSIS represents the next generation of national data for Medicaid and CHIP beneficiaries and the services they use. T-MSIS differs from MSIS in several important ways, including the timing of submissions (monthly versus quarterly) and the amount of content states report (nearly four times as many data elements, including several new segments).

The CCW team links the TAF research files by a unique CCW person identifier (the CCW beneficiary identifier; variable called the BENE_ID), allowing researchers to analyze information across the entire continuum of care, both within and across years, without using standard person identifiers such as MSIS_ID, Social Security number, or Medicare Health Insurance Claim number for dual-eligible beneficiaries. This unique CCW identifier follows an enrollee across states and other CCW research data sources. If the same person enrolls in Medicaid or CHIP in more than one state or is also a Medicare beneficiary, researchers may combine those records using the BENE_ID. For example, the CCW also contains Medicare enrollment data, fee-for-service (FFS) claims data, Medicare Advantage encounter data submitted by plans[3] to CMS, and assessment data (e.g., Minimum Data Set [MDS] and Outcome and Assessment Information Set [OASIS]). The BENE_ID facilitates analysis across all CMS data sources in the CCW.

There are seven types of data files available for TAF RIF, including the following:

1. Annual demographic and eligibility (DE)
2. Inpatient (IP) claims
3. Long-term care (LT) claims
4. Pharmacy (RX) claims
5. Other services (OT) claims
6. Annual managed care plan (APL)
7. Annual provider (APR)

This guide provides users with information that may clarify and requesting TAF research files. For example, various resources are available to assist TAF data users, as described in Chapter 7 — Further Assistance with Medicaid and CHIP Data. Record layouts for the TAF research files are available on the CCW Data Dictionaries webpage. Appendix A — List of Acronyms lists the acronyms used throughout the paper. Related to acronyms, the CCW team refers to TAF research files, or TAF RIFs, as “TAFs” throughout the rest of this user guide for brevity.

[3] Medicare Advantage encounter data files are available starting with 2015.

Chapter 1 — Introduction

CMS uses the CCW to develop and manage CMS research data resources. The CCW team creates TAF RIFs from enrollment and service-related data submitted by state Medicaid and CHIP programs to CMS. States send Medicaid and CHIP service data they obtain from either FFS claims or managed care encounter records. In 2016, the Medicaid or CHIP programs enrolled approximately 75 million people.[4] Accordingly, the associated enrollment and claims data files are very large.

The CCW's Medicare, Medicaid, and CHIP data are subject to the HIPAA Privacy Rule, the Privacy Act, and other federal government rules and regulations. Researchers need to obtain a DUA to obtain the files. Researchers can find additional details regarding the data request process in Chapter 6 — Receiving CCW Data.

A. Source Data

The administration of Medicaid and CHIP programs is a state-federal partnership. According to federal requirements, states administer Medicaid and CHIP. States are responsible for eligibility verification and enrollment and providing benefits to qualifying recipients within their state.[5]

CMS requires that states share key information regarding the Medicaid and CHIP programs. Since 1999, CMS has required states to submit electronic Medicaid claims and eligibility data files through the Medicaid Statistical Information System, or MSIS. States began planning for and transitioning to the T-MSIS as early as 2013. However, initial T-MSIS reporting dates (sometimes referred to as "cutover dates") vary by state.

Each state selected a starting calendar quarter for T-MSIS data. The state calibrated the starting quarter with the last submitted data. CMS required all states to cut over to the T-MSIS format by October 2015, although some began submitting in this format earlier. Appendix B — State Cutover to T-MSIS and Availability of MAX and TAF Research Files list state-specific cutover dates.

TAF RIF data are available for each state starting with the first full calendar year after submitting T-MSIS data (Table 1).[6] By 2016, all states[7] had cut over to T-MSIS. The 2017 and 2018 TAF RIFs contain all states (including Washington DC, Puerto Rico, and the Virgin Islands).

[4] US HHS, CMS Monthly Medicaid Enrollment Reports https://www.medicaid.gov/medicaid/national-medicaid-chip-program-information/medicaid-chip-enrollment-data/monthly-medicaid-chip-application-eligibility-determination-and-enrollment-reports-data/index.html (Accessed 12/09/2024)

[5] There is additional information regarding the state-federal partnership and state responsibilities for data collection and submission to CMS on the Medicaid.gov website e.g., https://www.medicaid.gov/state-overviews/scorecard/national-context/index.html (Accessed 12/09/2024)

[6] Iowa and Pennsylvania began submitting CHIP data before Medicaid data, but CMS did not include Iowa and Pennsylvania in the TAF until their Medicaid cutover date.

[7] The Virgin Islands was not submitting to T-MSIS, so CMS has not included the Virgin Islands in the TAF RIF until 2017.

Table 1. Count of states with TAFs in CCW, by year

	2014	2015	2016	2017 forward
Number of States	19*	31*	52*	53

* 2014+ includes DC. 2015+ includes Puerto Rico. 2016 includes all states and DC except for the Virgin Islands. Guam is not available in any of these annual files until 2023.

CMS has not included states that did not submit T-MSIS data for all 12 months of the year in the RIFs for that service year (or earlier service years). For these states, researchers should request MAX data files.[8] The next section further discusses the MAX files.

B. Changes from MAX Data

CMS produced the MAX files from state submitted MSIS data from 1999 to 2013. MAX data contains five types of files that correspond to the initial release of T-MSIS RIF file types:

1. Person summary file (PS) — like the annual demographic and eligibility file in TAF
2. Hospital inpatient file (IP)
3. Long-term care file (LT)
4. Prescription drug file (RX)
5. Other services file (OT)

CMS phased out MAX as states converted their data submissions to the T-MSIS format. Depending on the state(s) of interest, for 2014 and 2015, researchers may need to request MAX files for some states and TAFs for other states. Reference Appendix B — State Cutover to T-MSIS and Availability of MAX and TAF Research Files to determine the data file type available for each state and service year.

The new TAFs contain information on a broader set of individuals than MAX and include Medicaid FFS and managed-care enrolled, M-CHIP or S-CHIP enrolled, and Medicaid expansion populations when applicable. MAX does not include S-CHIP claims.[9] The CCW team uses the acronym CHIP to mean both M-CHIP and S-CHIP, unless otherwise specified. There are also more than twice as many data elements submitted by states through T-MSIS, compared to MSIS.

For example:

- New demographic information is available (e.g., household size),
- Additional program enrollment/benefits information is present (e.g., more waiver details and more managed care enrollment details), and
- Two-part claims: header — plus all claim line records (MAX limited claim line information to 26 lines)
- Additional files — for managed care plans (the APL file) and providers (the APR file)

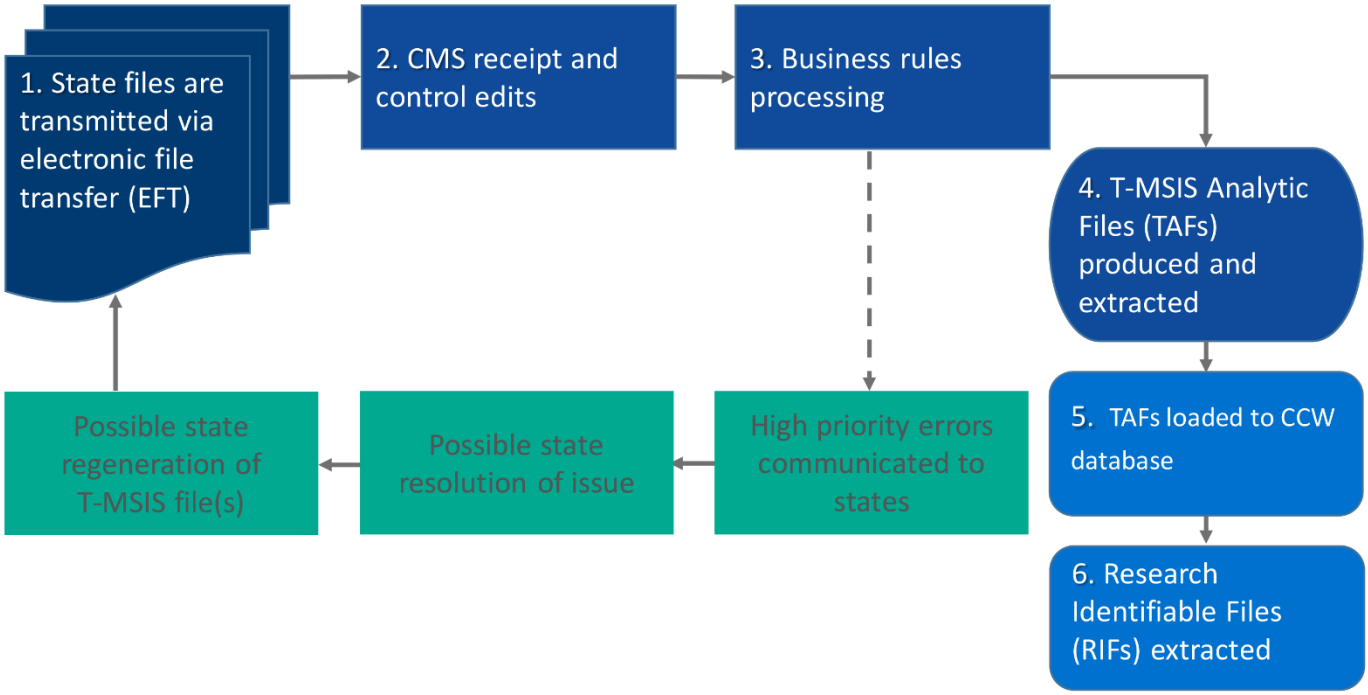
[8] For states that do not have a full 12 months of T-MSIS data, CMS produces MAX files. Reference https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAXGeneralInformation.html (Accessed 12/09/2024)

[9] MAX included M-CHIP claims but not S-CHIP claims.

C. Creation of TAF Research Files

This section of the document contains information about the Medicaid and CHIP source data used to create the TAF research files. States collect and process enrollment, claims, managed care plan and provider data, and submit that data to CMS via T-MSIS. CMS creates standard T-MSIS Analytic Files (TAFs) and sends them to the CCW, which loads them to the relational CCW database and then extracts standard RIFs to disseminate to data users. A visual depiction of the Medicaid and CHIP data flow is in Figure 1.

Figure 1. T-MSIS data processing flow



States submit transactional T-MSIS files to CMS each month. A small number of states submit data for non-Medicaid entities (e.g., CHIP or Third-Party Administrator [TPA]) separately.[10] Therefore, CMS refers to it as a submitting state "entity" (i.e., the variable called SUBMTG_STATE_CD) that researchers can roll up into the state variable (called STATE_CD). States may re-submit or re-send data for all service dates. A CMS database stores each claim record submitted by states and uses an algorithm to group all records into a unique service "claims family." An original claim and any claims that adjust the original claim make up a claim family, consisting of one or multiple claims depending on the number of adjustments. Within a family, any claim flagged as final action means that it is the final record for the service reported on all claims in the family. The TAF includes only the final action claims to help make the data more manageable for analytic purposes. Additional details regarding TAF final action, among various other key research topics, are available in a series of informational white papers called "Data Quality Assessments" and "Data Quality Snapshots."[11] Data Quality (DQ) Assessments and Snapshots are available to users in a new Data Quality tool called DQ Atlas, available on the Medicaid.gov website.

[10] The states that submit data for more than one entity are: PA, CHIP, WY CHIP, IA CHIP, and MT TPA.
[11] Reference "Additional data quality information." Final Action Status in T-MSIS Claims. https://www.medicaid.gov/dq-atlas/landing/resources (Accessed 12/09/2024)

States also submit monthly enrollment files to CMS. CMS accumulates and produces the TAF annual DE file, representing every beneficiary enrolled in a Medicaid or CHIP program anytime during the service year, as well as the four TAF claims files (inpatient, long-term care, other services, and pharmacy). CMS offloads (extracts) the TAF files and sends them to the CCW. The CCW team assigns a BENE_ID and partitions the data into calendar year RIFs based on service (or enrollment) date. Reference Chapter 1, section D. Assignment of a Beneficiary Identifier, below for more information on the beneficiary identifier. For early service years, CMS has produced more than one version of TAF RIFs because the states made significant improvements in their data quality over time. Appendix D — TAF RIF Releases lists the service years and their corresponding TAF RIF releases.

Although claim-level payment information is present for all FFS claims, the CMS Center for Medicaid and CHIP Services (CMCS) has required redaction of some payment fields for managed care claims due to the proprietary and confidential nature of this information. Additional details regarding the claims files contents are in Chapter 3, section B. Content and Description of Claims RIFs.

States submit monthly provider, managed care plan, and waiver files to CMS. CMS produces the TAF APL and APR files from that monthly data. The APL file contains information about each Medicaid and CHIP managed care plan or waiver entity that CMS authorized to operate in the state during the calendar year. The APR file contains information about the providers (facilities, groups, or individual practitioners) that are eligible to render services to Medicaid and CHIP beneficiaries for any month of a service year. Additional details regarding the APL and APR files are in Chapter 4 — Other Annual Files.

Researchers may obtain the record layouts for all TAF research files from the CCW Data Dictionaries webpage. Throughout this document, when the CCW team identifies a particular data variable by name, it often identifies the specific SAS® name by using all capitals.

D. Assignment of a Beneficiary Identifier

The CCW team assigns the beneficiary identifier (BENE_ID) to a unique person across distinct CCW data sources, thereby allowing researchers to track all beneficiary services over time. Further, the CCW team assigns Medicaid enrollees who are also Medicare program enrollees (or “dual eligible” beneficiaries), the same BENE_ID for both Medicaid and Medicare records.

To construct the BENE_ID, the CCW team developed internal cross-reference files consisting of historical Medicaid and Medicare enrollment information using CMS data sources such as the Enterprise Cross Reference (ECR) file. When the CCW team receives a new TAF enrollment file, they examine the MSIS_ID, STATE_CD, Social Security number (SSN), date of birth, sex, and other beneficiary identifying information to determine if a BENE_ID already exists.

If there is a single record that “best matches” the information in the Medicaid enrollment record, then the CCW team assigns the BENE_ID on that historical record to the Medicaid enrollment record. If there is no match or no “best match” after CCW exhausts a stringent matching process, the CCW team assigns a null (or missing) BENE_ID to the Medicaid enrollment record.

For any given year, approximately 4% of TAF records in the DE base file have a null (or missing) BENE_ID. Once the CCW team assigns a BENE_ID to a Medicaid or CHIP enrollment record for a particular year (except for those assigned to a null value), it does not change. When the CCW team receives a new Medicaid or CHIP enrollment file, it attempts to reassign those with missing BENE_IDs. The TAF research files do include the TAF DE records that are missing a BENE_ID.

The CCW team also assigns the BENE_ID from the Medicaid and CHIP enrollment record to each respective claim record (IP, LT, RX, OT). As with enrollment records, once the CCW team assigns a BENE_ID to a claims record, it does

not change. Additional guidance regarding the use of BENE_ID or MSIS_ID for linking TAFs together or for aggregating data appear later in this user guide (reference Chapter 5, section A. Linkage Variables and Joining Files and B. Summarizing Data).

The unique BENE_ID field is specific to the CCW and does not apply to any other identification system or data source. Before delivering data files to researchers, the CCW team encrypts this identifier. In addition, the CCW team encrypts all data files it sends to researchers (reference encryption information in Chapter 6 — Receiving CCW Data for details). It is important to note that each research request employs a different encryption key for the beneficiary identifier field and the data files.

Chapter 2 — Annual DE RIF

The annual DE TAF RIF is an annual file that CMS creates from the monthly state enrollment data. Chapter 1, section C. Creation of TAF Research Files, presents additional details. Each annual DE file includes all Medicaid and CHIP enrollees documented as enrolled in a Medicaid or CHIP program for at least one day of the year.

Upon enrollment, each state assigns its Medicaid and CHIP beneficiaries an MSIS identifier, also known as the MSIS_ID. Note that some states assign more than one MSIS_ID to the same beneficiary.

State-assigned MSIS_IDs are not ideal for research purposes because beneficiaries may have more than one MSIS_ID number over time due to enrollment in Medicaid and CHIP in the same or different states. The BENE_ID allows linkage of an individual's data both within and across states, years, different TAF RIF file types, and other CCW data sources. Some beneficiaries do not have a BENE_ID; additional details are in Chapter 5, section A. Linkage Variables and Joining Files.

The DE RIF contains several files — the "base" or core enrollment/demographic file, and six supplemental files: 1) Eligibility Dates, 2) Managed Care Enrollment, 3) Waiver Program Enrollment, 4) Money Follows the Person (MFP), 5) health home and State Plan Options (SPO), and 6) Disability and Need. The next section describes these files.

A. DE Base File

The DE RIF base is the primary eligibility file that nearly all researchers use. It contains approximately 200 variables with information on beneficiaries' demographic characteristics and details of their enrollment in Medicaid or CHIP. The DE data file has a record for each MSIS_ID within a submitting state entity (SUBMTG_STATE_CD); note that the SUBMTG_STATE_CD may roll up multiple state entities into a single state postal code (STATE_CD). As often as possible, the CCW team populates the BENE_ID correctly for a person across these submitting state entities. Since the same BENE_ID may apply to more than one MSIS_ID and SUBMTG_STATE_CD, data users should link the DE base with the DE supplemental files using the MSIS_ID and SUBMTG_STATE_CD to ensure the correct DE supplemental file records are associated with the selected DE base record. Chapter 5, section A. Linkage Variables and Joining Files, presents additional details, and there is an example in Table 21.

The DE base file includes variables related to demographics and geography (reference Table 2). It also includes monthly and annual enrollment information; how the beneficiary qualified for Medicaid or CHIP; information regarding eligibility group codes, enrollment in managed care, and indicators for enrollment in waivers (and/or Money Follows the Person demonstration, Health Home, etc.). Refer to Table 3 for examples of monthly and yearly enrollment variables.

The DE base file contains a small number of "dummy" records to represent beneficiaries who have claims data but no eligibility record in the source T-MSIS data from the states. Since most TAF RIF analyses use information from the DE enrollment files, CMS has created a dummy record in the DE base file for such an MSIS_ID that it populates only with the MSIS_ID, SUBMTG_STATE_CD and a few valid values. Researchers can identify the dummy enrollment records by a missing eligibility data indicator (MISG_ELGBLTY_DATA_IND) value of 1. The DE base file does not contain "dummy" records for most service tracking claims, because these records are usually non-person-specific lump-sum payments to providers or plans. CMS only creates dummy eligibility records for non-null/missing MSIS IDs that do not begin with an ampersand (&). Refer to Chapter 3, section C. 2. Person Claim Indicator for more information on service tracking claims.

Table 2. DE base file — examples of key demographic and geographic variables

DE base variables	Description	SAS variable name
Demographic information	Date of birth	BIRTH_DT
	Age (in years)	AGE
	Date of death	DEATH_DT
	Sex (biological) — latest in year	SEX_CD
Geographic information	State FIPS code for beneficiary home or mailing address — latest in year	BENE_STATE_CD
	County code for beneficiary home or mailing address — latest in year	BENE_CNTY_CD
	ZIP code for beneficiary home or mailing address — latest in year	BENE_ZIP_CD

Table 3. DE base file — examples of key enrollment variables

DE base variables	Description	SAS variable name
Monthly information	Medicaid enrollment days — (one variable for each month)	MDCD_ENRLMT_DAYS_01–12
	CHIP enrollment days — (one variable for each month)	CHIP_ENRLMT_DAYS_01–12
	Medicaid, Medicaid Expansion CHIP, or separate CHIP code — (one variable for each month)	CHIP_CD_01–12
	Scope of Medicaid or CHIP benefits — (one variable for each month)	RSTRCTD_BNFTS_CD_01–12
	Medicare-Medicaid dual eligibility code — (one variable for each month)	DUAL_ELGBL_CD_01–12
	Eligibility group code — (one variable for each month)	ELGBLTY_GRP_CD_01–12
	Reason for change in eligibility status — (one variable for each month)	ELGBLTY_CHG_RSN_CD_01–12
Yearly information	Medicaid enrollment days — total in year	MDCD_ENRLMT_DAYS_YR
	CHIP enrollment days — total in year	CHIP_ENRLMT_DAYS_YR

Identifying the type and level of benefits for enrollees is central to understanding the service use for the population when using claims data files. Some important subgroups that are of interest to many data users include: 1) Medicaid (Title XIX) enrolled, and 2) M-CHIP (Title XXI) enrolled, and 3) S-CHIP (Title XXI) enrolled. Data Tip 1 highlights an important data consideration.

Data Tip 1. How to identify Medicaid or CHIP enrollment

The monthly Medicaid, Medicaid Expansion CHIP, or Separate CHIP code field is one of the key fields to identify beneficiaries enrolled in Medicaid or CHIP for the month (CHIP_CD_01–12). Possible monthly values are:

1 = Medicaid-enrolled month
2 = Medicaid Expansion CHIP (M-CHIP) month
3 = Separate Title XXI CHIP (S-CHIP) enrolled month
4 = Individual was both Medicaid and separate CHIP eligible during the same month
Null/missing = not enrolled during the month

The CHIP_CD, along with the eligibility group code (ELGBLTY_GRP_CD_01–12), can provide enrollment information.

A SAS® coding example illustrates one method for identifying Medicaid or CHIP enrollees; of course, researchers may adapt this code to use any statistical software they prefer:

1) Begin by using the DE RIF, which the authors call "DE_BASE" in this example.
2) Create a variable called "group," where the values are Medicaid, M-CHIP, S-CHIP, and unclassified.
3) Use the latest CHIP_CD (CHIP_CD_LTST) and the latest eligibility group code (ELGBLTY_GRP_CD_LTST) to parse the data. For S-CHIP, the example checks to see if there were any CHIP enrollment days in the year (using CHIP_ENRLMT_DAYS_YR).
4) Save a file with the researcher's classifications — called "population."

```
data population;
      length group $20;
      set DE_BASE;
      if CHIP_CD_LTST = '1' and ELGBLTY_GRP_CD_LTST in ('01' '02' '03' '04' '05' '06' '07' '08' '09' '10' '11' '12' '13' '14'
          '15' '16' '17' '18' '19' '20' '21' '22' '23' '24' '25' '26' '27' '28' '29' '30' '31' '32' '33' '34' '35' '36' '37' '38' '39'
          '40''41' '42' '43' '44' '45' '46' '47' '48' '49' '50' '51' '52' '53' '54' '55' '56' '57' '58' '59' '60'
           '69' '70' '71' '72' '73' '74' '75') then group = "Medicaid";
          else if CHIP_CD_LTST = '2' then group = "M-CHIP";
          else if CHIP_CD_LTST = '3' and CHIP_ENRLMT_DAYS_YR > 0 then group = "S-CHIP";
          else group = "Unclassified";
run;
```

The restricted benefits code (RSTRCTD_BNFTS_CD_01–12) also provides information regarding the scope of Medicaid or CHIP benefits (e.g., full or partial benefits).

Starting in March 2020, individuals may have Medicaid enrollment due to the SARS-COVID-19 pandemic during the public health emergency (PHE) period. CMS implemented a new ELGBLTY_GRP_CD=76 "uninsured individual eligible for COVID-19 testing," and also added a RSTRCTD_BNFTS_CD_LTST=F "individual is eligible for Medicaid but is only entitled to restricted benefits for medical assistance for COVID-19 diagnostic products and any visit described as a COVID–19 testing-related service for which payment may be made under the state plan during any portion of the PHE, beginning March 18, 2020."

There is state-level variability in data quality related to the identification of Medicaid/CHIP enrollment, which may necessitate a variation on this logic to appropriately identify the population of interest for the researcher's study. To learn more about identifying populations, refer to the DQ Atlas for topic-specific assessments and benchmarking statistics by state or topic.

In addition to Medicaid/CHIP enrollment variables mentioned above, there are fields to identify state-specific eligibility groups (variables such as STATE_SPEC_ELGBLTY_GRP_CD_01–12). There are also some variables with an annual value.

These include variables CMS populates with the latest valid value in the source data (naming convention is *_LTST), and other variables that indicate whether a program or service ever applies during the year (naming convention is *_IND in conjunction with a SAS label containing the phrase "Ever in Calendar Year"). Refer to Table 4 for examples. Data users should refer to the TAF enrollment codebook to obtain additional details, including the valid values.

Table 4. DE base file — examples of key annual enrollment variables

DE base variables	Description	SAS variable name
Latest in year	Medicaid, Medicaid Expansion CHIP, or separate CHIP code — latest in year	CHIP_CD_LTST
	Scope of Medicaid or CHIP benefits — latest in year	RSTRCTD_BNFTS_CD_LTST
	Medicare-Medicaid dual eligibility code — latest in year	DUAL_ELGBL_CD_LTST
	Eligibility group code — latest in year	ELGBLTY_GRP_CD_LTST
Ever in year	Disability Indicator — difficulty concentrating — ever in calendar year	DSBLTY_DFCLTY_CNCNTRTNG_IND
	Disability Indicator — difficulty walking — ever in calendar year	DSBLTY_DFCLTY_WLKG_IND
	Disability Indicator — difficulty dressing or bathing — ever in calendar year	DSBLTY_DFCLTY_DRSNG_BATHNG_IND

Not all enrollees in the DE base file are in the supplemental files. To determine if a beneficiary has a record in a supplemental file, refer to the DE base indicator variables. Refer to Table 5. These nine fields are binary indicators, where 1 = beneficiary has a record in the supplemental file.

Table 5. DE base file — variables indicating records in DE supplemental file

Description	SAS variable name
Beneficiary Record in supplemental dates file	SPLMTL_DTS
Beneficiary Record in supplemental managed care file	SPLMTL_MC
Beneficiary Record in supplemental waiver file	SPLMTL_WVR
Beneficiary Record in supplemental home health and State Plan Option (SPO) file	SPLMTL_HLTH_HOME_SPO
Beneficiary Record in supplemental Money Follows Person (MFP) file	SPLMTL_MFP
Beneficiary Home and Community-Based Services (HCBS) record in supplemental disability file*	SPLMTL_DSB_HCBS
Beneficiary long-term services and supports (LTSS) record in supplemental disability file*	SPLMTL_DSB_LTSS
Beneficiary lock-in record in supplemental disability file*	SPLMTL_DSB_LCKIN
Beneficiary other needs record in supplemental disability file*	SPLMTL_DSB_OTHR

* Any of these four indicators means there is a record in the disability and need supplemental file.

B. DE Supplemental Files

Users can augment the DE base RIF using the six DE supplemental files. A description of the variables in the DE file is accessible on the CCW Data Dictionaries webpage. The following section briefly describes each of these six files (Table 6).

Table 6. DE supplemental files

DE supplemental file	Number of variables in RIF
1. Eligibility dates	8
2. Managed care (MC) enrollment	412*
3. Waiver program enrollment	268
4. Money Follows Person (MFP)	23
5. Health home and State Plan Options (SPO)	99
6. Disability and need	220

* Prior to August 2021, the MC enrollment supplemental file included 410 variables.

1. Eligibility Dates

This file may have more than one record per beneficiary; that is, there may be multiple records for an MSIS_ID (within a SUBMTG_STATE_CD) for a calendar year. Each record represents an enrollment spell for a Medicaid or CHIP beneficiary. A person may have more than one record in this supplemental file if there was more than one enrollment type code during the year (e.g., the person was CHIP for part of the year and Medicaid for part of the year). Another reason a person may have multiple enrollment periods within the year is that Medicaid covered a few months early in the year, did not cover for several months, and then covered again at the end of the year.

There is an enrollment start and end date (ENRLMT_START_DT and ENRLMT_END_DT) for each record. Note that enrollment days across Medicaid and CHIP may overlap. Additional information regarding enrollment benchmarking and enrollment patterns over time is available on the "Resources" page of the DQ Atlas.[12]

2. Managed Care Enrollment

The Managed Care Enrollment supplemental file contains one record for each beneficiary (identified by each unique combination of MSIS_ID and SUBMTG_STATE_CD) who enrolled in some form of managed care in any month during the calendar year. The file accommodates up to 16 managed care plan type codes and associated plan IDs for each calendar month: these arrays account for 384 of the 410 variables in the file (93.6%). CMS also constructs variables that summarize the count of months a beneficiary enrolls in each managed care plan type within the year.

3. Waiver Program Enrollment

The Waiver Program Enrollment supplemental file contains one record for each beneficiary (i.e., each unique combination of MSIS_ID and SUBMTG_STATE_CD) who enrolls in at least one waiver program in any month during the calendar year. The file accommodates 10 waiver type codes and associated waiver IDs for each month: these arrays account for 240 of the 268 variables in the file (89.5%). CMS also constructs variables that summarize the count of months a beneficiary enrolls in each type of waiver program within the year (Table 7).

[12] https://www.medicaid.gov/dq-atlas/landing/resources (Accessed 12/09/2024)

Table 7. DE Waiver Program Enrollment supplemental file — annual summary variables

Description	SAS variable name
1915(c) waiver months	WVR_1915C_MOS
1915(b) waiver months	WVR_1915B_MOS
1915(b)(c) waiver months	WVR_1915BC_MOS
1115 pharmacy waiver months	PHRMCY_WVR_1115_MOS
1115 disaster-related waiver months	DSTR_RLTD_WVR_1115_MOS
1115 family planning only waiver months	FMLY_PLNG_ONLY_WVR_1115_MOS
1115 Health Insurance Flexibility and Accountability Waiver months	HIFA_WVR_1115_MOS
1115 other type of waiver months	OTH_WVR_1115_MOS
Other waiver type months	OTH_WVR_TYPE_MOS

4. Money Follows the Person (MFP)

The MFP supplemental file contains one record for each beneficiary (identified by each unique combination of MSIS_ID and SUBMTG_STATE_CD) who participated in an MFP program in any month during the calendar year. Variables include monthly Money Follows Person (MFP) participant indicators (variables called MFP_IND_01–12). The file also contains constructed variables that identify the last known MFP status for the beneficiary within the year (Table 8). Additional information regarding the MFP Program is available on the Medicaid.gov website.[13]

Table 8. DE money follows person (MFP) supplemental file — annual summary variables

Description	SAS variable name
Money Follows Person (MFP) —lives with family or non-participant code	MFP_LVS_WTH_FMLY_CD
Money Follows Person (MFP) — qualified residence code	MFP_QLFYD_RSDNC_CD
Money Follows Person (MFP) — qualified institution code	MFP_QLFYD_INSTN_CD
Money Follows Person (MFP) — re-institutionalized reason code	MFP_RINSTLZD_RSN_CD
Money Follows Person (MFP) — participation ended reason code	MFP_PRTCPTN_END_RSN_CD

5. Health Home and State Plan Options (SPO)

The health home and SPO supplemental file contains one record for each beneficiary (identified by each unique combination of MSIS_ID and SUBMTG_STATE_CD) who participated in a Health Home, enrolled in an SPO, or had a Health Home chronic condition in any month during the calendar year. Variables include the name and provider identifier for the Health Home, and monthly Health Home Program participation indicator variables (called HLTH_HOME_PGM_IND_01–12). The file also has monthly variables that indicate enrollment in state plan options for each month, including Community First Choice, 1915i, 1915j, 1932a, and 1937 Alternative Benefit Plan. Additional information regarding these plans is available on the Medicaid.gov website.[14] The eight Health Home chronic condition indicators are: asthma, diabetes, heart disease, mental health, substance abuse, HIV/AIDs, overweight, and other conditions.

[13] The Medicaid.gov website explains MFP: https://www.medicaid.gov/medicaid/long-term-services-supports/money-follows-person/index.html (Accessed 12/09/2024)

[14] For example, CMS describes the Community First Choice (CFC) at: https://www.medicaid.gov/medicaid/home-community-based-services/home-community-based-services-authorities/community-first-choice-cfc-1915-k/index.html (Accessed 12/09/2024)

6. Disability and Need

The Disability and Need supplemental file contains one record for each beneficiary (identified by each unique combination of MSIS_ID and SUBMTG_STATE_CD) who met one of the following criteria:

- met the requirements for Long-Term Services and Supports (LTSS) program eligibility
- was eligible through the conception to birth option
- participated in other government assistance programs
- had any of the following — a Home and Community-Based Services (HCBS) chronic condition, a lock-in provider, a disabling condition, or third-party liability coverage

Within the file, there are ten 10 HCBS Chronic Condition indicators: aged, physical disability, intellectual disability, developmental disability, other disability, autism, mental illness, brain injury, HIV/AIDs, and technology-dependent medically fragile.[15] Suppose beneficiaries are "locked-in" to specific providers to better monitor services and reduce unnecessary or inappropriate utilization. In that case, there are variables for identifying up to three lock-in providers (provider ID -LCKIN_PRVDR_ID_1–3, each with an associated provider type code: LCKIN_PRVDR_TYPE_CD_1–3).

There are also monthly disability indicators for activities, as well as monthly variables related to enrollment in the following programs: Birth to Conception, Social Security Disability Insurance (SSDI), Supplemental Security Income (SSI), Temporary Assistance for Needy Families (TANF), and Third-Party Liability (TPL) insurance coverage. Refer to Table 9.

Table 9. DE Disability and Need supplemental file — monthly variables

Monthly disability variables — description	SAS variable name[16]
Disability indicator — difficulty concentrating (one variable for each month)	DSBLTY_DFCLTY_CNCNTRTNG_IND_01–12
Disability indicator — difficulty walking (one variable for each month)	DSBLTY_DFCLTY_WLKG_IND_01–12
Disability indicator — difficulty dressing or bathing (one variable for each month)	DSBLTY_DFCLTY_DRSNG_BTHNG_IND_01–12
Disability indicator — difficulty running errands alone (one variable for each month)	DSBLTY_DFCLTY_ERNDS_IND_01–12

Monthly program eligibility variables — description	SAS variable name
Birth to conception indicator (one variable for each month)	BIRTH_CNCPTN_IND_01–12
Social Security Disability Insurance (SSDI) indicator (one variable for each month)	SSDI_IND_01–12
Supplemental Security Income (SSI) indicator (one variable for each month)	SSI_IND_01–12
Supplemental Security Income (SSI) status code (one variable for each month)	SSI_STUS_CD_01–12

[15] Additional information is in DQ Atlas. Reference "Data Quality and Analytic Resource Downloads. Enrollment Benchmarking" topic on the "Resources" page. https://www.medicaid.gov/dq-atlas/landing/resources (Accessed 12/09/2024)
[16] These are monthly variables (01 thru 12).

Monthly program eligibility variables — description	SAS variable name
Supplemental Security Income (SSI) state supplement code (one variable for each month)	SSI_STATE_SPLMT_CD_01–12
Temporary assistance for needy families (TANF) cash code (one variable for each month)	TANF_CASH_CD_01–12
Third-party liability (TPL) insurance coverage indicator (one variable for each month)	TPL_INSRNC_CVRG_IND_01–12
Third-party liability — other coverage indicator (one variable for each month)	TPL_OTHR_CVRG_IND_01–12

Finally, there are monthly indicators for the overall level of care required to meet an individual's needs and determine LTSS program eligibility (CARE_LVL_STUS_CD_01–12). Related variables include up to three LTSS providers (LTSS_PRVDR_ID_1–3) during the year, and the associated monthly values for the level of care codes for each of these providers (LTSS_LVL_CD_1_01–12 for provider one through LTSS_LVL_CD_3_01–12 for provider three). In addition to the monthly level of care codes, there is a variable with the latest value for the level of care codes 1, 2, and 3 during the year (e.g., LTSS_LVL_CD_2_LTST stands for the Long-Term Services and Supports provider 2 level of care code — latest in year).

Chapter 3 — TAF Research File Claims

States submit FFS claims, managed care encounter claims, and supplemental payments for Medicaid, Medicaid-expansion CHIP, and Separate CHIP to T-MSIS. As stated in Chapter 1, section A. Source Data, the TAF claims files include all "final action" Medicaid and CHIP service records for a given year (e.g., all T-MSIS claims CMS determined to be final, as of the TAF creation date). The claims included in these files are active, final-action, non-voided, and non-denied claims.[17] (except for IL, reference Chapter 5, section C. 3. Known Issue with Illinois Claims)

It is important to note that there is significant state-by-state variability in the quality of claims and encounter data. Refer to the Data Quality documentation in the DQ Atlas tool and information later in Chapter 4 of this user guide regarding TAF Data Quality (reference, for example, Chapter 5, section C. Considerations for Selecting States for Analysis.

The monthly claims TAFs include only claims with a service end date in that month, along with their associated claim line records.

The TAF claims files are available for four settings:
1. Inpatient (IP)
2. Long-term care (LT)
3. Other services (OT)
4. Pharmacy (RX)

States submit the claims data for each of the settings as header and trailer records, which the CCW refers to as the "header" and "line" records in the RIF files. In addition, there are "occurrence" code files for IP, LT, and OT. The header file contains the core/basic service record, while the line file may contain multiple line items for a corresponding header record.[18] The header and line file records for a service link using the CLM_ID, which is a variable added by the CCW team specifically to link claim header and line records.

CMS sequentially numbers lines associated with a header claim using the sequential claim line number (LINE_NUM). Similarly, CMS sequentially numbers the occurrence code records for a header claim using the occurrence code sequence (variable called OCRNC_CD_SEQ). The CCW team adds the LINE_NUM and OCRNC_CD_SEQ variables to facilitate ease of use of the files. Later in this chapter, there are examples of how to associate lines or occurrence code records with the header claim.

A. Structure of Claims Files

1. Header Files

The header file contains claim header information such as the BENE_ID, MSIS_ID, claim type code, service begin (or admission) date, service end (or discharge) date, adjudication date, provider ID, managed care plan ID, waiver ID and type code, diagnosis codes, as well as the diagnosis-related group (DRG), and up to six procedure codes with associated dates (for IP records).

[17] "Non-denied" means the state did not deny the claim at the header level; but may have denied lines in the line file — e.g., the state didn't completely deny the claim; however, it denied some lines for these claims.
[18] It is possible, although not common, to have a header claim with no associated line records (e.g., nothing to join to within the line file).

The number of diagnosis code fields varies by file:

- for IP, there are up to 12 DX codes
- for LT, there are up to five
- for OT, there are up to two diagnosis codes
- RX has none

Claim-level payment information is also present for all FFS claims; however, the Center for Medicaid and CHIP Services (CMCS) at CMS has required redaction of some payment fields for managed care claims due to the proprietary and confidential nature of this information (refer to Appendix C — Managed Care Redacted Payment Fields). Redacted payment fields on the RIF include Medicaid paid amount, other insurance paid amounts (Medicare or a third party), and beneficiary liability amounts (deductibles and coinsurance). When summing expenditures using the claim files (either header or line) it is important to keep in mind that totals do not reflect the redacted managed care expenditures.

2. Line Files

Records in the line files are associated with the header claim (i.e., a particular CLM_ID). There are never "orphan" lines, (e.g., line records without an associated header claim record). However, it is possible to have a header claim record without any associated line records.

Line files have a separate record for each revenue center or ambulatory procedure (note that IP claims, the header record accommodates major procedures designated with a procedure code). Although providers often document the procedure using an ICD-9 or ICD-10-PCS code, states may use their own codes (refer to variables that identify the procedure code system/nomenclature used — PRCDR_CD_SYS_1–6). Revenue centers represent institutional cost centers which bill separate charges. For example, inpatient hospital claims may include revenue centers for emergency department (ED), intensive care, physical therapy, laboratory, pharmacy, blood, imaging, etc. For non-institutional settings, claim lines may represent distinct procedures (e.g., CPT or HCPCS codes) or National Drug Codes (NDCs).

It is common to use the line-level detail if researchers are interested in the OT or RX claims because the line record contains important information to identify the delivered services. The state assigns the type of service code (TOS_CD) at the line level to describe the service (while in the MAX claims, CMS assigns the type of service code, MSIS_TOS, at the header claim level).

The line files include the beginning and ending dates of service for the line, and information about the service provider and payment. CMS redacts the payment fields on the RIF line for managed care claims (refer to Appendix C — Managed Care Redacted Payment Fields).

Table 10 provides an example of how researchers identify lines associated with the header claim.

Table 10. Example of header and line records for an OT claim

BENE_ID	Header or line file	CLM_ID	LINE_NUM	SRVC_BGN_DT (on header)	SRVC_END_DT (on header)	Procedure code or revenue center code	Type of service code	BLG_PRVDR_ID
000001	header	123456	n/a	14MAY2016	14MAY2016	n/a	n/a	ABC117
000001	line	123456	1	(LINE_SRVC_BGN_DT) 14MAY2016	(LINE_SRVC_END_DT) 14MAY2016	99213	003	line (SRVC_PRVDR_ID) 007
000001	line	123456	2	14MAY2016	14MAY2016	36415	003	007
000001	line	123456	3	14MAY2016	14MAY2016	G0467	003	007
000885	header	987654	n/a	11JUL2016	11JUL2016	n/a	n/a	QRS999999
000885	line	987654	1	(LINE_SRVC_BGN_DT) 11JUL2016	(LINE_SRVC_BGN_DT) 11JUL2016	line revenue center code = 0260	002	line (SRVC_PRVDR_ID) 2224444
000885	line	987654	2	11JUL2016	11JUL2016	0335	002	2224444

In this example, there is a header claim followed by three associated line records for BENE_ID 000001. The line records indicate this beneficiary visited the Federally Qualified Health Center (FQHC, where the type of service code=003), where he is an established patient (procedure code, CPT4=99213) receiving a complete physical, with the collection of lab specimen (CPT code=36415). For the second example (Table 10), the line records indicate BENE_ID 000885 is receiving chemotherapy at a hospital outpatient department. Rather than procedure codes, this provider used revenue center codes to indicate the services provided (revenue center code 0260=IV infusion pump and 0335=Radiology, therapeutic chemotherapy IV).

Revenue codes and HCPCS codes are well-established standardized medical billing nomenclatures, while CMS creates the Medicaid type of service code nomenclature. Type of service valid values are in the TAF research files codebook on the CCW Data Dictionaries webpage.

3. Occurrence Code Files

Occurrence codes provide information when there is a special condition that may affect how the state paid for the service (e.g., date of an injury). States submit occurrence codes for about 46% of 2016 IP records, 6% of 2016 LT records, and 2% of 2016 OT records. The RX file does not have an occurrence code file. A claim may have up to 10 occurrence codes and dates, although only a subset of claims has a record in the occurrence code files. To use this file, link to the header claim file using the CLM_ID. CCW sequentially numbers occurrence records it associates with a header claim using the occurrence code sequence (OCRNC_CD_SEQ).

Occurrence codes may identify a significant event/date relating to an IP or LT stay (e.g., accident, medical condition, insurance-related, service-related, etc.). An occurrence code has a corresponding start date for the occurrence (variable called OCRNC_CD_START_DT), along with an ending date, when applicable (OCRNC_CD_END_DT). It is common for the occurrence start date to precede the dates of service on the claim. Table 11 presents an example of occurrence codes.

Table 11. Example of occurrence code records

BENE_ID	Header or occurrence file	CLM_ID	OCRNC_CD_SEQ	OCRNC_CD	SRVC_BGN_DT (on header)	SRVC_END_DT (on header)
000421	header	123456	n/a		27JUL2016	02AUG2016
000421	occurrence	123456	1	A2	(OCRNC_CD_START_DT) 01FEB2016	
000737	header	789012	n/a		10MAY2016	13MAY2016
000737	occurrence	789012	1	11	(OCRNC_CD_START_DT) 09MAY2016	(OCRNC_CD_END_DT) 09MAY2016
000737	occurrence	789012	2	18	01JAN2014	

In this example, an occurrence code (OCRNC_CD) is associated with the header claim for BENE_ID 000421 (A2), which means that the occurrence date represents the effective date for the insurance policy. The occurrence code start date is February 1, 2016, and the claim admission date is July 27, 2016.

BENE_ID 000737 also had occurrence codes associated with the header claim (Table 11). The first occurrence code (11) refers to the date of the onset of symptoms/illness. The second occurrence code (e.g., where OCRNC_CD_SEQ=2) is 18, which refers to the date of retirement for the patient/beneficiary. Occurrence codes are a standard medical billing nomenclature; reference the TAF research files codebook on the CCW Data Dictionaries webpage.

B. Content and Description of Claims RIFs

Claims for various types of providers and services are in different files. This section describes the contents of each of the four claims files. A complete list of variables is available on the CCW website in the TAF Research files record layout, and the associated codebook provides detailed variable descriptions and values.[19]

1. Inpatient (IP)

This file is for inpatient hospital services. Unlike the MAX files, where CMS summarizes claims into "stays," the IP TAF is at the claim-level, and more than one claim may represent a single hospital stay (e.g., for lengthy hospitalizations). The claim lines typically have the type of service code (TOS_CD) equal to '001'. Type of service codes typical in other settings may also

[19] Reference https://www.ccwdata.org/web/guest/data-dictionaries for the TAF research files codebooks.

be in inpatient lines. Refer to Table 12; the TAF codebook contains an exhaustive list of TOS_CD values. For additional information regarding the TOS_CDs in the data files, refer to the service use information section on the "Resources" page of the DQ Atlas.[20]

Key information on these claims includes diagnosis codes (up to 12) and the corresponding DRG. There is an indicator for each diagnosis code to flag whether the diagnosis was present on admission (POA). The IP file includes up to six major procedures and the associated procedure dates. The procedure code may be a standard ICD-9 or ICD-10 code, or a state-specific procedure code.

Table 12. Examples of the type of service code by file type

TOS code*	TOS description	IP	LT	OT	RX
001	Inpatient hospital services, other than services in an institution for mental diseases	X			
002	Outpatient hospital services			X	
003	Rural health clinic services			X	
006	Technical laboratory services			X	
008	Technical radiological services			X	
009	Nursing facility services for individuals aged 21 or older (other than services in an institution for mental disease)		X		
010	Early and periodic screening and diagnosis and treatment (EPSDT) services			X	
012	Physicians' services			X	
014	Outpatient substance abuse treatment services			X	
016	Home health services — nursing services			X	
028	Clinic services			X	
029	Dental services			X	
030	Physical therapy services (when not provided under home health services)			X	
033	Prescribed drugs				X
034	Over-the-counter medications				X
036	Medical equipment/prosthetic devices			X	X

* A full list of type of service codes is available in the codebook; a TOS code value may appear in more than one file type.

2. Long-Term Care (LT)

This file is for long-term care institutional claims, including nursing facilities, intermediate care facility services for individuals with intellectual disabilities, mental health facility services, and independent (free-standing) psychiatric wings of acute care hospitals.

The lines for these claims often have the type of service code (TOS_CD) equal to '009' and '044'–'048', although other type of service codes may also be present. Note that similar services may be in a non-institutional or home-based setting, and these claims are typically in the OT file.

Key information on the LT claims includes diagnosis codes (up to five), provider information, program and waiver information, and payment information (redacted for managed care encounter records).

[20] Reference "Data Quality and Analytic Resource Downloads. Service Use Information." Type of service – IP, LT, OT, and RX https://www.medicaid.gov/dq-atlas/landing/resources (Accessed 12/09/2024)

3. Other Services (OT)

This file contains information for a broad range of other services; all services not classified strictly as IP, LT, or RX. OT records can represent bills from either facilities or providers. These two broad groups use slightly different formats for their bills so that the information varies a little between claims based on a facility bill compared to claims based on a provider bill. Services in the OT TAF include but not limited to physician services, outpatient hospital services, dental services, other physician services (e.g., chiropractors, podiatrists, psychologists, optometrists, etc.), clinic services, laboratory services, X-ray services, sterilizations, home health services, personal support services, and managed care capitation payments.

Key information on the OT claims includes diagnosis codes (up to two), provider, and payment information (redacted for managed care encounter records). Much of the service information are on the claim lines — including procedures, drugs, etc. Refer to the example in Table 10. Note that this file may include many drugs, such as chemotherapy and other infused drugs or vaccines.

The lines for these claims have a broad range of type of service codes (TOS_CD). Refer to Table 12. A provider may submit a claim with multiple types of services, which means that each claim's line may have a different type of service code. Revenue center codes (REV_CNTR_CD) may appear on the lines for facility claims, whereas procedure codes (LINE_PRCDR_CD) are often on the provider's lines.

4. Pharmacy (RX)

This file contains information for both prescription and over-the-counter drugs covered by Medicaid or CHIP. In general, the file contains only the claims for drugs or other products (e.g., bandages or diabetic test strips) provided by a pharmacy.

The lines for these claims typically have the TOS_CD equal to 033 or 034, although other type of service codes may also be present. Refer to Table 12.

For RX claims, information regarding the drug dispensed is on the line file (e.g., the NDC). Some drugs appear in the OT file, such as infused drugs (including chemotherapies) and vaccines. In addition, drugs that an enrollee receives as part of hospitalization are not in this file. The enrollee pays for these drugs as part of the hospital stay. Similarly, there are rarely drug claims in the LT file since institutional facilities provide most of those services.

Note for some states, there may be a header RX claim without any associated line records. Additional information and tips for working with this file appear in Chapter 5, section A. Linkage Variables and Joining Files.

C. Key Concepts and Variables in Claims RIFs

The TAFs are very large. It is possible that, for a given study, smaller analytical files may be desirable to easily manipulate the data for the intended results. When creating an analytic file, it may be prudent to subset the claims files to retain only the key claims and variables of interest. For example, researchers may wish to only include claims for the study population that meet their criteria (e.g., states, duration of enrollment, type of program/benefit, age group, or who received a particular type of service [TOS_CD]).

The next section highlights some key TAF variables. Additional information regarding all the variables in the file is in the TAF research files codebook on the CCW Data Dictionaries webpage.

1. Type of Claim

Claim type code — this header claim field (CLM_TYPE_CD) identifies the kind of payment for the claim. This field distinguishes between claims that are for Medicaid or Medicaid expansion, S-CHIP, and other types of claims. Table 13 lists values.

Table 13. Claim type codes associated with Medicaid, Medicaid expansion, and S-CHIP

Description	Medicaid or expansion claims	S-CHIP claims	Other claims
FFS claim	1	A	U
Capitated payment	2	B	V
Encounter record	3	C	W
Service tracking*	4	D	X
Supplemental payment	5	E	Y

*Before August 2021, TAF RIFs did not include service tracking claims (i.e., where CLM_TYPE_CD=4, D, X) and claims with a missing CLM_TYPE_CD since they do not represent discrete services provided to patients.

2. Person Claim Indicator

Indicator of a claim for a person — there are instances where a claim does not include a valid person identifier and/or when a claim is not for a discrete service to an individual. Researchers may use the PRSN_CLM_IND variable to distinguish between claims for services for a person (PRSN_CLM_IND=1), versus claims that fit any of four scenarios: 1) missing MSIS_ID, 2) ampersand leading MSIS_ID (&MSIS_ID), 3) service tracking claim, and/or 4) missing claim type code (PRSN_CLM_IND=0).

Following are some scenarios that describe in more detail claims where the PRSN_CLM_IND is 0:

- Although CMS requires states to include an MSIS_ID on every claim, there are rare instances where this ID may be null/missing for data quality reasons.
- Some states pay an insurance premium for a family rather than an individual. The state may include an ampersand (&) in front of an MSIS_ID in these types of claims to indicate a multiple-person premium assistance payment.[21]
- Some states submit data files that include "service tracking claims" that are lump-sum payments to providers or plans (e.g., for drug rebates or disproportionate share hospital payments).[22] Researchers can identify these service tracking claims when the variable called CLM_TYPE_CD=4, D, or X.[23]

Researchers should decide whether to request these non-person claims for their research project since they are not claims for services rendered to a specific enrollee and are not discrete services provided to patients with a corresponding diagnosis or procedure information. If researchers are using the TAF claims data in the CCW VRDC, they

[21] Reference MACBIS memo: https://www.medicaid.gov/medicaid/data-and-systems/macbis/tmsis/tmsis-blog/entry/47574 and corresponding table: https://www.medicaid.gov/medicaid/data-and-systems/downloads/collection-systems/hipp-table-1.pdf (Accessed 12/09/2024)

[22] Additional details regarding service tracking claims are available on the DQ Atlas. For example, reference "Explore by Topic" and "Non-Claims Records": https://www.medicaid.gov/dq-atlas/landing/topics/single/map?topic=g7m85&tafVersionId=17 (Accessed 12/09/2024)

[23] Before August 2021, TAF RIFs did not include service tracking claims.

may filter out these non-person claims with a PRSN_CLM_IND value of 1. Most, but not all, of these non-person claims do NOT have an eligibility record in the DE file; that is, the MISG_ELGBLTY_DATA_IND field on the DE base file is not able to identify these records.

3. Reason for Service

The claims records include fields to help identify the diagnoses, procedures, and/or medications dispensed in an ambulatory setting. Various claims coding systems define these services as shown in Table 14.

- **Diagnosis** — the number of diagnosis code fields varies by file: 1) for IP, there are up to 12 diagnosis codes, 2) for LT, there are up to five, 3) for OT, there are up to two diagnosis codes, and 4) RX has none.
  - On October 1, 2015, CMS, following the Health Insurance Portability and Accountability Act (HIPAA), converted from the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) to the 10th Revision, Clinical Modification (ICD-10-CM) and Procedure Coding System (ICD-10-PCS).
  - Regardless of when a provider submits a claim payment, services that occurred before October 1, 2015, use ICD-9 codes for billing and all services provided on or after October 1, 2015, utilize ICD-10 codes.[24] States may have required providers to submit claims or encounter records with ICD-10 codes on a slightly different schedule. Within the TAF RIFs, there are header file variables (DGNS_VRSN_CD_1–X) that indicate whether a diagnosis code is ICD-9, ICD-10, or something else).

The cutover from ICD-9 to ICD-10 diagnosis codes in the TAF claims did not happen at the same time for all states. In September 2015, the claims files contained the ICD-9 diagnosis code almost 100% of the time in the IP header file, for example. In October 2015, only 1.96% of IP header claims reported ICD-9 codes, and 98.03% report ICD-10 codes. By January 2016, there are still 1.5% of IP header claims where the DGNS_VRSN_CD_1 indicates the use of an ICD-9 code.

Starting with 2020 claims, CMS maps the primary diagnosis code to the AHRQ Clinical Classifications Software Refined (CCSR) diagnosis category code. Reference the variable called DGNS_1_CCSR_CTGRY_CD on the claim header record for IP, LT, and OT files.

- **Procedures** — when examining procedure codes, researchers need to use the procedure code system variables (one for each of the procedure codes) called PRCDR_CD_SYS_1–X (six in IP header file; also, one on the OT line — LINE_PRCDR_CD_SYS). Reference Table 14. Starting with 2020 claims, CMS maps the line-level procedure code in the OT file to the AHRQ Clinical Classifications Software (CCS) procedure category code. Reference the variable called LINE_PRCDR_CCS_CTGRY_CD.
- **DRG** — CMS does not require states to use the Medicare DRG classification system. When a DRG code (variable called DRG_CD) appears on the claim, there is a corresponding DRG code system variable (called DRG_CD_SYS and associated description — DRG_DESC). Reference also Table 14.

[24] CMS. ICD-10 webpage. https://www.cms.gov/medicare/coding/icd10/index.html (Accessed 12/09/2024)

Table 14. Claims coding systems by file type

Type of code	IP	LT	OT	RX
ICD-9/ICD-10-CM diagnosis code*	X	X	X	
Diagnosis-related group (DRG)**	X			
Revenue center code	X	X	X	
Procedure code†	X	X	X	
National Drug Code (NDC)		X	X	X

* May use ICD-9 or ICD-10 codes, depending on the service date.

**DRG may be a state code (not Medicare DRG). Refer to the variable on the claim with the DRG called DRG_CD_SYS.

† May submit procedure codes using various formats, including ICD-9/ICD-10-PCS procedure, Current Procedural Terminology (CPT) code, Healthcare Common Procedure Coding System (HCPCS) code, or maybe a state code; refer to the PRCDR_CD_SYS_1–X variable.

- **Federally Assigned Service Category Code**— CMS derived a federally-assigned service category code (FASC) and added it to the claims during TAF production starting with 2020 claims RIF. CMS created a standard methodology to classify similar types of service use records across all claim files and across both FFS and managed care encounter records. It also allows for consistent identification of non-claim financial transactions, including managed care capitation records, other per-member-per-month payments, and disproportionate share hospital (DSH) payments. Additional details regarding the FASC are available in a DQ Brief *Assigning TAF Records to a Federally Assigned Service Category*.

4. Dates of Service

Service begin and end dates — the SRVC_BGN_DT is the beginning date of service on the claim/encounter record covering services rendered to the beneficiary (e.g., the claim record covers services starting on this date). The service begin date may not always correspond with the first date the beneficiary received services (e.g., it is not always equal to the admission date). This occurs when beneficiaries receive services over a long period, such as lengthy LT stays.

Similarly, the claim ending date of service (SRVC_END_DT) is the last day on the claim billing statement for services rendered to the beneficiary (e.g., the claim record covers services through this date). Note that this date may not always correspond with the last date the beneficiary received services (e.g., it is not always equal to the discharge date). RX claims do not have a service begin or end date on the header claim file; the prescription fill date indicates the date of service (variable called RX_FILL_DT).

For LT and OT claims, the CCW team utilizes the header claim variable SRVC_END_DT for claim selection. In the case of IP and RX claims, the CCW team uses header claim variables discharge date (DSCHRG_DT) and prescription fill date (RX_FILL_DT), respectively.

Some states do not always populate the header claim SRVC_END_DT or DSCHRG_DT. When this happens in the IP, LT, and OT RIFs, the CCW team utilizes a hierarchy of service date fields to populate the SRVC_END_DT (this is also the same hierarchy for TAF selection). For example, if the state doesn't populate the discharge date in an IP claim, the CCW team evaluates the maximum line service end date for the IP claim to populate the header SRVC_END_DT variable. The OT, LT, and IP claim types each have their own hierarchy. The variable SRVC_END_DT_CD identifies the service date variable value used to impute the SRVC_END_DT in the claim header file.

For IP, LT, and OT files, there are also service begin/end dates on the line file (variables called LINE_SRVC_BGN_DT and LINE_SRVC_END_DT). These may be of value if there is user interest in the beginning/ending date of service for the service line representing a home health visit, for example. Data users typically use these for many kinds of OT services analysis. RX line file claims do not have a line service begin or end date; use the prescription fill date variable (variable called RX_FILL_DT, which CCW copies from the RX header and includes in the RX line file).

Admission and discharge dates — IP and LT claims have admission (ADMSN_DT) and discharge (DSCHRG_DT) dates. There are times when a service date on the claim was missing. As mentioned above, in this circumstance, CCW uses the TAF selection algorithm mentioned previously to impute a SRVC_END_DT value to ensure claims are associated with the appropriate date for the data file. The CCW team includes a variable (called the service end date code — SRVC_END_DT_CD) to identify when and how CMS resources imputed the date; additional details are in the codebook.

Procedure date — the IP header claim has dates for each of the six procedure code fields. The date identified in the PRCDR_CD_DT_# applies to the procedure in the corresponding PRCDR_CD_#. Also, the OT line file has a procedure date variable (LINE_PRCDR_CD_DT) that corresponds to the line procedure code (LINE_PRCDR_CD).

5. Provider Information

The TAF claims have fields to accommodate several different types of providers. In addition, states may utilize various types of identification number. States may use a Medicaid-specific provider identification number (variables with the *_PRVDR_ID suffix), and/or CMS provider identifiers (variables with the *_PRVDR_NPI suffix). Refer to specific variable examples in Table 21. There are fields to accommodate provider identifiers on both the claim header and line files. Additionally, there are taxonomy codes in the claims files, which are a set of codes used to identify the type of provider and areas of specialization for health care providers. Starting with 2020 claims, CMS also derived the taxonomy code value by linking NPI to the National Plan and Provider Enumeration System (NPPES) for the billing provider (BLG_PRVDR_NPPES_TXNMY_CD).

Table 15 depicts the various types of provider IDs in the claims TAFs.

Table 15. Provider ID variables in claims

Provider	File	State-assigned identification number	CMS National Provider Identifier (NPI)	Taxonomy code	Provider type	Specialty code
Admitting	IP, LT	ADMTG_PRVDR_ID	ADMTG_PRVDR_NPI	ADMTG_PRVDR_TXNMY_CD	ADMTG_PRVDR_TYPE_CD	ADMTG_PRVDR_SPCLTY_CD
Billing	All	BLG_PRVDR_ID	BLG_PRVDR_NPI	BLG_PRVDR_TXNMY_CD*	BLG_PRVDR_TYPE_CD**	BLG_PRVDR_SPCLTY_CD
Referring	IP, LT, OT	RFRG_PRVDR_ID	RFRG_PRVDR_NPI	RFRG_PRVDR_TXNMY_CD†	RFRG_PRVDR_TYPE_CD	RFRG_PRVDR_SPCLTY_CD
Directing	OT	n/a	DRCTNG_PRVDR_NPI	DRCTNG_PRVDR_TXNMY_CD	n/a	n/a
Supervising	OT	n/a	SPRVSNG_PRVDR_NPI	SPRVSNG_PRVDR_TXNMY_CD	n/a	n/a
Prescribing	RX	PRSCRBNG_PRVDR_ID	PRSCRBNG_PRVDR_NPI	n/a	n/a	n/a
Dispensing	RX	DSPNSNG_PRVDR_ID	DSPNSNG_PRVDR_NPI	n/a	n/a	n/a
Line — servicing	IP, LT, and OT line	SRVC_PRVDR_ID	SRVC_PRVDR_NPI	SRVC_PRVDR_TXNMY_CD‡	SRVC_PRVDR_TYPE_CD	SRVC_PRVDR_SPCLTY_CD
Line — operating	IP line	n/a	OPRTG_PRVDR_NPI	n/a	n/a	n/a

* Starting with the 2020 claims file, for IP, LT, and OT file types CMS also derives the BLG_PRVDR_NPPES_TXNMY_CD.

** The RX file type does not have the billing provider type code on the header record.

† OT is the only file type that has the referring taxonomy code.

‡ Starting with the 2020 claims file, CMS also derives the SRVC_PRVDR_NPPES_TXNMY_CD. This derived NPPES taxonomy is available only in the OT line file.

The OT files contain some health home services. For these records, the OT header file includes information related to the health home provider NPI (HLTH_HOME_PRVDR_NPI) and the health home entity name (HLTH_HOME_ENT_NAME).

Starting in June 2021, CMS disseminates the Annual Provider (APR) file. This file includes many details regarding Medicaid/CHIP providers. The APR RIF section describes the file in detail.

Starting with 2020 claims, CMS derived a federally assigned service category code to classify similar types of service use records across all claim files and across both fee-for-service and managed care encounter records. Reference the FED_SRVC_CTGRY_CD field on the IP, LT, OT, and RX header claims.

6. Billing and Payment Information

CMS requires the CCW to employ claim-level redaction of managed care payment fields when the payer is a managed care plan (refer to Appendix C — Managed Care Redacted Payment Fields). Researchers can identify managed care claims using the field CLM_TYPE_CD and the values 3, C, or W. CMS considers all other claim type values as non-managed care and does not apply redaction. For the redacted claims records, the CCW team populates all the other RIF fields (i.e., applies no additional suppression for the managed care records).

The claims RIFs have several types of variables related to payment:

- The amount Medicaid paid for the service
- The amount the beneficiary had to pay for the service (in the form of deductibles or copayments)
- The amount paid by Medicare or a third party for the service.

States may also make additional payments to providers or managed care plans, such as disproportionate share hospital (DSH) payments or other lump-sum payments. These claims do not represent claims for services provided to individual beneficiaries. Refer also to Chapter 3, section C. 2. Person Claim Indicator.

7. Program/Plan/Waiver Types

The header claims (for all four claim types) have fields to capture other key applicable information, including:

- Program type code (PGM_TYPE_CD) — such as Early and Periodic Screening and Diagnosis and Treatment (EPSDT) or Money Follows the Person (MFP)
- Managed Care Plan identification number (MC_PLAN_ID)
- Waiver type code (WVR_TYPE_CD), which may indicate various 1915 and 1115 waivers or state plan options, and the associated waiver identification number (WVR_ID)
- Indicator insured is “Covered by another plan” (not Medicare or Medicaid) (OTHR_INSRNC_IND)
- 1115(A) Demonstration participation indicator (SECT_1115A_DEMO_IND) for CMMI demonstrations

States submit their Medicaid expenditure reports to CMS using a standard CMS-64 form,[25] and they submit CHIP expenditures on the CMS-21 form. The claim line files for each of the four claim types include fields to categorize the service in terms of the CMS-64 form and CMS-21 form:

- CMS-64 Form Category of Service for the Paid Claim (XIX_SRVC_CTGRY_CD)
- CMS-21 Form Category of Service for the Paid Claim (XXI_SRVC_CTGRY_CD)
- CMS-64 Form Code for Federal Reimbursement (CMS_64_FED_CTGRY_CD)

[25] Additional information regarding these reports is on the Medicaid.gov website https://www.medicaid.gov/medicaid/financial-management/state-expenditure-reporting-medicaid-chip/index.html (Accessed 12/09/2024)

Chapter 4 — Other Annual Files

A. Annual Managed Care Plan (APL) File RIF

CMS creates the Annual Managed Care Plan (APL) file from state-reported monthly managed care plan data. The file contains information about each Medicaid and CHIP managed care plan/entity[26] that CMS authorized to operate in the state during the calendar year. The information contained in the APL includes managed care name, type of managed care plan (e.g., comprehensive managed care, Primary Care Case Management (PCCM), Accountable Care Organizations (ACO), Health Home), the various locations of the managed care plan, the various service areas in which the managed care plan operates, operating authorities, and eligibility groups authorized to enroll in each plan.

The APL base file contains one record for each unique combination of submitting state entity code (i.e., the SUBMTG_STATE_CD) and state plan identification number (MC_PLAN_ID). The CCW team uniquely identifies each record with the CCW_APL_LINK_KEY. The file is available for each state that has managed care plans, eligibility, and claims RIFs for the year.

The APL includes five files, a base file and four supplemental files.

1. The **base** file contains one record for each unique managed care plan identifier (MC_PLAN_ID within a state; also denoted by a unique CCW_APL_LINK_KEY). This core file includes basic descriptive characteristics of each plan. Also included are indicators for whether the managed care plan has a record in one or more of the supplemental files (refer to Table 16).
2. The **location s**upplemental file contains records for every service location address associated with the plan (i.e., each unique CCW_APL_LINK_KEY). The file includes a flag for each month of the year to indicate when the location was active during the year.
3. The **service area** supplemental file contains records for all the geographic service areas each plan (i.e., each CCW_APL_LINK_KEY) covers. There is one record per service area name (MC_SAREA_NAME) for each plan.
4. The **population enrolled** supplemental file contains records for every eligibility group the state is authorized to enroll in each plan. There is one record per managed care population (MC_PLAN_POP) for each plan.
5. The **operating authority** supplemental file documents the operating authority and/or waiver identifiers under which the managed care plan is authorized to operate. It contains one record for every unique combination of managed care plan (MC_PLAN_ID), submitting state (SUBMTG_STATE_CD), waiver identifier (WVR_ID), and operating authority (OPRTG_AUTHRTY) that appears in any month.

Not all plans identified in the APL base file are in the supplemental files. To determine if a plan has a record in a supplemental file, refer to the supplemental indicator variables found in the APL base file (listed in Table 16). These four fields are binary indicators, where 1 = the plan has a record in the supplemental file. Further, most, but not all plans in the APL base and supplemental files are in the TAF beneficiary (i.e., the DE RIF) or claims files for the corresponding year.

[26] Although a managed care entity includes managed care plans, the file may also include arrangements that are not traditional managed care plans such as primary care case management programs. MC_PLAN_ID represents each managed care entity.

Table 16. APL base file — variables indicating records in APL supplemental file

Description	SAS variable name
Annual Managed Care Plan record in supplemental service address location file	SPLMTL_SRVC_ADDR_LCTN
Annual Managed Care Plan record in supplemental service area file	SPLMTL_SAREA
Annual Managed Care Plan record in supplemental enrolled population file	SPLMTL_POP_ENRLMT
Annual Managed Care Plan record in supplemental operating authority file	SPLMTL_OPRTG_AUTHRTY

Researchers join the APL base file to the supplemental file records by linking on the CCW_APL_LINK_KEY. *Note that there may be a one-to-many or one-to-none relationship between records in the APL base file and the APL supplemental files.*

Researchers may also link information from the APL file with the plans in the DE file (e.g., the Annual DE Managed Care Enrollment supplemental file includes MC_PLAN_ID_01_01 thru MC_PLAN_ID_16_12; there may be up to 16 plan identifiers per month for 12 months, so that a beneficiary theoretically could have as many as 192 occurrences of managed care plans in a given eligibility year). The first iteration of the managed care plan ID in the DE file (i.e., the MC_PLAN_ID_01_*MM*) for each of the calendar 12 months is the most well populated; however, the MC_PLAN _IDs found anywhere in the DE file links to the APL. Researchers may also link the APL file to the MC_PLAN_ID field in the header files of all claim file types.

Table 17 demonstrates a proposed linkage between APL and the DE file. As an example, if researchers are interested in knowing more about a beneficiary population enrolled in the MC plans. Researchers could start with creating a finder file from the TAF DE RIF enrollment data, where they obtain the MC_PLAN_IDs from the DE Managed Care Enrollment supplemental file. The example uses the first month's first occurrence of the MC_PLAN_ID (variable name MC_PLAN_ID_01_01), along with the associated MC plan type code (MC_PLAN_TYPE_01_01). This example uses only the first MC_PLAN_ID for simplicity; for a complete analysis of the beneficiary population of an MC plan, researchers would include all iterations of MC_PLAN_ID from the DE file. Researchers would then join to the APL base file by the MC_PLAN_ID and SUBMTG_STATE_CD to determine the plan name.

Table 17. Examples of joining the APL base file with the DE file.

SUBMTG_STATE_CD (DE Managed Care supplemental)	**MC_PLAN_ID_01_01** (DE Managed Care supplemental)	**MC_PLAN_TYPE_CD_01_01** (DE Managed Care supplemental)	**SUBMTG_STATE_CD** (APL base)	**MC_PLAN_ID** (APL base)	**CCW_APL_LINK_KEY** (APL base)	**MC_PLAN_NAME** (APL base)
12	1277889910	01	12	1277889910	98765	MIDWEST BEST PLAN
47	000456789	14	47	000456789	223344	HEALTHY SMILE, INC
22	031313131	17	22	031313131	555888	PACE EVERYWHERE

In this example, researchers can identify the plan names for plans previously identified in the enrollment data. It illustrates using the identifiers for three plans from the DE managed care file (the SUBMTG_STATE_CD and the MC_PLAN_ID fields) and retains the plan type code for these plans. The example joins this information to the APL base file. The first plan is MIDWEST BEST PLAN, which is a comprehensive managed care plan (where MC_PLAN_TYPE_CD=01). The second plan is HEALTHY SMILE, INC, which is a dental prepaid ambulatory health plan (PAHP; where MC_PLAN_TYPE_CD=14). The third plan is PACE EVERYWHERE, which is a Program of All-Inclusive Care for the Elderly (PACE) plan (where MC_PLAN_TYPE_CD=17).

If researchers need additional information from APL supplemental files, they join the APL base file to the supplemental files based on the CCW_APL_LINK_KEY.

B. Annual Provider (APR) File RIF

CMS creates the Annual Provider (APR) file from the monthly provider data submitted by the states. The file contains information about the providers that are eligible to render services to Medicaid and CHIP beneficiaries for any month of the year. Providers may be facilities, groups, or individual practitioners as a result the information available for a provider varies based on the provider type. The information contained in the APR file includes provider name, provider type, the provider group (if applicable), facility-specific information (if applicable, e.g., bed count), and the various Medicaid/CHIP programs/waivers/demonstrations, provider taxonomies, affiliated groups, locations, licenses, and identifiers (state-specific as well as national provider identifiers) associated with the provider.

The APR base file contains one record for each unique combination of submitting state entity code (i.e., the SUBMTG_STATE_CD) and state-specific provider identifier (SUBMTG_STATE_PRVDR_ID) for each provider active at any time during the calendar year. The CCW_APR_LINK_KEY uniquely identifies each of these state-entity-specific provider records. The file is available for states with eligibility and claims RIFs for the year.

The APR is comprised of nine files: a base file and eight supplemental files. All nine files have monthly indicators to identify the portion of the calendar year to which the information applies.

1. The **base** file contains one record for each state-specific provider identifier (SUBMTG_STATE_PRVDR_ID) within a state, also denoted by a unique CCW_APR_LINK_KEY. This is the primary file that includes a set of basic data elements that capture characteristics for each Medicaid and/or CHIP provider. The file includes variables that reflect the attributes of individual, group, or facility providers. Some examples of attributes include Medicaid and/or CHIP program enrollment indicators, ownership information, profit status, provider legal and/or tax names, whether the provider was accepting new patients and whether the facility was a teaching facility. Also included are indicators for whether the provider has a record in one or more of the supplemental files (refer to Table 18).
2. The **taxonomy** supplemental file contains records that describe a provider's classification and includes four provider classification schemes: 1) taxonomy, 2) provider specialty, 3) provider type, and 4) provider authorized categories of service. There may be more than one record per state-specific provider ID (i.e., per CCW_APR_LINK_KEY).
3. The **enrollment** supplemental file contains information about the provider's enrollment status and months in the Medicaid and/or CHIP program. There may be more than one record per state-specific provider ID. (i.e., per CCW_APR_LINK_KEY).
4. The **affiliated groups** supplemental file contains information about the unique, state-assigned identification number for the group or subpart with which the provider is associated during the calendar year. There may be more than one record per state-specific provider ID (i.e., per CCW_APR_LINK_KEY).
5. The **affiliated programs** supplemental file contains state-submitted health plan ID numbers and three categories of programs with which a provider may be affiliated during the calendar year: 1) Medicaid waiver, 2) health home, or 3) other. There may be more than one record per state-specific provider ID (i.e., per CCW_APR_LINK_KEY).
6. The **location** supplemental file contains a provider's location and address information for three types of locations: 1) billing location, 2) practice location, and 3) service location. The remaining three supplemental files (Licensing, Identifiers and Bed Type) contain the provider location identifier (PRVDR_LCTN_ID) associated with this record (i.e., for each distinct CCW_APR_LINK_KEY).
7. The **licensing** supplemental file contains licensing or accreditation information for a particular provider location (as indicated by the PRVDR_LCTN_ID field in the location supplemental file) for the provider (i.e., the

CCW_APR_LINK_KEY) for the calendar year. The file contains five types of licenses and accreditations: 1) state/county/municipality license, 2) Drug Enforcement Administration (DEA) license, 3) professional society accreditation, 4) Clinical Laboratory Improvement Amendment (CLIA) accreditation, and 5) other.

8. The **identifiers** supplemental file contains provider identifiers for a particular provider location (as indicated by the PRVDR_LCTN_ID field in the location supplemental file) for the provider (I.e., the CCW_APR_LINK_KEY) for the calendar year. The file contains six[27] types of identifiers: 1) state-specific Medicaid ID, 2) National Provider Identifier (NPI), 3) Medicare ID, 4) National Council for Prescription Drug Programs (NCPDP) ID, 5) federal tax ID, and 6) state tax ID.
9. The **bed type** supplemental file contains bed type variables for a facility provider location (as indicated by the PRVDR_LCTN_ID field in the location supplemental file) for the provider (I.e., the CCW_APR_LINK_KEY) for the calendar year. The file includes information for four bed types: 1) intermediate care facility for individuals with intellectual disabilities, 2) inpatient, 3) nursing facility, and 4) Title 18 skilled nursing facility.

Not all providers in the APR base file are in the supplemental files. To determine if a provider has a record in a supplemental file, refer to the APR base indicator variables (listed below in Table 18). These eight fields are binary indicators, where a value of 1 indicates that the plan has a record in the corresponding supplemental file. Further, most, but not all providers in the APR base and supplements are found in the TAF RIF claims data for the corresponding service year. This is because not every provider in the APR file provides services to Medicaid or CHIP beneficiaires during the year.

Table 18. APR base file — variables indicating records in APR supplemental file

Description	SAS variable name
Annual Provider record in supplemental taxonomy file	SPLMTL_TXNMY
Annual Provider record in supplemental enrollment file	SPLMTL_ENRLMT
Annual Provider record in supplemental affiliated groups file	SPLMTL_AFLTD_GRP
Annual Provider record in supplemental affiliated programs file	SPLMTL_AFLTD_PGM
Annual Provider record in supplemental provider address location file*	SPLMTL_ADDR_LCTN
Annual Provider record in supplemental provider license and accreditation file*	SPLMTL_LCNS_ACRDTN
Annual Provider record in supplemental identifier file*	SPLMTL_PRVDR_ID
Annual Provider record in supplemental bed type file*	SPLMTL_BED_TYPE

*These supplemental files may have more than one record for a provider location (PRVDR_LCNT_ID) related to a provider and submitting state entity code (represented by the CCW_APR_LINK_KEY).

Researchers may join the APR base file to the supplemental file records by linking on the CCW_APR_LINK_KEY. Note that there may be more than one APR supplemental record in any given supplemental file for a CCW_APR_LINK_KEY found in the base file (i.e., a one-to-many relationship). Information in four of the supplemental files relates to not only a provider but also to a particular provider location. The provider location ID (PRVDR_LCTN_ID) in the location file indicates details regarding the physical geographic location; the license, identifiers, and bed type supplemental files all have records for each PRVDR_LCTN_ID. That means that these four files do not have a 1:1 join to the base file since a provider may have more than one location.

[27] States may submit two additional types of identifiers: 7) Social Security number, and 8) other. However, CMS suppresses these in the APR due to privacy concerns.

Information from the APR may be associated with the various provider IDs that researchers can find on the claims (both header and line files). Within the TAF RIF claims files, there are fields for billing providers, servicing providers, referring providers, dispensing providers, admitting providers, lock-in providers, and many more (refer to Table 19 below). There are two key provider identifier types used throughout the TAF. The first is the state-assigned unique identifier for the provider entity, and the second is the NPI. These IDs generally do not appear in the same fields. For example, the claims header TAF contains two billing provider identifier variables: the state-assigned billing identifier as reported in the variable BLG_PRVDR_ID and the NPI billing identifier as reported the variable BLG_PRVDR_NPI. When researchers are joining claims data to the APR, they must be cognizant of associating the correct type of provider identifiers in the two files because the variables do not always have the same names in the claims data versus the APR. For example, researchers would join the claim BLG_PRVDR_ID to the APR variable SUBMTG_STATE_PRVDR_ID.

Researchers may also link information from the APR file with the providers in the DE file. Seven variables from the DE supplemental files are provider IDs (refer to Table 19 below): LCKIN_PRVDR_ID_1, LCKIN_PRVDR_ID_2, LCKIN_PRVDR_ID_3, LTSS_PRVDR_ID_1, LTSS_PRVDR_ID_2, and LTSS_PRVDR_ID_3 in the DE Disability and Need supplemental file, and HLTH_HOME_PRVDR_ID in the DE Health Home and State Plan Option supplemental file.

Although APR provider IDs may link to some of the provider IDs in the DE file, DE file provider IDs are generally sparsely populated. The CCW team recommends creating a finder file from the DE file containing provider ID(s) of interest before joining with APR data. Note the provider IDs from claims/DE RIFs have several possible field names (e.g., BLG_PRVDR_ID, RFRG_PRVDR_ID, SRVC_PRVDR_ID LCKIN_PRVDR_ID, HLTH_HOME_PRVDR_ID etc.) and in the APR a key variable is the SUBMTG_STATE_PRVDR_ID. Depending on researcher's software, they may have to rename the provider IDs to obtain a successful join. Refer to Table 19.

Table 19. Variables in the claims and ADE files containing provider IDs

Claim header files	Claim line files	ADE disability and need supplemental file	ADE health home and state plan supplemental file
BLG_PRVDR_ID, RFRG_PRVDR_ID, ADMTG_PRVDR_ID, PRSCRBNG_PRVDR_ID, and DSPNSNG_PRVDR_ID	SRVC_PRVDR_ID	LCKIN_PRVDR_ID_1, LCKIN_PRVDR_ID_2, LCKIN_PRVDR_ID_3, LTSS_PRVDR_ID_1, LTSS_PRVDR_ID_2, and LTSS_PRVDR_ID_3	HLTH_HOME_PRVDR_ID

If the researcher's objective is to identify providers who served beneficiaries, then the CCW team recommends linking the provider IDs from the claims files (either the header or line files) to the APR provider ID. For this linkage, researchers can use any of the provider IDs on the claims files. However, the CCW team has found that the billing provider ID (BLG_PRVDR_ID) is the ID type most frequently populated.

Table 20 illustrates a proposed linkage in using a small number of variables from the IP claim header file and joining a small number of variables from the APR base and location files.

As an example, if researchers would like to know the name of the provider providing inpatient services as well as the list of all associated locations (which requires use of the APR location supplemental file). The CCW team recommends that researchers start with a finder file from the TAF RIF claims data — obtaining the desired provider IDs from the file — and then join to the APR base file to determine the provider tax name (APR variable called PRVDR_TAX_NAME). To identify the provider location, use the CCW_APR_LINK_KEY from the APR base file to link to the APR location supplemental file to determine the city (PRVDR_CITY) and provider state (PRVDR_STATE) for the provider location (PRVDR_LCTN_ID).

Table 20. Examples of joining provider ID from the claim to the APR base file and the APR location supplemental file

CLM_ID	BLG_PRVDR_ID (claim)	SUBMTG_STATE_CD (claim)	SUBMTG_STATE_CD (APR – all files)	SUBMTG_STATE_PRVDR_ID (APR – all files)	CCW_APR_LINK_KEY (APR – all files)	PRVDR_TAX_NAME (APR base)	PRVDR_ADDR_LINE_1 (APR location supplemental)	PRVDR_CITY (APR location supplemental)	PRCTC_LCTN_IND (APR location supplemental)	PRVDR_STATE (APR location supplemental)	PRVDR_LCTN_ID (APR location supplemental)
1234	606060606060A	20	20	606060606060A	CD987654	BEST HOSPITAL NORTHEAST	242 MEDICAL DRIVE	JACKSON	1	20	00123
55555	357894523	17	17	357894523	A5A7258974	BIG UNIVERSITY HOSPITAL	100 MAIN ST	NEWPORT	1	17	0045
			17	357894523	A5A7258974	BIG UNIVERSITY HOSPITAL	5112 ADAMS AVE	NEWPORT	1	17	0046
33221	484946	12	12	484946	2255223344	CHILDREN'S MEDICAL	21054 PRESIDENT BLVD	DOVER	0	12	0992

The provider on the first claim is the BEST HOSPITAL NORTHEAST, which has only one location record (i.e., the APR base record links to only one APR location supplemental file record). The second claim links to two location records for the BIG UNIVERSITY HOSPITAL in the APR location supplemental file (i.e., there is a row for each PRVDR_LCTN_ID). Both locations are practice locations at different addresses within the same city. Note that some locations may only be active for part of the year, and there are monthly indicators in the location file to show whether the location was active during that month (variables PRVDR_LCTN_ACTV_IND_01-PRVDR_LCTN_ACTV_IND_12). The final example in the table for CHILDREN'S MEDICAL has only one billing location. As a result, the practice location binary indicator (PRCTC_LCTN_IND) is 0. Researchers could include the other two provider location type indicator variables, BLG_LCTN_IND and SRVC_LCTN_IND, in their results output for additional clarity.

There are many types of information available in the APR base and supplemental files. These conceptual examples simply illustrate some key information and how to join various files.

Chapter 5 — Using TAF Research Files

The CCW team encourages researchers to become familiar with the various CMS DQ Atlas topic assessments and resource documents[28] before working with the data files. For MAX users, there are major differences between the TAFs and the MAX files. This section of the document provides helpful hints and considerations for working with the TAF research files.

A. Linkage Variables and Joining Files

As the document discusses earlier, more than one MSIS_ID and SUBMTG_STATE_CD (which can roll up into states using the STATE_CD) may associate with the same BENE_ID if the enrollee is in Medicaid or CHIP in more than one state. In fact, the BENE_ID may also associate with more than one MSIS_ID in the same state if the eligible has multiple non-contiguous spells of Medicaid or CHIP enrollment. As such, when joining the DE enrollment files together (e.g., the DE base to any of the DE supplemental files), use the combination of MSIS_ID and SUBMTG_STATE_CD (or STATE_CD) to ensure that researchers are associating the correct supplemental records to the selected base record. A visual depiction of the rationale for using MSIS_ID and STATE_CD for this linkage is in Table 21.

Table 21. Examples of joining DE base file with supplemental file(s)

BENE_ID (DE base)	**MSIS_ID** (DE base)*	**STATE_CD** (DE base)	**BENE_ID** (Supplemental file)	**MSIS_ID** (Supplemental file)	**STATE_CD** (Supplemental file)	**WVR_1915C_MOS** (Supplemental file)**
000123	000ABCDE1357	NH	000123	000ABCDE1357	NH	9
000123	0762468WXYZ	NY	000123	0762468WXYZ	NY	1
000123	R99134ST8877	MA	000123	R99134ST8877	MA	0

* Within the data files, MSIS_ID is 32 alphanumeric characters, shortened in this example for brevity.
** This example assumes researchers are using the waiver supplemental file.

In this example, Medicaid or CHIP enrolls the same person (BENE_ID 000123) in three states during the year. To ensure researchers are joining the base enrollment record for that person to the associated DE supplemental file records (e.g., waiver program enrollment), they must use the MSIS_ID and STATE_CD. In this example, since there is more than one DE base record for the BENE_ID, it is impossible to accurately join these files using the BENE_ID — researchers risk joining the supplemental DE records to the wrong base records.

If researchers are using multiple years of data[29] and wish to identify the same person longitudinally, the CCW team

[28]DQ Atlas https://www.medicaid.gov/dq-atlas/welcome (Accessed 12/09/2024)
[29]CCW uses various agency beneficiary identifiers (such as HICN and SSN) in its CCW BENE_ID assignment algorithm, including a Medicaid-specific agency beneficiary identifier created by the TAF source data maintainer. Prior to the creation of the 2020 mature research files (released in October 2022 and referred to as 2020 Release 1 or R1 TAFs), CMS made updates to the Medicaid beneficiary identifier logic that they applied to all Medicaid and CHIP beneficiaries historically. This update impacted the CCW BENE_ID assignment for the 2020 R1 TAFs, resulting in the assignment of a new CCW BENE_ID to some Medicaid beneficiaries. This means that a beneficiary in the 2020 R1 TAFs may have a different CCW BENE_ID compared to their ID in the 2020 preliminary, 2019, or other prior versions of the TAFs. Researchers conducting longitudinal or time-series analysis involving the 2020 R1 TAFs and earlier service years should use the T-MSIS 2020 BENE_ID bridge file to map the 2020 R1 CCW BENE_ID to the

recommends using the BENE_ID[30] to identify the same person across states (e.g., if people move during the year). For longitudinal studies within a state, researchers may wish to use the MSIS_ID and state code. These linkage variables (BENE_ID, MSIS_ID and STATE_CD) appear in all TAF research files.

To link the header claim with associated lines, use the unique claim ID (CLM_ID), present on all header and line files. The system numbers the line file detail records associated with the claim sequentially (variable called line number LINE_NUM). Table 10 shows the linkage between the header claim and associated lines. Researchers can also join the header claim to the occurrence code file (available/populated for a subset of IP, LT, and OT claims) with the CLM_ID. Remember that the associated occurrence code records are in numerically sequential order (OCRNC_CD_SEQ).

Unlike the other three claim types, it is common to see RX header claims with no associated lines because the header claim alone is sufficient to describe a drug prescription/dispensing. There may also be RX claims with more than one associated line; this generally means the medication dispensed was a compound drug (e.g., more than one active ingredient/NDC).

The data files can be very large; therefore, the CCW team recommends creating an analytic file with a subset of variables to make the files more manageable. The team recommends that the users do NOT include all variables on the header and line files when combining the data.

B. Summarizing Data

Also recommended, is using the BENE_ID, when available, to identify all claims for the beneficiary for the year. If there is not a BENE_ID for the record, then use the combination of MSIS_ID and SUBMTG_STATE_CD. Similarly, when summarizing enrollment, although researchers join the DE base and supplemental files using the MSIS_ID and SUBMTG_STATE_CD (or STATE_CD), the CCW team recommends summarizing the data using the BENE_ID. Some scenarios are below, and Table 22 has an illustrative example:

BENE_IDs in the earlier versions of TAF data. For additional information about this file, refer to the *FACT SHEET: 2020 T-MSIS BENE_ID Bridge File* on the CCW secure website.

[30]Another impact of the upstream Medicaid beneficiary identifier logic changes is the algorithm assigned all instances of invalid Social Security numbers (SSN) (e.g., all zeros, eight zeros, or all eights) with the same date of birth (DOB) and sex the same Medicaid-specific agency beneficiary identifier. Invalid SSNs are a common data anomaly in state-reported T-MSIS data. The assignment of the same Medicaid beneficiary identifier to different Medicaid beneficiaries resulted, in limited cases, in the assignment of the same CCW BENE_ID to these different beneficiaries. This phenomenon can happen within a single state or across many states. CCW refers to these instances as "BENE_ID clusters." This issue is most common for new enrollees with invalid SSNs, including newborns, appearing in the upstream TAF for the first time in 2022 or later, although BENE_ID clusters also exist for earlier service years. Starting with the 2020 R1 and later TAF researcher files, if researchers identify TAF eligibility records where there are BENE_IDs associated to a large number of MSIS_ID and SUBMTG_STATE_CD combinations, it is prudent to assume these are not the same person and to use the MSIS_ID/SUBMTG_STATE_CD as the person identifiers for researcher's analysis. Researchers should also use MSIS_ID/SUBMTG_STATE_CD to join different TAF file types for the year (e.g., use the same variables for ADE and claims file joins and other analytics). Remember that there will always be BENE_IDs with more than one MSIS_ID/SUBMTG_STATE_CD if Medicaid/CHIP enrolled the person in more than one state during the year. It is also legitimate for a person to have multiple MSIS IDs within the same state, in the case of disenrollment/re-enrollments or state large-scale re-enumeration of their MSIS_ID nomenclature. The algorithm assigns large numbers of ADE records the same BENE_ID to different beneficiaries due to the clustering issue.

- **Summarizing services across states** — if looking at all the services received by a beneficiary during the year regardless of the state (in cases where beneficiaries were in more than one state), use the BENE_ID. If researchers are interested in services for a beneficiary in a single state, use BENE_ID within a STATE_CD.
- **De-duplicating enrollment within a state** — if researchers want to identify the precise number of beneficiaries in a state, use the BENE_ID (since someone might have more than one MSIS_ID within a state).
- **Enrollment across states** — if researchers are interested in the number of months of Medicaid enrollment across years for a beneficiary, use the BENE_ID.

Table 22. Examples of summarizing data for beneficiaries

BENE_ID	MSIS_ID*	STATE_CD	Months of Medicaid	Number inpatient claims
000123	000ABCDE1357	NH	3	1
000123	0762468WXYZ	NY	4	1
000123	R99134ST8877	MA	5	2
Total			**12**	**4**

* Within the data files, MSIS_ID is 32 alphanumeric characters, shortened in this example for brevity.

The example in Table 22 shows BENE_ID 000123 in Medicaid in three different states during the year. Accordingly, there are three different MSIS_ID and STATE_CD combinations. Looking only within a single state, one would obtain an inaccurate ascertainment of Medicaid enrollment and services for the beneficiary. The BENE_ID allows researchers to look across states for a more complete view of enrollment and services.

C. Considerations for Selecting States for Analysis

States have flexibility with program implementation and may expand the scope of benefits and how they deliver services (e.g., using waivers, etc.). Thus, it is important to understand how a specific state's program operates (e.g., managed care or FFS) and the range of benefits for the state(s) of interest. Various sources may help understand state Medicaid and CHIP programs, including the Medicaid.gov website.[31] Additionally, data quality varies by state, topic, and year. The data quality resources featured on the DQ Atlas website can provide insight into these issues.

States also have flexibility regarding when to convert from MSIS to T-MSIS, although all states converted to T-MSIS by October 2015.

1. State Cutover to T-MSIS

The CCW team did not create TAF research files for all states for 2014 or 2015 (refer to Table 1). A state needed a full calendar year of T-MSIS data for us to include in the TAFs. The states with TAFs, by year, are in Appendix B — State Cutover to T-MSIS and Availability of MAX and TAF Research Files. If there is not a TAF for the state(s) of interest, a MAX file should be available in most cases.

2. State CHIP and Eligibility Groups

The claims TAFs include more services than provided through the Medicaid and CHIP programs. States may have claims for services provided to other (or expansion) populations, depending on state-specific enrollment criteria.

[31] A resource for national and state Medicaid information is Medicaid.gov. https://www.medicaid.gov/medicaid/index.html (Accessed 12/09/2024)

Reference Medicaid.gov for State Overviews.

Within the DE TAF, researchers can identify populations enrolled in Medicaid, CHIP, expansion populations, or other state programs. Refer to Data Tip 1. Furthermore, within the claims RIFs, researchers can identify claims for Medicaid, CHIP, or other services using the claim type code (variable called CLM_TYPE_CD).

3. Known Issue with Illinois Claims

States have varying data quality levels (refer to the DQ Atlas website for state data quality information). Those wishing to use claims from Illinois, need to perform special processing. Refer to Data Tip 2.

Data Tip 2. Illinois — incremental claims

> Unlike the claims files from other states, Illinois does not limit its' claims files to final action records. When adjudicating claims, Illinois Medicaid processes incremental claims, rather than "replacement" claims. This means users must consider adjustments to claims (e.g., credits and debits) in conjunction with the original claim. For Illinois, they include all transactional claims/encounter records in the RIF. Additional information and guidance are available at the DQ Atlas website in the document *TAF Technical Documentation: How to Use Illinois Claims Data* — found on the Resources page, Data Quality and Analytic Resource Downloads area of the page, within the "Technical documentation for using TAF RIF data" section.

D. Cautions in Working with TAF RIFs

1. Claims without Enrollment

There are some Medicaid and CHIP claims for which there is no corresponding enrollment record for the beneficiary receiving the service. As a result, CMS has created a "dummy" enrollment record for these beneficiaries; however, there is no Medicaid/CHIP enrollment information in the dummy record. The "dummy" records only include an MSIS-ID and state code. No additional information (birth date, sex, or enrollment dates) is on the record. Researchers can identify the dummy enrollment records by a missing eligibility data indicator (MISG_ELGBLTY_DATA_IND) value of 1.

As discussed in Chapter 3, section C. 1. Type of Claim, RIFs now include service tracking payment records (e.g., DSH payments and bundled capitated payments). In some instances, notably if the MSIS_ID for the service tracking payment record does not have a leading ampersand (&), corresponding "dummy" enrollment records may appear in the DE base file for these records. If researchers are counting beneficiaries/enrollees in a state, the CCW team recommends removing the dummy enrollment records from the analysis. Do not simply count MSIS_IDs for each state because this results in an overcount of beneficiaries in the state.

Use the CLM_ID to identify every claim record in the data file, regardless of enrollment status.

2. Date Variable Transformations

T-MSIS source data submitted by states may occasionally include date values that are implausibly early. The CCW team recodes the T-MSIS date values before 1/1/1600 as 12/31/1599, due to CCW software defaults. The exception is for 1/1/1000, which is set to null in the RIFs.

E. TAF Data Quality Assessments and Other Supplemental Materials

CMS designs the TAF Data Quality Assessments to support TAF RIF data users. To evaluate the usability of the data, CMS analyzes various topics and variable-specific questions by state. The CCW team encourages researchers to use this information to design analyses and to inform analytic work. There are dozens of Data Quality assessments and other supplements,[32] organized into "themes" (e.g., enrollment, managed care, provider information, etc.). In addition to the issue assessments, there are state and topic "snapshots" designed to provide a quick, high-level evaluation of the quality of the state's data by topic area.

Chapter 6 — Receiving CCW Data

The CCW team provides the TAFs to academic researchers and certain government agencies to conduct approved research studies under a CMS Data Use Agreement (DUA). The CCW Medicaid and CHIP data contain identifiable information and are subject to the Privacy Act and other federal government rules and regulations (reference the ResDAC website for information on requesting Medicare and CHIP data http://www.resdac.org/). Anyone seeking approval for access to the Medicaid and CHIP data must ensure that the data can reliably support their research proposal.

External researchers have two options for accessing the data files — they may access them directly from the CCW Virtual Research Data Center (VRDC), or have their data shipped. For CCW VRDC users, the CCW team initiates their access upon receipt of the approved DUA and payment.

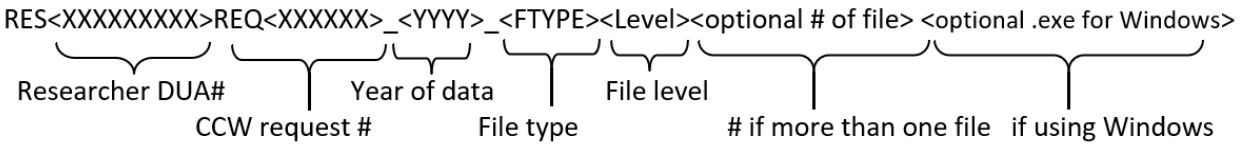
For data the CCW team ships, once the researcher's DUA is in place and CCW has received the payment for the files, data files are prepared and shipped to the requestor on either a USB external hard drive, DVD, or flash drive.

A. Format

The CCW team organizes and delivers files to researchers in the following format. There are several folders, each of which contains multiple files. The folders are:

- XXXX (folder with the CCW data request number)
- Extract file documentation

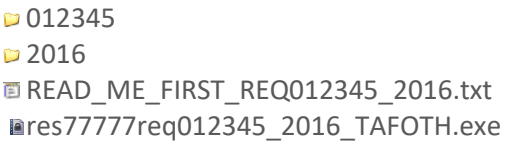
There may be additional folders if researchers have requested additional types of data. Researcher's data files are located within the folder with their CCW data request number (reference Chapter 6, section B. CCW Output Package — Data File Folder). There is a separate sub-folder for each year of data the researchers requested. The naming convention for data files is as follows:



For example, if the researchers DUA number was 77777, their CCW request number was 012345, and they obtained

[32] https://www.medicaid.gov/dq-atlas/welcome (Accessed 12/09/2024)

2016 T-MSIS OT data, their folders and data files would look like this:



Reference Table 23, Table 24, Table 25, and Table 26 to identify the types of claims (or other data files) as follows.

Table 23. T-MSIS RIF enrollment filenames

Enrollment file type	Base/Supplemental files	File level
DE base file	demog_elig_base	Base
Eligibility dates	demog_elig_dates	Supplement
Disability and need	demog_elig_disability	Supplement
Health home and State Plan Options (SPO)	demog_elig_hh_spo	Supplement
Managed care enrollment	demog_elig_mngd_care	Supplement
Money Follows Person (MFP)	demog_elig_mny_flw_prsn	Supplement
Waiver program enrollment	demog_elig_waiver	Supplement

Table 24. T-MSIS RIF claim filenames

Claims file type	Header/line/occurrence files	File level
Inpatient (IP)	inpatient_header inpatient_line inpatient_occurrence	Header Line Occurrence
Long-term care (LT)	long_term_header long_term_line long_term_occurrence	Header Line Occurrence
Other services (OT)	other_services_header other_services_line other_services_occurrence	Header Line Occurrence
Pharmacy (RX)	rx_header rx_line	Header Line

If the files are extremely large the CCW team may divide it into two or more files, in which case there is a sequential number at the end of the filename — such as “001,” “002,” to enumerate files of this type researchers receive, for example:

res77777req012345_2016_taf_other_services_header_001.dat
res77777req012345_2016 taf_other_services_line_001.dat
res77777req012345_2016 taf_other_services_line_002.dat

Table 25. Annual plan filenames

Annual Plan or Provider file type	Base/Supplemental files	File level
APL base file	mngd_care_plan_base	Base
Service address location	mngd_care_plan_location	Supplement
Service area	mngd_care_plan_srvc_area	Supplement
Enrolled population	mngd_care_plan_enrol_pop	Supplement
Operating authority	mngd_care_plan_oprtg_auth	Supplement

Table 26. Annual provider filenames

APR base file	prvdr_base	Base
Taxonomy	prvdr_taxonomy	Supplement
Enrollment	prvdr_enrollment	Supplement
Affiliated groups	prvdr_afltd_grps	Supplement
Affiliated programs	prvdr_afltd_pgms	Supplement
Address location	prvdr_location	Supplement
License and accreditation	prvdr_license	Supplement
Identifier	prvdr_identifiers	Supplement
Bed type	prvdr_bed_type	Supplement

B. CCW Output Package — Data File Folder*

There are three items in the data file folder:

1. **READ_ME_FIRST_REQXXXX_YYYY.txt** — this is a text file that describes the files contained in the output package. Filename Example: READ_ME_FIRST_REQ012345_2016.txt

2. **Claims files** —

   Table 27. Examples of the filenames, description, and unit of analysis

Claims files filename	Claims files file description	Claims files unit of analysis
res<DUA number>_req<XXX>_<YYYY>_ taf_other_services_header_<# files>	Other services (OT) header claim file	Claim
res<DUA number>_req<XXX>_<YYYY>_ other_services_line_<# files>	Other services line file	Line detail

3. **Reference code files —** the CCW team routinely delivers a reference code file for each of the claims file types (IP, LT, and OT; it is not applicable to the RX file).

   Table 28. Reference code files

Filename	File description	Unit of analysis
res<DUA number>_req<XXX>_<YYYY>_ OT _OCCR	Other services occurrence code file	Code detail

C. Content

Within each of the yearly data folders is a README file, which the CCW team recommends users read first. It is a text file that describes the files contained in the output package.

The CCW team packages these data files as encrypted self-decrypting archive (SDA) files. The CCW team sends the decryption password electronically via email. When researchers receive the data package (hard drive or DVD/CD), copy them from the shipping media to their local workspace. Note depending on the size of the data files, the CCW team ships some data on a hard drive on which users can decrypt. Using the password provided to the researcher via email, follow the decryption instructions enclosed in the data package. Researchers need to enter a password to extract each file.

After researchers extract the data files, they should compare their record count to the control counts that the CCW team obtained in producing their data file. These control counts are in the *.fts file. There is a separate .fts for each data file. The data files are in fixed column flat files. Researchers can use whatever analytic software they choose. For convenience, the CCW has included SAS read-in files. In addition to the raw data files, the SDA generates the following files for each executable in the output package (Table 29).

Table 29. Files contained within SDAs

Filename	File description
<filename>.fts	For each extracted data file, there is a corresponding transfer summary file. The names of these files correspond with the data filename [e.g., res<DUA number>_req<XXX>_<YYYY>_ other_services_line. fts]. This file transfer summary files contain: • Filename • File source • File transfer mode • Row length • File transfer format • Number of columns • Number of rows • File size
<filename>_<level>_v6.sas	Program to read data into a SAS version 6.x environment. For example, the file inpatient_header_claims_read_v6.sas reads the inpatient header claims data into a SAS version 6.x environment.
<filename>_<level>_v8.sas	Program to read data into a SAS version 8.x environment.
<filename>_<level>.dat	Naming convention for *.dat files is in Table 23 and Table 24.

Extract file documentation
Code Reference Sets.xls — describes the ICD-9 or -10 diagnosis and procedure codes, HCPC codes, revenue center, and other codes in the data files.
Decryption Instructions.pdf — instructions for decrypting/uncompressing the data files.
Tips on Getting Started with Data

Chapter 7 — Further Assistance with Medicaid and CHIP Data

There are multiple CCW Help Desk resources available to assist with technical and content questions and problems. Reference the section below to determine which CCW Help Desk to contact.

A. Ordering Data Files, General Contents Questions, and DUA

The Research Data Assistance Center (ResDAC) is the CMS contractor that helps researchers obtain the CCW data. The ResDAC website provides links to descriptions of the available CMS data, request procedures, supporting documentation, and other helpful resources. In particular, the ResDAC site hosts a range of TAF Data Quality briefs and other supplemental materials that CMS produces to describe state-level quality variation in commonly used variables and concepts within the TAF data files. Visit the ResDAC website at (https://www.resdac.org) for additional information.

Researchers can reach ResDAC by:
www.resdac.org
Email: resdac@umn.edu
Phone: 1-888-973-7322

If the researcher requires additional CMS data (data not available from CCW) to meet research objectives, or has any questions about other data sources, they can review all available CMS data by visiting the ResDAC website and contacting ResDAC for further assistance.

B. Questions After CCW has sent Data, or Granted Access

If researchers have already received the TAF RIFs and have questions or concerns regarding the data files, access to the CCW VRDC, or need assistance with the media, contact the CCW Help Desk:

www.ccwdata.org
Email: ccwhelp@ccwdata.org
Phone: 1-866-766-1915

All RIF record layouts, codebooks, and this user guide, are available on the https://www.ccwdata.org website.

Appendix A — List of Acronyms

Acronym	Definition
APL	Annual plan file
APR	Annual provider file
CCW	Chronic Conditions Warehouse
CHIP	Children's Health Insurance Program both M-CHIP and S-CHIP (unless specified otherwise)
CMCS	Center for Medicaid and CHIP Services
CMS	Centers for Medicare & Medicaid Services
DE	Demographic and eligibility file
DQ	Data Quality (Atlas)
DRG	Diagnosis related group
DUA	Data Use Agreement
FPL	Federal Poverty Level
FFS	Fee-for-Service
FPL	Federal Poverty Level
HCBS	Home and Community-Based Services
HCPCS	Healthcare Common Procedure Coding System
ICD-9	International Statistical Classification of Diseases and Related Health Problems, 9th revision
ICD-10	International Statistical Classification of Diseases and Related Health Problems, 10th revision
IP	Inpatient
LT	Long-term care file
LTSS	Long-term services and supports
MAX	Medicaid Analytic eXtract data files
MAX-T	MAX file produced with T-MSIS data
M-CHIP	Medicaid Expansion Children's Health Insurance Program
MFP	Money Follows the Person
MSIS	Medicaid Statistical Information System
NDC	National Drug Code
NPI	National Provider Identifier
OT	Other services
ResDAC	Research Data Analytic Center
RIF	Research Identifiable File
RX	Pharmacy
S-CHIP	Separate Children's Health Insurance Program
SDA	Self-decrypting archive
SPO	State plan options
SSI	Social Security Income
TAF	T-MSIS Analytic File
T-MSIS	Transformed Medicaid Statistical Information System
TPA	Third-party administrator
TPL	Third-party liability
VRDC	Virtual Research Data Center

Appendix B — State Cutover to T-MSIS and Availability of MAX and TAF Research Files

This table displays the calendar quarter when each state began submitting T-MSIS data, as well as whether TAF RIFs are available for each calendar year (CY). An additional appendix (Appendix D — TAF RIF Releases) contains information regarding TAF RIF releases for each CY data file.

State	ANSI state code (STATE_CD)	First TMSIS submission	MAX end	MAX-T* start	MAX-T* end	TAF start	TAF end
Alabama	AL	14-Jan	2013	N/A	N/A	2014	N/A
Alaska	AK	13-Oct	2013	N/A	N/A	2014	N/A
Arizona	AZ	14-Oct	2013	2014	2014	2015	N/A
Arkansas	AR	15-Apr	2013	2014	2015	2016	N/A
California	CA	15-Oct	2014	2015	2015	2016	N/A
Colorado	CO	11-Oct	2013	N/A	N/A	2014	N/A
Connecticut	CT	15-Apr	2013	2014	2015	2016	N/A
Delaware	DE	14-Jan	2013	N/A	N/A	2014	N/A
District of Columbia	DC	14-Jan	2013	N/A	N/A	2014	N/A
Florida	FL	13-Oct	2013	N/A	N/A	2014	N/A
Georgia	GA	15-Oct	2014	2015	2015	2016	N/A
Guam	GU	15-Oct	N/A	N/A	N/A	2023	N/A
Hawaii	HI	14-Oct	2013	2014	2014	2015	N/A
Idaho	ID	15-Oct	2014	2015	2015	2016	N/A
Illinois	IL	14-Jan	2013	N/A	N/A	2014	N/A
Indiana	IN	14-Oct	2013	2014	2014	2015	N/A
Iowa	IA	15-Oct	2014	2015	2015	2016	N/A
Kansas	KS	13-Jan	2013	N/A	N/A	2014	N/A
Kentucky	KY	14-Jul	2013	2014	2014	2015	N/A
Louisiana	LA	15-Oct	2014	2015	2015	2016	N/A
Maine	ME	14-Jan	2013	N/A	N/A	2014	N/A
Maryland	MD	14-Jan	2013	N/A	N/A	2014	N/A
Massachusetts	MA	14-Oct	2013	2014	2014	2015	N/A
Michigan	MI	15-Oct	2014	2015	2015	2016	N/A
Minnesota	MN	15-Oct	2014	2015	2015	2016	N/A
Mississippi	MS	15-Oct	2014	2015	2015	2016	N/A
Missouri	MO	15-Oct	2014	2015	2015	2016	N/A
Montana	MT	14-Jan	2013	N/A	N/A	2014	N/A
Montana TPA **	MT	16-Jan	N/A	N/A	N/A	2017	2019
Nebraska	NE	14-Jan	2013	N/A	N/A	2014	N/A

State	ANSI state code (STATE_CD)	First TMSIS submission	MAX end	MAX-T* start	MAX-T* end	TAF start	TAF end
Nevada	NV	14-Jan	2013	N/A	N/A	2014	N/A
New Hampshire	NH	14-Jan	2013	N/A	N/A	2014	N/A
New Jersey	NJ	15-Oct	2014	2015	2015	2016	N/A
New Mexico	NM	14-Jan	2013	N/A	N/A	2014	N/A
New York	NY	15-Jul	2013	2014	2015	2016	N/A
North Carolina	NC	13-Jul	2013	N/A	N/A	2014	N/A
North Dakota	ND	14-Jan	2013	N/A	N/A	2014	N/A
Ohio	OH	14-Oct	2013	2014	2014	2015	N/A
Oklahoma	OK	14-Oct	2013	2014	2014	2015	N/A
Oregon	OR	15-Jul	2013	2014	2015	2016	N/A
Pennsylvania	PA	15-Oct	2014	2015	2015	2016	N/A
Pennsylvania CHIP	PA	14-Jan	N/A	N/A	N/A	2016	N/A
Puerto Rico	PR	15-Jan	N/A	N/A	N/A	2015	N/A
Rhode Island	RI	12-Oct	2013	N/A	N/A	2014	N/A
South Carolina	SC	14-Jul	2013	2014	2014	2015	N/A
South Dakota	SD	15-Oct	2014	2015	2015	2016	N/A
Tennessee	TN	15-Oct	2014	2015	2015	2016	N/A
Texas	TX	14-Jul	2013	2014	2014	2015	N/A
Utah	UT	15-Oct	2014	2015	2015	2016	N/A
Vermont	VT	15-Oct	2014	2015	2015	2016	N/A
Virginia	VA	14-Apr	2013	2014	2014	2015	N/A
Virgin Islands	VI	17-Jan	N/A	N/A	N/A	2017	N/A
Washington	WA	15-Jan	2013	2014	2014	2015	N/A
West Virginia	WV	15-Oct	2014	N/A	N/A	2016	N/A
Wisconsin	WI	14-Jan	2013	N/A	N/A	2014	N/A
Wyoming	WY	15-Oct	2014	2015	2015	2016	N/A
Wyoming CHIP†	WY	15-Oct	N/A	N/A	N/A	2016	2020

* MAX-T is a version of the MAX file produced using T-MSIS, at least partly.
** MT TPA no longer submits data after 2019.
† WY CHIP is retired (i.e., not a separate SUBMTG_STATE_CD) and included in 56: Wyoming, after the 2020 data year. WY no longer differentiates between CHIP and non-CHIP.
N/A = No data available

Appendix C — Managed Care Redacted Payment Fields

For the TAF RIF, managed care payment fields are restricted. Below are the 36 fields in the claim types that have restricted access, according to CMCS.

- IP header
  1. Total Medicaid paid amount (MDCD_PD_AMT)
  2. Total Medicare deductible amount (MDCR_DDCTBL_PD_AMT)
  3. Total Medicare coinsurance amount (MDCR_COINSRNC_PD_AMT)
  4. Total billed amount (BILLED_AMT)
  5. Total allowed amount (MDCD_ALOWD_AMT)

- IP line
  6. Medicaid paid amount (LINE_MDCD_PD_AMT)
  7. Charge amount (REV_CNTR_CHRG_AMT)
  8. Allowed amount (LINE_MDCD_ALOWD_AMT)

- LT header
  9. Total Medicaid paid amount (MDCD_PD_AMT)
  10. Total Medicare deductible amount (MDCR_DDCTBL_PD_AMT)
  11. Total Medicare coinsurance amount (MDCR_COINSRNC_PD_AMT)
  12. Total billed amount (BILLED_AMT)
  13. Total allowed amount (MDCD_ALOWD_AMT)
  14. Aggregate room and board (MDCD_ACMDTN_PD_AMT)
  15. Aggregate ancillary amount (MDCD_ANCLRY_PD_AMT)

- LT line
  16. Medicaid paid amount (LINE_MDCD_PD_AMT)
  17. Charge amount (REV_CNTR_CHRG_AMT)
  18. Allowed amount (LINE_MDCD_ALOWD_AMT)

- RX header
  19. Total Medicaid paid amount (MDCD_PD_AMT)
  20. Total Medicare deductible amount (MDCR_DDCTBL_PD_AMT)
  21. Total Medicare coinsurance amount (MDCR_COINSRNC_PD_AMT)
  22. Total billed amount (BILLED_AMT)
  23. Total allowed amount (MDCD_ALOWD_AMT)

- RX line
  24. Medicaid paid amount (LINE_MDCD_PD_AMT)
  25. Medicare deductible amount (LINE_MDCR_DDCTBL_PD_AMT)
  26. Medicare coinsurance amount (LINE_MDCR_COINSRNC_PD_AMT)
  27. Billed amount (LINE_BILLED_AMT)
  28. Allowed amount (LINE_MDCD_ALOWD_AMT)

- OT header
  29. Total Medicaid paid amount (MDCD_PD_AMT)
  30. Total Medicare deductible amount (MDCR_DDCTBL_PD_AMT)
  31. Total Medicare coinsurance amount (MDCR_COINSRNC_PD_AMT)

32. Total billed amount (BILLED_AMT)
33. Total allowed amount (MDCD_ALOWD_AMT)

- OT line
  34. Medicaid paid amount (LINE_MDCD_PD_AMT)
  35. Billed amount (LINE_BILLED_AMT)
  36. Allowed amount (LINE_MDCD_ALOWD_AMT)

These 36 fields only have values suppressed for managed care claims. Researchers can identify managed care type claims using the field CLM_TYPE_CD = '3', 'C', 'W'. Researchers should consider all other values non-managed care and do not require restriction. The CCW team populates all other RIF fields for these claim records (e.g., the CCW team does not apply additional suppression for the managed care records).

Appendix D — TAF RIF Releases

Due to improvements in the quality of state data over time, CMS may decide to produce an updated RIF for a CY file. The below table presents the historical TAF RIF releases.

TAF RIF release	CCW data load	CCW RIF announcement	Notes
2014 (release 1)	Summer 2019	November 2019	No longer available for delivery
2015 (release 1)	Summer 2019	November 2019	No longer available for delivery
2016 (release 1)	Summer 2019	November 2019	No longer available for delivery
2017	May 2020	September 2020	No longer available for delivery
2018	June 2020	September 2020	No longer available for delivery
2014 (release 2)	July 2020	November 2020	
2015 (release 2)	July 2020	November 2020	
2016 (release 2)	July 2020	November 2020	
2019 (preliminary)*	October 2020	December 2020	No longer available for delivery
2014 APL and APR	March 2021	May 2021	
2015 APL and APR	March 2021	May 2021	
2016 APL and APR	March 2021	May 2021	
2017 (release 2)	May 2021	August 2021	Includes enrollment, claims, APL, and APR
2018 (release 2)	May 2021	August 2021	
2019	July 2021	September 2021	
2020 (preliminary)*	August 2021	November 2021	No longer available for delivery
2020	April 2022	October 2022	
2021 (preliminary)*	September 2022	December 2022	No longer available for delivery
2021	August 2023	December 2023	
2022 (preliminary)*	September 2023	December 2023	No longer available for delivery
2022	August 2024	December 2024	
2023 (preliminary)*	September 2024	December 2024	

*Data in the preliminary files are immature and the CCW team releases the files again when the data are fully mature.