Johnson & Johnson   
Common Data Model (CDM v5.2)

ETL Mapping Specification for TRUVEN (MDCD)

Qianli Ma, Erica Voss, Clair Blacketer

Version 8

December 12, 2017

**Table of Contents**

[0 Change Log 3](#_Toc475696898)

[1 Introduction 4](#_Toc475696899)

[1.1 Abbreviations 4](#_Toc475696900)

[1.2 Conventions Used in Document 4](#_Toc475696901)

[1.3 Processing Sequence Map 5](#_Toc475696902)

[2 Data Mapping 7](#_Toc475696903)

[2.1 Table Name: LOCATION 7](#_Toc475696904)

[2.2 Table Name: PERSON 8](#_Toc475696905)

[2.3 Table Name: OBSERVATION\_PERIOD 10](#_Toc475696906)

[2.4 Table Name: PAYER\_PLAN\_PERIOD 11](#_Toc475696907)

[2.5 Table Name: PROVIDER 13](#_Toc475696908)

[2.6 Table Name: VISIT\_OCCURRENCE 15](#_Toc475696909)

[2.7 Table Name: CONDITION\_OCCURRENCE 21](#_Toc475696910)

[2.8 Table Name: CONDITION\_ERA 25](#_Toc475696911)

[2.9 Table Name: PROCEDURE\_OCCURRENCE 27](#_Toc475696912)

[2.10 Table Name: PROCEDURE\_COST 32](#_Toc475696913)

[2.11 Table Name: DEATH 34](#_Toc475696914)

[2.12 Table Name: DRUG\_EXPOSURE 37](#_Toc475696915)

[2.13 Table Name: DRUG\_ERA 41](#_Toc475696916)

[2.14 Table Name: DRUG\_COST 43](#_Toc475696917)

[2.15 Table Name: OBSERVATION 45](#_Toc475696918)

[2.16 Table Name: ORGANIZATION 47](#_Toc475696919)

[2.17 Table Name: CARE\_SITE 47](#_Toc475696920)

[2.18 Table Name: COHORT 47](#_Toc475696921)

[2.19 Table Name: COHORT\_ATTRIBUTE 48](#_Toc475696922)

[2.20 Table Name: MEASUREMENT 48](#_Toc475696923)

[2.21 Table Name: SPECIMEN 50](#_Toc475696924)

[2.22 Table Name: DEVICE\_EXPOSURE 51](#_Toc475696925)

[2.23 Table Name: NOTE 52](#_Toc475696926)

[2.24 Table Name: DEVICE\_COST 53](#_Toc475696927)

[2.25 Table Name: DOSE\_ERA 53](#_Toc475696928)

[2.26 Table Name: FACT\_RELATIONSHIP 54](#_Toc475696929)

[2.27 Table Name: CDM\_SOURCE 54](#_Toc475696930)

[2.28 Table Name: VISIT\_COST 55](#_Toc475696931)

[2.29 Table Name: CDM\_DOMAIN\_META 57](#_Toc475696932)

[3 Code Snippets 60](#_Toc475696933)

[3.1 Vocabulary Mapping 60](#_Toc475696934)

[3.2 Domain IDs 62](#_Toc475696935)

# 0 Change Log

**Version 8**

*12/12/2017*

* Conversion to CDM v5.2.0

**Version 7.2**

*9/7/2017*

* Updated DEATH logic to using STCM to identify death records rather than an embedded code list

**Version 7.1**

*7/31/2017:*

* Added AWP (average wholesale price) from the DRUG\_CLAIMS table to the cost table [HIX-1319](http://rndusrahpbl07.rndus.na.jnj.com:9080/browse/HIX-1319?jql=project%20%3D%20HIX%20AND%20fixVersion%20%3D%20%22CDM%20Sprint%20201707%22)
* Updated procedure code mapping for ICD10PCS [HIX-1255](http://rndusrahpbl07.rndus.na.jnj.com:9080/browse/HIX-1255?jql=project%20%3D%20HIX%20AND%20fixVersion%20%3D%20%22CDM%20Sprint%20201707%22)
* Updated CONDITION\_START\_DATE to service date if coming from INPATIENT\_SERVICES/OUTPATIENT\_SERVICES [HIX-1274](http://rndusrahpbl07.rndus.na.jnj.com:9080/browse/HIX-1274)

**Version 7.01**

*4/19/2017:* Adding documentation to address issues found in builder 1.9.0.10

*3/27/2017:* Set unknown/missing PROVIDER\_IDs and CARE\_SITE\_IDs to NULL instead of 0 so as to not interfere with primary key creation (HIX-1223)

**Version 7.0**

*3/3/2017-* Logic added to address HIX issues 1195, 1180, 1135, 942 and new cost table for version 5.0.1

*2/24/2017* – Added ICD10 logic and birth month/day logic for people born in the same year as their first OBSERVATION\_PERIOD

# Introduction

The purpose of this document is to describe the Extract, Transform, Load (ETL) mapping of the licensed data from Truven’s Medicaid database (MDCD) into the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM). MDCD is very similar as Truven’s Commercial Claims and Encounters (CCAE) and Medicare (MDCR) with the following differences:

1. MDCD doesn't have HEALTH\_RISK\_ASSESSMENT and LAB tables, thus OBSERVATION table will not be built for MDCD.
2. MDCD’s ENROLLMENT\_DETAIL table does not have EGEOLOC (location information), thus LOCATION table will not be built for MDCD. Also this table does not have PLANKEY and ENROLID does not contain family enrollment information, thus PLAN\_SOURCE\_VALUE and FAMILY\_SOURCE\_VALUE of PAYER\_PLAN\_PERIOD table will be set as NULL. The RX coverage information for this table is stored in DRUGCOVG and only valid when MEDICARE = 0; and MEDICARE =1 means the member is eligible for both Medicaid and Medicare coverage.
3. MDCD’s INPATIENT\_SERVICES and OUTPATIENT\_SERVICES tables do not have DX5.
4. MDCD’s DRUG\_CLAIMS does not have RXMR (mail order indicator).
5. MDCD’s ENROLLMENT\_DETAIL table has STDRACE, which contains race and ethnicity information for PERSON table.
6. MDCD has LONG\_TERM\_CARE table, which contains monthly or half-monthly billing information for long term care and can be used to define long term care visits.

## Abbreviations

|  |  |
| --- | --- |
| **Table 1: Abbreviations** | |
| **Abbreviation** | **Description** |
| ETL | Extract, Transform, Load |
| CCAE | Commercial Claims and Encounters |
| MDCR | Medicare Supplemental |
| MDCD | Medicaid |
| OMOP | Observational Medical Outcomes Partnership |
| CDM | Common Data Model |
| FIPS | Federal Information Processing Standard |
| ICD9 | The International Classification of Diseases, Ninth Revision |
| CPT | Current Procedural Terminology |
| HCPCS | Healthcare Common Procedure Coding System |
| DRG | Diagnosis-Related Groups |
| MDC | Major Diagnostic Categories |

## Conventions Used in Document

The following conventions are used within this document:

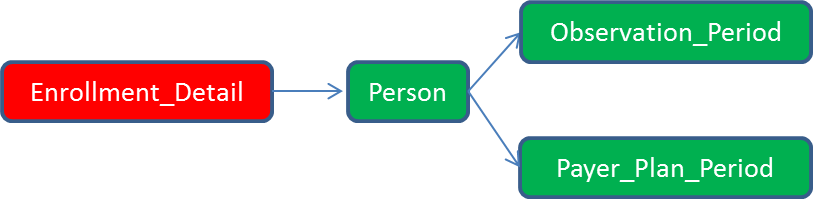
|  |  |
| --- | --- |
| **Table 2: Conventions** | |
| **Convention** | **Description** |
| - | Value does not exist |
| [X] | Value to be replaced or derived |
| *Italicized* | Referring to column in the table itself or another CDM table |

## Processing Sequence Map

This section describes the processing sequence to build CDM tables from Truven MDCD source data files.

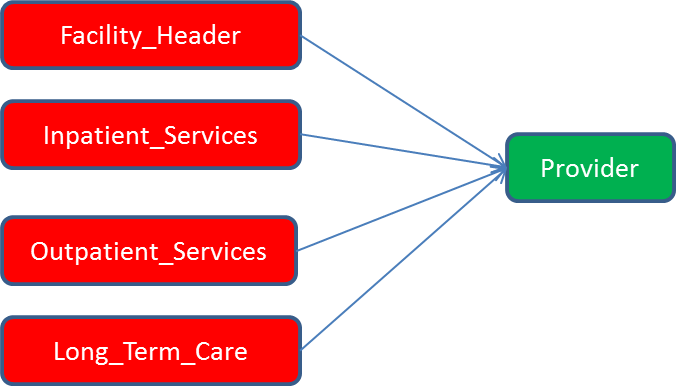
### Step I

Use MDCD ENROLLMENT\_DETAIL table to build PERSON, OBSERVATION\_PERIOD, and PAYER\_PLAN\_PERIOD tables:



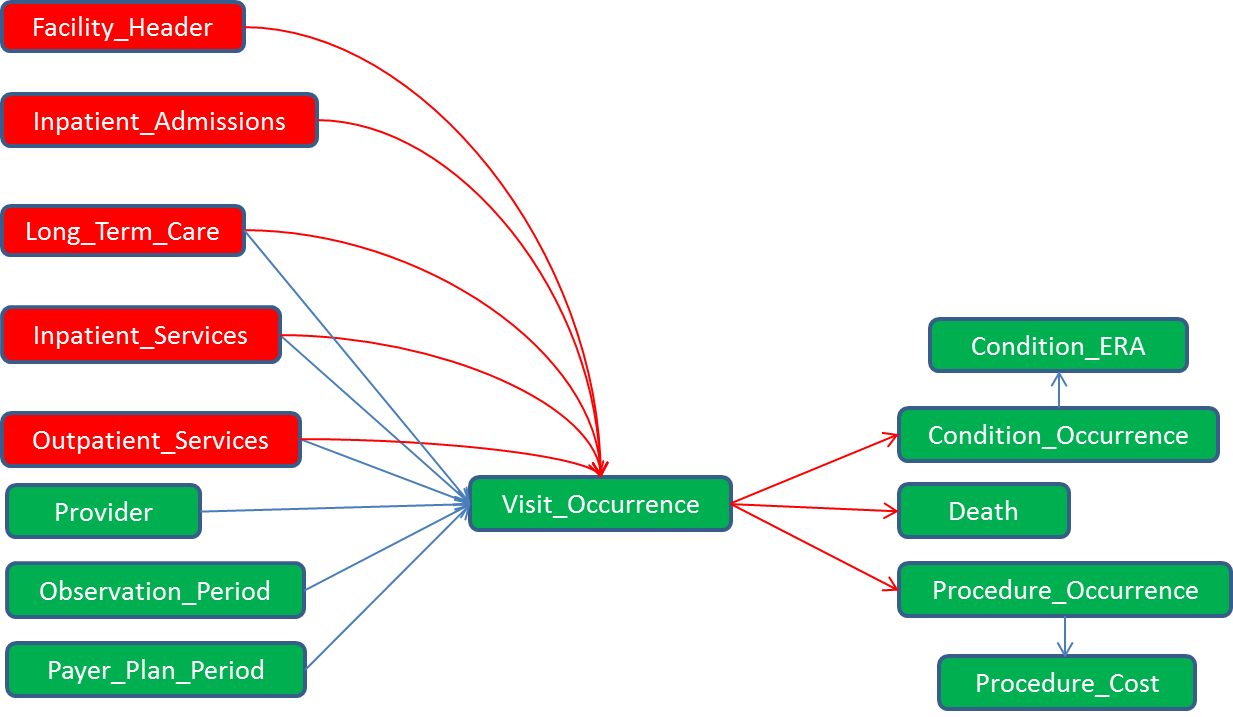
### Step II

Use MDCD FACILITY\_HEADER, INPATIENT\_SERVICES, OUTPATIENT\_SERVICES and LONG\_TERM\_CARE tables to build PROVIDER table:



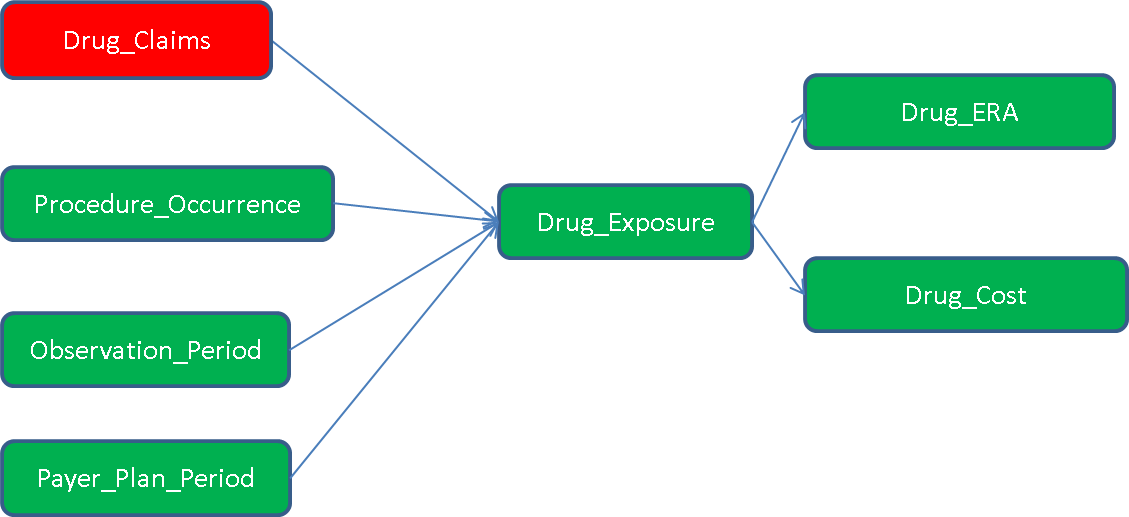
### Step III

Use OBSERVATION\_PERIOD, PAYER\_PLAN\_PERIOD, PROVIDER, and MDCD INPATIENT\_SERVICES, OUTPATIENT\_SERVICES and LONG\_TERM\_CARE tables to build VISIT\_OCCURRENCE. Then use VISIT\_OCCURRENCE, and MDCD INPATIENT\_SERVICES, OUTPATIENT\_SERVICES, LONG\_TERM\_CARE, FACILITY\_HEADER, INPATIENT\_ADMISSIONS tables to build CONDITION\_OCCURRENCE, CONDITION\_ERA, PROCEDURE\_OCCURRENCE, COST and DEATH tables:



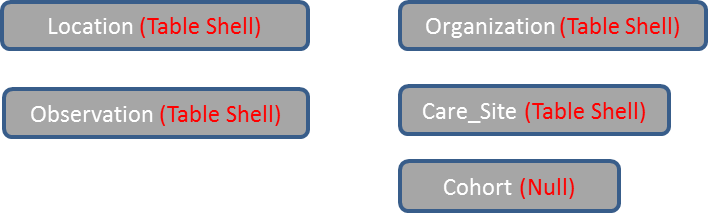
### Step IV

Use OBSERVATION\_PERIOD, PAYER\_PLAN\_PERIOD, PROCEDURE\_OCCURRENCE and Truven DRUG\_CLAIMS tables to build DRUG\_EXPOSURE, DRUG\_ERA and DRUG\_COST tables:



### Step V

Create table shells for LOCATION, OBSERVATION, ORGANIZATION and CARE\_SITE tables (No COHORT table will be populated):



# Data Mapping

This section describes how the source files are mapped into the CDM.

The following are a list of conventions used throughout all tables in this ETL:

* Since our data is already de-identified we do not need to mask our PERSON\_IDs.
* If a CONCEPT\_ID column cannot be mapped to a known code, it should be set to 0.
* If a non-CONCEPT\_ID column cannot be mapped to a known code, it should be set to NULL.
* Only pull records during observation periods where a person has both prescription benefits and medical benefits.
* Truven removes decimal points from ICD9 diagnosis and procedure codes so when mapping to the OMOP Vocabulary, the decimal points need to also be removed from the vocabulary in order to map between the source and the vocabulary.

## Table Name: LOCATION

MDCD does not have location information, so this table will only contain one value representing the fact that no location information will be captured.

| **Table 1: LOCATION** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| LOCATION\_ID | - | 0 |  |
| ADDRESS\_1 | - | NULL |  |
| ADDRESS\_2 | - | NULL |  |
| CITY | - | NULL |  |
| STATE | - | NULL |  |
| ZIP | - | NULL |  |
| COUNTY | **-** | NULL |  |
| LOCATION\_SOURCE\_VALUE | - | NULL |  |

## Table Name: PERSON

PERSON demographics are sourced from the MDCD ENROLLMENT\_DETAIL table.

Key conventions:

* Delete the following members: Year(GETDATE()) - max(DOBYR) < 90 AND max(DOBYR) > min(DOBYR) +2 or gender changed over different enrollment period.
* Only use records where the person has prescription benefits (DRUGCOVG =1) or eligible for both Medicaid and Medicare coverage (MEDICARE =1).
* If the person's other information changes (e.g. location, race, ethnicity), the last known record is used.
* Delete individuals whose DOBYR < 1900 or > the current year.
* After defining the patient’s DOBYR (the one on their last record), exclude any individual who was born > 1 year after their first enrollment period.
* The ENROLLMENT\_DETAIL table may store multiple records for each person. However, the CDM will only store one record per person in the PERSON table.
* For MONTH\_OF\_BIRTH, the *OBSERVATION\_PERIOD* table will need to be generated first. Then, the value can be derived from using YEAR\_OF\_BIRTH and OBSERVATION\_PERIOD\_START\_DATE only if the year of birth is equal to the year of the first OBSERVATION\_PERIOD\_START\_DATE.

| **Table 2: PERSON** | | | | |
| --- | --- | --- | --- | --- |
| **Destination Field** | **Source Field** | | **Applied Rule** | **Comment** |
| PERSON\_ID | **ENROLLMENT\_DETAIL:**  ENROLID | |  |  |
| GENDER\_CONCEPT\_ID | **ENROLLMENT\_DETAIL:**  SEX | | Map source values to their associated CONCEPT\_IDs:  1 - 8507  2 - 8532  If SEX is not 1 or 2 exclude that person. | The exclusion of a person by gender should happen on last enrollment record not just if they had one bad SEX record. |
| YEAR\_OF\_BIRTH | **ENROLLMENT\_DETAIL:**  DOBYR | | DOBYR needs to be > 1900 and <= current year. |  |
| MONTH\_OF\_BIRTH | - | | NULL |  |
| DAY\_OF\_BIRTH | - | | NULL |  |
| BIRTH\_DATETIME | - | | NULL |  |
| RACE\_CONCEPT\_ID | **ENROLLMENT\_DETAIL:**  STDRACE | | Map values of STDRACE to their associated CONCEPT\_IDs:  CASE  WHEN STDRACE ='1'  THEN **8527 /\*White\*/**  WHEN STDRACE ='2'  THEN **8516 /\*Black\*/**  ELSE **0 /\*OTHER\*/**  END AS RACE\_CONCEPT\_ID | Codes from MDCD:  1: White  2: Black  4: Hispanic  9: Other |
| ETHNICITY\_CONCEPT\_ID | **ENROLLMENT\_DETAIL:**  STDRACE | | Map values of STDRACE to their associated CONCEPT\_IDs:  CASE WHEN STDRACE ='4'  THEN **38003563 /\*Hispanic or Latino\*/**  ELSE **38003564**  END AS ETHNICITY\_CONCEPT\_ID |  |
| LOCATION\_ID | - | | NULL |  |
| PROVIDER\_ID | - | | NULL |  |
| CARE\_SITE\_ID | - | | NULL |  |
| PERSON\_SOURCE\_VALUE | **ENROLLMENT\_DETAIL:**  ENROLID | |  |  |
| GENDER\_SOURCE\_CONCEPT\_ID |  | | 0 |  |
| GENDER\_SOURCE\_VALUE | **ENROLLMENT\_DETAIL:**  SEX | |  |  |
| RACE\_SOURCE\_CONCEPT\_ID |  | | 0 |  |
| RACE\_SOURCE\_VALUE | **ENROLLMENT\_DETAIL:**  STDRACE |  | |  |
| ETHNICITY\_SOURCE\_CONCEPT\_ID |  | | 0 |  |
| ETHNICITY\_SOURCE\_VALUE | **ENROLLMENT\_DETAIL:**  STDRACE | |  |  |

## Table Name: OBSERVATION\_PERIOD

MDCD ENROLLMENT\_DETAIL table includes records that indicate a person’s enrollment for each month for the period covered by the claims data.

Enrollment entries are consolidated by combining records that indicate continuous enrollment over a period. Consolidation is done through the following steps:

* ENROLLMENT\_DETAIL records for each person are sorted in ascending order by DTSTART and DTEND.
* Periods of continuous enrollment are consolidated by combining monthly records as long as the time between the end of one enrollment period and the start of the next is 32 days or less (<=32).
* If there is a record with DTSTART > DTEND then set DTEND as equal to DTSTART

Key conventions:

* Only use records where the person has prescription benefits (DRUGCOVG =1) or eligible for both Medicaid and Medicare coverage (MEDICARE =1).
* The gap between observation periods needs to be 32 days or less (<=32).
* Remove duplicate records before assigning OBSERVATION\_PERIOD\_ID.

| **Table 3: OBSERVATION\_PERIOD** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| OBSERVATION\_PERIOD\_ID | - | System generated. |  |
| PERSON\_ID | **ENROLLMENT\_DETAIL:**  ENROLID |  |  |
| OBSERVATION\_PERIOD\_START\_DATE | **ENROLLMENT\_DETAIL:**  DTSTART | Minimum start date of a contiguous enrollment period. |  |
| OBSERVATION\_PERIOD\_END\_DATE | **ENROLLMENT\_DETAIL:**  DTEND | Maximum end date of a contiguous enrollment period. |  |
| PERIOD\_TYPE\_CONCEPT\_ID |  | 44814722-Period while enrolled in insurance |  |

## Table Name: PAYER\_PLAN\_PERIOD

For each person, create a record associating the person to the duration they were on a specific type of benefit plan (defined by MEDICARE, CAP and PLANTYP) according to MDCD ENROLLMENT\_DETAIL table.

Payer plan entries are consolidated by combining records that indicate continuous enrollment over a period for a specific payer plan. Consolidation is done through the following steps:

* Define the PAYER\_SOURCE\_VALUE (see logic in table below).
* ENROLLMENT\_DETAIL records for each person are sorted in ascending order by DTSTART, DTEND.
* Periods of continuous enrollment are consolidated by combining monthly records as long as the time between the end of one enrollment period and the start of the next is 32 days or less (<=32) and PAYER\_SOURCE\_VALUE has not changed.

Key conventions:

* Only use records where the person has prescription benefits (DRUGCOVG =1) or eligible for both Medicaid and Medicare coverage (MEDICARE =1).
* The gap between plan periods needs to be 32 days or less (<=32).
* Use MEDICARE (1 means eligible for both Medicaid and Medicare coverage), CAP (1 means capitated plan) and PLANTYP (e.g. HMO, PPO, etc.) to define PAYER\_SOURCE\_VALUE.
* Switch of type of plan (PAYER\_SOURCE\_VALUE) enrolled may cause the overlap of enroll periods on two plans: Sort ENROLLMENT\_DETAIL table by ENROLID, DTSTART and DTEND, and if there is some overlap between two coverage periods, always truncate the first one and discard the first record if DTEND <DSTART after truncation.
* Payer plans may break out differently than observation periods, the amount of observation time and payer plan period time may not match for each person.
* Remove duplicate records before assigning PAYER\_PLAN\_PERIOD\_ID.

| **Table 4: PAYER\_PLAN\_PERIOD** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| PAYER\_PLAN\_PERIOD\_ID | - | System generated. |  |
| PERSON\_ID | **ENROLLMENT\_DETAIL:**  ENROLID |  |  |
| PAYER\_PLAN\_PERIOD\_START\_DATE | **ENROLLMENT\_DETAIL:**  DTSTART | Minimum start date of a contiguous enrollment in a plan. |  |
| PAYER\_PLAN\_PERIOD\_END\_DATE | **ENROLLMENT\_DETAIL:**  DTEND | Maximum end date of a contiguous enrollment in a plan. |  |
| PAYER\_SOURCE\_VALUE | **ENROLLMENT\_DETAIL:**  MEDICARE, CAP and PLANTYP | Use PLANTYP string:  1 - Basic/major medical  2 - comprehensive  3 - EPO  4 - HMO  5 - POS  6 - PPO  7 -POS with capitation  8 - CDHP  9 -HDHP  If MEDICARE =’1’ then replace [1] with “D”; else replace [1] with “N”;  If CAP =’1’ then replace [2] with “C”; else replace [2] with “N”;  Replace [3] with “Medicaid”. | NULL: Unknown  1: [1] [2] [3] Basic/Major Medical  2: [1] [2] [3] Comprehensive  3: [1] [2] [3] EPO  4: [1] [2] [3] HMO  5: [1] [2] [3] POS  6: [1] [2] [3] PPO  7: [1] [2] [3] POS with Capitation  8: [1] [2] [3] CDHP  9: [1] [2] [3] HDHP |
| PLAN\_SOURCE\_VALUE | NULL |  |  |
| FAMILY\_SOURCE\_VALUE | - | NULL |  |

## Table Name: PROVIDER

The PROVIDER table contains a list of uniquely identified health care providers (physicians). MDCD does have some provider information however some of the providers listed by MDCD may also be considered care sites or organizations. Since there is no clear way to decipher between all items identified as providers by MDCD, regardless if they are truly organizations or care sites, they will be added to this table.

Key conventions:

* To build this table it is essentially a distinct listing of the entire provider IDs with all their associated specialties.

SELECT DISTINCT PROV\_ID, STDPROV

FROM FACILITY\_HEADER

UNION

SELECT DISTINCT PROV\_ID, STDPROV

FROM OUTPATIENT\_SERVICES

UNION

SELECT DISTINCT PROV\_ID, STDPROV

FROM INPATIENT\_SERVICES

UNION

SELECT DISTINCT PROV\_ID, STDPROV

FROM LONG\_TERM\_CARE

* Provider Specialty (STDPROV) is available in MDCD. We add mapping of MDCD provider specialty to OMOP concept - VOCABULARY\_ID = JNJ\_TRU\_P\_SPCLTY.
* Set SPECIALTY\_CONCEPT\_ID as 38004514 (Unknown Physician Specialty) if STDPROV is missing or can't be mapped.

| **Table 5: PROVIDER** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| PROVIDER\_ID | - | System generated. |  |
| PROVIDER\_NAME |  | NULL |  |
| NPI | - | NULL |  |
| DEA | - | NULL |  |
| SPECIALTY\_CONCEPT\_ID | **FACILITY\_HEADER:**  STDPROV  **OUTPATIENT\_SERVICES:**  STDPROV  **INPATIENT\_SERVICES:**  STDPROV  **LONG\_TERM\_CARE:**  STDPROV | Use the code in Section 3.1.2.  Filters:  WHERE SOURCE\_VOCABULARY\_ID IN ('JNJ\_TRU\_P\_SPCLTY')  AND TARGET\_STANDARD\_CONCEPT IS NOT NULL  AND TARGET\_INVALID\_REASON IS NULL | Set SPECIALTY\_CONCEPT\_ID as 38004514 (Unknown Physician Specialty) if STDPROV is missing or cannot be mapped |
| CARE\_SITE\_ID | 0 |  |  |
| YEAR\_OF\_BIRTH |  | NULL |  |
| GENDER\_CONCEPT\_ID |  | 0 |  |
| PROVIDER\_SOURCE\_VALUE | **FACILITY\_HEADER:**  PROV\_ID  **OUTPATIENT\_SERVICES:**  PROV\_ID  **INPATIENT\_SERVICES:**  PROV\_ID  **LONG\_TERM\_CARE:**  PROV\_ID |  |  |
| SPECIALTY\_SOURCE\_VALUE | **FACILITY\_HEADER:**  STDPROV  **OUTPATIENT\_SERVICES:**  STDPROV  **INPATIENT\_SERVICES:**  STDPROV  **LONG\_TERM\_CARE:**  STDPROV |  |  |
| SPECIALTY\_SOURCE\_CONCEPT\_ID |  | 0 |  |
| GENDER\_SOURCE\_VALUE |  | NULL |  |
| GENDER\_SOURCE\_CONCEPT\_ID |  | 0 |  |

## Table Name: VISIT\_OCCURRENCE

This table contains all person visits to health care providers, including inpatient, outpatient, and ER visits. A visit is an encounter for a patient at a point of care for duration of time. There could be several providers involved in the patient’s care during the visit. Only pull distinct values into this table.

Even though MDCD has defined inpatient visits, some inpatient visits still exist in OUTPATIENT\_SERVICES table. We kept those inpatient visits defined by MDCD, and also applied the logic derived to define inpatient visits versus emergency room visits from the following reference:   
*Scerbo, M., C. Dickstein, and A. Wilson, Health Care Data and SAS. 2001, Cary, NC: SAS Institute Inc.*

MDCD also has LONG\_TERM\_CARE table which contains monthly or half-monthly billing information for long term care visits.

Key conventions:

* Inpatient, ER and outpatient visits are only generated off the MDCD OUTPATIENT\_SERVICES and INPATIENT\_SERVICES tables, FACILITY\_HEADER and INPATIENT\_ADMISSIONS tables are not used.
* Long term care visits are generated off MDCD LONG\_TERM\_CARE table.
* **Extract records within observation periods where a person has both prescription benefits and medical benefits from MDCD INPATIENT\_SERVICES, OUTPATIENT\_SERVICES and LONG\_TERM\_CARE table. This data is stored in a table called TEMP\_MEDICAL that will be used to populate all future tables.**
  + Set Source Flag:
    - To retain the knowledge for later, set this flag to ‘I’ for all INPATIENT\_SERVICES records, ‘O’ to all OUTPATIENT\_SERVICES records and ‘L’ to all LONG\_TERM\_CARE records.
  + Add a new ‘DOMAIN’ field to TEMP\_MEDICAL by mapping all diagnoses (PDX, DX1-DX5) and procedures (PROC1) to the standard vocabulary and taking the resulting TARGET\_DOMAIN\_ID. There should be one domain for each diagnosis and procedure on each individual TEMP\_MEDICAL record.
    - For diagnosis codes, use the query in section 3.1.2 with the filters:

WHERE SOURCE\_VOCABULARY\_ID IN ('ICD9CM')

AND TARGET\_STANDARD\_CONCEPT IS NOT NULL

AND TARGET\_INVALID\_REASON IS NULL

* + - For procedure codes, use the query in section 3.1.2 with the filters:

WHERE SOURCE\_VOCABULARY\_ID IN ('ICD9Proc','HCPCS','CPT4')

AND TARGET\_STANDARD\_CONCEPT IS NOT NULL

AND TARGET\_INVALID\_REASON IS NULL

AND TARGET\_CONCEPT\_CLASS\_ID NOT IN ('HCPCS Modifier','CPT4 Modifier')

* Some codes will map to more than one standard concept so they will have more than one DOMAIN\_ID. For instance, the ICD9 code ‘250.02’ maps to two standard concepts; One with a DOMAIN\_ID of ‘Condition’ and one with a DOMAIN\_ID of ‘Measurement’
  + Start Date:
    - SVCDATE will be the user’s start date
    - If service falls during an OBSERVATION\_PERIOD but starts before the period, trim the start date by using the beginning of the OBSERVATION\_PERIOD
  + End Date:
    - Use TSVCDAT as the user’s end date, however if the record is NULL or blank use the users SVCDATE or if the SVCDATE > TSCVDAT use the SVCDATE
    - If service falls during an OBSERVATION\_PERIOD but ends after the period, trim the end date by using the end of the OBSERVATION\_PERIOD
  + Claim Type:
    - Data that comes from INPATIENT\_SERVICES:
      * Set all records to ‘IP’ unless they should be reclassified as ‘ER’.
      * Pull out ‘ER’ visits that exist within the inpatient records.

df.STDPLAC = 23

OR (df.REVCODE >= '0450' AND df.REVCODE <= '0459')

OR df.REVCODE = '0981'OR df.PROC1 IN ('99281','99282','99283','99284','99285')

* + - * We trust the inpatients records somewhat and do not look for outpatient records in this table.
    - Data that comes from the OUTPATIENT\_SERVICES:
      * Set all records to ‘OP’ unless they should be reclassified as ‘ER’ or ‘IP’
      * Pull out ‘ER’ visits that exist within the outpatient records. Use the logic from above.
      * Pull our inpatient records that exist within the outpatient records  
        /\*Room and Board Charges\*/  
        (REVCODE >= '0100' AND REVCODE <='0219')

/\*Labor Room and Delivery\*/  
OR (REVCODE >= '0720' AND REVCODE <='0729')  
/\*Inpatient Renal Dialysis\*/

OR (REVCODE >= '0800' AND REVCODE <='0809')

* + - Data that comes from the LONG\_TERM\_CARE: Set all records to ‘LTCP’
* Generate Visits:
  + For claim type = ‘LTCP’:
    - Sort data in ascending order by ENROLID, START\_DATE, END\_DATE, PROV\_ID, and STDPROV.
    - Then by ENROLID, collapse lines of claim as long as the time between the END\_DATE of one line and the START\_DATE of the next is <=1. Each consolidated long term care claim is considered as one long term care visit, set
      * MIN(START\_DATE) as VISIT\_START\_DATE
      * MAX(END\_DATE) as VISIT\_END\_DATE
      * ‘LTCP’ as PLACE\_OF\_SERVICE\_SOURCE\_VALUE
    - Following the step above, when VISIT\_START\_DATE is equal to VISIT\_END\_DATE, set the first date of month (VISIT\_START\_DATE) as VISIT\_START\_DATE and last date of month (VISIT\_START\_DATE) as VISIT\_END\_DATE. Sort data in ascending order by ENROLID, VISIT\_START\_DATE, VISIT\_END\_DATE, PROV\_ID, and STDPROV. For each ENROLID, collapse lines of claim as long as the time between the VISIT\_START\_DATE of one line and the VISIT\_END\_DATE of the next is <=1. Then each consolidated long term care claim is considered as one long term care visit:
      * MIN(VISIT\_START\_DATE) as VISIT\_START\_DATE
      * MAX(VISIT\_END\_DATE) as VISIT\_END\_DATE
      * ‘LTCP’ as PLACE\_OF\_SERVICE\_SOURCE\_VALUE
    - As you are collapsing records take the PROVID and STDPROV from the first claim line of each visit as VISIT\_PROVID and VISIT\_PROVSTD, this will be used later to assign providers associated to a visit.
  + For claim type = ‘IP’:
    - Sort data in ascending order by ENROLID, START\_DATE, END\_DATE, PROVID, and STDPROV.
    - Then by ENROLID, collapse lines of claim as long as the time between the END\_DATE of one line and the START\_DATE of the next is <=1.
    - Then each consolidated inpatient claim is considered as one inpatient visit, set
      * MIN(START\_DATE) as VISIT\_START\_DATE
      * MAX(END\_DATE) as VISIT\_END\_DATE
      * ‘IP’ as PLACE\_OF\_SERVICE\_SOURCE\_VALUE
    - As you are collapsing records take the PROVID and STDPROV from the first claim line of each visit as VISIT\_PROVID and VISIT\_PROVSTD, this will be used later to assign providers associated to a visit.
    - See if any ‘OP’ or ‘ER’ records occur during an ‘IP’ visit. These should be consolidated into that ‘IP’ visit, unless it is an ‘ER’ visit that starts and ends on the first day of the ‘IP’ visit.
  + For claim type = ‘ER’
    - Sort data in ascending order by ENROLID, START\_DATE, END\_DATE, PROVID, and STDPROV.
    - Then by ENROLID, collapse all ‘ER’ claims that start on the same day as one ER visit, then take START\_DATE as VISIT\_START\_DATE, MAX (END\_DATE) as VISIT\_END\_DATE, and ‘ER’ as PLACE\_OF\_SERVICE\_SOURCE\_VALUE.
    - As you are collapsing records take the PROVID and STDPROV from the first claim line of each visit as VISIT\_PROVID and VISIT\_PROVSTD, this will be used later to assign providers associated to a visit.
  + For claim type = ‘OP’
    - Sort data in ascending order by ENROLID, START\_DATE, PROVID, END\_DATE, and STDPROV.
    - Then by ENROLID and START\_DATE, collapse all ‘OP’ claims that have the same PROVID as one OP visit, then take START\_DATE as VISIT\_START\_DATE, MAX (END\_DATE) as VISIT\_END\_DATE, and ‘OP’ as PLACE\_OF\_SERVICE\_SOURCE\_VALUE.
    - As you are collapsing records take the PROVID and STDPROV from the first claim line of each visit as VISIT\_PROVID and VISIT\_PROVSTD, this will be used later to assign providers associated to a visit.
  + Three examples for defining VISIT\_OCCURRENCE can be found here:  
    
* Pull all additional information to move forward with building other tables:
  + Pull INPATIENT\_ADMISSIONS information needed to build future tables into TEMP\_INPATIENT\_ADMISSIONS. We take any INPATIENT\_ADMISSION that matches up to an ‘IP’ TEMP\_MEDICAL record on ENROLID, CASEID, and YEAR.
    - As with the TEMP\_MEDICAL table, create a new field called DOMAIN and map the diagnosis codes (PDX, DX1-DX15) and procedure codes (PPROC, PROC1-PROC15) on each record to the standard vocabulary and take the resulting DOMAIN\_ID.
      * For the diagnosis codes use the query in section 3.1.2 with the filters:

WHERE SOURCE\_VOCABULARY\_ID IN ('ICD9CM')

AND TARGET\_STANDARD\_CONCEPT IS NOT NULL

AND TARGET\_INVALID\_REASON IS NULL

* + - * For the procedure codes use the query in section 3.1.2 with the filters:

WHERE SOURCE\_VOCABULARY\_ID IN ('ICD9Proc','HCPCS','CPT4')

AND TARGET\_STANDARD\_CONCEPT IS NOT NULL

AND TARGET\_INVALID\_REASON IS NULL

AND TARGET\_CONCEPT\_CLASS\_ID NOT IN ('HCPCS Modifier','CPT4 Modifier')

* + Pull FACILITY\_HEADER information needed to build future tables into TEMP\_FACILITY\_HEADER. We take any FACILITY\_HEADER record that matches up to a TEMP\_MEDICAL record on ENROLID, FACHDID, and YEAR. However we only take the first row for each ENROLID, FACHDID, and YEAR giving the visit with highest priority preference if it can be linked to multiple visits: IP visit > ER visit > OP visit > LTCP visit.
    - As with the TEMP\_MEDICAL table, create a new field called DOMAIN and map the diagnosis codes (PDX, DX1-DX9) and procedure codes (PROC1-PROC6) on each record to the standard vocabulary and take the resulting DOMAIN\_ID.
      * For the diagnosis codes use the query in section 3.1.2 with the filters:

WHERE SOURCE\_VOCABULARY\_ID IN ('ICD9CM')

AND TARGET\_STANDARD\_CONCEPT IS NOT NULL

AND TARGET\_INVALID\_REASON IS NULL

* + - * For the procedure codes use the query in section 3.1.2 with the filters:

WHERE SOURCE\_VOCABULARY\_ID IN ('ICD9Proc','HCPCS','CPT4')

AND TARGET\_STANDARD\_CONCEPT IS NOT NULL

AND TARGET\_INVALID\_REASON IS NULL

AND TARGET\_CONCEPT\_CLASS\_ID NOT IN ('HCPCS Modifier','CPT4 Modifier')

* + The TEMP\_MEDICAL, TEMP\_INPATIENT\_ADMISSIONS and TEMP\_FACILITY\_HEADER all contain visit information and will be used to populate CONDITION\_OCCURRENCE, PROCEDURE\_OCCURRENCE and DEATH tables.

| **Table 6: VISIT\_OCCURRENCE** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| VISIT\_OCCURRENCE\_ID | - | System generated. |  |
| PERSON\_ID | **OUTPATIENT\_SERVICES:**ENROLID  **INPATIENT\_SERVICES:** ENROLID  **LONG\_TERM\_CARE:**  ENROLID |  |  |
| VISIT\_START\_DATE | **OUTPATIENT\_SERVICES:**  Min(SVCDATE)  **INPATIENT\_SERVICES:**  Min(SVCDATE)  **LONG\_TERM\_CARE:**  Min(SVCDATE) |  |  |
| VISIT\_START\_DATETIME | **-** | NULL |  |
| VISIT\_END\_DATE | **OUTPATIENT\_SERVICES:**  Max(SVCDATE)  **INPATIENT\_SERVICES:**  Max(TSVCDAT)  **LONG\_TERM\_CARE:**  Max(TSVCDAT) |  |  |
| VISIT\_END\_DATETIME | **-** | NULL |  |
| VISIT\_TYPE\_CONCEPT\_ID | **-** | All rows will have the CONCEPT\_ID 44818517 | CONCEPT\_ID 44818517 = ‘Visit derived from encounter on claim’ |
| VISIT\_CONCEPT\_ID | *VISIT\_SOURCE\_VALUE* | Map VISIT\_SOURCE\_VALUE to their associated CONCEPT\_IDs:  IP --- 9201  OP --- 9202  ER --- 9203  LTCP --- 42898160 |  |
| PROVIDER\_ID | - | Refer to logic above for assigning VISIT\_PROVID and VISIT\_PROVSTD, and map them to PROVIDER\_SOURCE\_VALUE and  SPECIALTY\_SOURCE\_VALUE in Provider table to extract Provider ID. | If there is no associated PROVIDER\_ID then set as NULL |
| CARE\_SITE\_ID | - | NULL |  |
| VISIT\_SOURCE\_VALUE | - | Use the logic mentioned above to define visit types, and value can be 'IP','ER','OP', 'LTCP'. |  |
| VISIT\_SOURCE\_CONCEPT\_ID | - | 0 |  |
| ADMITTING\_SOURCE\_CONCEPT\_ID | - |  |  |
| ADMITTING\_SOURCE\_VALUE | - |  |  |
| DISCHARGE\_TO\_CONCEPT\_ID | **INPATIENT\_SERVICES:**  DSTATUS | Use the code in Section 3.1.2.  Filters:  WHERE SOURCE\_VOCABULARY\_ID IN ('JNJ\_TRU\_DPOS')  AND TARGET\_STANDARD\_CONCEPT IS NOT NULL  AND TARGET\_INVALID\_REASON IS NULL |  |
| DISCHARGE\_TO\_SOURCE\_VALUE | **INPATIENT\_SERVICES:**  DSTATUS |  |  |
| PRECEDING\_VISIT\_OCCURRENCE\_ID | VISIT\_OCCURRENCE\_ID | For a given person, find the visit prior to this one and reference it here | A foreign key to the VISIT\_OCCURRENCE table of the visit immediately preceding this visit |

## Table Name: CONDITION\_OCCURRENCE

Conditions can be pulled from FACILITY\_HEADER, OUTPATIENT\_SERVICES, INPATIENT\_SERVICES, INPATIENT\_ADMISSIONS, or LONG\_TERM\_CARE tables. This table records individual instances of the conditions suffered by a person as extracted from the source data.

Key conventions:

* Use the TEMP\_MEDICAL, TEMP\_INPATIENT\_ADMISSIONS and TEMP\_FACILITY\_HEADER tables which are created during building the VISIT\_OCCURRENCE table to populate condition information. For each diagnosis code, always assign its associated VISIT\_START\_DATE as CONDITION\_START\_DATE, and use VISIT\_PROVID and VISIT\_PROVSTD to extract its ASSOCIATED\_PROVIDER\_ID from PROVIDER table.
* PDX appears in both MDCD INPATIENT\_SERVICES and INPATIENT\_ADMISSIONS tables. To avoid duplicates, extract PDX only from the TEMP\_INPATIENT\_ADMISSIONS table.
* For each VISIT\_OCCURRENCE\_ID, if multiple records for a same diagnosis code exist use the following logic to define priority and only keep the one with highest priority (1>2>3) and highest position (1>2>3…>15):

1. TEMP\_MEDICAL: priority =1; TEMP\_INPATIENT\_ADMISSIONS: priority =2; TEMP\_FACILITY\_HEADER: priority =3.
2. Assign CONDITION\_TYPE\_CONCEPT\_ID use the following logic:

| **Claim Type** | **Source** | **Position** | **CONDITION\_TYPE \_CONCEPT\_ID** | **CONCEPT\_NAME** |
| --- | --- | --- | --- | --- |
| IP | INPATIENT\_SERVICES and OUTPATIENT\_SERVICES TABLES | 1 (PDX) | 38000183 | Inpatient detail - primary |
| 2 (DX1) | 38000184 | Inpatient detail – 1st position |
| 3 (DX2) | 38000185 | Inpatient detail - 2rd position |
| 4 (DX3) | 38000186 | Inpatient detail - 3th position |
| 5 (DX4) | 38000187 | Inpatient detail - 4th position |
| 6 (DX5) | 38000188 | Inpatient detail - 5th position |
| INPATIENT\_ADMISSIONS (PDX excluded) and FACILITY\_HEADER TABLES | 1 (PDX) | 38000199 | Inpatient header - primary |
| 2 (DX1) | 38000200 | Inpatient header - 1st position |
| 3 (DX2) | 38000201 | Inpatient header - 2nd position |
| 4 (DX3) | 38000202 | Inpatient header - 3rd position |
| 5 (DX4) | 38000203 | Inpatient header - 4th position |
| 6 (DX5) | 38000204 | Inpatient header - 5th position |
| 7 (DX6) | 38000205 | Inpatient header - 6th position |
| 8 (DX7) | 38000206 | Inpatient header - 7th position |
| 9 (DX8) | 38000207 | Inpatient header - 8th position |
| 10 (DX9) | 38000208 | Inpatient header - 9th position |
| 11 (DX10) | 38000209 | Inpatient header - 10th position |
| 12 (DX11) | 38000210 | Inpatient header - 11th position |
| 13 (DX12) | 38000211 | Inpatient header - 12th position |
| 14 (DX13) | 38000212 | Inpatient header - 13th position |
| 15 (DX14) | 38000213 | Inpatient header - 14th position |
| ER or OP OR LTC | INPATIENT\_SERVICES, OUTPATIENT\_SERVICES TABLES, AND LONG\_TERM\_CARE TABLES | 1 (PDX) | 38000215 | Outpatient detail - 1st position |
| 2 (DX1) | 38000215 | Outpatient detail - 1st position |
| 3 (DX2) | 38000216 | Outpatient detail - 2nd position |
| 4 (DX3) | 38000217 | Outpatient detail - 3rd position |
| 5 (DX4) | 38000218 | Outpatient detail - 4th position |
| 6 (DX5) | 38000219 | Outpatient detail - 5th position |
| FACILITY\_HEADER TABLES | 2 (DX1) | 38000230 | Outpatient header - 1st position |
| 3 (DX2) | 38000231 | Outpatient header - 2nd position |
| 4 (DX3) | 38000232 | Outpatient header - 3rd position |
| 5 (DX4) | 38000233 | Outpatient header - 4th position |
| 6 (DX5) | 38000234 | Outpatient header - 5th position |
| 7 (DX6) | 38000235 | Outpatient header - 6th position |
| 8 (DX7) | 38000236 | Outpatient header - 7th position |
| 9 (DX8) | 38000237 | Outpatient header - 8th position |
| 10 (DX9) | 38000238 | Outpatient header - 9th position |

* Only keep records with valid ICD9 or ICD10 diagnoses, the DXVER variable in all tables indicates the ICD version to map to. If DXVER=9 then use ICD9CM; if DXVER=0 then use ICD10CM Else if before October 1, 2015 use ICD9 and if after October 1, 2015 use ICD10

1. ICD9 must start with 0-9, V or E, and non-numeric character is not allowed in other positions.
2. If ICD9 starts with 0-9 or V, length should be between 3 and 5; if starts with E, length should be between 4 and 5.
3. ICD10 must be between 3 and 7 digits

* Truven removes decimal points from ICD diagnosis so when mapping to the OMOP Vocabulary, the decimal points need to also be removed from the vocabulary in order to map between the source and the vocabulary.
* Remove duplicate records before assigning CONDITION\_OCCURRENCE\_ID.

| **Table 7: CONDITION\_OCCURRENCE** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| CONDITION\_ OCCURRENCE\_ID | - | System generated. |  |
| PERSON\_ID | **FACILITY\_HEADER:**  ENROLID  **OUTPATIENT\_SERVICES:**  ENROLID  **INPATIENT\_SERVICES:**  ENROLID  **INPATIENT\_ADMISSION:**  ENROLID  **LONG\_TERM\_CARE:**  ENROLID |  |  |
| CONDITION\_ CONCEPT\_ID | **FACILITY\_HEADER:**  DX1, DX2, DX3, DX4, DX5, DX6, DX7, DX8, DX9  **OUTPATIENT\_SERVICES:** DX1-DX4  **INPATIENT\_SERVICES:**  PDX, DX1-DX4  **LONG\_TERM\_CARE:**  DX1-DX4  **INPATIENT\_ADMISSION:** PDX, DX1, DX2, DX3, DX4, DX5, DX6, DX7, DX8, DX9, DX10, DX11, DX12, DX13, DX14, DX15 | Use the code in Section 3.1.2.  If DXVER=9 use the filter  WHERE SOURCE\_VOCABULARY\_ID IN ('ICD9CM')  AND TARGET\_STANDARD\_CONCEPT IS NOT NULL  AND TARGET\_INVALID\_REASON IS NULL  If DXVER=0 use the filter  WHERE SOURCE\_VOCABULARY\_ID IN ('ICD10','ICD10CM')  AND TARGET\_STANDARD\_CONCEPT IS NOT NULL  AND TARGET\_INVALID\_REASON IS NULL  From **HRA** table: These should already have been mapped to CONCPET\_IDs | As data is being assigned a CONCEPT\_ID check the domain. Only if the DOMAIN\_ID = “Condition” or if the CONCEPT\_ID = 0 can it land in the CONDITION\_OCCURRENCE table.  Other domains will need to go to their respective tables. For PROCEDURE\_OCCURRENCE and DRUG\_EXPOSURE treat the DX as an incoming procedure or drug with the lowest priority. MEASUREMENT and OBSERVATION data is described in more detail in their respective sections. |
| CONDITION\_START\_DATE | **OUTPATIENT\_SERVICES or INPATIENT\_SERVICES or LONG\_TERM\_CARE:**  SVCDATE  **ELSE:**  **VISIT\_OCCURRENCE** **–**VISIT\_START\_DATE |  |  |
| CONDITION\_START\_DATETIME | **-** | NULL |  |
| CONDITION\_END\_ DATE | - | NULL |  |
| CONDITION\_END\_DATETIME | - | NULL |  |
| CONDITION\_TYPE\_CONCEPT\_ID | - | Refer to logic above to assign CONDITION\_TYPE\_CONCEPT\_ID  If the record is coming from another table, like the OBSERVATION, keep the types that would have been assigned in that table. |  |
| STOP\_REASON | - | NULL |  |
| PROVIDER\_ID | - | Refer to logic in building VISIT\_OCCURRENCE table for assigning VISIT\_PROVID and VISIT\_PROVSTD, and map them to PROVIDER\_SOURCE\_VALUE and  SPECIALTY\_SOURCE\_VALUE in Provider table to extract associated Provider ID. | If there is no associated PROVIDER\_ID then set as NULL |
| VISIT\_OCCURRENCE\_ID | **VISIT\_OCCURRENCE:**  VISIT\_OCCURRENCE\_ID | Refer to logic in building VISIT\_OCCURRENCE table for linking with VISIT\_OCCURRENCE\_ID |  |
| CONDITION\_ SOURCE\_VALUE | **FACILITY\_HEADER:**  DX1, DX2, DX3, DX4, DX5, DX6, DX7, DX8, DX9  **OUTPATIENT\_SERVICES:**DX1-DX4  **INPATIENT\_SERVICES:**  PDX, DX1-DX4  **LONG\_TERM\_CARE:**  DX1-DX4  **INPATIENT\_ADMISSION:**  DX1, DX2, DX3, DX4, DX5, DX6, DX7, DX8, DX9, DX10, DX11, DX12, DX13, DX14, DX15 |  |  |
| CONDITION\_ SOURCE\_CONCEPT\_ ID | **-** | Use the code in Section 3.1.1.  Filters:  WHERE SOURCE\_VOCABULARY\_ID IN ('ICD9CM')  AND TARGET\_VOCABULARY\_ID IN ('ICD9CM') |  |
| CONDITION\_STATUS\_SOURCE\_VALUE | **-** | NULL |  |
| CONDITION\_STATUS\_CONCEPT\_ID | **0** | NULL |  |

## Table Name: CONDITION\_ERA

Similar to DRUG\_ERAs, CONDITION\_ERAs are chronological periods of condition occurrence. Combining individual CONDITION\_OCCURRENCEs into a single CONDITION\_ERA serves two purposes:

* + It allows aggregation of chronic conditions that require frequent ongoing care, instead of treating each Condition Occurrence as an independent event.
  + It allows aggregation of multiple, closely timed doctor visits for the same condition to avoid double-counting the Condition Occurrences.

Key conventions:

* The CONDITION\_ERAs are not aggregated to a higher-level class, therefore the CONDITION\_CONCEPT\_ID is unchanged.
* Use the following steps to build this table off CONDITION\_OCCURRENCE table:

1. Exclude records with CONDITION\_CONCEPT\_ID = 0.
2. Set CONDITION\_END\_DATE = CONDITION\_START\_DATE.
3. Sort CONDITION\_OCCURRENCE table by PERSON\_ID, CONDITION\_CONCEPT\_ID and CONDITION\_START\_DATE in ascending order.
4. Combine records as long as both PERSON\_ID and CONDITION\_CONCEPT\_ID don’t change and the time between CONDITION\_END\_DATE of one record and CONDITION\_START\_DATE of the next is 30 days or less (<=30).

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 8: CONDITION\_ERA** | | | |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| CONDITION\_ERA\_ID | - | System generated. |  |
| PERSON\_ID | PERSON\_ID |  |  |
| CONDITION\_CONCEPT\_ID | CONDITION\_CONCEPT\_ID | Do not build CONDITION\_ERAS where the CONDITION\_OCCURRENCE.CONDITION\_CONCEPT\_ID is 0. |  |
| CONDITION\_ERA\_START\_DATE | CONDITION\_START\_DATE |  | The start date for the condition era constructed from the individual instances of condition occurrences. It is the start date of the very first chronologically recorded instance of the condition. |
| CONDITION\_ERA\_END\_DATE | CONDITION\_END\_DATE |  | The end date for the condition era constructed from the individual instances of condition occurrences. It is the end date of the final continuously recorded instance of the condition. |
| CONDITION\_OCCURRENCE\_COUNT | - | Sum up the number of CONDITION\_OCCURRENCEs for this PERSON\_ID and this CONCEPT\_ID during the exposure window being built. |  |

## Table Name: PROCEDURE\_OCCURRENCE

Procedure data can be found in the FACILITY\_HEADER, OUTPATIENT\_SERVICES, INPATIENT\_SERVICES, and INPATIENT\_ADMISSIONS tables. These tables’ procedure columns have CPT-4s, HCPCs, and ICD9s procedure codes.

Key conventions:

* Use the TEMP\_MEDICAL, TEMP\_INPATIENT\_ADMISSIONS and TEMP\_FACILITY\_HEADER tables which are created during building the VISIT\_OCCURRENCE table to populate procedure information. The DOMAIN\_ID mappings created in the VISIT\_OCCURRENCE step will be used to help populate this table. Any code with a domain of ‘Procedure’ should be added to this table, regardless if it is a diagnosis (DX) field or procedure (PROC) field. PPROC appears in both Truven INPATIENT\_SERVICES and INPATIENT\_ADMISSIONS tables. To avoid duplicates, extract PPROC only from the TEMP\_INPATIENT\_ADMISSIONS table.

1. For procedure (PROC1) from **TEMP\_MEDICAL** table: Each record with a value in PROC1 in this table will be kept as one line in the COST table, and assign SVCDATE as PROCEDURE\_DATE.
   * For a **TEMP\_MEDICAL** record with no value in PROC1: If the record contains any cost information (value other than 0 in COB, COINS, COPAY, DEDUCT, NETPAY, PAY) but if PROC1 is blank then a record should be created in PROCEDURE\_OCCURRENCE with PROCEDURE\_CONCEPT\_ID = 0, else no record should be created.
2. For each row in TEMP\_MEDICAL table, assign a unique COST\_ID which will be used for creating COST table. Data will be sorted by VISIT\_OCCURRENCE\_ID, PROC1, PROCEDURE\_DATE, PROVID and STDPROV in ascending order, a unique procedure occurrence is defined by the combination of VISIT\_OCCURRENCE\_ID, PROC1 and PROCEDURE\_DATE, and always choose PROVID and STDPROV from the first line of each procedure occurrence as PPROV, PPROVCAT (associated provider id and its specialty for the procedure occurrence, will be used to populate ASSOCIATED\_PROVIDER\_ID). Assign Priority =1 and Position =1.
   * Only a procedure (PROC1) in **TEMP\_MEDICAL** will have a modifier in the **PROCMOD** and the modifier will only ever apply to a CPT4 code in PROC1. Refer to the table below for how to map these codes to standard concepts.
3. For procedure (PPROC, PROC1-PROC15) from **TEMP\_INPATIENT\_ADMISSIONS** table: Assign position =1 for PPROC and 2 for PROC1, etc., set VISIT\_END\_DATE as PROCEDURE\_DATE, VISIT\_PROVID and VISIT\_PROVSTD as PPROV and PROVCAT. Assign Priority =2.
4. For procedure (PROC1-PROC6) from **TEMP\_FACILITY\_HEADER** table: Assign position =2 for PROC1 and 3 for PROC2, etc., VISIT\_END\_DATE as PROCEDURE\_DATE, and VISIT\_PROVID and VISIT\_PROVSTD as PPROV and PROVCAT. Assign Priority =3.
5. For any procedure coming from a DX field in **TEMP\_MEDICAL** (PDX, DX1-DX5) assign position=2 for PDX, 3 for DX1, etc. set VISIT\_END\_DATE as PROCEDURE\_DATE, VISIT\_PROVID and VISIT\_PROVSTD as PPROV and PROVCAT. Assign priority = 4.
6. For any procedure coming from a DX field in **TEMP\_INPATIENT\_ADMISSIONS** (PDX, DX1-DX15) assign position=17 for PDX, 18 for DX1, etc. set VISIT\_END\_DATE as PROCEDURE\_DATE, VISIT\_PROVID and VISIT\_PROVSTD as PPROV and PROVCAT. Assign priority = 5.
7. For any procedure coming from a DX field in **TEMP\_FACILITY\_HEADER** (DX1-DX9) assign position=8 for DX1, 9 for DX2, etc. set VISIT\_END\_DATE as PROCEDURE\_DATE, VISIT\_PROVID and VISIT\_PROVSTD as PPROV and PROVCAT. Assign priority = 6.

* Merge procedure information extracted above as one table**. For one visit (identified by VISIT\_OCCURRENCE\_ID), if multiple records for a same procedure code exist on the same date**, use the following logic to define priority and only keep the one with highest priority (1>2>3) and highest position (1>2>3…). Then assign PROCEDURE\_TYPE\_CONCEPT\_ID use the following logic:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Type of Associated Visit** | **Source** | **Position** | **CONDITION\_ TYPE\_CONCEPT\_ID** | **CONCEPT\_NAME** |
| IP | INPATIENT\_SERVICES and OUTPATIENT\_SERVICES TABLES | 1 (PROC1) | 38000249 | Inpatient detail - 1st position |
| INPATIENT\_ADMISSIONS and FACILITY\_HEADER TABLES | 1 (PPROC) | 38000250 | Inpatient header - primary position |
| 2 (PROC1) | 38000251 | Inpatient header - 1st position |
| 3 (PROC2) | 38000252 | Inpatient header - 2nd position |
| 4 (PROC3) | 38000253 | Inpatient header - 3rd position |
| 5 (PROC4) | 38000254 | Inpatient header - 4th position |
| 6 (PROC5) | 38000255 | Inpatient header - 5th position |
| 7 (PROC6) | 38000256 | Inpatient header - 6th position |
| 8 (PROC7) | 38000257 | Inpatient header - 7th position |
| 9 (PROC8) | 38000258 | Inpatient header - 8th position |
| 10 (PROC9) | 38000259 | Inpatient header - 9th position |
| 11 (PROC10) | 38000260 | Inpatient header - 10th position |
| 12 (PROC11) | 38000261 | Inpatient header - 11th position |
| 13(PROC12) | 38000262 | Inpatient header - 12th position |
| 14 (PROC13) | 38000263 | Inpatient header - 13th position |
| 15(PROC14) | 38000264 | Inpatient header - 14th position |
| 16(PROC15) | 38000265 | Inpatient header - 15th position |
| ER or OP or LTCP | INPATIENT\_SERVICES, OUTPATIENT\_SERVICES TABLES and LONG\_TERM\_CARE | 1 (PROC1) | 38000267 | Outpatient detail - 1st position |
| FACILITY\_HEADER TABLES | 2 (PROC1) | 38000269 | Outpatient header - 1st position |
| 3 (PROC2) | 38000270 | Outpatient header - 2nd position |
| 4 (PROC3) | 38000271 | Outpatient header - 3rd position |
| 5 (PROC4) | 38000272 | Outpatient header - 4th position |
| 6 (PROC5) | 38000273 | Outpatient header - 5th position |
| 7 (PROC6) | 38000274 | Outpatient header - 6th position |

* Data from PPROC and PROC columns should create a record in this table (this is so we can associate PROCEDURE\_COSTS to the records). For example, if a PROC code is considered a CONDITION, a record will be written to the CONDITION\_OCCURRENCE table but a record will still be written to the PROCEDURE\_OCCURRENCE table just with a CONCEPT\_ID = 0.

| **Table 9: PROCEDURE\_OCCURRENCE** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| PROCEDURE\_OCCURRENCE\_ID | - | System generated. |  |
| PERSON\_ID | **FACILITY\_HEADER:**  ENROLID  **OUTPATIENT\_SERVICES:**  ENROLID  **INPATIENT\_SERVICES:**  ENROLID  **INPATIENT\_ADMISSIONS:**  ENROLID  **LONG\_TERM\_CARE:**  ENROLID |  |  |
| PROCEDURE\_CONCEPT\_ID | **FACILITY\_HEADER:**  PROC1, PROC2, PROC3, PROC4, PROC5, PROC6, DX1-DX9  **OUTPATIENT\_SERVICES:**  PROC1, PDX, DX1-DX4  **INPATIENT\_SERVICES:**  PROC1, PDX, DX1-DX4  **LONG\_TERM\_CARE:**  PROC1, DX1-DX4  **INPATIENT\_ADMISSIONS:**  PPROC, PROC1, PROC2, PROC3, PROC4, PROC5, PROC6, PROC7, PROC8, PROC9, PROC10, PROC11,  PROC12, PROC13, PROC14, PROC15, PDX, DX1-DX15 | Use Vocabulary map 3.1.2.  Filter when procedure code comes from PROC field:  WHERE SOURCE\_VOCABULARY\_ID IN ('ICD9Proc','HCPCS','CPT4',’ICD10PCS’)  AND TARGET\_STANDARD\_CONCEPT IS NOT NULL  AND TARGET\_INVALID\_REASON IS NULL  AND TARGET\_CONCEPT\_CLASS\_ID NOT IN ('HCPCS Modifier','CPT4 Modifier', ‘CPT4 Hierarchy’, ‘ICD10PCS Hierarchy’)  From **DX** field: These should have already been mapped to CONCEPT\_IDs | Procedures included in the DRUG\_EXPOSURE table are also included here.  Refer to [CONDITION\_OCCURRENCE](#_Table_Name:_) for description of how DOMAIN\_IDs are used to direct data to correct tables |
| PROCEDURE\_DATE | **For procedure codes from FACILITY\_HEADER / INPATIENT\_ADMISSIONS table:** use VISIT\_END\_DATE of its associated visit;  **For procedure code from PROC1 in OUTPATIENT\_SERVICES or INPATIENT\_SERVICES or LONG\_TERM\_CARE:**  SVCDATE. | Refer to key conventions for defining PROCEDURE\_DATE. |  |
| PROCEDURE\_DATETIME | **-** | NULL |  |
| PROCEDURE\_TYPE\_CONCEPT\_ID | - | Refer to key conventions for assigning PROCEDURE\_TYPE\_CONCEPT\_ID  If the record is coming from another table, like the CONDITION\_OCCURENCE, keep the types that would have been assigned in that table. | PROCEDURE\_OCCURRENCE types are found under VOCABULARY\_ID = ‘Procedure Type’ (OMOP Procedure Occurrence Type) |
| MODIFIER\_CONCEPT\_ID | **OUTPATIENT\_SERVICES:**  PROCMOD  **INPATIENT\_SERVICES:**  PROCMOD | Use Vocabulary filter 3.1.2  WHERE SOURCE\_CONCEPT\_CLASS\_ID IN ('CPT4 Modifier')  AND TARGET\_CONCEPT\_CLASS\_ID IN ('CPT4 Modifier') | If PROCMOD is blank then leave this field blank as well |
| QUANTITY | - | NULL | There is no quantity variable |
| PROVIDER\_ID | - | Refer to key conventions for defining PPROV and PPROVCAT, then map them to PROVIDER\_SOURCE\_VALUE and  SPECIALTY\_SOURCE\_VALUE in Provider table to extract associated Provider ID. | If there is no associated PROVIDER\_ID then set as NULL |
| VISIT\_OCCURRENCE\_ID | **VISIT\_OCCURRENCE**: VISIT\_OCCURRENCE\_ID | Refer to logic in building VISIT\_OCCURRENCE table for linking with VISIT\_OCCURRENCE\_ID |  |
| PROCEDURE\_SOURCE\_VALUE | **FACILITY\_HEADER:**  PROC1, PROC2, PROC3, PROC4, PROC5, PROC6, DX1-DX9  **OUTPATIENT\_SERVICES:**  PROC1, PDX, DX1-DX5  **INPATIENT\_SERVICES:** PROC1, PDX, DX1-DX5  **INPATIENT\_ADMISSIONS:**  PPROC, PROC1, PROC2, PROC3, PROC4, PROC5, PROC6, PROC7, PROC8, PROC9, PROC10, PROC11,  PROC12, PROC13, PROC14, PROC15, PDX, DX1-DX15 |  |  |
| PROCEDURE\_SOURCE\_CONCEPT\_ID | **-** | Use Vocab query 3.1.1.  Filter when procedure code comes from a PROC field:  WHERE SOURCE\_VOCABULARY\_ID IN ('ICD9Proc','HCPCS','CPT4',’ICD10PCS’)  AND TARGET\_VOCABULARY\_ID IN ('ICD9Proc','HCPCS','CPT4',’ICD10PCS’)  AND TARGET\_CONCEPT\_CLASS\_ID NOT IN ('HCPCS Modifier','CPT4 Modifier',’CPT4 Hierarchy’,’ICD10PCS Hierarchy’)  Incoming CONDITION\_OCCURRENCE records (PDX/DX#) should follow the SOURCE\_CONCEPT\_ID mapping from the CONDITION\_OCCURRENCE table. |  |
| MODIFIER\_SOURCE\_VALUE | **OUTPATIENT\_SERVICES:**  PROCMOD  **INPATIENT\_SERVICES:**  PROCMOD |  |  |

## Table Name: COST

The COST table captures all cost information for procedures, drugs, visits and devices. It replaces the PROCEDURE\_COST, DRUG\_COST, DEVICE\_COST and VISIT\_COST tables present in previous versions and was a new addition to CDM version 5.0.1

Key Conventions:

* Costs will be taken from TEMP\_INPATIENT\_ADMISSIONS, TEMP\_MEDICAL and DRUG\_CLAIMS and should be loaded after the VISIT\_OCCURRENCE table since the mapped domains will be needed in order to properly associate the costs with the correct fact table (CONDITION\_OCCURRENCE, PROCEDURE\_OCCURRENCE, DEVICE\_EXPOSURE, DRUG\_EXPOSURE, MEASUREMENT, OBSERVATION or VISIT\_OCCURRENCE).
* Each line of service from TEMP\_MEDICAL table will create one and only one record in the COST table since costs are associated with PROC1.
  + After mapping each PROC1 in TEMP\_MEDICAL to the correct CONCEPT\_ID and the correct DOMAIN during the VISIT\_OCCURRENCE step, use this information to populate COST\_EVENT\_ID and COST\_DOMAIN\_ID.
    - For example, if a code in PROC1 maps to a device then when reporting the cost, the COST\_EVENT\_ID will be the corresponding DEVICE\_EXPOSURE\_ID from the DEVICE\_EXPOSURE table and COST\_DOMAIN\_ID will be 41 which stands for ‘Device/Procedure’. See the below table for assigning COST\_DOMAIN\_ID.

| **Source** | **Position** | **Mapped Table** | **COST\_DOMAIN\_ID** | **Description** |
| --- | --- | --- | --- | --- |
| INPATIENT\_SERVICES and OUTPATIENT\_SERVICES TABLES | PROC1 | PROCEDURE\_OCCURRENCE | 10 | Procedure |
| PROC1 | CONDITION\_OCCURRENCE | 49 | Condition/Procedure |
| PROC1 | DEVICE\_EXPOSURE | 41 | Device/Procedure |
| PROC1 | DRUG\_EXPOSURE | 54 | Drug/Procedure |
| PROC1 | MEASUREMENT | 43 | Measurement/Procedure |
| PROC1 | OBSERVATION | 42 | Observation/Procedure |

* When taking cost information from TEMP\_INPATIENT\_ADMISSIONS we take any INPATIENT\_ADMISSION that matches up to an ‘IP’ TEMP\_MEDICAL record on ENROLID, CASEID, and YEAR.
  + All costs recorded from this table will have a COST\_DOMAIN\_ID = 8 (Visit) and the COST\_EVENT\_ID will be VISIT\_OCCURRENCE\_ID
* Drug cost information is pulled from the DRUG\_CLAIMS table at the same time the DRUG\_EXPOSURE is loaded and follows the same pull logic as described for DRUG\_EXPOSURE table.
  + Starts from the intermediate table created in step (3) of prescription drug for building DRUG\_EXPOSURE table.
  + Since the amount of observation time in OBSERVATION\_PERIOD may be greater than that in PAYER\_PLAN\_PERIOD table, use left join to avoid excluding records when pulling PAYER\_PLAN\_PERIOD\_ID. For those records fall out of PAYER\_PLAN\_PERIOD\_START\_DATE AND PAYER\_PLAN\_PERIOD\_END\_DATE, set PAYER\_PLAN\_PERIOD\_ID as NULL.

| **Table 10: COST** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| cost\_id | - | A system generated unique identifier for each cost record |  |
| cost\_event\_id | **CONDITION\_OCCURRENCE**: CONDITION\_OCCURRENCE\_ID  **PROCEDURE\_OCCURRENCE**: PROCEDURE\_OCCURRENCE\_ID  **DEVICE\_EXPOSURE**: DEVICE\_EXPOSURE\_ID  **DRUG\_EXPOSURE**: DRUG\_EXPOSURE\_ID  **MEASUREMENT**: MEASUREMENT\_ID  **OBSERVATION**: OBSERVATION\_ID  **VISIT\_OCCURRENCE**: VISIT\_OCCURRENCE\_ID | This allows the cost to be linked to the associated record. If a PROC1 code from the TEMP\_MEDICAL table maps to a CONCEPT\_ID with a domain of device, then this will be the DEVICE\_EXPOSURE\_ID assigned in the DEVICE\_EXPOSURE table.  Costs coming from the DRUG\_CLAIMS table will have COST\_EVENT\_ID = DRUG\_EXPOSURE\_ID  Only costs coming from the INPATIENT\_ADMISSIONS table will have a COST\_EVENT\_ID that corresponds with the VISIT\_OCCURRENCE\_ID |  |
| cost\_domain\_id | **-** | This reflects the provenance of the code. Refer to the above table for how to assign this variable |  |
| cost\_type\_concept\_id | - |  |  |
| currency\_concept\_id | **-** | This will be ‘**44818668**’ for all rows since this is a US claims database and paid in US Dollars |  |
| total\_charge | - | - |  |
| total\_cost | **DRUG\_CLAIMS:**  AWP | - |  |
| total\_paid | *PAID\_BY\_PAYER + PAID\_BY\_PATIENT + PAID\_BY\_PRIMARY* |  |  |
| paid\_by\_payer | **INPATIENT\_ADMISSIONS:** TOTNET  **OUTPATIENT\_SERVICES:** NETPAY  **INPATIENT\_SERVICES:**  NETPAY  **DRUG\_CLAIMS:**  SUM(NETPAY) |  |  |
| paid\_by\_patient | *PAID\_PATIENT\_COPAY +PAID\_PATIENT\_COINSURANCE +PAID\_PATIENT+DEDUCTIBLE* |  |  |
| paid\_patient\_copay | **INPATIENT\_ADMISSIONS:** TOTCOPAY  **OUTPATIENT\_SERVICES:** COPAY  **INPATIENT\_SERVICES:**  COPAY  **DRUG\_CLAIMS:**  SUM(COPAY) |  |  |
| paid\_patient\_coinsurance | **INPATIENT\_ADMISSIONS:** TOTCOINS  **OUTPATIENT\_SERVICES:** COINS  **INPATIENT\_SERVICES:**  COINS  **DRUG\_CLAIMS:**  SUM(COINS) |  |  |
| paid\_patient\_deductible | **INPATIENT\_ADMISSIONS:** TOTDED  **OUTPATIENT\_SERVICES:** DEDUCT  **INPATIENT\_SERVICES:**  DEDUCT  **DRUG\_CLAIMS:**  SUM(DEDUCT) |  |  |
| paid\_by\_primary | **INPATIENT\_ADMISSIONS:** TOTCOB  **OUTPATIENT\_SERVICES:**  COB  **INPATIENT\_SERVICES:**  COB  **DRUG\_CLAIMS:**  SUM(COB) |  |  |
| paid\_ingredient\_cost | **DRUG\_CLAIMS:**  SUM(INGCOST) |  |  |
| paid\_dispensing\_fee | **DRUG\_CLAIMS:**  SUM(DISPFEE) |  |  |
| payer\_plan\_period\_id |  | Lookup associated PAYER\_PLAN\_PERIOD\_ID. Look up by PERSON\_ID and PROCEDURE\_DATE. If there no match, put NULL. | There should only be one possible plan. |
| amount\_allowed | - | - |  |
| revenue\_code\_concept\_id | **OUTPATIENT\_SERVICES:**  REVCODE  **INPATIENT\_SERVICES:**  REVCODE | Use Vocab pull 3.1.1.  Filters:  WHERE SOURCE\_VOCABULARY\_ID IN ('Revenue Code')  AND TARGET\_VOCABULARY\_ID IN ('Revenue Code') |  |
| drg\_concept\_id | **INPATIENT\_ADMISSIONS**: DRG  **INPATIENT\_SERVICES**: DRG | Use Vocab pull 3.1.1.  Filters:  WHERE SOURCE\_VOCABULARY\_ID IN ('DRG')  AND SOURCE\_CONCEPT\_CLASS\_ID IN ('MS-DRG')  AND TARGET\_VOCABULARY\_ID IN ('DRG')  AND TARGET\_CONCEPT\_CLASS\_ID IN ('MS-DRG')  AND TSVCDAT >= TARGET\_VALID\_START\_DATE AND TSVCDAT <= TARGET\_VALID\_END\_DATE  AND TARGET\_STANDARD\_CONCEPT IS NOT NULL  AND TARGET\_INVALID\_REASON IS NULL | The filter to the left should be used for records coming from the INPATIENT\_SERVICES table only. When a cost record comes from the INPATIENT\_ADMISSIONS table replace TSVCDAT with DISDATE |
| revenue\_code\_source\_value | **OUTPATIENT\_SERVICES:**  REVCODE  **INPATIENT\_SERVICES:**  REVCODE |  |  |
| drg\_source\_value | **INPATIENT\_ADMISSIONS**: DRG  **INPATIENT\_SERVICES**: DRG | - |  |

## Table Name: DEATH

The Death table is designed to capture the time when a Person is deceased and cause of death. Death in Truven can be captured by discharge status in (20,40,41,12) (DSTATUS in FACILITY\_HEADER, INPATIENT\_ADMISSIONS and INPATIENT\_SERVICES) or condition in one of the various diagnosis fields containing codes indicating death.

Key conventions:

* Use the TEMP\_MEDICAL, TEMP\_INPATIENT\_ADMISSIONS and TEMP\_FACILITY\_HEADER tables which are created during building the VISIT\_OCCURRENCE table to populate death information.
  + To identify death records using ICD codes use the Vocab Pull 3.1.1 with the filter:
    - WHERE SOURCE\_VOCABULARY\_ID = ‘JNJ\_DEATH’
    - This can be done by mapping STCM.SOURCE\_CODE to the source codes in the various diagnosis fields
  + This is only for identification of the records as the TARGET\_CONCEPT\_ID = 0
* The date of death will be associated to the VISIT\_END\_DATE.
* The following table provides codes for identifying death from medical claim tables and mapping to DEATH\_TYPE\_CONCEPT\_ID:
* Only one record will be kept for each patient. If multiple lines of death records are captured, records with the latest date of death will be kept and the death captured by discharge status has higher priority than those captured by ICD 9 diagnosis codes. If there is a discharge status selected over an ICD9 diagnosis code, take that date.
* If there are health care visits after one month (32 days) of death date, delete this record.

| **Table 11: DEATH** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| PERSON\_ID | **FACILITY\_HEADER:**  ENROLID  **OUTPATIENT\_SERVICES:**  ENROLID  **INPATIENT\_SERVICES:**  ENROLID  **INPATIENT\_ADMISSION:**  ENROLID  **LONG\_TERM\_CARE:**  ENROLID |  |  |
| DEATH\_DATE | **VISIT\_OCCURRENCE**-  VISIT\_END\_DATE | Refer to logic in building VISIT\_OCCURRENCE table for linking with VISIT\_OCCURRENCE\_ID |  |
| DEATH\_DATETIME | **-** | NULL |  |
| DEATH\_TYPE\_CONCEPT\_ID | **FACILITY\_HEADER:**  DSTATUS, DX1, DX2, DX3, DX4, DX5, DX6, DX7, DX8, DX9  **INPATIENT\_SERVICES:**  DSTATUS, PDX, DX1, DX2, DX3, DX4  **INPATIENT\_ADMISSIONS:**  DSTATUS, PDX, DX1, DX2, DX3, DX4, DX5, DX6, DX7, DX8, DX9, DX10, DX11, DX12, DX13, DX14, DX15  **OUTPATIENT\_SERVICES:**  DX1, DX2, DX3, DX4  **LONG\_TERM\_CARE:**  DX1, DX2, DX3, DX4 | 38003566 for death captured by Discharge status, 38003567 for death captured by condition | Can be found under VOCABULARY\_ID = ‘Death Type’ |
| CAUSE\_OF\_DEATH\_CONCEPT\_ID | **-** | 0 |  |
| CAUSE\_OF\_DEATH\_SOURCE\_VALUE | **-** | NULL |  |
| CAUSE\_SOURCE\_CONCEPT\_ID | **-** | 0 |  |

## Table Name: DRUG\_EXPOSURE

Drug exposures are garnered from DRUG\_CLAIMS table but can also be extracted from procedure codes in PROCEDURE\_OCCURRENCE table.

Key conventions:

* Prescription drug:
  + - 1. Keep records with SVCDATE that fall within an OBSERVATION\_PERIOD available for this person.
      2. Remove duplicate records: Find all distinct records via the following columns: ENROLID, PAYER\_PLAN\_PERIOD\_ID, NDCNUM, FILL\_DATE, METQTY, DAYSUPP, COPAY, COINS, DEDUCT, NETPAY, COB, INGCOST, DISPFEE, AWP, REFILL.
      3. From the previous distinct, group by ENROLID, NDCNUM, SVCDATE and sum the following columns together for the DRUG\_EXPOSURE table: SUM(METQTY) as QUANTITY, SUM(DAYSUPP) as DAYS\_SUPPLY, SUM(COPAY) as PAID\_COPAY, SUM(COINS) as PAID\_COINSURANCE, SUM(DEDUCT) as PAID\_TOWARD\_DEDUCTIBLE,SUM(NETPAY) as PAID\_BY\_PAYER, SUM(COB) as PAID\_BY\_COORDINATION\_BENEFITS, SUM(INGCOST) as INGREDIENT\_COST, SUM(DISPFEE) as DISPENSING\_FEE, SUM(AWP) as AVERAGE\_WHOLESALE\_PRICE, **and MAX(REFILL) as REFILLS**. **Then set DAYS\_SUPPLY as 0 if it is negative or 365 if it is >365**. Assign a unique DRUG\_EXPOSURE\_ID for each record. This intermediate table will be used for populating DRUG\_COST table.
      4. When mapping prescription drug, map the 11-digit NDC code to SOURCE\_CODE in OMOP vocab first. If no mapping found, map the first 9 digits of NDC code to SOURCE\_CODE. To be considered a valid mapping SVCDATE must fall between SOURCE\_VALID\_START\_DATE and SOURCE\_VALID\_END\_DATE
* Procedure drug:

1. Refer to the documentation for VISIT\_OCCURRENCE on how to map procedures in **TEMP\_MEDICAL, TEMP\_INPATIENT\_ADMISSIONS,** and **TEMP\_FACILITY\_HEADER** to CONCEPT\_IDs. Any procedure with a DOMAIN\_ID of ‘Drug’ should be written to the DRUG\_EXPOSURE table.
2. Set VISIT\_PROVID and VISIT\_PROVSTD as PPROV and PROVCAT (associated provider id and its specialty for the procedure occurrence, will be used to populate ASSOCIATED\_PROVIDER\_ID)

| **Table 12: DRUG\_EXPOSURE** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| DRUG\_EXPOSURE\_ ID | - | System generated. |  |
| PERSON\_ID | ENROLID |  |  |
| DRUG\_CONCEPT\_ ID | **DRUG\_CLAIMS:**  NDCNUM  **TEMP tables :**  PROCEDURE\_SOURCE\_VALUE | Use Vocab map 3.1.2.  Filter if drug coming from **DRUG\_CLAIMS**:  WHERE SOURCE\_VOCABULARY\_ID IN ('NDC')  AND TARGET\_STANDARD\_CONCEPT IS NOT NULL  AND TARGET\_INVALID\_REASON IS NULL  AND SVCDATE BETWEEN SOURCE\_VALID\_START\_DATE AND SOURCE\_VALID\_END\_DATE  **TEMP tables:** These should have already been mapped to CONCEPT\_IDs with a DOMAIN\_ID of ‘Drug’ | For procedure drugs, the CONCEPT\_ID must have a DOMAIN\_ID of ‘Drug’ to be included. |
| DRUG\_EXPOSURE\_START\_DATE | **DRUG\_CLAIMS:**  SVCDATE  **For procedure drug codes from FACILITY\_HEADER / INPATIENT\_ADMISSIONS table:** use VISIT\_END\_DATE of its associated visit;  **For procedure drug code from PROC1 in OUTPATIENT\_SERVICES or INPATIENT\_SERVICES:**  SVCDATE. |  |  |
| DRUG\_EXPOSURE\_START\_DATETIME |  |  |  |
| VERBATIM\_END\_DATE |  | NULL |  |
| DRUG\_EXPOSURE\_END\_DATE | **DRUG\_CLAIMS:**  SVCDATE + SUM(DAYSUPP)  Else set to DRUG\_EXPOSURE\_START\_DATE |  |  |
| DRUG\_EXPOSURE\_END\_DATETIME | **-** | NULL |  |
| DRUG\_TYPE\_CONCEPT\_ID | **DRUG\_CLAIMS:**  NULL  **PROCEDURE\_OCCURRENCE:**  PROCEDURE\_TYPE\_CONCEPT\_ID | **DRUG\_CLAIMS:**  38000175 (Prescription dispensed in pharmacy)  **PROCEDURE\_OCCURRENCE**: if  PROCEDURE\_TYPE\_CONCEPT\_ID in (38000249, 38000250, 38000251, 38000252,  38000253, 38000254, 38000255,  38000256, 38000257, 38000258,  38000259, 38000260,38000261, 38000262, 38000263, 38000264,  38000265 )then 38000180 (Inpatient administration); else 38000179 (Physician administered drug, identified as procedure) | These CONCEPT\_IDs fall under DOMAIN\_ID = ‘Drug Type’ in CONCEPT table. |
| STOP\_REASON | - | NULL |  |
| REFILLS | **DRUG\_CLAIMS:**  REFILL  **TEMP tables:**  - | NULL for  **TEMP tables** | OMOP defines this column as the number of refills after the initial prescription. The initial prescription is not counted, values start with 0.  REFILL from DRUG\_CLAIM is defined as a number indicating whether this is the original prescription (0) or the refill number (e.g. 1, 2, etc.). |
| QUANTITY | **DRUG\_CLAIMS:**  SUM(METQTY)  **TEMP tables :** - | NULL for  **TEMP tables** |  |
| DAYS\_SUPPLY | **DRUG\_CLAIMS:**  SUM(DAYSUPP)  **TEMP tables :** - | NULL for  **TEMP\_tables** |  |
| SIG | - | NULL | "Sig" is short for the Latin, signetur, or "let it be labeled." |
| ROUTE\_CONCEPT\_ID | **-** | 0 |  |
| LOT\_NUMBER | **-** | NULL |  |
| PROVIDER\_ID | **DRUG\_CLAIMS:** -  **TEMP tables :** ASSOCIATED\_PROVIDER\_ID | NULL for prescription drug.  For data coming from **TEMP tables:**  Refer to key conventions for defining PPROV and PPROVCAT, then map them to PROVIDER\_SOURCE\_VALUE and  SPECIALTY\_SOURCE\_VALUE in Provider table to extract associated Provider ID. | If there is no associated PROVIDER\_ID then set as NULL |
| VISIT\_ OCCURRENCE\_ID | **DRUG\_CLAIMS:** -  **TEMP tables :** VISIT\_OCCURRENCE\_ID | NULL for prescription drug.  **TEMP tables :**  Refer to logic in building VISIT\_OCCURRENCE table for linking with VISIT\_OCCURRENCE\_ID |  |
| DRUG\_SOURCE\_ VALUE | **DRUG\_CLAIMS:**  NDCNUM  **TEMP\_MEDICAL, TEMP\_INPATIENT\_ADMISSIONS,** and **TEMP\_FACILITY\_HEADER :** PPROC, PROC1-PROC15 |  |  |
| DRUG\_SOURCE\_ CONCEPT\_ID | **DRUG\_CLAIMS:**  NDCNUM  **TEMP tables :**  PPROC, PROC1-PROC15 | Filter when drug coming from **DRUG\_CLAIMS**:  WHERE SOURCE\_VOCABULARY\_ID IN ('NDC')  AND TARGET\_VOCABULARY\_ID IN ('NDC')  Incoming PROCEDURE\_OCCURRENCE records should follow the SOURCE\_CONCEPT\_ID mapping from the source table. | Use Vocab pull 3.1.1. |
| ROUTE\_SOURCE\_VALUE | **-** | 0 |  |
| DOSE\_UNIT\_SOURCE\_VALUE | **-** | 0 |  |

## Table Name: DRUG\_ERA

A DRUG\_ERA is defined as a span of time when the person is assumed to be exposed to a particular drug. Successive periods of DRUG\_EXPOSUREs are combined under certain rules to produce continuous DRUG\_ERAs. The DRUG\_ERA table is populated by pulling from the DRUG\_EXPOSURE table within the CDM. DRUG\_ERAs are consolidated to their respective ingredient off the DRUG\_EXPOSURE table. A drug era is therefore understood as exposure to a certain compound over a certain period of time. There will only be one type of persistence window (duration that is allowed to elapse between drug exposures) applied to this CDM, which is 30 days.

Key conventions:

* Use the following steps to build this table off DRUG\_EXPOSURE table:

1. Exclude records with DRUG\_CONCEPT\_ID = 0.
2. Use the following logic to map DRUG\_CONCEPT\_ID to ingredient levels:

Select distinct A.concept\_id as Drug\_EXPOSURE\_concept\_id,

C.concept\_id as ingredient\_concept\_id

FROM CONCEPT C

JOIN CONCEPT\_ANCESTOR CA

ON CA.ancestor\_concept\_id = C.concept\_id

and c.vocabulary\_id = 'RxNorm'

and c.concept\_class\_id = 'Ingredient'

and invalid\_reason is null

JOIN concept A

ON CA.descendant\_CONCEPT\_ID = A.CONCEPT\_ID

1. Replace the values of DRUG\_CONCEPT\_ID with their ingredient CONCEPT\_IDs identified by step 2), and exclude records with DRUG\_CONCEPT\_IDs that can’t be mapped to ingredient level.
2. Calculate DRUG\_EXPOSURE\_END\_DATE: If DRUG\_TYPE\_CONCEPT\_ID in (38000175, 38000176) then set to DRUG\_EXPOSURE\_START\_DATE+DAYS\_SUPPLY,

Else set to DRUG\_EXPOSURE\_START\_DATE.

1. Sort DRUG\_EXPOSURE table by PERSON\_ID, DRUG\_CONCEPT\_ID, DRUG\_EXPOSURE\_START\_DATE and DRUG\_EXPOSURE\_END\_DATE in ascending order.
2. Combine records as long as both PERSON\_ID and DRUG\_CONCEPT\_ID don’t change and the time between DRUG\_EXPOSURE\_END\_DATE of one record and DRUG\_EXPOSURE\_START\_DATE of the next is 30 days or less (<=30).

* Compound drugs can created multiple ERAs since they can be mapped to multiple ingredients.

| **Table 13: DRUG\_ERA** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| DRUG\_ERA\_ID | - | System generated. |  |
| PERSON\_ID | **DRUG\_EXPOSURE:**  PERSON\_ID |  |  |
| DRUG\_CONCEPT\_ID | **DRUG\_EXPOSURE:**  DRUG\_CONCEPT\_ID | Use the logic above to map to ingredient CONCEPT\_ID and exclude records. |  |
| DRUG\_ERA\_START\_DATE | **DRUG\_EXPOSURE:**  DRUG\_EXPOSURE\_START\_DATE | The start date for the drug era constructed from the individual instances of drug exposures. It is the start date of the very first chronologically recorded instance of utilization of a drug. |  |
| DRUG\_ERA\_END\_DATE | **DRUG\_EXPOSURE**:  DAYS\_SUPPLY, DRUG\_EXPOSURE\_START\_DATE, DRUG\_TYPE\_CONCEPT\_ID | The end date for the drug era constructed from the individual instance of drug exposures. It is the end date of the final continuously recorded instance of utilization of a drug. |  |
| DRUG\_EXPOSURE\_COUNT | - | Sum up the number of DRUG\_EXPOSURES for this PERSON\_ID and this CONCEPT\_ID during the exposure window being built. |  |
| GAP\_DAYS | - | Sum of the days in the drug\_era that were not covered by a drug\_exposure\_record |  |

## Table Name: OBSERVATION

OBSERVATIONS describe symptoms, clinical observations, lab test etc. about the person.

Key conventions:

* For data coming from a DX or PROC field in **TEMP\_MEDICAL, TEMP\_INPATIENT\_ADMISSIONS, TEMP\_FACILITY\_HEADER (TEMP tables)** with the DOMAIN\_ID ‘Observation’**:**

1. Remove any duplicate records
2. Set VISIT\_PROVID and VISIT\_PROVSTD as PPROV and PROVCAT

* We will use the field **DSTATUS** in TEMP\_INPATIENT\_ADMISSIONS and TEMP\_FACILITY\_HEADER tables that are created during the building of the VISIT\_OCCURRENCE table to map discharge status
  + For all records the CONCEPT\_ID = 4202605 (Discharge Status)
  + Use the below table to map VALUE\_AS\_STRING

|  |  |
| --- | --- |
| **DSTATUS** | **VALUE\_AS\_STRING** |
| 01 | Discharged to home self care |
| 02 | Transfer to short term hospital |
| 03 | Transfer to SNF |
| 04 | Transfer to ICF |
| 05 | Transfer to other facility |
| 06 | Discharged home under care |
| 07 | Left against medical advice |
| 08-19 | Other alive status |
| 20 | Died |
| 21 | Discharged/transferred to court/law enforcement |
| 30-39 | Still patient |
| 40-42 | Other died status |
| 43 | Discharged/transferred to federal hospital |
| 50 | Discharged to home (from Hospice) |
| 51 | Transfer to med fac (from Hospice) |
| 61 | Transfer to Medicare approved swing bed |
| 62 | Transferred to inpatient rehab facility (IRF) |
| 63 | Transferred to long term care hospital (LTCH) |
| 64 | Transferred to nursing facility Medicaid certified |
| 65 | Transferred to psychiatric hospital or unit |
| 66 | Transferred to critical access hospital (CAH) |
| 70 | Transfer to another facility NEC |
| 71 | Transfer/referred to other facility for outpt svcs |
| 72 | Transfer/referred to this facility for outpt svcs |
| 99 | Transfer (Hospital ID MDST change) |

Additionally we add a Mental Health and Substance Abuse flags to the OBSERVATION table. In Truven in the ENROLLMENT\_DETAIL table there is information about when an individual will have this type of coverage. We will add a 1 to VALUE\_AS\_NUMBER when the coverage exists and else a 0. Make sure to take distinct records per ENROLID, DTEND, MHSACOVG.

| **Table 15: OBSERVATION** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| OBSERVATION\_ID | - | System generated. |  |
| PERSON\_ID | ENROLID |  |  |
| OBSERVATION\_CONCEPT\_ID | **TEMP tables:**  PDX, DX1-DX15, PPROC, PROC1-PROC15  **ENROLLMENT\_DETAIL:**  -  **DSTATUS:**  **-** | **TEMP tables:** These should have already been mapped to **CONCEPT\_IDS**  **ENROLLMENT\_DETAIL:**  900000010  **DSTATUS:**  4202605 | Use Vocab pull 3.1.2.  Refer to [CONDITION\_ OCCURRENCE](#_Table_Name:_) for description of how DOMAIN\_IDs are used to direct data to correct tables |
| OBSERVATION\_DATE | **For observations from FACILITY\_HEADER / INPATIENT\_ADMISSIONS table:** use VISIT\_END\_DATE of its associated visit;  **For observation from PROC1 in OUTPATIENT\_SERVICES or INPATIENT\_SERVICES:**  SVCDATE.  **ENROLLMENT\_DETAIL:**  DTEND |  |  |
| OBSERVATION\_DATETIME | - | NULL |  |
| OBSERVATION\_TYPE\_CONCEPT\_ID | - | **TEMP Tables**:  If the record is coming from another table, like the CONDITION\_OCCURENCE, keep the types that would have been assigned in that table.  **ENROLLMENT\_DETAIL:**  900000009- Mental Health Substance Abuse Coverage Indicator  **DSTATUS:** 38000280 – Observation recorded from EHR |  |
| VALUE\_AS\_NUMBER | **TEMP tables:** 0  **ENROLLMENT\_DETAIL:**  MHSACOVG | **ENROLLMENT\_DETAIL:**  CASE WHEN MHSACOVG IS NULL THEN 0 ELSE MHSACOVG END |  |
| VALUE\_AS\_STRING | **TEMP tables:  -**  **ENROLLMENT\_DETAIL:**  -  **DSTATUS:**  **-** | **DSTATUS:** Refer to table above for how to map the discharge status to a string |  |
| VALUE\_AS\_CONCEPT\_ID | 0 |  |  |
| QUALIFIER\_CONCEPT\_ID | 0 |  |  |
| UNIT\_CONCEPT\_ID | 0 |  |  |
| PROVIDER\_ID | **TEMP tables:**  PPROV  **ENROLLMENT\_DETAIL:**  - | **TEMP Tables**: Map STDPROV and PROVID  to PROVIDER\_SOURCE\_VALUE and  SPECIALTY\_SOURCE\_VALUE in Provider table to extract associated Provider ID. | If there is no associated PROVIDER\_ID then set as NULL |
| VISIT\_OCCURRENCE\_ID | **TEMP tables:** VISIT\_OCCURRENCE\_ID  **ENROLLMENT\_DETAIL:**  - | **TEMP tables**:  Refer to logic in building VISIT\_OCCURRENCE table for linking with VISIT\_OCCURRENCE\_ID |  |
| OBSERVATION\_SOURCE\_VALUE | **TEMP tables:** PDX, DX1-DX15, PPROC, PROC1-PROC15  **ENROLLMENT\_DETAIL:**  -  **DSTATUS:** DSTATUS | **TEMP tables:**  For records incoming from PROCEDURE\_OCCURRENC or CONDITION\_OCCURRENCE use the respective SOURCE\_VALUES from those tables.  **ENROLLMENT\_DETAIL:**  “MHSACOVG” |  |
| OBSERVATION\_SOURCE\_CONCEPT\_ID | **TEMP tables:** PDX, DX1-DX15, PPROC, PROC1-PROC15 | **TEMP tables:**  For records incoming from PROCEDURE\_OCCURRENC or CONDITION\_OCCURRENCE use the respective SOURCE\_VALUE\_CONCEPT\_IDs from those tables. |  |
| UNIT\_SOURCE\_VALUE | - |  |  |
| QUALIFIER\_SOURCE\_VALUE | - |  |  |

## Table Name: ORGANIZATION

The ORGANIZATION table contains a list of uniquely identified health care organizations (hospitals, clinics, practices, etc.). This is the highest level of the health care hierarchy. Most organizations contain multiple CARE\_SITEs, but each CARE\_SITE belongs to one Organization. Location information about the organization is stored in the location table and referenced by the foreign key ORGANIZATION\_LOCATION\_ID.

Truven does not have clear organization information so this table will only contain one value representing the fact no organization information would be captured.

| **Table 16: ORGANIZATION** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| ORGANIZATION\_ID | NULL |  |  |
| PLACE\_OF\_SERVICE\_CONCEPT\_ID | 0 |  |  |
| LOCATION\_ID | NULL |  |  |
| ORGANIZATION\_SOURCE\_VALUE | - | NULL |  |
| PLACE\_OF\_SERVICE\_SOURCE\_VALUE | - | NULL |  |

## Table Name: CARE\_SITE

The CARE\_SITE table contains a list of uniquely identified points of care, or an individual clinical location within an organization. Each CARE\_SITE belongs to one ORGANIZATION.

Truven does not have clear care site information so this table will only contain one value representing the fact that no care site information will be captured.

| **Table 17: CARE\_SITE** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| CARE\_SITE\_ID | NULL |  |  |
| LOCATION\_ID | NULL |  |  |
| ORGANIZATION\_ID | NULL |  |  |
| PLACE\_OF\_SERVICE\_CONCEPT\_ID | 0 |  |  |
| CARE\_SITE\_SOURCE\_VALUE | - | NULL |  |
| PLACE\_OF\_SERVICE\_SOURCE\_VALUE | - | NULL |  |

## Table Name: COHORT

Cohort defines groups of entities that are exposed to common circumstances. At this time, we will not incorporate any special logic to create cohorts however in the future this logic will be provided from the epidemiology team.

| **Table 18: COHORT** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| COHORT\_ID |  |  |  |
| COHORT\_CONCEPT\_ID |  |  |  |
| COHORT\_START\_DATE |  |  |  |
| COHORT\_END\_DATE |  |  |  |
| SUBJECT\_ID |  |  |  |
| STOP\_REASON |  |  |  |

## Table Name: COHORT\_ATTRIBUTE

The COHORT\_ATTRIBUTE table contains attributes associated with each subject within a cohort, as defined by a given set of criteria for a duration of time.

| **Table 19: COHORT\_ATTRIBUTE** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| COHORT\_DEFINITION\_ID |  |  |  |
| SUBJECT\_ID |  |  |  |
| COHORT\_START\_DATE |  |  |  |
| COHORT\_END\_DATE |  |  |  |
| ATTRIBUTE\_DEFINITION\_ID |  |  |  |
| VALUE\_AS\_NUMBER |  |  |  |
| VALUE\_AS\_CONCEPT\_ID |  |  |  |

## Table Name: MEASUREMENT

The Measurement table contains lab tests, vital signs, etc. that were conducted during a valid OBSERVATION\_PERIOD. Lab values are sourced from the **TEMP\_MEDICAL, TEMP\_INPATIENT\_ADMISSIONS,** and **TEMP\_FACILITY\_HEADER** tables. Units of measure are also mapped to standard concept identifiers in the dictionary for standardization. Lab observations from **TEMP\_MEDICAL, TEMP\_INPATIENT\_ADMISSIONS,** and **TEMP\_FACILITY\_HEADER (TEMP tables)** are codes in the DX and PROC fields that have been mapped to standard concepts with DOMAIN\_IDs of ‘Measurement’.

| **Table 19: MEASUREMENT** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| MEASUREMENT\_ID | - | System Generated |  |
| PERSON\_ID | ENROLID |  |  |
| MEASUREMENT\_CONCEPT\_ID | **TEMP tables:**  PDX, DX1-DX15, PPROC, PROC1-PROC15 | **TEMP tables:** These should already have been mapped to CONCEPT\_IDs | Refer to [CONDITION\_OCCURRENCE](#_Table_Name:_) for description of how DOMAIN\_IDs are used to direct data to correct tables |
| MEASUREMENT\_ DATE | From **TEMP\_MEDICAL:**  SVCDATE  **For measurements from TEMP\_FACILITY\_HEADER / TEMP\_INPATIENT\_ADMISSIONS table:** use VISIT\_END\_DATE of its associated visit; |  |  |
| MEASUREMENT\_DATETIME | - | NULL |  |
| MEASUREMENT\_ TIME | - | NULL |  |
| MEASUREMENT\_ TYPE\_CONCEPT\_ID | - | If the record is coming from another table, like the CONDITION\_OCCURENCE, keep the types that would have been assigned in that table. |  |
| OPERATOR\_ CONCEPT\_ID | - | 0 |  |
| VALUE\_AS\_NUMBER | **TEMP Tables:** NULL |  |  |
| VALUE\_AS\_ CONCEPT\_ID | **TEMP Tables/HRA:** 0 |  |  |
| UNIT\_CONCEPT\_ID | **TEMP Tables/HRA:** 0 |  |  |
| RANGE\_LOW | **TEMP Tables/HRA:** NULL |  |  |
| RANGE\_HIGH | **TEMP Tables/HRA:** NULL |  |  |
| PROVIDER\_ID | PROVID, STDPROV | Map STDPROV and PROVID  to PROVIDER\_SOURCE\_VALUE and  SPECIALTY\_SOURCE\_VALUE in Provider table to extract associated Provider ID. | If there is no associated PROVIDER\_ID then set as NULL |
| VISIT\_OCCURRENCE\_ID | **Measurement from TEMP tables use VISIT\_OCCURRENCE**: VISIT\_OCCURRENCE\_ID | Refer to logic in building VISIT\_OCCURRENCE table for linking with VISIT\_OCCURRENCE\_ID |  |
| MEASUREMENT\_ SOURCE\_VALUE | **TEMP Tables:** PDX, DX1-DX15, PPROC, PROC1-PROC15 | **TEMP Tables:** Code as it appears in the table |  |
| MEASUREMENT\_SOURCE\_CONCEPT\_ID | **TEMP Tables:** PDX, DX1-DX15, PPROC, PROC1-PROC15 | Incoming CONDITION\_OCCURRENCE (PDX/DX#) or PROCEDURE\_OCCURRENCE records (PPROC, PROC#) should follow the SOURCE\_CONCEPT\_ID mapping from the CONDITION\_OCCURRENCE table. |  |
| UNIT\_SOURCE\_ VALUE | **TEMP Tables/HRA:** NULL |  |  |
| VALUE\_SOURCE\_ VALUE | **TEMP Tables:** NULL |  |  |

## Table Name: SPECIMEN

The SPECIMEN table contains information identifying biological samples from a person.

Truven does not have clear specimen information so this table will only contain one value representing the fact that no specimen information will be captured.

| **Table 20: SPECIMEN** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| SPECIMEN\_ID | 0 |  |  |
| PERSON\_ID | 0 |  |  |
| SPECIMEN\_CONCEPT\_ID | 0 |  |  |
| SPECIMEN\_TYPE\_CONCEPT\_ID | 0 |  |  |
| SPECIMEN\_DATE | - | NULL |  |
| SPECIMEN\_DATETIME | - | NULL |  |
| QUANTITY | - | NULL |  |
| UNIT\_CONCEPT\_ID | 0 |  |  |
| ANATOMIC\_SITE\_CONCEPT\_ID | 0 |  |  |
| DISEASE\_STATUS\_CONCEPT\_ID | 0 |  |  |
| SPECIMEN\_SOURCE\_ID | 0 |  |  |
| SPECIMEN\_SOURCE\_VALUE | - | NULL |  |
| UNIT\_SOURCE\_VALUE | - | NULL |  |
| ANATOMIC\_SITE\_SOURCE\_VALUE | - | NULL |  |
| DISEASE\_STATUS\_SOURCE\_VALUE | - | NULL |  |

## Table Name: DEVICE\_EXPOSURE

The **DEVICE\_EXPOSURE** table will be populated with records from **TEMP\_MEDICAL, TEMP\_INPATIENT\_ADMISSIONS** and **TEMP\_FACILITY\_HEADER** where the PROC code was mapped to a standard concept with a DOMAIN\_ID of ‘Device’

| **Table 21: DEVICE\_EXPOSURE** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| DEVICE\_EXPOSURE\_ID | - | System Generated |  |
| PERSON\_ID | ENROLID |  |  |
| DEVICE\_CONCEPT\_ID | PPROC, PROC1-PROC15 | These codes should have already been mapped to CONCEPT\_IDs to find the DOMAIN\_ID of ‘Device’ | Refer to [CONDITION\_OCCURRENCE](#_Table_Name:_) for description of how DOMAIN\_IDs are used to direct data to correct tables |
| DEVICE\_EXPOSURE\_START\_DATE | **VISIT\_OCCURRENCE** **–**VISIT\_START\_DATE |  |  |
| DEVICE\_EXPOSURE\_START\_DATETIME | **-** | NULL |  |
| DEVICE\_EXPOSURE\_END\_DATE | - | NULL |  |
| DEVICE\_EXPOSURE\_END\_DATETIME | - | NULL |  |
| DEVICE\_TYPE\_CONCEPT\_ID | - | If the record is coming from another table, like the PROCEDURE\_OCCURENCE, keep the types that would have been assigned in that table. |  |
| UNIQUE\_DEVICE\_ID | - |  |  |
| QUANTITY | 0 |  |  |
| PROVIDER\_ID | PROVID, STDPROV | Map STDPROV and PROVID  to PROVIDER\_SOURCE\_VALUE and  SPECIALTY\_SOURCE\_VALUE in Provider table to extract associated Provider ID. | If there is no associated PROVIDER\_ID then set as NULL |
| VISIT\_OCCURRENCE\_ID | **VISIT\_OCCURRENCE** **–**VISIT\_OCCURRENCE\_ID | Refer to logic in building VISIT\_OCCURRENCE table for linking with VISIT\_OCCURRENCE\_ID |  |
| DEVICE\_SOURCE\_VALUE | PPROC. PROC1-PROC15 |  |  |
| DEVICE\_SOURCE\_CONCEPT\_ID | PPROC. PROC1-PROC15 | WHERE SOURCE\_VOCABULARY\_ID IN ('ICD9Proc','HCPCS','CPT4')  AND TARGET\_VOCABULARY\_ID IN ('ICD9Proc','HCPCS','CPT4')  AND TARGET\_CONCEPT\_CLASS\_ID NOT IN ('HCPCS Modifier','CPT4 Modifier') | Use the code in Section 3.1.1 |

## Table Name: NOTE

The NOTE table captures free text information recorder by a provider about a patient.

Truven does not have clear note information so this table will only contain one value representing the fact that no note information will be captured.

| **Table 22: NOTE** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| NOTE\_ID | 0 |  |  |
| PERSON\_ID | 0 |  |  |
| NOTE\_DATE | - | NULL |  |
| NOTE\_DATETIME | - | NULL |  |
| NOTE\_TYPE\_CONCEPT\_ID | 0 |  |  |
| NOTE\_CLASS\_CONCEPT\_ID | 0 | NULL |  |
| NOTE\_TITLE | - | NULL |  |
| NOTE\_TEXT | - | NULL |  |
| ENCODING\_CONCEPT\_ID | 0 | NULL |  |
| LANGUAGE\_CONCEPT\_ID | 0 | NULL |  |
| PROVIDER\_ID | 0 |  |  |
| NOTE\_SOURCE\_VALUE | - | NULL |  |
| VISIT\_OCCURRENCE\_ID | 0 |  |  |

## Table Name: NOTE\_NLP

The NOTE\_NLP table encodes all output of NLP on clinical notes. Each row represents a single extracted term from a note

Truven does not have clear note information so this table will only contain one value representing the fact that no note information will be captured.

| **Table 22: NOTE\_NLP** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| NOTE\_NLP\_ID | 0 |  |  |
| NOTE\_ID | 0 |  |  |
| SECTION\_CONCEPT\_ID | 0 | NULL |  |
| SNIPPET | - | NULL |  |
| OFFSET | - | NULL |  |
| LEXICAL\_VARIANT | - | NULL |  |
| NOTE\_NLP\_CONCEPT\_ID | 0 | NULL |  |
| NOTE\_NLP\_SOURCE\_CONCEPT\_ID | 0 | NULL |  |
| NLP\_SYSTEM | - | NULL |  |
| NLP\_DATE | - | NULL |  |
| NLP\_DATETIME | - | NULL |  |
| TERM\_EXISTS | - | NULL |  |
| TERM\_TEMPORAL | - | NULL |  |
| TERM\_MODIFIERS | - | NULL |  |

## Table Name: FACT\_RELATIONSHIP

The FACT\_RELATIONSHIP table captures the relationships between facts in any table of the CDM. These can be person relationships, care site relationships, indication relationships, usage relationships, or facts derived from one another

Truven does not have clear relationship information so this table will only contain one value representing the fact that no relationship information will be captured.

| **Table 25: FACT\_RELATIONSHIP** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| DOMAIN\_CONCEPT\_ID\_1 | 0 |  |  |
| FACT\_ID\_1 | 0 |  |  |
| DOMAIN\_CONCEPT\_ID\_2 | 0 |  |  |
| FACT\_ID\_2 | 0 |  |  |
| RELATIONSHIP\_CONCEPT\_ID | 0 |  |  |

## Table Name: CDM\_SOURCE

The CDM\_SOURCE table contains detail about the source database and the process used to transform the data into the OMOP Common Data Model.

| **Table 26: CDM\_SOURCE** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| CDM\_SOURCE\_NAME | - | Truven Health MarketScan® Multi-State Medicaid Database |  |
| CDM\_SOURCE\_ABBREVIATION | - | MDCD |  |
| CDM\_HOLDER |  | Janssen R&D |  |
| SOURCE\_DESCRIPTION |  | The Truven Health MarketScan® Research Databases contain individual-level, de-identified, healthcare claims information from employers, health plans, hospitals, Medicare, and Medicaid programs. Since their creation in the early 1990s, the MarketScan Databases have grown into one of the largest collections of de-identified patient-level data in the nation. These databases reflect the real-world of treatment patterns and costs by tracking millions of patients as they travel through the healthcare system offering detailed information about all aspects of care. Data from individual patients are integrated from all providers of care, maintaining all healthcare utilization and cost record connections at the patient level. Used primarily for research, these databases are fully HIPAA compliant. Research using MarketScan data has been widely publicized in peer-reviewed journals.  This retrospective claims analysis utilized data from the Truven Health MarketScan® Medicaid Database. These data included health insurance claims for Medicaid enrollees from multiple states across the continuum of care (e.g. inpatient, outpatient, outpatient pharmacy) as well as variables of interest for Medicaid populations, such as ethnicity and Medicare eligibility. |  |
| SOURCE\_DOCUMENTATION\_REFERENCE |  | http://hicoe.jnj.com/DataSources/Truven/mdcd |  |
| CDM\_ETL\_REFERENCE |  | http://www.ohdsi.org/web/wiki/ doku.php?id=documentation:example\_etls |  |
| SOURCE\_RELEASE\_DATE |  | SELECT VERSION\_DATE  FROM [\_Version] | Get from the source tables. |
| CDM\_RELEASE\_DATE |  | SELECT CONVERT(VARCHAR(10), GETDATE(),102) | Get the date the run completes on. |
| CDM\_VERSION | - | V5.0 |  |
| VOCABULARY\_VERSION | - | SELECT VOCABULARY\_VERSION  FROM vocabulary  WHERE VOCABULARY\_ID = 'None' | Taken from the Vocabulary loaded into the CDM. |

## Table Name: CDM\_DOMAIN\_META

| **Table 4: CDM\_DOMAIN\_META** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| DOMAIN\_ID | DOMAIN\_ID from the table below |  |  |
| DESCRIPTION | DESCRIPTION from the lookup table below |  |  |

|  |  |  |
| --- | --- | --- |
| **Meta Data Lookup Table** | | |
| **TABLE NAME** | **DOMAIN\_ID** | **DESCRIPTION** |
| PERSON | Person | Person information is pulled from insurance enrollment data where the individual both has medical and prescription benefits or the individual has both Medicaid and Medicare eligibility. The month of birth is not provided however for enrollees who start their enrollment the year they are born we extrapolate their month of birth from the month where their enrollment starts, for the majority of patients only year of birth is available. Persons who change gender over their enrollments or change year of birth are excluded. |
| OBSERVATION\_PERIOD | Observation Period | An observation period is a representation of when a patient was enrolled in a health insurance plan and had prescription benefits. Periods of continuous enrollment are consolidated by combining monthly records as long as the time between the end of one enrollment period and the start of the next is 32 days or less. |
| CARE\_SITE | Care Site | There is not clear care site information in this source so no data will be captured within this table. |
| VISIT\_OCCURRENCE | Visit | A standardized definition of visit logic is applied to our U.S. claims data. The data vendors do apply methods to define inpatient, outpatient, ER and long term care visits however we have found inconsistencies between the data vendors of claims data. Inpatient visits defined by Truven remain as IP unless they have a revenue code suggesting it is actually an ER visit. Additionally outpatient service charges that are associated to room and board are categorized as inpatient. ER or outpatient services during the middle of an inpatient stay are associated to that inpatient stay. To learn additional information please refer to the publication: Voss EA, Ma Q, Ryan PB. The impact of standardizing the definition of visits on the consistency of multi-database observational health research. BMC Med Res Methodol. 2015 Mar 8;15:13. doi: 10.1186/s12874-015-0001-6. PubMed PMID: 25887092; PubMed Central PMCID: PMC4369827. |
| PROVIDER | Provider | Unique list of health care providers (physicians). Truven does provide some provider information however some of the providers listed by Truven may also be considered care sites or organizations. Since there is not clear way to decipher between all items identified as providers by Truven, regardless if they are truly organizations or care sites, they will be added to this table. |
| DEATH | Death | Death in Truven can be captured at discharge from an inpatient visits or in some cases by diagnosis code. The death data in this source should not be considered complete, for example if a patient left a hospital and later died at home that would not be captured. Additionally if a death was recorded however if the patient continues to have services charges after 30 days of the death date we assume the death data was faulty. |
| CONDITION\_OCCURRENCE | Condition | Condition records are primarily recorded as codified claims data (e.g. ICD9 or ICD10 records that are submitted associated with a service). |
| DRUG\_EXPOSURE | Drug | Drug exposure records are primarily recorded as codified claims data (e.g. an NDC code or a procedure code that includes a drug). If the OMOP Vocabulary deems a code of a non-traditional drug centric vocabulary is in fact a drug exposure, the record will move to this table (e.g. CPT4- 90690- “Typhoid vaccine, live, oral” maps to drug concept in the OMOP Vocabularies so the CDM\_BUILDER will move the record to the DRUG\_EXPOSURE table instead of the procedure table). |
| PROCEDURE\_OCCURRENCE | Procedure | Procedure occurrence records are recorded as codified claims data (e.g. a CPT4 code or ICD9 procedure code). If the OMOP Vocabulary deems a procedure code to be of a type of another domain (e.g. CPT4- 90690- “Typhoid vaccine, live, oral” maps to drug concept in the OMOP Vocabularies so the CDM\_BUILDER will move the record to the DRUG\_EXPOSURE table instead of the procedure table) however in the case of the primary procedure code those will always write a record to this table in order to maintain cost data. |
| MEASUREMENT | Measurement | Truven MDCD does not contain lab data, however, if the OMOP Vocabulary deems a code of a non-traditional measurement centric vocabulary is in fact a measurement, the record will move to this table (e.g. ICD9- V85.22- “Body Mass Index 26.0-26.9, adult” usually thought of as a diagnosis code maps to a measurement concept in the OMOP Vocabularies so the CDM\_BUILDER will move the record to the MEASUREMENT table). |
| OBSERVATION | Observation | Codified data that is not a diagnosis, drug exposure, procedure, or measurement will become an observation. |

# Code Snippets

## Vocabulary Mapping

Use this code to map source codes to concept ids; change the source\_vocabulary\_id and target\_vocabulary\_id as needed.

### Source to Source

WITH CTE\_VOCAB\_MAP AS (

       SELECT c.concept\_code AS SOURCE\_CODE, c.concept\_id AS SOURCE\_CONCEPT\_ID, c.CONCEPT\_NAME AS SOURCE\_CODE\_DESCRIPTION,

                        c.vocabulary\_id AS SOURCE\_VOCABULARY\_ID, c.domain\_id AS SOURCE\_DOMAIN\_ID, c.concept\_class\_id AS SOURCE\_CONCEPT\_CLASS\_ID,

            c.VALID\_START\_DATE AS SOURCE\_VALID\_START\_DATE, c.VALID\_END\_DATE AS SOURCE\_VALID\_END\_DATE, c.invalid\_reason AS SOURCE\_INVALID\_REASON,

            c.concept\_ID as TARGET\_CONCEPT\_ID, c.concept\_name AS TARGET\_CONCEPT\_NAME, c.vocabulary\_id AS TARGET\_VOCABULARY\_ID, c.domain\_id AS TARGET\_DOMAIN\_ID,

                        c.concept\_class\_id AS TARGET\_CONCEPT\_CLASS\_ID, c.INVALID\_REASON AS TARGET\_INVALID\_REASON,

            c.STANDARD\_CONCEPT AS TARGET\_STANDARD\_CONCEPT

       FROM CONCEPT c

       UNION

       SELECT source\_code, SOURCE\_CONCEPT\_ID, SOURCE\_CODE\_DESCRIPTION, source\_vocabulary\_id, c1.domain\_id AS SOURCE\_DOMAIN\_ID, c2.CONCEPT\_CLASS\_ID AS SOURCE\_CONCEPT\_CLASS\_ID,

                                        c1.VALID\_START\_DATE AS SOURCE\_VALID\_START\_DATE, c1.VALID\_END\_DATE AS SOURCE\_VALID\_END\_DATE,stcm.INVALID\_REASON AS SOURCE\_INVALID\_REASON,

                                        target\_concept\_id, c2.CONCEPT\_NAME AS TARGET\_CONCEPT\_NAME, target\_vocabulary\_id, c2.domain\_id AS TARGET\_DOMAIN\_ID, c2.concept\_class\_id AS TARGET\_CONCEPT\_CLASS\_ID,

                     c2.INVALID\_REASON AS TARGET\_INVALID\_REASON, c2.standard\_concept AS TARGET\_STANDARD\_CONCEPT

       FROM source\_to\_concept\_map stcm

              LEFT OUTER JOIN CONCEPT c1

                     ON c1.concept\_id = stcm.source\_concept\_id

              LEFT OUTER JOIN CONCEPT c2

                     ON c2.CONCEPT\_ID = stcm.target\_concept\_id

       WHERE stcm.INVALID\_REASON IS NULL

)

SELECT \*

FROM CTE\_VOCAB\_MAP

/\*EXAMPLE FILTERS\*/

WHERE SOURCE\_VOCABULARY\_ID IN ('ICD9CM')

AND TARGET\_VOCABULARY\_ID IN ('ICD9CM')

### Source to Standard Terminology

WITH CTE\_VOCAB\_MAP AS (

       SELECT c.concept\_code AS SOURCE\_CODE, c.concept\_id AS SOURCE\_CONCEPT\_ID, c.concept\_name AS SOURCE\_CODE\_DESCRIPTION, c.vocabulary\_id AS SOURCE\_VOCABULARY\_ID,

                           c.domain\_id AS SOURCE\_DOMAIN\_ID, c.CONCEPT\_CLASS\_ID AS SOURCE\_CONCEPT\_CLASS\_ID,

                                                   c.VALID\_START\_DATE AS SOURCE\_VALID\_START\_DATE, c.VALID\_END\_DATE AS SOURCE\_VALID\_END\_DATE, c.INVALID\_REASON AS SOURCE\_INVALID\_REASON,

                           c1.concept\_id AS TARGET\_CONCEPT\_ID, c1.concept\_name AS TARGET\_CONCEPT\_NAME, c1.VOCABULARY\_ID AS TARGET\_VOCABUALRY\_ID, c1.domain\_id AS TARGET\_DOMAIN\_ID, c1.concept\_class\_id AS TARGET\_CONCEPT\_CLASS\_ID,

                           c1.INVALID\_REASON AS TARGET\_INVALID\_REASON, c1.standard\_concept AS TARGET\_STANDARD\_CONCEPT

       FROM CONCEPT C

             JOIN CONCEPT\_RELATIONSHIP CR

                        ON C.CONCEPT\_ID = CR.CONCEPT\_ID\_1

                        AND CR.invalid\_reason IS NULL

                        AND cr.relationship\_id = 'Maps To'

              JOIN CONCEPT C1

                        ON CR.CONCEPT\_ID\_2 = C1.CONCEPT\_ID

                        AND C1.INVALID\_REASON IS NULL

       UNION

       SELECT source\_code, SOURCE\_CONCEPT\_ID, SOURCE\_CODE\_DESCRIPTION, source\_vocabulary\_id, c1.domain\_id AS SOURCE\_DOMAIN\_ID, c2.CONCEPT\_CLASS\_ID AS SOURCE\_CONCEPT\_CLASS\_ID,

                                        c1.VALID\_START\_DATE AS SOURCE\_VALID\_START\_DATE, c1.VALID\_END\_DATE AS SOURCE\_VALID\_END\_DATE,

                     stcm.INVALID\_REASON AS SOURCE\_INVALID\_REASON,target\_concept\_id, c2.CONCEPT\_NAME AS TARGET\_CONCEPT\_NAME, target\_vocabulary\_id, c2.domain\_id AS TARGET\_DOMAIN\_ID, c2.concept\_class\_id AS TARGET\_CONCEPT\_CLASS\_ID,

                     c2.INVALID\_REASON AS TARGET\_INVALID\_REASON, c2.standard\_concept AS TARGET\_STANDARD\_CONCEPT

       FROM source\_to\_concept\_map stcm

              LEFT OUTER JOIN CONCEPT c1

                     ON c1.concept\_id = stcm.source\_concept\_id

              LEFT OUTER JOIN CONCEPT c2

                     ON c2.CONCEPT\_ID = stcm.target\_concept\_id

       WHERE stcm.INVALID\_REASON IS NULL

)

SELECT \*

FROM CTE\_VOCAB\_MAP

/\*EXAMPLE FILTERS\*/

WHERE SOURCE\_VOCABULARY\_ID IN ('NDC')

AND TARGET\_VOCABUALRY\_ID IN ('RxNORM')

## Domain IDs

|  |  |
| --- | --- |
| **Domain Mappings** | |
| **domain\_id** | **Maps to table** |
| Unit |  |
| Gender |  |
| Place of Service |  |
| Note Type |  |
| Race |  |
| Meas Value Operator |  |
| Relationship |  |
| Provider Specialty |  |
| Device | Device Exposure |
| Spec Disease Status |  |
| Drug | Drug exposure |
| Specimen |  |
| Route |  |
| Spec Anatomic Site |  |
| Observation | Observation |
| Metadata |  |
| Meas Value |  |
| Measurement | Measurement |
| Procedure | Procedure\_Occurrence |
| Condition | Condition\_Occurrence |