

Technical Specifications

September 2021

Identifying Beneficiaries with a Treated Substance Use Disorder (SUD)



**This page has been left blank for double-sided copying.**

TABLE OF CONTENTS

I Description 1

II Logic overview 2

III Detailed technical specifications 7

IV Limitations 15

**This page has been left blank for double-sided copying.**

TABLES

1 Logic overview for identifying beneficiaries with a treated SUD 3

2 DE variables to retain 7

3 IP variables to retain 8

4 LT variables to retain 9

5 OT variables to retain 10

6 RX variables to retain 11

7 TAF variables name corresponding to each code type 11

8 Values of claims qualification rule flag 12

9 Values of SUD claims-level condition flags 12

10 Code type to TAF variable crosswalk for professional claims 13

11 SUD claims-level condition flags for prescription drugs 13

12 Service date variables for each TAF file 14

**This page has been left blank for double-sided copying.**

I. Description

These technical specifications provide a general logic for using Medicaid and CHIP medical and pharmacy claims data to identify beneficiaries with a treated substance use disorder (SUD). More specifically, they describe how to identify beneficiaries who received Medicaid- or Children’s Health Insurance Program (CHIP)-funded services for alcohol, tobacco, opioids, and illicit drug use, including cannabis, hallucinogens, stimulants, inhalants, and sedatives. They also identify beneficiaries treated for more than one SUD and are designed to be applied to the following population:

* Beneficiaries eligible for comprehensive benefits through Medicaid or CHIP: beneficiaries who qualify for only limited benefits (such as emergency Medicaid or family planning only) are unlikely to receive Medicaid-funded SUD treatment regardless of need.
* Beneficiaries enrolled for the full calendar year (12 months of enrollment with no more than one gap that does not exceed 45 days): we may not be capturing all service use for beneficiaries with gaps in enrollment.
* Beneficiaries up to age 64 not dually enrolled in Medicare: we suggest excluding beneficiaries dually enrolled in Medicaid and Medicare because Medicaid is the secondary payer for this population. When Medicaid is a secondary payer, the primary payer may pay some SUD treatment services received and there may be no record in the Medicaid claims data of the beneficiary receiving these services. In addition, we assume most users may not have access to Medicare administrative data necessary for capturing SUD services delivered to the beneficiary.

II. Logic overview

To identify beneficiaries with a treated SUD using the Transformed Medicaid Statistical Information System (T-MSIS) Analytic File (TAF) Research Identifiable File (RIF) data, the technical specifications lay out a nine-step algorithm described in Table 1 below. For each step, we provide the relevant data files needed, the logic and purpose for the step, and information about how to implement the step in the TAF RIF. The algorithm relies on the TAF Demographic and Eligibility (DE), inpatient (IP), long-term care (LT), other services (OT) and prescription drug (RX) files.

To identify the target population, users will identify the analytic sample, identify SUD treatment claims for that sample, and determine whether a beneficiary had enough claims to be classified as having a SUD. To identify claims for SUD treatment, the algorithm requires (1) a diagnosis code indicating a SUD and (2) a type of bill, revenue, and procedure code or a national drug code that indicates that the service involved the provision of SUD treatment (see the accompanying documentation: “Identifying Beneficiaries with a Treated SUD: Reference Codes” for a full list of relevant codes).[[1]](#footnote-2) The type of bill, revenue, procedure, and national drug codes are used to record the types of service rendered, which include inpatient care, residential treatment, outpatient visits, and pharmaceutical and medication-assisted treatment.After identifying claims for SUD treatment, the algorithm specifies the number of claims on different dates of service that are required to identify a beneficiary with a treated SUD. The algorithm identifies a beneficiary as having a treated SUD if the individual has one inpatient or prescription drug claim for SUD treatment or two outpatient or residential claims for SUD treatment on different dates of service. When multiple claims are used to determine whether a beneficiary should be included in the population, the algorithm requires the claims to be in the same SUD diagnostic category. For example, if a beneficiary has one outpatient claim that has the diagnosis code for a tobacco use disorder and another that has the diagnosis code for an alcohol use disorder but no other claims for SUD services, then the beneficiary would not be classified as having a SUD.[[2]](#footnote-3)

Table 1. Logic overview for identifying beneficiaries with a treated SUD

|  |  |  |  |
| --- | --- | --- | --- |
| Step | Relevant TAF RIF | Logic and purpose | Implementation using TAF RIF |
| 1. Identify beneficiaries in enrollment file who qualify for inclusion | DE file | We recommend including only those beneficiaries continuously enrolled for all or nearly all of the calendar year with full scope Medicaid or CHIP benefits, and for whom Medicaid was the primary payer. Most Medicaid and CHIP beneficiaries are enrolled for the full year, defined as 12 months of enrollment with no more than one gap that does not exceed 45 days. Since we do not have a full record of all service use for partial-year beneficiaries, we recommend excluding this relatively small group.  Beneficiaries who qualify for only limited benefits (such as emergency Medicaid or family planning only) are unlikely to receive Medicaid-funded SUD treatment regardless of need, and their inclusion has the potential to bias estimates of SUD prevalence downward. When Medicaid is a secondary payer—most commonly, for individuals dually eligible for Medicare, but also for those with private insurance coverage—the primary payer is likely to pay for any SUD treatment services received and there may be no record in the Medicaid claims data of the beneficiary receiving these services. This would similarly bias estimates of SUD prevalence downward. | To identify continuously enrolled beneficiaries, include only those with one or two distinct stints of enrollment. Use ENRLMT\_START\_DT and ENRLMT\_END\_CY\_DT in the DE file to identify beneficiaries who had no more than one gap in enrollment during the calendar year that did not exceed 45 days.  To identify beneficiaries with full scope benefits in TAF, retain only DE enrollment records in which RSTRCTED\_BNFTS\_CD is equal to 1 (eligible for full-scope Medicaid or CHIP benefits), 4 (eligible for restricted benefits for pregnancy-related services), 5 (eligible for other restricted benefits) or 7 (eligible for an alternative package of benchmark-equivalent Medicaid coverage).  To exclude dually eligible beneficiaries, drop DE records in which DUAL\_ELGBL\_CD is equal to 1 (full dual), 2 (partial dual), or 3 (other dual).  To exclude beneficiaries over the age of 64, drop beneficiaries with AGE > 64. |
| 2. Merge claims to enrollment data to create an annual file | DE, IP, LT, OT, and RX files | Only those claims matching to a beneficiary who qualifies for inclusion in the analytic population should be retained. Claims matching to beneficiaries with partial-year enrollment, partial benefits, who were dually eligible, or who were over the age of 64 should not be examined for SUD service use. | Merge the DE file after implementing step 1 with the relevant monthly claim file for each of the IP, LT, and OT files. This creates an annual file with all medical claims incurred by included beneficiaries during the calendar year.  All TAF claims are monthly files. Select the most recent file version date for each month using the variables IP\_VRSN, LT\_VRSN, OT\_VRSN, and RX\_VRSN. Use the encrypted MSIS\_ID to link the DE to each claim file. Any claim records that did not match to an MSIS\_ID in the DE file will be excluded. |
| 3. Link line-level records to header records in the claims files | IP, LT, OT, and RX files | Each monthly IP, LT, and OT TAF consists of two files: the header-level file and the line-level file. Certain line-level data elements are required to identify claims or encounters for SUD services, including revenue code (on facility claims only), CPT or HCPCS procedure code, National Drug Code (NDC), service begin date, and service end date. Additional header-level data elements are also required to identify SUD services, including diagnosis code, ICD procedure code (on inpatient facility claims only), bill type (on facility claims only), and place of service (on professional claims only). As a result, each header-level record must be linked with its associated line-level records to evaluate whether the claim qualifies as the delivery of a SUD service. | Link the header and line-level files on the basis of (1) the maximum value of production data run ID (DA\_RUN\_ID) available in both files and (2) the claim ID (CLM\_ID). |
| 4. Restrict claims files to FFS claims and encounter records | IP, LT, and OT files | After linking the header-level and line-level claims, exclude certain types of claim records based on the type of claim code:  The TAF header-level file contains all payments qualifying for federal matching funds that were made by the Medicaid agency on behalf of an enrolled beneficiary, including monthly capitation payments, service tracking claims, and supplemental payments. Since we were interested only in records representing a specific service delivered to a single identified beneficiary, restrict the TAF claim records to those that represent FFS claims or managed care encounters. | Retain header-level claims where CLM\_TYPE\_CD is equal to one of the following values:  1: A FFS Medicaid or Medicaid-expansion claim  3: Medicaid or Medicaid-expanding managed care encounter record  A: Separate CHIP (Title XXI) FFS claim  C: Separate CHIP (Title XXI) encounter record |
| 5. Identify facility claims for SUD services and classify by SUD condition | IP, LT, and OT files | Claims submitted on a facility (UB-04) form are populated with a different set of fields compared with claims submitted on a professional (CMS-1500) form. As a result, the criteria for identifying SUD services on facility claims differ from the criteria used on professional claims. For example, the type of bill and revenue code fields are only coded on facility claims.  Facility claims for medical services must include a relevant SUD diagnosis from the reference codes to be used to identify beneficiaries with a SUD. A single inpatient facility claim or two non-inpatient facility claims (including outpatient and subacute care claims) are required to flag a beneficiary as having a SUD. | Evaluate overnight facility claims, which in the IP and LT TAFs, using the SUD service identification rules contained on the “SUD Facility claims" sheet (purple tab) in the reference codes workbook. These rules use a combination of diagnosis code and type of bill, revenue code, ICD-9 and ICD-10 procedure codes, or CPT codes to identify SUD services.  Evaluate outpatient facility claims, which in the OT TAF, using the SUD service identification rules contained in the “SUD Facility claims” sheet (purple tab) in the reference codes workbook. These rules use a combination of diagnosis code and revenue codes, HCPCS codes, or CPT codes to identify SUD services.  All facility claims require one of the condition-specific SUD diagnoses (green tabs) in any position to qualify as a SUD service. |
| 6. Identify professional claims for SUD services and classify by SUD condition | OT file | Certain fields, including place of service, are available only on professional claims. As with facility claims, the algorithm generally requires a professional claim to include a SUD diagnosis from the reference codes to be used to identify beneficiaries with a SUD. A single professional claim with a place of service indicating inpatient or two professional claims with any other place of service are required to flag a beneficiary as having a SUD. | Evaluate professional claims, which in the OT TAF, using the SUD service identification rules contained on the “SUD-professional” sheet (purple tab) in the reference codes workbook. These rules use a combination of diagnosis code (green tabs) and CPT or HCPCS codes to identify SUD services. The place of service code on the claim is used to differentiate between services provided to inpatients (which require only 1 qualifying claim) versus all other service locations (which require 2 or more qualifying claims for a beneficiary to be identified as belonging in the SUD population). |
| 7. Identify prescription drug claims for SUD and classify by SUD condition | IP, LT, OT and RX files | Prescription drug claims have a different set of fields available compared with medical claims. Detailed information on the specific drug filled is identified through the NDC on the claim. Unlike with medical claims, diagnosis codes are not available on drug claims and, as a result, are not required to identify SUD pharmacy claims.  The list of NDCs used to identify and flag SUD pharmacy claims parallels the set used by the Chronic Conditions Warehouse (CCW) to flag beneficiaries with a SUD.  Naltrexone is used to treat both AUD (alcohol use disorder) and OUD (opioid use disorder). As a result, this set of codes is included for both AUD and OUD. Following the logic used in the CCW, beneficiaries with a naltrexone fill are classified as having an OUD only if they have no other medical or pharmacy claims indicating an AUD. | The majority of prescription drug claims are located in the RX TAF, although some physician-administered drug claims may be found in the IP, LT, and OT TAFs. Use all four claims files to identify claims that have an NDC code matching one of the codes contained in the “Alcohol-Rx,” “Tobacco-Rx,” or “Opioids-Rx” sheets of the reference codes (blue tabs). |
| 8. De-duplicate medical claims to count no more than one qualifying claim per day | Constructed SUD claim file | Using only the claims identified in steps 4, 5, and 6 to construct a SUD claim file, de-duplicate to count no more than one claim per beneficiary per day within the categories of inpatient facility, long-term care facility, or outpatient and professional claims. This is required because many outpatient or office-based visits can result in multiple separate claims for the same service or set of services. For example, an office-based visit with a primary care physician may generate a distinct claim from the physician for the professional component of the visit and from the outpatient hospital that owns the physician practice for the facility component of the visit. Similarly, a follow-on laboratory test ordered by the physician may generate yet more professional and facility claims on the same day, although many researchers would consider these services to be part of the same visit.  Prescription drug claims should not be de-duplicated, as each claim corresponds to a single drug fill and split claims for the same service do not occur in the prescription drug data. | Collapse claims for the same beneficiary (identified by SUBMTG\_STATE\_CD and MSIS\_ID) on the same date of service into a single record that retains all flags from each claim used to identify a SUD claim and specific categories of SUD conditions. The date of service is defined using DSCHRG\_DT on the IP file, SRVC\_ENDNG\_DT on the LT and OT files, and RX\_FIL\_DT on the RX file. |
| 9. Flag beneficiaries receiving SUD services | DE file and constructed SUD claim file | Roll up claims to the beneficiary level to identify and flag Medicaid beneficiaries receiving services for a SUD. To be identified in the SUD population, a beneficiary must have one qualifying medical claim from an inpatient setting, one qualifying prescription drug claim, or two qualifying medical claims on different dates of service that occurred in a non-inpatient setting. Certain screening codes identified in the reference codes can count as one qualifying non-inpatient SUD service but must be paired with another non-screening service to include the beneficiary in the SUD population.  Since this algorithm is intended to identify beneficiaries with specific SUD conditions, beneficiaries identified through non-inpatient claims must have the same condition coded on both claims to qualify for inclusion. That means that a beneficiary with two non-inpatient claims for different conditions (for example, one AUD screening claim and one claim indicating tobacco use disorder) would not qualify for inclusion in the SUD population. | Loop through the set of prescription drug and medical claims that are flagged as being SUD services (where SUD\_CLM\_FLAG is equal to 1, 2, or 3). Flag beneficiaries as being part of the SUD population (POP\_SUD = 1) if they have   1. one claim with SUD\_CLM\_FLG = 1, or 2. two claims with SUD\_CLM\_FLG = 2 and the same condition-specific flag on both claims, or 3. one claim with SUD\_CLM\_FLG = 2 and one with SUD\_CLM\_FLG = 3, and the same condition-specific flag on both claims. |

DE = demographics and eligibility; FFS = fee-for-service; CPT = Current Procedural Terminology; HCPCS = Healthcare Common Procedure Coding System; IP = inpatient; LT = long term care; OT = other services; RX = prescription drug; TAF = T-MSIS Analytic File

III. Detailed technical specifications

The detailed technical specifications below describe how the logic is applied to the TAF RIF.

Identify beneficiaries in DE who qualify for inclusion

1. Keep only the variables from the annual DE TAF that are needed for the algorithm (Table 2).

Table 2. DE variables to retain

|  |  |
| --- | --- |
| TAF variable name(s) | Description |
| DA\_RUN\_ID | A unique identifier that identifies the TAF production run that produced the TAF file. |
| RFRNC\_YR | Year of the reporting period. |
| DE\_VRSN | Indicator representing the iteration of the file. |
| MSIS\_ID | The encrypted state-assigned unique identification number used to identify and Medicaid/CHIP enrolled beneficiary and any claims submitted to the system. |
| SUBMTG\_STATE\_CD | The ANSI numeric state code for the U.S. state, territory, or the District of Columbia that has submitted the data. |
| DUAL\_ELGBL\_CD\_*mm* | A flag to indicate whether and how a beneficiary was dually-eligible. |
| RSTRCTD\_BNFTS\_CD\_*mm* | A flag that indicates the scope of Medicaid or CHIP benefits to which an individual is entitled. |
| AGE | Indicates beneficiary age in years during the last month of enrollment. |
| ENRLMT\_START\_DT | Indicates the date at which a beneficiary’s enrollment in Medicaid or CHIP became effective. |
| ENRLMT\_END\_\_DT | Indicates the date at which a beneficiary’s enrollment in Medicaid or CHIP ended. |

1. Using ENRLMT\_START\_DT and ENRLMT\_END\_DT, restrict to beneficiaries (identified by MSIS\_ID) with 12 months of continuous enrollment with no more than one gap in enrollment during the calendar year. Exclude beneficiaries whose gap in enrollment exceeds 45 days.
2. Next, restrict to beneficiaries who are not dually eligible for Medicaid and Medicare by dropping beneficiaries (identified by MSIS\_ID) for whom DUAL\_ELGBL\_CD\_1-12 = (1, 2, or 3) for any month in the year.
3. Restrict to beneficiaires who are 64 or younger by dropping beneficiaries for whom AGE > 64.
4. Lastly, restrict to beneficiaries who are eligible for the full scope of Medicaid benefits by keeping only those beneficiaries for whom RSTRCTD\_BNFTS\_CD\_1-12 = (1, 4, 5, or 7) for all 12 months in the year.

Merge claims to enrollment data to create an annual file

In this step we will conduct initial processing of the four types of monthly claims files to limit their size and to facilitate linking. We will then merge the claims files with the DE annual file created in step 1 to create a single large analytic file.

1. Read in all monthly medical and pharmacy claims records for the year from the IP, OT, LT, and RX claims header and line files, restricting the set of variables retained to those identified in Tables 3-6 below. Note that each monthly TAF consists of two files: the header-level file and the line-level file. The tables below note whether the variables that should be retained are located on the header file, the line file, or both files.
2. Drop from the claims files any record that does not match to an MSIS\_ID included in the base annual enrollment file created in step 1.

Table 3. IP variables to retain

|  |  |
| --- | --- |
| Variable name | Variable description |
| Header and line | |
| DA\_RUN\_ID | A unique identifier that identifies the TAF production run that produced the TAF file. |
| CLM\_ID | CCW claim identifier used to link headers and lines |
| IP\_FIL\_DT | Year and month of the reporting period. |
| IP\_VRSN | Indicator representing the iteration of the file. |
| MSIS\_ID | The encrypted state-assigned unique identification number used to identify and Medicaid/CHIP enrolled beneficiary and any claims submitted to the system. |
| SUBMTG\_STATE\_CD | The ANSI numeric state code for the U.S. state, territory, or the District of Columbia that has submitted the data. |
| Header only | |
| ADMSN\_DT | The date on which the recipient was admitted to the hospital. |
| DSCHRG\_DT | The date on which the recipient was discharged from the hospital. |
| BILL\_TYPE\_CD | A data element corresponding with UB-04 form locator FL4 that classifies the claim as to the type of facility (2nd digit), type of care (3rd digit) and the billing record's sequence in the episode of care (4th digit). |
| DGNS\_CD\_\* (1-12) | ICD-9/10-CM code found on the claim. |
| DGNS\_VRSN\_CD\_\* (1-12) | A flag that identified the coding system (ICD-9 or ICD-10) used for the diagnosis code. |
| CLM\_TYPE\_CD | A data element identifying what kind of payment is covered and distinguishes between claims that are for Medicaid or Medicaid-expansion, S-CHIP, and other types of claims. |
| PRCDR\_CD\_\* | A procedure code (ICD-9, ICD-10, CPT, HCPCS or other) used by the state to identify the procedures performed during the hospital stay referenced by this claim. |
| PRCDR\_CD\_DT\_\* (1-6) | The date upon which the procedure was performed. |
| SRVC\_BGN\_DT | For services received during a single encounter with a provider, the date the service covered by this claim was received. |
| SRVC\_ENDG\_DT | For services received during a single encounter with a provider, the date the service covered by this claim was received. |
| Line only | |
| LINE\_SRVC\_BGN\_DT | For services received during a single encounter with a provider, the date the service covered by this claim was received. |
| LINE\_SRVC\_ENDG\_DT | For services received during a single encounter with a provider, the date the service covered by this claim was received. |
| REV\_CNTR\_CD | A code which identifies a specific accommodation, ancillary service or billing calculation. |
| NDC | A code in National Drug Code (NDC) format indicating the drug, device, or medical supply covered by this claim. |

Table 4. LT variables to retain

|  |  |
| --- | --- |
| Variable name | Variable description |
| Header and line | |
| DA\_RUN\_ID | A unique identifier that identifies the TAF production run that produced the TAF file. |
| CLM\_ID | CCW claim identifier used to link headers and lines |
| LT\_FIL\_DT | Year and month of the reporting period. |
| LT\_VRSN | Indicator representing the iteration of the file. |
| MSIS\_ID | The encrypted state-assigned unique identification number used to identify and Medicaid/CHIP enrolled beneficiary and any claims submitted to the system. |
| SUBMTG\_STATE\_CD | The ANSI numeric state code for the U.S. state, territory, or the District of Columbia that has submitted the data. |
| Header only | |
| ADMSN\_DT | The date on which the recipient was admitted to the hospital. |
| DSCHRG\_DT | The date on which the recipient was discharged from the hospital. |
| BILL\_TYPE\_CD | A data element corresponding with UB-04 form locator FL4 that classifies the claim as to the type of facility (2nd digit), type of care (3rd digit) and the billing record's sequence in the episode of care (4th digit). (Note that the 1st digit is always zero.) |
| DGNS\_CD\_\* (1-12) | ICD-9/10-CM code found on the claim. |
| DGNS\_VRSN\_CD\_\* (1-12) | A flag that identified the coding system (ICD-9 or ICD-10) used for the diagnosis code. |
| CLM\_TYPE\_CD | A data element identifying what kind of payment is covered and distinguishes between claims that are for Medicaid or Medicaid-expansion, S-CHIP, and other types of claims. |
| SRVC\_BGN\_DT | For services received during a single encounter with a provider, the date the service covered by this claim was received. |
| SRVC\_END\_DT | For services received during a single encounter with a provider, the date the service covered by this claim was received. |
| Line only | |
| REV\_CNTR\_CD | A code which identifies a specific accommodation, ancillary service or billing calculation. |
| LINE\_SRVC\_BGN\_DT | For services received during a single encounter with a provider, the date the service covered by this claim was received. |
| LINE\_SRVC\_END\_DT | For services received during a single encounter with a provider, the date the service covered by this claim was received. |
| NDC | A code in National Drug Code (NDC) format indicating the drug, device, or medical supply covered by this claim |

Table 5. OT variables to retain

|  |  |
| --- | --- |
| Variable name | Variable description |
| Header and line | |
| DA\_RUN\_ID | A unique identifier that identifies the TAF production run that produced the TAF file. |
| CLM\_ID | CCW claim identifier used to link headers and lines |
| OT\_FIL\_DT | Year and month of the reporting period. |
| OT\_VRSN | Indicator representing the iteration of the file. |
| MSIS\_ID | The encrypted state-assigned unique identification number used to identify and Medicaid/CHIP enrolled beneficiary and any claims submitted to the system. |
| SUBMTG\_STATE\_CD | The ANSI numeric state code for the U.S. state, territory, or the District of Columbia that has submitted the data. |
| Header only | |
| BILL\_TYPE\_CD | A data element corresponding with UB-04 form locator FL4 that classifies the claim as to the type of facility (2nd digit), type of care (3rd digit) and the billing record's sequence in the episode of care (4th digit). (Note that the 1st digit is always zero.) |
| POS\_CD | A code indicating where the service was performed. |
| DGNS\_CD\_\* (1-2) | ICD-9/10-CM code found on the claim. |
| DGNS\_VRSN\_CD\_\* (1-2) | A flag that identified the coding system (ICD-9 or ICD-10) used for the diagnosis code. |
| CLM\_TYPE\_CD | A data element identifying what kind of payment is covered and distinguishes between claims that are for Medicaid or Medicaid-expansion, S-CHIP, and other types of claims. |
| SRVC\_BGN\_DT | For services received during a single encounter with a provider, the date the service covered by this claim was received. |
| SRVC\_END\_DT | For services received during a single encounter with a provider, the date the service covered by this claim was received. |
| Line only | |
| LINE\_PRCDR\_CD | A procedure code (ICD-9, ICD-10, CPT, HCPCS or other) used by the state to identify the procedures performed during the hospital stay referenced by this claim. |
| LINE\_PRCDR\_CD\_SYS | A flag that identifies the coding system used for the procedure code. |
| LINE\_PRCDR\_CD\_DT | The date upon which the procedure was performed. |
| LINE\_SRVC\_BGN\_DT | For services received during a single encounter with a provider, the date the service covered by this claim was received. |
| LINE\_SRVC\_END\_DT | For services received during a single encounter with a provider, the date the service covered by this claim was received. |
| REV\_CNTR\_CD | A code which identifies a specific accommodation, ancillary service or billing calculation. |
| NDC | A code in National Drug Code (NDC) format indicating the drug, device, or medical supply covered by this claim. |

Table 6. RX variables to retain

|  |  |
| --- | --- |
| Variable name | Variable description |
| Header and line | |
| DA\_RUN\_ID | A unique identifier that identifies the TAF production run that produced the TAF file. |
| CLM\_ID | CCW claim identifier used to link headers and lines |
| RX\_FIL\_DT | Year and month of the reporting period. |
| RX\_VRSN | Indicator representing the iteration of the file. |
| MSIS\_ID | The encrypted state-assigned unique identification number used to identify and Medicaid/CHIP enrolled beneficiary and any claims submitted to the system. |
| SUBMTG\_STATE\_CD | The ANSI numeric state code for the U.S. state, territory, or the District of Columbia that has submitted the data. |
| Header only | |
| RX\_FILL\_DT | Prescription fill date. |
| CLM\_TYPE\_CD | A data element identifying what kind of payment is covered and distinguishes between claims that are for Medicaid or Medicaid-expansion, S-CHIP, and other types of claims. |
| Line only | |
| NDC | A code in National Drug Code (NDC) format indicating the drug, device, or medical supply covered by this claim. |

Link line-level records to header records in the claim files

1. Merge the header records onto the line records for each of the four types of monthly claims files (IP, LT, OT, and RX TAF) for all 12 months of the year using the CLM\_ID. At the completion of this step, there should be 48 merged claim files: 12 monthly IP, 12 monthly LT, 12 monthly OT, and 12 monthly RX.

Restrict claims files to FFS claims and encounter records

1. Drop from each of the 48 files any records that represent payments that cannot be tied to specific services delivered to a beneficiary or have an unknown program type (not Medicaid or S-CHIP). The dropped records are those with the following values for CLM\_TYPE\_CD:
   * + CLM\_TYPE\_CD = 2 or B (capitation payments)
     + CLM\_TYPE\_CD = 4 or D (service tracking claims)
     + CLM\_TYPE\_CD = U, V, W, X, or Y (claims from an unknown program type)

Identify facility claims for SUD services and classify by SUD condition

1. In the IP and LT files, identify claims for SUD services as those with one or more lines that match one of the identification rules on the “SUD-overnight facility” sheet in the reference codes. See Table 7 for the TAF variable names corresponding to each field specified by the code type (type of bill, revenue code, ICD-9 procedure codes, ICD-10 procedure codes, and CPT codes).

Table 7. TAF variables name corresponding to each code type

|  |  |
| --- | --- |
| Code type | TAF variable name |
| Type of bill codes | BILL\_TYPE\_CD |
| Revenue codes | REV\_CD |
| ICD-9 procedure codes | PRCDR\_CD\_\*, where PRCDR\_CD\_SYS\_\* = 02 |
| ICD-10 procedure codes | PRCDR\_CD\_\*, where PRCDR\_CD\_SYS\_\* = 07 |
| CPT codes | PRCDR\_CD\_\*, where PRCDR\_CD\_SYS\_\* = 01 |
| HCPCS codes | PRCDR\_CD\_\*, where PRCDR\_CD\_SYS\_\* = 06 |

1. Among the claims identified in step 5A, further identify those header records that also have a diagnosis code in any position corresponding to one of the specified SUD diagnoses from any of the condition-specific tabs (coded green). For these claims, set the following flags:
   * + Set the flag SUD\_CLM\_FLAG as noted in Table 8. If more than one line within the claim qualifies for a SUD\_CLM\_FLAG value, set the minimum value across lines within the claim.

Table 8. Values of claims qualification rule flag

|  |  |
| --- | --- |
| Claims qualification rule | Flag value |
| 1+ claims | SUD\_CLM\_FLAG = 1 |
| 2+ claims on different dates of service | SUD\_CLM\_FLAG = 2 |
| Screening code - special requirements | SUD\_CLM\_FLAG = 3 |

* + - Next, create a condition-specific flag as noted in Table 9, based on the table in which the diagnosis code on the qualifying claim was found.

Table 9. Values of SUD claims-level condition flags

|  |  |
| --- | --- |
| Diagnosis code tab | Flag value |
| Alcohol Dx | ALCHL\_SUD\_DSRDR = 1 |
| Tobacco Dx | TBCCO\_SUD\_DSRDR = 1 |
| Cannabis Dx | CNNBS\_SUD\_DSRDR = 1 |
| Caffeine Dx | CFFNE\_SUD\_DSRDR = 1 |
| Hallucinogens Dx | HLLCNGN\_SUD\_DSRDR = 1 |
| Inhalants Dx | INHLNTS\_SUD\_DSRDR = 1 |
| Opioids Dx | OPIOIDS\_SUD\_DSRDR = 1 |
| Sedatives Dx | SHA\_SUD\_DSRDR = 1 |
| Stimulants Dx | STMLNTS\_SUD\_DSRDR = 1 |
| Other and Unk Dx | OTHER\_SUD\_DSRDR = 1 |
| Polysubstance Dx | PLYSBSTNCE\_SUD\_DSRDR = 1 |

1. In the OT file, identify facility claims for SUD services as those with one or more lines that match one of the requirements on the “SUD-outpatient facility” sheet. See Table 7 for the TAF variable name corresponding to each field specified by the code type (revenue codes, HCPCS codes, and CPT codes).
2. Among the claims identified in step 5C, further identify those header records that also have a diagnosis in any position corresponding to one of the specified SUD diagnoses from any of the condition-specific tabs (coded green). For these claims, set the SUD\_CLM\_FLAG and condition-specific flags as noted in Table 8 and Table 9, respectively. If more than one line within the claim qualifies for a SUD\_CLM\_FLAG value, set the minimum value across lines within the claim.

Identify professional claims for SUD services and classify by SUD condition

1. In the OT file, identify professional claims for SUD services as those that match one or more of the requirements on the “SUD-professional” sheet in the reference codes (coded purple). See Table 10 for the TAF variable name corresponding to each field specified by the code types (revenue codes, HCPCS codes, CPT codes, and place of service codes).

Table 10. Code type to TAF variable crosswalk for professional claims

|  |  |
| --- | --- |
| Code type | TAF variable name |
| CPT codes | LINE\_PRCDR\_CD\_\* & LINE\_PRCDR\_SYS\_\* = 01 |
| HCPCS codes | LINE\_PRCDR\_CD\_\* & LINE\_PRCDR\_SYS\_\* = 06 |
| Place of service codes | POS\_CD = 21 or 51 |

1. Among the claims identified in step 6A, further identify those header records that also have a diagnosis in any position corresponding to one of the specified SUD diagnoses from any of the condition-specific tabs (coded green). For these claims, set the SUD\_CLM\_FLAG and condition-specific flags as noted in Table 8 and Table 9 above, respectively. If more than one line within the claim qualifies for a SUD\_CLM\_FLAG value, set the minimum value across lines within the claim.

Identify prescription drug claims for SUD and classify by SUD condition

1. In all claims files (RX, IP, LT, and OT), identify SUD prescription claims as those with an NDC\_CODE value that matches to one of the NDC codes in the condition-specific prescription drug tabs for alcohol, tobacco, and opioids.
2. For the claims identified in step 7A, set a SUD\_CLM\_FLAG = 1. Additionally, set the condition-specific flags as noted in Table 11 based on the table in which the NDC code was found. The one exception is for NDC values that map to the non-proprietary drug name “naltrexone”. Since that drug is used to treat both opioid use disorder and alcohol use disorder, instead of assigning a condition flag, create a new flag NLTRXNE\_RX = 1 for these claims.[[3]](#footnote-4)

Table 11. SUD claims-level condition flags for prescription drugs

|  |  |
| --- | --- |
| Prescription drug code tab | Flag value |
| Alcohol Rx | ALCHL\_SUD\_DSRDR = 1 |
| Tobacco Rx | TBCCO\_SUD\_DSRDR = 1 |
| Opioids Rx | OPIOIDS\_SUD\_DSRDR = 1 |

De-duplication of medical claims to count no more than one qualifying claim per day

1. Use the service dates defined in Table 12 below for each file type to identify records within the file that share the same date of service for a unique beneficiary (as identified by MSIS\_ID). When all the lines within a claim do not share the same value for the service date value, take the maximum value across all lines to determine the service date for the whole claim.

Table 12. Service date variables for each TAF file

|  |  |
| --- | --- |
| File | Service date variable |
| IP | DSCHRG\_DTa |
| LT | SRVC\_END\_DTb |
| OT | SRVC\_END\_DTc |

a If this value is missing, use ADMSN\_DATE, SRVC\_BGN\_DT, or SRV\_END\_DT on the header level or LINE\_SRVC\_END\_DT or LINE\_SRVC\_BGN\_DT on the line level.

b If this value is missing, use SRVC\_BGN\_DT, DISCHRG\_DT, or ADMSN\_DATE on the header level, or LINE\_SRVC\_END\_DT or LINE\_SRVC\_BGN\_DT on the line level.

c If this value is missing, use SRVC\_BGN\_DT on the header level or LINE\_SRVC\_END\_DT or LINE\_SRVC\_BGN\_DT on the line level.

1. For any claims within a file that share the same SUBMTG\_STATE\_CD, MSIS\_ID, and service date, collapse the claims to a single record and ensure that record takes:
   * + The minimum value of the SUD\_CLM\_FLAG across all collapsed claims
     + The minimum non-missing value for each condition-specific flag
     + The maximum value of any binary variables

Flag beneficiaries receiving SUD services

1. Keeping only the claims with a non-missing SUD\_CLM\_FLG value, concatenate the IP, LT, OT, and RX claims and sort by state and beneficiary. Loop through the flagged claims and assign new beneficiary-level condition-specific flags called POP\_SUD\_\* if the beneficiary meets any one of the following criteria:
   * + One claim with a condition-specific flag set to 1 (e.g., ALCHL\_SUD\_DSRDR = 1)
     + Two claims with different dates of service with both condition-specific flags set to 2 (e.g., ALCHL\_SUD\_DSRDR = 2 on both claims)
     + Two claims with different dates of service, with the condition-specific flag set to 2 on one claim and to 3 on the other (e.g., ALCHL\_SUD\_DSRDR = 2 on the first claim and ALCHL\_SUD\_DSRDR = 3 on the second claim)
     + Apply this logic to all \*\_SUD\_DSRDR flags.
2. For beneficiaries with one or more claims where NLTRXNE\_RX = 1, use the following logic to set the beneficiary-level condition-specific flags:
   * + If POP\_SUD\_ALCHL is set to missing, then set POP\_SUD\_OPIOIDS = 1
     + If POP\_SUD\_ALCHL = 1, then keep POP\_SUD\_ALCHL = 1 and do not change the value of the POP\_SUD\_OPIOIDS flag
3. Set POP\_SUD\_PLYSBSTNCE\_SUD=1 for all beneficiaries with POP\_SUD\_\* flags for more than one condition type (e.g., if POP\_SUD\_ALCHL=1 and POP\_SUD\_OPIOIDS=1).
4. For all beneficiaries with a beneficiary-level condition-specific POP\_SUD\_\* flag set to 1, also set POP\_SUD = 1.

IV. Limitations

This algorithm relies on administrative claims data to identify Medicaid beneficiaries receiving SUD-related services. The use of administrative claims data has advantages such as ease of access and accuracy regarding type and timing of services. However, there are important limitations to using administrative claims data to identify SUD services. First, it can identify only those Medicaid beneficiaries who received treatment for a SUD as part of a Medicaid-funded service. In addition, the accuracy of diagnoses listed on claims depends on the provider’s clinical knowledge of SUDs, the resources available to the provider for evaluating the patient’s condition, and administrative factors such as accurate transcription from medical records/notes to electronic billing systems. Also, despite concerted efforts to combat the stigma associated with SUD, some clinicians may be reluctant to list a SUD diagnosis on a medical claim, and many individuals with SUD do not seek treatment. Thus, results from administrative claims data are likely to underestimate the proportion of Medicaid recipients with SUD.

Another limitation of administrative claims data is the potential for misidentification of individuals who receive services related to substance use but do not meet diagnostic criteria for a SUD (for example, a patient who receives a SUD screening or assessment). To limit misidentification, the logic requiring a beneficiary to have at least two services for the same SUD on different dates to be included in the population, with the exception of inpatient SUD treatment and pharmacy fills for certain types of medication-assisted treatment that are not used for non-SUD conditions.

Treatment for SUD includes inpatient, outpatient, long-term care, and medication-related services. State Medicaid programs vary widely in terms of which SUD-related services are covered and in which settings. For example, some state Medicaid programs cover all FDA-approved medications for treating OUD, whereas other programs cover only some of these medications. Therefore, if making state-to-state comparisons, it is important to consider whether the differences in the number of beneficiaries receiving SUD services are driven by the services each state has chosen to cover under its Medicaid program—and thus the accessibility of different forms of SUD treatment.

Julia Baller1, Allison Barrett1, Preeti Gill1, Andres Arguello1, Puja Nair1, Kimberly Proctor2, Jessie Parker2. Jeffrey Galecki “Substance Use Disorder (SUD) Tool 1: Identifying Beneficiaries with a Treated SUD On-demand Software Documentation.” Baltimore, MD: CMS, February 2021.

Reviewers: Keith Branham2, Carol V. Irvin1

1Mathematica, 2Centers for Medicare and 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](mailto:MACBISData@cms.hhs.gov)

1. Some claims also require a place-of-service code for identification. [↑](#footnote-ref-2)
2. This methodology may result in the under-reporting of beneficiaries who have only one treatment episode because diagnosis codes do not always capture untreated disorders. Similarly, beneficiaries who only require a lower level of treatment for their SUD may not be captured in administrative data. [↑](#footnote-ref-3)
3. When claims are rolled up to the beneficiary level, these claims will be assigned to either the alcohol or opioid use disorder categories based on the presence or absence of other claims for alcohol use disorder for the beneficiary. [↑](#footnote-ref-4)