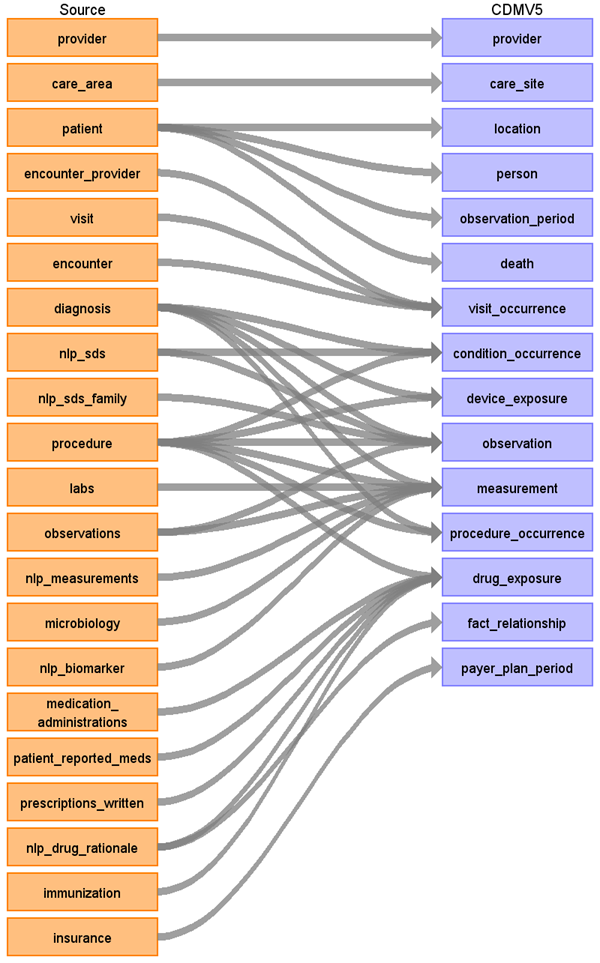
CDM ETL for Optum Oncology EHR

Chris Knoll, Ajit Londhe, Anthony Sena, Erica Voss

# Changelog

|  |  |
| --- | --- |
| Date | Comments |
| 6/10/2016 | Phase 1 ETL with the following tables: LOCATION, CARE\_SITE, PROVIDER, PERSON, OBSERVATION\_PERIOD, DEATH, VISIT\_OCCURRENCE, CONDITION\_OCCURRENCE, DRUG\_EXPOSURE |
| 7/13/2016 | Person table changes - no patients with unknown gender or unknown birth year  New death table rule to prevent death status for patients with data 32+ days after date\_of\_death |
| 7/20/2016 | Changed drug\_exposure\_start\_date mapping in medication\_administration → drug\_exposure to use admin\_date/admin\_time. Added mappings for drug\_exposure, device\_exposure, procedure\_occurence, measurement and observation |
| 8/11/2016 | Adjusted all STCM file names. Added CDM\_Domain\_Meta and CDM\_Source. |
| 8/24/2016 | Adjusted logic of prescriptions\_written → drug\_exposure (quantities, unit source value). |
| 9/15/2016 | Cleanup of logic and comments fields. More details on concept\_id mappings, referring to appendices as appropriate |

# Source Data Mapping Approach to CDMV5



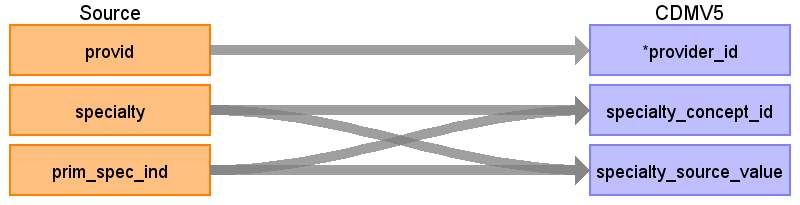
## Default Domains

The following table lists the default domain for a given source table/source column where a CDM row should be inserted with an unmapped concept ID (0) if no other domains were found for this source row.

|  |  |
| --- | --- |
| **Table.Column** | **Default Domain** |
| diagnosis.diagnosis\_cd | Condition |
| medication\_administratons.ndc | Drug |
| prescriptions\_written.ndc | Drug |
| immunization.ndc | Drug |
| procedure.proc\_code | Procedure |
| patient\_reported\_meds.ndc | Drug |
| labs.test\_name | Measurement |
| observations.obs\_type | Observation |
| nlp\_measurements | Measurement |

## Table name: provider

### Reading from provider



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| provider\_id | provid |  |  |
| provider\_name |  |  |  |
| npi |  |  |  |
| dea |  |  |  |
| specialty\_concept\_id | specialty |  | Use mapping JNJ\_OPTUM\_EHR\_SPCLTY.csv. |
| care\_site\_id |  |  |  |
| year\_of\_birth |  |  |  |
| gender\_concept\_id |  |  |  |
| provider\_source\_value |  |  |  |
| specialty\_source\_value | specialty |  |  |
| specialty\_source\_concept\_id | specialty |  | Use mapping JNJ\_OPTUM\_EHR\_SPCLTY.csv. |
| gender\_source\_value |  |  |  |
| gender\_source\_concept\_id |  |  |  |

Use the following query to select the provid, specialty to use for creating the provider table:

SELECT provid, specialty

FROM (

 select provid, specialty, prim\_spec\_ind, row\_number() over (      partition by provid, specialty order by prim\_spec\_ind desc) ordinal

from native.provider

) P

WHERE P.ordinal = 1

## Table name: care\_site

The Optum Oncology EHR does not contain site specific information. The care\_area information relates to divisions of organizations and not physical locations, so we are not producing a care\_site mapping for this ETL.

## Table name: location

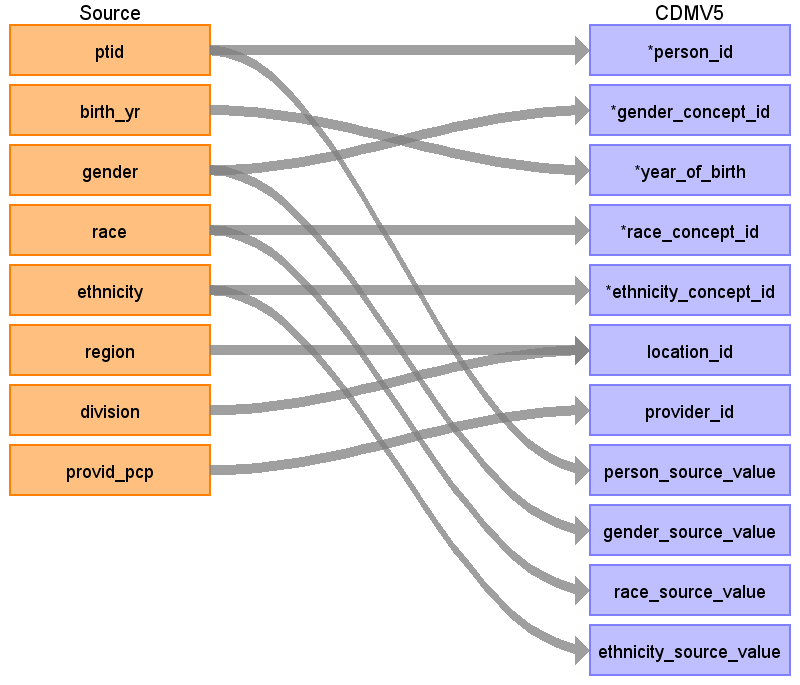
### Reading from patient



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| location\_id |  |  | Autogen |
| address\_1 |  |  |  |
| address\_2 |  |  |  |
| city |  |  |  |
| state |  |  |  |
| zip |  |  |  |
| county |  |  |  |
| location\_source\_value | division  region | {region}\_{division} | Concatenate region\_division region to create unique location\_id |

## Table name: person

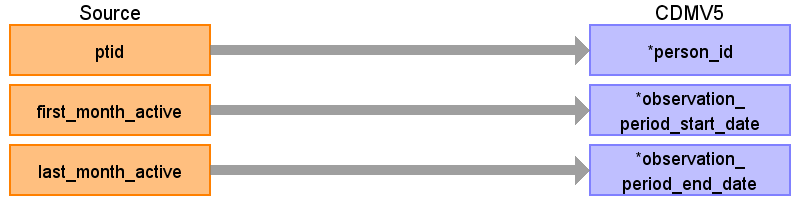
### Reading from patient



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| person\_id | ptid |  | Remove ‘PT’ prefix: cast(replace(ptid, 'PT','') as bigint) |
| gender\_concept\_id | gender | If gender = ‘Unknown’ then remove  Else if gender = 'Male' then 8507  else if gender = 'Female' then 8532 |  |
| year\_of\_birth | birth\_yr | if birth\_yr == 'Unknown' then remove  Else if birth\_yr <> 'Unknown' then  if birth\_yr = '1927 and earlier' then 1927  else birth\_yr |  |
| month\_of\_birth |  |  |  |
| day\_of\_birth |  |  |  |
| time\_of\_birth |  |  |  |
| race\_concept\_id | race | if 'Caucasian' then 8527  else if 'African American' then 8516  else if 'Asian' then 8515  else 0 |  |
| ethnicity\_concept\_id | ethnicity | if 'Hispanic' then 38003563  else 0 |  |
| location\_id | region  division | Look up from Location table |  |
| provider\_id | provid\_pcp | Look up from Provider table. In the case when the value is “” in the source, set the provider\_id to NULL. |  |
| care\_site\_id |  |  |  |
| person\_source\_value | ptid |  |  |
| gender\_source\_value | gender |  |  |
| gender\_source\_concept\_id |  |  |  |
| race\_source\_value | race |  |  |
| race\_source\_concept\_id |  |  |  |
| ethnicity\_source\_value | ethnicity |  |  |
| ethnicity\_source\_concept\_id |  |  |  |

## Table name: observation\_period

### Reading from patient



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| observation\_period\_id |  |  |  |
| person\_id | ptid |  |  |
| observation\_period\_start\_date | first\_month\_active | First day of first\_month\_active |  |
| observation\_period\_end\_date | last\_month\_active | Last day of last\_month\_active  Add 1 month to last month active |  |
| period\_type\_concept\_id | 38000280 |  | Observation recorded from EHR |

## Table name: death

Reading from patient



**A person’s death should be removed if any data available > 30 days after date\_of\_death**

|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| person\_id |  |  |  |
| death\_date | date\_of\_death | if patient.date\_of\_death is not null, then use the last day of the month of death: Same logic as observation\_period\_end\_date:  Add 1 to month of death and subtract 1 day. Example: month of death of 200402 becomes Feb + 1 month = March 1st - 1 day = Feb 29th, 2004. |  |
| death\_type\_concept\_id |  |  |  |
| cause\_concept\_id |  |  |  |
| cause\_source\_value |  |  |  |
| cause\_source\_concept\_id |  |  |  |

## Table name: visit\_occurrence

When creating visits, create a temporary mapping table named **TEMP\_VISIT** to associate encids from the native to VISIT\_OCCURRENCE in cdm. This lookup will be needed when ETLing the CONDITION\_OCCURRENCE, DRUG\_EXPOSRE, PROCEDURE\_OCCURRENCE, and any native table which references an encounter\_id that maps into a CDM table with a visit\_id.

**TEMP\_VISIT table query**

select Q.visitid source\_visit\_id, Q.encid cdm\_visit\_id, e.encid as source\_encounterid

from

(

select v.visitid, e.encid, row\_number() over (partition by v.visitid order by e.interaction\_date) as ordinal

from native.visit v

join native.encounter e on v.visitid = e.visitid

) Q

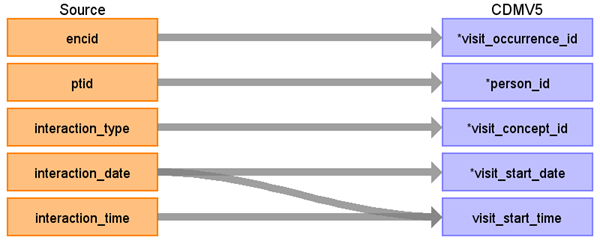
JOIN native.encounter e on e.visitid = Q.visitid

where Q.ordinal = 1

### Reading from encounter

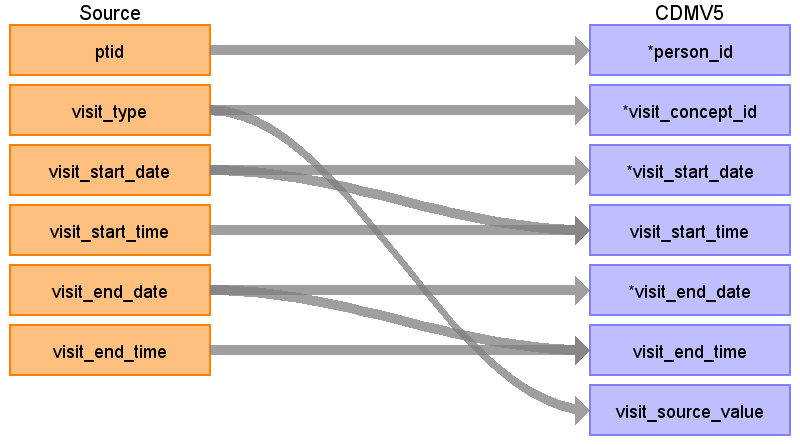
**Note: when pulling in encounters, only pull in encounters not associated with a visit:**

LEFT JOIN visit v on e.visitid = v.visitid where v.visitid is null



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| visit\_occurrence\_id | encid |  |  |
| person\_id | ptid |  |  |
| visit\_concept\_id | interaction\_type | STCM Lookup JNJ\_OPTUM\_EHR\_VISIT.csv |  |
| visit\_start\_date | interaction\_date |  |  |
| visit\_start\_time | interaction\_time |  |  |
| visit\_end\_date | interaction\_date |  |  |
| visit\_end\_time |  |  |  |
| visit\_type\_concept\_id |  | 44818518 | Visit derived from EHR record |
| provider\_id | encid | Look up in Provider table using encounter\_provider.provid | When joining to the provider table, employ the following logic to ensure only 1 provider is associated with the visit:   * If there is > 1 provider, prioritize the provider where provider\_role == ‘ATTENDING’.   + If there is > 1 record after applying this filter, sort by the specialty and pick the first. * If there is > 1 provider and 0 meet the criteria of provider\_role == ‘ATTENDING’, sort by the role & specialty and pick the first. |
| care\_site\_id |  |  |  |
| visit\_source\_value | encid |  |  |
| visit\_source\_concept\_id | interaction\_type | STCM Lookup JNJ\_OPTUM\_EHR\_VISIT.csv |  |

### Reading from visit



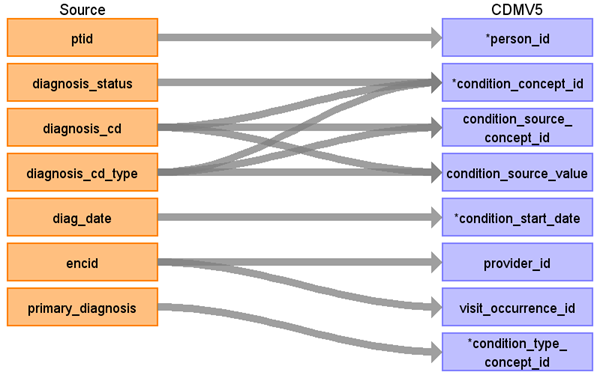
|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| visit\_occurrence\_id | visitid | Lookup visitid in TEMP\_VISIT and use cdm\_visit\_id |  |
| person\_id | ptid |  |  |
| visit\_concept\_id | visit\_type | Case ‘Inpatient’ : 9201 Case ‘Emergency Patient’: 9203 Case ‘Observation Patient’: 9201  ELSE: 0 |  |
| visit\_start\_date | visit\_start\_date |  |  |
| visit\_start\_time | visit\_start\_time | Use the date portion from visit\_start\_date |  |
| visit\_end\_date | visit\_end\_date |  |  |
| visit\_end\_time | visit\_end\_time | Use the date portion from visit\_end\_date |  |
| visit\_type\_concept\_id |  | 44818518 | Visit derived from EHR record |
| provider\_id | visitid | Use visitid to first encounter.encid associated, then to encounter\_provider.provid | To avoid duplication, apply the same logic described when transforming the encounter table to visit\_occurrence above. |
| care\_site\_id |  |  |  |
| visit\_source\_value | visit\_type |  |  |
| visit\_source\_concept\_id |  | Use first encid associated to visit. |  |

## Table name: condition\_occurrence

Any row mapped to a target concept must be in the Condition domain when inserting into CONDITION\_OCCURRENCE.

**This is the only table where duplicate records should be collapsed.**

### Reading from diagnosis

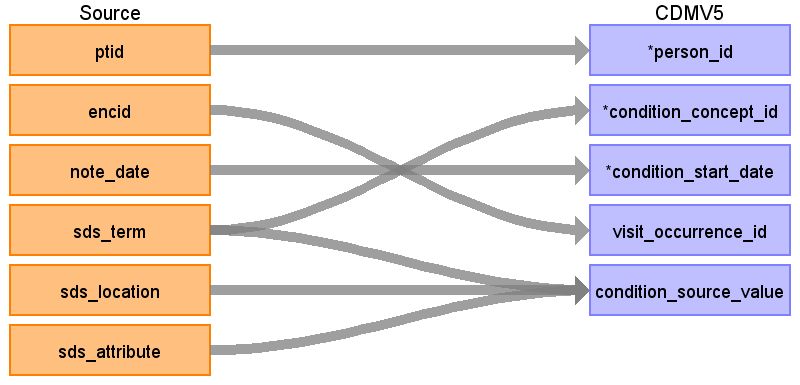


**Only use diagnosis records with DIAGNOSIS\_STATUS = 'Diagnosis of'**

|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| condition\_occurrence\_id |  | autogen |  |
| person\_id | ptid |  |  |
| condition\_concept\_id | diagnosis\_cd\_type  Diagnosis\_cd  diagnosis\_status | Use Appendix 1.1.2.  If diagnosis\_status = ‘Diagnosis of’  Then  { Case diagnosis\_cd\_type = ‘ICD9’:  Use lookup for ICD9CM Case diagnosis\_cd\_type = ‘ICD10’:  Use lookup for ICD10CM Case diagnosis\_cd\_type = ‘SNOMED’:  Use lookup for SNOMED CASE ‘’:   Lookup code in ICD9CM, ICD10CM, SNOMED CASE ELSE: 0 }  Else exclude record | For diagnosis\_cd\_type = ICD9, strip dot from lookup  For diagnosis\_cd\_type = ICD10, leave dots in lookup. |
| condition\_start\_date | diag\_date |  |  |
| condition\_end\_date |  |  |  |
| condition\_type\_concept\_id | primary\_diagnosis | If primary\_diagnosis = ‘1’ then 44786627 Else 44786629 |  |
| stop\_reason |  |  |  |
| provider\_id | encid | join to encounter\_provider to attempt to lookup the provid | To avoid duplication, apply the same logic described when transforming the encounter table to visit\_occurrence above. |
| visit\_occurrence\_id | encid | If encid is found in TEMP\_VISIT, use cdm\_visit\_id, else use encid |  |
| condition\_source\_value | diagnosis\_cd  diagnosis\_cd\_type | Use diagnosis\_cd if available. Otherwise use diagnosis\_cd\_type |  |
| condition\_source\_concept\_id | diagnosis\_cd  diagnosis\_cd\_type | Use Appendix 1.1.1, and obtain the source\_vocabulary\_id and target\_vocabulary\_id for the where clause using this:  If diagnosis\_status = ‘Diagnosis of’  Then  {  Case diagnosis\_cd\_type = ‘ICD9’:  Use lookup for ICD9CM  Case diagnosis\_cd\_type = ‘ICD10’:  Use lookup for ICD10CM  Case diagnosis\_cd\_type = ‘SNOMED’:  Use lookup for SNOMED  CASE ‘’:  Lookup code in ICD9CM, ICD10CM, SNOMED  CASE ELSE: 0  }  else exclude record | For diagnosis\_cd\_type = ICD9, strip dot from lookup  For diagnosis\_cd\_type = ICD10, leave dots in lookup. |

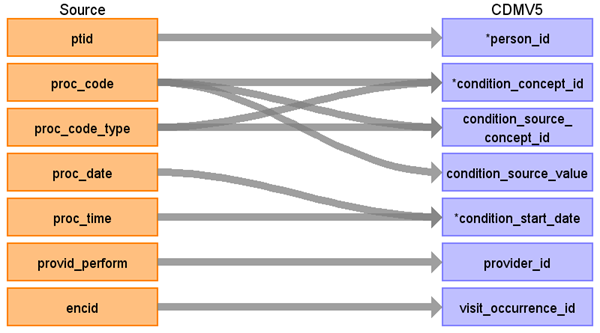
### Reading from nlp\_sds (DO NOT LOAD)

Phase 1 note: do not perform this map into Condition\_occurrence for phase 1.



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| condition\_occurrence\_id |  |  |  |
| person\_id | ptid |  |  |
| condition\_concept\_id | sds\_term |  |  |
| condition\_start\_date | note\_date |  |  |
| condition\_end\_date |  |  |  |
| condition\_type\_concept\_id |  |  |  |
| stop\_reason |  |  |  |
| provider\_id |  |  |  |
| visit\_occurrence\_id | encid |  |  |
| condition\_source\_value | sds\_term  sds\_location  sds\_attribute |  |  |
| condition\_source\_concept\_id |  |  |  |

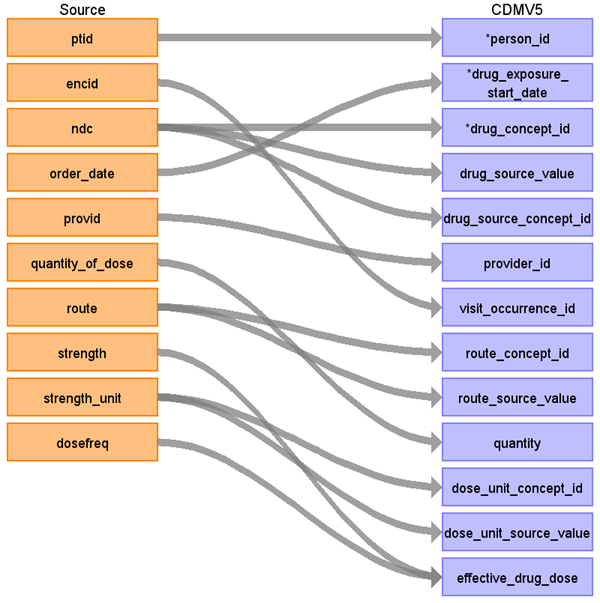
### Reading from procedure



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| condition\_occurrence\_id |  | Autogen |  |
| person\_id | ptid |  |  |
| condition\_concept\_id | proc\_code  proc\_code\_type | Use Appendix 1.1.2  Case proc\_code\_type: “CPT4” : use CPT4 Vocab “HCPCS”: use HCPCS Vocab “ICD10”: user ICD10PCS “ICD9”: use ICD9Proc Vocab Else: 0 |  |
| condition\_start\_date | proc\_date |  |  |
| condition\_source\_value | proc\_code |  |  |
| condition\_source\_concept\_id | proc\_code  proc\_code\_type | Use Appendix 1.1.1.  Case proc\_code\_type: “CPT4” : use CPT4 Vocab “HCPCS”: use HCPCS Vocab “ICD9”: use ICD9Proc Vocab Else: 0 |  |
| provider\_id | provid\_perform | Use Provider Lookup | To avoid duplication, apply the same logic described when transforming the encounter table to visit\_occurrence above. |
| visit\_occurrence\_id | encid |  |  |
| condition\_end\_date |  |  |  |
| condition\_type\_concept\_id |  | 0 |  |
| stop\_reason |  |  |  |

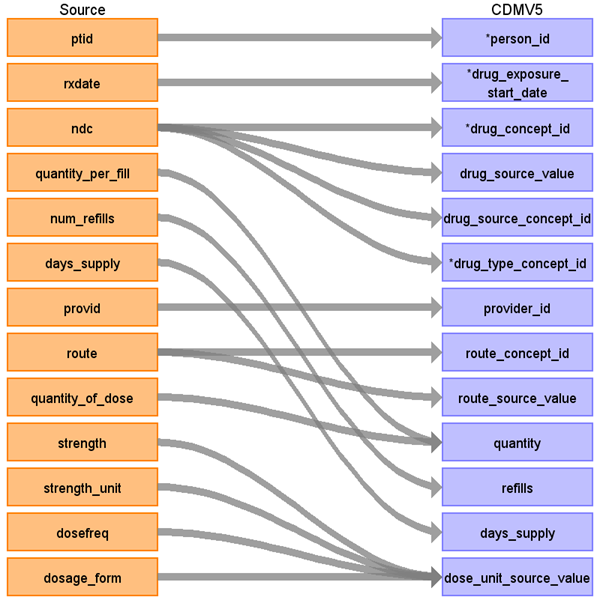
## Table name: drug\_exposure

### Reading from medication\_administrations



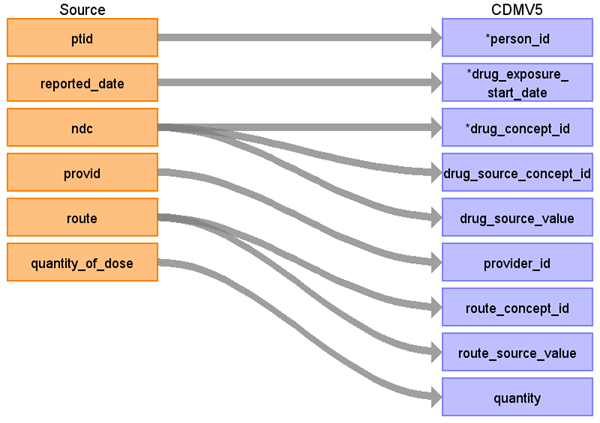
|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| drug\_exposure\_id |  | autogen |  |
| person\_id | ptid |  |  |
| drug\_concept\_id | ndc | Use Appendix 1.1.2  WHERE Vocabulary\_ID = ‘NDC’ |  |
| drug\_exposure\_start\_date | order\_date |  |  |
| drug\_exposure\_end\_date |  |  |  |
| drug\_type\_concept\_id |  | 38000180 | Inpatient Administration |
| stop\_reason |  |  |  |
| refills |  |  |  |
| quantity | quantity\_of\_dose |  |  |
| days\_supply |  |  |  |
| sig |  |  |  |
| route\_concept\_id | route | STCM Lookup JNJ\_OPTUM\_EHR\_ROUTE.csv |  |
| effective\_drug\_dose | quantity\_of\_dose  dosefreq  strength |  | Leave empty for Phase 1, to revisit in phase 2 |
| dose\_unit\_concept\_id | strength\_unit |  | Leave empty for Phase 1, to revisit in phase 2 |
| lot\_number |  |  |  |
| provider\_id | Provid  encid | If the provid is not specified, look at the encounter associated with this record. If there is a provider specified, used that value here. | To avoid duplication, apply the same logic described when transforming the encounter table to visit\_occurrence above. |
| visit\_occurrence\_id | encid |  |  |
| drug\_source\_value | ndc |  |  |
| drug\_source\_concept\_id | ndc |  |  |
| route\_source\_value | route |  |  |
| dose\_unit\_source\_value | strength\_unit |  |  |

### Reading from prescriptions\_written



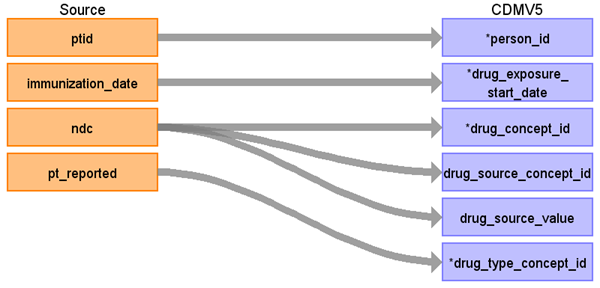
|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| drug\_exposure\_id |  |  |  |
| person\_id | ptid |  |  |
| drug\_exposure\_start\_date | rxdate |  |  |
| drug\_concept\_id | ndc |  |  |
| drug\_source\_value | ndc |  |  |
| drug\_source\_concept\_id | ndc |  |  |
| drug\_type\_concept\_id | ndc | 38000177 |  |
| provider\_id | provid |  | To avoid duplication, apply the same logic described when transforming the encounter table to visit\_occurrence above. |
| visit\_occurrence\_id |  |  |  |
| route\_concept\_id | route | Lookup route value in the STCM Lookup JNJ\_OPTUM\_EHR\_ROUTE |  |
| route\_source\_value | route |  |  |
| quantity | quantity\_per\_fill  quantity\_of\_dose | IF quantity\_per\_fill != null THEN  cdm.quantity = split(quantity\_per\_fill).first()  ELSE IF quantity\_of\_dose != null THEN  cdm.quantity = quantity\_per\_dose  ELSE 0; |  |
| refills | num\_refills |  |  |
| days\_supply | days\_supply |  |  |
| dose\_unit\_concept\_id |  |  |  |
| dose\_unit\_source\_value | strength  strength\_unit  dosage\_form  dosefreq | Concatenate: strength + strength\_unit + dosage\_form + dosefreq with a semi-colon for delimiter |  |
| effective\_drug\_dose |  |  |  |
| stop\_reason |  |  |  |
| sig |  |  |  |
| lot\_number |  |  |  |
| drug\_exposure\_end\_date |  |  |  |

### Reading from patient\_reported\_meds



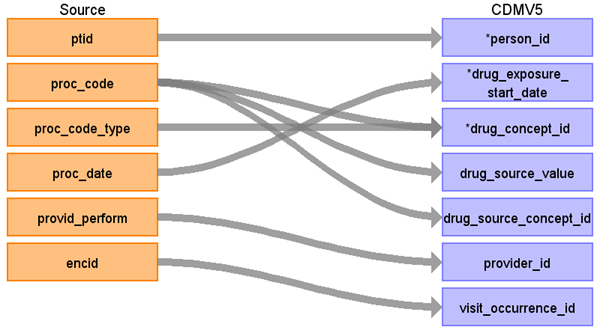
|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| drug\_exposure\_id |  |  |  |
| person\_id | ptid |  |  |
| drug\_exposure\_start\_date | reported\_date |  |  |
| drug\_concept\_id | ndc |  |  |
| drug\_source\_concept\_id | ndc |  |  |
| drug\_source\_value | ndc |  |  |
| drug\_type\_concept\_id |  | 44787730 | Patient Self-Reported Medication |
| provider\_id | provid |  | To avoid duplication, apply the same logic described when transforming the encounter table to visit\_occurrence above. |
| route\_concept\_id | route | Lookup route value in the STCM Lookup JNJ\_OPTUM\_EHR\_ROUTE |  |
| route\_source\_value | route |  |  |
| quantity | quantity\_of\_dose |  |  |
| refills |  |  |  |
| days\_supply |  |  |  |
| dose\_unit\_concept\_id |  |  |  |
| dose\_unit\_source\_value |  |  |  |
| effective\_drug\_dose |  |  |  |
| visit\_occurrence\_id |  |  |  |
| stop\_reason |  |  |  |
| sig |  |  |  |
| lot\_number |  |  |  |
| drug\_exposure\_end\_date |  |  |  |

### Reading from immunization



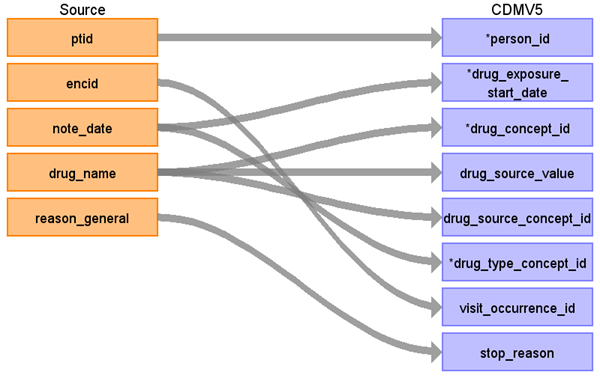
|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| drug\_exposure\_id |  |  |  |
| person\_id | ptid |  |  |
| drug\_concept\_id | ndc  pt\_reported | where pt\_reported = 'N' |  |
| drug\_exposure\_start\_date | immunization\_date |  |  |
| drug\_exposure\_end\_date |  |  |  |
| drug\_type\_concept\_id |  | 43542358 | Physician administered drug (identified from EHR observation) |
| stop\_reason |  |  |  |
| refills |  |  |  |
| quantity |  |  |  |
| days\_supply |  |  |  |
| sig |  |  |  |
| route\_concept\_id |  |  |  |
| effective\_drug\_dose |  |  |  |
| dose\_unit\_concept\_id |  |  |  |
| lot\_number |  |  |  |
| provider\_id |  |  | To avoid duplication, apply the same logic described when transforming the encounter table to visit\_occurrence above. |
| visit\_occurrence\_id |  |  |  |
| drug\_source\_value | ndc |  |  |
| drug\_source\_concept\_id | ndc |  |  |
| route\_source\_value |  |  |  |
| dose\_unit\_source\_value |  |  |  |

### Reading from procedure



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| drug\_exposure\_id |  |  |  |
| person\_id | ptid |  |  |
| drug\_exposure\_start\_date | proc\_date |  |  |
| drug\_concept\_id | proc\_code  proc\_code\_type |  | Only map rows where the proc\_code and proc\_code\_type map to a concept in the vocabulary that has domain\_id = ‘drug’ |
| drug\_source\_value | proc\_code |  |  |
| drug\_source\_concept\_id | proc\_code |  |  |
| drug\_type\_concept\_id |  | 38000275 | EHR order list entry |
| provider\_id | provid\_perform |  | To avoid duplication, apply the same logic described when transforming the encounter table to visit\_occurrence above. |
| visit\_occurrence\_id | encid |  |  |
| route\_concept\_id |  |  |  |
| route\_source\_value |  |  |  |
| quantity |  |  |  |
| refills |  |  |  |
| days\_supply |  |  |  |
| dose\_unit\_concept\_id |  |  |  |
| dose\_unit\_source\_value |  |  |  |
| effective\_drug\_dose |  |  |  |
| stop\_reason |  |  |  |
| sig |  |  |  |
| lot\_number |  |  |  |
| drug\_exposure\_end\_date |  |  |  |

### Reading from nlp\_drug\_rationale



Only map the rows that match the following query:

select \*

from nlp\_drug\_rationale

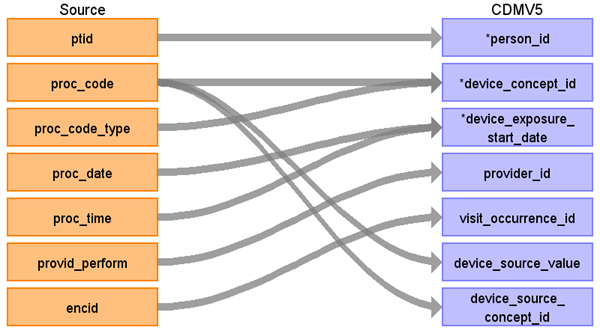
where (drug\_action is null or drug\_action not in ('STOP', 'TAKE.NOT', 'ON.NOT', 'SWITCH FROM', 'AVOID', 'REMOVE', 'REMAIN OFF', 'START.NOT', 'GET.NOT', 'TRY.NOT'))

AND (sentiment is null or sentiment not in ('REJECT', 'NEED.NOT', 'WANT.NOT'))

|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| drug\_exposure\_id |  |  |  |
| person\_id | ptid |  |  |
| drug\_exposure\_start\_date | note\_date |  |  |
| drug\_concept\_id | drug\_name | STCM Lookup JNJ\_OPTUM\_NLP\_DRUG.csv |  |
| drug\_source\_value | drug\_name |  |  |
| drug\_source\_concept\_id | drug\_name |  |  |
| drug\_type\_concept\_id | note\_date | 38000281 = Observation recorded from EHR with text result |  |
| provider\_id |  |  |  |
| visit\_occurrence\_id | encid |  |  |
| route\_concept\_id |  |  |  |
| route\_source\_value |  |  |  |
| quantity |  |  |  |
| refills |  |  |  |
| days\_supply |  |  |  |
| dose\_unit\_concept\_id |  |  |  |
| dose\_unit\_source\_value |  |  |  |
| effective\_drug\_dose |  |  |  |
| stop\_reason | reason\_general |  |  |
| sig |  |  |  |
| lot\_number |  |  |  |
| drug\_exposure\_end\_date |  |  |  |

## Table name: device\_exposure

### Reading from procedure





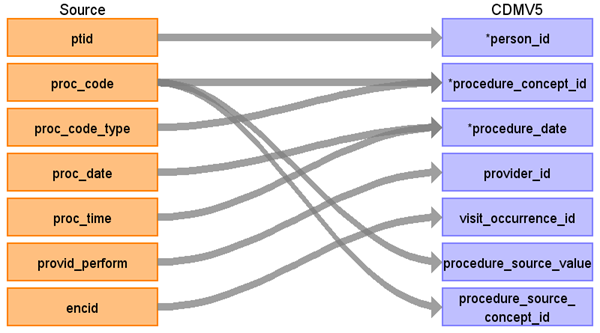
|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| device\_exposure\_id |  |  | Auto-incremented |
| person\_id | ptid |  |  |
| device\_concept\_id | proc\_code |  | Domain\_id = Device for those procedures that have device concepts in the proc\_code (mainly HCPCS) |
| device\_exposure\_start\_date | Proc\_date |  |  |
| device\_exposure\_end\_date |  |  |  |
| device\_type\_concept\_id |  |  |  |
| unique\_device\_id |  |  |  |
| quantity |  |  |  |
| provider\_id | provid |  | To avoid duplication, apply the same logic described when transforming the encounter table to visit\_occurrence above. |
| visit\_occurrence\_id | encid |  | Lookup using visit\_occurrence mapping logic |
| device\_source\_value | proc\_code |  |  |
| device\_source\_concept\_id | Proc\_code mapped to concept\_id |  | Source code mapped to vocabulary |

### Reading from diagnosis

|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| device\_exposure\_id |  |  |  |
| person\_id | ptid |  |  |
| device\_concept\_id | diagnosis\_cd  diagnosis\_status  diagnosis\_cd\_type |  | Domain\_id = Device for those procedures that have device concepts in the proc\_code (mainly HCPCS) |
| device\_exposure\_start\_date | diag\_date |  |  |
| device\_exposure\_end\_date |  |  |  |
| device\_type\_concept\_id | primary\_diagnosis |  | If primary\_diagnosis = ‘1’ then 44786627  Else 44786629 |
| unique\_device\_id |  |  |  |
| quantity |  |  |  |
| provider\_id | Encid -> provider\_id |  | Use this field to map up to the encounter\_provider. To avoid duplication, apply the same logic described when transforming the encounter table to visit\_occurrence above. |
| visit\_occurrence\_id | encid |  | Use the visit\_occurrence mapping logic |
| device\_source\_value | diagnosis\_cd |  |  |
| device\_source\_concept\_id | diagnosis\_cd |  |  |

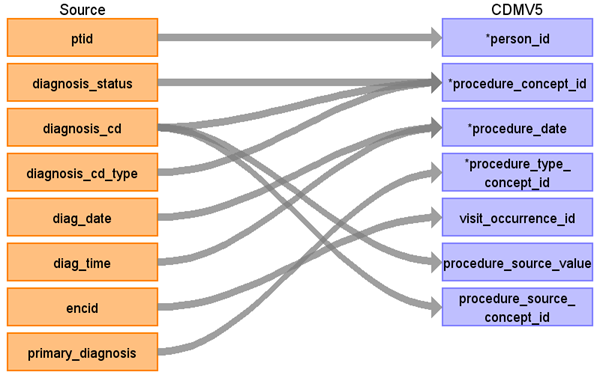
## Table name: procedure\_occurrence

### Reading from procedure



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| procedure\_occurrence\_id |  |  |  |
| person\_id | ptid |  |  |
| procedure\_concept\_id | proc\_code  proc\_code\_type |  | Lookup to the vocabulary using the proc\_code\_type to find the proper vocabulary\_id |
| procedure\_date | proc\_date |  |  |
| procedure\_type\_concept\_id | 38000275 |  | EHR order list entry |
| modifier\_concept\_id |  |  |  |
| quantity |  |  |  |
| provider\_id | provid\_perform |  | To avoid duplication, apply the same logic described when transforming the encounter table to visit\_occurrence above. |
| visit\_occurrence\_id | encid |  |  |
| procedure\_source\_value | proc\_code |  |  |
| procedure\_source\_concept\_id | proc\_code |  |  |
| qualifier\_source\_value |  |  |  |

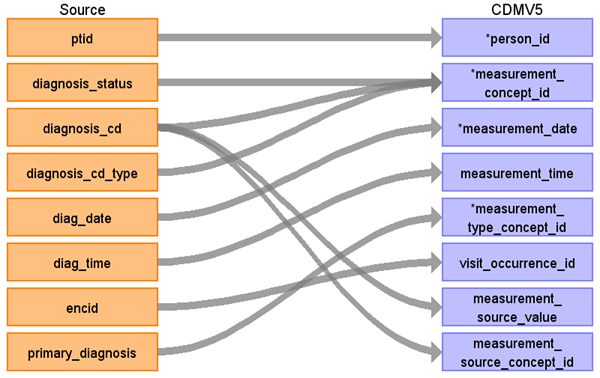
### Reading from diagnosis



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| procedure\_occurrence\_id |  |  |  |
| person\_id | ptid |  |  |
| procedure\_concept\_id | diagnosis\_status  diagnosis\_cd  diagnosis\_cd\_type |  | Diagnosis\_status = ‘Diagnosis Of’ and diagnosis\_cd must map to domain = procedure  Where the target vocabulary\_id = diagnaosis\_cd\_type. |
| procedure\_date | diag\_date  diag\_time |  |  |
| procedure\_type\_concept\_id | 42865906 |  | ‘Condition Procedure’ |
| modifier\_concept\_id |  |  |  |
| quantity |  |  |  |
| provider\_id |  |  | To avoid duplication, apply the same logic described when transforming the encounter table to visit\_occurrence above. |
| visit\_occurrence\_id | encid |  |  |
| procedure\_source\_value | diagnosis\_cd |  |  |
| procedure\_source\_concept\_id | diagnosis\_cd |  |  |
| qualifier\_source\_value |  |  |  |

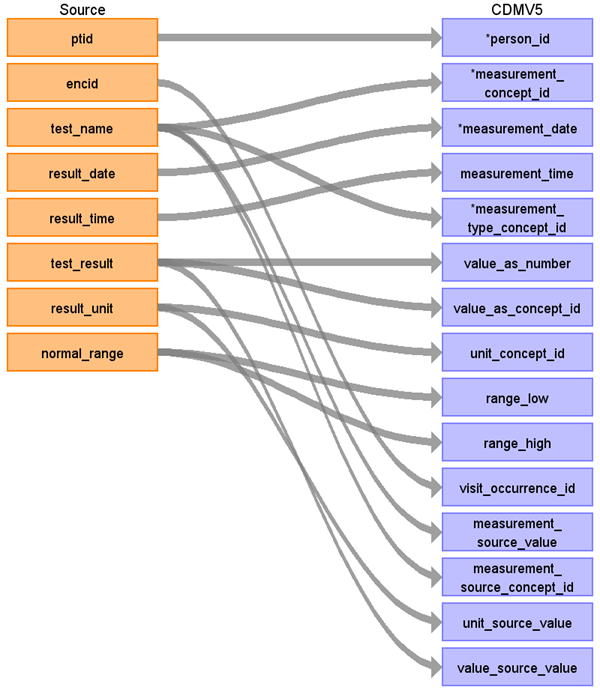
## Table name: measurement

### Reading from diagnosis



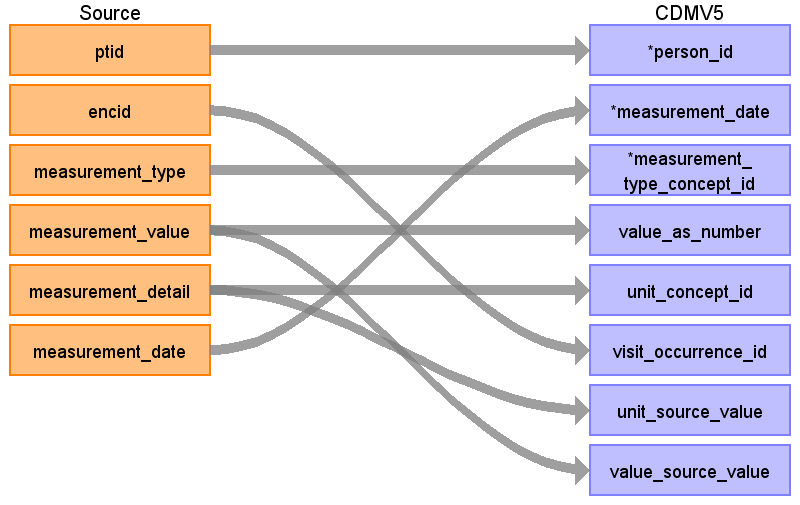
|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| measurement\_id |  |  |  |
| person\_id | ptid |  |  |
| measurement\_concept\_id | diagnosis\_status  diagnosis\_cd  diagnosis\_cd\_type | Use Appendix 1.1.2 to obtain the source\_vocabulary\_id for the where clause  If diagnosis\_status = ‘Diagnosis of’  Then  {  Case diagnosis\_cd\_type = ‘ICD9’:  Use lookup for ICD9CM  Case diagnosis\_cd\_type = ‘ICD10’:  Use lookup for ICD10CM  Case diagnosis\_cd\_type = ‘SNOMED’:  Use lookup for SNOMED  CASE ‘’:  Lookup code in ICD9CM, ICD10CM, SNOMED  CASE ELSE: 0  } | For diagnosis\_cd\_type = ICD9, strip dot from lookup  For diagnosis\_cd\_type = ICD10, leave dots in lookup. |
| measurement\_date | diag\_date |  |  |
| measurement\_time | diag\_time |  |  |
| measurement\_type\_concept\_id | primary\_diagnosis | if primary\_diagnosis = ‘1’ then 44786627  else 44786629 |  |
| operator\_concept\_id |  |  |  |
| value\_as\_number |  |  |  |
| value\_as\_concept\_id |  |  |  |
| unit\_concept\_id |  |  |  |
| range\_low |  |  |  |
| range\_high |  |  |  |
| provider\_id |  |  | To avoid duplication, apply the same logic described when transforming the encounter table to visit\_occurrence above. |
| visit\_occurrence\_id | encid |  |  |
| measurement\_source\_value | diagnosis\_cd |  |  |
| measurement\_source\_concept\_id | diagnosis\_cd | Use Appendix 1.1.1 to obtain the source\_vocabulary\_id for the where clause  If diagnosis\_status = ‘Diagnosis of’  Then  {  Case diagnosis\_cd\_type = ‘ICD9’:  Use lookup for ICD9CM  Case diagnosis\_cd\_type = ‘ICD10’:  Use lookup for ICD10CM  Case diagnosis\_cd\_type = ‘SNOMED’:  Use lookup for SNOMED  CASE ‘’:  Lookup code in ICD9CM, ICD10CM, SNOMED  CASE ELSE: 0  } | For diagnosis\_cd\_type = ICD9, strip dot from lookup  For diagnosis\_cd\_type = ICD10, leave dots in lookup. |
| unit\_source\_value |  |  |  |
| value\_source\_value |  |  |  |

### Reading from labs



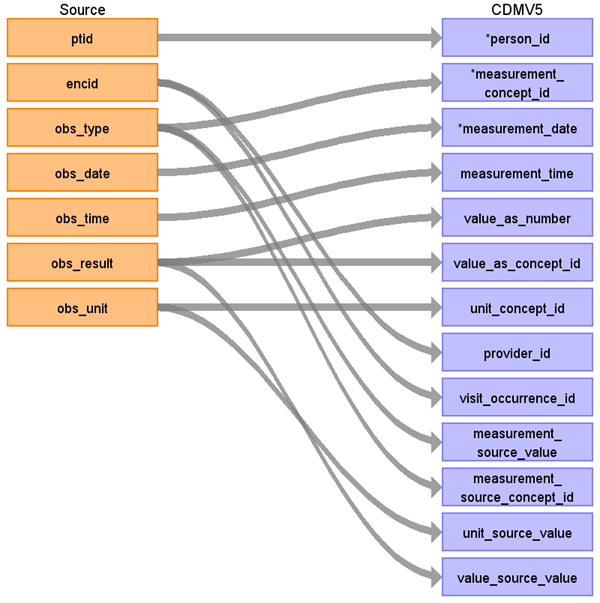
|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| measurement\_id |  |  |  |
| person\_id | ptid |  |  |
| measurement\_concept\_id | test\_name |  | STCM Lookup JNJ\_OPTUM\_EHR\_LABNAME.csv |
| measurement\_date | result\_date |  |  |
| measurement\_time | result\_time |  |  |
| measurement\_type\_concept\_id | 44818702 |  | Lab result |
| operator\_concept\_id | relative\_indicator |  | When relative\_indicator is NULL/empty, leave the value as NULL. These will map to domain\_id = 'Meas Value Operator' in the vocabulary.  CASE WHEN relative\_incicator == ‘<=’ THEN 4171754  CASE WHEN relative\_incicator == ‘>=’ THEN 4171755  CASE WHEN relative\_incicator == ‘<’ THEN 4171756  CASE WHEN relative\_incicator == ‘=’ THEN 4172703  CASE WHEN relative\_incicator == ‘>’ THEN 4172704  ELSE 0 |
| value\_as\_number | test\_result |  | Only do this if test\_result is numeric |
| value\_as\_concept\_id | test\_result |  | If test\_result is NOT NUMERIC THEN  STCM Lookup JNJ\_OPTUM\_EHR\_LABRES.csv |
| unit\_concept\_id | result\_unit |  | Map to UCUM vocabulary using a CASE-SENSITIVE matching |
| range\_low | normal\_range |  | Parse lower bound (split on hyphen), first piece |
| range\_high | normal\_range |  | Parse upper bound (split on hyphen), second piece |
| provider\_id | encid |  | Look up provider from the encounter. To avoid duplication, apply the same logic described when transforming the encounter table to visit\_occurrence above. |
| visit\_occurrence\_id | encid |  |  |
| measurement\_source\_value | test\_name |  |  |
| measurement\_source\_concept\_id | test\_name | 0 |  |
| unit\_source\_value | result\_unit |  |  |
| value\_source\_value | test\_result |  |  |

### Reading from nlp\_measurements



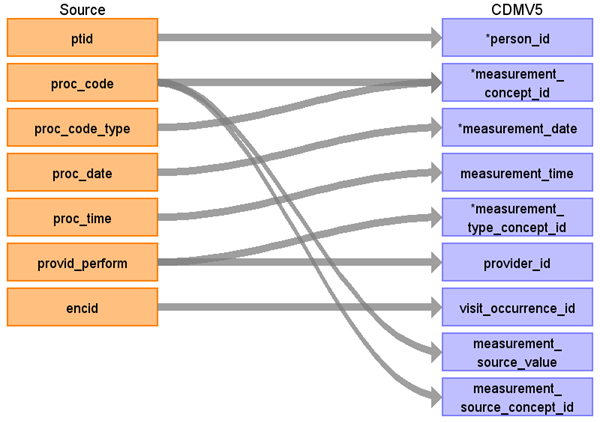
|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| measurement\_id |  |  |  |
| person\_id | ptid |  |  |
| measurement\_concept\_id | measurement\_type |  | STCM Lookup JNJ\_OPTUM\_EHR\_LABNAM.csv  This STCM Lookup should be a consolidated list of text values found in the source “labs” and “nlp\_measurements” tables so we don’t have duplicate concept\_ids. |
| measurement\_date | measurement\_date |  |  |
| measurement\_time |  |  |  |
| measurement\_type\_concept\_id | 45754907 |  | Derived value |
| operator\_concept\_id |  |  |  |
| value\_as\_number | measurement\_value |  | Only do this if measurement\_value is numeric |
| value\_as\_concept\_id | measurement\_value |  |  |
| unit\_concept\_id | measurement\_detail |  | Map to UCUM vocabulary using a CASE-SENSITIVE matching |
| range\_low |  |  |  |
| range\_high |  |  |  |
| provider\_id |  |  | To avoid duplication, apply the same logic described when transforming the encounter table to visit\_occurrence above. |
| visit\_occurrence\_id | encid |  |  |
| measurement\_source\_value |  |  |  |
| measurement\_source\_concept\_id |  |  |  |
| unit\_source\_value | measurement\_detail |  |  |
| value\_source\_value | measurement\_value |  |  |

### Reading from observations



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| measurement\_id |  |  |  |
| person\_id | ptid |  |  |
| measurement\_concept\_id | obs\_type |  | STCM Lookup JNJ\_OPTUM\_EHR\_OBTYPE.csv  Create a source to concept map (STCM) for the native observations that should map to the CDM measurement table. Then when mapping rows to the the CDM Measurement table, only push over the rows that are in the STCM via an inner join. When the target is the CDM observation table, do a left join and bring over the values that are NOT in the STCM. |
| measurement\_date | obs\_date |  |  |
| measurement\_time | obs\_time |  |  |
| measurement\_type\_concept\_id | 45754907 |  | Derived value |
| operator\_concept\_id | 0 |  |  |
| value\_as\_number | obs\_result |  |  |
| value\_as\_concept\_id | obs\_result |  |  |
| unit\_concept\_id | obs\_unit |  |  |
| range\_low |  |  |  |
| range\_high |  |  |  |
| provider\_id | encid |  | Map encid to encounter\_provider. To avoid duplication, apply the same logic described when transforming the encounter table to visit\_occurrence above. |
| visit\_occurrence\_id | encid |  |  |
| measurement\_source\_value | obs\_type |  |  |
| measurement\_source\_concept\_id | obs\_type |  |  |
| unit\_source\_value | obs\_unit |  |  |
| value\_source\_value | obs\_result |  |  |

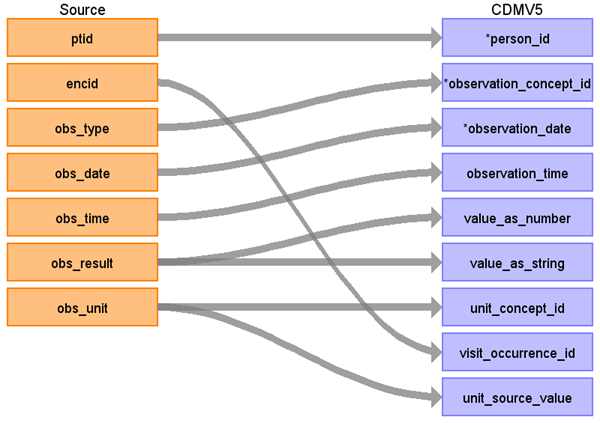
### Reading from procedure



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| measurement\_id |  |  |  |
| person\_id | ptid |  |  |
| measurement\_concept\_id | proc\_code  proc\_code\_type |  | Lookup to the vocabulary using the proc\_code\_type to find the proper vocabulary\_id |
| measurement\_date | proc\_date |  |  |
| measurement\_time | proc\_time |  |  |
| measurement\_type\_concept\_id | provid\_perform |  | 44818701  45754907    Derived value |
| operator\_concept\_id |  |  |  |
| value\_as\_number |  |  |  |
| value\_as\_concept\_id |  |  |  |
| unit\_concept\_id |  |  |  |
| range\_low |  |  |  |
| range\_high |  |  |  |
| provider\_id | provid\_perform |  | To avoid duplication, apply the same logic described when transforming the encounter table to visit\_occurrence above. |
| visit\_occurrence\_id | encid |  |  |
| measurement\_source\_value | proc\_code |  |  |
| measurement\_source\_concept\_id | proc\_code |  |  |
| unit\_source\_value |  |  |  |
| value\_source\_value |  |  |  |

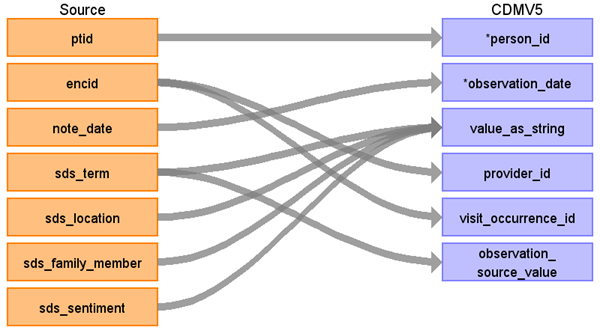
## Table name: observation

### Reading from observations



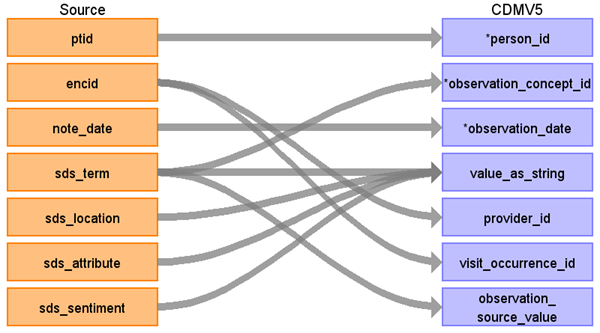
|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| observation\_id |  |  |  |
| person\_id | ptid |  |  |
| observation\_concept\_id | obs\_type | CASE WHEN ‘SMOKE’ THEN <concept\_id>  CASE WHEN ‘ALCOHOL’ THEN <concept\_id>  CASE WHEN ‘EXERCISE’ THEN <concept\_id>  ELSE 0 | STCM Lookup JNJ\_OPTUM\_EHR\_OBTYPE.csv  Create a source to concept map (STCM) for the native observations that should map to the CDM measurement table. Then when mapping rows to the the CDM Measurement table, only push over the rows that are in the STCM via an inner join. When the target is the CDM observation table, do a left join and bring over the values that are NOT in the STCM and use 0 for the concept id. |
| observation\_date | obs\_date |  |  |
| observation\_time | obs\_time |  |  |
| observation\_type\_concept\_id |  |  | 38000280 |
| value\_as\_number | obs\_result |  | When obs\_result can be treated as a decimal, place the value in this field |
| value\_as\_string | obs\_result |  | When the obs\_result cannot be treated as a decimal, please the value in this field. |
| value\_as\_concept\_id |  |  |  |
| qualifier\_concept\_id |  |  |  |
| unit\_concept\_id | obs\_unit |  | Map to UCUM vocabulary using a CASE-SENSITIVE matching |
| provider\_id | encid |  | To avoid duplication, apply the same logic described when transforming the encounter table to visit\_occurrence above. |
| visit\_occurrence\_id | encid |  |  |
| observation\_source\_value |  |  |  |
| observation\_source\_concept\_id |  |  |  |
| unit\_source\_value | obs\_unit |  |  |
| qualifier\_source\_value |  |  |  |

### Reading from nlp\_sds\_family



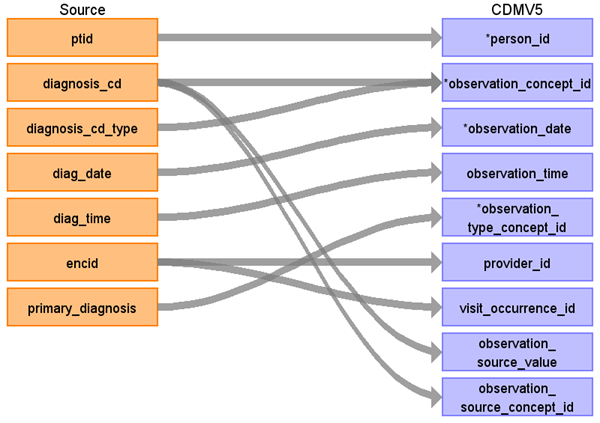
|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| observation\_id |  |  |  |
| person\_id | ptid |  |  |
| observation\_concept\_id |  |  |  |
| observation\_date | note\_date |  |  |
| observation\_time |  |  |  |
| observation\_type\_concept\_id |  | 38000280 |  |
| value\_as\_number |  |  |  |
| value\_as\_string | sds\_term  sds\_location  sds\_family\_member  sds\_sentiment | Concatenate the fields: sds\_term, sds\_location, sds\_family\_member and sds\_sentiment    Use a semi-colon as the delimiter |  |
| value\_as\_concept\_id |  |  |  |
| qualifier\_concept\_id |  |  |  |
| unit\_concept\_id |  |  |  |
| provider\_id | encid | To avoid duplication, apply the same logic described when transforming the encounter table to visit\_occurrence above. |  |
| visit\_occurrence\_id | encid |  |  |
| observation\_source\_value | sds\_term |  |  |
| observation\_source\_concept\_id |  |  |  |
| unit\_source\_value |  |  |  |
| qualifier\_source\_value |  |  |  |

### Reading from nlp\_sds



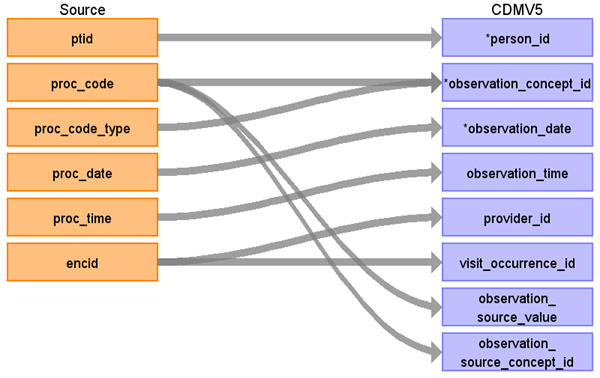
|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| observation\_id |  |  |  |
| person\_id | ptid |  |  |
| observation\_concept\_id | sds\_term |  |  |
| observation\_date | note\_date |  |  |
| observation\_time |  |  |  |
| observation\_type\_concept\_id |  | 38000280 |  |
| value\_as\_number |  |  |  |
| value\_as\_string | sds\_term  sds\_location  sds\_attribute  sds\_sentiment | Concatenate sds\_term + sds\_location + sds\_attribute + sds\_sentiment    Use a "semi-colon" for all concatenation |  |
| value\_as\_concept\_id |  |  |  |
| qualifier\_concept\_id |  |  |  |
| unit\_concept\_id |  |  |  |
| provider\_id | encid | To avoid duplication, apply the same logic described when transforming the encounter table to visit\_occurrence above. |  |
| visit\_occurrence\_id | encid |  |  |
| observation\_source\_value | sds\_term | NLP\_SDS |  |
| observation\_source\_concept\_id |  |  |  |
| unit\_source\_value |  |  |  |
| qualifier\_source\_value |  |  |  |

### Reading from diagnosis



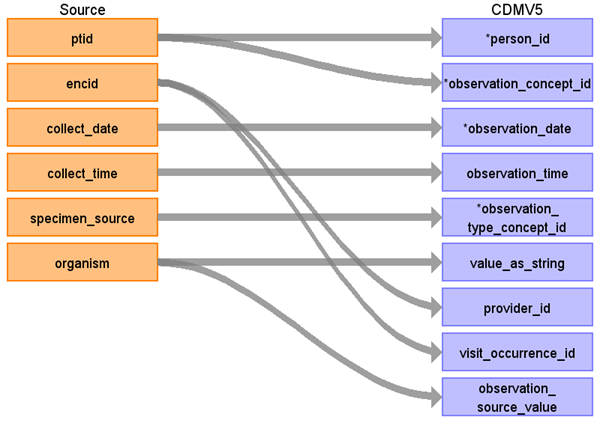
|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| observation\_id |  |  |  |
| person\_id | ptid |  |  |
| observation\_concept\_id | diagnosis\_cd  diagnosis\_cd\_type |  |  |
| observation\_date | diag\_date |  |  |
| observation\_time | diag\_time |  |  |
| observation\_type\_concept\_id | primary\_diagnosis |  | If primary\_diagnosis = ‘1’ then 44786627  Else 44786629 |
| value\_as\_number |  |  |  |
| value\_as\_string |  |  |  |
| value\_as\_concept\_id |  |  |  |
| qualifier\_concept\_id |  |  |  |
| unit\_concept\_id |  |  |  |
| provider\_id | encid |  | To avoid duplication, apply the same logic described when transforming the encounter table to visit\_occurrence above. |
| visit\_occurrence\_id | encid |  |  |
| observation\_source\_value | diagnosis\_cd |  |  |
| observation\_source\_concept\_id | diagnosis\_cd |  |  |
| unit\_source\_value |  |  |  |
| qualifier\_source\_value |  |  |  |

### Reading from procedure



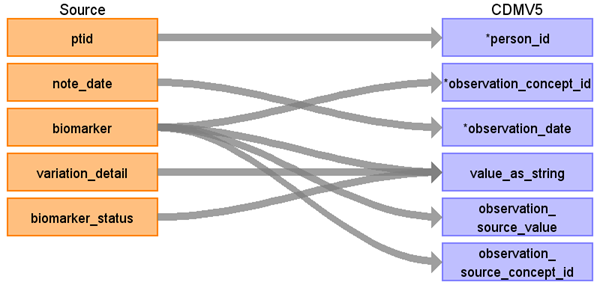
|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| observation\_id |  |  |  |
| person\_id | ptid |  |  |
| observation\_concept\_id | proc\_code  proc\_code\_type | Look for procedures that map to the vocabulary with domain\_id = observation |  |
| observation\_date | proc\_date |  |  |
| observation\_time | proc\_time |  |  |
| observation\_type\_concept\_id |  |  |  |
| value\_as\_number |  |  |  |
| value\_as\_string |  |  |  |
| value\_as\_concept\_id |  |  |  |
| qualifier\_concept\_id |  |  |  |
| unit\_concept\_id |  |  |  |
| provider\_id | encid |  | To avoid duplication, apply the same logic described when transforming the encounter table to visit\_occurrence above. |
| visit\_occurrence\_id | encid |  |  |
| observation\_source\_value | proc\_code |  |  |
| observation\_source\_concept\_id | proc\_code |  |  |
| unit\_source\_value |  |  |  |
| qualifier\_source\_value |  |  |  |

### Reading from microbiology



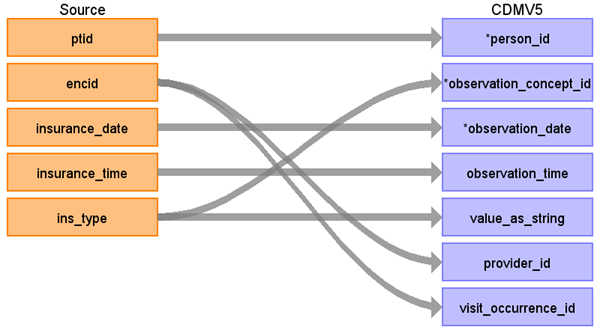
|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| observation\_id |  |  |  |
| person\_id | ptid |  |  |
| observation\_concept\_id | ptid | 4155370 Microbiology |  |
| observation\_date | collect\_date |  |  |
| observation\_time | collect\_time |  |  |
| observation\_type\_concept\_id | specimen\_source | 44818702 Lab result  44818703 Pathology finding |  |
| value\_as\_number |  |  |  |
| value\_as\_string | organism |  |  |
| value\_as\_concept\_id |  |  |  |
| qualifier\_concept\_id |  |  |  |
| unit\_concept\_id |  |  |  |
| provider\_id | encid |  |  |
| visit\_occurrence\_id | encid |  |  |
| observation\_source\_value | organism |  |  |
| observation\_source\_concept\_id |  |  |  |
| unit\_source\_value |  |  |  |
| qualifier\_source\_value |  |  |  |

### Reading from nlp\_biomarker



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| observation\_id |  |  |  |
| person\_id | ptid |  |  |
| observation\_concept\_id | biomarker | 0 |  |
| observation\_date | note\_date |  |  |
| observation\_time |  |  |  |
| observation\_type\_concept\_id |  |  |  |
| value\_as\_number |  |  |  |
| value\_as\_string | biomarker\_status  variation\_detail  biomarker | Concatenate biomarker + variation\_detail + biomarker\_status |  |
| value\_as\_concept\_id |  |  |  |
| qualifier\_concept\_id |  |  |  |
| unit\_concept\_id |  |  |  |
| provider\_id |  |  | To avoid duplication, apply the same logic described when transforming the encounter table to visit\_occurrence above. |
| visit\_occurrence\_id |  |  |  |
| observation\_source\_value | biomarker | BIOMARKER |  |
| observation\_source\_concept\_id | biomarker |  |  |
| unit\_source\_value |  |  |  |
| qualifier\_source\_value |  |  |  |

### Reading from Insurance



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| observation\_id |  |  |  |
| person\_id | ptid |  |  |
| observation\_concept\_id | ins\_type | 45877222 | Insurance status/requirement |
| observation\_date | insurance\_date |  |  |
| observation\_time | insurance\_time |  |  |
| observation\_type\_concept\_id |  |  |  |
| value\_as\_number |  |  |  |
| value\_as\_string | ins\_type |  |  |
| value\_as\_concept\_id |  |  |  |
| qualifier\_concept\_id |  |  |  |
| unit\_concept\_id |  |  |  |
| provider\_id | encid |  | To avoid duplication, apply the same logic described when transforming the encounter table to visit\_occurrence above. |
| visit\_occurrence\_id | encid |  |  |
| observation\_source\_value |  |  |  |
| observation\_source\_concept\_id |  |  |  |
| unit\_source\_value |  |  |  |
| qualifier\_source\_value |  |  |  |

## Table Name: CDM\_DOMAIN\_META

|  |  |  |  |
| --- | --- | --- | --- |
| **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 |  |  |

## Tables Not Mapped:

|  |  |  |
| --- | --- | --- |
| **Meta Data Lookup Table for OPTUM ONCOLOGY HER** | | |
| **TABLE NAME** | **DOMAIN\_ID** | **DESCRIPTION** |
| PERSON | Person |  |
| OBSERVATION\_PERIOD | Observation Period |  |
| CARE\_SITE | Care Site |  |
| VISIT\_OCCURRENCE | Visit |  |
| PROVIDER | Provider |  |
| DEATH | Death |  |
| CONDITION\_OCCURRENCE | Condition |  |
| DRUG\_EXPOSURE | Drug |  |
| PROCEDURE\_OCCURRENCE | Procedure |  |
| MEASUREMENT | Measurement |  |
| OBSERVATION | Observation |  |

### Table name: cdm\_source

|  |  |  |  |
| --- | --- | --- | --- |
| SOURCE\_DOCUMENTATION\_REFERENCE |  | <https://jnj.sharepoint.com/sites/PHM-GCSP-RND/RWE/Pages/DataSourceDetails.aspx?DataSourceItem=36> |  |
| 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: payer\_plan\_period

### Table name: procedure\_cost

### Table name: visit\_cost

### Table name: drug\_cost

### Table name: device\_cost

### Table name: drug\_era

### Table name: dose\_era

### Table name: condition\_era

### Table name: cohort

### Table name: cohort\_definition

### Table name: cohort\_attribute

### Table name: attribute\_definition

# Appendix

# 1 Code Snippets

## 1.1 Vocabulary Mapping

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

### 1.1.1 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')

### 1.1.2 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\_VOCABULARY\_ID IN ('RxNORM')

### 1.1.3 Source to Maps to Value

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 Value'

JOIN CONCEPT C1

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

AND C1.INVALID\_REASON IS NULL

)

SELECT \*

FROM CTE\_VOCAB\_MAP

/\*EXAMPLE FILTERS\*/

WHERE SOURCE\_CODE = 'V87.43'

## Source tables

### Table: patient

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Most freq. value | Comment |
| ptid | character | List truncated... | Create new integer value unique to PTID |
| birth\_yr | character varying | 1930 and Earlier |  |
| gender | character varying | Male |  |
| race | character varying | Caucasian |  |
| ethnicity | character varying | Not Hispanic |  |
| region | character varying | South | ACTION: confirm it's clean mapping from region to division. If it is, then just bring division as it's more granular. |
| division | character varying | South Atl/West South Crl |  |
| avg\_hh\_income | double precision | 39005.0 |  |
| pct\_college\_educ | double precision | 26.0 |  |
| deceased\_indicator | character varying | 0 |  |
| date\_of\_death | character |  |  |
| provid\_pcp | character varying |  |  |
| idn\_indicator | character | 1 |  |
| first\_month\_active | character varying | 200601 |  |
| last\_month\_active | character varying | 201506 |  |
| notes\_eligible | character | 1 |  |
| has\_notes | character | 1 |  |
| sourceid | character varying | S0034 |  |
| source\_data\_through | character varying | 201506 |  |

### Table: diagnosis

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Most freq. value | Comment |
| ptid | character | List truncated... |  |
| encid | character varying |  |  |
| diag\_date | date | 2014-05-02 |  |
| diag\_time | timestamp without time zone | 1900-01-01.000000 |  |
| diagnosis\_cd | character varying |  |  |
| diagnosis\_cd\_type | character varying | ICD9 | To save time, create reference table with source concept and standard concept |
| diagnosis\_status | character varying | Diagnosis of | Only use 'Diagnosis of' in CONDITION\_OCCURRENCE |
| poa | character | 0 |  |
| admitting\_diagnosis | character | 0 |  |
| discharge\_diagnosis | character | 0 |  |
| primary\_diagnosis | character | 0 |  |
| problem\_list | character | N |  |

### Table: care\_area

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Most freq. value | Comment |
| ptid | character | PT095853463 |  |
| encid | character varying | List truncated... |  |
| carearea | character varying | UNKNOWN CARE AREA |  |
| carearea\_date | date |  |  |
| carearea\_time | timestamp without time zone | 1900-01-01.000000 |  |

### Table: encounter\_provider

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Most freq. value | Comment |
| encid | character varying | List truncated... |  |
| provid | character varying | 70371 |  |
| provider\_role | character varying | ATTENDING |  |

### Table: provider

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Most freq. value | Comment |
| provid | character varying | List truncated... |  |
| specialty | character varying | Unspecified |  |
| prim\_spec\_ind | character | 1 |  |

### Table: visit

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Most freq. value | Comment |
| ptid | character | PT091048451 |  |
| visitid | bigint | List truncated... |  |
| visit\_type | character varying | Inpatient |  |
| visit\_start\_date | date | 2012-08-13 |  |
| visit\_start\_time | timestamp without time zone | 1900-01-01.000000 | delete the null time 1900-01-01.000000 21,485 23.0 |
| visit\_end\_date | date | 2014-07-25 |  |
| visit\_end\_time | timestamp without time zone | 1900-01-01.000000 |  |
| discharge\_disposition | character varying | 01 DISCHARGED TO HOME OR SELF CARE |  |
| admission\_source | character varying | Referred by physician or self referral; non-healthcare facility point of origin |  |
| drg | character varying |  |  |

### Table: encounter

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Most freq. value | Comment |
| ptid | character | List truncated... |  |
| visitid | bigint |  |  |
| encid | character varying | List truncated... |  |
| interaction\_type | character varying | Office or clinic patient | Outpatient, Inpatient, ER, LTC. "Prescriptions and refills" don't sound like a valid encounter. Shouldn't be just going to the pharmacy. "other" --> unknown |
| interaction\_date | date | 2014-06-16 |  |
| interaction\_time | timestamp without time zone | 1900-01-01.000000 |  |

### Table: nlp\_sds

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Most freq. value | Comment |
| ptid | character | List truncated... |  |
| encid | character varying |  |  |
| note\_date | date | 2013-04-23 |  |
| sds\_term | character varying | pain | usagi. restrict to condition, only use high quality mapping. everything else, drop it. |
| sds\_location | character varying |  |  |
| sds\_attribute | character varying |  |  |
| sds\_sentiment | character varying |  | need to express only 'positive' findings in CONDITION\_OCCURRENCE |
| note\_section | character varying |  |  |

### Table: medication\_administrations

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Most freq. value | Comment |
| ptid | character | PT093154612 |  |
| encid | character varying |  |  |
| drug\_name | character varying | SODIUM CHLORIDE 0.9 % |  |
| ndc | character varying | 00338004904 |  |
| ndc\_source | character varying | Derived |  |
| order\_date | date | 2014-10-07 |  |
| order\_time | timestamp without time zone | 1900-01-01.000000 |  |
| admin\_date | date |  |  |
| admin\_time | timestamp without time zone |  |  |
| provid | character varying |  |  |
| quantity\_of\_dose | character varying |  |  |
| route | character varying | Intravenous |  |
| strength | character varying |  |  |
| strength\_unit | character varying |  |  |
| dosage\_form | character varying |  |  |
| dosefreq | character varying |  |  |
| generic\_desc | character varying | SODIUM CHLORIDE |  |
| drug\_class | character varying | Intravenous nutritional therapy; electrolyte; trace element; metal; vitamin; alone or combinations |  |

### Table: patient\_reported\_meds

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Most freq. value | Comment |
| ptid | character | PT155237089 |  |
| reported\_date | date | 2009-03-27 |  |
| drug\_name | character varying |  |  |
| ndc | character varying | 49999035930 |  |
| ndc\_source | character varying | Direct |  |
| provid | character varying |  |  |
| route | character varying | Oral |  |
| quantity\_of\_dose | character varying |  |  |
| strength | character varying |  |  |
| strength\_unit | character varying | mg |  |
| dosage\_form | character varying | Tabs |  |
| dosefreq | character varying |  |  |
| generic\_desc | character varying |  |  |
| drug\_class | character varying | Salicylates |  |

### Table: prescriptions\_written

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Most freq. value | Comment |
| ptid | character | PT118881881 |  |
| rxdate | date | 2014-02-24 |  |
| rxtime | timestamp without time zone | 1900-01-01.000000 |  |
| drug\_name | character varying |  |  |
| ndc | character varying | 00406035705 |  |
| ndc\_source | character varying | Direct |  |
| provid | character varying |  |  |
| route | character varying | Oral |  |
| quantity\_of\_dose | character varying |  |  |
| strength | character varying |  |  |
| strength\_unit | character varying | mg |  |
| dosage\_form | character varying | Tabs |  |
| dosefreq | character varying |  |  |
| quantity\_per\_fill | character varying | 30 |  |
| num\_refills | double precision | 0.0 |  |
| days\_supply | double precision |  |  |
| generic\_desc | character varying |  |  |
| drug\_class | character varying | HMG & CoA reductase inhibitors (statins) |  |

### Table: nlp\_drug\_rationale

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Most freq. value | Comment |
| ptid | character | PT191017751 |  |
| encid | character varying |  |  |
| note\_date | date | 2013-04-23 |  |
| note\_section | character varying | MEDICATIONS |  |
| drug\_name | character varying | ASPIRIN |  |
| drug\_action | character varying | N/A |  |
| drug\_action\_preposition | character varying | OF |  |
| reason\_general | character varying |  |  |
| sentiment | character varying |  |  |
| sentiment\_who | character varying |  |  |

### Table: immunization

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Most freq. value | Comment |
| ptid | character | PT194773862 |  |
| immunization\_date | date |  |  |
| immunization\_desc | character varying | INFLUENZA VIRUS VACCINE; INACTIVATED |  |
| mapped\_name | character varying | Influenza Inactivated Vaccine |  |
| ndc | character varying | 49281001350 |  |
| ndc\_source | character varying | Derived |  |
| pt\_reported | character |  |  |

### Table: procedure

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Most freq. value | Comment |
| ptid | character | List truncated... |  |
| encid | character varying |  |  |
| proc\_date | date | 2012-10-22 |  |
| proc\_time | timestamp without time zone | 1900-01-01.000000 |  |
| proc\_code | character varying |  |  |
| proc\_desc | character varying |  |  |
| proc\_code\_type | character varying | CPT4 |  |
| provid\_perform | character varying |  |  |
| provid\_order | character varying |  |  |
| betos\_code | character varying |  |  |
| betos\_desc | character varying |  |  |

### Table: nlp\_measurements

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Most freq. value | Comment |
| ptid | character | PT130234983 |  |
| encid | character varying |  |  |
| note\_date | date | 2013-04-23 |  |
| measurement\_type | character varying | DBP |  |
| measurement\_value | character varying | 2 |  |
| measurement\_detail | character varying |  |  |
| note\_section | character varying |  |  |
| measurement\_date | date |  |  |

### Table: insurance

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Most freq. value | Comment |
| ptid | character | PT130555350 |  |
| encid | character varying | List truncated... |  |
| insurance\_date | date | 2008-09-15 |  |
| insurance\_time | timestamp without time zone | 1900-01-01.000000 |  |
| ins\_type | character varying | Medicare |  |

### Table: labs

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Most freq. value | Comment |
| ptid | character | PT143890838 |  |
| encid | character varying |  |  |
| test\_name | character varying | O2 saturation.oximetry |  |
| order\_date | date |  |  |
| order\_time | timestamp without time zone |  |  |
| collected\_date | date |  |  |
| collected\_time | timestamp without time zone |  |  |
| result\_date | date |  |  |
| result\_time | timestamp without time zone | 1900-01-01.000000 |  |
| test\_result | character varying | Negative |  |
| relative\_indicator | character varying |  |  |
| result\_unit | character varying | % |  |
| normal\_range | character varying |  |  |
| evaluated\_for\_range | character | N |  |
| value\_within\_range | character | U |  |

### Table: observations

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Most freq. value | Comment |
| ptid | character | PT143890838 |  |
| encid | character varying |  |  |
| obs\_type | character varying | SBP |  |
| obs\_date | date | 2014-06-10 |  |
| obs\_time | timestamp without time zone | 1900-01-01.000000 |  |
| obs\_result | character varying | 18 |  |
| obs\_unit | character varying | mm Hg |  |
| evaluated\_for\_range | character | N |  |
| value\_within\_range | character | U |  |

### Table: microbiology

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Most freq. value | Comment |
| ptid | character | PT117858758 |  |
| encid | character varying |  |  |
| order\_date | date |  |  |
| order\_time | timestamp without time zone | 1900-01-01.000000 |  |
| collect\_date | date |  |  |
| collect\_time | timestamp without time zone | 1900-01-01.000000 |  |
| receive\_date | date |  |  |
| receive\_time | timestamp without time zone |  |  |
| result\_date | date | 2015-05-22 |  |
| result\_time | timestamp without time zone | 1900-01-01.000000 |  |
| result\_status | character varying |  |  |
| specimen\_source | character varying | Urine |  |
| organism | character varying |  |  |
| mapped\_organism\_found | character varying |  |  |
| mapped\_organism\_excluded | character varying |  |  |
| culture\_growth | character varying |  |  |
| culture\_value | character varying |  |  |
| culture\_unit | character varying |  |  |
| antibiotic | character varying |  |  |
| mapped\_antibiotic | character varying |  |  |
| sensitivity | character varying |  |  |

### Table: nlp\_biomarker

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Most freq. value | Comment |
| ptid | character | PT084542965 |  |
| note\_date | date | 2013-04-23 |  |
| biomarker | character varying | CD20 |  |
| variation\_detail | character varying |  |  |
| biomarker\_status | character varying | positive |  |

### Table: nlp\_sds\_family

|  |  |  |  |
| --- | --- | --- | --- |
| Field | Type | Most freq. value | Comment |
| ptid | character | PT084524827 |  |
| encid | character varying |  |  |
| note\_date | date | 2014-07-23 |  |
| sds\_term | character varying | cancer |  |
| sds\_location | character varying |  |  |
| sds\_family\_member | character varying | who=family |  |
| sds\_sentiment | character varying |  |  |
| note\_section | character varying | family medical history |  |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 38000175 | Prescription dispensed in pharmacy | Type Concept | Drug Type | Drug Type | S | OMOP generated | 1970-01-01 | 2099-12-31 |  |
| 38000176 | Prescription dispensed through mail order | Type Concept | Drug Type | Drug Type | S | OMOP generated | 1970-01-01 | 2099-12-31 |  |
| 38000177 | Prescription written | Type Concept | Drug Type | Drug Type | S | OMOP generated | 1970-01-01 | 2099-12-31 |  |
| 38000178 | Medication list entry | Type Concept | Drug Type | Drug Type | S | OMOP generated | 1970-01-01 | 2099-12-31 |  |
| 38000179 | Physician administered drug (identified as procedure) | Type Concept | Drug Type | Drug Type | S | OMOP generated | 1970-01-01 | 2099-12-31 |  |
| 38000180 | Inpatient administration | Type Concept | Drug Type | Drug Type | S | OMOP generated | 1970-01-01 | 2099-12-31 |  |
| 38000181 | Drug era - 0 days persistence window | Type Concept | Drug Type | Drug Type | S | OMOP generated | 1970-01-01 | 2099-12-31 |  |
| 38000182 | Drug era - 30 days persistence window | Type Concept | Drug Type | Drug Type | S | OMOP generated | 1970-01-01 | 2099-12-31 |  |
| 44787730 | Patient Self-Reported Medication | Type Concept | Drug Type | Drug Type | S | OMOP generated | 1970-01-01 | 2099-12-31 |  |
| 43542356 | Physician administered drug (identified from EHR problem list) | Type Concept | Drug Type | Drug Type | S | OMOP generated | 1970-01-01 | 2099-12-31 |  |
| 43542357 | Physician administered drug (identified from referral record) | Type Concept | Drug Type | Drug Type | S | OMOP generated | 1970-01-01 | 2099-12-31 |  |
| 43542358 | Physician administered drug (identified from EHR observation) | Type Concept | Drug Type | Drug Type | S | OMOP generated | 1970-01-01 | 2099-12-31 |  |
| 44777970 | Randomized Drug | Type Concept | Drug Type | Drug Type | S | OMOP generated | 1970-01-01 | 2099-12-31 |  |