****

**Japan Medical Data Center (JMDC 2018-)**

**to**

**OMOP Common Data Model (CDM V5.3.0)**

**ETL Mapping Specification**

**Version 4.0**

**25-Feb-2015**

**July 30, 2018**

**Martijn Schuemie,**

**Amy Matcho,**

**Chris Knoll,**

**Erica Voss**

# Table of contents

Contents

[1. Table of contents 2](#_Toc510013864)

[2. Introduction 3](#_Toc510013865)

[3. Source Data Mapping Approach 4](#_Toc510013866)

[4. Data Mapping 5](#_Toc510013867)

[Table name: cost 5](#_Toc510013868)

[Table name: visit\_occurrence 9](#_Toc510013869)

[Table name: person 12](#_Toc510013870)

[Table name: observation\_period 14](#_Toc510013871)

[Table name: death 15](#_Toc510013872)

[Table name: condition\_occurrence 17](#_Toc510013873)

[Table name: measurement 21](#_Toc510013874)

[Table name: observation 24](#_Toc510013875)

[Table name: procedure\_occurrence 28](#_Toc510013876)

[Table name: drug\_exposure 32](#_Toc510013877)

[Table name: care\_site 35](#_Toc510013878)

[Table name: provider 36](#_Toc510013879)

[Table name: cdm\_source 37](#_Toc510013880)

[Table name: cdm\_domain\_meta 37](#_Toc510013881)

[5. Source Independent Data Mapping 40](#_Toc510013882)

[Table Name: DRUG\_ERA 40](#_Toc510013883)

[Table Name: CONDITION\_ERA 40](#_Toc510013884)

[6. Appendices 41](#_Toc510013885)

[Drug mapping 41](#_Toc510013886)

[Specialty mapping 42](#_Toc510013887)

[Measurement mapping 42](#_Toc510013888)

# Introduction

This document reflects the requirements, assumptions, business rules and transformations for the Japan Medical Data Center (JMDC) implementation of the Common Data Model Version 5.3.0 (CDM) as implemented by Janssen. This document assumes JMDC follows the structure and conventions used in 2018 onwards, which is a different structure than the one used before that.

The purpose of this document is to describe the ETL mapping of the licensed data from JMDC into the OMOP Common Data Model.

Japanese universal multi-payer health-care system covering virtually all citizens was started in 1961. Currently, there are more than 3000 insurers in Japan which may be divided into several main groups. The government‐managed health insurance (kyo-kai-kenpo) consisting of a single (largest) insurer (central government) covers about 35 million people who are workers in small and medium-size enterprises and their dependents. Area-basis public health insurances (koku-min-ken-ko-ho-ken) run by 1,881 municipal governments (e.g., cities, towns and villages) cover 38 million people. Eighty five cooperative insurances (kyo-sai-kumi-a-i) cover 9 million who are mainly civil servants and their dependents. A total of 1,431 Society-Managed Health Insurances (kumi-a-i-sho-kan-ken-ko-ho-ken) cover 29 million who are workers in relatively large enterprises and their dependents. A new scheme starting from 2008 called as “Medical Care System for People Aged 75 and over” run by 47 prefectures covers 15 million old people aged 75 or older.

JMDC database consists of data from 60 Society-Managed Health Insurances covering workers aged 18 to 65 and their dependents (children younger than 18 years old and elderly people older than 65 years old). The old people (particularly those aged 66 or older) are less representative as compared with whole population in the nation. When estimated among the people who are younger than 66 years old, the proportion of children younger than 18 years old in JMDC is approximately the same as the proportion in the whole nation. JMDC data includes data on membership status of the insured people and claims data provided by insurers under contract. Claims data are derived from monthly claims issued by clinics, hospitals and community pharmacies. The number of claims issued and added to JMDC database is about 800,000 per month. The size of JMDC population is 1.9 million, 1.5% of about 120 million people in the whole nation. A total of 60 insurances may be classified by data period as follows:

<data period> <N of insurances> <Start year>

10 years 4 2005

7 years 5 2008

6 years 16 2009

5 years 7 2010

4 years 11 2009 or 2011

3 years 7 2009 or 2012

2 years 4 2009 to 2013

1 year 5 2009 (one year)

This document is based on the OMOP ETL Specifications. General information that is covered by the OMOP ETL Specification will not be covered in this document, but a detailed discussion of the JMDC specific aspects of mapping and converting data to the standard CDM is provided.

The document is composed of two main sections:

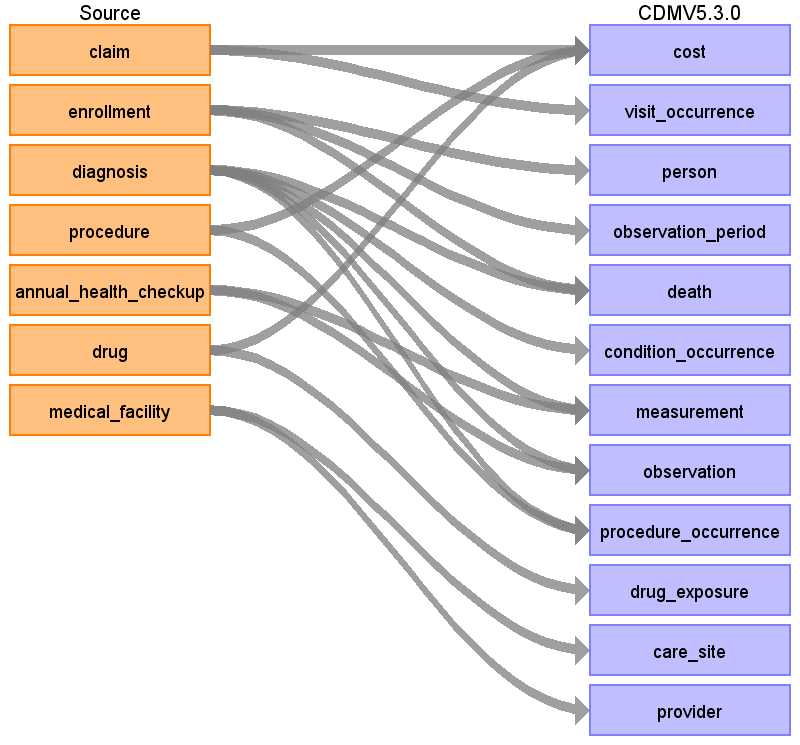
* Source Data Mapping. Describes major tables of the CDM schema and special data handling required for each table.
* Source Independent Data Mapping. Describes mapping process of the Drug and Condition Era’s.

In each section, the tables and their mapping are individually reviewed along with any source specific rules and exceptions.

The intended audience for this document will include both researchers that want to use the experience and learning in order to incorporate them into their own CDM construction.

# Source Data Mapping Approach

The figure below represents the general approach to mapping the source data tables that are available in JMDC to the CDM data schema. The orange boxes represent the input tables in JMDC and the blue boxes represent the resulting CDM tables.



# Data Mapping

## Table name: cost

Reading from claim



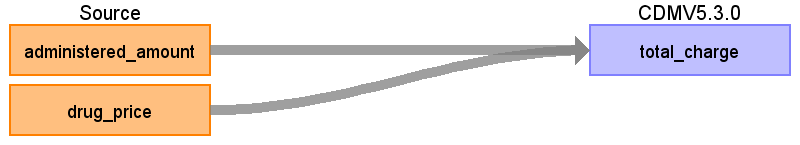
|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| cost\_id |  |  | Auto generate |
| cost\_event\_id | claim\_id |  | From claim: use claim\_id, remove ‘C’ prefix (= visit\_occurrence\_id). |
| cost\_domain\_id |  |  | Visit |
| cost\_type\_concept\_id |  |  | 5031 (Amount paid by the patient or reimbursed by the payer) |
| currency\_concept\_id |  |  | 44818592 (Japanese yen) |
| total\_charge |  |  |  |
| total\_cost |  |  |  |
| total\_paid | total\_point |  | Amount = point \* 10 Yen |
| paid\_by\_payer |  |  |  |
| paid\_by\_patient |  |  |  |
| paid\_patient\_copay |  |  |  |
| paid\_patient\_coinsurance |  |  |  |
| paid\_patient\_deductible |  |  |  |
| paid\_by\_primary |  |  |  |
| paid\_ingredient\_cost |  |  |  |
| paid\_dispensing\_fee |  |  |  |
| payer\_plan\_period\_id |  |  |  |
| amount\_allowed |  |  |  |
| revenue\_code\_concept\_id |  |  |  |
| revenue\_code\_source\_value |  |  |  |
| drg\_concept\_id |  |  |  |
| drg\_source\_value |  |  |  |

Reading from procedure



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| cost\_id |  |  | Auto generate |
| cost\_event\_id |  |  | From procedure: use procedure\_occurrence\_id. |
| cost\_domain\_id |  |  | Procedure |
| cost\_type\_concept\_id |  |  | 5032 (Amount charged to the patient or the payer by the provider, list price) |
| currency\_concept\_id |  |  | 44818592 (Japanese yen) |
| total\_charge | number\_of\_times |  | total\_charge = number\_of\_times \* 10 \* procedure\_standard\_point. (Join to procedure\_master table on both standardized\_procedure\_id and standardized\_procedure\_version) |
| total\_cost |  |  |  |
| total\_paid |  |  | Amount = point \* 10 Yen |
| paid\_by\_payer |  |  |  |
| paid\_by\_patient |  |  |  |
| paid\_patient\_copay |  |  |  |
| paid\_patient\_coinsurance |  |  |  |
| paid\_patient\_deductible |  |  |  |
| paid\_by\_primary |  |  |  |
| paid\_ingredient\_cost |  |  |  |
| paid\_dispensing\_fee |  |  |  |
| payer\_plan\_period\_id |  |  |  |
| amount\_allowed |  |  |  |
| revenue\_code\_concept\_id |  |  |  |
| revenue\_code\_source\_value |  |  |  |
| drg\_concept\_id |  |  |  |
| drg\_source\_value |  |  |  |

Reading from drug

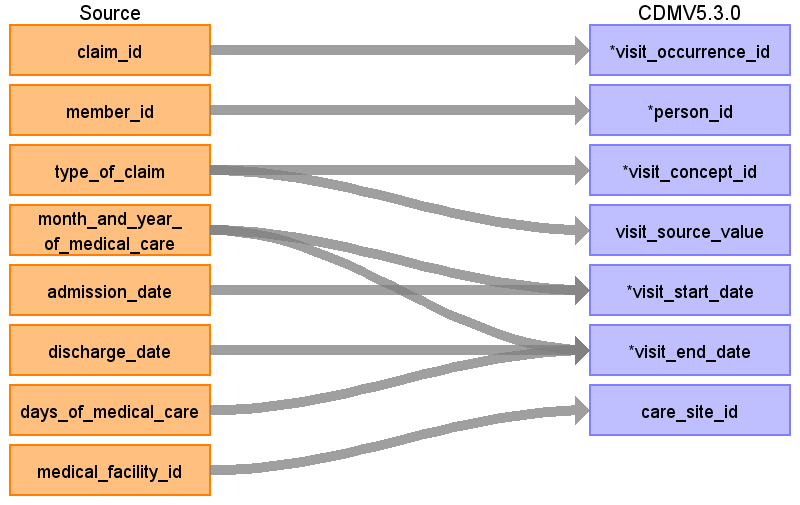


|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| cost\_id |  |  | Auto generate |
| cost\_event\_id |  |  | From drug: use drug\_exposure\_id. |
| cost\_domain\_id |  |  | Drug |
| cost\_type\_concept\_id |  |  | 5032 (Amount charged to the patient or the payer by the provider, list price) |
| currency\_concept\_id |  |  | 44818592 (Japanese yen) |
| total\_charge | drug\_price  administered\_amount |  | total\_charge = drug\_price \* administered\_amount |
| total\_cost |  |  |  |
| total\_paid |  |  | Amount = point \* 10 Yen |
| paid\_by\_payer |  |  |  |
| paid\_by\_patient |  |  |  |
| paid\_patient\_copay |  |  |  |
| paid\_patient\_coinsurance |  |  |  |
| paid\_patient\_deductible |  |  |  |
| paid\_by\_primary |  |  |  |
| paid\_ingredient\_cost |  |  |  |
| paid\_dispensing\_fee |  |  |  |
| payer\_plan\_period\_id |  |  |  |
| amount\_allowed |  |  |  |
| revenue\_code\_concept\_id |  |  |  |
| revenue\_code\_source\_value |  |  |  |
| drg\_concept\_id |  |  |  |
| drg\_source\_value |  |  |  |

## Table name: visit\_occurrence

Each unique claim id for a patient will be assigned to a visit, **except for pharmacy claims**. In claims only the year + month is provided, day information is not available. However, oftentimes a claim will be associated with a piece of information that does have a specific date. Specific dates can come from prescription date, procedure date, admission date, discharge date, or the start of medical care date. This information will be used to infer the visit date with more precision. (e.g. if it’s a 2 day visit, and it has a procedure taking place on the 20th, we might assume the visit starts on the 20th). If no specific date is available, we will assume the visit start and any events associated with the visit fall on the 15th of the month. We choose the middle of the month because accidental reversal of temporality (where the order of events is switched because one piece of information did have a date, and another didn’t) is just as likely to occur in one direction as the other.

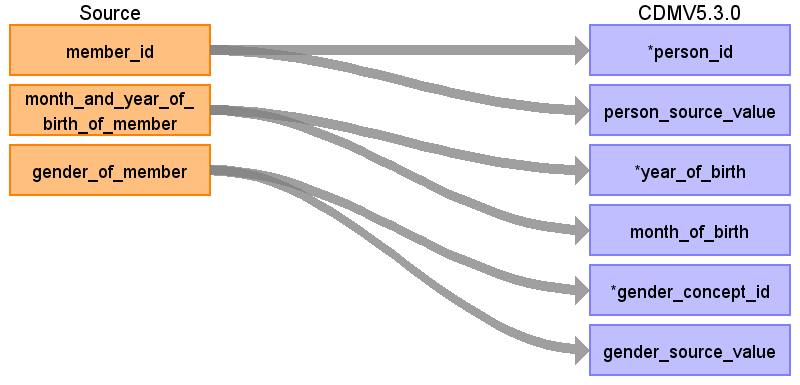
Reading from claim



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| visit\_occurrence\_id | claim\_id | Remove ‘C’ prefix |  |
| person\_id | member\_id | Remove 'M' prefix |  |
| visit\_concept\_id | type\_of\_claim | DPC and inpatient 9201 Outpatient and Pharmacy 9202 |  |
| visit\_source\_value | type\_of\_claim |  |  |
| visit\_start\_date | month\_and\_year\_of\_medical\_care  admission\_date | If the claim is associated with a diagnosis, and (1) that Diagnosis.date\_of\_medical\_care\_start falls within the Month\_and\_year\_of\_medical\_care and (2) there is no other diagnoses with the same level 4 ICD-10 code from the same institution for the same member with the same Diagnosis.date\_of\_medical\_care\_start, then the Diagnosis.date\_of\_medical\_care\_start is used as the visit\_start\_date.  If a claim has an admission date, and the admission date falls within the Month\_and\_year\_of\_medical\_care , set the visit\_start\_date to the admission date.  If a claim has an admission date, and the admission date falls before the Month\_and\_year\_of\_medical\_care , set the visit\_start\_date to the 1st day of Month\_and\_year\_of\_medical\_care.  If the claim is associated with a Drug.prescription\_date or Procedure.procedure\_date, use the minimum of those dates and the visit\_start\_date so far. (if no date was established so far, just take the minimum of the procedure and prescription dates).  If no visit\_start\_date has been identified at that point, set it to the 15th of the Month\_and\_year\_of\_medical\_care. |  |
| visit\_end\_date | month\_and\_year\_of\_medical\_care  days\_of\_medical\_care  discharge\_date | Start by setting the visit\_end\_date to the visit\_start\_date + days\_of\_medical\_care.  If the discharge\_date is specified, use the discharge\_date as the visit\_start\_date.  Take the minimum of the visit\_start\_date so far and the end of the month\_and\_year\_of\_medical care. |  |
| care\_site\_id | medical\_facility\_id |  |  |
| visit\_start\_datetime |  |  |  |
| visit\_end\_datetime |  |  |  |
| visit\_type\_concept\_id |  |  |  |
| provider\_id |  |  |  |
| visit\_source\_concept\_id |  |  |  |
| admitting\_source\_concept\_id |  |  |  |
| admitting\_source\_value |  |  |  |
| discharge\_to\_concept\_id |  |  |  |
| discharge\_to\_source\_value |  |  |  |
| preceding\_visit\_occurrence\_id |  |  |  |

## Table name: person

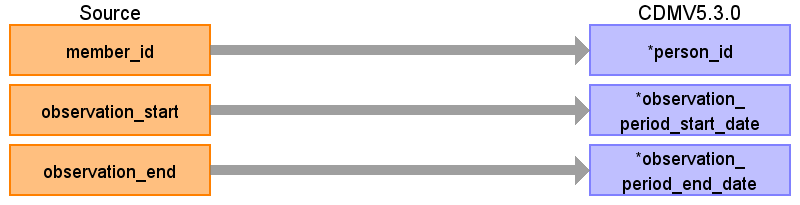
Reading from enrollment



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| person\_id | member\_id | Remove 'M' prefix |  |
| person\_source\_value | member\_id |  |  |
| year\_of\_birth | month\_and\_year\_of\_birth\_of\_member | Take first 4 digits |  |
| month\_of\_birth | month\_and\_year\_of\_birth\_of\_member | Take last 2 digits (01 is January) |  |
| gender\_concept\_id | gender\_of\_member | Male = 8507 Female = 8532 |  |
| gender\_source\_value | gender\_of\_member |  |  |
| day\_of\_birth |  |  |  |
| birth\_datetime |  |  |  |
| race\_concept\_id |  |  |  |
| ethnicity\_concept\_id |  |  |  |
| location\_id |  |  |  |
| provider\_id |  |  |  |
| care\_site\_id |  |  |  |
| gender\_source\_concept\_id |  |  |  |
| race\_source\_value |  |  |  |
| race\_source\_concept\_id |  |  |  |
| ethnicity\_source\_value |  |  |  |
| ethnicity\_source\_concept\_id |  |  |  |

## Table name: observation\_period

Reading from enrollment



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| observation\_period\_id |  |  |  |
| person\_id | member\_id | Remove 'M' prefix |  |
| observation\_period\_start\_date | observation\_start | Set to first of month (E.g. '200802' becomes 1st of February 2008) |  |
| observation\_period\_end\_date | observation\_end | Set to last day of month (E.g. '200801' becomes 31st of January 2008) |  |
| period\_type\_concept\_id |  |  | 44814722 Period while enrolled in insurance |

## Table name: death

There are two sources of death status: enrollment and the diagnosis table. Death is coded as ‘outcome = 3’ in the diagnosis table, and as ‘withdrawal\_death\_flag = true’ in the enrollment table. To make sure we have at most one death per person, when there are multiple death records per person, we will take the latest from diagnose if available, else the date from enrollment. The reason for this is that the date from diagnosis might be specific to date, whereas the information from enrollment status is always at the month level and therefore assumed to be at the end of the month.

Reading from enrollment



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| person\_id | member\_id | Remove 'M' prefix |  |
| death\_date | observation\_end |  |  |
| death\_datetime |  |  |  |
| death\_type\_concept\_id |  |  | From enrollment: 38003565 (Payer enrollment status "Deceased") From diagnosis: 38003567 (Medical claim diagnostic code indicating death) |
| cause\_concept\_id |  |  |  |
| cause\_source\_value |  |  |  |
| cause\_source\_concept\_id |  |  |  |

Reading from diagnosis



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| person\_id | member\_id | Remove 'M' prefix |  |
| death\_date | month\_and\_year\_of\_medical\_care | Use end of visit/claim date |  |
| death\_datetime |  |  |  |
| death\_type\_concept\_id |  |  | From enrollment: 38003565 (Payer enrollment status "Deceased") From diagnosis: 38003567 (Medical claim diagnostic code indicating death) |
| cause\_concept\_id |  |  |  |
| cause\_source\_value |  |  |  |
| cause\_source\_concept\_id |  |  |  |

## Table name: condition\_occurrence

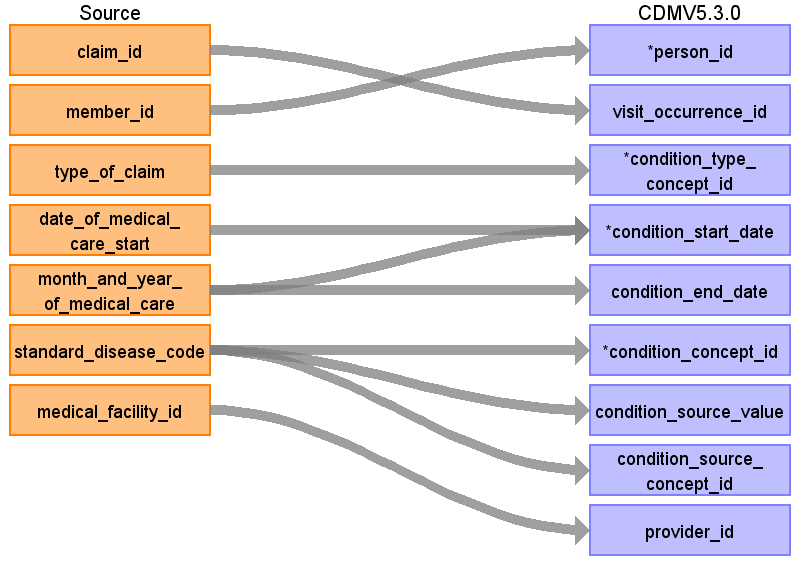
Two types of conditions will be captured in JMDC: (1) ICD-10 diagnoses with start dates of the associated visit (see visit\_occurrence section for how these are determined), and (2) ICD-10 diagnoses with start dates of the “date\_of\_medical\_care\_start”. This second type of condition record can be considered to be a ‘condition era’ type record, as the condition start date will in some cases be the actual start of the condition as opposed to a service date related to the condition. See mapping table below for specific logic.

This condition-era type record is possible because in Japanese claims, the “date\_of\_medical\_care\_start” is specified for almost all conditions and is the first day when that hospital/clinic started to provide health-care service for that condition. This date can be very old (e.g., decades ago).

The “date\_of\_medical\_care\_start” is often the date of true occurrence of the condition though the date can be later than the true occurrence date as a new start date will be created if the patient is referred to a different clinic/hospital. Even in the same clinic/hospital, a new start date may be given to the same condition of the same patient because of several reasons. For instance, if the patient had an acute condition again after the recovery from the previous same acute condition, a new start date may be created as two episodes were judged independent. Even for the chronic condition, a new start date may be created if the service has not been given for certain period (e.g., more than several months). The latter may occur according to the policy of handling patient’s conditions which may differ between clinics/hospitals. A new start date may also be created from an administrative view (e.g., the service fee for a new condition is a little higher than that for an old condition).

Reading from diagnosis

When an ICD10 code in the diagnosis table maps to a concept in the Condition domain a record should be created in the condition\_occurrence table, except when the field ‘Suspicion flag’ equals 1.

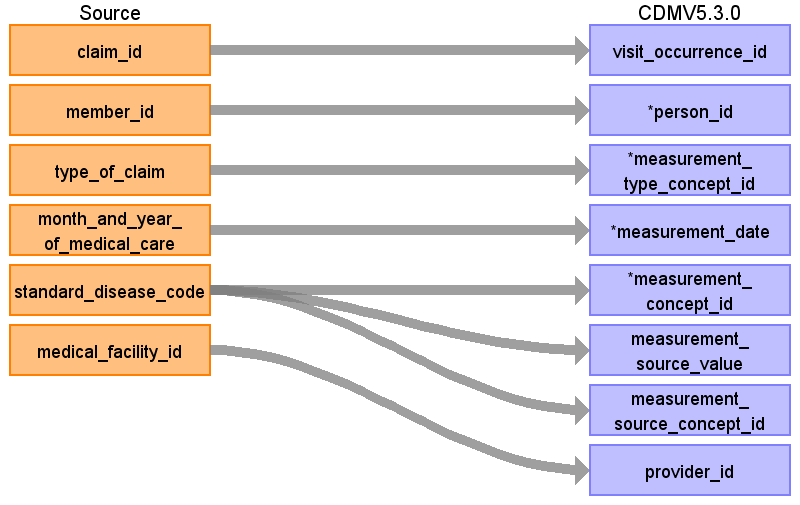


|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| condition\_occurrence\_id |  |  |  |
| person\_id | member\_id | Remove 'M' prefix |  |
| visit\_occurrence\_id | claim\_id | Remove ‘C’ prefix |  |
| condition\_type\_concept\_id | type\_of\_claim | Outpatient: 38000215 (Outpatient detail - 1st position) InPatient or DPC: 38000184 (Inpatient detail - 1st position)  condition-era type record 38000246 (Condition era - 0 days persistence window) |  |
| condition\_start\_date | month\_and\_year\_of\_medical\_care  date\_of\_medical\_care\_start | Use start of visit.  Create additional condition occurrences (condition 'eras') if date\_of\_medical\_care\_start is earlier than the start of month\_and\_year of\_medical\_care. If date\_of\_medical\_care\_start precedes the observation\_period\_start date, set start to observation\_period\_start\_date. |  |
| condition\_end\_date | month\_and\_year\_of\_medical\_care | For condition era type records, the end date should be the end of the last diagnosis for that patient that has the specific date\_of\_medical\_care\_start, medical\_facility\_id, and standard\_disease\_code.  For all other records set to null. |  |
| condition\_concept\_id | standard\_disease\_code |  | Lookup icd10\_level4\_code in diagnosis\_master table, and use vocab to map to standard concept. Remove '-' prior to mapping (e.g. 'I50-' should map to 'I50'), and ignore period (e.g. 'I500' should map to 'I50.0') |
| condition\_source\_value | standard\_disease\_code |  | Lookup icd10\_level4\_code in diagnosis\_master table  Lookup icd10\_level4\_code |
| condition\_source\_concept\_id | standard\_disease\_code |  | Lookup icd10\_level4\_code in diagnosis\_master table, and use vocab to map to source concept. Remove '-' prior to mapping (e.g. 'I50-' should map to 'I50'), and ignore period (e.g. 'I500' should map to 'I50.0')  Lookup icd10\_level4\_code |
| provider\_id | medical\_facility\_id |  | Use dummy provider corresponding to the institute |
| condition\_start\_datetime |  |  |  |
| condition\_end\_datetime |  |  |  |
| stop\_reason |  |  |  |
| visit\_detail\_id |  |  |  |
| condition\_status\_source\_value |  |  |  |
| condition\_status\_concept\_id |  |  |  |

## Table name: measurement

Reading from diagnosis

When an ICD10 code in the diagnosis table maps to a concept in the Measurement domain a record should be created in the measurement table.



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| measurement\_id |  |  |  |
| visit\_occurrence\_id | claim\_id | Remove ‘C’ prefix |  |
| person\_id | member\_id | Remove 'M' prefix |  |
| measurement\_type\_concept\_id | type\_of\_claim | Outpatient: 38000215 (Outpatient detail - 1st position) Inpatient or DPC: 38000184 (Inpatient detail - 1st position) | 38000277 (Lab observation with numeric result) for continuous fields 38000279 (Lab observation concept code result) for ECG 900000004 (Observation text) for ophthalmoscopies '44814721' |
| measurement\_date | month\_and\_year\_of\_medical\_care | Use derived visit\_start\_date |  |
| measurement\_concept\_id | standard\_disease\_code |  | Lookup icd10\_level4\_code in diagnosis\_master table, and use vocab to map to standard concept. |
| measurement\_source\_value | standard\_disease\_code |  | Lookup icd10\_level4\_code in diagnosis\_master table |
| measurement\_source\_concept\_id | standard\_disease\_code |  | Lookup icd10\_level4\_code in diagnosis\_master table, and use vocab to map to source concept. |
| provider\_id | medical\_facility\_id | Use dummy provider corresponding to the institute | Use dummy provider corresponding to the institute |
| measurement\_datetime |  |  |  |
| operator\_concept\_id |  |  |  |
| value\_as\_number |  |  |  |
| value\_as\_concept\_id |  |  | From Health checkups: from mapping table. Else 4181412 (Present) |
| unit\_concept\_id |  |  | From mapping table |
| range\_low |  |  | From Health checkups: take from reference file |
| range\_high |  |  | From Health checkups: take from reference file |
| visit\_detail\_id |  |  |  |
| unit\_source\_value |  |  | From mapping table |
| value\_source\_value |  |  |  |

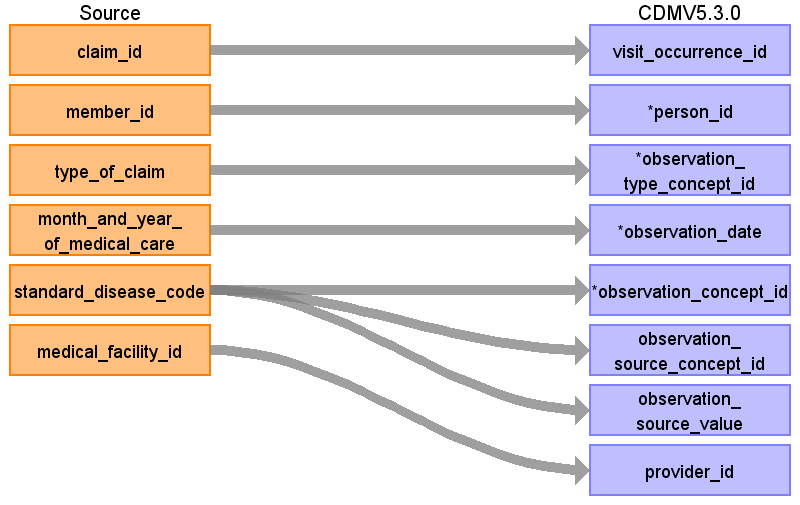
Reading from annual\_health\_checkup

Most information in the health checkups table should be entered in the measurement table (see health checkups mapping table).

## Table name: observation

Reading from diagnosis

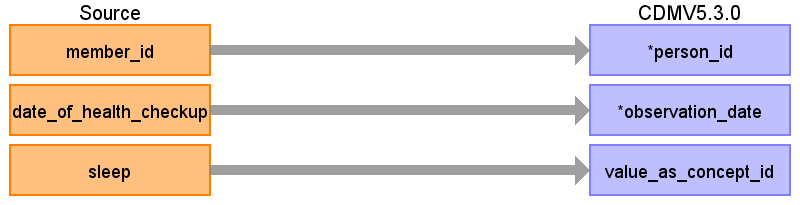
When an ICD10 code in the diagnosis table maps to a concept in the Observation domain a record should be created in the observation table.



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| observation\_id |  |  |  |
| visit\_occurrence\_id | claim\_id | Remove ‘C’ prefix |  |
| person\_id | member\_id | Remove 'M' prefix |  |
| observation\_type\_concept\_id | type\_of\_claim | Outpatient: 38000215 (Outpatient detail - 1st position) Inpatient or DPC: 38000184 (Inpatient detail - 1st position) |  |
| observation\_date | month\_and\_year\_of\_medical\_care | Use visit\_start\_date |  |
| observation\_concept\_id | standard\_disease\_code |  | Lookup icd10\_level4\_code in diagnosis\_master table, and use vocab to map to standard concept. Remove '-' prior to mapping (e.g. 'I50-' should map to 'I50'), and ignore period (e.g. 'I500' should map to 'I50.0') |
| observation\_source\_concept\_id | standard\_disease\_code |  | Lookup icd10\_level4\_code in diagnosis\_master table, and use vocab to map to source concept. Remove '-' prior to mapping (e.g. 'I50-' should map to 'I50'), and ignore period (e.g. 'I500' should map to 'I50.0') |
| observation\_source\_value | standard\_disease\_code |  | Lookup icd10\_level4\_code in diagnosis\_master table |
| provider\_id | medical\_facility\_id |  | Use dummy provider corresponding to the institute |
| observation\_datetime |  |  |  |
| value\_as\_number |  |  |  |
| value\_as\_string |  |  |  |
| value\_as\_concept\_id |  |  | From diagnoses:value from “Maps to value” relationship or else 4181412 (present) |
| qualifier\_concept\_id |  |  |  |
| unit\_concept\_id |  |  |  |
| visit\_detail\_id |  |  |  |
| unit\_source\_value |  |  |  |
| qualifier\_source\_value |  |  |  |

Reading from annual\_health\_checkup

Information in the sleeping field should go to the observation table, as well as information in fields that are not mapped to specific concepts.

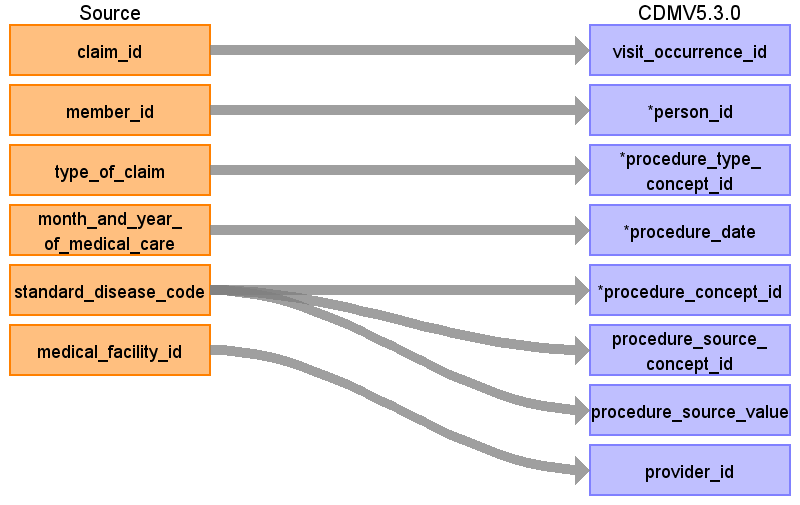


|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| observation\_id |  |  |  |
| visit\_occurrence\_id |  |  |  |
| person\_id | member\_id | Remove 'M' prefix |  |
| observation\_type\_concept\_id |  |  |  |
| observation\_date | date\_of\_health\_checkup |  |  |
| observation\_concept\_id |  |  |  |
| observation\_source\_concept\_id |  |  |  |
| observation\_source\_value |  |  |  |
| provider\_id |  |  |  |
| observation\_datetime |  |  |  |
| value\_as\_number |  |  |  |
| value\_as\_string |  |  |  |
| value\_as\_concept\_id | sleep | 1 = 4188539 (yes), 2 = 4188540 (no) | From diagnoses:value from “Maps to value” relationship or else 4181412 (present) |
| qualifier\_concept\_id |  |  |  |
| unit\_concept\_id |  |  |  |
| visit\_detail\_id |  |  |  |
| unit\_source\_value |  |  |  |
| qualifier\_source\_value |  |  |  |

## Table name: procedure\_occurrence

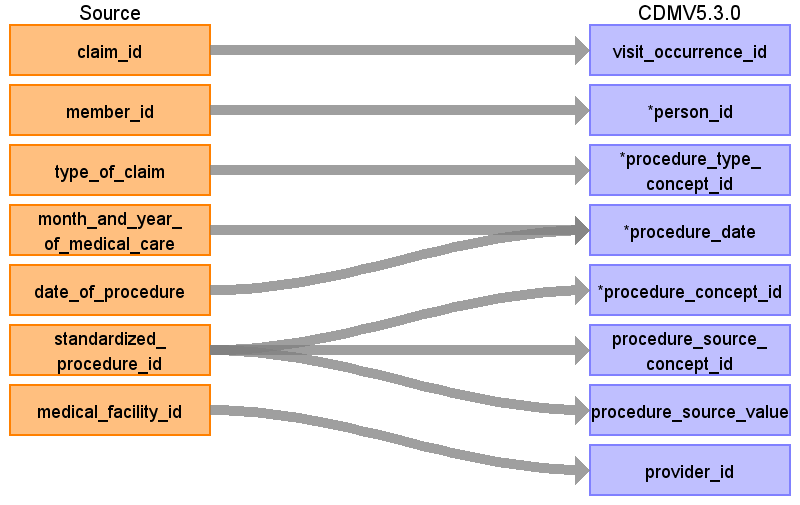
Reading from diagnosis

When an ICD10 code in the diagnosis table maps to a concept in the Procedure domain a record should be created in the procedure\_occurrence table.



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| procedure\_occurrence\_id |  |  |  |
| visit\_occurrence\_id | claim\_id | Remove ‘C’ prefix |  |
| person\_id | member\_id | Remove 'M' prefix |  |
| procedure\_type\_concept\_id | type\_of\_claim | Outpatient: 38000215 (Outpatient detail - 1st position) Inpatient or DPC: 38000184 (Inpatient detail - 1st position) |  |
| procedure\_date | month\_and\_year\_of\_medical\_care | Use visit\_start\_date |  |
| procedure\_concept\_id | standard\_disease\_code |  | Lookup icd10\_level4\_code in diagnosis\_master table, and use vocab to map to standard concept. Remove '-' prior to mapping (e.g. 'I50-' should map to 'I50'), and ignore period (e.g. 'I500' should map to 'I50.0') |
| procedure\_source\_concept\_id | standard\_disease\_code |  | Lookup icd10\_level4\_code in diagnosis\_master table, and use vocab to map to source concept. Remove '-' prior to mapping (e.g. 'I50-' should map to 'I50'), and ignore period (e.g. 'I500' should map to 'I50.0') |
| procedure\_source\_value | standard\_disease\_code |  | Lookup icd10\_level4\_code in diagnosis\_master table |
| provider\_id | medical\_facility\_id |  | Use dummy provider corresponding to the institute |
| procedure\_datetime |  |  |  |
| modifier\_concept\_id |  |  |  |
| quantity |  |  |  |
| visit\_detail\_id |  |  |  |
| modifier\_source\_value |  |  |  |

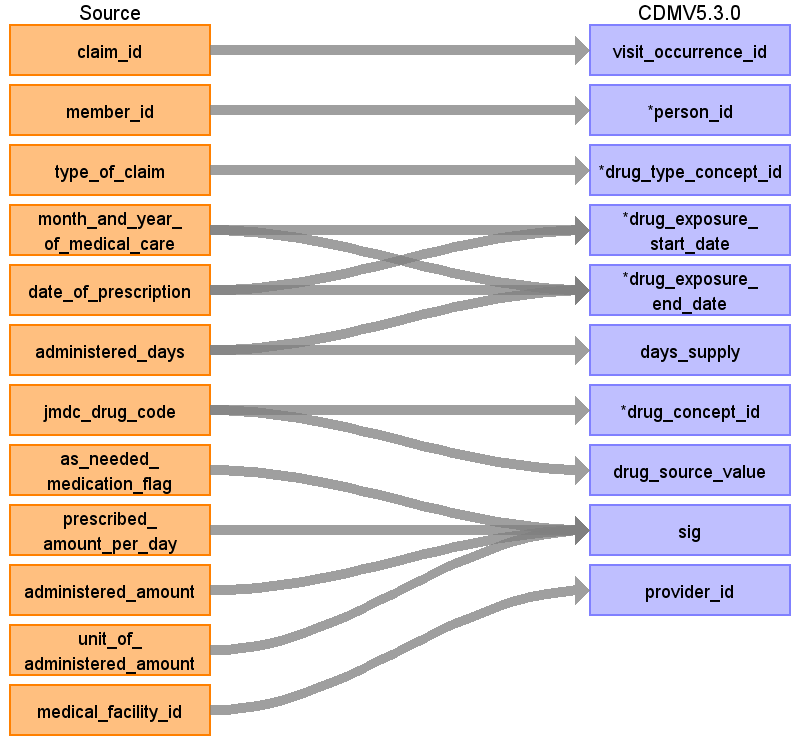
Reading from procedure



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| procedure\_occurrence\_id |  |  |  |
| visit\_occurrence\_id | claim\_id | Remove ‘C’ prefix |  |
| person\_id | member\_id | Remove 'M' prefix |  |
| procedure\_type\_concept\_id | type\_of\_claim | Outpatient: 38000267 (Outpatient detail - 1st position) Inpatient or DPC: 38000249 (Inpatient detail - 1st position) |  |
| procedure\_date | month\_and\_year\_of\_medical\_care  date\_of\_procedure | Use date of procedure when populated, otherwise use start of visit |  |
| procedure\_concept\_id | standardized\_procedure\_id |  |  |
| procedure\_source\_concept\_id | standardized\_procedure\_id |  |  |
| procedure\_source\_value | standardized\_procedure\_id |  |  |
| provider\_id | medical\_facility\_id |  | Use dummy provider corresponding to the institute |
| procedure\_datetime |  |  |  |
| modifier\_concept\_id |  |  |  |
| quantity |  |  |  |
| visit\_detail\_id |  |  |  |
| modifier\_source\_value |  |  |  |

## Table name: drug\_exposure

Reading from drug



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| drug\_exposure\_id |  |  |  |
| visit\_occurrence\_id | claim\_id | Remove ‘C’ prefix |  |
| person\_id | member\_id | Remove 'M' prefix |  |
| drug\_type\_concept\_id | type\_of\_claim | Pharmacy, Outpatient: 38000175 (Prescription dispensed in pharmacy) Inpatient or DPC: 38000180 (Inpatient administration) |  |
| drug\_exposure\_start\_date | date\_of\_prescription  month\_and\_year\_of\_medical\_care | Use date of prescription if available, otherwise set to start of visit. |  |
| drug\_exposure\_end\_date | month\_and\_year\_of\_medical\_care  date\_of\_prescription  administered\_days | Drug\_exposure\_end\_date = drug\_exposure\_start\_date + min(administered\_days,180) |  |
| days\_supply | administered\_days | If value > 180, set to 180 (occurs in 74 prescriptions) |  |
| drug\_concept\_id | jmdc\_drug\_code |  | Look up in drug mapping table |
| drug\_source\_value | jmdc\_drug\_code |  |  |
| sig | as\_needed\_medication\_flag  prescribed\_amount\_per\_day  administered\_amount  unit\_of\_administered\_amount | <prescribed\_amount\_per\_day> <administered\_amount> per day (<as\_needed\_medication\_flag  >?as needed, <administered\_amount> < unit\_of\_administered\_amount> total | Combine the four fields to create a sig string: |
| provider\_id | medical\_facility\_id |  | Use the dummy providers we created per institution. |
| quantity |  |  |  |
| dose\_unit\_source\_value |  |  |  |
| drug\_source\_concept\_id |  |  |  |
| drug\_exposure\_end\_datetime |  |  |  |
| verbatim\_end\_date |  |  |  |
| stop\_reason |  |  |  |
| refills |  |  |  |
| drug\_exposure\_start\_datetime |  |  |  |
| route\_concept\_id |  |  |  |
| lot\_number |  |  |  |
| visit\_detail\_id |  |  |  |
| route\_source\_value |  |  |  |

## Table name: care\_site

Reading from medical\_facility

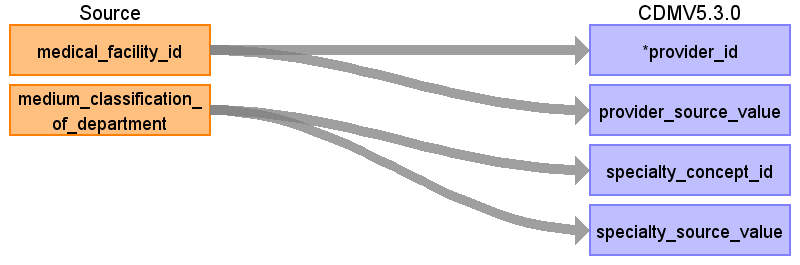


|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| care\_site\_id | medical\_facility\_id | Remove ‘F’ prefix |  |
| care\_site\_source\_value | medical\_facility\_id |  |  |
| care\_site\_name |  |  |  |
| place\_of\_service\_concept\_id |  |  |  |
| location\_id |  |  |  |
| place\_of\_service\_source\_value |  |  |  |

## Table name: provider

Reading from medical\_facility

We create a dummy provider for every institution, so we can at least capture specialty for diagnoses and procedures (which are mapped to institution, but not to physician).



|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| provider\_id | medical\_facility\_id | Remove ‘F’ prefix |  |
| provider\_source\_value | medical\_facility\_id |  |  |
| specialty\_concept\_id | medium\_classification\_of\_department |  | Use mapping table |
| specialty\_source\_value | medium\_classification\_of\_department |  |  |
| provider\_name |  |  |  |
| npi |  |  |  |
| dea |  |  |  |
| care\_site\_id |  |  |  |
| year\_of\_birth |  |  |  |
| gender\_concept\_id |  |  |  |
| specialty\_source\_concept\_id |  |  |  |
| gender\_source\_value |  |  |  |
| gender\_source\_concept\_id |  |  |  |

## Table name: cdm\_source

|  |  |  |  |
| --- | --- | --- | --- |
| Destination Field | Source Field | Logic | Comment |
| cdm\_source\_name |  | JMDC (Japan Medical Data Center) database |  |
| cdm\_source\_abbreviation |  | JMDC |  |
| source\_description |  | JMDC database consists of data from 60 Society-Managed Health Insurances covering workers aged 18 to 65 and their dependents (children younger than 18 years old and elderly people older than 65 years old). The old people (particularly those aged 66 or older) are less representative as compared with whole population in the nation. When estimated among the people who are younger than 66 years old, the proportion of children younger than 18 years old in JMDC is approximately the same as the proportion in the whole nation. JMDC data includes data on membership status of the insured people and claims data provided by insurers under contract. Claims data are derived from monthly claims issued by clinics, hospitals and community pharmacies. |  |
| source\_documentation\_reference |  |  |  |
| cdm\_etl\_reference |  | https://github.com/OHDSI/ETL-CDMBuilder |  |
| cdm\_version |  | 5.2.0 |  |

## Table name: cdm\_domain\_meta

|  |  |  |
| --- | --- | --- |
| **TABLE NAME** | **DOMAIN\_ID** | **DESCRIPTION** |
| PERSON | Person | JMDC covers workers aged 18 to 65 and their dependents (children younger than 18 years old and elderly people older than 65 years old). The old people (particularly those aged 66 or older) are less representative as compared with whole population in the nation. When estimated among the people who are younger than 66 years old, the proportion of children younger than 18 years old in JMDC is approximately the same as the proportion in the whole nation.  Only the year of birth is available, so not the day or month. |
| OBSERVATION\_PERIOD | Observation Period | The observation period is defined as the time of enrollment in the health insurance. If the member is a dependent, the enrollment depends on the enrollment of the main beneficiary. The observation is truncated by the time when JMDC had a contract with the respective insurer. |
| CARE\_SITE | Care Site | Care sites in JMDC are institutions where care is provided, typically a department in a hospital. |
| VISIT\_OCCURRENCE | Visit | Each insurance claim that is not a pharmacy claim is translated into a visit. Because a claim is always for costs in a specific month, a visit is always restricted to one month, even if the underlying real visit is longer than a month. Only the year and month is provided for each claim, day information is not available. However, oftentimes a claim will be associated with a piece of information that does have a specific date. Specific dates can come from prescription date, procedure date, admission date, discharge date, or the start of medical care date. This information will be used to infer the visit date with more precision. (e.g. if it’s a 2 day visit, and it has a procedure taking place on the 20th, we might assume the visit starts on the 20th). If no specific date is available, we will assume the visit start and any events associated with the visit fall on the 15th of the month. We choose the middle of the month because accidental reversal of temporality (where the order of events is switched because one piece of information did have a date, and another didn’t) is just as likely to occur in one direction as the other. |
| PROVIDER | Provider | In JMDC providers are derived from two sources: institutions and physicians. Institutions (typically hospital departments) are included because diagnoses and procedures are linked to institutions but not to physicians, and this way we can preserve the specialty of the department that recorded a diagnosis or procedure. |
| DEATH | Death | There are two sources of death status: enrollment and the diagnosis table. To make sure we have at most one death per person, when there are multiple death records per person, we will take the latest from diagnose if available, else the date from enrollment. The reason for this is that the date from diagnosis might be specific to date, whereas the information from enrollment status is always at the month level and therefore assumed to be at the end of the month. |
| CONDITION\_OCCURRENCE | Condition | In JMDC we distinguish between two different types of conditions, both derived from the diagnoses associated with a claim. The first have their start date set to the date of diagnose, which is often taken to be the visit start date. The second type uses the ‘date of start of medical care’ that is recorded for virtually every diagnosis in JMDC. This start date often precedes the actual visit where the diagnosis is recorded, and sometimes even precedes start of enrollment. This second type of conditions is used to construct condition eras by linking diagnoses from the same institution for the same diagnose code with the same start of medical care. |
| PROCEDURE\_OCCURRENCE | Procedure | Procedures performed during hospitalization. For procedures the exact day of the procedure (relative to the visit start date) is available. Some of the procedure occurrences are derived from diagnose codes using the vocabulary. |
| DRUG\_EXPOSURE | Drug | Drug exposures include a mix of prescriptions and dispensings as records on the insurance claims. |
| MEASUREMENT | Measurement | Most employees in Japan are required by law to undergo annual health checkups. For some but not all insurance companies included in JMDC the results of these checkups are available as measurements. |
| OBSERVATION | Observation | During annual health screenings information about sleeping is collected that cannot go into the measurement table and is therefore stored here. Some observations are derived from some of the diagnosis codes. |
| COST | Cost | Only the total amount paid for a claim is known with certainty. Since we map most claims directly to visits, we have reflected these costs as costs per visit. |

# Source Independent Data Mapping

## Table Name: DRUG\_ERA

The DRUG\_ERA table is defined as a span of time when a person is assumed to be using a particular drug. Successive periods of such drug exposure may be combined to produce one continuous drug era. The DRUG\_ERA table is populated from the DRUG\_EXPOSURE table within the CDM. Drug eras are consolidated to their respective ingredient off the DRUG\_EXPOSURE table, eliminating all information about strength and administration route. 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.

| **Destination Field** | **Source Field** | **Applied Rule** | **Comment** |
| --- | --- | --- | --- |
| DRUG\_ERA\_ID |  | System Generated |  |
| PERSON\_ID | PERSON\_ID |  |  |
| DRUG\_CONCEPT\_ID | DRUG\_CONCEPT\_ID | Do not create DRUG\_ERAs where the DRUG\_EXPOSURE.DRUG\_CONCEPT\_ID is 0.  Use the map below to map DRUG\_EXPOSURE.DRUG\_CONCEPT\_ID to the ingredient level DRUG\_CONCEPT\_ID used in the DRUG\_ERA. |  |
| DRUG\_ERA\_START\_DATE | DRUG\_EXPOSURE\_START\_DATE |  |  |
| DRUG\_ERA\_END\_DATE | Drug\_exposure\_end\_date |  |  |
| DRUG\_TYPE\_CONCEPT\_ID |  | Apply a 30 day persistence window and label as CONCEPT\_ID 38000182 (Drug era - 30 days persistence window). | Falls under CONCEPT\_VOCABULARY\_ID = 36 - Drug Exposure Type. |
| DRUG\_EXPOSURE\_COUNT |  | Sum up the number of DRUG\_EXPOSURES for this PERSON\_ID and this CONCEPT\_ID during the exposure window being built. |  |

## Table Name: CONDITION\_ERA

Condition Era table is constructed through an aggregation of individual Condition Occurrences recorded in the CONDITION\_OCCURRENCE table.

| **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\_TYPE\_CONCEPT\_ID | - | Apply a 30 day persistence window and label as CONCEPT\_ID 38000247 (Condition era - 30 days persistence window). | Falls under CONCEPT\_VOCABULARY\_ID = 37 - OMOP Condition Occurrence Type. |
| CONDITION\_OCCURRENCE\_COUNT | - | Sum up the number of CONDITION\_OCCURRENCEs for this PERSON\_ID and this CONCEPT\_ID during the exposure window being built. |  |

# Appendices

## Drug mapping

A drug map was automatically constructed based on the data available in the Drug table. The process is described in this document:



The full mapping can be found here:



## Specialty mapping

Specialties were manually mapped from the values of the ‘Medium classification of department’ fields to OMOP Specialty concepts.



## Measurement mapping

Measurements as found in the ‘Health checkups’ table were manually mapped to LOINC concepts, and units were mapped to UCUM concepts.

