CMS
CENTERS FOR MEDICARE & MEDICAID SERVICES

IAP | Medicaid Innovation Accelerator Program

Identifying Pregnant and Postpartum Beneficiaries in Medicaid and CHIP Administrative Data

Technical Specifications

January 2023

MACBIS
Medicaid and CHIP
Business Information Solutions

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

Contents

Tables

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

I. Description

This guide provides technical specifications with a general logic for using Medicaid and Children's Health Insurance Program (CHIP) administrative data to identify beneficiaries who were pregnant or postpartum in a calendar year. The Maternal and Infant Health (MIH): Identifying Pregnant and Postpartum Beneficiaries in Medicaid and CHIP Administrative Data Technical Specifications enable users to identify beneficiaries in four pregnancy-related categories, as shown in Table 1:

1. Beneficiaries pregnant at any point during the calendar year (Category 1)
2. Beneficiaries who had a live birth during the calendar year (Category 2)
3. Beneficiaries who had a miscarriage, stillbirth, or termination during the calendar year (Category 3)
4. Beneficiaries who had a delivery during the calendar year but the outcome (live birth, stillbirth, miscarriage, or termination) could not be determined from administrative data (Category 4)

These specifications are designed to use one calendar year of data. The target population includes female beneficiaries ages 8 to 64[1] ever enrolled in Medicaid or CHIP (fee-for-service or managed care) during the calendar year. The specifications do not impose continuous enrollment requirements (that is, contiguous months of enrollment as a recipient of full Medicaid or CHIP benefits), because their purpose is to identify and produce counts of pregnant and postpartum beneficiaries covered by Medicaid and CHIP regardless of their length of enrollment. Additionally, some women may be eligible for Medicaid or CHIP coverage and benefits just because they are pregnant and so would not have a full year of continuous enrollment in either program. Users can impose limits on continuous enrollment at their discretion.

These specifications rely on a list of procedure, revenue, and diagnosis codes to identify pregnancy and delivery-related claims. (See the accompanying documentation: Maternal and Infant Health (MIH): Identifying Pregnant and Postpartum Beneficiaries in Medicaid and CHIP Administrative Data Reference Codes for a full list of relevant codes.) After identifying relevant headers and lines on claims and creating claims-level indicator variables, users roll up to the beneficiary-level to calculate the number of beneficiaries in each pregnancy-related category. Each beneficiary is counted once in each category that applies to her, even if she has multiple pregnancies in the calendar year.

[1] Although the specifications are intended to identify outcomes for female beneficiaries of reproductive age (ages 15 to 49), we define the age range as 8 to 64 to capture more pregnancies, including those outside the traditional range, and to exclude clear coding errors for beneficiaries in the pediatric or geriatric age ranges.

Table 1. Descriptions of MIH categories

Beneficiary-level categories	Description	Notes and considerations
Category 1: Ever pregnant	Identifies women pregnant at any time during the calendar year, including those still pregnant at the end of the calendar year and those who had a claim indicating delivery, live birth, stillbirth, miscarriage, or termination during the calendar year.	All women included in Category 2: live birth, Category 3: miscarriage, stillbirth, or termination, and Category 4: delivery outcome unknown will also be included in this category.
Category 2: Live birth	Identifies women who had a claim for a live birth during the calendar year.	Categories 2: live birth and 3: miscarriage, stillbirth, or termination are not mutually exclusive. Women who have a miscarriage, stillbirth, or termination *and* a live birth during the calendar year will be included in both categories. Similarly, women who have a multiple pregnancy (twins, triplets, and so on) with different birth outcomes (such as one live birth and one stillbirth) will be included in both categories if the claims files contain codes associated with both outcomes.
Category 3: Miscarriage, stillbirth, or termination	Identifies women who had a claim indicating miscarriage, stillbirth, or termination during the calendar year.	Category 2: live birth and Category 3: miscarriage, stillbirth, or termination are not mutually exclusive. Women who have a miscarriage, stillbirth, or termination *and* a live birth during the calendar year will be included in both categories. Similarly, women who have a multiple pregnancy (twins, triplets, etc.) with different birth outcomes (such as one live birth and one stillbirth) will be included in both categories if the claims files contain codes associated with both outcomes.
Category 4: Delivery outcome unknown	Identifies women with claims indicating that a delivery occurred but the claims do not provide enough information to classify the delivery as a Category 2: live birth or Category 3: miscarriage, stillbirth, or termination.	Category 4: delivery outcome unknown is a catchall category that is mutually exclusive with Categories 2: live birth and 3: miscarriage, stillbirth, or termination. By definition, women with claims that have enough detail to be in Category 2 or Category 3 will never be included in Category 4.

II. Logic overview

These technical specifications lay out a five-step algorithm to identify pregnant and postpartum beneficiaries using the Transformed Medicaid Statistical Information System (T-MSIS) Analytic Files (TAF) Research Identifiable Files (RIF) data, described in Table 2 below. At a high level, these specifications include the following steps:

1. Identify the target population (step 1)
2. Identify pregnancy- and delivery-related claims (steps 2–4)
3. Roll up claims-level variables to construct beneficiary-level variables that align with the four pregnancy-related categories (step 5)

For each step, the specifications 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 Annual Demographic and Eligibility (DE), Inpatient (IP), and Other Services (OT) files.

Table 2. Logic overview for identifying pregnant and postpartum beneficiaries

Steps	Relevant TAF RIF	Logic and purpose	Implementation using TAF RIF
1. Identify beneficiaries in the enrollment file who qualify for inclusion	DE file	Identify beneficiaries who were ever enrolled in Medicaid or CHIP in the calendar year, are female, and are ages 8 to 64, to include in the analysis.	Use the reference year (RFRNC_YR) to identify records that correspond to the calendar year of interest. Select the most recent file version using the variable DE_VRSN. Drop records in which there is no eligibility information (MISG_ELGBLTY_DATA_IND = 1) as these claim records do not contain any enrollment information about the beneficiary. Limit to female beneficiaries (SEX_CD = F). Calculate beneficiary age (using BIRTH_DT) and limit to ages 8 to 64.
2. Merge claims and enrollment data to create an annual IP file and an annual OT file for beneficiaries who qualify for inclusion	IP, OT, and DE files	Only those claims matching to a beneficiary who qualifies for inclusion in the target population should be retained. To reduce run-time, we recommend creating a finder file of beneficiaries from Step 1 and only pulling claims that match their beneficiary ID.	Read in IP and OT header records for beneficiaries identified in the target population created in Step 1, based on MSIS identification number (MSIS_ID) and state (SUBMTG_STATE_CD). Select the most recent file version date for each month using the variables IP_VRSN and OT_VRSN. Drop records that represent payments and cannot be tied to a specific service delivered to the beneficiary (CLM_TYPE_CD = 2, 4, B, D, V, X). Stack the monthly IP and OT claim files (keeping the files separate) to create an annual IP file and an annual OT file. Stack months for January through December of the calendar year (for example, 2020 would be 01012020 through 12312020).
3. Join the header and line-level files	IP and OT files	Each monthly IP and OT TAF RIF 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 pregnant and postpartum beneficiaries, including revenue codes (on facility claims only) and CPT or HCPCS procedure codes. Additional header-level data elements are also required to identify pregnant and postpartum beneficiaries, including diagnosis codes and ICD procedure codes (on inpatient facility 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 a pregnancy or postpartum service.	Separately for the stacked IP and OT files, link the header- and line-level files using the unique combination MSIS ID (MSIS_ID), submitting state code (SUBMTG_STATE_CD), and claim identifier (CLM_ID).

Table 2 (*continued*)

Steps	Relevant TAF RIF	Logic and purpose	Implementation using TAF RIF
4. Identify claims for each MIH outcome	IP and OT files	Identify and create claim-level flags for each of the MIH categories by using diagnosis codes, procedure codes, and revenue center codes.	Using the files created in Step 3, create a binary (0/1) claim-level flag for each of the MIH categories from the Maternal and Infant Health (MIH): Identifying Pregnant and Postpartum Beneficiaries in Medicaid and CHIP Administrative Data Reference Codes list. In the IP file, look for claims with any of the MIH-related codes from the reference codes list in the diagnosis codes (DGNS_CD_1 to DGNS_CD_12), admitting diagnosis codes (ADMTG_DGNS_CD), procedure codes (PRCDR_1_CD - PRCDR_6_CD), and revenue center codes (REV_CNTR_CD). In the OT file, look for claims with any of the MIH-related diagnosis codes, procedure codes, and revenue codes from the reference codes list in the DGNS_CD_1 and DGNS_CD_2 variables, procedure codes (LINE_PRCDR_CD), and revenue center codes (REV_CNTR_CD). If the user identifies an MIH-related claim, set the corresponding claim-level flag equal to 1 (e.g., if the user identifies a code from the "Ever pregnant" code list on a claim line, set the flag for "ever pregnant" to 1 for that claim). Otherwise, set the claim-level flag equal to 0. Drop all claim-level records where all flags are equal to 0.
5. Create beneficiary-level indicator variables	IP, OT, and DE files	Roll up claims to the beneficiary level to identify and flag the pregnant and postpartum categories for Medicaid and CHIP beneficiaries.	Stack the IP and OT files into a combined file, keeping only MSIS_ID, STATE_CD, CLM_ID and the constructed claim-level flags. Roll up the claims-level flags to create deduplicated annual beneficiary-level variables for the following categories: 1. Category 1: ever pregnant. Equal to 1 if either the IP or OT annual beneficiary-level flags for "ever pregnant" is equal to 1, otherwise set to 0. 2. Category 2: live birth. Equal to 1 if either the IP or OT annual beneficiary-level flags for "live birth" is equal to 1, otherwise set to 0. 3. Category 3: miscarriage, stillbirth, or termination. Equal to 1 either the IP or OT annual beneficiary-level flags for "miscarriage, stillbirth, or termination" is equal to 1, otherwise set to 0. 4. Category 4: delivery outcome unknown. Equal to 1 if either the IP or OT annual beneficiary-level flags for "labor & delivery, outcome unknown" is equal to 1 and the IP and OT annual beneficiary-level flags for both "live birth" and "miscarriage, stillbirth, or termination" are equal to 0, otherwise set to 0.

[a] DE = demographic and eligibility file; IP = inpatient file; OT= other services file; TAF= Transformed Medicaid Statistical Information System Analytic File; RIF= research Identifiable file

[b] All beneficiaries with at least 1 IP or OT claim with codes in the "1: ever pregnant" tab of the Identifying Pregnant and Postpartum Beneficiaries: Reference Codes list will be categorized as pregnant during the calendar year. Pregnancy outcomes are also identified based on claim codes for three additional categories listed in the tabs: "2: live birth," "3: miscarriage, stillbirth, or termination," and "4: delivery outcome unknown." Women identified in these three categories will always be classified as pregnant. Note that the live birth and miscarriage, stillbirth, or termination categories are not mutually exclusive, as it is possible for a beneficiary to have two separate pregnancies with different outcomes during the calendar year. Additionally, women who have a multiples pregnancy (twins, triplets, and so on) could have different outcomes for the same pregnancy. Only women who are identified as pregnant during the calendar year who have delivery-related claims with an unknown outcome and who also do not have claims for a live birth or a miscarriage, stillbirth, or termination during the calendar year are classified in Category 4: delivery outcome unknown.

III. Detailed technical specifications

The detailed technical specifications below describe how the algorithm to identify pregnant and postpartum beneficiaries can be applied to the TAF RIF.

1. Identify beneficiaries in the DE file who qualify for inclusion

Conduct initial processing of the DE claims file to limit its size and to facilitate linking.

a. Read in the DE file and use the reference year (RFRNC_YR) to identify records that correspond to the four-digit calendar year of interest (for example, to analyze calendar year 2020, read in data where RFRNC_YR = 2020). Select the most recent file version date using the variable DE_VRSN Keep only the DE variables that are needed for the analysis (Table 3).
b. Remove records in which there are no eligibility data (MISG_ELGBLTY_DATA_IND = 1), as these claim records do not contain information for how long a beneficiary was enrolled in Medicaid or CHIP.
c. Limit the file to female beneficiaries (SEX_CD = F).
d. Calculate a beneficiary's age during the calendar year by calculating January 1 in given year - BIRTH_DT. Limit the file to beneficiaries ages 8 to 64.

Table 3. DE variables to retain

TAF RIF variable name(s)	Description
BIRTH_DT	Beneficiary's date of birth; most recent in the calendar and all prior years
DE_VRSN	Indicator representing the iteration of the file
MISG_ELGBLTY_DATA_IND	A flag to indicate that the person had claims for the year but no eligibility information
MSIS_ID	A state-assigned unique identification number used to identify a Medicaid/CHIP-enrolled beneficiary
RFRNC_YR	The year of the data file
SEX_CD	The beneficiary's biological sex; most recent in the calendar and all prior years
SUBMTG_STATE_CD	The numeric state code for the U.S. state, territory, or the District of Columbia that has submitted the data

2. Identify claims for beneficiaries who qualify for inclusion

Conduct initial processing of the monthly IP and OT claims files to limit their size and to facilitate linking.

a. Read in all monthly IP and OT header records for the year separately for each file, restricting the set of variables retained to those identified in Tables 4 and 5 below for the most recent version of the file (based on _VRSN). 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. Only read in records for beneficiaries who qualify for inclusion in Step 1, based on MSIS identification number (MSIS_ID) and state (SUBMTG_STATE_CD).
b. Drop records that represent payments and cannot be tied to a specific service delivered to the beneficiary (CLM_TYPE_CD = 2, 4, B, D, V, X).

c. Stack the monthly IP and OT claim files (keeping the files separate) to create an annual IP file and an annual OT file. Stack months for January through December of the calendar year (for example, 2020 would be 01012020 through 12312020).

Table 4. IP variables to retain

Variable name	Variable description
Header and line	
IP_VRSN	Indicator representing the iteration of the file
CLM_ID	The unique identification number for the claim
MSIS_ID	The encrypted state-assigned unique identification number used to identify a 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	
ADMTG_DGNS_CD	ICD-10-CM admitting diagnosis code
CLM_TYPE_CD	A code indicating what kind of payment is covered in this claim.
DGNS_CD_1 - DGNS_CD_12	ICD-10-CM code found on the claim
PRCDR_1_CD - PRCDR_6_CD	A procedure code (CPT or HCPCS) used by the state to identify the procedures performed during the hospital stay referenced by this claim
Line only	
REV_CNTR_CD	A code that identifies a specific accommodation, ancillary service, or billing calculation

Table 5. OT variables to retain

Variable name	Variable description
Header and line	
OT_VRSN	Indicator representing the iteration of the file
CLM_ID	The unique identification number for the claim
MSIS_ID	The encrypted state-assigned unique identification number used to identify a 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	
CLM_TYPE_CD	A code indicating what kind of payment is covered in this claim
DGNS_CD_1 - DGNS_CD_2	ICD-10-CM admitting diagnosis code
Line only	
LINE_PRCDR_CD	A procedure code (CPT or HCPCS) used by the state to identify the procedures performed during the hospital stay referenced by this claim
REV_CNTR_CD	A code that identifies a specific accommodation, ancillary service, or billing calculation

3. Join the header- and line-level files

Link each header-level record with its associated line-level records to evaluate whether the claim qualifies as a pregnancy or postpartum service.

a. To link header-level records with associated line-level records, separately for the IP and OT files, link the header- and line-level files using the beneficiary identifier (MSIS_ID), the submitting state code (SUBMTG_STATE_CD), and claim identification (CLM_ID).

4. Identify claims for each MIH outcome

Identify and create claim-level flags for each of the MIH categories by using diagnosis codes, procedure codes, and revenue center codes.

a. Using the files created in Step 3, create a binary (0/1) claim-level flag for each of the MIH categories from the Identifying Pregnant and Postpartum Beneficiaries: Reference Codes list. If the claim has any of the procedure codes, diagnosis codes, or revenue codes for the relevant category, set the claim-level flag equal to 1. Otherwise, set the claim-level flag equal to 0.
    i. In the IP file, look for claims with any of the MIH-related codes from the reference codes list in the diagnosis codes (DGNS_CD_1 to DGNS_CD_12), admitting diagnosis codes (ADMTG_DGNS_CD), procedure codes (PRCDR_1_CD - PRCDR_6_CD), and revenue center codes (REV_CNTR_CD).
    ii. In the OT file, look for claims with any of the MIH-related diagnosis codes, procedure codes, and revenue codes from the reference codes list in the DGNS_CD_1 and DGNS_CD_2 variables, procedure codes (LINE_PRCDR_CD), and revenue center codes (REV_CNTR_CD).
b. If the user identifies an MIH-related procedure, diagnosis, or revenue code, set the corresponding claim-level flag equal to 1 (for example, if the user identifies a code from the “ever pregnant” code list on a claim line, then set the flag for “ever pregnant” to 1 for that claim line). Otherwise, set the claim-level flag equal to 0.
c. Drop all claim-level records where all flags are equal to 0.

5. Create a beneficiary-level file

Roll up claims to the beneficiary-level to identify and flag the pregnant and postpartum categories for Medicaid and CHIP beneficiaries.

a. Stack the IP and OT claim-level variables into a combined IP and OT file, keeping only the MSIS ID, STATE_CD, CLM_ID and the constructed claim-level flags. Roll up the claims-level flags to create annual beneficiary-level variables that take the maximum of each value for each category, for the following categories:
    i. Category 1: Ever pregnant. Equal to 1 if either the IP or OT annual beneficiary-level flags for “ever pregnant” is equal to 1, otherwise set to 0.
    ii. Category 2: Live birth. Equal to 1 if either the IP or OT annual beneficiary-level flags for “live birth” is equal to 1, otherwise set to 0.
    iii. Category 3: Miscarriage, stillbirth, or termination. Equal to 1 either the IP or OT annual beneficiary-level flags for “miscarriage, stillbirth, or termination” is equal to 1, otherwise set to 0.

iv. Category 4: Delivery outcome unknown. Equal to 1 if either the IP or OT annual beneficiary-level flags for “labor & delivery, outcome unknown” is equal to 1 and the IP and OT annual beneficiary-level flags for both “live birth” and “miscarriage, stillbirth, or termination” are equal to 0, otherwise set to 0.

IV. Limitations

These technical specifications have the following limitations:

- **The algorithm does not identify unique pregnancy episodes within the same category.** As a result, the algorithm should not be used to count the number of pregnancies, live births, or miscarriages during the calendar year. It should only be used to count the *number of beneficiaries* in each of these categories. For instance, women who give birth twice during the calendar year (for example, in January and December) will be counted only once in Category 1: ever pregnant and once in Category 2: live birth. Similarly, women with a multiple birth (twins, triplets, and so on) will be counted only once in Category 1: ever pregnant and once in Category 2: live birth, despite giving birth to more than one infant.
- **The algorithm does not identify pregnant or postpartum beneficiaries who did not receive pregnancy-related services or diagnoses during the calendar year.** This limitation is most likely to affect women who became pregnant toward the end of the calendar year or those who were early in their pregnancy during the calendar year and less likely to receive frequent prenatal services relative to women in the later stages of pregnancy. Similarly, the results may underestimate women who had a miscarriage, stillbirth, or termination (Category 3), because not all miscarriages or terminations will result in a medical claim or be covered by Medicaid, or both. Claims for terminations are unlikely to be included in the TAF RIF because, other than a few limited exceptions, states are prohibited from using federal Medicaid funding to pay for terminations.[2]
- **Coding errors and data quality issues can affect the accuracy of the results.** The logic of the algorithm assumes that the procedure codes, diagnosis codes, and revenue codes are used correctly; however, there is evidence of coding errors and other limitations in some states' TAF RIF data. Errors in the data will result in the misclassification of beneficiaries. Further, data quality issues introduced by states during the reporting process may affect the accuracy of the results. States with serious data quality issues are not included in the final Medicaid public use file output.

[2] See https://www.congress.gov/bill/113th-congress/senate-bill/142.

Emily Harrison[1], Lauren Hula[1], Amanda Mims[1], Mattan Alalouf[1]. “Maternal and Infant Health: Identifying Pregnant and Postpartum Beneficiaries Technical Specifications.” Baltimore, MD: CMS, January 2023.

Reviewers: Jeffrey Galecki[2], Julia Baller[1], and Ben Schneedecker[1]

[1] Mathematica, [2] Centers 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