Janssen Research & Development, Pharmaceutical Companies   
of Johnson & Johnson   
Common Data Model (CDM v5.0)

ETL Mapping Specification for SEER Medicare

Amy Matcho, Clair Blacketer, Jennifer Duryea & Jenna Reps

## Table of Contents

[Table of Contents 2](#_Toc437601205)

[1 Introduction 3](#_Toc437601206)

[1.1 Abbreviations 3](#_Toc437601207)

[1.2 Conventions Used in this Document 3](#_Toc437601208)

[2 Data Mapping 4](#_Toc437601209)

[2.0 Sequence Map 4](#_Toc437601210)

[2.1 Table name: LOCATION 4](#_Toc437601211)

[2.2 Table name: PERSON 4](#_Toc437601212)

[2.3 Table name: PAYER\_PLAN\_PERIOD 6](#_Toc437601213)

[2.4 Table name: OBSERVATION\_PERIOD 7](#_Toc437601214)

[2.5 Table name: CARE\_SITE 8](#_Toc437601215)

[2.6 Table name: PROVIDER 9](#_Toc437601216)

[2.7 Table name: VISIT\_OCCURRENCE 10](#_Toc437601217)

[2.8 Table name: CONDITION\_OCCURRENCE 13](#_Toc437601218)

[2.9 Table name: CONDITION\_ERA 18](#_Toc437601219)

[2.10 Table name: PROCEDURE\_OCCURRENCE 18](#_Toc437601220)

[2.11 Table name: PROCEDURE\_COST 22](#_Toc437601221)

[2.12 Table name: DEATH 22](#_Toc437601222)

[2.13 Table name: DRUG\_EXPOSURE 23](#_Toc437601223)

[2.14 Table name: DRUG\_COST 23](#_Toc437601224)

[2.15 Table name: DRUG\_ERA 23](#_Toc437601225)

[2.16 Table name: DEVICE\_EXPOSURE 23](#_Toc437601226)

[Appendix 1: FIPS State Codes 26](#_Toc437601227)

[Appendix 2: SSA State Codes 28](#_Toc437601228)

[Appendix 3: Source to Source Vocab Query 30](#_Toc437601229)

[Appendix 4: Source to Standard Vocab Query 31](#_Toc437601230)

# 1 Introduction

The purpose of this document is to describe the Extract, Transform, Load (ETL) mapping of the licensed SEER Medicare data from the National Cancer Institute into the Observation Health Data Sciences and Informatics (OHDSI) Common Data Model (CDM) V5.

## Abbreviations

|  |  |
| --- | --- |
| **Table 1: Abbreviations** | |
| **Abbreviation** | **Description** |
| ETL | Extract, Transform, Load |
| OHDSI | Observational Health Data Sciences and Informatics |
| CDM | Common Data Model |
| FIPS | Federal Information Processing Standard |
| ICD9 | The International Classification of Diseases, Ninth Revision |
| PEDSF | Patient Entitlement and Diagnosis Summary File |
| CPT | Current Procedural Terminology |
| HCPCS | Healthcare Common Procedure Coding System |
| DRG | Diagnosis-Related Groups |
| MDC | Major Diagnostic Categories |

## Conventions Used in this Document

The following conventions are used within this document:

|  |  |
| --- | --- |
| **Table 1: Abbreviations** | |
| **Convention** | **Description** |
| - | Value does not exist |
| [X] | Value to be replaced or derived |
| *Italicized* | Referring to column in the table itself |

# Data Mapping

This section details how the source files are mapped into the CDM. The SEER pedsf file that contains demographic and registry information was pivoted from 2000+ plus columns into 4 columns: Patient\_ID, Name (name of the former variable), Value, Label (label of the former variable). Any pedsf variable referred to will be located in the Name column and the associated value for that variable will be in the Value column.

## Sequence Map

TBD

## 2.1 Table name: LOCATION

This table is built off of the PEDSF file, which houses the state information based on the Medicare enrollment file in the year of first diagnosis at age 65 or older or the last diagnosis if never 65.

Key conventions:

* The PEDSF file uses the FIPS system to code for state and county values. These FIPS values will need to be mapped to concept IDs
  + A list of state FIPS codes and corresponding state abbreviations are available in [Appendix 1](#_Appendix_1:_FIPS)

| **Table 1: LOCATION** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Logic** | **Comment** |
| LOCATION\_ID | - | System generated |  |
| ADDRESS\_1 | - | NULL |  |
| ADDRESS\_2 | - | NULL |  |
| CITY | - | NULL |  |
| STATE | *LOCATION\_SOURCE\_VALUE* | Convert FIPS state codes to state abbreviations using [Appendix 1](#_Appendix_1:_FIPS) |  |
| ZIP | - | NULL |  |
| COUNTY | - | NULL |  |
| LOCATION\_SOURCE\_VALUE | **PEDSF.**NAME = State |  | Take from the name column in the pivoted pedsf file |

## 2.2 Table name: PERSON

Key Conventions:

* Person demographics are sourced from the pedsf file. From this point on it will be understood that any variable referred to in the pedsf file will acutally be pedsf.NAME = *variable.*
* If a person was born before 1900 or after 2015 then they should be deleted.
* Any person with an unknown gender should be deleted

| **Table 2: PERSON** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Logic** | **Comment** |
| PERSON\_ID | **PEDSF**.PATIENT\_ID | NULL |  |
| YEAR\_OF\_BIRTH | **PEDSF**.BIRTHYR |  | This is the birth year according to Medicare records |
| MONTH\_OF\_BIRTH | **PEDSF.** BIRTHM |  | This is the month of birth according to Medicare records |
| GENDER\_CONCEPT\_ID | *GENDER\_SOURCE\_VALUE* | Map source values to their associated CONCEPT\_IDs:  1=8507  2=8532 |  |
| GENDER\_SOURCE\_VALUE | **PEDSF**.M\_SEX | 1=Male  2=Female |  |
| GENDER\_SOURCE\_CONCEPT\_ID | - | 0 |  |
| RACE\_CONCEPT\_ID | *RACE\_SOURCE\_VALUE* | Map source values to their associated CONCEPT\_IDs:  0=0  1=8527  2=8516  3=0  4=8515  5=0  6=8657 |  |
| RACE\_SOURCE\_VALUE | **PEDSF**.RACE | 0=Unknown  1=White  2=Black  3=Other  4=Asian  5=Hispanic  6=North American Native |  |
| RACE\_SOURCE\_CONCEPT\_ID | - | 0 |  |
| ETHNICITY\_CONCEPT\_ID | *RACE\_SOURCE\_VALUE* | Map source\_values to their associated CONCEPT\_IDs  0=0  1=38003564  2=38003564  3=38003564  4=38003564  5=38003563  6=38003564 |  |
| ETHNICITY\_SOURCE\_VALUE | **PEDSF.**RACE | 0=Unknown  1=White  2=Black  3=Other  4=Asian  5=Hispanic  6=North American Native |  |
| ETHNICITY\_SOURCE\_CONCEPT\_ID | **-** | 0 |  |
| LOCATION\_ID | **PEDSF.**STATE | Map STATE to LOCATION\_SOURCE\_VALUE in Location table then extract its associated LOCATION\_ID | Use [Appendix 1](#_Appendix_1:_FIPS) to map the FIPS codes to state abbreviations and then walk back to LOCATION\_ID |
| PROVIDER\_ID | - | NULL |  |
| DAY\_OF\_BIRTH | - | NULL |  |
| TIME\_OF\_BIRTH | - | NULL |  |
| CARE\_SITE\_ID | - | NULL |  |
| PERSON\_SOURCE\_VALUE | **-** | NULL |  |

## 2.3 Table name: PAYER\_PLAN\_PERIOD

This table will be sourced from the pedsf file. Part A, Part B, Part C (Part A+Part B) indicators for plan\_source\_values will be created from the name=’MON1’1-‘MON252’ fields where MON1 represents January 1991. The plan period should be an era created from values in MON1-MON252. For instance, if all values of MON1-MON252=3 (in the ‘value’ field of pedsf) then the start of the plan era will be January 1, 1991 and the end of the era will be December 31, 2011 and plan\_source\_value=’Part A and Part B’. If MON1-MON4=1 and MON5-MON253=3 then there would be two plan period eras created with 1st era start date = January 1, 1991 and era end date=April 30, 1991 and plan\_source\_value would be ‘Part A Only’. The 2nd era start date would be May 1, 1991-Dec 31, 2011 with plan\_source\_value of ‘Part A and B’. Do not create era for month value=0 or ‘Not entitled’.

Mon1-MonN: 0 = Not entitled, 1 = Part A Only, 2 = Part B Only, 3 = Part A and B

There will also be plan eras created from name=’GHO1’-‘GHO252’ in the pedsf file. A person would not have had HMO coverage if ‘GHO1’=0 during January 2011. Examples of plans eras to create: if all values of GHO1-GHO252=1 then the start of the plan era will be January 1, 1991 and the end of the era will be December 31, 2011 and plan\_source\_value=’HMO member’. If GHO-GHO4=B and GHO5-GHO252=0 then there would be one plan period era created with 1st era start date = January 1, 1991 and era end date=April 30, 1991 and plan\_source\_value would be ‘HMO Member’. Do not create era for month value=0 or ‘Not an HMO Member’.

Gho1=GhoN: 0 = ‘Not an HMO Member’, 1 or 2 or 4 or A or B or C =’ HMO Member’

There will also be plan periods for Part D enrollment which are drug benefits. Name= plan06\_01- plan06\_12 in pedsf file contains an indicator if the patient had Part D coverage in each month of the year 2006. There are values in pedsf following this format for years 2006-2011. If ‘value’ = H, R, S, or E then the patient had Part D coverage for that month. For example if there are only two records of the type for a patient: name=’plan06\_01’=’H’ and name=’plan06\_02’=’H’ then the start of the plan era would be January 1, 2006 and the end of the era would be February 28, 2006. Plan source\_value would be ‘Part D Coverage’.

These specificiations will cause overlapping payer plan periods.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 3: PAYER\_PLAN\_PERIOD** | | | |
| **Destination Field** | **Source Field** | **Logic** | **Comment** |
| PAYER\_PLAN\_PERIOD\_ID | - | System Generated |  |
| PERSON\_ID | PEDSF.PATIENT\_ID | NULL |  |
| PAYER\_PLAN\_PERIOD\_START\_DATE | PEDSF.VALUE FOR NAME=MON1-MON252,GHO1-GHO252,PLAN06\_01-PLAN06\_12, PLAN07\_01-PLAN07\_12, PLAN08\_01-PLAN08\_12, PLAN09\_01-PLAN09\_12, PLAN10\_01-PLAN10\_12, PLAN11\_01-PLAN11\_12, | Minimum start date of a continuous plan era | See logic above |
| PAYER\_PLAN\_END\_DATE | PEDSF.VALUE FOR NAME=MON1-MON252,GHO1-GHO252,PLAN06\_01-PLAN06\_12, PLAN07\_01-PLAN07\_12, PLAN08\_01-PLAN08\_12, PLAN09\_01-PLAN09\_12, PLAN10\_01-PLAN10\_12, PLAN11\_01-PLAN11\_12, | Maximum end date of a continuous plan era | See logic above |
| PAYER\_SOURCE\_VALUE |  | ‘Medicare’ |  |
| PLAN\_SOURCE\_VALUE | PEDSF.VALUE FOR NAME=MON1-MON252,GHO1-GHO252,PLAN06\_01-PLAN06\_12, PLAN07\_01-PLAN07\_12, PLAN08\_01-PLAN08\_12, PLAN09\_01-PLAN09\_12, PLAN10\_01-PLAN10\_12, PLAN11\_01-PLAN11\_12, |  | Name=Mon1-MonN: value=1 = Part A Only, 2 = Part B Only, 3 = Part A and B  Name=Gho1=GhoN:value= 1 or 2 or 4 or A or B or C =’ HMO Member’  Name=plan06\_01- plan06\_12  plan07\_01- plan07\_12  plan08\_01- plan08\_12 plan09\_01- plan09\_12 plan10\_01- plan10\_12 plan11\_01- plan11\_12: ‘value’ = H, R, S, or E‘=’Part D Coverage’ |
| FAMILY\_SOURCE\_VALUE |  | NULL |  |

## 2.4 Table name: OBSERVATION\_PERIOD

This table will be based upon eras derived from payer\_plan\_period table. Eras of enrollment in ‘Part A and B’ coverage and no enrollment in HMO coverage (lack of ‘HMO member’ era for same time period) will be created from payor\_plan\_periods. For example, if a patient had payor\_plan\_period era of ‘Part A and B’ coverage from 1/1/2000 to 12/31/2013 and no ‘HMO member’ eras during that same time period then their observation period would be 1/1/2000 to 12/31/2013. If the same patient had an ‘HMO member’ era from 1/1/2005-12/31/2013 then their observation period would be 1/1/2000-12/31/2004.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 4: OBSERVATION\_PERIOD** | | | |
| **Destination Field** | **Source Field** | **Logic** | **Comment** |
| OBSERVATION\_PERIOD\_ID | - | System Generated |  |
| PERSON\_ID | PEDSF.PATIENT\_ID | NULL |  |
| OBSERVATION\_PERIOD\_START\_DATE | PEDSF.VALUE FOR NAME=MON1-MON252,GHO1-GHO252 | ‘Enrollment in Part A and B and no HMO coverage’ era start from payer\_plan\_period | See logic above |
| OBSERVATION\_PERIOD\_END\_DATE | PEDSF.VALUE FOR NAME=MON1-MON252,GHO1-GHO252 | ‘Enrollment in Part A and B and no HMO coverage’ era end from payer\_plan\_period | See logic above |
| PERIOD\_TYPE\_CONCEPT\_ID |  | ‘44814725’ ‘Period inferred by algorithm’ |  |

## 2.5 Table name: CARE\_SITE

This table will be sourced from SEER tables MEDPAR, OUTSAF

Key conventions:

1. Use the first two digits of the provider id as the SSA state code, then map the state codes to the state abbreviation and then map to the correction LOCATION\_ID
   1. If the first two digits do not map to an SSA state code, set LOCATION\_ID = 0
2. If a provider is in both the MEDPAR and OUTSAF tables set PLACE\_OF\_SERVICE\_CONCEPT\_ID = 8717 (inpatient hospital)

Note: This table uses SSA state codes to define location. These codes are listed in [Appendix 2](#_Appendix_2:_SSA)

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 5: CARE\_SITE** | | | |
| **Destination Field** | **Source Field** | **Logic** | **Comment** |
| CARE\_SITE\_ID | - | System Generated |  |
| CARE\_SITE\_NAME | - | NULL | The name is not listed in the files, only the Medicare provider ID |
| PLACE\_OF\_SERVICE\_CONCEPT\_ID |  | Provider ID from MEDPAR = 8717  Provider ID from OUTPAT = 8756 | This field is derived from which table the provider id is taken from. If a provider is in both MEDPAR and OUTSAF, set PLACE\_OF\_SERVICE\_CONCEPT\_ID = 8717 |
| LOCATION\_ID | **MEDPAR**.PROVIDER  **OUTPAT**.PROVIDER  (first two digits) | Convert SSA state codes to state abbreviations and then map to associated LOCATION\_ID | Use the first two digits of the PROVIDER field as the SSA state code. |
| CARE\_SITE\_SOURCE\_VALUE | **MEDPAR**.PROVIDER  **OUTPAT**.PROVIDER |  |  |
| PLACE\_OF\_SERVICE\_SOURCE\_VALUE |  | Provider ID from MEDPAR = “Inpatient Hospital”  Provider ID from OUTPAT = “Outpatient Hospital” | If a provider id is in both MEDPAR and OUTSAF, set this as “Inpatient Hospital” |

## 2.6 Table name: PROVIDER

This table will be sourced from SEER tables OUTSAF, NCH and DME

Key Concepts:

Each provider of interest in the three tables: ordering (ORD), performing (PRF) or attending (AT) has both a national provider identification number (NPI) and a unique physician identification number (UPIN). We are most interested in the NPI but if that is blank then look to the UPIN for identification of the provider. Each provider has multiple specialties that we want to preserve so to build this table it is essentially a distinct listing of the each provider ID with all their associated specialties.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 6: PROVIDER** | | | |
| **Destination Field** | **Source Field** | **Logic** | **Comment** |
| PROVIDER\_ID | - | System Generated | Auto generate for each unique provider\_source\_value |
| PROVIDER\_NAME | - | NULL |  |
| NPI | **DME**.ORD\_NPI  **NCH**.PRF\_NPI,  **OUTPAT**.AT\_NPI |  | Refer to logic for PROVIDER\_SOURCE\_VALUE. If the NPI number is used to identify the provider then put that number here, otherwise NULL |
| DEA | - | NULL |  |
| SPECIALITY\_CONCEPT\_ID | *SPECIALITY\_SOURCE\_VALUE* | OUTPAT=NULL  Map source\_values to their associated TARGET\_CONCEPT\_IDs using the vocab query in [Appendix 4](#_Appendix_4:_Source) | Use filters:  WHERE SOURCE\_VOCABULARY\_ID IN ('Specialty')  AND TARGET\_VOCABUALRY\_ID IN ('Specialty')  NOTE: You will need to strip the leading zeros from the source values before mapping  Set SPECIALTY\_CONCEPT\_ID as 38004514 (Unknown Physician Specialty) if the SPECIALTY\_SOURCE\_VALUE is NULL or cannot be mapped |
| CARE\_SITE\_ID | **OUTPAT**.PROVIDER | DME, NCH=NULL  Map PROVIDER to CARE\_SITE\_SOURCE\_VALUE in CARE\_SITE table then extract its associated CARE\_SITE\_ID |  |
| YEAR\_OF\_BIRTH | - | NULL |  |
| GENDER\_CONCEPT\_ID | - | 0 |  |
| PROVIDER\_SOURCE\_VALUE | **NCH**.PRF\_NPI or PERUPIN,  **DME**. ORD\_NPI or ORD\_UPIN,  **OUTSAF**.AT\_NPI or AT\_UPIN | * If **NCH.**PRF\_NPI is NULL then use **NCH.**PERUPIN * If **DME.** ORD\_NPI is NULL then use **DME.** ORD\_UPIN * If **OUTSAF**.AT\_NPI is NULL then use **OUTSAF**.AT\_UPIN |  |
| SPECIALITY\_SOURCE\_VALUE | **DME**.HCFASPEC,  **NCH**.HCFASPEC,  **OUTSAF** = NULL |  |  |
| SPECIALITY\_SOURCE\_CONCEPT\_ID | - | 0 |  |
| GENDER\_SOURCE\_VALUE | - | NULL |  |
| GENDER\_SOURCE\_CONCEPT\_ID | - | 0 |  |

## **2.7** Table name: VISIT\_OCCURRENCE

This table will be sourced from SEER tables MEDPAR, OUTSAF, NCH and DME

Key Conventions:

One visit can be made up of claims from multiple tables. The hierarchy is as follows: MEDPAR > OUTSAF > NCH, DME. A unique visit is defined by: a unique claim (link\_num DME and NCH), one record in MEDPAR or claim lines with the same link\_num and start and end dates in OUTSAF. When evaluating OUTSAF claims as to whether they should be part of an IP visit or if it should become a stand-alone OP visit, disregard any claim lines with CENTER = 0001. This is a summary line for cost purposes. When evaluating NCH and DME claims any claim line with PMTDNLCD = 0 should be deleted as this indicates the charge was denied.

* Start by identifying inpatient visits:One record in MEDPAR should become one visit; if a visit does not have a discharge date (DIS\_M, DIS\_D, DIS\_Y are blank) then delete this record
  + MEDPAR claims are either SNF or IP. If SNFIND = ‘N’ then this becomes a SNF visit with VISIT\_CONCEPT\_ID =42898160 else if SNFIND is other than ‘N’ then this becomes an inpatient visit VISIT\_CONCEPT\_ID = 9201.
    - After identifying SNF visits (SNFIND=’N’), if two visits have the same value for PROVIDER and the discharge date (DIS\_D, DIS\_M, DIS\_Y) of one is within two days of the admit date (ADM\_D, ADM\_M, ADM\_Y) of another then collapse the two records into one visit.
  + If an OUTSAF claim line start date (CENDD, CENDM, CENDY or FROM\_DTD, FROM\_DTM, FROM\_DTY) or end date (THRU\_DTD, THRU\_DTM, THRU\_DTY) is entirely within MEDPAR admission and discharge dates, the claim line should become part of that MEDPAR VISIT\_OCCURRENCE . If an OUTSAF claim line starts within a MEDPAR visits but ends after the MEDPAR visit then it should become its own visit. The same should be done if an OUTSAF claim starts before a MEDPAR visit but ends within it.
  + If an NCH or DME claim line has a start and end date entirely within a MEDPAR visit then it should become part of the MEDPAR VISIT\_OCCURRENCE under these conditions:
    - After identifying that an NCH or DME claim line should be rolled into a MEDPAR visit, if the MEDPAR claim has SNFIND = ‘N’ it is a SNF visit and the NCH or DME variable PLCSRVC must be in (4,31)
    - If the MEDPAR visit has SNFIND other than ‘N’ the NCH or DME variable PLCSRVC must be in (3,21,51,52,55,56,61,65) to be rolled into the MEDPAR VISIT\_OCCURRENCE
* After identifying OUTSAF claim lines that should be rolled into MEDPAR visits the remaining claims should become stand-alone visits.
  + Claim lines with CENTER <> ‘0001’ that have the same link\_num and the same start and end dates should become one visit.
  + After identifying the stand-alone OUTSAF visits, apply the following logic to determine if it is an ER visit and should be given VISIT\_CONCEPT\_ID = 9203:
  + If any claim line in the visit has a revenue center code (CENTER) between ‘0450’ and ‘0459’ or CENTER = ‘0981’
  + or if any claim line in the visit has a HCPCS (HCPCS) code of ’99281’, ‘99282’, ‘99283’, ‘99284’, ‘99285’
  + If the above conditions for an ER visit are met and an NCH or DME claim line has a start and end date entirely within the OUTSAF claim start and end dates then the NCH or DME claim line should be rolled into the OUTSAF visit only if the NCH or DME variable PLCSRVC is in (23, 41, 42)
* If an OUTSAF visit does not meet the above criteria for an ER visit then it should be given VISIT\_CONCEPT\_ID = 9202 (Outpatient visit).
  + - If an OUTSAF visit is given VISIT\_CONCEPT\_ID = 9202 and an NCH claim line or DME claim line start and end dates are entirely contained within the OUTSAF visit start and end dates then the NCH/DME claim should be rolled into the OUTSAF claim only if the following conditions are met:
      * The NCH or DME claim line has PLCSRVC in (5,8,9,22,26,53,54,60,62,71,72)
  + If an NCH or DME claim line has a start and end date entirely within a OUTSAF claim but it does not meet the above conditions to be rolled up then apply the following logic:
    - If the variable PLCSRVC is in (23, 41, 42) then this should become a separate visit with VISIT\_CONCEPT\_ID = 9203. Otherwise they will become their own visits using the below logic:
* This should now leave NCH claims and DME claims that do not roll into MEDPAR or OUTSAF visits. Each unique claim should be its own visit with VISIT\_CONCEPT\_ID = 9202.
* If any visit occurs before the patient’s date of birth then that visit and all corresponding records (condition, procedure, drug\_exposure, device\_exposure, measurement, observation) should be deleted
* If any visit occurs >= 30 days after the patient’s death then that visit and all corresponding records (condition, procedure, drug\_exposure, device\_exposure, measurement, observation) should be deleted

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 7: VISIT\_OCCURRENCE** | | | |
| **Destination Field** | **Source Field** | **Logic** | **Comment** |
| VISIT\_OCCURRENCE\_ID |  | System generated |  |
| PERSON\_ID | Patient\_ID |  |  |
| VISIT\_CONCEPT\_ID |  | Refer to above logic above for how to assign this variable |  |
| VISIT\_START\_DATE | **MEDPAR**.ADM\_D, ADM\_M, ADM\_Y  **NCH**.FREXPENM, FREXPEND, FREXPENY  **OUTSAF**.CENM, CEND, CENY OR FROM\_DTM, FROM\_DTD or FROM\_DTY  **DME**.FREXPENM, FREXPEND, FREXPENY | An OUTSAF visit is defined as claim lines having the same LINK\_NUM and start and end dates. If CENDM, CENDD, and CENDY are <> ‘00’ then use them as the visit start and end date. If they are equal to ‘00’ then use FROM\_DTM, FROM\_DTD and FROM\_DTY as VISIT\_START\_DATE |  |
| VISIT\_START\_TIME |  | NULL |  |
| VISIT\_END\_DATE | **MEDPAR**.DIS\_D, DIS\_M, DIS\_Y  **NCH**.LSEXPENM, LSEXPEND, LSEXPENY  **OUTSAF.**CEND, CENY OR THRU\_DTM, THRU\_DTD, THRU\_DTY  **DME**. LSEXPENM, LSEXPEND, LSEXPENY | An OUTSAF visit is defined as claim lines having the same LINK\_NUM and start and end dates. If CENDM, CENDD, and CENDY are <> ‘00’ then use them as the start and end date. If they are equal to ‘00’ then use THRU\_DTM, THRU\_DTD, THRU\_DTY as the VISIT\_END\_DATE |  |
| VISIT\_END\_TIME | - | NULL |  |
| VISIT\_TYPE\_CONCEPT\_ID |  |  | All visits should have a VISIT\_TYPE\_CONCEPT\_ID = 44818518 |
| PROVIDER\_ID | **NCH**.PRF\_NPI or PERUPIN,  **DME**. ORD\_NPI or ORD\_UPIN,  **OUTSAF**.AT\_NPI or AT\_UPIN | * If **NCH.**PRF\_NPI is NULL then use **NCH.**PERUPIN * If **DME.** ORD\_NPI is NULL then use **DME.** ORD\_UPIN * If **OUTSAF**.AT\_NPI is NULL then use **OUTSAF**.AT\_UPIN |  |
| CARE\_SITE\_ID | **MEDPAR**, **OUTSAF**.PROVIDER |  | Use source code to lookup care\_site\_id |
| VISIT\_SOURCE\_VALUE | **-** | NULL |  |
| VISIT\_SOURCE\_CONCEPT\_ID | - | 0 |  |

## 2.8 Table name: CONDITION\_OCCURRENCE

Key conventions

* This table will be sourced from MEDPAR, OUTSAF, NCH and DME. Refer to the logic in the VISIT\_OCCURRENCE table for defining visits; it will be used for assigning VISIT\_OCCURRENCE\_ID. Any code in a diagnosis field, HCPCS field or surgical field that maps to a concept with DOMAIN\_ID =’ Condition’ should go in this table.
* If any condition occurs >= 30 days after death then that record should be deleted

|  |  |  |  |
| --- | --- | --- | --- |
| **Source Table** | **Position/ Source Field** | **CONDITION\_ TYPE\_CONCEPT\_ID** | **CONCEPT\_NAME** |
| MEDPAR | DGN\_CD1 | 38000200 | Inpatient header – 1st position |
| DGN\_CD2 | 38000201 | Inpatient header – 2nd position |
| DGN\_CD3 | 38000202 | Inpatient header – 3rd position |
| DGN\_CD4 | 38000203 | Inpatient header – 4th position |
| DGN\_CD5 | 38000204 | Inpatient header – 5th position |
| DGN\_CD6 | 38000205 | Inpatient header – 6th position |
| DGN\_CD7 | 38000206 | Inpatient header – 7th position |
| DGN\_CD8 | 38000207 | Inpatient header – 8th position |
| DGN\_CD9 | 38000208 | Inpatient header – 9th position |
| DGN\_CD10 | 38000209 | Inpatient header – 10th position |
| DGN\_CD11 | 38000210 | Inpatient header – 11th position |
| DGN\_CD12 | 38000211 | Inpatient header – 12th position |
| DGN\_CD13 | 38000212 | Inpatient header – 13th position |
| DGN\_CD14 | 38000213 | Inpatient header – 14th position |
| DGN\_CD15-DGN\_CD25 | 38000214 | Inpatient header – 15th position |
| OUTSAF | DGN\_CD1 or E1DGNSCD | 38000230 | Outpatient header – 1st position |
| DGN\_CD2 or  EDGNSD1 | 38000231 | Outpatient header – 2nd position |
| DGN\_CD3 or EDGNSD2 | 38000232 | Outpatient header – 3rd position |
| DGN\_CD4 or EDGNSD3 | 38000233 | Outpatient header – 4th position |
| DGN\_CD5 or EDGNSD4 | 38000234 | Outpatient header – 5th position |
| DGN\_CD6 or EDGNSD5 | 38000235 | Outpatient header – 6th position |
| DGN\_CD7 or EDGNSD6 | 38000236 | Outpatient header – 7th position |
| DGN\_CD8 | 38000237 | Outpatient header – 8th position |
| DGN\_CD9 | 38000238 | Outpatient header – 9th position |
| DGN\_CD10 | 38000239 | Outpatient header – 10th position |
| DGN\_CD11 | 38000240 | Outpatient header – 11th position |
| DGN\_CD12 | 38000241 | Outpatient header – 12th position |
| DGN\_CD13 | 38000242 | Outpatient header – 13th position |
| DGN\_CD14 | 38000243 | Outpatient header – 14th position |
| DGN\_CD15-DGN\_CD25 | 38000244 | Outpatient header – 15th position |
| NCH | LINEDIAG | 45756843 | Carrier claim detail – 1st position |
| DME | LINEDIAG | 38000215 | Outpatient detail – 1st position |
| NCH | DGN\_CD1 | 45756835 | Carrier claim header – 1st position |
| DGN\_CD2 | 45756836 | Carrier claim header – 2nd position |
| DGN\_CD3 | 45756837 | Carrier claim header – 3rd position |
| DGN\_CD4 | 45756838 | Carrier claim header – 4th position |
| DGN\_CD5 | 45756839 | Carrier claim header – 5th position |
| DGN\_CD6 | 45756840 | Carrier claim header – 6th position |
| DGN\_CD7 | 45756841 | Carrier claim header – 7th position |
| DGN\_CD8-DGN\_CD12 | 45756842 | Carrier claim header – 8th position |
| DME | DGN\_CD1 | 38000230 | Outpatient header – 1st position |
| DGN\_CD2 | 38000231 | Outpatient header – 2nd position |
| DGN\_CD3 | 38000232 | Outpatient header – 3rd position |
| DGN\_CD4 | 38000233 | Outpatient header – 4th position |
| DGN\_CD5 | 38000234 | Outpatient header – 5th position |
| DGN\_CD6 | 38000235 | Outpatient header – 6th position |
| DGN\_CD7 | 38000236 | Outpatient header – 7th position |
| DGN\_CD8 | 38000237 | Outpatient header – 8th position |
| DGN\_CD9 | 38000238 | Outpatient header – 9th position |
| DGN\_CD10 | 38000239 | Outpatient header – 10th position |
| DGN\_CD11 | 38000240 | Outpatient header – 11th position |
| DGN\_CD12 | 38000241 | Outpatient header – 12th position |
| MEDPAR | 1 (SRGCDE1) | 38000184 | Inpatient detail - 1st position |
| 2 (SRGCDE 2) | 38000185 | Inpatient detail - 2nd position |
| 3 (SRGCDE 3) | 38000186 | Inpatient detail - 3rd position |
| 4 (SRGCDE 4) | 38000187 | Inpatient detail - 4th position |
| 5 (SRGCDE 5) | 38000188 | Inpatient detail - 5th position |
| 6 (SRGCDE 6) | 38000189 | Inpatient detail - 6th position |
| 7 (SRGCDE 7) | 38000190 | Inpatient detail - 7th position |
| 8 (SRGCDE 8) | 38000191 | Inpatient detail - 8th position |
| 9 (SRGCDE 9) | 38000192 | Inpatient detail - 9th position |
| 10 (SRGCDE 10) | 38000193 | Inpatient detail - 10th position |
| 11 (SRGCDE 11) | 38000194 | Inpatient detail - 11th position |
| 12 (SRGCDE 12) | 38000195 | Inpatient detail - 12th position |
| 13 (SRGCDE 13) | 38000196 | Inpatient detail - 13th position |
| 14 (SRGCDE 14) | 38000197 | Inpatient detail - 14th position |
| 15(SRGCDE 15) | 38000198 | Inpatient detail - 15th position |
| 16(SRGCDE16) | 44818709 | Inpatient detail - 16th position |
| 17(SRGCDE17) | 44818710 | Inpatient detail - 17th position |
| 18(SRGCDE18) | 44818711 | Inpatient detail - 18th position |
| 19(SRGCDE19) | 44818712 | Inpatient detail - 19th position |
| 20(SRGCDE20-SRGCDE25) | 44818713 | Inpatient detail – 20th position |
| OUTSAF | HCPCS | 38000215 | Outpatient detail – 1st position |
| NCH | HCPCS | 45756843 | Carrier claim detail – 1st position |
| DME | HCPCS | 38000215 | Outpatient detail - 1st position |

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 11: CONDITION\_OCCURRENCE** | | | |
| **Destination Field** | **Source Field** | **Logic** | **Comment** |
| CONDITION\_OCCURRENCE\_ID | - | System Generated | Auto generate for each unique provider\_source\_value |
| PERSON\_ID | **MEPDAR.**Patient\_ID  **OUTSAF.**Patient\_ID  **NCH.**Patient\_ID  **DME.**Patient\_ID |  |  |
| CONDITION\_CONCEPT\_ID | Surgical Fields:  **MEDPAR.**SRGCDE1-SRGCDE25  HCPCS fields:  **OUTSAF.**HCPCS  **NCH.**HCPCS  **DME.**HCPCS  Diagnosis fields:  **MEDPAR.**DGN\_CD1-DGN\_CD25  **OUTSAF.** DGN\_CD1-DGN\_CD25  **OUTSAF.**E1DGNSCD  **OUTSAF.**EDGNSD1-EDGNSD6  **NCH.**LINEDIAG  **NCH.**DGN\_CD1-DGN\_CD12  **DME.**LINEDIAG  **DME.**DGN\_CD1-DGN\_CD12 | Map source\_values to their associated TARGET\_CONCEPT\_IDs using the vocab query in [Appendix 4](#_Appendix_4:_Source) | For codes from surgical and HCPCS fields use 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')  For codes from diagnosis fields use filters:  WHERE SOURCE\_VOCABULARY\_ID IN ('ICD9CM')  AND TARGET\_VOCABUALRY\_ID IN ('SNOMED') |
| CONDITION\_START\_DATE | Surgical Fields:  **MEDPAR.**  SG\_DT1-SG\_DT25  HCPCS Fields:  **NCH**.FREXPENM, FREXPEND, FREXPENY  **OUTSAF.**CENM, CEND, CENY OR FROM\_DTM, FROM\_DTD, FROM\_DTY  **DME**. FROM\_DTM, FROM\_DTD, FROM\_DTY  Diagnosis fields:  **MEDPAR**.ADM\_D, ADM\_M, ADM\_Y  **NCH**.FREXPENM, FREXPEND, FREXPENY OR FROM\_DTM, FROM\_DTD, FROM\_DTY  **OUTSAF.** FROM\_DTM, FROM\_DTD, FROM\_DTY  **DME**. FREXPENM, FREXPEND, FREXPENY OR FROM\_DTM, FROM\_DTD, FROM\_DTY | For a HCPCS field in the **OUTSAF** file, if CENM, CEND, CENY = ‘00’ then use FROM\_DTM, FROM\_DTD, FROM\_DTY as the condition start date  For a diagnosis code coming from OUTSAF (DGN\_CD1-DGN\_CD25) then use FROM\_DTM, FROM\_DTD, FROM\_DTY as the condition start date  For diagnosis codes coming from NCH or DME, if the code is **NCH.**LINEDIAG or **DME.**LINEDIAG then use FREXPENM, FREXPEND, FREXPENY, otherwise for DGN\_CD1-DGN\_CD12 use FROM\_DTM,FROM\_DTD, FROM\_DTY |  |
| CONDITION\_END\_DATE | **Surgical Fields:**  **MEDPAR.**  SG\_DT1-SG\_DT25  **HCPCS Fields:**  **NCH**.LSEXPENM, LSEXPEND, LSEXPENY  **OUTSAF.**CENM, CEND, CENY OR THRU\_DTM, THRU\_DTD, THRU\_DTY  **DME**. THRU\_DTM, THRU\_DTD, THRU\_DTY  **Diagnosis fields:**  **MEDPAR**.DIS\_D, DIS\_M, DIS\_Y  **NCH**.LSEXPENM, LSEXPEND, LSEXPENY OR THRU\_DTM, THRU\_DTD, THRU\_DTY  **OUTSAF.** THRU\_DTM, THRU\_DTD, THRU\_DTY  **DME**. LSEXPENM, LSEXPEND, LSEXPENY OR THRU\_DTM, THRU\_DTD, THRU\_DTY | For a HCPCS field in the **OUTSAF** file, if CENM, CEND, CENY = ‘00’ then use THRU\_DTM, THRU\_DTD, THRU\_DTY as the device end date  For a diagnosis code coming from OUTSAF (DGN\_CD1-DGN\_CD25) then use THRU\_DTM, THRU\_DTD, THRU\_DTY as the condition end date  For diagnosis codes coming from NCH or DME, if the code is **NCH.**LINEDIAG or **DME.**LINEDIAG then use LSEXPENM, LSEXPEND, LSEXPENY, otherwise use THRU\_DTM, THRU\_DTD, THRU\_DTY |  |
| CONDITION\_TYPE\_CONCEPT\_ID |  | Refer to above tables and documentation to map this field |  |
| STOP\_REASON |  |  |  |
| PROVIDER\_ID | **NCH**.PRF\_NPI or PERUPIN  **NCH**.HCFASPEC  **DME**. ORD\_NPI or ORD\_UPIN  **DME**.HCFASPEC  **OUTSAF**.AT\_NPI or AT\_UPIN | * If **NCH.**PRF\_NPI is NULL then use **NCH.**PERUPIN * If **DME.** ORD\_NPI is NULL then use **DME.** ORD\_UPIN * If **OUTSAF**.AT\_NPI is NULL then use **OUTSAF**.AT\_UPIN   When procedure is in DME or NCH use both UPIN/NPI and HCFASPEC to map to the correct PROVIDER\_ID | Map these values back to PROVIDER\_ID using the PROVIDER table |
| VISIT\_OCCURRENCE\_ID | **VISIT\_OCCURRENCE:** VISIT\_OCCURRENCE\_ID | Refer to logic in table 2.6 to assign this value |  |
| CONDITION\_SOURCE\_VALUE | Surgical Fields:  **MEDPAR.**SRGCDE1-SRGCDE25  HCPCS fields:  **OUTSAF.**HCPCS  **NCH.**HCPCS  **DME.**HCPCS  Diagnosis fields:  **MEDPAR.**DGN\_CD1-DGN\_CD25  **OUTSAF.** DGN\_CD1-DGN\_CD25  **OUTSAF.**E1DGNSCD  **OUTSAF.**EDGNSD1-EDGNSD6  **NCH.**LINEDIAG  **NCH.**DGN\_CD1-DGN\_CD12  **DME.**LINEDIAG  **DME.**DGN\_CD1-DGN\_CD12 |  |  |
| CONDITION\_SOURCE\_CONCEPT\_ID | Surgical Fields:  **MEDPAR.**SRGCDE1-SRGCDE25  HCPCS fields:  **OUTSAF.**HCPCS  **NCH.**HCPCS  **DME.**HCPCS  Diagnosis fields:  **MEDPAR.**DGN\_CD1-DGN\_CD25  **OUTSAF.** DGN\_CD1-DGN\_CD25  **OUTSAF.**E1DGNSCD  **OUTSAF.**EDGNSD1-EDGNSD6  **NCH.**LINEDIAG  **NCH.**DGN\_CD1-DGN\_CD12  **DME.**LINEDIAG  **DME.**DGN\_CD1-DGN\_CD12 | Map source\_values to their associated SOURCE\_CONCEPT\_ID using the vocab query in [Appendix 3](#_Appendix_3:_Source) | From surgical or HCPCS fields:  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')  From diagnosis fields:  WHERE SOURCE\_VOCABULARY\_ID IN ('ICD9CM')  AND TARGET\_VOCABUALRY\_ID IN ('ICD9CM') |

## 2.9 Table name: CONDITION\_ERA

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 9: 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. |  |

## 2.10 Table name: PROCEDURE\_OCCURRENCE

Key Conventions:

* This table will be sourced from MEDPAR, OUTSAF, NCH and DME. Refer to the logic in the VISIT\_OCCURRENCE table for defining visits; it will be used for assigning VISIT\_OCCURRENCE\_ID. Any code in a diagnosis field, HCPCS field or surgical field that maps to a concept with DOMAIN\_ID =’ Procedure’ should go in this table.
* If a code in a HCPCS field from OUTSAF or NCH is mapped to a standard concept with a domain other than procedure a record will be written to the corresponding table (measurement, device, etc.), however, a PROCEDURE\_OCCURRENCE record is also written with PROCEDURE\_CONCEPT\_ID = 0. This will be used later when mapping PROCEDURE\_COST records.
* If any procedure occurs >= 30 days after the patient’s death then that record should be deleted

After assigning VISIT\_OCCURRENCE\_ID, use this logic to map PROCEDURE\_TYPE\_CONCEPT\_ID:

|  |  |  |  |
| --- | --- | --- | --- |
| **Source Table** | **Position/ Source Field** | **PROCEDURE\_ TYPE\_CONCEPT\_ID** | **CONCEPT\_NAME** |
| MEDPAR and OUTSAF | DGN\_CD1-DGN\_CD25 | 42865906 | Condition Procedure |
| NCH and DME | LINEDIAG or DGN\_CD1-DGN\_CD12 | 42865906 | Condition Procedure |
| MEDPAR | 1 (SRGCDE1) | 38000251 | Inpatient header - 1st position |
| 2 (SRGCDE 2) | 38000252 | Inpatient header - 2nd position |
| 3 (SRGCDE 3) | 38000253 | Inpatient header - 3rd position |
| 4 (SRGCDE 4) | 38000254 | Inpatient header - 4th position |
| 5 (SRGCDE 5) | 38000255 | Inpatient header - 5th position |
| 6 (SRGCDE 6) | 38000256 | Inpatient header - 6th position |
| 7 (SRGCDE 7) | 38000257 | Inpatient header - 7th position |
| 8 (SRGCDE 8) | 38000258 | Inpatient header - 8th position |
| 9 (SRGCDE 9) | 38000259 | Inpatient header - 9th position |
| 10 (SRGCDE 10) | 38000260 | Inpatient header - 10th position |
| 11 (SRGCDE 11) | 38000261 | Inpatient header - 11th position |
| 12 (SRGCDE 12) | 38000262 | Inpatient header - 12th position |
| 13 (SRGCDE 13) | 38000263 | Inpatient header - 13th position |
| 14 (SRGCDE 14) | 38000264 | Inpatient header - 14th position |
| 15(SRGCDE 15- SRGCDE25) | 38000265 | Inpatient header - 15th position |
| OUTSAF | HCPCS | 38000215 | Outpatient detail – 1st position |
| NCH | HCPCS | 45756843 | Carrier detail – 1st position |
| DME | HCPCS | 38000215 | Outpatient detail – 1st position |
| OUTSAF | DGN\_CD1-DGN\_CD25, E1DGNSCD, EDGNSD1-EDGNSD6 | 42865906 | Condition Procedure |
| NCH and DME | LINEDIAG or DGN\_CD1-DGN\_CD12 | 42865906 | Condition Procedure |

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 8: PROCEDURE\_OCCURRENCE** | | | |
| **Destination Field** | **Source Field** | **Logic** | **Comment** |
| PROCEDURE\_OCCURRENCE\_ID | - | System Generated | Auto generate for each unique provider\_source\_value |
| PERSON\_ID | **MEPDAR.**Patient\_ID  **OUTSAF.**Patient\_ID  **NCH.**Patient\_ID  **DME.**Patient\_ID |  |  |
| PROCEDURE\_CONCEPT\_ID | Surgical Fields:  **MEDPAR.**SRGCDE1-SRGCDE25  HCPCS fields:  **OUTSAF.**HCPCS  **NCH.**HCPCS  **DME.**HCPCS  Diagnosis fields:  **MEDPAR.**DGN\_CD1-DGN\_CD25  **OUTSAF.** DGN\_CD1-DGN\_CD25  **OUTSAF.**E1DGNSCD  **OUTSAF.**EDGNSD1-EDGNSD6  **NCH.**LINEDIAG  **NCH.**DGN\_CD1-DGN\_CD12  **DME.**LINEDIAG  **DME.**DGN\_CD1-DGN\_CD12 | Map source\_values to their associated TARGET\_CONCEPT\_IDs using the vocab query in [Appendix 4](#_Appendix_4:_Source) | For codes from HCPCS or Surgical fields use 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')  For codes from diagnosis fields use filters:  WHERE SOURCE\_VOCABULARY\_ID IN ('ICD9CM')  AND TARGET\_VOCABUALRY\_ID IN ('SNOMED') |
| PROCEDURE\_DATE | Surgical Fields:  **MEDPAR.**  SG\_DT1-SG\_DT25  HCPCS Fields:  **NCH**.FREXPENM, FREXPEND, FREXPENY  **OUTSAF.**CENM, CEND, CENY OR FROM\_DTM, FROM\_DTD, FROM\_DTY  **DME**. FROM\_DTM, FROM\_DTD, FROM\_DTY  Diagnosis fields:  **MEDPAR**.ADM\_D, ADM\_M, ADM\_Y  **NCH**.FREXPENM, FREXPEND, FREXPENY OR FROM\_DTM, FROM\_DTD, FROM\_DTY  **OUTSAF.** FROM\_DTM, FROM\_DTD, FROM\_DTY  **DME**. FREXPENM, FREXPEND, FREXPENY OR FROM\_DTM, FROM\_DTD, FROM\_DTY | For a HCPCS field in the **OUTSAF** file, if CENM, CEND, CENY = ‘00’ then use FROM\_DTM, FROM\_DTD, FROM\_DTY as the procedure date  For a diagnosis code coming from OUTSAF (DGN\_CD1-DGN\_CD25) then use FROM\_DTM, FROM\_DTD, FROM\_DTY as the procedure date  For diagnosis codes coming from NCH or DME, if the code is **NCH.**LINEDIAG or **DME.**LINEDIAG then use FREXPENM, FREXPEND, FREXPENY, otherwise for DGN\_CD1-DGN\_CD12 use FROM\_DTM,FROM\_DTD, FROM\_DTY |  |
| PROCEDURE\_TYPE\_CONCEPT\_ID |  | Refer to above tables and documentation to map this field |  |
| MODIFIER\_CONCEPT\_ID | **OUTSAF.**MF1  **NCH.**MF1  **DME.**MF1 | Modifiers only exist on records where the procedure code comes from the HCPCS field.  Map source\_values to their associated TARGET\_CONCEPT\_IDs using the vocab query in [Appendix 4](#_Appendix_4:_Source)  If there is no value for this field then set to 0 | Use the filter:  WHERE SOURCE\_VOCABULARY\_ID IN ('HCPCS', 'CPT4')  AND TARGET\_VOCABUALRY\_ID IN ('HCPCS', 'CPT4')  and TARGET\_CONCEPT\_CLASS\_ID in ('CPT4 Modifier', 'HCPCS Modifier') |
| QUANTITY | **OUTSAF.**UNIT  **NCH.**MTUSCNT  **DME.**MTUSCNT | NULL | Note that about 0.03% of the time the MTUSCNT comes across as a FLOAT. We have rounded this value to an INT. |
| PROVIDER\_ID | **NCH**.PRF\_NPI or PERUPIN  **NCH**.HCFASPEC  **DME**.ORD\_NPI or ORD\_UPIN  **DME**.HCFASPEC  **OUTSAF**.AT\_NPI or AT\_UPIN | * If **NCH.**PRF\_NPI is NULL then use **NCH.**PERUPIN * If **DME.** ORD\_NPI is NULL then use **DME.** ORD\_UPIN * If **OUTSAF**.AT\_NPI is NULL then use **OUTSAF**.AT\_UPIN   When procedure is in DME or NCH use both UPIN/NPI and HCFASPEC to map to the correct PROVIDER\_ID | Map these values back to PROVIDER\_ID using the PROVIDER table |
| VISIT\_OCCURRENCE\_ID | **VISIT\_OCCURRENCE:** VISIT\_OCCURRENCE\_ID | Refer to logic in table 2.6 to assign this value |  |
| PROCEDURE\_SOURCE\_VALUE | Surgical Fields:  **MEDPAR.**SRGCDE1-SRGCDE25  HCPCS fields:  **OUTSAF.**HCPCS  **NCH.**HCPCS  **DME.**HCPCS  Diagnosis fields:  **MEDPAR.**DGN\_CD1-DGN\_CD25  **OUTSAF.** DGN\_CD1-DGN\_CD25  **OUTSAF.**E1DGNSCD  **OUTSAF.**EDGNSD1-EDGNSD6  **NCH.**LINEDIAG  **NCH.**DGN\_CD1-DGN\_CD12  **DME.**LINEDIAG  **DME.**DGN\_CD1-DGN\_CD12 |  |  |
| PROCEDURE\_SOURCE\_CONCEPT\_ID | Surgical Fields:  **MEDPAR.**SRGCDE1-SRGCDE25  HCPCS fields:  **OUTSAF.**HCPCS  **NCH.**HCPCS  **DME.**HCPCS  Diagnosis fields:  **MEDPAR.**DGN\_CD1-DGN\_CD25  **OUTSAF.** DGN\_CD1-DGN\_CD25  **OUTSAF.**E1DGNSCD  **OUTSAF.**EDGNSD1-EDGNSD6  **NCH.**LINEDIAG  **NCH.**DGN\_CD1-DGN\_CD12  **DME.**LINEDIAG  **DME.**DGN\_CD1-DGN\_CD12 | Map source\_values to their associated SOURCE\_CONCEPT\_ID using the vocab query in [Appendix 3](#_Appendix_3:_Source) | For codes from HCPCS or surgical fields use 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')  From diagnosis fields:  WHERE SOURCE\_VOCABULARY\_ID IN ('ICD9CM')  AND TARGET\_VOCABUALRY\_ID IN ('ICD9CM') |
| QUALIFIER\_SOURCE\_VALUE | **OUTSAF.**MF1  **NCH.**MF1  **DME.**MF1 | NULL |  |

## 2.11 Table name: PROCEDURE\_COST

Key conventions:

* Procedure costs are only taken from the OUTSAF and NCH tables and are only associated with HCPCS fields
* If a code in a HCPCS field from OUTSAF or NCH is mapped to a standard concept with a domain other than procedure a record will be written to the corresponding table (measurement, device, etc.), however, a PROCEDURE\_OCCURRENCE record is also written with PROCEDURE\_CONCEPT\_ID = 0. This is so a PROCEDURE\_COST record will be written; the costs are left in the PROCEDURE\_COST table rather than moving them to other cost tables because the claim was originally billed as a procedure regardless of the domain it eventually maps to.

Please note: This logic will cause each procedure coming from a HCPCS field to be duplicated. If an analysis is conducted using CONCEPT\_IDs then the duplication will not matter but if an analysis is conducted using source codes the table must be de-duped before proceeding with analysis.

Also, the field TOTAL\_PAID will reflect to total amount paid to the provider, NOT the total amount Medicare paid. To find what Medicare alone paid use PAID\_BY\_PAYER

| **Table 11: PROCEDURE\_COST** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| PROCEDURE\_COST\_ID | - | System Generated |  |
| PROCEDURE\_OCCURRENCE\_ID | **PROCEDURE\_OCCURRENCE**: PROCEDURE\_OCCURRENCE\_ID |  |  |
| CURRENCY\_CONCEPT\_ID | **-** | This will be ‘44818668’ for all rows since this is a US claims database and paid in US Dollars |  |
| PAID\_COPAY | - |  |  |
| PAID\_COINSURANCE | **OUTSAF:** WAGEADJ  **NCH:** COINAMT |  |  |
| PAID\_TOWARD\_DEDUCTIBLE | **OUTSAF:** REVDCTBL  **NCH:** LDEDAMT |  |  |
| PAID\_BY\_PAYER | **OUTSAF:** PAY  **NCH:** LINEPMT |  |  |
| PAID\_BY\_COORDINATION\_BENEFITS | **OUTSAF:** REV\_MSP1  **NCH:** LPRPAYAT |  |  |
| TOTAL\_OUT\_OF\_POCKET | PAID\_COPAY+PAID\_COINSURANCE+PAID\_TOWARD\_DEDUCTIBLE |  |  |
| TOTAL\_PAID | PAID\_BY\_COORDINATION\_OF\_BENEFITS+TOTAL\_OUT\_OF\_POCKET + PAID\_BY\_PAYER |  |  |
| REVENUE\_CODE\_CONCEPT\_ID | **OUTSAF:** CENTER | Use code in [Appendix 3](#_Appendix_3:_Source)  Filters:  WHERE SOURCE\_VOCABULARY\_ID IN ('Revenue Code')  AND TARGET\_VOCABULARY\_ID IN ('Revenue Code') |  |
| 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. |
| REVENUE\_CODE\_SOURCE\_VALUE | **OUTSAF:** CENTER |  |  |

## 2.12 Table name: DEATH

Key conventions:

* This will be sourced from PEDSF and MEDPAR only
* The cause of death will come from the PEDSF variables:
  + PEDSF.cod89v
    - If icd\_code = 8 then it indicates the cause of death is coded in ICD8 and will not be mapped
    - If icd\_code = 9 then it indicates the cause of death is coded in ICD9 and should be mapped from cod89v according to the filters below
  + PEDSF.cod10v
    - If icd\_code = 1 then it indicates the cause of death is coded in ICD10 and should be mapped from cod10v according to the filters below. As with the ICD9 codes, the decimal should be removed from cod10v before mapping.
* This is the only table that will use ICD10 codes

| **Table 12: Death** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Logic** | **Comment** |
| PERSON\_ID | PEDSF.patient\_id  MEDPAR.patient\_id | Select patient\_id from PEDSF where PEDSF.dod\_flg!=0  union  Select patient\_id from MEDPAR where himasind not = ‘ ‘ | The logic picks out the patients who have a recording of death |
| DEATH\_DATE | (1)PEDSF.med\_dody, PEDSF.med\_dodm, PEDSF.med\_dodd  (3)PEDSF.ser\_dody, PEDSF.ser\_dodm  (2)MEDPAR.dod\_y, MEDPAR.dod\_m, MEDPAR.dod\_d | If PEDSF.med\_dody is not NULL construct date from:   * Year: PEDSF.med\_dody, * Month: PEDSF.med\_dodm * Day: PEDSF.med\_dodd   Else if MEDPAR.dod\_y is not NULL construct date from:   * Year: MEDPAR.dod\_y, * Month: MEDPAR.dod\_m, * Day: MEDPAR.dod\_d   Else if PEDSF.ser\_dody is not NULL construct date from:   * Year: PEDSF.ser\_dody, * Month: PEDSF.ser\_dodm, * Day: use 1   Else date is NULL | If day is missing use 1st  If month is missing use January |
| DEATH\_TYPE\_CONCEPT\_ID | MEDPAR.himasind | DEATH\_DATE source = PEDSF then use 38003565  Else if   * MEDPAR.himasind=V use 38003567 * MEDPAR.himasind=B use 38003567 * MEDPAR.himasind=N use 38003566 |  |
| CAUSE\_CONCEPT\_ID | PEDSF.cod89v  PEDSF.cod10v | Map source\_values to their associated TARGET\_CONCEPT\_IDs using the vocab query in [Appendix 4](#_Appendix_4:_Source)  Otherwise set to 0 | IF PEDSF.icd\_code=1 and PEDSF.cod10v!=’0000’ then map cod10v using the filter:  WHERE SOURCE\_VOCABULARY\_ID IN ('ICD10CM') AND TARGET\_VOCABUALRY\_ID IN ('SNOMED')  IF PEDSF.icd\_code=9 then map cov89v using the filter:  WHERE SOURCE\_VOCABULARY\_ID IN ('ICD9CM') AND TARGET\_VOCABUALRY\_ID IN ('SNOMED')  If PEDSF.icd\_code=8 then set to 0 |
| CAUSE\_SOURCE\_VALUE | PEDSF.cod89v  PEDSF.cod10v |  | NULL if death only recorded in MEDPAR |
| CAUSES\_SOURCE\_CONCEPT\_ID | PEDSF.cod89v  PEDSF.cod10v | Map source\_values to their associated SOURCE\_CONCEPT\_ID using the vocab query in [Appendix 3](#_Appendix_3:_Source)  Otherwise set to 0 | IF PEDSF.icd\_code=1 and PEDSF.cod10v!=’0000’ use filter:  WHERE SOURCE\_VOCABULARY\_ID in (’ICD10CM’) AND TARGET\_VOCABUALRY\_ID in (’ICD10CM’)  IF PEDSF.icd\_code=9 then map cov89v using the filter:  WHERE SOURCE\_VOCABULARY\_ID IN ('ICD9CM') AND TARGET\_VOCABUALRY\_ID IN ('ICD9CM')  If PEDSF.icd\_code=8 then set to 0 |

## 2.13 Table name: DRUG\_EXPOSURE

This table will be sourced from PDESAF, OUTSAF, NCH and DME. Any PROD\_SRVC\_ID field in PDESAF (NDC codes) that maps to a concept with DOMAIN\_ID =’ Drug’ should go in this table as well as any value in the HCPCS field in OUTSAF, NCH or DME that maps to a concept with DOMAIN\_ID = ‘Drug’.

* If any drug\_exposure record occurs >= 30 days after death then that record should be deleted

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 13: DRUG\_EXPOSURE** | | | |
| **Destination Field** | **Source Field** | **Logic** | **Comment** |
| DRUG\_EXPOSURE\_ID | - | System Generated | Auto generate for each unique provider\_source\_value |
| PERSON\_ID | **PDESAF.**Patient\_ID  **OUTSAF.**Patient\_ID  **NCH.**Patient\_ID  **DME.**Patient\_ID |  |  |
| DRUG\_CONCEPT\_ID | **PDESAF**.PROD\_SRVC\_ID  HCPCS fields:  **OUTSAF.**HCPCS  **NCH.**HCPCS  **DME.**HCPCS | Map source\_values to their associated TARGET\_CONCEPT\_IDs using the vocab query in [Appendix 4](#_Appendix_4:_Source) | For codes from PROD\_SRVC\_ID field use filter:  SOURCE\_VOCABULARY\_ID='NDC'  AND TARGET\_VOCABUALRY\_ID='RxNORM'  For codes from HCPCS fields use 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') |
| DRUG\_EXPOSURE\_START\_DATE | **PDESAF**.SRVC\_MON  **PDESAF**.SRVC\_DAY  **PDESAF**.SRVC\_YR    HCPCS Fields:  **NCH**.FREXPENM, FREXPEND, FREXPENY  **OUTSAF.**CENM, CEND, CENY OR FROM\_DTM, FROM\_DTD, FROM\_DTY  **DME**. FROM\_DTM, FROM\_DTD, FROM\_DTY | SRVC\_MON+ SRVC\_DAY+ SRVC\_YR  For a HCPCS field in the **OUTSAF** file, if CENM, CEND, CENY = ‘00’ then use FROM\_DTM, FROM\_DTD, FROM\_DTY as the procedure date |  |
| DRUG\_EXPOSURE\_END\_DATE |  | NULL |  |
| DRUG\_TYPE\_CONCEPT\_ID | - | For a code from PROD\_SVC\_ID:  38000175-Prescription dispensed in pharmacy  HCPCS fields:  38000179-Physician administered drug (identified as procedure) |  |
| STOP\_REASON |  | NULL |  |
| REFILLS |  | NULL |  |
| QUANTITY | **PDESAF**.QTY\_DSPNSD\_NUM  **OUTSAF.**UNIT  **NCH.**MTUSCNT  **DME.**MTUSCNT | NULL |  |
| DAYS\_SUPPLY | **PDESAF**.DAYS\_SUPLY\_NUM | NULL |  |
| SIG |  | NULL |  |
| ROUTE\_CONCEPT\_ID |  | 0 |  |
| EFFECTIVE\_DRUG\_DOSE |  | NULL |  |
| DOSE\_UNIT\_CONCEPT\_ID |  | 0 |  |
| LOT\_NUMBER |  | NULL |  |
| PROVIDER\_ID | **NCH**.PRF\_NPI or PERUPIN  **NCH**.HCFASPEC  **DME**.ORD\_NPI or ORD\_UPIN  **DME**.HCFASPEC  **OUTSAF**.AT\_NPI or AT\_UPIN | * If **NCH.**PRF\_NPI is NULL then use **NCH.**PERUPIN * If **DME.** ORD\_NPI is NULL then use **DME.** ORD\_UPIN * If **OUTSAF**.AT\_NPI is NULL then use **OUTSAF**.AT\_UPIN   When procedure is in DME or NCH use both UPIN/NPI and HCFASPEC to map to the correct PROVIDER\_ID | Map these values back to PROVIDER\_ID using the PROVIDER table |
| VISIT\_OCCURRENCE\_ID | For code coming from a HCPCS field  **VISIT\_OCCURRENCE:** VISIT\_OCCURRENCE\_ID | Refer to logic in table 2.6 to assign this value |  |
| DRUG\_SOURCE\_VALUE | **PDESAF**.PROD\_SRVC\_ID  HCPCS fields:  **OUTSAF.**HCPCS  **NCH.**HCPCS  **DME.**HCPCS |  |  |
| DRUG\_SOURCE\_CONCEPT\_ID | **PDESAF**.PROD\_SRVC\_ID  HCPCS fields:  **OUTSAF.**HCPCS  **NCH.**HCPCS  **DME.**HCPCS | Map source\_values to their associated SOURCE\_CONCEPT\_ID using the vocab query in [Appendix 3](#_Appendix_3:_Source) | SOURCE\_VOCABULARY\_ID='NDC'  AND TARGET\_VOCABUALRY\_ID='NDC'  For codes from HCPCS or surgical fields use 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') |
| ROUTE\_SOURCE\_VALUE |  | NULL |  |
| DOSE\_UNIT\_SOURCE\_VALUE |  | NULL |  |

## 2.14 Table name: DRUG\_COST

This table will be sourced from PDESAF

| **Table 14: DRUG\_COST** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| DRUG\_COST\_ID | - | System Generated |  |
| DRUG\_EXPOSURE\_ID | **DRUG\_EXPOSURE**:  DRUG\_EXPOSURE\_ID |  |  |
| CURRENCY\_CONCEPT\_ID | **-** | This will be ‘44818668’ for all rows since this is a US claims database and paid in US Dollars |  |
| PAID\_COPAY | - |  |  |
| PAID\_COINSURANCE | - |  |  |
| PAID\_TOWARD\_DEDUCTIBLE | - |  |  |
| PAID\_BY\_PAYER | TOT\_RX\_CST\_AMT |  |  |
| PAID\_BY\_COORDINATION\_BENEFITS | - |  |  |
| TOTAL\_OUT\_OF\_POCKET | PT\_PAY\_AMT |  |  |
| TOTAL\_PAID | TOT\_RX\_CST\_AMT+PT\_PAY\_AMT |  |  |
| INGREDIENT\_COST | - |  |  |
| DISPENSING\_FEE | - |  |  |
| AVERAGE\_WHOLESALE\_PRICE | **-** |  |  |
| 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. |

## 2.15 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.
3. Sort DRUG\_EXPOSURE table by PERSON\_ID, DRUG\_CONCEPT\_ID, DRUG\_EXPOSURE\_START\_DATE and DRUG\_EXPOSURE\_END\_DATE in ascending order.
4. 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 create multiple ERAs since they can be mapped to multiple ingredients.

| **Table 15: 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 |  |

## 2.16 Table name: DEVICE\_EXPOSURE

Key Conventions:

This table will be sourced from MEDPAR, OUTSAF, NCH and DME. Refer to the logic above for defining visits; it will be used for assigning VISIT\_OCCURRENCE\_ID. Any code in a diagnosis field, HCPCS field or surgical field that maps to a concept with DOMAIN\_ID =’ Device’ should go in this table.

* If any device exposure occurs >= 30 days after the patient’s death then that record should be deleted

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 16: DEVICE\_EXPOSURE** | | | |
| **Destination Field** | **Source Field** | **Logic** | **Comment** |
| DEVICE\_EXPOSURE\_ID | - | System Generated | Auto generate for each unique provider\_source\_value |
| PERSON\_ID | **MEPDAR.**Patient\_ID  **OUTSAF.**Patient\_ID  **NCH.**Patient\_ID  **DME.**Patient\_ID |  |  |
| DEVICE\_CONCEPT\_ID | Surgical Fields:  **MEDPAR.**SRGCDE1-SRGCDE25  HCPCS fields:  **OUTSAF.**HCPCS  **NCH.**HCPCS  **DME.**HCPCS  Diagnosis fields:  **MEDPAR.**DGN\_CD1-DGN\_CD25  **OUTSAF.** DGN\_CD1-DGN\_CD25  **OUTSAF.**E1DGNSCD  **OUTSAF.**EDGNSD1-EDGNSD6  **NCH.**LINEDIAG  **NCH.**DGN\_CD1-DGN\_CD12  **DME.**LINEDIAG  **DME.**DGN\_CD1-DGN\_CD12 | Map source\_values to their associated TARGET\_CONCEPT\_IDs using the vocab query in [Appendix 4](#_Appendix_4:_Source) | For codes from Surgical or HCPCS fields use 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')  For codes from diagnosis fields use filters:  WHERE SOURCE\_VOCABULARY\_ID IN ('ICD9CM')  AND TARGET\_VOCABUALRY\_ID IN ('SNOMED') |
| DEVICE\_EXPOSURE\_START\_DATE | Surgical Fields:  **MEDPAR.**SG\_M1-SG\_M25, SG\_D1-SG\_D25, SG\_Y1-SG\_Y25  HCPCS Fields:  **NCH**.FREXPENM, FREXPEND, FREXPENY  **OUTSAF.**CENM, CEND, CENY OR FROM\_DTM, FROM\_DTD, FROM\_DTY  **DME**. FROM\_DTM, FROM\_DTD, FROM\_DTY  Diagnosis fields:  **MEDPAR**.ADM\_D, ADM\_M, ADM\_Y  **NCH**.FREXPENM, FREXPEND, FREXPENY OR FROM\_DTM, FROM\_DTD, FROM\_DTY  **OUTSAF.**CENM, CEND, CENY OR FROM\_DTM, FROM\_DTD, FROM\_DTY  **DME**. FREXPENM, FREXPEND, FREXPENY OR FROM\_DTM, FROM\_DTD, FROM\_DTY | For a HCPCS field in the **OUTSAF** file, if CENM, CEND, CENY = ‘00’ then use FROM\_DTM, FROM\_DTD, FROM\_DTY as the device start date  For diagnosis codes coming from NCH or DME, if the code is **NCH.**LINEDIAG or **DME.**LINEDIAG then use FREXPENM, FREXPEND, FREXPENY, otherwise use FROM\_DTM, FROM\_DTD, FROM\_DTY |  |
| DEVICE\_EXPOSURE\_END\_DATE | Surgical Fields:  **MEDPAR.**SG\_DT1-SG\_DT25  HCPCS Fields:  **NCH**.LSEXPENM, LSEXPEND, LSEXPENY  **OUTSAF.**CENM, CEND, CENY OR THRU\_DTM, THRU\_DTD, THRU\_DTY  **DME**. THRU\_DTM, THRU\_DTD, THRU\_DTY  Diagnosis fields:  **MEDPAR**.DIS\_D, DIS\_M, DIS\_Y  **NCH**.LSEXPENM, LSEXPEND, LSEXPENY OR THRU\_DTM, THRU\_DTD, THRU\_DTY  **OUTSAF.**CENM, CEND, CENY OR THRU\_DTM, THRU\_DTD, THRU\_DTY  **DME**. LSEXPENM, LSEXPEND, LSEXPENY OR THRU\_DTM, THRU\_DTD, THRU\_DTY | For a HCPCS field in the **OUTSAF** file, if CENM, CEND, CENY = ‘00’ then use THRU\_DTM, THRU\_DTD, THRU\_DTY as the device end date  For diagnosis codes coming from NCH or DME, if the code is **NCH.**LINEDIAG or **DME.**LINEDIAG then use LSEXPENM, LSEXPEND, LSEXPENY, otherwise use THRU\_DTM, THRU\_DTD, THRU\_DTY |  |
| DEVICE\_TYPE\_CONCEPT\_ID | - | 44818705 (Inferred from procedure claim) |  |
| UNIQUE\_DEVICE\_ID | **0** | - |  |
| QUANTITY | - | NULL |  |
| PROVIDER\_ID | **NCH**.PRF\_NPI or PERUPIN  **NCH**.HCFASPEC  **DME**. ORD\_NPI or ORD\_UPIN  **DME**.HCFASPEC  **OUTSAF**.AT\_NPI or AT\_UPIN | * If **NCH.**PRF\_NPI is NULL then use **NCH.**PERUPIN * If **DME.** ORD\_NPI is NULL then use **DME.** ORD\_UPIN * If **OUTSAF**.AT\_NPI is NULL then use **OUTSAF**.AT\_UPIN   When device is in DME or NCH use both UPIN/NPI and HCFASPEC to map to the correct PROVIDER\_ID | Map these values back to PROVIDER\_ID using the PROVIDER table |
| VISIT\_OCCURRENCE\_ID | **VISIT\_OCCURRENCE:** VISIT\_OCCURRENCE\_ID | Refer to logic in table 2.6 to assign this value |  |
| DEVICE\_SOURCE\_VALUE | Surgical Fields:  **MEDPAR.**SRGCDE1-SRGCDE25  HCPCS fields:  **OUTSAF.**HCPCS  **NCH.**HCPCS  **DME.**HCPCS  Diagnosis fields:  **MEDPAR.**DGN\_CD1-DGN\_CD25  **OUTSAF.** DGN\_CD1-DGN\_CD25  **OUTSAF.**E1DGNSCD  **OUTSAF.**EDGNSD1-EDGNSD6**NCH.**LINEDIAG  **NCH.**DGN\_CD1-DGN\_CD12  **DME.**LINEDIAG  **DME.**DGN\_CD1-DGN\_CD12 |  |  |
| DEVICE\_SOURCE\_CONCEPT\_ID | Surgical Fields:  **MEDPAR.**SRGCDE1-SRGCDE25  HCPCS fields:  **OUTSAF.**HCPCS  **NCH.**HCPCS  **DME.**HCPCS  Diagnosis fields:  **MEDPAR.**DGN\_CD1-DGN\_CD25  **OUTSAF.** DGN\_CD1-DGN\_CD25  **OUTSAF.**E1DGNSCD  **OUTSAF.**EDGNSD1-EDGNSD6  **NCH.**LINEDIAG  **NCH.**DGN\_CD1-DGN\_CD12  **DME.**LINEDIAG  **DME.**DGN\_CD1-DGN\_CD12 | Map source\_values to their associated SOURCE\_CONCEPT\_ID using the vocab query in [Appendix 3](#_Appendix_3:_Source) | 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')  For codes from diagnosis fields use filters:  WHERE SOURCE\_VOCABULARY\_ID IN ('ICD9CM')  AND TARGET\_VOCABUALRY\_ID IN ('ICD9CM') |

## 2.17 Table name: DEVICE\_COST

Key conventions:

* Device costs are only taken from the DME table and are only associated with HCPCS fields
* If a code in a DME HCPCS field is mapped to a standard concept with a domain other than Device a record will be written to the corresponding table (measurement, procedure, etc.), however, a DEVICE\_EXPOSURE record is also written with DEVICE\_EXPOSURE\_ID = 0. This is so a DEVICE\_COST record will be written; the costs are left in the DEVICE\_COST table rather than moving them to other cost tables because the claim was originally billed as a device regardless of the domain it eventually maps to.

Please note: This logic will cause each device coming from a DME HCPCS field to be duplicated. If an analysis is conducted using CONCEPT\_IDs then the duplication will not matter but if an analysis is conducted using source codes the table must be de-duped before proceeding with analysis.

| **Table 17: DEVICE\_COST** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| DEVICE\_COST\_ID | - | System Generated |  |
| DEVICE\_EXPOSURE\_ID | **DEVICE\_EXPOSURE**: DEVICE\_EXPOSURE\_ID |  |  |
| CURRENCY\_CONCEPT\_ID | **-** | This will be ‘44818668’ for all rows since this is a US claims database and paid in US Dollars |  |
| PAID\_COPAY | - |  |  |
| PAID\_COINSURANCE | COINAMT |  |  |
| PAID\_TOWARD\_DEDUCTIBLE | LDEDAMT |  |  |
| PAID\_BY\_PAYER | LINEPMT |  |  |
| PAID\_BY\_COORDINATION\_BENEFITS | LPRPAYAT |  |  |
| TOTAL\_OUT\_OF\_POCKET | PAID\_COPAY+PAID\_COINSURANCE+PAID\_TOWARD\_DEDUCTIBLE |  |  |
| TOTAL\_PAID | PAID\_BY\_COORDINATION\_OF\_BENEFITS+TOTAL\_OUT\_OF\_POCKET + PAID\_BY\_PAYER |  |  |
| REVENUE\_CODE\_CONCEPT\_ID | **-** |  |  |
| PAYER\_PLAN\_PERIOD\_ID | - | Lookup associated PAYER\_PLAN\_PERIOD\_ID. Look up by PERSON\_ID and DEVICE\_EXPOSURE\_START\_DATE. If there no match, put NULL. | There should only be one possible plan. |
| REVENUE\_CODE\_SOURCE\_VALUE |  |  |  |

## 2.18 Table name: VISIT\_COST

Key conventions:

* Visit costs are costs associated with IP visits coming from the MEDPAR table OR
  + Claim lines in OUTSAF where the HCPCS field is blank and the CENTER field is not equal to 0001
  + Claim lines in DME where the HCPCS field is blank

| **Table 18: VISIT\_COST** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| VISIT\_COST\_ID | - | System Generated |  |
| VISIT\_OCCURRENCE\_ID | **VISIT\_OCCURRENCE**: VISIT\_OCCURRENCE\_ID |  |  |
| CURRENCY\_CONCEPT\_ID | **-** | This will be ‘44818668’ for all rows since this is a US claims database and paid in US Dollars |  |
| PAID\_COPAY | - |  |  |
| PAID\_COINSURANCE | **MEDPAR.**COINAMT  **OUTSAF.**WAGEADJ  **DME**.COINAMT |  |  |
| PAID\_TOWARD\_DEDUCTIBLE | **MEDPAR.**INPATDED+BLOODDED  **OUTSAF.**REVDCTBL  **DME**.LDEDAMT |  |  |
| PAID\_BY\_PAYER | **MEDPAR.**REIMBAMT+PASSTHRU  **OUTSAF.**PAY  **DME**.LINEPMT |  |  |
| PAID\_BY\_COORDINATION\_BENEFITS | **MEDPAR.**PRIPYAMT  **OUTSAF.**REV\_MSP1  **DME**.LPRPAYAT |  |  |
| TOTAL\_OUT\_OF\_POCKET | PAID\_COPAY+PAID\_COINSURANCE+PAID\_TOWARD\_DEDUCTIBLE |  |  |
| TOTAL\_PAID | PAID\_BY\_COORDINATION\_OF\_BENEFITS+TOTAL\_OUT\_OF\_POCKET + PAID\_BY\_PAYER |  |  |
| REVENUE\_CODE\_CONCEPT\_ID | **OUTSAF:** CENTER | Use code in [Appendix 3](#_Appendix_3:_Source)  Filters:  WHERE SOURCE\_VOCABULARY\_ID IN ('Revenue Code')  AND TARGET\_VOCABULARY\_ID IN ('Revenue Code') |  |
| PAYER\_PLAN\_PERIOD\_ID | - | Lookup associated PAYER\_PLAN\_PERIOD\_ID. Look up by PERSON\_ID and VISIT\_START\_DATE. If there is no match, put NULL. | There should only be one possible plan. |
| REVENUE\_CODE\_SOURCE\_VALUE | **OUTSAF:** CENTER |  |  |

## 2.19 Table name: MEASUREMENT

Key conventions:

* This table will be populated with rows from the NCH table where the variable HGB\_RSLT is not in (' .', ' ' or '0000').
  + Take distinct HGB\_RSLT by VISIT\_OCCURRENCE\_ID
* This table will be sourced from MEDPAR, OUTSAF, NCH and DME. Refer to the logic in the VISIT\_OCCURRENCE table for defining visits; it will be used for assigning VISIT\_OCCURRENCE\_ID. Any code in a diagnosis field, HCPCS field or surgical field that maps to a concept with DOMAIN\_ID =’ Measurement’ should go in this table.
* If any measurement record occurs >= 30 days after death then that record should be deleted

| **Table 19: MEASUREMENT** | | | |
| --- | --- | --- | --- |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| MEASUREMENT\_ID | - | System generated |  |
| PERSON\_ID | PATIENT\_ID |  |  |
| MEASUREMENT\_CONCEPT\_ID | HGB\_TYPE OR  Surgical Fields:  **MEDPAR.**SRGCDE1-SRGCDE25  HCPCS fields:  **OUTSAF.**HCPCS  **NCH.**HCPCS  **DME.**HCPCS  Diagnosis fields:  **MEDPAR.**DGN\_CD1-DGN\_CD25  **OUTSAF.** DGN\_CD1-DGN\_CD25  **OUTSAF.**E1DGNSCD  **OUTSAF.**EDGNSD1-EDGNSD6  **NCH.**LINEDIAG  **NCH.**DGN\_CD1-DGN\_CD12  **DME.**LINEDIAG  **DME.**DGN\_CD1-DGN\_CD12 | If HGB\_TYPE = R1 then 30000963  If HGB\_TYPE = R2 then 3009542  Otherwise  Map source\_values to their associated TARGET\_CONCEPT\_IDs using the vocab query in [Appendix 4](#_Appendix_4:_Source) | For codes from HCPCS or Surgical fields use 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')  For codes from diagnosis fields use filters:  WHERE SOURCE\_VOCABULARY\_ID IN ('ICD9CM')  AND TARGET\_VOCABUALRY\_ID IN ('SNOMED') |
| MEASUREMENT\_DATE | **If Measurement from HGB\_TYPE:**  **VISIT\_OCCURRENCE.**VISIT\_START\_DATE  Surgical Fields:  **MEDPAR.**  SG\_DT1-SG\_DT25  HCPCS Fields:  **NCH**.FREXPENM, FREXPEND, FREXPENY  **OUTSAF.**CENM, CEND, CENY OR FROM\_DTM, FROM\_DTD, FROM\_DTY  **DME**. FROM\_DTM, FROM\_DTD, FROM\_DTY  Diagnosis fields:  **MEDPAR**.ADM\_D, ADM\_M, ADM\_Y  **NCH**.FREXPENM, FREXPEND, FREXPENY OR FROM\_DTM, FROM\_DTD, FROM\_DTY  **OUTSAF.** FROM\_DTM, FROM\_DTD, FROM\_DTY  **DME**. FREXPENM, FREXPEND, FREXPENY OR FROM\_DTM, FROM\_DTD, FROM\_DTY | For a HCPCS field in the **OUTSAF** file, if CENM, CEND, CENY = ‘00’ then use FROM\_DTM, FROM\_DTD, FROM\_DTY as the measurement date  For a diagnosis code coming from OUTSAF (DGN\_CD1-DGN\_CD25) then use FROM\_DTM, FROM\_DTD, FROM\_DTY as the measurement date  For diagnosis codes coming from NCH or DME, if the code is **NCH.**LINEDIAG or **DME.**LINEDIAG then use FREXPENM, FREXPEND, FREXPENY, otherwise for DGN\_CD1-DGN\_CD12 use FROM\_DTM,FROM\_DTD, FROM\_DTY |  |
| MEASUREMENT\_TIME | - |  |  |
| MEASUREMENT\_TYPE\_CONCEPT\_ID | - | This will be 44818702 for all rows (lab result) |  |
| OPERATOR\_CONCEPT\_ID | - | 0 |  |
| VALUE\_AS\_NUMBER | HGB\_RSLT |  |  |
| VALUE\_AS\_CONCEPT\_ID | - | 0 |  |
| UNIT\_CONCEPT\_ID | - | 0 |  |
| RANGE\_LOW | - |  |  |
| RANGE\_HIGH | - |  |  |
| PROVIDER\_ID | **VISIT\_OCCURRENCE.**PROVIDER\_ID |  |  |
| VISIT\_OCCURRENCE\_ID | **VISIT\_OCCURRENCE.**VISIT\_OCCURRENCE\_ID |  |  |
| MEASUREMENT\_SOURCE\_VALUE | HGB\_RSLT  OR  Surgical Fields:  **MEDPAR.**SRGCDE1-SRGCDE25  HCPCS fields:  **OUTSAF.**HCPCS  **NCH.**HCPCS  **DME.**HCPCS  Diagnosis fields:  **MEDPAR.**DGN\_CD1-DGN\_CD25  **OUTSAF.** DGN\_CD1-DGN\_CD25  **OUTSAF.**E1DGNSCD  **OUTSAF.**EDGNSD1-EDGNSD6  **NCH.**LINEDIAG  **NCH.**DGN\_CD1-DGN\_CD12  **DME.**LINEDIAG  **DME.**DGN\_CD1-DGN\_CD12 |  |  |
| MEASUREMENT\_SOURCE\_CONCEPT\_ID | Surgical Fields:  **MEDPAR.**SRGCDE1-SRGCDE25  HCPCS fields:  **OUTSAF.**HCPCS  **NCH.**HCPCS  **DME.**HCPCS  Diagnosis fields:  **MEDPAR.**DGN\_CD1-DGN\_CD25  **OUTSAF.** DGN\_CD1-DGN\_CD25  **OUTSAF.**E1DGNSCD  **OUTSAF.**EDGNSD1-EDGNSD6  **NCH.**LINEDIAG  **NCH.**DGN\_CD1-DGN\_CD12  **DME.**LINEDIAG  **DME.**DGN\_CD1-DGN\_CD12 | Map source\_values to their associated SOURCE\_CONCEPT\_ID using the vocab query in [Appendix 3](#_Appendix_3:_Source) | For codes from HCPCS or surgical fields use 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')  From diagnosis fields:  WHERE SOURCE\_VOCABULARY\_ID IN ('ICD9CM')  AND TARGET\_VOCABUALRY\_ID IN ('ICD9CM') |
| UNIT\_SOURCE\_VALUE | - |  |  |
| VALUE\_SOURCE\_VALUE | - |  |  |

## 2.20 Table name: OBSERVATION

Key Conventions:

* This table will be sourced from PEDSF, MEDPAR, OUTSAF, NCH and DME.
* For MEDPAR, OUTSAF, NCH and DME refer to the logic in the VISIT\_OCCURRENCE table for defining visits; it will be used for assigning VISIT\_OCCURRENCE\_ID. Any code in a diagnosis field, HCPCS field or surgical field that maps to a concept with DOMAIN\_ID =’ Observation’ should go in this table.
* Records in the PEDSF table are based on person, not on visit so there will be no need to assign a VISIT\_OCCURRENCE\_ID to records coming from that table.
  + In order to map codes in **PEDSF.**NAME to their associated TARGET\_CONCEPT\_IDs you will need to collate the srcTable column and name column like this:
    - srcTable+’\_’+NAME
  + This is how the values were stored in the vocabulary because different source tables (srcTable) may have different mappings based on which type of cancer they are describing
* One record should be created per patient\_id and srcTable value identifying which source tables each person belongs to
  + select distinct patient\_id, srcTable from pedsf
  + After identifying which source tables each patient belongs to, create one record per unique patient\_id and source table with the following fields:
    - OBSERVATION\_ID = System generated
    - PERSON\_ID = patient\_id
    - OBSERVATION\_CONCEPT\_ID = 0
    - OBSERVATION\_DATE = MIN(OBSERVATION\_PERIOD\_START\_DATE)
    - OBSERVATION\_TIME = 0
    - OBSERVATION\_TYPE\_CONCEPT\_ID = 0
    - VALUE\_AS\_NUMBER = 0
    - VALUE\_AS\_STRING = value in srcTable column
    - VALUE\_AS\_CONCEPT\_ID = 0
    - QUALIFIER\_CONCEPT\_ID = 0
    - UNIT\_CONCEPT\_ID = 0
    - PROVIDER\_ID = 0
    - VISIT\_OCCURRENCE\_ID = 0
    - OBSERVATION\_SOURCE\_VALUE = ‘srcTable’
    - OBSERVATION\_SOURCE\_CONCEPT\_ID = 0
    - UNIT\_SOURCE\_VALUE = NULL
    - QUALIFIER\_SOURCE\_VALUE = NULL
* If any observation record occurs >= 30 days after the patient’s death then that record should be deleted

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 20: OBSERVATION** | | | |
| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| OBSERVATION\_ID | - | System generated |  |
| PERSON\_ID | Patient\_ID |  |  |
| OBSERVATION\_CONCEPT\_ID | Surgical Fields:  **MEDPAR.**SRGCDE1-SRGCDE25  HCPCS fields:  **OUTSAF.**HCPCS  **NCH.**HCPCS  **DME.**HCPCS  Diagnosis fields:  **MEDPAR.**DGN\_CD1-DGN\_CD25  **OUTSAF.** DGN\_CD1-DGN\_CD25  **OUTSAF.**E1DGNSCD  **OUTSAF.**EDGNSD1-EDGNSD6  **NCH.**LINEDIAG  **NCH.**DGN\_CD1-DGN\_CD12  **DME.**LINEDIAG  **DME.**DGN\_CD1-DGN\_CD12  PEDSF fields:  PEDSF.srcTable+’\_’+PEDSF.Name | Map source\_values to their associated TARGET\_CONCEPT\_IDs using the vocab query in [Appendix 4](#_Appendix_4:_Source) | For codes from HCPCS or Surgical fields use 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')  For codes from diagnosis fields use filters:  WHERE SOURCE\_VOCABULARY\_ID IN ('ICD9CM')  AND TARGET\_VOCABUALRY\_ID IN ('SNOMED')  For codes from PEDSF fields use filters:  WHERE SOURCE\_VOCABULARY\_ID IN ('JNJ\_SEER\_REGISTRY') |
| OBSERVATION\_DATE | Surgical Fields:  **MEDPAR.**  SG\_DT1-SG\_DT25  HCPCS Fields:  **NCH**.FREXPENM, FREXPEND, FREXPENY  **OUTSAF.**CENM, CEND, CENY OR FROM\_DTM, FROM\_DTD, FROM\_DTY  **DME**. FROM\_DTM, FROM\_DTD, FROM\_DTY  Diagnosis fields:  **MEDPAR**.ADM\_D, ADM\_M, ADM\_Y  **NCH**.FREXPENM, FREXPEND, FREXPENY OR FROM\_DTM, FROM\_DTD, FROM\_DTY  **OUTSAF.** FROM\_DTM, FROM\_DTD, FROM\_DTY  **DME**. FREXPENM, FREXPEND, FREXPENY OR FROM\_DTM, FROM\_DTD, FROM\_DTY  PEDSF fields:  Min(OBSERVATION\_PERIOD\_START\_DATE) | For a HCPCS field in the **OUTSAF** file, if CENM, CEND, CENY = ‘00’ then use FROM\_DTM, FROM\_DTD, FROM\_DTY as the measurement date  For a diagnosis code coming from OUTSAF (DGN\_CD1-DGN\_CD25) then use FROM\_DTM, FROM\_DTD, FROM\_DTY as the measurement date  For diagnosis codes coming from NCH or DME, if the code is **NCH.**LINEDIAG or **DME.**LINEDIAG then use FREXPENM, FREXPEND, FREXPENY, otherwise for DGN\_CD1-DGN\_CD12 use FROM\_DTM,FROM\_DTD, FROM\_DTY |  |
| OBSERVATION\_TIME | 0 |  |  |
| OBSERVATION\_TYPE\_CONCEPT\_ID | 0 | If the record is coming from another table, like the CONDITION\_OCCURENCE, keep the types that would have been assigned in that table. |  |
| VALUE\_AS\_NUMBER | 0 |  |  |
| VALUE\_AS\_STRING | **PEDSF**.fields: Use the value in the PEDSF.value field |  |  |
| VALUE\_AS\_CONCEPT\_ID | 0 |  |  |
| QUALIFIER\_CONCEPT\_ID | 0 |  |  |
| UNIT\_CONCEPT\_ID | 0 |  |  |
| PROVIDER\_ID | **PEDSF** fields: 0  **All other fields:**  **VISIT\_OCCURRENCE.**PROVIDER\_ID |  |  |
| VISIT\_OCCURRENCE\_ID | **PEDSF** fields: 0  **All other fields:**  **VISIT\_OCCURRENCE.**VISIT\_OCCURRENCE\_ID |  |  |
| OBSERVATION\_SOURCE\_VALUE | Surgical Fields:  **MEDPAR.**SRGCDE1-SRGCDE25  HCPCS fields:  **OUTSAF.**HCPCS  **NCH.**HCPCS  **DME.**HCPCS  Diagnosis fields:  **MEDPAR.**DGN\_CD1-DGN\_CD25  **OUTSAF.** DGN\_CD1-DGN\_CD25  **OUTSAF.**E1DGNSCD  **OUTSAF.**EDGNSD1-EDGNSD6**NCH.**LINEDIAG  **NCH.**DGN\_CD1-DGN\_CD12  **DME.**LINEDIAG  **DME.**DGN\_CD1-DGN\_CD12  PEDSF fields:  PEDSF.srcTable+’\_’+PEDSF.Name |  |  |
| OBSERVATION\_SOURCE\_CONCEPT\_ID | Surgical Fields:  **MEDPAR.**SRGCDE1-SRGCDE25  HCPCS fields:  **OUTSAF.**HCPCS  **NCH.**HCPCS  **DME.**HCPCS  Diagnosis fields:  **MEDPAR.**DGN\_CD1-DGN\_CD25  **OUTSAF.** DGN\_CD1-DGN\_CD25  **OUTSAF.**E1DGNSCD  **OUTSAF.**EDGNSD1-EDGNSD6  **NCH.**LINEDIAG  **NCH.**DGN\_CD1-DGN\_CD12  **DME.**LINEDIAG  **DME.**DGN\_CD1-DGN\_CD12  PEDSF fields:  PEDSF.srcTable+’\_’+PEDSF.Name | Map source\_values to their associated SOURCE\_CONCEPT\_ID using the vocab query in [Appendix 3](#_Appendix_3:_Source) | 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')  For codes from diagnosis fields use filters:  WHERE SOURCE\_VOCABULARY\_ID IN ('ICD9CM')  AND TARGET\_VOCABUALRY\_ID IN ('ICD9CM')  For codes from PEDSF fields use filters:  WHERE SOURCE\_VOCABULARY\_ID IN ('JNJ\_SEER\_REGISTRY') |
| UNIT\_SOURCE\_VALUE | - |  |  |
| QUALIFIER\_SOURCE\_VALUE | - |  |  |

## 2.21 Table name: CDM\_SOURCE

**CDM\_SOURCE.SOURCE\_DESCRIPTION:**

The SEER-Medicare data reflect the linkage of two large population-based sources of data that provide detailed information about Medicare beneficiaries with cancer. The data come from the Surveillance, Epidemiology and End Results (SEER) program of cancer registries that collect clinical, demographic and cause of death information for persons with cancer and the Medicare claims for covered health care services from the time of a person's Medicare eligibility until death.

This database contains claims and registry data for patients with B-cell cancers, namely leukemia, lymphoma, and multiple myeloma as well as data for a random sample of Medicare beneficiaries.

|  |  |  |
| --- | --- | --- |
| **Table 21: CDM\_SOURCE** | | |
| **TABLE NAME** | **DOMAIN\_ID** | **DESCRIPTION** |
| PERSON | Person | Person data is pulled from the SEER registry information. Any person born before 1900 or after 2015 is removed. |
| OBSERVATION\_PERIOD | Observation Period | An observation period is a representation of when a patient was enrolled in Medicare Part A and Medicare Part B and not enrolled in an HMO. This is due to the fact that HMO plans are handled by third party vendors and the resulting claims are not available in the SEER-Medicare database. |
| CARE\_SITE | Care Site | The only care site information available is the state in which the care was given so care site represents a state rather than an institution. |
| LOCATION | Location | Location in SEER-Medicare represents the states that participate in the cancer registry. |
| COHORT | Cohort | Any cohorts created using this database are housed here. |
| VISIT\_OCCURRENCE | Visit | A standardized definition of visit logic is applied to our U.S. claims data. SEER-Medicare applies methods to define inpatient and outpatient visits however we had to create logic to attribute carrier and device claims to visits. Inpatient visits defined by Medicare remain as IP unless they have a indicator suggesting it is actually a long-term care visit. Outpatient, physician, and device services during the middle of an inpatient stay are associated to that inpatient stay. Remaining outpatient claims are considered outpatient visits unless they have a revenue code suggesting they are actually ER visits. Any physician and device services during the middle of an outpatient visit are attributed to that outpatient stay. |
| PROVIDER | Provider | Unique list of health care providers (physicians). SEER-Medicare does provide some provider information however not all is available. Providers are first identified by their national provider number and if that is not available then their unique physician identification number is used. |
| DEATH | Death | SEER-Medicare endeavors to reconcile death data from claims with publicly available information and death certificates where possible. If a death date is listed in the registry file then that date is used first otherwise the death date from the claims information is used. |
| 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 | The only lab data available in SEER-Medicare is a hemoglobin test, 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 | All of the registry data collected by SEER is currently housed in this table as well as any codified data that is not a diagnosis, drug exposure, procedure, or measurement will become an observation. In order to determine the source table a patient came from (leukemia, lymphoma, multiple myeloma, 5 percent random sample) the variable OBSERVATION\_SOURCE\_VALUE will show the name of the source table collated with the name of the original variable as it appeared in the registry. There are over 2,000 registry variables available per patient. |

# Appendix 1: FIPS State Codes

|  |  |  |
| --- | --- | --- |
| **State Abbreviation** | **FIPS Code** | **State Name** |
| AK | 02 | ALASKA |
| AL | 01 | ALABAMA |
| AR | 05 | ARKANSAS |
| AS | 60 | AMERICAN SAMOA |
| AZ | 04 | ARIZONA |
| CA | 06 | CALIFORNIA |
| CO | 08 | COLORADO |
| CT | 09 | CONNECTICUT |
| DC | 11 | DISTRICT OF COLUMBIA |
| DE | 10 | DELAWARE |
| FL | 12 | FLORIDA |
| GA | 13 | GEORGIA |
| GU | 66 | GUAM |
| HI | 15 | HAWAII |
| IA | 19 | IOWA |
| ID | 16 | IDAHO |
| IL | 17 | ILLINOIS |
| IN | 18 | INDIANA |
| KS | 20 | KANSAS |
| KY | 21 | KENTUCKY |
| LA | 22 | LOUISIANA |
| MA | 25 | MASSACHUSETTS |
| MD | 24 | MARYLAND |
| ME | 23 | MAINE |
| MI | 26 | MICHIGAN |
| MN | 27 | MINNESOTA |
| MO | 29 | MISSOURI |
| MS | 28 | MISSISSIPPI |
| MT | 30 | MONTANA |
| NC | 37 | NORTH CAROLINA |
| ND | 38 | NORTH DAKOTA |
| NE | 31 | NEBRASKA |
| NH | 33 | NEW HAMPSHIRE |
| NJ | 34 | NEW JERSEY |
| NM | 35 | NEW MEXICO |
| NV | 32 | NEVADA |
| NY | 36 | NEW YORK |
| OH | 39 | OHIO |
| OK | 40 | OKLAHOMA |
| OR | 41 | OREGON |
| PA | 42 | PENNSYLVANIA |
| PR | 72 | PUERTO RICO |
| RI | 44 | RHODE ISLAND |
| SC | 45 | SOUTH CAROLINA |
| SD | 46 | SOUTH DAKOTA |
| TN | 47 | TENNESSEE |
| TX | 48 | TEXAS |
| UT | 49 | UTAH |
| VA | 51 | VIRGINIA |
| VI | 78 | VIRGIN ISLANDS |
| VT | 50 | VERMONT |
| WA | 53 | WASHINGTON |
| WI | 55 | WISCONSIN |
| WV | 54 | WEST VIRGINIA |
| WY | 56 | WYOMING |

# Appendix 2: SSA State Codes

|  |  |
| --- | --- |
| 01 | Alabama |
| 02 | Alaska |
| 03 | Arizona |
| 04 | Arkansas |
| 05 | California |
| 06 | Colorado |
| 07 | Connecticut |
| 08 | Delaware |
| 09 | District of Columbia |
| 10 | Florida |
| 11 | Georgia |
| 12 | Hawaii |
| 13 | Idaho |
| 14 | Illinois |
| 15 | Indiana |
| 16 | Iowa |
| 17 | Kansas |
| 18 | Kentucky |
| 19 | Louisiana |
| 20 | Maine |
| 21 | Maryland |
| 22 | Massachusetts |
| 23 | Michigan |
| 24 | Minnesota |
| 25 | Mississippi |
| 26 | Missouri |
| 27 | Montana |
| 28 | Nebraska |
| 29 | Nevada |
| 30 | New Hampshire |
| 31 | New Jersey |
| 32 | New Mexico |
| 33 | New York |
| 34 | North Carolina |
| 35 | North Dakota |
| 36 | Ohio |
| 37 | Oklahoma |
| 38 | Oregon |
| 39 | Pennsylvania |
| 40 | Puerto Rico |
| 41 | Rhode Island |
| 42 | South Carolina |
| 43 | South Dakota |
| 44 | Tennessee |
| 45 | Texas |
| 46 | Utah |
| 47 | Vermont |
| 48 | Virgin Islands |
| 49 | Virginia |
| 50 | Washington |
| 51 | West Virginia |
| 52 | Wisconsin |
| 53 | Wyoming |
| 54 | Africa |
| 55 | California |
| 56 | Canada & Islands |
| 57 | Central America and West Indies |
| 58 | Europe |
| 59 | Mexico |
| 60 | Oceania |
| 61 | Philippines |
| 62 | South America |
| 63 | U.S. Possessions |
| 64 | American Samoa |
| 65 | Guam |
| 66 | Commonwealth of the Northern Marianas Islands |
| 67 | Texas |
| 68 | Florida (eff. 10/2005) |
| 69 | Florida (eff. 10/2005) |
| 70 | Kansas (eff. 10/2005) |
| 71 | Louisiana (eff. 10/2005) |
| 72 | Ohio (eff. 10/2005) |
| 73 | Pennsylvania (eff. 10/2005) |
| 74 | Texas (eff. 10/2005) |
| 80 | Maryland (eff. 8/2000) |
| 97 | Northern Marianas |
| 98 | Guam |
| 99 | 99 = With 000 county code is American Samoa; otherwise unknown |

# Appendix 3: Code Snippets

## 3.1: Source to Source Vocab Query

/\*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')

## 3.2: Source to Standard Vocab Query

/\*Source to Standard\*/

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')