

COVID-19 Clinical Data Warehouse Data Dictionary
Based on
OMOP Common Data Model Specifications Version 5.3

Contents

| | | |
|----|--|----|
| 1 | PERSON | 3 |
| 2 | OBSERVATION_PERIOD | 5 |
| 3 | SPECIMEN | 6 |
| 4 | DEATH | 7 |
| 5 | VISIT_OCCURRENCE | 8 |
| 6 | PROCEDURE_OCCURRENCE | 9 |
| 7 | DRUG_EXPOSURE | 10 |
| 8 | DEVICE_EXPOSURE | 12 |
| 9 | CONDITION_OCCURRENCE | 13 |
| 10 | MEASUREMENT | 15 |
| 11 | OBSERVATION | 17 |
| | Standardized Health System Data Tables | 19 |
| 12 | LOCATION | 19 |
| 13 | CARE_SITE | 20 |
| 14 | PROVIDER | 21 |

This document is a modification of an existing OHDSI document. Fields with grey strike through fonts are excluded from the data pulled.

The NIH COVID-19 Data Warehouse is an NIH data sharing resource, operated under a contract containing clinical and imaging data from individuals who have received a Coronavirus Disease 2019 (“COVID-19”) tested or whose symptoms are consistent with COVID-19. Data will also be collected from individuals infected with pathogens such as SARS 1, MERS, and H1N1 to support comparative studies.

Note the tables and fields contain both “required” and “non-required data”. The required information will be expected as part of the data pull on each patient where non-required fields are optional.

The field Person_ID is a unique identifier of an individual patient however it is NOT the institutions individual medical record number, (MRN). The person_id is generated at the time of de-identification and will be used to concatenate individual’s data in the longitudinal data warehouse.

Institutions are given two options in the data transfer agreement

Option 1: Supply Personal Health Information, (PHI) PHI that includes unaltered dates of service, zip AND a hash patient identifier. The hash patient identifier is generated by mathematical algorithm using a combination of personal identifying information (DOB, MRN, Zip, etc.). The hash is held by a trusted 3rd party that can be used to re-identify patients. Patient identification allows for linkages to other data types like images or linkages to future prospective studies

Option 2: Supply a limited data set (dates of service and zip are not altered)

1 PERSON

The person domain contains records that uniquely identify each patient in the source data who is time at-risk to have clinical observations recorded within the source systems. Each person record has associated demographic attributes which are assumed to be constant for the patient throughout the course of their periods of observation. All other patient-related data domains have a foreign-key reference to the person domain.

| Field | Required | Type | Description |
|--------------------------|----------|--------------|---|
| person_id | Yes | integer | A unique identifier for each person. |
| gender_concept_id | Yes | integer | A foreign key that refers to a standard concept identifier in the Standardized Vocabularies for the gender of the person. |
| year_of_birth | Yes | integer | The year of birth of the person. For data sources with date of birth, the year is extracted. For data sources where the year of birth is not available, the approximate year of birth is derived based on any age group categorization available. |
| race_concept_id | Yes | integer | A foreign key that refers to a standard concept identifier in the Standardized Vocabularies for the race of the person. |
| ethnicity_concept_id | Yes | integer | A foreign key that refers to the standard concept identifier in the Standardized Vocabularies for the ethnicity of the person. |
| location_id | No | integer | A foreign key to the place of residency for the person in the location table, where the detailed address information is stored. |
| provider_id | No | integer | A foreign key to the primary care provider the person is seeing in the provider table. |
| care_site_id | No | integer | A foreign key to the site of primary care in the care_site table, where the details of the care site are stored. |
| person_source_value | No | varchar (50) | An (encrypted) key derived from the person identifier in the source data. This is necessary when a use case requires a link back to the person data at the source dataset. |
| gender_source_value | No | varchar (50) | The source code for the gender of the person as it appears in the source data. The person's gender is mapped to a standard gender concept in the Standardized Vocabularies; the original value is stored here for reference. |
| gender_source_concept_id | No | Integer | A foreign key to the gender concept that refers to the code used in the source. |
| race_source_value | No | varchar (50) | The source code for the race of the person as it appears in the source data. The person race is mapped to a standard race concept in the Standardized Vocabularies and the original value is stored here for reference. |
| race_source_concept_id | No | Integer | A foreign key to the race concept that refers to the code used in the source. |

| | | | |
|-----------------------------|----|--------------|--|
| ethnicity_source_value | No | varchar (50) | The source code for the ethnicity of the person as it appears in the source data. The person ethnicity is mapped to a standard ethnicity concept in the Standardized Vocabularies and the original code is, stored here for reference. |
| ethnicity_source_concept_id | No | Integer | A foreign key to the ethnicity concept that refers to the code used in the source. |

4.1.1 CONVENTIONS

- Valid Gender, Race and Ethnicity Concepts belong to the "Demographic" domain.
- Person source data attributes are race, gender, and ethnicity.
- Ethnicity in the OMOP CDM follows the OMB Standards for Data on Race and Ethnicity: Only distinctions between Hispanics and Non-Hispanics are made.
- Additional information is stored through references to other tables about the home address (location_id) and the primary care provider.
- The provider refers to the primary care provider (General Practitioner).
- The care site refers to where the provider typically provides the primary care.
- All persons are required to have a valid gender and year of birth.
- The person table requires only one value for each attribute. While it is possible for a person to change genders, locations, and providers over time, it is the responsibility of the data holder to select the one value to use in the CDM.

2 OBSERVATION_PERIOD

The observation period domain contains records which uniquely define the spans of time for which a person is at-risk to have clinical events recorded within the source systems. One person may have one or more disjoint observation periods, during which times analyses may assume that clinical events would be captured if observed, and outside of which no clinical events may be recorded.

| Field | Required | Type | Description |
|-------------------------------|----------|---------|---|
| observation_period_id | Yes | integer | A unique identifier for each observation period. |
| person_id | Yes | integer | A foreign key identifier to the person for whom the observation period is defined. The demographic details of that person are stored in the person table. |
| observation_period_start_date | Yes | date | The start date of the observation period for which data are available from the data source. |
| observation_period_end_date | Yes | date | The end date of the observation period for which data are available from the data source. |
| period_type_concept_id | Yes | Integer | A foreign key identifier to the predefined concept in the Standardized Vocabularies reflecting the source of the observation period information |

i. CONVENTIONS

1. Each Person can have more than one valid OBSERVATION_PERIOD record, but no two observation periods can overlap in time for a given person.
2. During an Observation Period, any clinical event that happens to the patient is expected to be recorded. Conversely, the absence of data indicates that no clinical events occurred to the patient.
3. No clinical data are valid outside an active Observation Period. Clinical data that refer to a time outside (diagnoses of previous conditions such as "Old MI" or medical history) of an active Observation Period are recorded as Observations. The date of the Observation is the first day of the first Observation Period of a patient.
4. For claims data, observation periods are inferred from the enrollment periods to a health benefit plan.

3 SPECIMEN

The specimen domain contains the records identifying each biological sample from a person.

| Field | Required | Type | Description |
|-----------------------------|----------|--------------|---|
| specimen_id | Yes | integer | A unique identifier for each specimen. |
| person_id | Yes | integer | A foreign key identifier to the person for whom the specimen is recorded. |
| specimen_concept_id | Yes | integer | A foreign key referring to a standard concept identifier in the Standardized Vocabularies for the specimen. |
| specimen_type_concept_id | Yes | integer | A foreign key referring to the predefined concept identifier in the Standardized Vocabularies reflecting the system of record from which the specimen was represented in the source data. |
| specimen_date | Yes | date | The date the specimen was obtained from the person. |
| specimen_time | No | time | The time on the date when the specimen was obtained from the person. |
| quantity | No | float | The amount of specimen collection from the person during the sampling procedure |
| unit_concept_id | No | integer | A foreign key to a standard concept identifier for the unit associated with the numeric quantity of the specimen collection. |
| anatomic_site_concept_id | No | integer | A foreign key to a standard concept identifier for the anatomic location of specimen collection. |
| disease_status_concept_id | No | integer | A foreign key to a standard concept identifier for the disease status of specimen collection. |
| specimen_source_id | No | varchar (50) | The specimen identifier as it appears in the source data. |
| specimen_source_value | No | varchar(50) | The specimen value as it appears in the source data. This value is mapped to a standard concept in the Standardized Vocabularies and the original code is, stored here for reference. |
| unit_source_value | No | varchar(50) | The information about the unit as detailed in the source. |
| anatomic_site_source_value | No | varchar(50) | The information about the anatomic site as detailed in the source. |
| disease_status_source_value | No | varchar(50) | The information about the disease status as detailed in the source. |

a. CONVENTIONS

- i. Anatomic site is coded at the most specific level of granularity possible, such that higher level classifications can be derived using the Standardized Vocabularies

4 DEATH

The death domain contains the clinical event for how and when a person dies. A person can have up to one record if the source systems contain evidence that s/he is deceased, such as:

- i. Condition Code in the Header or Detail information of claims
- ii. Status of enrollment into a health plan
- iii. Explicit record in EHR data

Living patients should not contain any information in the death table.

| Field | Required | Type | Description |
|-------------------------|----------|-------------|--|
| person_id | Yes | integer | A foreign key identifier to the deceased person. The demographic details of that person are stored in the person table. |
| death_date | Yes | date | The date the person was deceased. If the precise date including day or month is not known or not allowed, December is used as the default month, and the last day of the month the default day. |
| death_type_concept_id | Yes | integer | A foreign key referring to the predefined concept identifier in the Standardized Vocabularies reflecting how the death was represented in the source data. |
| cause_concept_id | No | integer | A foreign key referring to a standard concept identifier in the Standardized Vocabularies for conditions. |
| cause_source_value | No | varchar(50) | The source code for the cause of death as it appears in the source data. This code is mapped to a standard concept in the Standardized Vocabularies and the original code is, stored here for reference. |
| cause_source_concept_id | No | Integer | A foreign key to the concept that refers to the code used in the source. Note, this variable name is abbreviated to ensure it will be allowable across database platforms. |

5.4.1 CONVENTIONS

- Each Person may have more than one record of death in the source data. It is the task of the Extract Transform and Load (ETL) to pick the most plausible or most accurate records to be aggregated and stored as a single record in the Death table.
- If the Death Date cannot be precisely determined from the data, the best approximation should be used.

5 VISIT_OCCURRENCE

The visit domain contains the spans of time a person continuously receives medical services from one or more providers at a care site in a given setting within the health care system. Visits are classified into 4 settings: outpatient care, inpatient confinement, emergency room, and long-term care. Persons may transition between these settings over the course of an episode of care. If applicable, relationships between visits within an episode of care may be represented in the FACT_RELATIONSHIP table.

Visits are recorded in various data sources in different forms with varying levels of standardization. For example:

- i. Medical Claims include Inpatient Admissions, Outpatient Services, and Emergency Room visits.
- ii. Electronic Health Records may capture Person visits as part of the activities recorded.

| Field | Required | Type | Description |
|-------------------------|----------|-------------|--|
| visit_occurrence_id | Yes | integer | A unique identifier for each person's visit or encounter at a healthcare provider. |
| person_id | Yes | integer | A foreign key identifier to the person for whom the visit is recorded. The demographic details of that person are stored in the person table. |
| visit_concept_id | Yes | integer | A foreign key that refers to a visit concept identifier in the Standardized Vocabularies. |
| visit_start_date | Yes | date | The start date of the visit. |
| visit_start_time | No | time | The time the visit started. |
| visit_end_date | Yes | date | The end date of the visit. If this is a one-day visit the end date should match the start date. |
| visit_end_time | No | time | The time the visit ended. |
| visit_type_concept_id | Yes | Integer | A foreign key to the predefined concept identifier in the Standardized Vocabularies reflecting the type of source data from which the visit record is derived. |
| provider_id | No | integer | A foreign key to the provider in the provider table who was associated with the visit. |
| care_site_id | No | integer | A foreign key to the care site in the care site table that was visited. |
| visit_source_value | No | Varchar(50) | The source code for the visit as it appears in the source data. |
| visit_source_concept_id | No | Integer | A foreign key to a concept that refers to the code used in the source. |

5.5.1 CONVENTIONS

- A Visit Occurrence is recorded for each visit to a healthcare facility.
- Valid Visit Concepts belong to the "Visit" domain.
- Standard Visit Concepts are defined as Inpatient Visit, Outpatient Visit, Emergency Room Visit and Long-Term Care Visit. Source concepts from place of service vocabularies are mapped into these standard visit concepts in the Standardized Vocabularies.
- Each Visit is standardized by assigning a corresponding Concept Identifier based on the type of facility visited and the type of services rendered.
- At any one day, there could be more than one visit.
- One visit may involve multiple providers, in which case the ETL must specify how a single provider id is selected or leave the provider_id field null.
- One visit may involve multiple care sites, in which case the ETL must specify how a single care_site id is selected or leave the care_site_id field null.

6 PROCEDURE_OCCURRENCE

The procedure domain contains records of activities or processes ordered by and/or carried out by a healthcare provider on the patient to have a diagnostic and/or therapeutic purpose.

Procedures are present in various data sources in different forms with varying levels of standardization. For example:

6.5.1 Medical Claims include CPT-4, ICD-9-CM (Procedures), and HCPCS procedure codes that are submitted as part of a claim for health services rendered, including procedures performed.

6.5.2 Electronic Health Records that capture CPT-4, ICD-9-CM (Procedures), HCPCS or OPCS-4 procedures as orders.

| Field | Required | Type | Description |
|-----------------------------|----------|-------------|---|
| procedure_occurrence_id | Yes | integer | A system-generated unique identifier for each procedure occurrence. |
| person_id | Yes | integer | A foreign key identifier to the person who is subjected to the procedure. The demographic details of that person are stored in the person table. |
| procedure_concept_id | Yes | integer | A foreign key that refers to a standard procedure concept identifier in the Standardized Vocabularies. |
| procedure_date | Yes | date | The date on which the procedure was performed. |
| procedure_type_concept_id | Yes | integer | A foreign key to the predefined concept identifier in the Standardized Vocabularies reflecting the type of source data from which the procedure record is derived. |
| modifier_concept_id | No | integer | A foreign key to a standard concept identifier for a modifier to the procedure (e.g. bilateral) |
| quantity | No | integer | The quantity of procedures ordered or administered. |
| provider_id | No | integer | A foreign key to the provider in the provider table who was responsible for carrying out the procedure. |
| visit_occurrence_id | No | integer | A foreign key to the visit in the visit table during which the procedure was carried out. |
| procedure_source_value | No | varchar(50) | The source code for the procedure as it appears in the source data. This code is mapped to a standard procedure concept in the Standardized Vocabularies and the original code is stored here for reference. Procedure source codes are typically ICD-9-Proc, CPT-4, HCPCS or OPCS-4 codes. |
| procedure_source_concept_id | No | integer | A foreign key to a procedure concept that refers to the code used in the source. |
| qualifier_source_value | No | varchar(50) | The source code for the qualifier as it appears in the source data. |

6.5.3 CONVENTIONS

- Valid Procedure Concepts belong to the "Procedure" domain. Procedure Concepts are based on a variety of vocabularies: SNOMED-CT, ICD-9-Proc, CPT-4, HCPCS and OPCS-4.
- Procedures are expected to be carried out within one day.
- Procedures could involve the application of a drug, in which case the procedural component is recorded in the procedure table and simultaneously the administered drug in the drug exposure table when both the procedural component and drug are identifiable.
- If the quantity value is omitted, a single procedure is assumed.
- The Procedure Type defines from where the Procedure Occurrence is drawn or inferred; for administrative claims records, the type indicates whether a Procedure was primary or secondary and their relative positioning within a claim.
- The Visit during which the procedure was performed is recorded through a reference to the VISIT_OCCURRENCE table. This information is not always available.
- The Provider carrying out the procedure is recorded through a reference to the PROVIDER table. This information is not always available.

7 DRUG_EXPOSURE

The drug exposure domain captures records about the inferred utilization of a biochemical substance with a physiological effect when ingested or otherwise introduced into the body. Drugs include prescription and over-the-counter medicines, vaccines, and large-molecule biologic therapies. Drug exposure is inferred from clinical events associated with orders, prescriptions written, pharmacy dispensing, procedural administrations, and other patient-reported information.

Drug Exposure records are recorded from a variety of source information:

7.5.1 The "Prescription" section of an EHR captures prescriptions written by physicians or from electronic ordering systems

7.5.2 The "Medication list" section of an EHR for both non-prescription products and medications prescribed by other providers

7.5.3 Prescriptions filled at dispensing providers such as pharmacies, and then captured in reimbursement claim systems

7.5.4 Drugs administered as part of a Procedure, such as chemotherapy or vaccines

Only drugs with active pharmaceutical ingredients are recorded. Radiological devices ingested or applied locally do not count as drugs.

| Field | Required | Type | Description |
|--------------------------|----------|-------------|---|
| drug_exposure_id | Yes | Integer | A system-generated unique identifier for each drug utilization event. |
| person_id | Yes | Integer | A foreign key identifier to the person who is subjected to the drug. The demographic details of that person are stored in the person table. |
| drug_concept_id | Yes | Integer | A foreign key that refers to a standard concept identifier in the Standardized Vocabularies for the drug concept. |
| drug_exposure_start_date | Yes | date | The start date for the current instance of drug utilization. Valid entries include a start date of a prescription, the date a prescription was filled, or the date on which a drug administration procedure was recorded. |
| drug_exposure_end_date | No | date | The end date for the current instance of drug utilization. It is not available from all sources. |
| drug_type_concept_id | Yes | integer | A foreign key to the predefined concept identifier in the Standardized Vocabularies reflecting the type of drug exposure recorded. It indicates how the drug exposure was represented in the source data: as medication history, filled prescriptions, etc. |
| stop_reason | No | varchar(20) | The reason the medication was stopped, where available. Reasons include regimen completed, changed, removed, etc. |
| Refills | No | integer | The number of refills after the initial prescription. The initial prescription is not counted, values start with 0. |
| Quantity | No | float | The quantity of drug as recorded in the original prescription or dispensing record. |
| days_supply | No | integer | The number of days of supply of the medication as recorded in the original prescription or dispensing record. |
| Sig | No | CLOB | The directions ("signetur") on the drug prescription as recorded in the original prescription (and printed on the container) or |

| | | | |
|------------------------|----|-------------|---|
| route_concept_id | No | integer | A foreign key to a predefined concept in the Standardized Vocabularies reflecting the route of administration. |
| effective_drug_dose | No | float | Numerical value of drug dose for this drug_exposure record. |
| dose_unit_concept_id | No | integer | A foreign key to a predefined concept in the Standardized Vocabularies reflecting the unit the effective_drug_dose value is expressed. |
| lot_number | No | varchar(50) | An identifier to determine where the product originated |
| provider_id | No | integer | A foreign key to the provider in the provider table who initiated (prescribed) the drug exposure. |
| visit_occurrence_id | No | integer | A foreign key to the visit in the visit table during which the drug exposure initiated. |
| drug_source_value | No | varchar(50) | The source code for the drug as it appears in the source data. This code is mapped to a standard drug concept in the Standardized Vocabularies and the original code is, stored here for reference. |
| drug_source_concept_id | No | Integer | A foreign key to a drug concept that refers to the code used in the source. |
| route_source_value | No | varchar(50) | The information about the route of administration as detailed in the source. |
| dose_unit_source_value | No | varchar(50) | The information about the dose unit as detailed in the source. |

7.5.5 CONVENTIONS

- Valid Drug Concepts belong to the "Drug" domain. Most Concepts in the Drug domain are based on RxNorm, but some may come from other sources. Concepts are members of the Clinical Drug or Pack, Branded Drug or Pack, Drug Component or Ingredient classes.
- Source drug identifiers, including NDC codes, Generic Product Identifiers, etc. are mapped to standard drug Concepts in the Standardized Vocabularies (e.g., based on RxNorm). When the Drug Source Value of the code cannot be translated into standard Drug Concept IDs, a Drug exposure entry is stored with only the corresponding source_concept_id and drug_source_value and a drug_concept_id of 0.
- The Drug Concept with the highest content of information is preferred during the mapping process: Concept Classes Branded Drug or Pack, followed by Clinical Drug, followed by Drug Component, and only if no other information is available the Ingredient. If only the drug class is known, no drug record should be written.
- A Drug Type is assigned to each Drug Exposure to track from what source the data were drawn or inferred.
- The Effective Drug Dose and the Dose Unit Concepts are provided in cases when the dose is explicitly provided, as it is typically for pediatric and chemotherapeutic treatments, and can only refer to a single active ingredient. Combination products which have doses for each ingredient need to be recorded as separate records
- If possible, the visit in which the drug was prescribed or delivered is recorded through a reference to the visit table.

8 DEVICE_EXPOSURE

The device exposure domain captures records about a person's inferred exposure to a foreign physical object or instrument that which is used for diagnostic or therapeutic purposes through a mechanism beyond chemical action. Devices include implantable objects (e.g. pacemakers, stents, artificial joints), durable medical equipment and supplies (e.g. bandages, crutches, syringes), and other instruments used in medical procedures (e.g. sutures, defibrillators).

| Field | Required | Type | Description |
|----------------------------|----------|-------------|--|
| device_exposure_id | Yes | integer | A system-generated unique identifier for each device exposure. |
| person_id | Yes | integer | A foreign key identifier to the person who is subjected to the procedure. The demographic details of that person are stored in the person table. |
| device_concept_id | Yes | integer | Only the DI portion of the UDI would be captured as a Concept in the Standardized Vocabularies. |
| device_exposure_start_date | Yes | date | The date the device or supply was applied or used. |
| device_exposure_end_date | No | date | The date the device or supply was removed from use. |
| device_type_concept_id | Yes | integer | Provenance for the data, e.g. procedure device, from registry, etc. |
| unique_device_id | No | varchar(50) | The entire UDI or equivalent. |
| quantity | No | integer | The number of individual devices used for the exposure |
| provider_id | No | integer | A foreign key to the provider in the provider table who was responsible for using the device. |
| visit_occurrence_id | No | integer | A foreign key to the visit in the visit table during which the device was used. |
| device_source_value | No | varchar(50) | The source code for the device as it appears in the source data. This code is mapped to a standard device concept in the Standardized Vocabularies and the original code is stored here for reference. |
| device_source_concept_id | No | integer | A foreign key to a device concept that refers to the code used in the source. |

8.5.1 CONVENTIONS

- Valid Device Concepts belong to the "Device" domain.
- The distinction between devices or supplies and procedures are sometimes blurry, but the former are physical objects while the latter are actions, often to apply a device or supply.
- For medical devices that are regulated by the FDA, a Unique Device Identification (UDI) is required if available in the data source and is recorded in the unique_device_id field.
- The DI portion of that UDI is used to define concepts in the CONCEPT table. However, devices are also defined based on other source vocabularies, like HCPCS.
- The Visit during which the device was first used is recorded through a reference to the VISIT_OCCURRENCE table. This information is not always available.
- The Provider exposing the patient to the device is recorded through a reference to the PROVIDER table. This information is not always available.

9 CONDITION_OCCURRENCE

The condition occurrence domain captures records of clinical observations of a person suggestive of the existence of disease or a medical condition based on diagnoses, signs and/or symptoms observed by a provider or reported by a patient.

Conditions are recorded in different sources and levels of standardization. For example:

9.5.1 Medical claims data include ICD-9-CM diagnosis codes that are submitted as part of a claim for health services and procedures.

9.5.2 EHRs may capture Person conditions in the form of diagnosis codes and symptoms as ICD-9-CM codes but may not have a way to capture out-of-system conditions.

| Field | Required | Type | Description |
|-----------------------------|----------|-------------|--|
| condition_occurrence_id | Yes | integer | A unique identifier for each condition occurrence event. |
| person_id | Yes | integer | A foreign key identifier to the person who is experiencing the condition. The demographic details of that person are stored in the person table. |
| condition_concept_id | Yes | integer | A foreign key that refers to a standard condition concept identifier in the Standardized Vocabularies. |
| condition_start_date | Yes | date | The date when the instance of the condition is recorded. |
| condition_end_date | No | date | The date when the instance of the condition is considered to have ended. If this information is not available, set to NULL. |
| condition_type_concept_id | Yes | integer | A foreign key to the predefined concept identifier in the Standardized Vocabularies reflecting the source data from which the condition was recorded, the level of standardization, and the type of occurrence. For example, conditions may be defined as primary or secondary diagnoses, problem lists and person statuses. |
| stop_reason | No | varchar(20) | The reason, if available, that the condition was no longer recorded, as indicated in the source data. Valid values include discharged, resolved, etc. Note that a stop_reason does not necessarily imply that the condition is no longer occurring. |
| provider_id | No | integer | A foreign key to the provider in the provider table who was responsible for determining (diagnosing) the condition. |
| visit_occurrence_id | No | integer | A foreign key to the visit in the visit table during which the condition was determined (diagnosed). |
| condition_source_value | No | varchar(50) | The source code for the condition as it appears in the source data. This code is mapped to a standard condition concept in the Standardized Vocabularies and the original code is, stored here for reference. Condition source codes are typically ICD-9-CM diagnosis codes from medical claims or discharge status/visit diagnosis codes from EHRs. |
| condition_source_concept_id | No | integer | A foreign key to a condition concept that refers to the code used in the source. |

5.9.1 CONVENTIONS

- Valid Condition Concepts belong to the "Condition" domain. Standard Condition Concepts are based on SNOMED-CT.
- Condition records are typically inferred from diagnostic codes recorded in the source data. Such code system, like ICD-9-CM, ICD-10-CM, read etc., provide a comprehensive coverage of conditions. However, if the code does not define a condition, but rather an observation or a procedure, then such information is not stored in the CONDITION_OCCURRENCE table, but in the respective tables instead.
- Source Condition identifiers are mapped to Standard Concepts for Conditions in the Standardized Vocabularies. When the source code cannot be translated into a Standard Concept, a CONDITION_OCCURRENCE entry is stored with only the corresponding source_concept_id (if available) and source_value and a condition_concept_id of 0.
- Family history and past diagnoses ("history of") are not recorded in the CONDITION_OCCURRENCE table. Instead, they are listed in the OBSERVATION table.
- Codes written in the process of establishing the diagnosis, such as "question of" of and "rule out", are not represented here. Instead, they are listed in the OBSERVATION table, if they are used for analyses.
- A Condition Occurrence Type is assigned based on the data source and type of condition attribute, including:
 - ICD-9-CM Primary Diagnosis from Inpatient and Outpatient Claims
 - ICD-9-CM Secondary Diagnoses from Inpatient and Outpatient Claims
 - Clinician diagnoses or problem Concepts from EHRs

10 MEASUREMENT

A measurement is the capture of a structured value (numerical or categorical) obtained through systematic examination of a person or sample. The Measurement domain captures measurement orders and measurement results. The measurement domain can contain laboratory results, vital signs, quantitative findings from pathology reports, etc.

| Field | Required | Type | Description |
|-----------------------------|----------|---------|---|
| measurement_id | Yes | integer | A unique identifier for each measurement. |
| person_id | Yes | integer | A foreign key identifier to the person about whom the measurement was recorded. The demographic details of that person are stored in the person table. |
| measurement_concept_id | Yes | integer | A foreign key to the standard measurement concept identifier in the Standardized Vocabularies. |
| measurement_date | Yes | date | The date of the Measurement. |
| measurement_time | No | time | The time of the Measurement |
| measurement_type_concept_id | Yes | integer | A foreign key to the predefined concept identifier in the Standardized Vocabularies reflecting the type of data on which the measurement record is based. |
| operator_concept_id | No | integer | A foreign key identifier to the mathematical operator that is applied to the value_as_number. Operators are <, ≤, =, ≥, > |
| value_as_number | No | float | A measurement stored as a number. This is applicable to measurement where the result is expressed as a numeric value. |
| value_as_concept_id | No | integer | A foreign key to a measurement stored as a concept identifier. This is applicable to measurements where the result can be expressed as a standard concept from the Standardized Vocabularies (e.g., positive/negative, present/absent, low/high, etc.). |
| unit_concept_id | No | integer | A foreign key to a standard concept identifier of measurement units in the Standardized Vocabularies. |
| range_low | No | float | The lower limit of the normal range of the measurement. The lower range is assumed to be in the same units of measure as the measurement value. |
| range_high | No | float | The upper limit of the normal range of the measurement. The lower range is assumed to be in the same units of measure as the measurement value. |
| provider_id | No | integer | A foreign key to the provider in the provider table who was responsible for making the measurement. |
| visit_occurrence_id | No | integer | A foreign key to the visit in the visit table during which the measurement was recorded. |

| | | | |
|-------------------------------|----|-------------|---|
| measurement_source_value | No | varchar(50) | The measurement name as it appears in the source data. This code is mapped to a standard concept in the Standardized Vocabularies and the original code is, stored here for reference. |
| measurement_source_concept_id | No | integer | A foreign key to a concept that refers to the code used in the source. |
| unit_source_value | No | varchar(50) | The source code for the unit as it appears in the source data. This code is mapped to a standard unit concept in the Standardized Vocabularies and the original code is, stored here for reference. |
| value_source_value | No | varchar(50) | The source value associated with the structured value stored as numeric or concept. This field can be used in instances where the source data are transformed to produce the structured value. |

10.5.1 CONVENTIONS

- Valid Measurement Concepts for both the measure (measurement_concept_id) and the measure result (value_as_concept) belong to the "Observation" domain. Measurement Concepts are based mostly on the LOINC vocabulary, with some additions from SNOMED-CT.
- Measurements are stored as attribute value pairs, where the attribute is the measure and the value represent the result. The value can be a concept (stored in value_as_concept), or a numerical value (value_as_number). The availability of a result is not mandatory.
- If reference ranges for upper and lower limit of normal as provided (typically by a laboratory) are stored in the range_high and range_low fields. Ranges have the same unit as the value_as_number.
- The Visit during which the observation was made is recorded through a reference to the VISIT_OCCURRENCE table. This information is not always available.
- The Provider making the observation is recorded through a reference to the PROVIDER table. This information is not always available.

11 OBSERVATION

The observation domain captures any clinical facts about a patient obtained in the context of examination, questioning or a procedure. The observation domain supports capture of data not represented by other domains, including unstructured measurements, medical history, and family history.

| Field | Required | Type | Description |
|-------------------------------|----------|--------------|--|
| observation_id | Yes | integer | A unique identifier for each observation. |
| person_id | Yes | integer | A foreign key identifier to the person about whom the observation was recorded. The demographic details of that person are stored in the person table. |
| observation_concept_id | Yes | integer | A foreign key to the standard observation concept identifier in the Standardized Vocabularies. |
| observation_date | Yes | date | The date of the observation. |
| observation_datetime | No | time | The time of the observation. |
| observation_type_concept_id | Yes | integer | A foreign key to the predefined concept identifier in the Standardized Vocabularies reflecting the type of the observation. |
| value_as_number | No | float | The observation result stored as a number. This is applicable to observations where the result is expressed as a numeric value. |
| value_as_string | No | varchar(60) | The observation result stored as a string. This is applicable to observations where the result is expressed as verbatim text. |
| value_as_concept_id | No | Integer | A foreign key to an observation result stored as a concept identifier. This is applicable to observations where the result can be expressed as a standard concept from the Standardized Vocabularies (e.g., positive/negative, present/absent, low/high, etc.). |
| qualifier_concept_id | No | integer | A foreign key to a standard concept identifier for a qualifier (e.g., severity of drug-drug interaction alert) |
| unit_concept_id | No | integer | A foreign key to a standard concept identifier of measurement units in the Standardized Vocabularies. |
| provider_id | No | integer | A foreign key to the provider in the provider table who was responsible for making the observation. |
| visit_occurrence_id | No | integer | A foreign key to the visit in the visit table during which the observation was recorded. |
| visit_detail_id | No | integer | The VISIT_DETAIL record during which the Observation occurred. For example, if the Person was in the ICU at the time the VISIT_OCCURRENCE record would reflect the overall hospital stay and the VISIT_DETAIL record would reflect the ICU stay during the hospital visit. |
| observation_source_value | No | varchar (50) | The observation code as it appears in the source data. This code is mapped to a standard concept in the Standardized Vocabularies and the original code is, stored here for reference. |
| observation_source_concept_id | No | integer | A foreign key to a concept that refers to the code used in the source. |

| | | | |
|------------------------|----|-------------|---|
| unit_source_value | No | varchar(50) | The source code for the unit as it appears in the source data. This code is mapped to a standard unit concept in the Standardized Vocabularies and the original code is, stored here for reference. |
| qualifier_source_value | No | varchar(50) | The source value associated with a qualifier to characterize the observation |

11.5.1 CONVENTIONS

- Valid Observation Concepts for the object (observation_concept_id) belong to the "Observation" domain. Observation Concepts are based mostly on the LOINC vocabulary, with some additions from SNOMED-CT.
- Valid Observation Concepts and the finding (value_as_concept_id) are not enforced by a domain but should be Standard Concepts.
- Observations must have an object represented as a concept, and a finding, represented as a concept, a numerical value or a verbatim string. There should be no observations records without an associated value. Observations which appear to be suggestive statements of positive assertion should have a recorded value as concept of 'Yes'.
- Observations obtained using standardized methods (e.g. laboratory assays) that produce discrete results are recorded by preference in the MEASUREMENT table.
- The Visit during which the observation was made is recorded through a reference to the VISIT_OCCURRENCE table. This information is not always available.
- The Provider making the observation is recorded through a reference to the PROVIDER table. This information is not always available.

Standardized Health System Data Tables

These tables describe the healthcare provider system responsible for administering the healthcare of the patient, rather than the demographic or clinical events the patient is involved in.

Below provides an entity-relationship diagram highlighting the tables within the Standardized Health System portion of the OMOP Common Data Model:

12 LOCATION

The Location table represents a generic way to capture physical location or address information. Locations are used to define the addresses for Persons and Care Sites.

| Field | Required | Type | Description |
|-----------------------|----------|-------------|---|
| location_id | Yes | integer | A unique identifier for each geographic location. |
| zip | No | varchar(9) | The zip or postal code. For US addresses, valid zip codes can be 3, 5 or 9 digits long, depending on the source data. |
| location_source_value | No | varchar(50) | The verbatim information that is used to uniquely identify the location as it appears in the source data. |

12.5.1 CONVENTIONS

- Each address or Location is unique and is present only once in the table.
- Locations do not contain names. In order to construct a full address that can be used on the Postal Service, the address information from the Location needs to be combined with information from the Care Site. The Person table does not contain name information.
- All fields in the Location tables contain the verbatim data in the Source. None of them are mandatory, but a valid Location record should at least contain either a Location Name or Location Zip.
- Zip codes are handled as strings of up to 9 characters length. For US addresses, these represent either a 3-digit abbreviated Zip code as provided by many Sources for Patient protection reasons, or the full 5-digit Zip code or the 9-digit (ZIP + 4) codes are recorded. Unless for specific reasons, analytical methods should expect and utilize only the first 3 digits. For international addresses, different rules apply.

13 CARE_SITE

The Care Site table contains a list of uniquely identified physical or organizational units where healthcare delivery is practiced (offices, wards, hospitals, clinics, etc.).

| Field | Required | Type | Description |
|-------------------------------|----------|--------------|---|
| care_site_id | Yes | Integer | A unique identifier for each organization. Here, an organization is defined as a collection of one or more care sites that share a single EHR database. |
| care_site_name | No | varchar(255) | The description of the care site |
| place_of_service_concept_id | No | Integer | A foreign key that refers to a place of service concept identifier in the Standardized Vocabularies. |
| location_id | No | Integer | A foreign key to the geographic location of the administrative offices of the organization in the location table, where the detailed address information is stored. |
| care_site_source_value | No | varchar(50) | The identifier for the organization in the source data, stored here for reference. |
| place_of_service_source_value | No | varchar(50) | The source code for the place of service as it appears in the source data, stored here for reference. |

14.5.2 CONVENTIONS

- There can be hierarchical and business relationships between Care Sites (e.g., wards can belong to clinics, which can in turn belong to hospitals, which in turn can belong to hospital systems, which in turn can belong to HMOs). These relationships should be defined in the FACT_RELATIONSHIP table.
- The Care Site Source Value typically contains the name of the Care Site.
- The Place of Service Concepts belongs to the Domain "Provider". These Concepts are based on a catalog maintained by the CMS

14 PROVIDER

The Provider table contains a list of uniquely identified health care providers. These are typically physicians, nurses, etc.

| Field | Required | Type | Description |
|-----------------------------|----------|-------------|---|
| provider_id | Yes | Integer | A unique identifier for each provider. |
| specialty_concept_id | No | Integer | A foreign key to a standard provider's specialty concept identifier in the Standardized Vocabularies. |
| care_site_id | No | Integer | A foreign key to the main care site where the provider is practicing. |
| provider_source_value | No | varchar(50) | The identifier used for the provider in the source data, stored here for reference. |
| specialty_source_value | No | varchar(50) | The source code for the provider specialty as it appears in the source data, stored here for reference. |
| specialty_source_concept_id | No | integer | A foreign key to a concept that refers to the code used in the source. |

14.5.1 CONVENTIONS

- Providers are not duplicated in the table.
- Valid Specialty Concepts for both the test (measurement_concept_id) belong to the "Provider" domain. The Specialty Concepts are based on the CDC specialty classification.
- This table is used to represent fixed relationship between Providers and Care Sites. Providers are also linked to Care Sites through Condition, Procedure and Visit records

OHDSI Release notes v5.3

CDM v5.3 ⁸⁹ <https://github.com/OHDSI/CommonDataModel/releases/tag/v5.3.0>

This version includes the following changes:

- [#64](#) ¹ Removes the datetime fields from OBSERVATION_PERIOD
- [#70](#) ⁴ Adds the VISIT_DETAIL table
- [#79](#) Adds the METADATA table
- [#92](#) ² Fixes modifier typo in PROCEDURE_OCCURRENCE
- [#120](#) ² Adds the following fields to PAYER_PLAN_PERIOD:
 - PAYER_CONCEPT_ID
 - PAYER_SOURCE_CONCEPT_ID
 - PLAN_CONCEPT_ID
 - PLAN_SOURCE_CONCEPT_ID
 - SPONSOR_CONCEPT_ID
 - SPONSOR_SOURCE_CONCEPT_ID
 - STOP_REASON_CONCEPT_ID
 - STOP_REASON_SOURCE_VALUE
 - STOP_REASON_SOURCE_CONCEPT_ID