Memorandum

To: David Moeny, Sarah Dutcher, Efe Eworuke, Patricia Bright, Jamila Mwidau, Lucia Menegussi, Jamal Jones, Terrence Lee

From: Sentinel Operations Center and Duke University Department of Population Health Sciences

CC: James Mork, Nick Williams

Re: TAF Characteristics Memo

Comments: Version 1.0.1

TAF Characteristics Memo: All-States Medicaid

Principal Investigator: Bradley Hammill, DrPH[1]

Prepared by: Michael Stagner[1], Steven J. Lippmann, PhD[1]; Jessica E. Pritchard, PhD[1]; Bradley Hammill, DrPH[1]

Author Affiliations: [1]Department of Population Health Sciences, Duke University School of Medicine, Durham, NC

Originally Submitted: November 11, 2021
Update: October 22, 2024

DISCLAIMER TO READERS

- The content in this document is technical and intended for use by scientists, analysts, and programmers, in various areas of expertise.
- Source data files obtained from CMS by other researchers may have different file names, different partitioning schemes (e.g., annual RIF), different samples of the data (i.e., not 100% Fee-For-Service and Comprehensive Managed Care) and possibly different variables and/or variable names.
- Upon review of this document, there is no mechanism for technical support by Duke University, the Sentinel Operations Center, ResDAC, CMS, or by the U.S. Food and Drug Administration (FDA), for answering questions that may arise.
- In no event shall any individual, the Duke University Department of Population Health Sciences, the Sentinel Operations Center located at Harvard Pilgrim Health Care Institute, nor the FDA, be liable for any damages whatsoever relating to the use, misuse, or inability to make use of this document (including, without limitation, damages for loss of profits or revenue, business interruption, loss of information, or any other loss).

Table of Contents

INDEX OF TABLES AND FIGURES

Abbreviations

Abbreviation	Description
ACA	Patient Protection and Affordable Care Act
APR	Annual Provider File
BHO	Behavioral Health Organization
CHIP	Children's Health Insurance Program
CCW	Chronic Conditions Warehouse
CMC	Comprehensive managed care
CMS	Centers for Medicare & Medicaid Services
CY	Calendar year
DE	Demographic and eligibility
DPHS	Duke University Department of Population Health Sciences
DQ	Data Quality (in the context of the DQ Atlas)
DUA	Data Use Agreement
ETL	Extract, transform, and load
FFS	Fee-for-service
HCUP	Healthcare Cost and Utilization Project
ID	Identifier
IP	Inpatient
LT	Long-term care
M-CHIP	Medicaid Expansion CHIP
MACBIS	Medicaid and CHIP Business Information Solutions
MAX	Medicaid Analytic eXtract
MEC	Minimum essential coverage
MIL	Mother-infant linkage
MSIS	Medicaid Statistical Information System
NPI	National Provider Identifier
OT	Other services
PCCM	Primary Care Case Management
RIF	Research Identifiable File
RX	Pharmacy
S-CHIP	Separate CHIP
SCDM	Sentinel Common Data Model
TAF	T-MSIS Analytic Files
T-MSIS	Transformed Medicaid Statistical Information System
VRDC	Virtual Research Data Center

Jurisdiction Abbreviations

Abbreviation	Description
AK	Alaska
AL	Alabama
AR	Arkansas
AZ	Arizona
CA	California
CO	Colorado
CT	Connecticut
DC	District of Columbia
DE	Delaware
FL	Florida
GA	Georgia
GU	Guam
HI	Hawaii
IA	Iowa
ID	Idaho
IL	Illinois
IN	Indiana
KS	Kansas
KY	Kentucky
LA	Louisiana
MA	Massachusetts
MD	Maryland
ME	Maine
MI	Michigan
MN	Minnesota
MO	Missouri
MS	Mississippi

Abbreviation	Description
MT	Montana
NC	North Carolina
ND	North Dakota
NE	Nebraska
NH	New Hampshire
NJ	New Jersey
NM	New Mexico
NV	Nevada
NY	New York
OH	Ohio
OK	Oklahoma
OR	Oregon
PA	Pennsylvania
PR	Puerto Rico
RI	Rhode Island
SC	South Carolina
SD	South Dakota
TN	Tennessee
TX	Texas
UT	Utah
VA	Virginia
VI	Virgin Islands
VT	Vermont
WA	Washington
WI	Wisconsin
WV	West Virginia
WY	Wyoming

I. Introduction

The Medicaid/CHIP data comprises information about beneficiary enrollment and eligibility, utilization and claims, and expenditures for people covered by Medicaid or the Children's Health Insurance Program (CHIP). States and territories of the United States administer these health insurance programs and submit their enrollment and claims data to the federal government in the Transformed Medicaid Statistical Information System (T-MSIS) format. T-MSIS is a standardized format that was first implemented in 2014, with all states complying by 2016. An earlier submission system known as the Medicaid Statistical Information System (MSIS) was used previous to T-MSIS, from which the Medicaid Analytic eXtract (MAX) files were created. Because the T-MSIS source data are not suitable for research due to their complexity and frequent, rolling updates, the Centers for Medicare and Medicaid Services (CMS) creates the T-MSIS Analytic Files (TAFs) from the T-MSIS data supplied by the individual jurisdictions, and releases these files to researchers in the TAF Research Identifiable Files (RIF) format.

This memo describes the characteristics and quality of the 100% Medicaid TAF RIFs housed in the CMS Virtual Research Data Center (VRDC). It also describes the preliminary recommendations of the staff in the Department of Population Health Sciences (DPHS) at Duke University regarding the extraction, transformation, and loading (ETL) of these data into the Sentinel Common Data Model (SCDM) as part of Sentinel's eventual Medicaid/CHIP DataMart.
This memo consists of the following sections:

- **TAF Characteristics:**
  - **Source Data**: This section describes the provenance, content, coverage, structure, and update schedule of the 100% TAF RIF data stored within the VRDC. This is applicable to the TAF RIF source data from 2014 through 2018.
  - **Data Quality**: This section summarizes data quality (DQ) information published by MACBIS (Medicaid and CHIP Business Information Solution) about the 100% TAF RIF data and classifies each DQ metric based on its perceived relevance to the SCDM.
- **ETL Characteristics**: This section describes the recommended inclusion/exclusion criteria for the first Medicaid/CHIP ETL and plans for monitoring data quality in future ETLs. This section also addresses other pertinent information regarding working with the TAF RIF data, including the expected process of the Mother-Infant Linkage (MIL).

In this memo, there are a few conventions we will use:

- We will refer to "jurisdictions" instead of "states," since the District of Columbia (DC) and several territories of the United States, like Puerto Rico and the U.S. Virgin Islands, also administer Medicaid and CHIP programs and submit data to CMS.
- We will refer to "claims" when discussing the healthcare utilization information in the TAF RIFs, acknowledging that some of this information is generated by fee-for-service (FFS) insurance products (i.e., true claims data), while other information is generated by managed care insurance products (i.e., encounter data).
- Unless specifically noted, we will use "Medicaid/CHIP" to refer to both the Medicaid and CHIP programs.
- Finally, we refer to the "Medicaid/CHIP data" as a shorthand for all the Medicaid and CHIP data under the control of CMS, since the 100% TAF RIF data in the VRDC are the end product of an internal process of CMS data transformations.

II. TAF Characteristics

A. Source Data

1. Types of Medicaid and CHIP Programs in the Source Data

Beneficiaries in the Medicaid/CHIP data are either Medicaid (Title XIX of the Social Security Act) or CHIP (Title XXI of the Social Security Act) recipients. Medicaid and CHIP, combined, provide health coverage to millions of Americans of all ages, including eligible low-income adults, children, pregnant women, elderly adults and people with disabilities. In 2010, the Patient Protection and Affordable Care Act (ACA) included a provision to expand Medicaid eligibility to adults up to age 64 with incomes up to 138% of the federal poverty level. (Prior to the ACA, Medicaid was generally never available to non-disabled adults under age 65 unless they had minor children.) As of August 2021, 38 states and Washington, DC have adopted this income-based Medicaid eligibility expansion under the ACA. States may administer their CHIP programs by either expanding their current Medicaid program eligibility (Medicaid Expansion CHIP, or M-CHIP), creating a new program altogether (Separate CHIP, or S-CHIP), or pursuing a combination of these two approaches. Determination of Medicaid and CHIP eligibility for patients varies widely by state.

Additionally, some Medicaid/CHIP beneficiaries may also be eligible for and enrolled in Medicare at the same time. While some data for these dually-eligible beneficiaries will appear in the Medicaid/CHIP data, not all information about service utilization is expected to be there, as Medicare is the primary payer for their care. In the process of creating the Medicaid/CHIP and CMS ETLs for Sentinel, Patient Identifiers (PatIDs) will be generated using a method that does not allow for linkage of beneficiaries between the two sets of SCDM tables. Each ETL will, however, maintain a crosswalk table of SCDM PatIDs and source data Beneficiary Identifiers (BENE_IDs) that could be used to link Medicare and Medicaid/CHIP patients between ETLs on an ad-hoc basis.

2. Scope of Benefits

The healthcare benefits available to Medicaid/CHIP beneficiaries can vary greatly between jurisdictions, and even within a single jurisdiction since there are dozens of eligibility categories with different coverages. Benefits packages associated with each eligibility category are classified as either full-scope, comprehensive, or limited.

- **Full-scope** benefits include all services provided by a jurisdiction's Medicaid/CHIP program, both mandatory (e.g., inpatient and outpatient hospital care, physician care) and optional (e.g., prescription drugs, dental care, physical therapy). Even though prescription drug coverage is technically optional, all jurisdictions provide this coverage as part of their full-scope benefits. Appendix A lists benefits included in full-scope coverage.
- **Comprehensive** benefits cover those services necessary to fulfill the minimum essential coverage (MEC) mandated by the ACA, including acute-care services, prescription drugs, and laboratory services. Appendix A lists benefits included in comprehensive coverage.
- **Limited** benefits provide narrowly-defined coverage to beneficiaries, usually tailored to specific services or conditions (e.g. family planning, emergency services).

S-CHIP programs are required to offer comprehensive benefit packages (at a minimum), and many jurisdictions provide full-scope benefits to S-CHIP beneficiaries. As a result, there is not as much variety in S-CHIP benefits within jurisdictions compared to that found in Medicaid and M-CHIP programs.

The TAF RIFs contain monthly indicators for each beneficiary that denote the type of Medicaid/CHIP program the beneficiary is enrolled in, along with the scope of benefits provided. The majority of Medicaid/CHIP benefits packages include both medical and drug coverage.

3. Creation of the TAF RIF Files

As noted in the introduction, CMS creates the TAFs from the T-MSIS data supplied by the individual jurisdictions and releases redacted RIF files based on these T-MSIS data to researchers. The transformation of the T-MSIS source data to the released TAF RIFs occurs in several steps: T-MSIS data → TAF Interim Files → TAF Unredacted RIFs → TAF Redacted RIFs. While the T-MSIS data are constantly updated and not research-ready, the different versions of the TAFs are used for internal (to CMS) or external research by different entities.

Table 1. Main differences in content and access between the different TAF files

	TAF Interim Files (Unredacted)	TAF Unredacted RIFs	TAF Redacted RIFs
What is the data format used by these files?	T-MSIS source data format	Research format	Research format
Do these files include the CCW Beneficiary ID for tracking beneficiaries across jurisdictions and year?	No	Yes	Yes
Do these files include personally identifiable information (PII) like name, full address, etc.?	Yes, some PII	No	No
Do these files include service tracking records and excluded claims records?	Yes, both service tracking and excluded claims records are comingled with standard claims records	Both service tracking and excluded claims records are available in separate data sets from standard claims	Service tracking records are available in separate data sets (beginning with 2017 and 2018 TAF RIF Release 2). Excluded Claims records are not available.
May Research DUAs use these files?	No	No	Yes, must use these files

	TAF Interim Files (Unredacted)	TAF Unredacted RIFs	TAF Redacted RIFs
Who may use these files?	Selected users with approved justifications may have access. Users with access to the Interim files also get access to the Unredacted RIFs, and vice versa	Selected users with approved justifications may have access. Users with access to the Interim files also get access to the Unredacted RIFs, and vice versa	Most other (non-research DUA) users will use these files
Is MACBIS approval required to access these files?	Yes	Yes	No

4. Available Files

The TAF RIF data include several different data files.

- Beneficiary information - eligibility categories, scope of benefits, enrollment windows, demographics, and death information - is contained in the annual **Demographic and Eligibility (DE) Files**. The DE files comprise a Base file and six supplemental files: Eligibility Dates, Managed Care Enrollment, Waiver Program Enrollment, Money Follows the Person, Health Home and State Plan Options, and Disability and Need. For the purposes of transforming the Medicaid/CHIP data into the SCDM format, DPHS will be utilizing the Base and Eligibility Dates files only.
- Drug utilization information about prescription and over-the-counter drugs filled at a pharmacy is found in the **Pharmacy (RX) files**.
- Medical utilization information, based on both FFS claims and managed care encounter data for healthcare services delivered, is contained in either the **Inpatient (IP), Long-Term Care (LT), or Other Services (OT) files**. The OT file can include outpatient, dental, home health, durable medical equipment, and many other categories of healthcare services.
- Provider and facility information is found in the **Annual Provider File (APR).** The term "Provider" is used broadly and includes individual practitioners, groups, and facilities. These data are stored relationally as a base provider file and eight supplemental files. This set of files may be useful in populating certain SCDM fields, such as facility location and provider specialty, although we do not currently have access to these files for Medicaid/CHIP ETL 1. Some of this information, however, is also present in the claims files.

Documentation for all redacted TAF RIFs can be found under the Medicaid section at https://www.ccwdata.org/web/guest/data-dictionaries. The "Record Layout" spreadsheets available here list which variables appear in each dataset; and the PDF "Codebook" documents describe each field in detail, including valid value sets, when appropriate. For DE documentation, see the "T-MSIS Analytic File (TAF) Demographic and Eligibility" links; for all medical utilization documentation (RX, IP, LT, & OT), see the "T-MSIS Analytic File (TAF) Claims" links; and for Provider file documentation, see the "T-MSIS Analytic File (TAF) Annual Provider" links.

All redacted TAF RIFs are available in yearly SAS™ libraries (named TAFR*yy*, where *yy* indicates the service year) within the VRDC.

Table 2. Naming conventions for redacted TAF RIFs within the VRDC

Source file category	Source file type	Filename[1]
Demographic & eligibility (DE) files	Base	DEMOG_ELIG_BASE
Demographic & eligibility (DE) files	Enrollment dates	DEMOG_ELIG_DATES
Pharmacy (RX) claims	Header records	RX_HEADER_*mm*
Pharmacy (RX) claims	Line records	RX_LINE_*mm*
Inpatient (IP) claims	Header records	INPATIENT_HEADER_*mm*
Inpatient (IP) claims	Line records	INPATIENT_LINE_*mm*
Long-term care (LT) claims	Header records	LONG_TERM_HEADER_*mm*
Long-term care (LT) claims	Line records	LONG_TERM_LINE_*mm*
Other services (OT) claims	Header records	OTHER_SERVICES_HEADER_*mm*
Other services (OT) claims	Line records	OTHER_SERVICES_LINE_*mm*

Drug and medical service utilization data are stored in the TAF RIFs as header and line files. Header records contain summary information related to a single claim. Line records include additional details related to specific, individual services billed for a claim. The information found in header and line records is directly correlated to the layout of the institutional (UB-04) and professional (CMS-1500) claims forms submitted by health providers. Header and line files are stored in monthly files within the yearly SAS™ libraries and contain data for all jurisdictions with T-MSIS data in that year.

5. Release Schedule/Data Maturity

CMS has not yet specified a release schedule for the TAF RIFs, but has generally committed to the following plan: they will release a preliminary version of the TAF RIFs for each service year based on information submitted to CMS through June 30 of the following year, for a minimum data maturity of six months. (The 2019 TAF RIFs were the first available as a preliminary release.) Preliminary TAF RIFs are not considered fully mature for research purposes since jurisdictions are still submitting T-MSIS service data from that calendar year when those files are created. Once the TAF RIF data for a service year are fully mature and include at least 12 months of runout, CMS will make an initial release of the TAF RIF data (Release 1). As the quality of submitted enrollment and claims data improves, however, CMS may periodically release updates to these files. For example, the 2014–2018 TAF RIF data currently available in the VRDC have all been updated (as Release 2) from their initial release. The timing of these releases is shown in Table 3 below.

Table 3. Release schedule of the currently available 2014-2019 TAF RIF data

TAF RIF Year	Preliminary Release	Release 1	Release 2
2014	--	Nov-2019	Nov-2020
2015	--	Nov-2019	Nov-2020
2016	--	Nov-2019	Nov-2020
2017	--	Sep-2020	Sep-2021

[1] *mm* indicates a month-specific file (e.g., 01 for January)

TAF RIF Year	Preliminary Release	Release 1	Release 2
2018	--	Sep-2020	Sep-2021
2019	Dec-2020	Sep-2021	--
2020	Nov-2021		

Once a new release of TAF RIF data is available, the new data replace the old data in the VRDC. In the past, the prior release has been made available in a separate SAS™ library for several months after their replacement in the main SAS™ library.

6. Time Period Coverage

The TAF RIF data cover years 2014 and later. As noted above, data from 2014 to 2018 are available as Release 2, and data from 2019 has been released as Release 1. Prior to 2014, all jurisdictions submitted Medicaid/CHIP data to CMS in the older Medicaid Statistical Information System (MSIS) format. CMS used those data from 1999 to 2015 to create the MAX files. The MAX data file format is substantially different from the TAF RIF format and will not be used for this project.

7. Geographic Coverage

Jurisdictions appear in the TAF RIF data beginning in the first full calendar year they began submitting Medicaid/CHIP data to CMS in the T-MSIS format. The 2014 TAF RIF contain data from 18 states, plus the District of Columbia (DC). The 2015 TAF RIF contain data from 29 states, plus DC and Puerto Rico (PR). In 2016, all states were required to use the T-MSIS format, so the 2016 TAF RIF contain data for all 50 states, plus DC and PR. The 2017 and subsequent TAF RIFs contain data for all 50 states, DC, PR, and the US Virgin Islands.

Appendix B lists the jurisdictions included in the 2014 – 2018 TAF RIFs by year.

8. Record Count Estimates

To provide a sense of the size of the current 2014–2018 TAF RIFs, Tables 4 and 5 present record counts for individual source data files.

Table 4. Demographic file counts

TAF RIF Year	Beneficiaries	Deaths
2014	20,720,816	197,430
2015	46,901,744	422,307
2016	93,490,003	833,360
2017	93,942,808	852,075
2018	93,214,883	853,456

Table 5. Utilization header file counts

TAF RIF Year	Inpatient (IP)	Long-term Care (LT)	Other Services (OT)[2]	Pharmacy (RX)
2014	3,042,145	4,942,352	654,186,541	138,970,363
2015	7,704,645	12,053,498	1,569,326,899	392,890,423

[2] Roughly 45% of these claims are capitated payment claims, not service use claims.

TAF RIF Year	Inpatient (IP)	Long-term Care (LT)	Other Services (OT)[2]	Pharmacy (RX)
2016	14,971,216	34,075,228	3,806,640,636	846,807,453
2017	15,057,001	31,244,988	4,061,021,732	840,361,355
2018	14,570,126	31,145,958	4,223,973,259	841,144,276

Not all records in all tables will be used for Medicaid/CHIP ETL 1, given the recommended exclusion criteria in Section IV.

B. Data Quality

1. Overview

Since T-MSIS data are generated and submitted by individual jurisdictions to CMS, data quality within the TAF RIF can vary by year within a jurisdiction or between FFS and managed care service use data within a jurisdiction/year. For this reason, data quality at the jurisdiction/year/plan level will need to be considered when deciding which data to include or exclude in the final Medicaid/CHIP ETL tables. This variability in data quality across jurisdictions represents a significant difference from the Medicare RIF data, where all submitted claims are subject to a common set of rules.

Internal consistency in data completeness and quality within jurisdiction/year/plan also needs to be considered because inconsistent data quality could yield inaccurate reporting of enrollment or utilization that could potentially bias Sentinel query results. Poor data quality can occur in a number of ways. For example, claims data for a given jurisdiction/year may be considered low quality on its own, or claims data for a given jurisdiction/year could be considered high quality, but the corresponding enrollment data might be lower quality. In both cases it could be determined necessary to exclude the data for the whole jurisdiction/year from an ETL release. For comprehensive managed care (CMC) plans, claims data in a given jurisdiction/year/plan could be considered low quality also, but in those instances, it would only be necessary to exclude data for managed care plans in that jurisdiction/year.

2. Source for Information about Data Quality

CMS conducts ongoing data quality assessments on the TAF RIF data to detect data quality differences across jurisdictions, years, and FFS vs. managed care plans. A collection of interactive Data Quality (DQ) Assessment maps and tables, along with several technical guidance documents, are provided by CMS in the DQ Atlas (https://www.medicaid.gov/dq-atlas/welcome). DQ Assessments summarize the data quality within the TAF RIF data for a specific topic by jurisdiction, year, plan, and TAF RIF release version. Since assessments are conducted for each release of a TAF RIF, each Medicaid/CHIP ETL will need to use the most current assessments; changes in the assessed data quality for a given jurisdiction/year/plan may impact the inclusion/exclusion of the corresponding TAF RIF data for each Medicaid/CHIP ETL iteration.

DQ Assessment topics that may be relevant to the Medicaid/CHIP ETL (i.e., topics unrelated to payments and non-claim records) are grouped into the following categories: Enrollment Benchmarking, Enrollment Patterns over Time, Beneficiary Information, Claim Files Completeness, Service Use Information, and Provider Information. For each jurisdiction, the data quality of each topic is assessed using methodology described in the DQ Atlas. As part of this methodology, topic-specific thresholds have been established that allow the categorization of a jurisdiction/year/plan's data on that topic as Low Concern, Medium

Concern, High Concern, or Unusable. Topics may also be assessed as Unclassified when a topic is not relevant to a jurisdiction (e.g., a managed care-related topic for a jurisdiction without a managed care plan) or when there is no relevant data for assessing a topic (e.g., topics related to ICD-10 codes are assessed as Unclassified prior to 2016).

The majority of DQ Atlas topics assess demographics, enrollment, and claims data/variables that will be critically important to the Medicaid/CHIP ETL and the Sentinel project. In assigning "importance" levels for DQ Atlas topics, however, we are attempting to specifically determine whether any widespread data quality issues might exist that would necessitate all of the data for a jurisdiction/year/plan be excluded entirely from the Medicaid/CHIP ETL (see Section III.D). For example, DQ Atlas topics related to enrollment patterns over time (as described below) are considered "very important" in deciding whether to exclude a jurisdiction/year/plan in the Medicaid/CHIP ETL, since problems with enrollment patterns may call into question all of the data for a jurisdiction/year/plan. On the other hand, demographic variable topic assessments are considered as "limited importance" for inclusion in the Medicaid/CHIP ETL (despite being critical to the SCDM) because missing values for these types of variables (such as age, sex, or race) would be handled on a beneficiary level, rather than jurisdiction level.

The following subsections summarize each DQ assessment topic and rate each topic specifically on their perceived importance to inclusion in the Medicaid/CHIP ETL. Summaries, backgrounds, and assessment methodologies for each DQ Atlas topic can be found on the DQ Atlas site at https://www.medicaid.gov/dq-atlas/landing/topics/info.

a. Enrollment Benchmarking

These topics assess the actual versus expected numbers of enrollees in multiple Medicaid/CHIP programs, based on an external data source. Program enrollment numbers that are assessed in the DQ Atlas include all Medicaid, M-CHIP/S-CHIP, dual-enrolled in Medicare and Medicaid, CMC, and Medicaid expansion populations. Enrollment benchmarking topics focus on higher-level population and program enrollment counts, but the Medicaid/CHIP ETL will include beneficiaries based upon scope-of-benefits, regardless of enrollment program. As such and as summarized in table 6, these enrollment benchmarking topics will not be useful for making inclusion/exclusion decisions.

Table 6. Enrollment benchmarking importance to inclusion in Medicaid/CHIP ETL

DQ Topic: Enrollment	Importance to Inclusion in Medicaid/CHIP ETL
CMC Plans	Not important
PCCM Programs	
BHO Plans	
M-CHIP and S-CHIP	
Adult Expansion	
Newly Eligible Adult	
Total Medicaid and CHIP	
Medicaid-Only Enrollment	
Dually Enrolled in Medicare	

b. Enrollment Patterns Over Time

These topics assess potential data quality issues related to a beneficiary's eligibility dates within a given year. The lengths of enrollment gaps are analyzed for beneficiaries who dis-enroll and re-enroll within a calendar year. Overlapping enrollment spans are assessed for beneficiaries who appear to be enrolled in Medicaid and CHIP at the same time. The number of enrollment spans per beneficiary is assessed for beneficiaries that have either only one, or more than two enrollments span within the year. While the overwhelming majority of Medicaid/CHIP beneficiaries are expected to have only one enrollment span per year (i.e., most beneficiaries will not have a gap in enrollment), jurisdictions reporting 100% of beneficiaries with only one span might have a problem with their enrollment data. We consider all three of these enrollment pattern topics as critical to making inclusion/exclusion decisions.

Table 7. Enrollment over time importance to inclusion in Medicaid/CHIP ETL

DQ Topic: Enrollment over Time	Importance to Inclusion in Medicaid/CHIP ETL
Number of Enrollment Spans	Very important
Length of Enrollment Gaps	
Overlapping Enrollment Spans	

c. Beneficiary Information

These DQ topics assess missing, incomplete, and incorrect data in the TAF RIFs' demographic and eligibility files. Demographic topics include assessments of age, gender, race/ethnicity, income, and ZIP code variables. Some of these topics are deemed of limited importance to inclusion in the Medicaid/CHIP ETL because the relevant SCDM fields are optional, instead of required. As such, even though these variables are vital to Sentinel project analyses and incompleteness in any of the variables will need to be dealt with at a beneficiary level, DQ Atlas assessments for these demographic variable topics will not be used to exclude entire jurisdiction/year/plans. Eligibility topics include assessments of eligibility group code (which classifies the reason for a beneficiary's Medicaid/CHIP eligibility), CHIP code (which distinguishes between Medicaid and CHIP enrollment), dual Medicare/Medicaid code (which is critical for identifying dually-enrolled beneficiaries), and restricted benefits code (which is key for identifying beneficiaries with full-scope/comprehensive benefits) variables.

Table 8. Beneficiary information importance to inclusion in Medicaid/CHIP ETL

DQ Atlas Topic: Beneficiary Information		Importance to Inclusion in Medicaid/CHIP ETL
Demographics	Gender	Limited importance
	Age	Limited importance
	ZIP Code	Limited importance
	Race and Ethnicity	Limited importance
	Income	Not important
Eligibility Codes	Dual Eligibility Code	Very important
	Restricted Benefits Code	Very important
	CHIP Code	Limited importance
	Eligibility Group Code	Not important

d. Claim File Completeness

These topics assess issues related to claims volumes, service usage, and comprehensive managed care plan encounters. Claims volumes (across all claims types) and CMC plan encounter volumes are assessed for header and line records, as well as the average number of line records per header, compared against the national median. Service users are assessed by analyzing the percentage of total beneficiaries in a jurisdiction with claims data in the TAF RIFs. Any of these assessments may indicate problems with how jurisdictions reported claims, encounter, or eligibility data and will be critical in determining included/excluded beneficiaries. CMC-specific topics will inform the exclusion of managed care beneficiaries from the Medicaid/CHIP ETL. We have chosen not to use the benchmarking IP stays topic as an exclusion criterion for the Medicaid/CHIP ETL since this assessment compares actual TAF inpatient stay volumes to the expected number of stays according to external benchmark data from the Healthcare Cost and Utilization Project (HCUP). Instead, we will prioritize internal TAF assessments (i.e., those assessments looking strictly within the TAF data) when recommending exclusions based upon inpatient data concerns.

Table 9. Claim file completeness importance to inclusion in Medicaid/CHIP ETL

DQ Topic: Claims File Completeness		Importance to Inclusion in Medicaid/CHIP ETL
Benchmarking Inpatient Stays - IP		Limited importance
Service Users	IP OT RX	Very important
Claims Volume	IP LT OT RX	Very important
CMC Plan Encounters	IP LT OT RX	Very important

e. Service Use Information

These topics assess missing, incomplete, or invalid data for variables in claims files related to encounters, such as diagnosis and procedure codes, types of service, place of service, type of bill, and admission/discharge dates. Diagnosis codes, procedure codes, type of service codes, and admission dates are all critical components in the SCDM, so data quality issues with any of these topics may require jurisdiction/year/plan exclusion. However, procedure codes are not always expected in outpatient institutional claims, so we would not want to exclude a jurisdiction/year/plan based upon this specific topic. Problems regarding place of service and type of bill should be monitored, but the information contained in these variables might be found elsewhere in the TAF RIF data. Problems with discharge dates also require monitoring, and missing dates will need to be dealt with at a beneficiary level. DQ Atlas assessments for discharge date topics, however, will not be used to exclude entire jurisdiction/year/plans.

Table 10. Service use information importance to inclusion in Medicaid/CHIP ETL

DQ Topic: Service Use Information		Importance to Inclusion in Medicaid/CHIP ETL
Bill Types[3]	IP LT OT	Limited importance
Generic Indicator - RX		Not important
Admission Dates	IP LT	Very important
Discharge Dates	IP LT	Limited importance
Diagnosis Codes	IP LT OT	Very important
Types of Service	IP OT LT RX	Very important Very important Very important Not important
Place of Service		Limited importance
Procedure Codes	OT Professional IP OT Institutional	Very important Very important Limited importance

f. Provider Information

These topics assess incomplete and incorrect provider variables (e.g., National Provider Identifier [NPI] and provider type) found in the TAF RIF claims files. Missingness across provider NPIs should be monitored, but incomplete data is not necessarily reason for exclusion from the ETL. Provider type and hospital type assessments are not critical since other fields may be used to supplement missing or "other" type values.

Table 11. Provider information importance to inclusion in Medicaid/CHIP ETL

DQ Topic: Provider Information		Importance to Inclusion in Medicaid/CHIP ETL
Hospital Type - IP		Not important
Provider NPIs	Servicing Provider NPI - OT Prescribing Provider NPI - RX Dispensing Provider NPI – RX	Limited importance
Billing Provider NPIs	IP LT	Limited importance

[3] "Type of Bill" is a specific topic assessed by the DQ Atlas concerning the BILL_TYPE_CD variable found in the header records of IP, LT, and OT files. This variable contains data related to type of facility and type of care.

	OT RX	
Billing Provider Types	IP LT OT	Not important

3. Data Quality Importance to Inclusion in Medicaid/CHIP ETL

Certain DQ Atlas assessment topics are more relevant to inclusion in the Medicaid/CHIP ETL than others. We categorized each DQ topic as either "very important," "limited importance," or "not important" based on the perceived impact that poor data quality in that domain would have on the usability and integrity of the Medicaid/CHIP ETL. Table 12 summarizes the importance that we assigned to the set of DQ topics.

Exclusion recommendations are based upon the Release 2 assessments for all years. Summaries, backgrounds, and assessment methodologies for each DQ Atlas topic can be found on the DQ Atlas site at https://www.medicaid.gov/dq-atlas/landing/topics/info.

Table 12. Summary of DQ topics by importance to inclusion in Medicaid/CHIP ETL

Assigned Importance	Data Quality Topic
Very important	• Enrollment Spans • Eligibility Codes: Dual eligibility, restricted benefits • Service Users • Admission Dates • Claims Volume • Diagnosis Codes • Types of Service: IP, OT, LT • CMC Plan Encounters • Procedure Codes: OT Professional, IP
Limited importance	• Demographics: Age, gender, ZIP code, race, ethnicity • Eligibility Codes: CHIP • Benchmarking Inpatient Stays • Bill Types • Provider NPIs • Billing Provider NPIs • Discharge Dates • Place of Service • Procedure Codes: OT Institutional
Not important	• Overall Enrollment Plan Benchmarking • Demographics: Income • Eligibility Codes: Eligibility Group • Hospital Type - IP • Billing Provider Types • Generic Indicator - RX • Types of Service: RX

III. ETL Characteristics

This section describes the DPHS plans and recommendations for Medicaid/CHIP ETL 1.

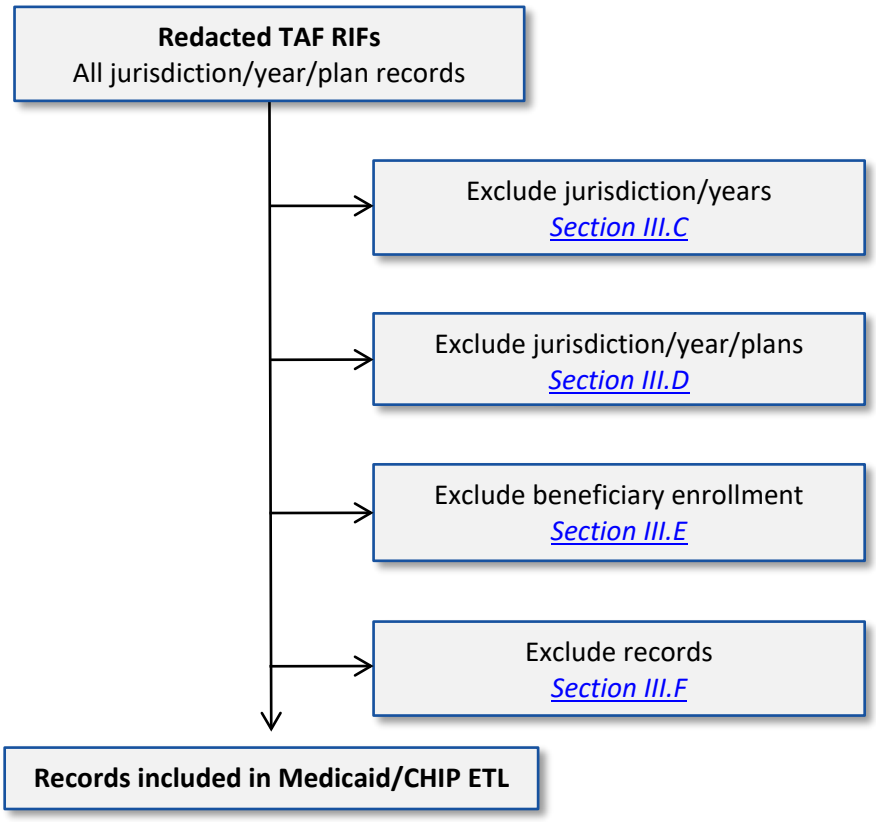
A. Source Data

Medicaid/CHIP ETL 1 will utilize the following files from the CY (Release 2) 2014–2018 100% redacted TAF RIFs: Demographic and Eligibility, Inpatient utilization, Long Term Care utilization, Other Services utilization, and Pharmacy utilization. All of these files have been approved for use by DPHS under CMS DUA RSCH-2017-50370.
Preliminary CY 2019 TAF RIF files were released by CMS in December 2020, with a final release in September 2021, but these preliminary data files are not approved for use on the DUA listed above and will not be utilized in Medicaid/CHIP ETL 1. The Annual Provider file is also not approved for use on the DUA listed above and will not be utilized in Medicaid/CHIP ETL 1. Both the Provider file and the preliminary TAF RIF releases may be considered for future inclusion in Medicaid/CHIP ETLs.

B. General Inclusion Criteria

All jurisdiction/year/plans and beneficiaries are to be included in the Medicaid/CHIP ETL except when subject to any of the following recommended exclusions. After all TAF RIF data are read-in, exclusions will be applied to full jurisdiction/years, jurisdiction/year/managed-care plans, beneficiaries, and claim records, according to the flowchart below. Exclusions applied to each exclusion level are described in detail in the sections following the flowchart.

Figure 1. Flowchart of record selection for inclusion in the Medicaid/CHIP ETL



C. Jurisdiction/Year-Level Exclusions

By combining the DQ Atlas assessment data with our ranking of the importance of each DQ topic, we were able to develop a repeatable method of programmatically determining which jurisdiction/years to exclude from a given Medicaid/CHIP ETL. We anticipate using this method to generate a control file that instructs the ETL programs which jurisdiction/years to include in a given Medicaid/CHIP ETL.

Recommendation: We recommend excluding any jurisdiction/year with an "Unusable" assessment for any single topic deemed "Very Important" to inclusion in the Medicaid/CHIP ETL in Section II.B. Any other concern level (low, medium, or high) assessment for a DQ Atlas topic would not lead to a recommendation for exclusion. Unclassified assessments occur when a topic is not relevant to a certain jurisdiction or year and are similarly not considered when recommending jurisdiction/year exclusions. A "Very Important" determination regarding a DQ Atlas topic is an attempt to specifically decide whether any widespread data quality issues might exist that would require all of the data for a jurisdiction/year/plan to be excluded entirely from the Medicaid/CHIP ETL and should not be construed as an indicator of a topic or variable's importance to the Sentinel project or SCDM. These decisions would yield the numbers of fully excluded jurisdictions in Table 13.

D. Jurisdiction/Year/Plan-Level Exclusions

Specific to CMC plans, the DQ Atlas assesses the claim file completeness of CMC plan encounters for the IP, LT, OT, and RX files. These assessments are important in determining whether a jurisdiction/year may have data issues specifically related to managed care plans.

Recommendation: For jurisdiction/years listed as having unusable data for CMC plans (as assessed in the DQ Atlas by examining header and line record claims volumes), and no other problems that would lead to the entire jurisdiction/year being excluded, we recommend excluding only managed care beneficiaries, rather than all beneficiaries. Any other concern level (low, medium, or high) assessment for a CMC plan encounter topic would not lead to a recommendation for exclusion. Unclassified assessments occur for these topics in jurisdiction/years without a managed care plan and are similarly not considered when recommending jurisdiction/year/plan exclusions. These decisions would yield the numbers of jurisdictions with CMC plans excluded in Table 13.

Table 13. Number of jurisdiction/plans recommended to be included and excluded in Medicaid/CHIP ETL1, by year

TAF RIF Year	Jurisdictions Fully Included	Jurisdictions Fully Excluded	Jurisdictions with CMC Plans Excluded (FFS plans still included)
2014	14	1	4
2015	23	4	4
2016	39	10	3
2017	44	7	2
2018	41	9	3

A full list of each jurisdiction's inclusion/exclusion status by year/plan is shown in Appendix D. This appendix also indicates which data quality topic(s) led to the exclusion of jurisdictions each given year.

The inclusion/exclusion status, based on the DQ Atlas assessments, of a given jurisdiction must be re-determined whenever there is a new data release for a year because data quality is specific to each CY TAF RIF release.

E. Beneficiary-Level Exclusions

Recommendation: We recommend excluding beneficiaries during periods of enrollment with limited benefits enrollment and during periods of enrollment when dual eligibility status is missing. Not all of the healthcare utilization for these beneficiaries will be captured in the TAF RIF data, since Medicare is the primary payer for dual-eligible beneficiaries. Dual eligibility codes in the TAF RIF can be found at https://resdac.org/cms-data/variables/medicare-medicaid-dual-eligibility-code-latest-year.

Table 14. Yearly percentage of beneficiaries by dual Medicare/Medicaid enrollment status

Dual Eligibility Code	Eligibility	2014	2015	2016	2017	2018
Missing	-	7.4%	6.3%	9.2%	5.5%	5.2%
00	Medicaid only	78.4%	80.2%	78.2%	81.7%	81.7%
01 - 10	Dual eligible	14.2%	13.5%	12.6%	12.8%	13.1%

Recommendation: We recommend excluding beneficiaries during periods of enrollment with limited benefits coverage and during periods of enrollment when benefit coverage information is missing. We expect to see complete healthcare utilization information only for beneficiaries with full-scope or comprehensive coverage. This exclusion ensures that only beneficiaries with both medical and drug coverage are included in the Medicaid/CHIP ETL 1, since there is no simple way in the TAF RIF data to determine coverage separately (unlike the Part A/B and Part D coverage indicators found in Medicare). Scope of benefits codes in the TAF RIF can be found at https://resdac.org/cms-data/variables/scope-medicaid-or-chip-benefits-latest-year.

Table 15. Yearly percentage of beneficiaries by benefits package

Restricted Benefits Code	Scope	2014	2015	2016	2017	2018
Missing	-	3.1%	4.9%	3.6%	3.3%	2.9%
0	-	<0.1%	<0.1%	0.2%	<0.1%	<0.1%
1	Full	83.9%	84.2%	81.6%	82.0%	82.7%
2	Limited	0.4%	1.0%	1.9%	1.9%	1.9%
3	Limited	4.5%	4.0%	3.1%	3.3%	3.4%
4	Comprehensive[4]	0.5%	0.8%	0.5%	0.5%	0.5%
5	Limited	1.8%	1.0%	1.1%	1.0%	1.1%
6	Limited	2.2%	2.1%	3.1%	3.0%	3.0%
7	Comprehensive	3.5%	2.0%	4.8%	5.0%	4.5%
A	Comprehensive	<0.1%	<0.1%	<0.1%	<0.1%	<0.1%
C	Limited	-	<0.1%	<0.1%	<0.1%	<0.1%
D	Comprehensive	<0.1%	<0.1%	<0.1%	<0.1%	<0.1%
E	Limited	-	-	<0.1%	<0.1%	<0.1%

F. Record-Level Exclusions

Recommendation: There are records in the TAF RIF data labeled as capitated payment, supplemental payment, and service tracking claims. All of these "claims" are specifically related to payments/expenditures and do not reflect specific healthcare services. We recommend excluding these records from the ETL since all SCDM-relevant information already exists in other records. We also recommend excluding records with missing information about claim type.

Table 16. Percentage of Medicaid claim type code by data file for 2018

Medicaid Claim Type Code	Included/Excluded	IP	LT	OT	RX
Missing	Excluded	<0.1%	<0.1%	<0.1%	-
FFS Claim	Included	34.7%	61.8%	20.4%	24.1%
Capitated Payment	Excluded	-	-	52.2%	<0.1%
Managed Care Encounter	Included	64.6%	26.5%	26.7%	75.9%
Service Tracking Claim	Excluded	0.1%	<0.1%	0.3%	<0.1%
Supplemental Payment	Excluded	0.6%	11.6%	0.4%	<0.1%

[4] Category 4 beneficiaries are only entitled to restricted benefits for pregnancy-related services, but all jurisdictions except Arkansas, Idaho, and South Dakota elect to offer a comprehensive benefits package to these individuals. Beneficiaries from these three states will be excluded for this benefits category.

Specific to Illinois are records labeled "adjustment claims." The Illinois Medicaid claims processing system is unable to void and replace claims, which causes the state to submit both original and adjustment claims to CMS, rather than one single "final action" claim, as is done in other jurisdictions. For the purposes of generating the Medicaid/CHIP ETL, only the original claim data will be used for Illinois Medicaid claims. A full accounting of how CMS recommends using Illinois TAF RIF claims data can be found at https://resdac.org/TAF-data-quality-resources/TAFTechDoc-IL. The recommendations for Medicaid/CHIP ETL 1 are aligned with those suggested by CMS.

G. Mother-Infant Linkage

We will attempt to link all births in the Medicaid/CHIP data, primarily through the use of the encrypted T-MSIS case number (MSIS_CASE_NUM in the TAF RIF). This number is a jurisdiction-assigned unique ID used to identify a Medicaid/CHIP case, which often acts as a family identifier. Prior research on mother-infant linkage in the Medicaid MAX data from 2000 to 2007 found that linkage rates varied by state, due to differences in how states assigned case numbers to beneficiaries. Appendix C shows these prior results. Their overall linkage rate for inpatient deliveries across the 50 states was 55.6%, but fully 40 states had a linkage rate over 70%. The overall linkage rate for outpatient deliveries across the 50 states was 23.8%, and only four states had a linkage rate over 70%. It is possible that states have changed how they assign case numbers in the years since this study was conducted, but we anticipate similar results from the more recent TAF RIF data, with linkage being straightforward within jurisdictions that assign the same T-MSIS case number to both the mother and infant for a birth, and more difficult within jurisdictions that do not utilize the MSIS_CASE_NUM for this purpose. We will evaluate other approaches to performing the mother-infant linkage, if needed.

H. ETL Size Estimate

In order to estimate the size of the Medicaid/CHIP ETL 1 SCDM tables, we used the size of the Medicare SCDM tables (annual or all-year, as appropriate) scaled by the relative sizes of the Medicaid/CHIP and Medicare source tables that feed those SCDM tables. These scaling factors will be imperfect, since the set of tables in both data sources does not align perfectly but should still be useful. The calculations are shown in Table 17. In the end, we believe all the Medicaid/CHIP ETL 1 SCDM tables will require just under 1 TB of disk space.

Table 17. Size estimate calculations for the Medicaid/CHIP ETL 1 SCDM tables

Year(s)	SCDM Table	SCDM Table Size, Medicare (GB)	TAF RIF: Medicare Source Table Size Ratio	SCDM Table Size, Medicaid/CHIP (GB, expected)
Combined	ENROLLMENT	2	150%	3
	DEMOGRAPHIC	2	150%	3
	DEATH	0.23	30%	0.07
	FACILITY	0.01	100%	0.01
	PROVIDER	0.02	100%	0.02
2014	DISPENSING	32	9%	2.9
	ENCOUNTER	36	14%	5.0
	DIAGNOSIS	120	14%	16.8
	PROCEDURE	110	14%	15.4
2015	DISPENSING	32	26%	8.3

Year(s)	SCDM Table	SCDM Table Size, Medicare (GB)	TAF RIF: Medicare Source Table Size Ratio	SCDM Table Size, Medicaid/CHIP (GB, expected)
	ENCOUNTER	36	37%	13.3
	DIAGNOSIS	120	37%	44.4
	PROCEDURE	110	37%	40.7
2016	DISPENSING	32	57%	18.2
	ENCOUNTER	36	89%	32.0
	DIAGNOSIS	120	89%	106.8
	PROCEDURE	110	89%	97.9
2017	DISPENSING	32	57%	18.2
	ENCOUNTER	36	90%	32.4
	DIAGNOSIS	120	90%	108.0
	PROCEDURE	110	90%	99.0
2018	DISPENSING	32	57%	18.2
	ENCOUNTER	36	94%	33.8
	DIAGNOSIS	120	94%	112.8
	PROCEDURE	110	94%	103.4
			TOTAL GB	**934**

Appendix A: Full-Scope and Comprehensive Medicaid Benefits

Benefit Type	Full-Scope Benefits	Comprehensive Benefits
Hospital Services	• Inpatient hospital services • Outpatient hospital services	• Emergency services • Hospitalization
Clinic Services	• Rural health clinic services • Federally qualified health center services • Clinic services	• Ambulatory patient services • Pediatric services
Long-term Inpatient Care	• Nursing Facility Services • Health Homes for Enrollees with Chronic Conditions – Section 1945	
Home Health	• Home health services • Private duty nursing services • Personal Care	
Provider	• Physician services • Other practitioner services	
Maternity	• Family planning services • Nurse Midwife services • Certified Pediatric and Family Nurse Practitioner services • Freestanding Birth Center services (when licensed or otherwise recognized by the state) • Tobacco cessation counseling for pregnant women	• Pregnancy, maternity, and newborn care
Transportation	• Transportation to medical care	
Drug	• Prescription Drugs	• Prescription drugs
Mental Health	• Services for individuals aged 65 or older in an Institution for Mental Disease (IMD) • Inpatient psychiatric services for individuals under age 21	• Mental health and substance use disorder services, including behavioral health treatment
Diagnostic, Screening, and Rehabilitation	• Laboratory and X-ray services • Early and Periodic Screening, Diagnostic, and Treatment Services • Physical therapy (EPSDT) • Occupational therapy • Chiropractic services • Speech, hearing and language disorder services • Respiratory care services • Podiatry services • Tuberculosis Related Services • Other diagnostic, screening, preventive and rehabilitative services • Prosthetics	• Rehabilitative and habilitative services and devices • Laboratory services

Benefit Type	Full-Scope Benefits	Comprehensive Benefits
Dental	• Dental Services • Dentures	• Pediatric oral care
Vision	• Optometry services • Eyeglasses	• Pediatric vision care
Miscellanea	• Hospice • Case management • State Plan Home and Community Based Services- 1915(i) • Self-Directed Personal Assistance Services- 1915(j) • Community First Choice Option- 1915(k) • Other services approved by the Secretary	• Preventive and wellness services and chronic disease management

Appendix B: Jurisdictions Included in the TAF RIF by Year

Jurisdiction	2014	2015	2016	2017	2018
AK	X	X	X	X	X
AL	X	X	X	X	X
AR			X	X	X
AZ		X	X	X	X
CA			X	X	X
CO	X	X	X	X	X
CT			X	X	X
DC	X	X	X	X	X
DE	X	X	X	X	X
FL	X	X	X	X	X
GA			X	X	X
HI		X	X	X	X
IA			X	X	X
ID			X	X	X
IL	X	X	X	X	X
IN		X	X	X	X
KS	X	X	X	X	X
KY		X	X	X	X
LA			X	X	X
MA		X	X	X	X
MD	X	X	X	X	X
ME	X	X	X	X	X
MI			X	X	X
MN			X	X	X
MO			X	X	X
MS			X	X	X
MT	X	X	X	X	X
NC	X	X	X	X	X
ND	X	X	X	X	X
NE	X	X	X	X	X
NH	X	X	X	X	X
NJ			X	X	X
NM	X	X	X	X	X
NV	X	X	X	X	X
NY			X	X	X
OH		X	X	X	X
OK		X	X	X	X
OR			X	X	X
PA			X	X	X
PR		X	X	X	X
RI	X	X	X	X	X
SC		X	X	X	X
SD			X	X	X

Jurisdiction	2014	2015	2016	2017	2018
TN			X	X	X
TX		X	X	X	X
UT			X	X	X
VA		X	X	X	X
VI				X	X
VT			X	X	X
WA		X	X	X	X
WI	X	X	X	X	X
WV			X	X	X
WY			X	X	X

Appendix C: Published Example of Successful Mother-Infant Linkage

The following is adapted from Palmsten, et. al., which described mother-infant linkage using 2000–2007 Medicaid Analytic eXtract (MAX) data. MAX files were the precursor to TAF RIFs. This example is provided to illustrate the feasibility of mother-infant linkage in Medicaid claims. The provided counts are the number of deliveries, while percentage linked is the percentage of deliveries linked to at least one infant.

	Inpatient Deliveries		Outpatient Deliveries	
State	Count	% Linked	Count	% Linked
AK	31,553	83.4	7,383	41.7
AL	335,584	83.9	3,177	54.0
AR	133,629	21.7	111,902	2.0
CA	905,520	78.9	417,779	38.3
CO	116,192	79.3	12,198	47.0
CT	20,782	0.0	1,970	0.0
DC	3,927	80.4	915	26.4
DE	10,149	93.2	23,841	70.6
FL	543,810	71.6	443,337	7.1
GA	415,916	35.9	100,357	19.1
HI	27,633	77.6	11,758	65.1
IA	86,326	73.0	21,308	55.8
ID	61,086	74.6	8,050	46.8
IL	458,715	72.3	156,915	13.5
IN	222,255	91.7	178,780	18.0
KS	80,230	90.9	12,787	65.1
KY	175,466	81.1	49,526	40.8
LA	270,481	88.2	138,652	12.6
MA	81,289	89.4	17,975	33.4
MD	1,470,560	9.3	13,016	10.8
ME	15,566	93.0	3,909	25.6
MI	177,475	70.7	71,736	26.6
MN	91,875	93.8	70,072	67.2
MO	253,912	54.8	39,871	31.2
MS	184,337	90.2	93,213	11.6
MT	27,084	0.0	4,093	0.0
NC	484,133	17.3	44,800	9.3
ND	16,679	96.0	3,334	36.2
NE	25,261	79.7	8,158	54.6
NH	22,699	94.0	3,519	46.6
NJ	89,215	84.4	88,993	53.4
NM	93,901	85.4	28,986	72.7
NV	31,258	89.5	11,527	75.4
NY	642,194	52.1	107,573	34.3
OH	252,528	94.2	154,680	17.1
OK	156,994	87.9	37,478	55.5
OR	98,716	88.0	23,427	60.2
PA	111,772	93.0	17,485	31.2
RI	46,047	91.1	8,280	31.4
SC	199,463	16.1	116,129	1.7
SD	30,083	93.8	3,170	50.1
TN	168,714	80.0	71,995	69.3
TX	831,729	9.1	486,918	1.0
UT	51,991	95.7	9,264	81.6
VA	151,876	87.1	62,204	58.6
VT	18,386	91.9	2,807	56.8
WA	117,179	84.4	52,704	63.9
WI	154,097	94.0	33,966	36.9
WV	40,739	85.8	6,237	47.5
WY	20,999	93.9	4,114	39.4
Total	***10,058,005***	***55.6***	***3,402,268***	***23.8***

Appendix D: Recommendations by Jurisdiction and Year

The table below indicates which jurisdiction/year/plans will be included (✓) in Medicaid/CHIP ETL 1 and which will be excluded (×). We will include all jurisdiction/year/plans, unless they have "Unusable" data for any DQ Atlas topic categorized as being "Very Important," per the data quality criteria described in Section III.C and III.D. The reasons for exclusion are noted in parentheses below and are indexed in table footnotes. This table only speaks to jurisdiction/year/plan-level exclusions. Within included jurisdiction/year/plans, additional exclusions at the beneficiary-level and record-level will be made according to the criteria described in Sections III.E and III.F, respectively.

✓ indicates data included for this jurisdiction/year/plan
× indicates data excluded for this jurisdiction/year/plan. Reasons for exclusion are denoted in parentheses and indexed below the table.
– indicates no data in TAF for this jurisdiction/year/plan
Y indicates that the years of included data are contiguous.
N indicates that the years of included data are not continuous.

Jurisdiction	FFS or CMC Plan	2014	2015	2016	2017	2018	Years Included	Continuity
AK	FFS	✓	✓	✓	✓	✓	5	Y
	CMC	✓	✓	✓	✓	✓	5	Y
AL	FFS	× (a)	× (a)	× (a)	× (a)	× (a)	0	-
	CMC	× (a)	× (a)	× (a)	× (a)	× (a)	0	-
AR	FFS	–	–	× (a)	✓	✓	2	Y
	CMC	–	–	× (a)	✓	✓	2	Y
AZ	FFS	–	✓	✓	✓	✓	4	Y
	CMC	–	✓	✓	✓	✓	4	Y
CA	FFS	–	–	✓	✓	✓	3	Y
	CMC	–	–	✓	✓	✓	3	Y
CO	FFS	✓	✓	✓	✓	✓	5	Y
	CMC	× (b: LT, OT)	× (b: OT)	× (b: IP, OT)	✓	✓	2	Y
CT	FFS	–	–	✓	✓	✓	3	Y
	CMC	–	–	✓	✓	✓	3	Y

Jurisdiction	FFS or CMC Plan	2014	2015	2016	2017	2018	Years Included	Continuity
DC	FFS	✓	× (c)	✓	✓	✓	4	N
	CMC	✓	× (c)	✓	✓	✓	4	N
DE	FFS	✓	✓	✓	✓	✓	5	Y
	CMC	× (b: IP, OT)	× (b: IP, OT)	✓	✓	✓	3	Y
FL	FFS	✓	✓	✓	✓	✓	5	Y
	CMC	✓	✓	✓	✓	× (b: OT)	4	Y
GA	FFS	–	–	✓	✓	× (d: IP)	2	Y
	CMC	–	–	✓	✓	× (d: IP)	2	Y
HI	FFS	–	✓	✓	✓	✓	4	Y
	CMC	–	✓	✓	✓	✓	4	Y
IA	FFS	–	–	✓	✓	✓	3	Y
	CMC	–	–	✓	✓	✓	3	Y
ID	FFS	–	–	✓	✓	✓	3	Y
	CMC	–	–	✓	✓	× (b: IP, OT)	2	Y
IL	FFS	✓	✓	× (e: LT)	✓	✓	4	N
	CMC	✓	✓	× (e: LT)	✓	✓	4	N
IN	FFS	–	✓	✓	✓	✓	4	Y
	CMC	–	✓	✓	✓	✓	4	Y
KS	FFS	✓	✓	✓	✓	✓	5	Y
	CMC	✓	✓	✓	✓	✓	5	Y

Jurisdiction	FFS or CMC Plan	2014	2015	2016	2017	2018	Years Included	Continuity
KY	FFS	–	× (c)	× (c, f: IP)	× (c)	× (c)	0	-
	CMC	–	× (c)	× (c, f: IP)	× (c)	× (c)	0	-
LA	FFS	–	–	✓	✓	✓	3	Y
	CMC	–	–	✓	✓	✓	3	Y
MA	FFS	–	✓	✓	✓	✓	4	Y
	CMC	–	× (b: IP)	× (b: IP)	× (b: IP)	× (b: IP)	0	-
MD	FFS	✓	✓	× (e: IP, LT, OT; f: IP)	× (e: IP, LT, OT; f: IP)	× (f: IP)	2	Y
	CMC	✓	✓	× (e: IP, LT, OT; f: IP)	× (e: IP, LT, OT; f: IP)	× (f: IP)	2	Y
ME	FFS	✓	✓	✓	✓	✓	5	Y
	CMC	✓	✓	✓	✓	✓	5	Y
MI	FFS	–	–	× (a)	✓	✓	2	Y
	CMC	–	–	× (a)	✓	✓	2	Y
MN	FFS	–	–	✓	✓	✓	3	Y
	CMC	–	–	✓	✓	✓	3	Y
MO	FFS	–	–	✓	✓	✓	3	Y
	CMC	–	–	✓	✓	✓	3	Y
MS	FFS	–	–	× (g: RX)	✓	✓	2	Y

Jurisdiction	FFS or CMC Plan	2014	2015	2016	2017	2018	Years Included	Continuity
	CMC	–	–	× (g: RX)	✓	✓	2	Y
MT	FFS	✓	✓	✓	✓	✓	5	Y
	CMC	✓	✓	✓	✓	✓	5	Y
NC	FFS	✓	✓	✓	✓	✓	5	Y
	CMC	✓	✓	✓	✓	✓	5	Y
ND	FFS	✓	✓	✓	✓	✓	5	Y
	CMC	× (b: LT, OT)	× (b: LT, OT)	× (b: OT)	✓	✓	2	Y
NE	FFS	✓	✓	✓	✓	✓	5	Y
	CMC	✓	✓	✓	✓	✓	5	Y
NH	FFS	✓	✓	✓	✓	✓	5	Y
	CMC	✓	✓	✓	✓	✓	5	Y
NJ	FFS	–	–	✓	✓	✓	3	Y
	CMC	–	–	✓	✓	✓	3	Y
NM	FFS	✓	✓	✓	✓	✓	5	Y
	CMC	✓	✓	✓	✓	✓	5	Y
NV	FFS	✓	✓	✓	✓	✓	5	Y
	CMC	✓	✓	✓	✓	✓	5	Y
NY	FFS	–	–	✓	✓	✓	3	Y
	CMC	–	–	✓	✓	✓	3	Y
OH	FFS	–	✓	✓	✓	× (h: IP, OT)	3	Y
	CMC	–	✓	✓	✓	× (h: IP, OT)	3	Y
OK	FFS	–	✓	✓	✓	✓	4	Y
	CMC	–	✓	✓	✓	✓	4	Y

Jurisdiction	FFS or CMC Plan	2014	2015	2016	2017	2018	Years Included	Continuity
OR	FFS	–	–	✓	✓	✓	3	Y
	CMC	–	–	✓	✓	✓	3	Y
PA	FFS	–	–	✓	✓	✓	3	Y
	CMC	–	–	✓	✓	✓	3	Y
PR	FFS	–	✓	× (f: IP)	× (c, f: IP)	✓	2	N
	CMC	–	✓	× (f: IP)	× (c, f: IP)	✓	2	N
RI	FFS	✓	✓	✓	✓	× (g: IP)	4	Y
	CMC	× (b: IP)	✓	✓	× (b: IP)	× (b: IP, g: IP)	2	Y
SC	FFS	–	✓	× (e: LT)	× (e: LT)	× (e: LT)	1	Y
	CMC	–	✓	× (e: LT)	× (e: LT)	× (e: LT)	1	Y
SD	FFS	–	–	✓	✓	✓	3	Y
	CMC	–	–	✓	✓	✓	3	Y
TN	FFS	–	–	✓	× (e: IP)	× (e: IP)	1	Y
	CMC	–	–	✓	× (e: IP)	× (e: IP)	1	Y
TX	FFS	–	✓	✓	✓	✓	4	Y
	CMC	–	✓	✓	✓	✓	4	Y
UT	FFS	–	–	× (a, d: LT)	× (a, f: OT)	× (a, f: OT)	0	-

Jurisdiction	FFS or CMC Plan	2014	2015	2016	2017	2018	Years Included	Continuity
	CMC	–	–	× (a, d: LT)	× (a, f: OT)	× (a, f: OT)	0	-
VA	FFS	–	✓	✓	✓	✓	4	Y
	CMC	–	✓	✓	✓	✓	4	Y
VI	FFS	–	–	–	✓	✓	2	Y
	CMC	–	–	–	✓	✓	2	Y
VT	FFS	–	–	✓	✓	✓	3	Y
	CMC	–	–	✓	✓	✓	3	Y
WA	FFS	–	✓	✓	✓	✓	4	Y
	CMC	–	✓	✓	✓	✓	4	Y
WI	FFS	✓	× (d: LT)	✓	✓	✓	4	**N**
	CMC	✓	× (d: LT)	✓	✓	✓	4	**N**
WV	FFS	–	–	✓	✓	✓	3	Y
	CMC	–	–	✓	✓	✓	3	Y
WY	FFS	–	–	✓	✓	✓	3	Y
	CMC	–	–	✓	✓	✓	3	Y

Index of DQ Atlas topics used for jurisdiction/year/plan exclusion because assessed as unusable:

a) Dual Eligibility Code
b) Comprehensive Managed Care Plan Encounters
c) Number of Enrollment Spans
d) Admission Date
e) Diagnosis Codes
f) Procedure Codes
g) Claims Volume
h) Type of Service

Index of claim types:

IP Inpatient
LT Long-Term Care
OT Other Services
RX Pharmacy

Summary of recommendations:

Jurisdictions with 5 Years of Continuous Inclusion[5]	AK, KS, ME, MT, NC, NE, NH, NM, NV
Jurisdictions with < 5 Years of Continuous Inclusion	AR, AZ, CA, CO, CT, DE, FL, GA, HI, IA, ID, IN, LA, MD, MI, MN, MO, MS, ND, NJ, NY, OH, OK, OR, PA, RI, SC, SD, TN, TX, VA, VI, VT, WA, WV, WY
Jurisdictions with Years of Intervening Exclusions	DC, IL, PR, WI
Jurisdictions without Any Inclusion	AL, KY, MA, UT

[5] In this table, an Included Year is counted only when both FFS and CMC are included.

Bibliography

2019 Preliminary Medicaid and CHIP Data Now Available (2020) *ResDAC*. Available at: https://resdac.org/cms-news/2019-preliminary-medicaid-and-chip-data-now-available (Accessed: 29 September 2021).

Barrett, A. *et al.* (2019) *Production of the TAF Research Identifiable Files (RIFs)*. 9010. Baltimore, MD: Centers for Medicare and Medicaid Services, p. 42. Available at: https://www.medicaid.gov/dq-atlas/downloads/supplemental/9010_Production_of_TAF_RIF.pdf.

Chowdhury, S., Kuncaitis, S. and Collier, J. (2020) *TAF Technical Guidance: How to Use Illinois Claims Data*. Baltimore, MD: Centers for Medicare and Medicaid Services, p. 27. Available at: https://resdac.org/TAF-data-quality-resources/TAFTechDoc-IL.

Christensen, A. *et al.* (2020) *TAF Technical Guidance: Claims Files*. Baltimore, MD: Centers for Medicare and Medicaid Services, p. 32.

'CMS Chronic Condition Warehouse (CCW) Data Dictionary TAF Annual Provider File - Base File' (2021). Chronic Condition Warehouse. Available at: https://www2.ccwdata.org/documents/10280/19022436/record-layout-taf-annual-provider.xlsx.

'CMS Chronic Condition Warehouse (CCW) Data Dictionary TAF Demographic and Enrollment - Base File' (2020). Chronic Condition Warehouse. Available at: https://www2.ccwdata.org/documents/10280/19022436/record-layout-taf-demographic-eligibility.xlsx.

'CMS Chronic Condition Warehouse (CCW) Data Dictionary TAF Inpatient (IP) Claims Files' (2019). Chronic Condition Warehouse. Available at: https://www2.ccwdata.org/documents/10280/19022436/record-layout-taf-claims.xlsx.

CMS Guidance: Reporting Restricted-Benefits-Code in the T-MSIS Eligible File (2020) *Medicaid.gov*. Available at: https://www.medicaid.gov/medicaid/data-and-systems/macbis/tmsis/tmsis-blog/entry/53953 (Accessed: 29 September 2021).

CMS Medicaid Data: Linking Mothers to Infants (2018) *ResDAC*. Available at: https://resdac.org/articles/cms-medicaid-data-linking-mothers-infants (Accessed: 29 September 2021).

Codebook: T-MSIS Analytic Files (TAF) Annual Provider (APR) File. version 1.0 (2021). Chronic Condition Warehouse. Available at: https://www2.ccwdata.org/documents/10280/19022436/codebook-taf-annual-provider.pdf.

Codebook: T-MSIS Analytic Files (TAF) Claims Research Identifiable Files (RIFs). version 1.1 (2021). Chronic Condition Warehouse. Available at: https://www2.ccwdata.org/documents/10280/19022436/codebook-taf-claims.pdf.

Codebook: T-MSIS Analytic Files (TAF) Demographic and Eligibility Research Identifiable Files (RIFs). version 1.2 (2020). Chronic Condition Warehouse. Available at: https://www2.ccwdata.org/documents/10280/19022436/codebook-taf-demographic-eligibility.pdf.

Contractors (2021) *CMS.gov*. Available at: https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/Data-Disclosures-Data-Agreements/DUA-Contractors (Accessed: 28 September 2021).

Czajka, J.L. (2012) *Movement of Children Between Medicaid and CHIP, 2005-2007*. 4, p. 7. Available at: https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/Downloads/Medicaid_and_CHIP_Transitions.pdf (Accessed: 23 September 2021).

DQ Atlas (no date) *Medicaid.gov*. Available at: https://www.medicaid.gov/dq-atlas/welcome (Accessed: 24 September 2021).

Enterprise Privacy Policy Engine (EPPE) | CMS (no date) *CMS.gov*. Available at: https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/Data-Disclosures-Data-Agreements/Enterprise-Privacy-Policy-Engine (Accessed: 28 September 2021).

Essential Health Benefits - HealthCare.gov Glossary (no date) *HealthCare.gov*. Available at: https://www.healthcare.gov/glossary/essential-health-benefits/ (Accessed: 29 September 2021).

Introduction to the Transformed Medicaid Statistical Information System (T-MSIS) Analytic Files (TAF) (2021). Centers for Medicare and Medicaid Services. Available at: https://www.medicaid.gov/medicaid/data-and-systems/downloads/macbis/taf-introduction.pdf (Accessed: 24 September 2021).

Mandatory & Optional Medicaid Benefits (no date) *Medicaid.gov*. Available at: https://www.medicaid.gov/medicaid/benefits/mandatory-optional-medicaid-benefits/index.html (Accessed: 29 September 2021).

Medicaid and CHIP Business Information Solution (no date) *Medicaid.gov*. Available at: https://www.medicaid.gov/medicaid/data-systems/medicaid-and-chip-business-information-solution/index.html (Accessed: 28 September 2021).

Medicaid and CHIP in the Territories (2021). Washington, DC: Medicaid and CHIP Payment and Access Commission, p. 12. Available at: https://www.macpac.gov/publication/medicaid-and-chip-in-the-territories/ (Accessed: 23 September 2021).

Medicaid and CHIP T-MSIS Analytic Files Data Release (2019) *CMS.gov*. Available at: https://www.cms.gov/newsroom/fact-sheets/medicaid-and-chip-t-msis-analytic-files-data-release (Accessed: 24 September 2021).

Natzke, B. *et al.* (2021) *Use of the Same Medicaid Identification Number for Mother and Newborn Services in 2016*. TAF research brief 7011. Baltimore, MD: Centers for Medicare and Medicaid

Services, p. 7. Available at: https://www.medicaid.gov/dq-atlas/downloads/supplemental/7011_Shared_Medicaid_IDs_2016.pdf (Accessed: 23 September 2021).

Nolan, L. *et al.* (2020) *TAF Technical Guidance: Annual Demographic and Eligibility (DE) File*. Baltimore, MD: Centers for Medicare and Medicaid Services, p. 80. Available at: https://www.resdac.org/TAF-data-quality-resources/TAFTechGuide-DEF.

Nolan, L. *et al.* (2021) *Identifying Beneficiaries with Full-Scope, Comprehensive, and Limited Benefits in the TAF*. 4151. Baltimore, MD: Centers for Medicare and Medicaid Services, p. 13. Available at: https://www.medicaid.gov/dq-atlas/downloads/supplemental/4151_Scope_of_Benefits.pdf.

Palmsten, K. *et al.* (2013) 'Harnessing the Medicaid Analytic eXtract (MAX) to Evaluate Medications in Pregnancy: Design Considerations', *PloS One*, 8(6), p. e67405. doi:10.1371/journal.pone.0067405.

Prescription Drugs (no date) *Medicaid.gov*. Available at: https://www.medicaid.gov/medicaid/prescription-drugs/index.html (Accessed: 29 September 2021).

Service Tracking Claims in Medicaid T-MSIS Analytic File (TAF) Research Identifiable Files (RIF) (2021) *ResDAC*. Available at: https://resdac.org/articles/service-tracking-claims-medicaid-t-msis-analytic-file-taf-research-identifiable-files-rif (Accessed: 29 September 2021).

Simon, T. (2021) 'TAF Comparison'.

Singer, E. *et al.* (2020) *T-MSIS Analytic Files Frequently Asked Questions*. Baltimore, MD: Centers for Medicare and Medicaid Services.

State Summary: 2016 TAF DQ Topical Snapshot (no date). Centers for Medicare and Medicaid Services, p. 24. Available at: https://requests.resdac.org/sites/resdac.umn.edu/files/TAF_DQ_Snapshot_2016_0.pdf (Accessed: 23 September 2021).

'TAF_DQ_by_State_by_Topic_2014-2018_external.xlsx' (2021). Center Medicare and Medicaid Services.

T-MSIS Analytic Files (TAF) Research Identifiable Files (RIFs). version 1.4 (2021). Chronic Condition Warehouse (Chronic Condition Warehouse User Guide). Available at: https://www2.ccwdata.org/documents/10280/19002246/ccw-taf-rif-user-guide.pdf.

Whitney, M. *et al.* (2020) *Unique Beneficiary Identifiers in the TAF RIF in 2016*. TAF data quality brief 3031. Baltimore, MD: Centers for Medicare and Medicaid Services, p. 9. Available at: https://www.medicaid.gov/dq-atlas/downloads/supplemental/3031_BENE_ID_2016.pdf (Accessed: 23 September 2021).

Winsor, D. (2021) 'Questions about TAF RIF data (DCH-627)'.