TAF Technical Documentation: Annual Provider (APR) File

August 2021

TABLE OF CONTENTS

I. Introduction 1
II. Structure and contents of the APR file 1
A. Overview 1
B. Structure 2
C. Linking the APR files 7
III. Sources of provider information in the TAF 9
IV. Linking APR to claims files 13
V. Linking TAF to external provider data 16
VI. Identifying and classifying providers 16
References 18

TABLES AND FIGURES

Table 1. Contents of each APR file 3

Figure 1. APR base and supplemental files: Relationships and recommended linking methods 8

Table 2. Information in claims and in the APR 10

Table 3. Recommended approach to link claim headers and/or lines to APR records 13

Table 4. Alternative approach to link claim headers and/or lines to APR records (NPI) 14

Table 5. Recommended approach to link claim headers to APR records for a specific location ID 15

I. Introduction

States administer Medicaid and the Children's Health Insurance Program (CHIP) and share the responsibility for funding and program administration with the federal government.[1] Along with enrollment and claims information, each state must report standardized data on all participating Medicaid and CHIP providers to the federal Transformed Medicaid Statistical Information System (T-MSIS). The Centers for Medicare and Medicaid Services (CMS) administers T-MSIS to improve quality of care and program integrity and to meet stakeholders' needs. Although states submit a wide variety of information to T-MSIS, the system is not optimized for conducting analyses. To meet this need, CMS constructs a research-optimized version of T-MSIS data called the T-MSIS Analytic Files (TAF).[2,3] Information on the completeness and quality of key TAF data elements can be accessed through *DQ Atlas,* available at https://www.medicaid.gov/dq-atlas/welcome. Specific topics relevant to each section of this technical documentation are noted in the footnotes.

The TAF are released as TAF Research Identifiable Files (RIF).[4] The TAF RIF includes (1) annual files containing demographic and eligibility information for all Medicaid- and CHIP-eligible beneficiaries; (2) monthly claims files containing service use and payment records; (3) annual files containing information for all managed care entities; and (4) annual files containing information for all providers authorized to deliver services to Medicaid and CHIP beneficiaries, as well as providers whose authorization is pending, denied, or terminated. The TAF Annual Provider file (APR), the provider component of the TAF that is also available as a TAF RIF, is the focus of this technical documentation.

II. Structure and contents of the APR file

A. Overview

The APR is a new file that CMS is making available publicly. Based on data submitted to CMS from states through the T-MSIS process, the APR captures information about all providers authorized by a state to render services to Medicaid and CHIP beneficiaries at any point in the calendar year, as well as providers whose authorization is pending, denied, or terminated, regardless of whether or how often the providers billed the state for services. Because providers contracted to deliver care through managed care plans are required to enroll as participating providers with the state, the APR generally captures these providers as well as those participating in Medicaid or CHIP on an FFS basis.

---

1 For more information about the Medicaid and CHIP programs, see the CMS website: https://www.Medicaid.gov.

2 For more information about TAF, see the T-MSIS Analytic Files website at: https://www.medicaid.gov/medicaid/data-systems/macbis/medicaid-chip-research-files/transformed-medicaid-statistical-information-system-t-msis-analytic-files-taf/index.html

3 More information on TAF production is available at: https://www.medicaid.gov/dq-atlas/downloads/supplemental/9010_Production_of_TAF_RIF.pdf

4 During the transformation into RIFs, some TAF data elements are suppressed, changed, or renamed. For more details on the difference between the pre-RIF and RIF version of the TAF data, including a crosswalk of variable names, see "Production of the TAF Research Identifiable Files (RIFs)," available in the Resources section of *DQ Atlas*.

The APR includes the full range of providers that participate in Medicaid and CHIP, including individual providers, groups, and facilities. In the TAF APR, a facility is defined as an organization, institution, place, building, or agency that furnishes, conducts, and operates health care services for the prevention, diagnosis, or treatment of human disease, pain, or injury. Examples include hospitals, nursing facilities, home health agencies, schools, or transportation organizations. A group is defined as two or more providers (for example, physicians, advanced practice nurses, physical therapists, or dentists) who work together and share facilities; the providers may have different specialties. An individual provider[5] is defined as a person who provides medical or nonmedical services (CMS n.d.[a]). Nonmedical services may include personal care services, respite services, and non-emergency medical transportation. Providers that do not provide medical services are often referred to as "atypical" providers.

B. Structure

The APR includes information about the characteristics, locations, taxonomies and classifications, affiliated groups and programs, licensing and accreditations, and (for facilities) bed types for Medicaid- or CHIP-eligible providers, along with other provider-related identifiers. The TAF APR consists of the base file and eight supplemental files that can be linked to the base file. Table 1 provides an overview of the types of information in each file.

The base file serves as the primary ("parent") file, and the affiliated groups, affiliated programs, taxonomy, enrollment, and location supplemental files are subordinate to it ("child" files) (Table 1). Likewise, the licensing, identifiers, and bed type files ("grandchild" files) are subordinate to the location file. Each base file record may link to one or more records in each "child" file, so long as the information is applicable to the base file record and reported by the state. Likewise, each location file record may link to one or more records in each "grandchild" file, so long as the information is applicable to the provider and location and is reported by the state. For example, location file records for certain types of facility providers would typically link to one or more corresponding bed type file records, but group or individual providers would not link to any records in the bed type file because this information does not apply to these providers. The base file includes data elements that indicate whether the provider has a corresponding record in each supplemental file.

Two data elements can be used to identify providers throughout the TAF. The first is the state-assigned unique identifier for the provider entity, which is typically the identifier used in the state's Medicaid Management Information System when processing claims (CMS n.d.[c]). The second is the National Provider Identifier (NPI), which is the unique, 10-digit identification number that the NPPES assigns to

[5] Individual providers include two distinct groups. The first, incorporated sole practitioners, are sole practitioners with legal separation between themselves and their practice. The second, nonincorporated sole practitioners (also known as sole proprietors), are sole practitioners who operate with no legal distinction between themselves and their practice. The latter have characteristics associated with both individual providers and organizations.

Table 1. Contents of each APR file

File	Content	Relationship to base (parent) file[a]
Base	• The "primary" APR file to which the APR supplemental files can be linked. • Contains one record for each provider in the state, as defined by the unique combination of state-assigned provider identifier (SUBMTG_STATE_PRVDR_ID) and submitting state code (SUBMTG_STATE_CD). • Each base file record may represent an individual provider, a group of providers, or a facility. • Includes variables that reflect the attributes of these providers (e.g., ownership information, profit status, provider name, whether the provider is accepting new patients, and whether the facility is a teaching facility). • Includes flags, constructed from the taxonomy file, indicating whether the provider falls under one of several classification categories (e.g., behavioral health and social service providers). • Includes flags indicting the provider's enrollment status (active, denied, pending, terminated) and the programs enrolled (Medicaid and CHIP) • Includes up to 2 NPIs per state-assigned provider ID	n.a.
Affiliated groups	• The affiliated groups file captures the group identifier for providers affiliated with a larger provider group or subpart. For these providers, this file contains records that list the state-assigned provider identifiers associated with the affiliated group; each of these group providers should also be present as its own record in the base file. • If a provider is affiliated with more than one group during the year, the provider's base file record might link to multiple records in this supplemental file. If a provider is not affiliated with a larger group (or the state does not report this affiliation), the provider's base file record will not match to any records in this file.	Child
Affiliated programs	• The affiliated programs file captures the health plan, Medicaid waiver, home health entity, and other programs with which a provider is affiliated. • When a provider is affiliated with a health plan, the affiliated program type code (AFLTD_PGM_TYPE_CD) in the record in this file will be "2," and the record will contain the state-assigned managed care plan ID in the affiliated program ID field (AFLTD_PGM_ID), which can be linked to the TAF Annual Plan file. • When a provider is affiliated with a Medicaid waiver, the affiliated program type code in the record will be "3," and the affiliated program ID will be the state-assigned waiver number. • When the provider is affiliated with a health home entity, the affiliated program type code in the record will be "4," and the affiliated program ID will be the health home entity's name. • Other (unspecified) affiliations are represented by records with an affiliated program type code of "5." • Many providers will not have a corresponding record in this supplemental file. Conversely, if a provider is affiliated with more than one program during the year, their base file record could match to multiple records in this supplemental file.	Child

File	Content	Relationship to base (parent) file[a]
Taxonomy	• The taxonomy file captures classification information for Medicaid and CHIP providers. • An indicator in each record (PRVDR_CLSFCTN_TYPE_CD) designates the classification scheme that the code (PRVDR_CLSFCTN_CD) in the record represents: a value of "1" indicates a taxonomy code, "2" indicates a provider specialty code, "3" indicates a provider type code, and "4" indicates a provider authorized category of service code. • This supplemental file is expected to contain at least one record for every provider in the base file. It could contain up to four records if a state reported information for all four classification schemes. Furthermore, if a state reported multiple classification values for a given provider (e.g., multiple taxonomy codes) or different classification values for a provider during different months of the year, all valid values will be retained as separate records in this file. In these cases, a record in the base file could link to more than four records in this supplemental file. • At a minimum, states should report the taxonomy code (which uses the National Plan and Provider Enumeration System (NPPES) code set) for providers with a National Provider Identifier (NPI), and they should report the authorized category of service for providers that do not qualify for an NPI (those rendering nonmedical services) (CMS n.d.[b]).	Child
Enrollment	• The enrollment file can be used to identify providers' effective dates of enrollment with a state's Medicaid program, CHIP, or both. • As with other supplemental files, if a state submits different information about the provider's enrollment in different months of the calendar year, all information is retained as separate records. This could result in an individual base file record that links to more than one record in the enrollment file.	Child
Location	• The location file captures information about the service, practice, or billing address associated with each of a provider's locations. Each location has a unique location ID (record) in this file. • Each location in this file has corresponding information for one (and only one) address, if reported by the state, with the exception of records with a Location ID of '000' which have all address variables set to missing.[c] When a location ID is submitted in T-MSIS with multiple addresses, address-related variables are populated in the TAF in a hierarchal manner, as follows: – Address-related variables are populated based on the most recent **service location** (if reported), which is the location where a provider delivers services to patients. States are required to submit this information for all providers. – If service location was not reported, address-related variables are populated based on the most recent **practice location**, which represents the physical address of a provider's practice. In general, TAF users may use records representing service and practice locations interchangeably. – If service and practice location were not reported, address-related variables are populated based on the most recent **billing location**, which is where a provider receives payment. In many cases, a provider's billing address differs from the location where services are delivered. • TAF users can determine which location type a location record represents by using the three provider location type indicators in the file: (1) billing location, (2) practice location, or (3) service location. • Every base file record is expected to link to one or more records in this supplemental file, but some states do not report location information for every provider.	Child

File	Content	Relationship to base (parent) file[a]
Licensing	• The licensing file captures licensing or accreditation information for providers. • Each license could be specific to a particular location (for example, a hospital with multiple service locations); in these cases, each record in the location supplemental file will link to a different record in the licensing supplemental file. Licenses that are not location-specific will have “000” in the location ID field. • This file can specify up to five license or accreditation types. TAF users can determine which scheme a licensing record represents by using the license type code (LCNS_TYPE_CD) variable. A license type code value of “1” indicates a state, county, or municipality professional or business license; “2” indicates a U.S. Drug Enforcement Administration license; “3” indicates a professional society accreditation; “4” indicates a Clinical Laboratory Improvement Amendments accreditation; and “5” indicates another type of license or accreditation. • Most base file records are expected to have at least one corresponding record in this supplemental file. However, certain types of atypical providers who do not have a formal license are not expected to appear in this file.	Grandchild
Identifiers	• The identifiers file is designed to capture other identifiers associated with a provider (for example, NPIs and Medicare IDs), in addition to the state-assigned provider identifier that serves as the basis for organizing the APR. • Each identifier may be specific to a particular location. Identifiers that are not location-specific will have “000” in the location ID field. • There are eight identifier types that states may elect to report for providers. TAF users can determine which identifier a record represents by using the identifier type code (PRVDR_ID_TYPE_CD) variable. An identifier type of “1” indicates a state-assigned Medicaid provider ID; “2 indicates an NPI; “3” indicates a Medicare ID; “4” indicates a National Council for Prescription Drug Programs ID; “5” indicates a federal tax ID; “6” indicates a state tax ID; “7” indicates a Social Security number, and “8” indicates another type of identifier. [b] • All base file records are expected to have at least one corresponding record in this supplemental file.	Grandchild
Bed type	• The bed type file specifies the bed type associated with a facility provider, if applicable. Base file records representing individual and group providers will not have a corresponding record in this supplemental file. • Bed type information is almost always location-specific. Bed type information that is not location-specific will have “000” in the location ID field. • There are four possible bed types; TAF users can determine which bed type a record represents by using the bed type code (BED_TYPE_CD) variable. A Bed type of “1” indicates an intermediate care facility for individuals with intellectual disabilities; “2” indicates an inpatient facility; “3” indicates a nursing facility; and “4” indicates a Title 18 skilled nursing facility.	Grandchild

[a] This column indicates the relationship of each supplemental file to the APR base file. The affiliated groups, affiliated programs, taxonomy, enrollment, and location (“child”) files are subordinate to the base (“parent”) file. Likewise, the identifiers, licenses, and bed type (“grandchild” files) are subordinate to the location file because the variables in those APR files will often have information that differs by provider location ID.
[b] Social Security Numbers and “other” identifiers are available in the TAF but are excluded from the TAF RIF.
[c] A Location ID of ’000’ indicates that the provider has one or more corresponding licensing, identifier, or bed type records that apply to all locations. Records with a Location ID of ‘000’ have all address variables set to missing in the location file. If the state did not submit a record with Location ID '000' for the provider in the location file but did submit licensing, identifier, or bed type records with Location ID '000', the TAF production process creates a "dummy" record (with Location ID = '000') in the location file to allow a data user to link to the provider’s base file record to the location file and to the corresponding records in the licensing, identifier, or bed type files that apply to all of the provider’s locations. The TAF production process does not create dummy location records for licensing, identifier, or bed type records with a Location ID other than ‘000’, in the event the state reports grandchild records without a corresponding record in the location file.
n.a. = not applicable

each provider covered by the Health Insurance Portability and Accountability Act (HIPAA).[6] The APR is constructed at the level of the state-assigned provider identifier such that each record in the base file—the APR's primary file—contains one record for each state-assigned provider identifier (SUBMTG_STATE_PRVDR_ID). When linking to other files or counting providers, users of the APR file need to use both the submitting state code (SUBMTG_STATE_CD) and the state-assigned provider identifier, as two different states could possibly assign different providers the same ID number.[7]

Each record in the APR's base file can represent an individual provider, a group of providers, or a facility, and the base file contains variables that reflect the attributes of these providers.[8] Examples of attributes include ownership information, profit status, various provider names, whether the provider is accepting new patients, and whether the facility is a teaching facility. In addition, the base file has flags indicating whether the provider falls under one of several classification categories (for example, behavioral health and social service providers). For some TAF users, the information in the base file may be sufficient, depending on the analytic need. Other users might want to link to one or more of the APR's supplemental files, which support and enhance the information in the base file, to examine a particular topic area in greater depth (such as provider affiliations, locations, and classifications).

In a small number of states, the CHIP program or the third-party administrator (TPA) submitted T-MSIS records separately from the Medicaid program. In some cases, more than one agency in a state reports eligibility and claims data to T-MSIS, and the data from each reporting entity have separate submitting state codes. As of 2021, four states have separate reporting entities: Wyoming and Wyoming CHIP (SUBMTG_STATE_CD 56 and 93, respectively), Montana and Montana Third-Party Administrator (TPA) (SUBMTG_STATE_CD 30 and 94, respectively), Iowa and Iowa CHIP (SUBMTG_STATE_CD 19 and 96, respectively), and Pennsylvania and Pennsylvania CHIP (SUBMTG_STATE_CD 42 and 97, respectively).[9] As part of the production of the TAF RIF, records from different data submitters in the

6 Some providers do not have an NPI in the APR. Certain "atypical" providers—those that are authorized to deliver services to Medicaid and CHIP beneficiaries but do not provide "health care," as the term is defined in HIPAA 45 C.F.R. 160.103—are not eligible to receive an NPI (CMS 2006). Atypical providers reimbursed by Medicaid include those that offer taxi services, home and vehicle modifications, and respite services. TAF APR records for these providers are therefore not expected to contain NPIs. Additionally, some states do not report into T-MSIS the NPIs of all providers who are eligible for one.

7 Because providers can participate in more than one state's Medicaid program, it is also possible for more than one TAF APR record (each with different submitting state codes) to represent a single provider. For national-level analyses that are expected to include providers operating across states, TAF users may want to deduplicate these providers using NPIs.

8 TAF users can determine whether a provider is an individual, group, or facility by using the base file's facility/group/individual code data element (FAC_GRP_INDVDL_CD).

9 In TAF files produced after February 1, 2021 the submitting state code distinguishes between the Medicaid and non-Medicaid entities in Montana, Pennsylvania, and Wyoming; Iowa Medicaid began reporting separate CHIP records in its T-MSIS submissions and all TAF created after February 1, 2021, will exclude records with SUBMTG_STATE_CD = '96' (Iowa CHIP). In May 2020, Montana Medicaid began reporting TPA records in its T-MSIS submissions and all TAF created after July 1, 2021 exclude records from the Montana TPA. In October 2020, Wyoming Medicaid began reporting separate CHIP records in its T-MSIS submissions.

same state are assigned the same state code. However, users of non-RIF versions of the TAF should make sure to include records with both codes for analyses in those states.[10]

C. Linking the APR files

All nine APR files can be linked together using unique keys constructed based on various data elements. Although it is possible to use just the base file for some provider analyses, TAF users can link the information in the taxonomy, enrollment, affiliated groups, affiliated programs, and location supplemental files ("child" files) to the base records using PR_LINK_KEY.[11] Some records in the base file might have no matching records in one or more of these supplemental files. This could occur if the information is not applicable to the provider; for example, an independent practitioner who was not affiliated with any health plan or Medicaid waiver program might not have any records in the affiliated program file. It could also occur when information is applicable to the provider but the state did not report it to T-MSIS.

Furthermore, some records in the base file might link to more than one record in a supplemental file. For example, a hospital with multiple campuses might have one record in the base file but more than one record in the location supplemental file. TAF users should keep in mind the possibility of matching base records to no records or to multiple records in the supplemental files when designing an approach for appending information onto base records. Figure 1 shows the relationships between the APR files and the recommended methods of linking these files.

The APR location supplemental file is unique in that it is both a "child" and a "parent" file. The location file is subordinate to the base file, just like the other "child" files. But it also has three supplemental files ("grandchild" files) that are subordinate to it because the variables in those files often have values that differ by provider location ID. The grandchild files are the licensing, identifiers, and bed type supplemental files. Although it is possible to use just the location file for some provider analyses (such as for studying access to care), TAF users can link the licensing, identifiers, and bed type data to the location records using PR_LOC_LINK_KEY.[12] When the information in the grandchild files applies to all locations (that is, it is not location specific), the location ID in the record will be "000." All provider records in the base file with any corresponding record in at least one of the three grandchild files with location ID of "000" will have a record in the location file. This is because the TAF production process creates a dummy location

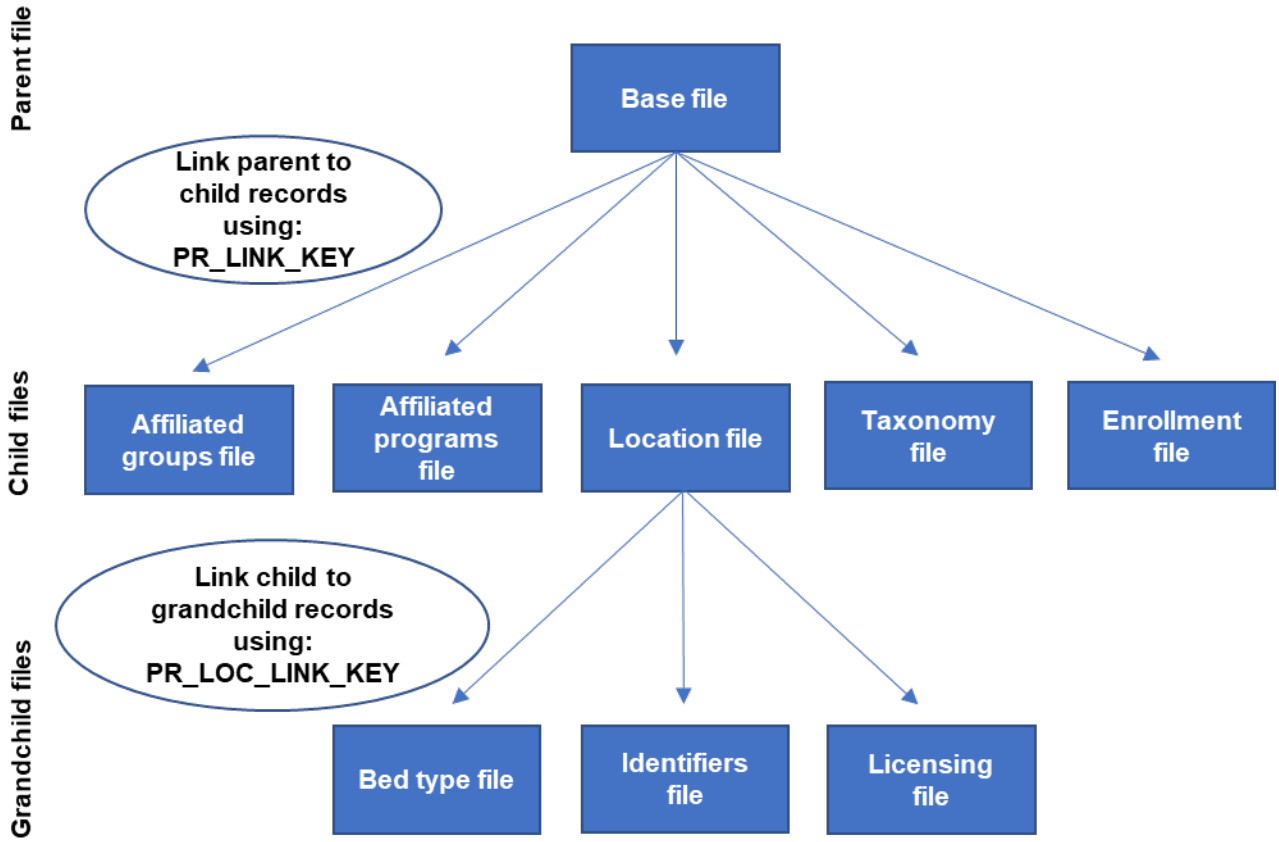
[10] Users of non-RIF versions of the TAF interested in analyzing information at a state level in these states should conduct all file linking and tabulations at the Medicaid or non-Medicaid entity level. However, for provider-level analyses, users will need to concatenate the provider records for the Medicaid entity and the non-Medicaid entity before summarizing the results by state. Because there is no T-MSIS guidance preventing one entity from using the same identification number as another entity for a different provider, TAF users should not attempt to link provider records across entities in the event one or more records with the same state-assigned provider ID are submitted by both entities. TAF users interested in de-duplicating provider records submitted by the Medicaid entity and the non-Medicaid entity may consider doing so using other data elements such as the NPI.

[11] PR_LINK_KEY is a constructed variable consisting of the following variables, concatenated and separated by hyphens, to create a unique link key for the five APR TAF that might be linked to the base file: DA_RUN_ID, PR_FIL_DT, PR_VRSN, SUBMTG_STATE_CD, and SUBMTG_STATE_PRVDR_ID.

[12] PR_LOC_LINK_KEY is a constructed variable consisting of the following variables, concatenated and separated by hyphens, to create a unique link key for the three APR supplemental grandchild files that might be linked to the APR supplemental location file: DA_RUN_ID, PR_FIL_DT, PR_VRSN SUBMTG_STATE_CD, SUBMTG_STATE_PRVDR_ID, PRVDR_LCTN_ID.

record (with a location ID of "000") for any provider with a record in the base file and at least one record in a grandchild file with a location ID of "000," but without a corresponding record reported by the state in the location file. These dummy records allow TAF users to link records in the base file to the location file and to the subset of grandchild records that apply to all of a provider's locations.

Figure 1. APR base and supplemental files: Relationships and recommended linking methods



Because of the parent, child, and grandchild relationships, and the unique consideration that grandchild records with location ID "000" contain information applicable to all locations, the process for linking base and supplemental file records will differ depending on the information needed. For example:

- A TAF user might not be interested in location-specific information but wants to find *all* license records associated with provider ID 12345 in State A. In this scenario, the user would find the provider base record in State A for provider ID 12345 and use the PR_LINK_KEY to select *all* provider location records in State A for this provider ID. The user would then use the PRV_LOC_LINK_KEYs from the selected location records to link to records in the provider license file.
- Alternatively, a TAF user might be interested in the license records associated with only one of a provider's multiple locations (in this scenario, location "00001" for provider ID 12345 in State A). Because provider license records for location ID "000" apply to all locations (including location ID

"00001"), the user should select the corresponding license records for this provider ID for both location IDs "000" and "00001."

Note that TAF users cannot directly link APR base records to the "grandchild" records in the licensing, identifier, or bed type supplemental files using the available link keys. This is because the base file does not share a common link key variable with these supplemental files. Users can link these files by first linking the base file to the location file records and then linking the location file records to the licensing, identifier, or bed type records, following the process described above. Or, if users are not interested in location-specific information, they can link the base records directly to the licensing, identifier, or bed type records using the same method used to link APR files to non-APR TAF files (such as claims files) described in Section VI—that is, by linking the records from each respective file using the file date, submitting state code, and state-assigned provider ID. This method would only work for users who do not need location-specific information and for analyses where it is appropriate to examine all available licensing, identifier, or bed type information associated with a provider (for example, a study of individual providers for which a user wants to append to the base file all available identifiers associated with that provider).[13]

III. Sources of provider information in the TAF

The focus and scope of the information reported for providers in the APR file differ significantly from that reported in claims. The FFS and encounter records in the claims files (IP, LT, OT, and RX) include basic information about the providers associated with the claim (for instance, state-assigned provider ID, NPI, and for certain providers on the claim [for instance, the billing provider] , classification information), as well as information about the types of services billed or rendered by the provider and payments received. Unlike the claims file, the APR is not limited to providers that billed for or rendered services to Medicaid and CHIP beneficiaries. Rather, it also contains providers that are authorized but did not bill for or render services during the year, and includes those whose authorization is pending, denied, or terminated. In addition, the APR includes more detailed information about these providers than claims do. The APR is designed to capture all taxonomy codes applicable to a provider. Table 2 compares the provider information available in claims versus in the APR.

TAF users will need to determine whether the provider information in claims is sufficient for their analysis or whether linking to the APR is necessary. For example, for users intending to examine beneficiary receipt of services delivered by primary care providers, the provider information in the claims may be sufficient. On the other hand, users intending to examine beneficiary receipt of services delivered by primary care providers by the provider's geographic location (for instance, to examine the adequacy of provider networks) would need to link to the APR.

Table 2. Information in claims and in the APR

Type of information	Information in claims	Information in the APR
Services provided	FFS claims and encounter records describe the type of service billed or rendered by the provider in the following fields: procedure code, Healthcare Common Procedure Coding System rate, revenue center, and bill type code.	n.a.
Medicaid and CHIP payments received	FFS claims, supplemental payments,[a] and service tracking claims[b] include the amounts paid by the state Medicaid or CHIP agency to the provider. Many states report per member, per month payments issued directly to providers as capitated payments (such as payments for primary care case management or health homes).[c]	n.a.
Provider identifiers	The IP, LT, OT, and RX files include fields for the state-assigned provider identifier and NPI (unless otherwise noted) for the following providers associated with the claim: ● Provider identifier fields for the admitting, billing, referring, and servicing providers are available in the IP and LT files. ● An identifier field for the operating provider (NPI only) is also available in the IP file. ● Provider identifier fields for the billing, referring, and servicing providers are available in the OT file. A provider identifier field for health home providers, the provider under direction, and the provider under supervision are also available in this file (NPI only). ● Provider identifier fields for the billing, dispensing, prescribing, and servicing providers are available in the RX file.	The APR includes the state-assigned provider identifier in the base file and in each supplemental file. The APR also includes up to two NPIs associated with the provider in the base file; any additional NPIs are included in the identifiers supplemental file. The APR includes additional identifiers beyond those appearing in the claims files for a given year. These additional providers include those who are authorized to render services to Medicaid and CHIP beneficiaries but did not bill for or render services during the year, and providers whose authorization is pending, denied, or terminated.
Provider location	The FFS and encounter header records in the IP, LT, OT, and RX files include the location ID that corresponds to the servicing provider (on the claim line in the IP, LT, and OT files) or the dispensing provider number (on the claim header in the RX file), which may be linked to the APR to obtain full address information.	The APR's location supplemental file includes one address for each of a provider's locations, including street address, city, state, zip code, and county code. This information reflects the location ID's service location, if reported by the state. If the service location is not reported, this information reflects the practice location, and if the practice location is not reported, this information reflects the location ID's billing location. Location records with a location ID of "000" do not have any address information because they apply to all of a provider's locations. TAF users can identify which type of location is reflected in a particular location record using the following flags in the file: PRVDR_ADR_SRVC_IND (service location), PRVDR_ADR_PRCTC_IND (practice location), and PRVDR_ADR_BLG_IND (billing location).

Type of information	Information in claims	Information in the APR
Provider taxonomy[d]	The FFS and encounter records in the IP, LT, OT, and RX files include taxonomy for certain providers in a claim (most notably, for the billing provider in all claims and the servicing provider in OT claims). Most states require providers with an NPI to include the taxonomy code on the claim form. A provider can have more than one applicable taxonomy code, but for a given claim submission, the provider will submit the taxonomy code most closely associated with the services provided for that claim.	Taxonomy codes are available in the APR's taxonomy supplemental file. Taxonomy file records that contain a taxonomy code will have a provider classification type code (PRVDR_CLSFCTN_TYPE_CD) of "1"; the corresponding code in that record (PRVDR_CLSFCTN_CD) is the taxonomy code. All providers with an NPI should have at least one taxonomy code that appears in this file; however, some states report this information incompletely. The APR is designed to capture all taxonomy codes applicable to a provider.
Provider specialty[e]	The FFS and encounter records in the IP, LT, OT, and RX files include provider specialty for certain providers in a claim (most notably, for the billing provider in all claims and the servicing provider in OT claims). Unlike taxonomy codes, which providers are generally required to include on the claim form, provider specialty codes are provided by the state on its T-MSIS claims submissions.	Provider specialty codes are available in the APR's taxonomy supplemental file. Taxonomy file records that contain a provider specialty code will have a provider classification type code (PRVDR_CLSFCTN_TYPE_CD) of "2"; the corresponding code in that record (PRVDR_CLSFCTN_CD) is the specialty code.
Provider type[f]	The FFS and encounter records in the IP, LT, OT, and RX files include provider type for certain providers in a claim (most notably, for the billing provider in all claims and the servicing provider in OT claims). Unlike taxonomy codes, which providers are generally required to include on the claim form, provider type codes are provided by the state on its T-MSIS claims submissions.	Provider type codes are available in the APR's taxonomy supplemental file. Taxonomy file records that contain a provider type code will have a provider classification type code (PRVDR_CLSFCTN_TYPE_CD) of "3"; the corresponding code in that record (PRVDR_CLSFCTN_CD) is the provider type code.
Authorized category of service[g]	n.a.	This classification scheme is used in the APR only. TAF users can append this information to a provider's base file record by selecting taxonomy file records that contain authorized category-of-service codes (which will have PRVDR_CLSFCTN_TYPE_CD = 4); the corresponding code in that record (PRVDR_CLSFCTN_CD) is the authorized category-of-service code. States must report this classification type, at a minimum, for providers who do not qualify for an NPI. The authorized category-of-service code set is available in the TAF APR Codebook (Chronic Condition Warehouse 2021).
Provider name	n.a.	The APR's base file includes the first and last name of each provider. TAF users who want to link a provider's identifier with the name of the individual, group, or facility would need to obtain the provider's name from the APR file.
Facility/group/individual code	n.a.	The APR's base file includes the facility/group/individual code, which classifies each provider as a facility, a group of practitioners, or an individual practitioner.

Type of information	Information in claims	Information in the APR
Additional provider information	n.a.	The APR's base and supplemental files include additional information about the characteristics, affiliated groups, affiliated programs, licensing/accreditations, bed types (facility providers only), and other identifiers associated with the provider. See Table 1 for more information.

[a] Supplemental payments are payments to providers above the capitation fee or set payment rate for services delivered to a specific Medicaid or CHIP beneficiary.
[b] Service tracking claims are lump-sum payments to providers that cannot be attributed to a specific Medicaid or CHIP beneficiary.
[c] Capitated payments are fixed per beneficiary per month payments made by states on behalf of Medicaid and CHIP beneficiaries.
[d] Provider taxonomy is one of three classification schemes (along with provider specialty and provider type) used across the claims and APR TAF to describe a provider's classification or area of specialization. Taxonomy codes are unique, standardized, 10-character codes maintained by the National Uniform Claim Committee that designate a provider's classification or area of specialization. Providers select one or more taxonomy codes that most closely describe their classification or area of specialization and submit the selected code when applying for an NPI. The taxonomy code set is available in the claims and APR data dictionaries.
[e] Provider specialty describes a provider's classification or area of specialization, but with less granularity than the taxonomy code. The provider specialty code set is available in the claims and APR data dictionaries.
[f] Provider type describes a provider's classification or area of specialization, but with less granularity than the taxonomy code. The provider type code set is available in the claims and APR data dictionaries.
[g] Authorized category of service is classification scheme that describes the types of services (medical or nonmedical) a provider is authorized to deliver.
n.a. = not applicable

IV. Linking APR to claims files

The APR can be linked to other TAF (including the TAF annual Demographic and Eligibility file; the TAF Annual Plan file; and the IP, LT, OT, and RX files). The APR was designed to be linked to these other TAF using state-assigned provider identifiers, along with other fields. Users who need to link claim headers and/or lines to APR records should use a combination of three data elements, as described in Table 3: (1) the first four characters (representing the year) of the file date, (2) the submitting state code, and (3) the state-assigned provider ID.

Table 3. Recommended approach to link claim headers and/or lines to APR records

Data elements that should be used to link claim headers and/or lines to APR records	Data element names	Notes
First 4 characters of the file date, which represents the year of data (e.g. 2020)	• APR: **PR_FIL_DT** • IP: **IP_FIL_DT** • LT: **LT_FIL_DT** • RX: **RX_FIL_DT** • OT: **OT_FIL_DT**	The APR includes a full calendar year of data, so the file date will always be a four-character year. In contrast, the IP, LT, OT, and RX are monthly files, and the file dates are therefore a year and month combination. Linking on the first four characters of the file date is necessary to ensure a full year of APR and claims data are linked.
Submitting state code	**SUBMTG_STATE_CD**	For a mapping of submitting state codes to states, refer to the TAF APR Codebook (Chronic Condition Warehouse 2021).
State-assigned provider ID	• APR: **SUBMTG_STATE_PRVDR_ID** • IP, OT, LT, RX: ***_PRVDR_NUM** in the claim headers or claim lines, where * represents the provider data element to be linked (BLG = billing, RFRG = referring, ADMTG = admitting, SRVCNG = servicing, PRSCRBNG = prescribing, and DSPNSNG_PD = dispensing)	The state-assigned provider ID on the APR should be linked to the state-assigned provider ID for the relevant provider on the claim (for example, the provider ID representing the billing provider).

The NPI is present in most claims and is also available as a field in the APR base file. So when linking claims to APR records using the state-assigned provider ID yields an inadequate linkage rate, TAF users could use the NPI to link the remaining records. But they should be cautious in doing so, for several reasons:

- The same NPI could be present in multiple APR base file records (each with a different state-assigned provider identifier), which would result in a many-to-many merge when using this data element to link files. For instance, a state could create separate provider IDs (leading to multiple base file records) for each of a provider's multiple locations, all of which share a common NPI. Or for providers in a group practice, a state might incorrectly populate the group practice's NPI and the individual provider's NPI on base file records representing these individual providers.
- Conversely, some APR base records have more than one NPI associated with them. This can occur because of mergers between provider organizations or because subparts or different locations of an organization each have their own NPI.

- Many APR records have missing NPIs (although there is considerable variation across states), so linking claims to the APR using NPIs will usually still result in a linkage rate of less than 100 percent. Note that because certain provider fields in claims (operating providers in the IP file and health home provider, provider under direction, and provider under supervision in the OT file) have an associated NPI variable, but not a state-assigned provider ID variable, these providers can only be linked to the APR using the NPI.

TAF users who want to use the NPI to link claim headers and/or lines to APR records should use a combination of three data elements, as described in Table 4: (1) the first four characters (representing the year) of the file date, (2) the submitting state code, and (3) the NPI.

Table 4. Alternative approach to link claim headers and/or lines to APR records (NPI)

Data elements that should be used to link claim headers and/or lines to APR records	Data element names	Notes
First 4 characters of the file date, which represents the year of data (e.g. 2020)	• APR: **PR_FIL_DT** • IP: **IP_FIL_DT** • LT: **LT_FIL_DT** • RX: **RX_FIL_DT** • OT: **OT_FIL_DT**	The APR includes a full calendar year of data, so the file date will always be a four-character year. In contrast, the IP, LT, OT, and RX are monthly files, and the file dates are therefore a year and month combination. Linking on the first four characters of the file date is necessary to ensure a full year of APR and claims data are linked.
Submitting state code	**SUBMTG_STATE_CD**	For a mapping of submitting state codes to states, refer to the TAF APR Codebook (Chronic Condition Warehouse 2021).
NPI	• APR base file: PRVDR_NPI_01 or PRVDR_NPI_02 [a] • IP, OT, LT, RX: ***_PRVDR_NPI_NUM** in the claim headers or claim lines, where * represents the provider data element to be linked (BLG = billing, RFRG = referring, ADMTG = admitting, SRVCNG = servicing, PRSCRBNG = prescribing, and DSPNSNG_PD = dispensing)	One or both NPIs on the APR should be linked to the NPI for the relevant provider on the claim (for example, the NPI representing the billing provider).

[a] If a provider has more than two NPIs (which is uncommon), only the first two appear in the APR base file. All of a provider's NPIs appear in the identifiers supplemental file. TAF users intending to link the APR identifiers supplemental file to claim headers or lines must instead link these records using the identifier file's PRVDR_ID where PRVDR_ID_TYPE_CD = "2."

TAF users might want to link certain APR supplemental files to claims based on location ID, a variable included only in the APR location, identifiers, licenses, and bed type supplemental files and in the IP, LT, OT, and RX header records. The location ID in the IP, LT, and OT header records corresponds to the servicing provider (located on the claim line in each file), whereas the location ID in the RX header records corresponds to the dispensing provider (located on the claim header in this file). For example, users might want to link information between the APR and claims for only one of a provider's multiple locations. However, we recommend that users perform this type of file linking only when (1) the location ID variable is populated in both files to be linked and (2) users want to access the location, licenses, identifiers, or bed type data from the APR. Users can perform this type of linking using the following fields, as described in Table 5: (1) the first four characters (representing the year) of the file date, (2) the submitting state code, (3) the state-assigned provider ID, and (4) the provider's location ID. When linking APR supplemental files to claims based on location ID, users should always include location ID of "000" if present, which indicates that the information applies to all of the provider's locations.

Table 5. Recommended approach to link claim headers to APR records for a specific location ID

Data elements that should be used to link claim headers to APR records	Data element names	Notes
First 4 characters of the file date, which represents the year of data (e.g. 2020)	• APR: **PR_FIL_DT** • IP: **IP_FIL_DT** • LT: **LT_FIL_DT** • RX: **RX_FIL_DT** • OT: **OT_FIL_DT**	The APR includes a full calendar year of data, so the file date will always be a four-character year. In contrast, the IP, LT, OT, and RX are monthly files, and the file dates are therefore a year and month combination. Linking on the first four characters of the file date is necessary to ensure a full year of APR and claims data are linked.
Submitting state code	**SUBMTG_STATE_CD**	For a mapping of submitting state codes to states, refer to the TAF APR Codebook (Chronic Condition Warehouse 2021).
State-assigned provider ID	• APR: **SUBMTG_STATE_PRVDR_ID** • IP, OT, LT, RX: ***_PRVDR_NUM** in the claim headers or claim lines, where * represents the provider data element to be linked (BLG = billing, RFRG = referring, ADMTG = admitting, SRVCNG = servicing, PRSCRBNG = prescribing, and DSPNSNG_PD = dispensing)	The state-assigned provider ID on the APR should be linked to the state-assigned provider ID for the relevant provider on the claim (for example, the provider ID representing the billing provider).
Provider location ID	**PRVDR_LCTN_ID**	The provider location ID on the APR (available only on the location, bed type, identifiers, and licenses supplemental files) should be linked to the provider location ID on the claim headers. The location ID in the IP, LT, and OT header records corresponds to the servicing provider (located on the claim line in these files), whereas the location ID in the RX header records corresponds to the dispensing provider (located on the claim header in this file).

V. Linking TAF to external provider data

In most instances, TAF users who want to link the APR to external, non-TAF files will need to use the NPI. For example, users might want to link to the publicly available NPPES NPI registry to obtain more information about a provider's characteristics. Much of the provider information in the NPPES overlaps with that in the APR (for example, taxonomy code and the designation of the provider as an individual or an organization, the latter corresponding to facilities and groups in the APR). However, the NPPES has this information populated for every provider because it is required when applying for an NPI, whereas some states do not have all of this information in the APR.[14] The NPPES also includes additional details about a provider's characteristics, such as an individual provider's credentials (M.D., N.P., P.A., R.N., and so on).

But as noted previously, users must exercise caution when linking with NPIs because slightly less than half of the providers in the APR have an NPI, whereas other providers have more than one NPI. The rates of APR records with an NPI vary considerably by state. TAF users interested in analyzing data for a particular state (or group of states) will want to verify that the percentage of APR records with an NPI is sufficient for linking for the states of interest.[15] If linking the APR to other external data sets using NPIs, note that multiple APR base records could share the same NPI and that some base records could have more than one NPI. In certain cases, TAF users might also find other identifiers available in the identifiers supplemental file (such as the Medicare ID or the National Council for Prescription Drug Programs ID) useful for linking, depending on the application.

Some users might prefer to link TAF claim records directly to external data sets, such as NPPES (using the NPI), rather than linking claims to the APR. In this case, users should examine the proportion of records in the claims files with a nonmissing NPI for the specific provider fields of interest (such as billing or servicing).[16] For most states, provider fields in claims have more complete NPI information than the APR has.

VI. Identifying and classifying providers

The APR captures information about the providers authorized by a state to render services to Medicaid and CHIP beneficiaries at any point in the calendar year, along with providers whose authorization is pending, denied, or terminated. Most TAF users will want to restrict their analysis to providers that were ever active in the year or active during the months of the analysis. The APR's active enrollment indicator (PRVDR_ENRLMT_STUS_ACTV_IND) shows the status of each provider's enrollment in Medicaid and

[14] For more information about the percentage of APR records with a facility/group/individual code or a taxonomy code reported, see the "Facility/Group/Individual Code," "Group and Individual Providers—Classification Types," and "Facility Providers—Classification Types" topics in *DQ Atlas* at http://www.medicaid.gov/dq-atlas.

[15] For more information about the percentage of APR records with an NPI, see the "National Provider Identifier" topic in *DQ Atlas* at http://www.medicaid.gov/dq-atlas.

[16] For more information about the percentage of claims with missing NPI for select provider fields, see the "Billing Provider NPI—IP," "Billing Provider NPI—LT," "Billing Provider NPI—OT," "Servicing Provider NPI—OT," "Billing Provider NPI—RX," "Prescribing Provider NPI—RX," and "Dispensing Provider NPI—RX" topics in *DQ Atlas* at http://www.medicaid.gov/dq-atlas.

CHIP—"active," "pending," "denied," or "terminated"—on a monthly basis.[17] Note that many providers with an "active" status did not render or bill for Medicaid or CHIP services during the year, and states sometimes report a nonactive status for providers with paid claims during the year.[18] TAF users who want to identify providers that actively rendered or billed for services during the year will need to link to the claims files (as described in Section VI) to determine whether the provider appeared as a billing, servicing, etc. provider on a claim (depending on the analytic question to be answered) rather than rely on the active enrollment indicator.

The APR's taxonomy supplemental file can be used to identify a provider's area of specialization using up to four possible classification schemes: three that are also used across the TAF claims files (taxonomy, specialty, and provider type) and a fourth (authorized category of service) that is included in the APR only. Maintained by the National Uniform Claim Committee, taxonomy codes are standardized, unique 10-character codes that designate a provider's classification or area of specialization. Providers select one or more taxonomy codes that most closely describe their classification or area of specialization and must submit the selected code when applying for a NPI. Provider specialty, provider type, and authorized category of service also designate a provider's classification or area of specialization, but with less granularity than taxonomy code does. Analysis of various topics, such as the geographic distribution of specialists or primary care physicians, requires information about provider specialization. The APR base file also includes flags, constructed from the same T-MSIS source data underlying the taxonomy file, indicating whether the provider's area of specialization falls within a certain category (for example, behavioral health and social service providers).

States are encouraged to report all classification types for which information is available. But according to CMS guidance, states should report, at a minimum, taxonomy codes for providers with an NPI and authorized categories of service for providers without an NPI (CMS n.d.[b]). Certain "atypical" providers—those that are authorized to deliver services to Medicaid and CHIP beneficiaries but do not provide "health care," as the term is defined in HIPAA 45 C.F.R. 160.103—are not eligible to receive an NPI (CMS 2006). Atypical providers reimbursed by Medicaid include those that offer taxi services, home and vehicle modifications, and respite services. These providers' TAF APR records are therefore not expected to contain NPIs.

[17] A provider's Medicaid or CHIP enrollment can be denied because of multiple provider numbers, an invalid license, failure to meet eligibility requirements, or other reasons. A provider's enrollment status can be pending license verification, rate determination, a valid NPI, status approval, documentation, or fulfillment of other requirements. A provider's Medicaid or CHIP enrollment status can also be terminated for various reasons, including noncompliance, change of ownership, an expired or revoked license, voluntary termination, and death of the provider. If the provider is eligible and approved for enrollment in Medicaid or CHIP (not pending, denied, or terminated), the state considers the provider as actively enrolled in the state's Medicaid program or CHIP.

[18] For more information on the accuracy of the active provider indicator for identifying providers with FFS claims or managed care encounters in TAF, see the topics "Linkage of Claims to APR" and "Active Enrollment Status Indicator" in *DQ Atlas* at http://www.medicaid.gov/dq-atlas.

References

Centers for Medicare & Medicaid Services. “November 2020 Medicaid and CHIP Enrollment Data Highlights.” Available at https://www.medicaid.gov/medicaid/program-information/medicaid-and-chip-enrollment-data/report-highlights/index.html. Accessed June 25, 2021.

CMS. “CMS Guidance: Reporting Provider Facility-Group-Individual-Code in T-MSIS.” Baltimore, MD: CMS, n.d.(a). Available at https://www.medicaid.gov/medicaid/data-and-systems/macbis/tmsis/tmsis-blog/entry/51240. Accessed May 27, 2021.

CMS. “CMS Guidance: Best Practice for Reporting PROV-CLASSIFICATION-TYPE and PROV-CLASSIFICATION-CODE in the T-MSIS Provider File.” Baltimore, MD: CMS, n.d.(b). Available at https://www.medicaid.gov/medicaid/data-and-systems/macbis/tmsis/tmsis-blog/entry/47562. Accessed May 27, 2021.

CMS. “CMS Guidance: Reporting Provider Identifiers in T-MSIS.” Baltimore, MD: CMS, n.d.(c). Available at https://www.medicaid.gov/medicaid/data-and-systems/macbis/tmsis/tmsis-blog/entry/50507. Accessed May 27, 2021.

CMS. *DQ Atlas*. n.d. Available at https://www.medicaid.gov/dq-atlas/. Accessed May 27, 2021.

CMS. Memo to state Medicaid directors. SMDL #06-020. September 19, 2006. Available at https://www.medicaid.gov/Federal-Policy-Guidance/downloads/SMD091906b.pdf. Accessed May 27, 2021.

Chronic Condition Warehouse. CODEBOOK: T-MSIS Analytic Files (TAF) Annual Provider (APR) File. Version 1.0. June 2021. Available at https://www2.ccwdata.org/documents/10280/19022436/codebook-taf-annual-provider.pdf. Accessed June 28, 2021.

Kristin Andrews Lemos[1], Allison Barrett[1], Kimberly Proctor[2], and Jessie Parker[2]. “TAF Technical Documentation: Annual Provider (APR) File.” Baltimore, MD: CMS, 2021.

Reviewers: Tawney Simon[2], Ali Fokar[2], Laura Nolan[1], Brian Johnston[1], Julie Sykes[1]

[1]Mathematica, [2]Centers for Medicare & Medicaid Services, Center for Medicaid and CHIP Services, Data and Systems Group, Division of Business and Data Analytics

Correspondence should be addressed to MACBISData@cms.hhs.gov