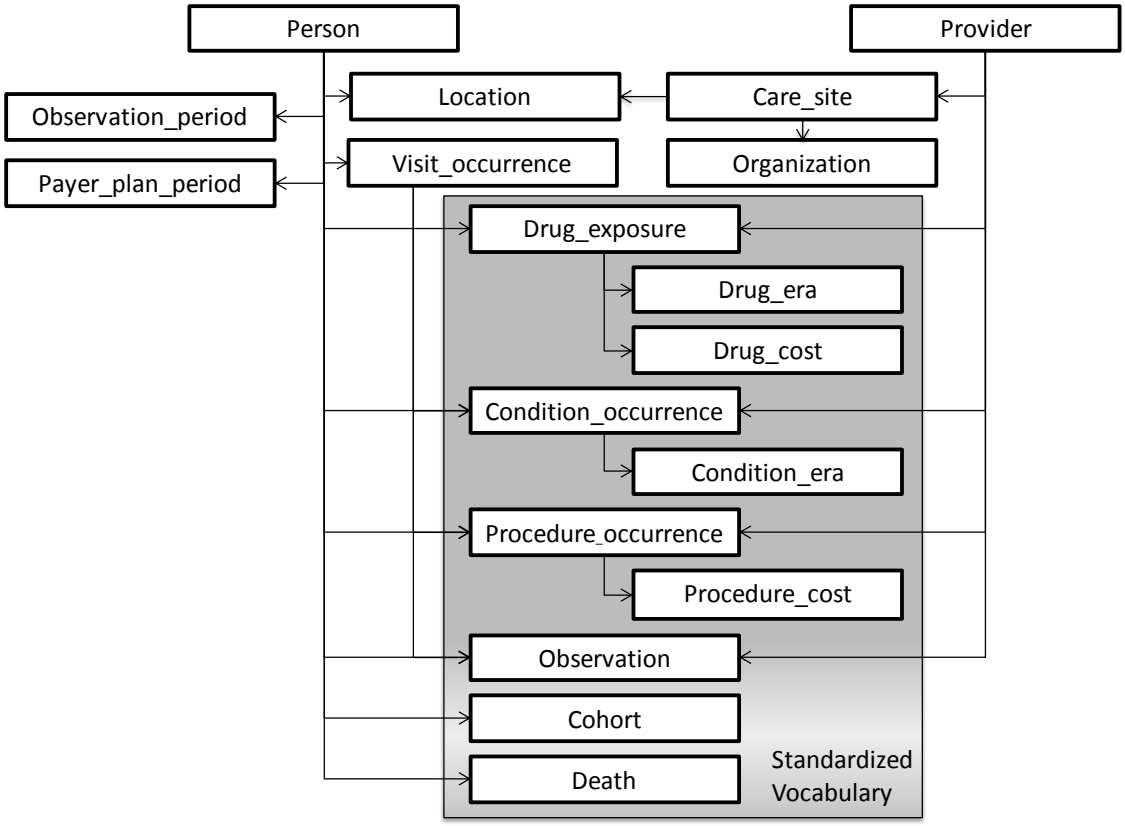
# Observational Medical Outcomes Partnership
# Common Data Model Specifications
# Version 4.0

**April 2012**



# License

# TABLE OF CONTENTS



# 1 Background

The Observational Medical Outcomes Partnership (OMOP, http://omop.fnih.org) is a public-private partnership designed to protect human health by improving the monitoring of medical, such as drugs or other regulated medical products, for safety and effectiveness. OMOP is funded and managed through the Foundation for the National Institutes of Health (FNIH), and draws on the expertise and resources of a large community from the pharmaceutical industry, academic institutions, non-profit organizations, the Food and Drug Administration (FDA), and other federal agencies.

The partnership began in the fourth quarter of 2008 to conduct research to determine the contribution and utility of using existing healthcare databases to identify and evaluate the effects of medical products.. OMOP's approach to its methodological research is the empirical evaluation of the performance of various analytical methods (http://omop.fnih.org/MethodsLibrary) for estimating the association between treatment and outcome across multiple disparate observational data sources. OMOP established a network of disparate data sources, both administrative claims and electronic health records that were maintained in a central research labor externally at a distributed partner.[1]

To achieve the research objective, OMOP created a suite of tools, such as a data model, experimental protocols, and database evaluation tools, which are available freely to the public domain. This has to purpose of encouraging collaborations within the community of scientific investigators. All project results are also made public in accordance with the public health mission of the partnership. These include comprehensive reports on scientific and technical findings, lessons learned, and peer-reviewed articles on the experimental findings by OMOP's sponsored investigators.

As part of the tool set, OMOP initially developed[2], and now improved and enhanced, a common structure and framework for organizing and standardizing observational data. The updated OMOP Common Data Model can accommodate use cases to perform research related to medical treatment outcome studies, including medical device safety, comparative effectiveness, and healthcare quality. This document describes the design and technical specifications of the OMOP Common Data Model (version 4).

## 1.1 The Role of the Common Data Model

No single observational data source is likely to be sufficient to meet all expected outcome analysis needs, so there is interest in assessing and analyzing multiple data sources concurrently. The OMOP Common Data Model (CDM) however is not intended to be an integration mechanism for multiple source datasets into a large pool. Instead, a separate CDM instance is expected to be generated for each source dataset, and summary results from each data source can be combined within a central coordinating center.

The CDM needs to support the conduct of research to identify and evaluate associations between interventions (drug exposure, procedures, healthcare policy changes etc.) and outcomes caused by these interventions (condition occurrences, procedures, drug exposure etc.). Outcomes can be efficacious (benefit) or adverse (risk). Often times, specific cohorts (e.g., myocardial infarction, acute liver failure) may be defined for treatments or outcomes, using clinical events (diagnoses, observations, procedures, etc.) that occur in predefined temporal relationships to each other. The CDM, combined with a method for

[1] Stang PE, Ryan PB, Racoosin JA, Overhage JM, Hartzema AG, Reich C, et al. Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. *Ann Intern Med*. 2010 Nov 2;153(9):600-6.
[2] Overhage JM, Ryan PB, Reich CG et al. Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc.* 2012;19:54-60.

standardizing its content (via the Vocabulary), will ensure that research methods can be systematically applied to produce meaningfully comparable results.

All analysis methods and code (e.g., SAS, SQL, or R programs) used to execute OMOP research protocols was developed for the CDM, with the purpose of enabling a common set of procedures to be applied to or be “portable” across participating data sources.

## 1.2 Design Principles

The CDM is designed to store observational data to allow for research, under the following principles:

1. **Data protection.** The CDM is aims at providing data storage optimal for analysis, instead of reflecting transactions in the course of patient care. In addition, all data that might jeopardize the identity and protection of patients, such as names, precise birthdays etc. are limited. Exceptions are possible where the research expressly requires more detailed information, such as precise birth dates for the study of infants.
2. **Reuse of existing models.** In designing the CDM, industry-leading data modeling efforts are leveraged, such as HL7 RIM, the HIMSS EHR Definitional Model, the i2b2 Hive framework, the HMORN Virtual Data Warehouse, etc.
3. **Design of domains.** The domains are modeled in a person-centric relational data model, where for each record the identity of the person and a date is captured as a minimum.
4. **Standard vocabulary.** To standardize the content of those records, the CDM relies on a Standard Vocabulary containing all necessary and appropriate corresponding standard healthcare concepts.
5. **Reuse of existing vocabularies.** If possible, these concepts are leveraged from national or industry standardization or vocabulary definition organizations or initiatives, such as the National Library of Medicine, the Department of Veterans' Affairs, the Center of Disease Control and Prevention, etc.
6. **Technology neutrality.** The CDM does not require a specific technology. It can be realized in any relational database, such as Oracle, MySQL etc., or as SAS analytical datasets. The tools the OMOP team or collaborators publish will be instantiated in a specific technology (OMOP uses both Oracle and SAS to store and analyze data) and may require some small adaptation if other technologies are utilized.
7. **Scalability.** The CDM is optimized for data processing and computational analysis to accommodate data sources that vary in size, up to and including databases with tens of millions of persons and billions of clinical observations.

## 1.3 Data Model for Data Tables

The CDM includes all observational data elements that are relevant for the identification of demographic information, health care interventions and outcomes. These data domains are comprised of the following:

- Person including demographic information
- Exposure to drug
- Occurrence of conditions (diagnoses)
- Administration of diagnostic and therapeutic procedures
- Other clinical observations, like lab results, diagnostic tests, signs and symptoms, etc.
- Periods of observation
- Patient visits to points of care
- Occurrence and cause of death
- Cost of drugs
- Cost of procedures, including inpatient and outpatient
- Information about the healthcare providers
- Information about the point of care and payer plan coverage
- Definition of cohorts for research purposes

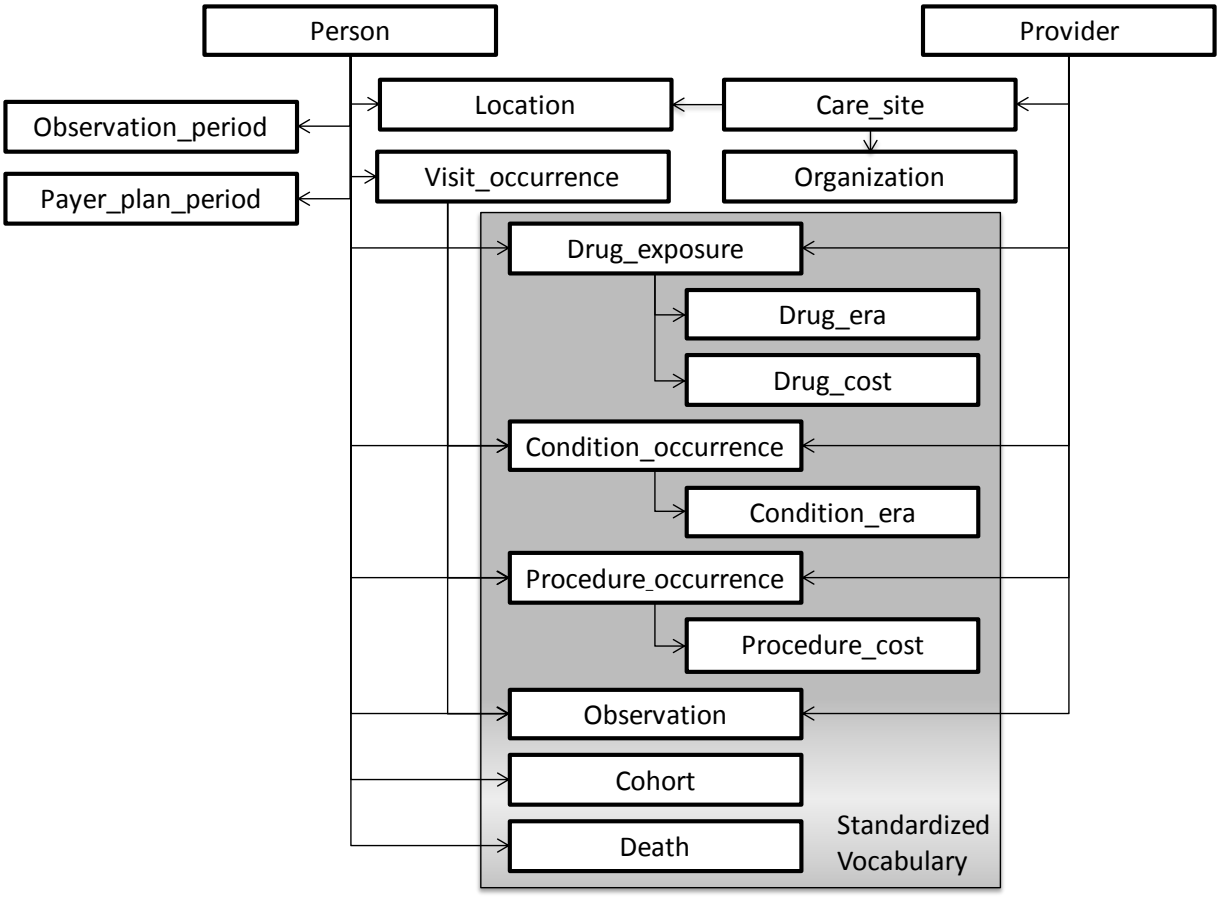
To represent these domains, the CDM contains 18 data tables:

| Table name | Description |
|---|---|
| Person | Demographic information about a Person |
| Drug Exposure | Association between a Person and a Drug at a specific time |
| Drug Era | Association between a Person and a Drug over a specific time period |
| Condition Occurrence | A diagnosis or condition that has been recorded about a person at a certain time |
| Condition Era | A diagnosis or condition over a period of time |
| Observation Period | Time intervals during which health care information, such as drugs, conditions, and other clinical observations, may be available |
| Observation | Observations are clinical facts, such as laboratory tests, signs/symptoms, which are not captured within other CDM tables |
| Procedure Occurrence | Procedures carried out on the Person |
| Visit Occurrence | Visits for health care services of the Person |
| Death | Time and cause of death of the Person |
| Drug Cost | For each Drug Exposure record additional information about cost and payments |
| Procedure Cost | For each Procedure additional information about cost and payments |
| Location | Physical addresses of patients, organizations and care sites |
| Provider | Information about health care providers |
| Organization | Information about health care organizations |
| Care Site | Information about the site of care |
| Payer Plan Period | Information about the coverage plan of the person |
| Cohort | Person, Provider or Visit cohorts |

The CDM defines table structures for each of the data in a Person and Provider-centric model. Almost all tables have foreign keys into the Person table and a date. This allows for a longitudinal view on all the healthcare-relevant events. In addition, Providers carrying out health care are linked to many of the events as well. Both are linked to healthcare organizations (hospitals, independent physician associations), care sites (doctor's offices, hospital departments etc.) and physical locations (addresses).

Diagram 1: CDM Conceptual Model




## 1.4 Data Model for Standard Vocabulary

The Standard Vocabulary is a semantic network containing all of the Concepts, Concept-to-Concept Relationships and other metadata necessary to describe the meanings and structures of the data within the CDM. The Vocabulary will accommodate Concepts for each of the entities of interest relative to drugs, conditions, procedures, visits, demographics, etc. The Conceptual data for the OMOP Vocabulary is a standardized format designed to integrate and standardize terminologies for observational analysis.

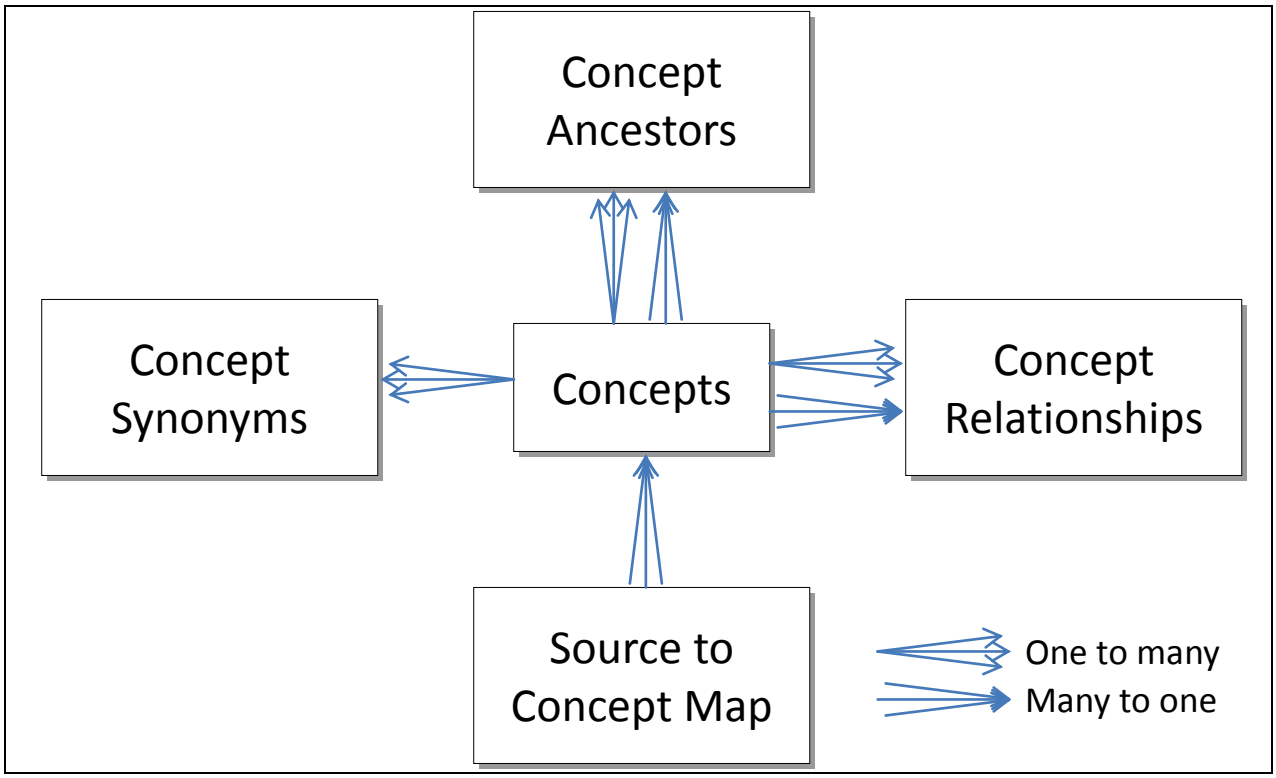
The Concepts are stored in the Vocabulary:

| Table name | Description |
|---|---|
| Concept | Contains all Concepts across domains, including the name (description), their source (called Concept Type), their identifier in the source (called Source Code), their level in the hierarchy and their class. |
| Concept Synonym | Contains any number of synonymous names (descriptions) for the Concepts. |
| Concept Relationship | Contains the Relationship between any two Concepts and the type of the Relationship. |
| Concept Ancestor | Contains the ancestry relationship of Concepts that have one or several hierarchical Relationships among them for fast lookup of hierarchical tree information, such as drug classes. |

| Table name | Description |
|---|---|
| Vocabulary | Contains a list of sources for the various Concepts. Many Concepts are derived from national or industry initiatives, such as the ICD diagnostic codes. Others are created by OMOP. |
| Source To Concept Map | Mapping between codes used in source data and Concepts in the Standard Vocabulary. |
| Relationship | List of all types of Relationships in the Concept Relationship Table and their names (descriptions). |

Diagram 2: The diagram below depicts its internal organization:




For example, the Vocabulary contains a single standard Concept with the Concept code 4249983, which means “Acute Myocardial Infarction, Anterolateral Wall, Initial Episode of Care.” The mappings from the Concept to its various source-specific representations are captured in the Source-to-Concept Map section of the Vocabulary. The Concept Relationship table defines all direct inter-Concept relationships (e.g., parent-child, composite-component, etc.). Finally, Concept Ancestor contains all direct and hierarchical ancestral relationships (e.g. multi-step relations, such as grandparent-child) between any Concepts.

## 1.5 Data in a physical CDM instance

Loading a CDM instance from a source dataset standardizes the data, both in format and in representation, to ensure that data management tools and analytical methods applied to the CDM instance will be portable to any other CDM instance. This transformation of source data to CDM data is performed in the following ETL process:

1. Extraction of source data
2. Transformation (i.e., reformatting) the source dataset content to conform to the CDM table structures. This involves the lookup of the source data in the Source To Concept Map and the identification of the corresponding standard Concept.
3. Loading into the CDM and making available to researchers

## 1.6 Data Model Conventions

There are a number of implicit and explicit conventions that have been adopted in the CDM. Developers of methods that run methods against the CDM need to understand these conventions.

- **Difference between Concept IDs and Source Values.** Many tables contain equivalent information twice: As a Source Value and as a Concept ID. Concepts are CDM-specific entities that have unique IDs across all domains, while Source Values or Source codes might overlap between domains. All Concept IDs are stored as foreign keys to the Concept table in the Vocabulary, where all the detailed information (name, relationships, types etc.) can be found. Source Values are only provided verbatim from the source for convenience and QA purposes. Source Values are optional, while Concepts are mandatory. Source Values may contain information that is only meaningful in the context of a specific data source. If a Source Value cannot be mapped to a Concept, the Concept ID value of 0 is used to represent an unmapped source value.

**Note:** Only the Concept IDs should be used for standardized analysis purposes. Source values should be used only for QA and reporting purposes.

- **Difference between Type Concepts** (ending in _type_concept_id) **and general Concepts** (ending in _concept_id). Many tables contain special fields for Type Content IDs. These are specialty Concepts with the purpose of indicating where the data are derived from in the source. For example, the Type Concept field can be used to distinguish a Drug Exposure record that is derived from a pharmacy dispensing claim from a Drug Exposure record indicative of a prescription written in an EHR
- **Other Identifiers.** The table contains a number of fields ending in "_id". This indicates a foreign key to another table, where the information is stored.
- **Precision of data types.** All precisions are provided at the minimal required value. For example, numbers for zip codes are 9 characters long. However, the data model can be instantiated with higher precision if necessary.

# 2 Changes between Versions 2.0 and 4.0

CDM Version 2.0 has been in production in a variety of different organizations and has shown its utility for outcome research purposes. A proposed updated CDM version is based on CDM V2.0 experiences and feedback from collaborators. In order to facilitate application to broader Comparative Effectiveness Research (CER) and Health Economics studies, a number of additional tables were necessary. In addition, the CDM Version 2.0 was also reviewed for idiosyncrasies, unused fields and other possible improvements to the existing schema, and the resulting changes have been incorporated into CDM Version 4.0.

## 2.1 Changes to existing tables

The changes can be summarized into four categories:

- The precision of fields were modified to reflect the value ranges found in observational data.
- Field names were changed to reflect a new convention: All verbatim data from the source stored in the CDM now end in "source_value".
- Reference values that used to be stored in a separate table (name ending in "_ref" or "_type") are now part of the vocabulary. The reference tables are obsolete. The fields referring to these tables (names ending in "_type") are now renamed to end in "type_concept_id". The exceptions to this are the reference tables of the Vocabulary itself.
- Some field names were revised to no longer contain the table name as a suffix in the field name.

| Changed table | Change |
|---|---|
| Vocabulary | Changed name of table from vocabulary_ref |
| Relationship | Changed name of table from relationship_type |

| Table | Changed field | Change |
|---|---|---|
| person | person_source_value | Changed name from source_person_key |
| person | gender_source_value | Changed name from source_gender_code |
| person | race_source_value | Changed name from source_race_code |
| drug_exposure | drug_type_concept_id | Changed name from drug_ exposure_type. Reference values are now stored as concepts in the concept table. |
| drug_exposure | quantity | Changed name from drug_quantity. Changed field size to 4. |
| drug_exposure | drug_source_value | Changed name from source_drug _code. Changed field size to 50. |
| drug_era | drug_type_concept_id | Changed name from drug_exposure_type. Reference values are now stored as concepts in the concept table. |
| drug_era | drug_exposure_count | Changed field size to 4 |
| condition_occurrence | condition_type_concept_id | Changed name from condition_occurrence_type. Reference values are now stored as concepts in the concept table. Mortality is no longer a condition occurrence type, but instead recorded in the separate death table. |
| condition_occurrence | condition_source_value | Changed name from source_condition_code. Changed field size to 50. |
| condition_era | condition_type_concept_id | Changed name from condition_occurrence_type. Reference values are now stored as concepts in the concept table. |

| Table | Changed field | Change |
|---|---|---|
| condition_era | condition_occurrence_count | Changed field size to 4 |
| visit_occurrence | visit_source_value | Changed name from source_visit_code |
| visit_occurrence | place_of_service_concept_id | Changed name from visit_concept_id |
| procedure_occurrence | procedure_type_concept_id | Changed name from procedure_occurrence_type. Reference values are now stored as concepts in the concept table. |
| procedure_occurrence | procedure_source_value | Changed name from source_procedure_code. Changed size to field 50. |
| observation | observation_id | Changed name from obs_occurrence_id |
| observation | observation_date | Changed name from obs_date |
| observation | value_as_number | Changed name from obs_value_as_number |
| observation | value_as_string | Changed name from obs_value_as_string |
| observation | value_as_concept_id | Changed name from obs_value_as_concept_id |
| observation | unit_concept_id | Changed name from obs_unit_concept_id |
| observation | range_low | Changed name from obs_range_low |
| observation | range_high | Changed name from obs_range_high |
| observation | observation_type_concept_id | Changed name from obs_type. Reference values are now stored as concepts in the concept table. |
| observation | observation_source_value | Changed name from source_obs_code. Changed field size to 50. |
| concept | vocabulary_id | Changed name from concept_vocabulary_code. Changed from string to integer. |
| concept_synonym | concept_synonym_name | Changed name from description_name |
| concept_relationship | relationship_id | Changed name from relationship_type |
| source_to_concept_map | source_vocabulary_id | Changed name from source_vocabulary_code |
| source_to_concept_map | target_vocabulary_id | Changed name from target_vocabulary_code |
| vocabulary | vocabulary_id | Changed name from vocabulary_code |
| Relationship | relationship_id | Changed name from relationship_type |
| Relationship | relationship_name | Changed name from relationship_description |

## 2.2 New Fields and Tables

A number of new tables where introduced to reflect the additional information that will be stored in the CDM. Death information will now be stored in a separate death table. This is a significant change from CDM V2.0, where death information could be a condition_occurrence, condition_occurrence_type or observation_period status. Other tables were added to store information about the point of care and the providers, as well as the cost and coverage of care.

| Table | Description |
|---|---|
| death | Contains information about a person's death and the associated reasons |
| drug_cost | Stores the one or more cost records for a person's drug exposure |
| procedure_cost | Stores the one or more cost records for a person's procedure or visit |
| location | Stores geographic location (addresses) |
| provider | Stores the descriptors and details of the health care provider |
| organization | Stores the details of a healthcare organization |
| care_site | Stores information about the care site that is part of an organization and where a provider delivers healthcare |
| payer_plan_period | Stores information about the Person’s coverage plan |
| cohort | Stores cohorts of Persons, Providers or Visits |

In addition to new tables, a number of fields were added to existing tables:

| Table | Field | Description |
|---|---|---|
| person | month_of_birth | Month and day of birth were added as optional fields to the existing year of birth. |
| person | day_of_birth | |
| person | ethnicity_concept_id | The race and ethnicity information was split into two fields: The actual race_concept_id and corresponding source_value and the ethnicity. |
| person | ethnicity_source_value | |
| person | location_id | The location concepts were replaced with references to records in a separate location table. |
| person | provider_id | Information about a person's primary care provider, the organization that provider belongs to and the site of where the care happens where added. |
| person | care_site_id | |
| drug_exposure | sig | The directions ("signetur") on the drug prescription. |

| Table | Field | Description |
|---|---|---|
| drug_exposure | prescribing_provider_id | A foreign key to the provider who prescribed the medication. |
| drug_exposure | relevant_condition_concept_id | The information about the condition that led to the drug exposure event. This is only relevant for some data sources. |
| drug_exposure | visit_occurrence_id | A foreign key to the visit. In V2.0, visits were not directly linked to drug exposure events. |
| condition_occurrence | associated_provider_id | A foreign key to the provider who was responsible for determining (diagnosing) the condition. |
| condition_occurrence | visit_occurrence_id | A foreign key to the visit. In V2.0, visits were not directly linked to condition occurrences. |
| visit_occurrence | care_site_id | A foreign key to source information about the care site that was visited. |
| procedure_occurrence | provider_id | A foreign key to the provider who was responsible for carrying out the procedure. |
| procedure_occurrence | visit_occurrence_id | A foreign key to the visit during which the procedure was carried out. |
| procedure_occurrence | relevant_condition_concept_id | The information about the condition that led to the procedure. This is only relevant for some data sources. |
| observation | observation_time | The time of the observation in addition to the existing date field. |
| observation | visit_occurrence_id | A foreign key to the visit during which the observation was recorded. |
| observation | relevant_condition_concept_id | The information about the condition that led to the observation. This is only relevant for some data sources. |
| observation | unit_source_value | The source code for the unit as it appears in the source data. This code is mapped to a standard unit concept in the vocabulary and the original code is, stored here for reference. |
| concept | valid_start_date | The date when the instance of the concept is first recorded. |
| concept | valid_end_date | The date when the instance of the concept is last recorded. |
| concept | invalid_reason | The reason the concept is not valid. |
| concept_relationship | valid_start_date | The date when the instance of the relationship is first recorded. |

| Table | Field | Description |
|---|---|---|
| concept_relationship | valid_end_date | The date when the instance of the relationship is last recorded. |
| concept_relationship | invalid_reason | The reason the relationship is not valid. |
| source_to_concept_map | primary_map | To indicate the primary of several alternative records mapping the same source_code, mapping_type and target_concept_id. |
| source_to_concept_map | valid_start_date | The date when the instance of the concept is first recorded. |
| source_to_concept_map | valid_end_date | The date when the instance of the concept is last recorded. |
| source_to_concept_map | invalid_reason | The reason the concept is not valid. |

## 2.3 Obsolete Fields and Tables

The following tables and fields have become obsolete and are not part of the CDM:

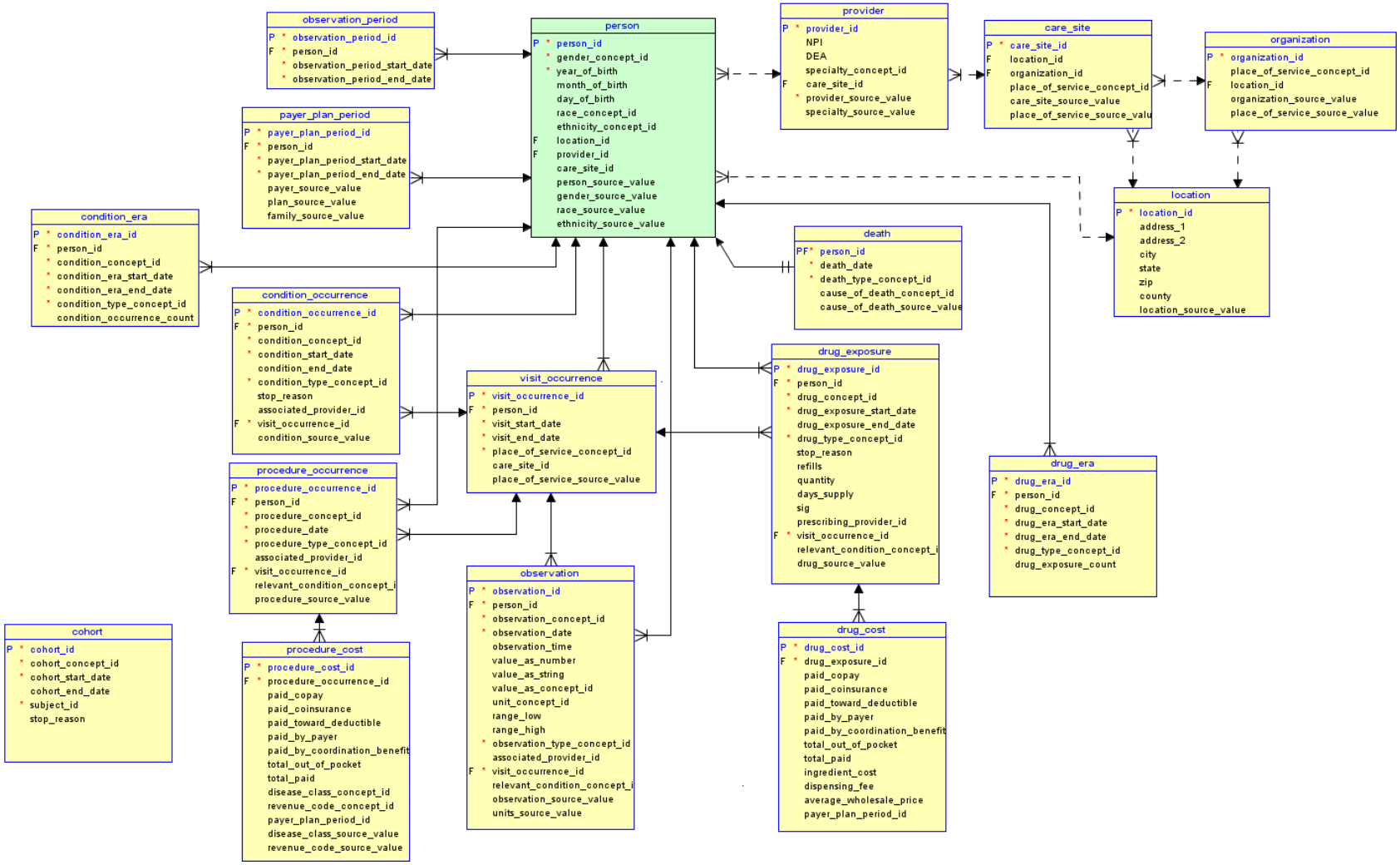
| Obsolete table | Rationale for becoming obsolete |
|---|---|
| drug_exposure_ref | Drug exposure types are no longer stored in this separate reference table, but as concepts in the vocabulary. |
| condition_occurrence_ref | Condition types are no longer stored in this separate reference table, but as concepts in the vocabulary. |
| proc_occurrence_ref | Procedure types are no longer stored in this separate reference table, but as concepts in the vocabulary. |
| observation_type_ref | Observation types are no longer stored in this separate reference table, but as concepts in the vocabulary. |

| Table | Obsolete field | Rationale for becoming obsolete |
|---|---|---|
| person | location_concept_id | The location is no longer stored as a concept for 3-digit zip codes, census regions and states, but in the location table. |
| person | source_location_code | The location is no longer stored as a concept for 3-digit zip codes, census regions and states, but in the location table. |

| Table | Obsolete field | Rationale for becoming obsolete |
|---|---|---|
| condition_occurrence | dx_qualifier | Only diagnoses occurring at the time of the record are allowed in the condition_occurrence table. Qualifiers like "family history of", "history of", "recurrence", "risk of", "rule-out" and similar qualifiers for diagnoses are no longer represented in the condition table. Indicators for "hospitalization" are recorded in the visit table. |
| condition_era | Confidence | Confidence values have not been proven practical or useful, and are therefore abandoned. |
| observation_period | rx_data_availability | Observation periods are expected to cover durations for which data of all dimensions (drugs, medical and hospital) are captured. |
| observation_period | dx_data_availability | |
| observation_period | hospital_data_availability | |
| observation_period | confidence | Confidence values have not been proven practical or useful, and are therefore abandoned. |
| observation_period | person_status_concept_id | Information previously stored in this field such as "Active", "Inactive", "Obsolete", "Deceased", etc. are either removed during ETL, captured as an observation, or stored in the death table. |
| source_to_concept_map | source_to_concept_map_id | Unique record identifier not required. |
| concept_relationship | concept_relationship_id | Unique record identifier not required. |
| concept_ancestor | concept_ancestor_map_id | Unique record identifier not required. |

# 3 Data Tables

Diagram 3: The entity-relationship diagram (ERD) of the CDM data tables and relationships between them. All relationships are not displayed.



## 3.1 Person

The Person table is one of the basic four mandatory dimensions of analysis, and when combined with the Drug Exposure, Condition, Observation, and Procedure entities, presents the framework for active drug surveillance. The source data for the Person table comes from person demographics data that will be de-identified to ensure HIPAA compliance. The extent of these data varies by data source. The Person table attribute values are stored as standard Concept codes mapped to the original (i.e., "raw") source values.

| Field | Required | Type Precision | Standard | Description |
|---|---|---|---|---|
| person_id | Yes | integer | | A system-generated unique identifier for each person. |
| gender_concept_id | Yes | integer | HL7 | A foreign key that refers to a standard concept identifier in the vocabulary for the gender of the person. |
| year_of_birth | Yes | number(4) | | 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. |
| month_of_birth | No | number(2) | | The month of birth of the person. For data sources that provide the precise date of birth, the month is extracted and stored in this field. |
| day_of_birth | No | number(2) | | The day of the month of birth of the person. For data sources that provide the precise date of birth, the day is extracted and stored in this field. |
| race_concept_id | No | integer | OMB, CDC | A foreign key that refers to a standard concept identifier in the vocabulary for the race of the person. |
| ethnicity_concept_id | No | integer | OMB | A foreign key that refers to the standard concept identifier in the vocabulary 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 primary care site in the care site table, where the details of the care site are stored. |
| person_source_value | No | string(50) | | An encrypted key derived from the person identifier in the source data. This is necessary when a drug safety issue requires a link back to the person data at the source dataset. No value with any medical or demographic significance must be stored. |

| Field | Required | Type Precision | Standard | Description |
|---|---|---|---|---|
| gender_source_value | No | string(50) | | The source code for the gender of the person as it appears in the source data. The person gender is mapped to a standard gender concept in the vocabulary and the corresponding concept identifier is, stored here for reference. |
| race_source_value | No | string(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 vocabulary and the original code is, stored here for reference. |
| ethnicity_source_value | No | string(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 vocabulary and the original code is, stored here for reference. |

### 3.1.1 BUSINESS RULES

- Person data will remain de-identified as much as possible to comply with the Design Principles. Accordingly, the precise date of birth will only be stored if other measures are taken to protect the patient information. Only the year of birth is mandatory, and no identifiers are stored that could be used to re-identify the Person data.
- The granularity of the Person data from the source system will be maintained. There will be no consolidation or aggregation of individual Person records.
- Standard attributes will be stored as Concept codes. Original source values will be mapped to the corresponding standard Concept codes in the Vocabulary.
- Person source data attributes are race, gender, and ethnicity. Additional information is stored through references to other tables about the home address (location) and the primary care provider: the provider, care site and organization.

### 3.1.2 EXAMPLE OF A LOADED TABLE

The following shows two typical records of a person table:

| Field | Record Example 1 | Record Example 2 |
|---|---|---|
| person_id | 121107 | 127260 |
| gender_concept_id | 8532 | 8532 |
| year_of_birth | 1932 | 1933 |
| month_of_birth | 5 | |
| day_of_birth | 31 | |
| race_concept_id | 8527 | 8558 |
| ethnicity_concept_id | 0 | 38003563 |
| location_id | integer foreign key | |
| provider_id | integer foreign key | |

| Field | Record Example 1 | Record Example 2 |
|---|---|---|
| care_site_id | integer foreign key | |
| person_source_value | 57bcc40b9080b35f781bc87dd8dc77b7 | cd759452e9577d52354cc327b86f0760 |
| gender_source_value | Female | Female |
| race_source_value | Caucasian | Dominican |
| ethnicity_source_value | | Hispanic |

All IDs are generated integers. Year of Birth is mandatory; however, for studies on infants the Month and Day might be necessary. The Person Source Value could be realized as a one-way hash key using MD5 hashing.

| Source Person Identifier | Source Person Key |
|---|---|
| 121107 | 57bcc40b9080b35f781bc87dd8dc77b7 |
| 127260 | cd759452e9577d52354cc327b86f0760 |

## 3.2 Drug Exposure

Drug Exposure contains individual records that reflect drug utilization from within the source data. Drug Exposure indicators include drug details (captured as standard Concept identifiers in the Vocabulary), drug quantity, number of days supply, period of exposure, and prescription refill data. Drug Exposure is recorded in a variety of ways.

- The “Prescription” section of an EHR captures prescriptions written by physicians.
- Other drugs (both non-prescription products and medications prescribed by other providers) used by a Person are recorded in the “Medications” section of the EHR.
- Administrative claim systems capture prescriptions filled at dispensing providers.
- Drug Exposure information as a by-product of certain procedure codes (i.e., procedure codes that refer to the administration of certain drugs, such as chemotherapy or vaccines).

The drug product is indicated in the CDM by standard drug Concepts from the Vocabulary. The standard Concept identifier for a drug is stored with the drug reference data; however, the Concept hierarchy and therapeutic class categorizations from the source data are not stored with the drug exposure data but can be obtained through the Vocabulary.

| Field | Required | Type Precision | Standard | 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 | RxNorm | A foreign key that refers to a standard concept identifier in the vocabulary 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. |

| Field | Required | Type Precision | Standard | Description |
|---|---|---|---|---|
| drug_type_concept_id | Yes | integer | OMOP | A foreign key to the predefined concept identifier in the vocabulary 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 | string(20) | | The reason the medication was stopped, where available. Reasons include regimen completed, changed, removed, etc. |
| refills | No | number(3) | | The number of refills after the initial prescription. The initial prescription is not counted, values start with 0. |
| quantity | No | number(4) | | The quantity of drug as recorded in the original prescription or dispensing record. |
| days_supply | No | number(4) | | The number of days of supply of the medication as recorded in the original prescription or dispensing record. |
| sig | No | string(500) | | The directions ("signetur") on the drug prescription as recorded in the original prescription (and printed on the container) or dispensing record. |
| prescribing_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. |
| relevant_condition_concept_id | No | integer | SNOMED | A foreign key to the predefined concept identifier in the vocabulary reflecting the condition that was the cause for initiation of the drug exposure. Note that this is not a direct reference to a specific condition record in the condition table, but rather a condition concept in the vocabulary. |
| drug_source_value | No | string(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 vocabulary and the original code is, stored here for reference. |

### 3.2.1 BUSINESS RULES

- Source drug identifiers, including NDC codes, Generic Product Identifiers, etc. are mapped to standard drug Concepts in the Vocabulary (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 Drug Source Value and a Concept ID of 0.
- A Drug Type is assigned to each Drug Exposure to track from what source the data were drawn or inferred.
- If possible, the visit in which the drug was prescribed or delivered is recorded through a reference to the visit table.
- The Relevant Condition Concept is defined as the condition that is associated with the Drug Exposure. This is typically the indication, but could also be the condition to be determined for Drug Exposure as a diagnostic. This information is not typically available.
- As a minimum, the Person ID, Drug Concept ID, Start Date and Drug Type need to be available for a valid record.

### 3.2.2 EXAMPLE OF A LOADED TABLE

Consider the following example of care for patients receiving Warfarin treatment:

Patient 121801, from claims records:

- 14-Aug-07: Pharmacy claim for 30 tablets of Warfarin 7.5 MG Oral Tablet NDC 59772038670

Patient 121798, from EHR records:

- 29-Nov-08: Single dose of "Warfarin 7.5 MG Oral Tablet" GPI 83200030200320, listed in the medication list, stopped for reason "Other", indication ICD9 427.32 "Atrial flutter", during visit
- 19-Jan-09: Prescription written for "Warfarin 2.5 MG Oral Tablet" GPI 83200030200310, 30 tablets, one refill, over phone
- 25-Jun-10: Prescription written for "Warfarin 7.5 MG Oral Tablet" GPI 83200030200320, once daily, over phone
- 16-Jul-10: Prescribed medication stopped, over phone

Patient 234513, from EHR records:

- 13-Jan-09: Vaccination against measles during visit at doctor's office, ICD-9-Procedure code 99.45.

The above data are represented in the Drug Exposure table below. Note, there is no required date format date, but an example convention is shown.

| Field | Record Example 1 | Record Example 2 | Record Example 3 |
|---|---|---|---|
| drug_exposure_id | 150807907 | 32830 | 32832 |
| person_id | 121801 | 121798 | 121798 |
| drug_concept_id | 1310213 | 1310217 | 1310213 |
| drug_exposure_start_date | 14-Aug-07 | 29-Nov-08 | 19-Jan-09 |
| drug_exposure_end_date | | 29-Nov-08 | |
| drug_type_concept_id | 38000175 | 38000178 | 38000177 |
| stop_reason | | Other | |
| refills | 0 | | 1 |
| quantity | 30 | | |
| days_supply | | | 30 |
| sig | | | |
| prescribing_provider_id | | integer foreign key | |
| visit_occurrence_id | | integer foreign key | |
| relevant_condition_concept_id | | 314665 | |
| drug_source_value | 59772038670 | 83200030200320 | 83200030200310 |

| Field | Record Example 4 | Record Example 5 |
|---|---|---|
| drug_exposure_id | 131428827 | 90234867 |
| person_id | 121798 | 234513 |
| drug_concept_id | 1310217 | 594249 |
| drug_exposure_start_date | 25-Jun-08 | 13-Jan-09 |
| drug_exposure_end_date | 16-Jun-08 | |
| drug_type_concept_id | 38000177 | 38000179 |
| stop_reason | Removed | |

| Field | Record Example 4 | Record Example 5 |
|---|---|---|
| refills | | |
| quantity | | |
| days_supply | | |
| sig | qd | |
| prescribing_provider_id | | |
| visit_occurrence_id | | |
| relevant_condition_concept_id | | |
| drug_source_value | 83200030200320 | 99.45 |

Drug Exposure ID is an auto-generated integer for the record. The Person ID is a foreign key to the Person table. The Drug Concept IDs are for Warfarin 2.5 and 7.5 mg tablets. The Start Date is the date of the source record. The End Date is only given for the one dose exposure in Example 2. The Drug Type Concept IDs stand for "Prescription dispensed in pharmacy", "Medication list entry", "Prescription written" and "Physician administered drug (identified as procedure)". The Stop Reason is copied verbatim from the source. The number of Refills, Quantity, Days Supply and Sig are verbatim from the source.

Example 2 has been recorded as part of a Visit, so that the respective Visit ID and Provider ID foreign keys are provided. The indication "Atrial flutter" is listed as the respective Condition Concept ID, that is inferred by a Vocabulary lookup in the table Source To Concept Map table for Source Value='427.32' and with Mapping Type equal to 'CONDITION'. Note that this is not a foreign key to an actual Condition Occurrence record, but to a Condition Concept in the Vocabulary. It remains up to the ETL to ensure there is such a record in the table. This is in contrast to the Visit Occurrence ID or Prescribing Provider ID, which link directly to the corresponding Visit Occurrence where the Drug Exposure was initiated. Finally, the Drug Source Values are verbatim NDC, GPI and Procedure codes (ICD-9-Proc in this example).

## 3.3 Drug Era

A Drug Era is defined as a span of time when the Person is assumed to be exposed to a particular drug. A Drug Era is not the same as a Drug Exposure: Exposures are individual records corresponding to the source when drug was delivered to the Person, while successive periods of Drug Exposures are combined under certain rules to produce continuous Drug Eras.

| Field | Required | Type Precision | Standard | Description |
|---|---|---|---|---|
| drug_era_id | Yes | integer | | A system-generated unique identifier for each drug era. |
| person_id | Yes | integer | | A foreign key identifier to the person who is subjected to the drug during the drug era. The demographic details of that person are stored in the person table. |
| drug_concept_id | Yes | integer | RxNorm | A foreign key that refers to a standard concept identifier in the vocabulary for the drug concept. |
| drug_era_start_date | Yes | date | | The start date for the drug era constructed from the individual instances of drug exposures. It is the start date of the very first chronologically recorded instance of utilization of a drug. |
| drug_era_end_date | Yes | date | | The end date for the drug era constructed from the individual instance of drug exposures. It is the end date of the final continuously recorded instance of utilization of a drug. |

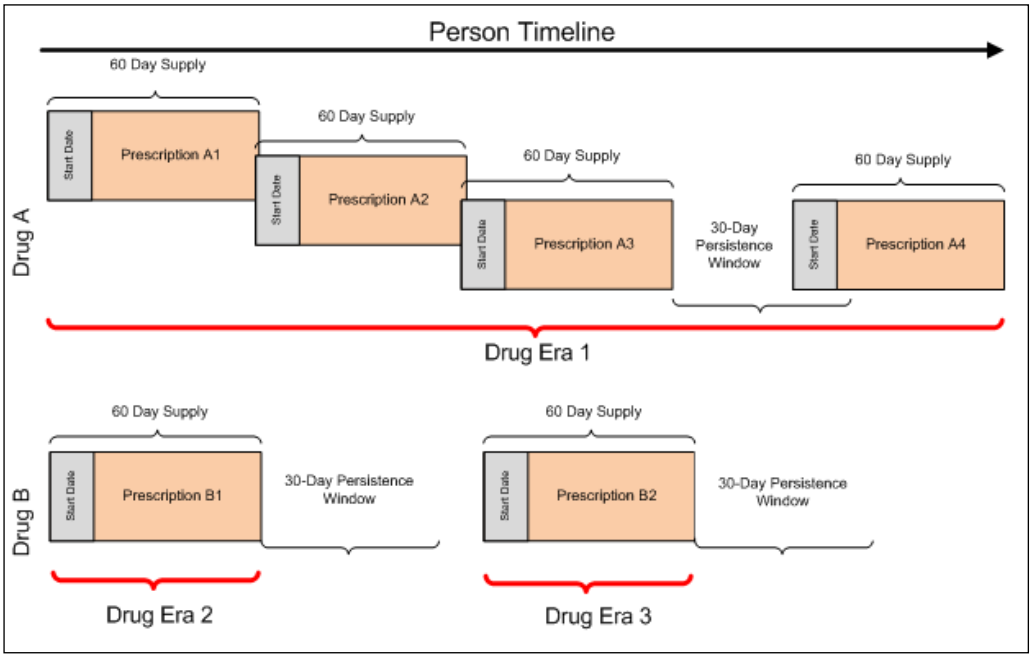
 

| Field | Required | Type Precision | Standard | Description |
|---|---|---|---|---|
| drug_type_concept_id | Yes | integer | OMOP | A foreign key to the predefined concept identifier in the vocabulary reflecting the parameters used to construct the drug era. |
| drug_exposure_count | No | number(4) | | The number of individual drug exposure occurrences used to construct the drug era. |

### 3.3.1 BUSINESS RULES

- Each Drug Era corresponds to one or many Drug Exposures that form a continuous interval.
- The Drug Concept ID is the same for all component Drug Exposures. However, Eras can also be built to aggregate drugs to, e.g., ingredients or drug class eras. In that case, the aggregate concept ID is populated into the Drug Concept ID field and all Drug Exposure records are consolidated into the Era that are children of that Concept as defined in the Concept Ancestor table in the Vocabulary.
- The Drug Era Start Date is the start date of the first Drug Exposure.
- The Drug Era End Date is the end date of the last Drug Exposure.
- The End Date of each Drug Exposure is either taken from the field Drug Exposure End Date or, as it is typically not available, inferred using the following rules:
  - For pharmacy prescriptions claims, the date when the drug was dispensed and the number of days of supply are used to extrapolate the End Date for the Drug Exposure.
  - For EHRs, the medications data often include the start and end dates for the medication. If not, the start and end dates are assumed identical.
  - For Procedure Drugs, usually the drug is administered on a single date (i.e., the administration date).
- To determine which Drug Exposures are combined to a Drug Era, the individual records are stacked together to form a continuous Era, as long as the Start Date of the following record precedes or is at the same date as the End Date of the previous record plus a period of "slack" called Persistence Window. Consider a Person who is taking two drugs: Drug A and Drug B (see Diagram 4). The Person has had four prescriptions for Drug A (A1, A2, A3, A4), each with a sixty-day supply. The Person has also had two prescriptions for Drug B (B1, B2).

Diagram 4: Drug Era Example



To define the Drug Era for Drug A, the timing, duration, overlap, and persistence of the Person's prescriptions for Drug A must be considered. A2 was filled before the expected completion of A1. Similarly, A3 was filled before the expected completion of A2. A4 was filled after A3 was completed, but within the Persistence Window for Drug A. Therefore, the four prescriptions for Drug A will be consolidated into a single Drug Era (Drug Era1), with the start for prescription A1 recorded as the start date for the consolidated record and the end date for prescription A4 recorded as the end date.

As the Persistence Window was exceeded between filling the two prescriptions for Drug B, they are defined as two distinct Drug Eras. The start and end dates for Drug Era2 and Drug Era3 are the start and end dates for prescriptions B1 and B2, respectively.

The choice of Persistence Window is up to the user of the Drug Eras. OMOP uses a standard Persistence Window of thirty days. Note, that for eras built using 30 day-Persistence Windows no additional 30 days is being added at the end of the last Drug Exposure.

### 3.3.2 EXAMPLE OF A LOADED TABLE

Consider the following example excerpt from the Drug Exposure table:

| Field | Example 1 | Example 2 | Example 3 | Example 4 | Example 5 | Example 6 |
|---|---|---|---|---|---|---|
| drug_exposure_id | 1001 | 1002 | 1003 | 1004 | 1005 | 1006 |
| person_id | 121107 | 127260 | 127260 | 127260 | 127260 | 127260 |
| drug_concept_id | 1310216 | 1310213 | 1310213 | 1310213 | 1310217 | 1310217 |
| drug_exposure_start_date | 9-May-03 | 30-Apr-03 | 27-Jul-03 | 22-Aug-03 | 7-Sep-03 | 2-Oct-03 |
| drug_exposure_end_date | | 30-Apr-03 | 27-Jul-03 | 22-Aug-03 | | |
| drug_type_concept_id | 38000175 | 38000178 | 38000178 | 38000178 | 38000178 | 38000178 |
| stop_reason | | Regimen Completed | | | | Regimen Completed |
| refills | | | | | 1 | 1 |
| quantity | | | | | 30 | 90 |
| days_supply | | | | | 30 | 90 |
| sig | | | | | | |
| prescribing_provider_id | | | | | | |
| visit_occurrence_id | | | | | | |
| relevant_condition_concept_id | | | | | | |
| drug_source_value | 00179139370 | 83200030200310 | 83200030200310 | 83200030200310 | 83200030200320 | 83200030200320 |

The above example uses the following Drug Concepts from the Vocabulary:

| Concept ID | Concept Description | Concept Level |
|---|---|---|
| 1310213 | Warfarin 2.5 MG Oral Tablet | 1 |
| 1310216 | Warfarin 6 MG Oral Tablet | 1 |
| 1310217 | Warfarin 7.5 MG Oral Tablet | 1 |

The Concept Ancestor hierarchy in the Vocabulary indicates that the above Drug Concepts are children of the following Ingredient Drug Class Concept:

| Concept ID | Code | Description | Level |
|---|---|---|---|
| 1310149 | 11289 | Warfarin | 2 |

Drug Eras with a Persistence Window of 30 days are represented as the following Drug Type Concept:

| Concept ID | Description |
|---|---|
| 38000182 | Drug era - 30 days persistence window |

The Drug Eras constructed from the above data, based on aggregation at the Ingredient level and using a 30-day Persistence Window would be reflected in the Drug Era table as follows:

| Field | Example 1 | Example 2 | Example 3 |
|---|---|---|---|
| drug_era_id | 20001 | 20002 | 20003 |
| person_id | 121107 | 127260 | 127260 |
| drug_concept_id | 1310149 | 1310149 | 1310149 |
| drug_era_start_date | 09-May-03 | 30-Apr-03 | 27-Jul-03 |
| drug_era_end_date | 08-Jun-03 | 30-Apr-03 | 31-Dec-03 |
| drug_type_concept_id | 38000182 | 38000182 | 38000182 |
| drug_exposure_count | 1 | 1 | 4 |

Patient 121107 has a pharmacy claim of Warfarin 6 MG Oral Tablets as the only record. No End Date or other details are given, and therefore a 30 day exposure is assumed ending on the 8-Jun-03.

Patient 127260 has several medical history entries that are aggregated into two eras: the first (Example 2) is inferred from the Example 2 in the hypothetical Drug Exposure table. The end date is 30-Apr-03, and even a 30 day Persistence Window will not reach the subsequent Example 3 with a Drug Exposure on the 27-Jun-03. Examples 3-6, however, are aggregated into a single Era, as all End Dates are either after the subsequent Start Dates (Example 5), or after adding the Persistence Window of 30 days to the record (Examples 3 and 4).

Note that this era ends at the calculated End Date (02-Oct-03 + 90 days supply = 31-Dec-03), and no additional Persistence Window is added to the end, even though Persistence Windows were required for its construction.

### 3.3.3 DRUG EXPOSURE TYPES

Drug Types record from what source the Drug Exposure was captured and what kind of recording was used. It also defines the parameters of Drug Era construction:

- Prescription Written (from Electronic Health Records)
- Medication History (from Electronic Health Records)
- Filled Prescription (from Pharmacy Claims)
- Drug from Procedure Code (from Medical Claims and Electronic Health Records)
- Drug Era using a Persistence Window

Drug Exposure Types are stored as Concepts in the Vocabulary. They have been generated based on Thomson and GE References. Different databases might need additional concepts to reflect information that is not currently captured. In that case, OMOP will be providing new Drug Types upon request.

The following Drug Types are currently part of the Vocabulary:

| Concept ID | Concept Name | Concept Level | Concept Class | Vocabulary ID |
|---|---|---|---|---|
| 38000175 | Prescription dispensed in pharmacy | 1 | Drug Exposure Type | 36 |
| 38000176 | Prescription dispensed through mail order | 1 | Drug Exposure Type | 36 |
| 38000177 | Prescription written | 1 | Drug Exposure Type | 36 |
| 38000178 | Medication list | 1 | Drug Exposure Type | 36 |
| 38000179 | Physician administered drug (identified as procedure) | 1 | Drug Exposure Type | 36 |
| 38000180 | Inpatient administration | 1 | Drug Exposure Type | 36 |
| 38000181 | Drug era - 0 days persistence window | 1 | Drug Exposure Type | 36 |
| 38000182 | Drug era - 30 days persistence window | 1 | Drug Exposure Type | 36 |

## 3.4 Condition Occurrence

Condition Occurrences record individual instances of the conditions suffered by Persons as extracted from source data. Conditions are recorded in various data sources in different forms with varying levels of standardization. For example:

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

Condition Occurrences are analyzed based on standard condition Concepts in the Vocabulary.

| Field | Required | Type Precision | Standard | Description |
|---|---|---|---|---|
| condition_occurrence_id | Yes | integer | | A system-generated 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. |

| Field | Required | Type Precision | Standard | Description |
|---|---|---|---|---|
| condition_concept_id | Yes | integer | SNOMED | A foreign key that refers to a standard condition concept identifier in the vocabulary. |
| 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. This is not typically recorded. |
| condition_type_concept_id | Yes | integer | OMOP | A foreign key to the predefined concept identifier in the vocabulary reflecting the source data from which the condition was recorded, the level of standardization, and the type of occurrence. Conditions are defined as primary or secondary diagnoses, problem lists and person statuses. |
| stop_reason | No | string(20) | | The reason, if available, that the condition was no longer recorded, as indicated in the source data. Valid values include discharged, resolved, etc. |
| associated_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 | string(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 vocabulary 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/disposition codes from EHRs. |

### 3.4.1 BUSINESS RULES

The approach to extraction of Condition Occurrence data is based on the individual data source, but the following guidelines are common to all data sources.

- Source attributes mapped to conditions are checked for standardization. If the source attributes are available as standard diagnosis codes (e.g., ICD-9-CM Diagnosis Codes) or specific discharge status codes, then they are mapped to standard Concepts in the Vocabulary.
- If the source data are not coded to a national or international standard, then a finite listing of attribute values is created and mapped to standard condition Concepts in the Vocabulary.

- 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
    - Problem Concepts from EHRs
- As a minimum, the Person ID, Condition Concept ID, Start Date and Condition Type need to be available for a valid record.
- Special handling is necessary for source data in which Person condition entries are updated by expiration of the current entry and addition of an updated entry. In such cases, only the final version of the record is extracted for inclusion in the CDM.

### 3.4.2 EXAMPLE OF A LOADED TABLE

Consider the following of Conditions recorded for Person 127260:

- 30-May-03: ICD9 787.02 "Nausea", reported over the phone
- 29-Jul-03: ICD9 787.02 "Nausea", reported over the phone
- 23-Aug-03: ICD9 531.01 "Acute gastric ulcer without hemorrhage or perforation without obstruction", visit at the doctor's office

The following Concepts in the Vocabulary correspond to the meanings of the problems captured in the source data and can be obtained by mapping the ICD-9-CM codes using the Source To Concept Map table:

| Concept ID | Concept Description |
|---|---|
| 4197598 | Gastric Ulcer |
| 31967 | Nausea |

Assuming that all records are originated as entries in the EHR Problem list, the following Condition Occurrence Type applies:

| Concept ID | Concept Description |
|---|---|
| 38000245 | EHR problem list entry |

The Condition Occurrence table, loaded with the above data, would appear as follows:

| Field | Example 1 | Example 2 | Example 3 |
|---|---|---|---|
| condition_occurrence_id | 3003 | 3004 | 3005 |
| person_id | 127260 | 127260 | 127260 |
| condition_concept_id | 31967 | 31967 | 4197598 |
| condition_start_date | 30-May-03 | 29-Jul-03 | 23-Aug-03 |
| condition_end_date | | | |
| condition_type_concept_id | 38000245 | 38000245 | 38000245 |
| stop_reason | | | |
| associated_provider_id | integer foreign key | integer foreign key | integer foreign key |
| visit_occurrence_id | | | integer foreign key |
| condition_source_value | 787.02 | 787.02 | 531.01 |

## 3.5 Condition Era

Similar to Drug Eras, Condition Eras are chronological periods of Condition Occurrence. Combining individual Condition Occurrences into a single Condition Era serves two purposes:

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

For example, consider a Person who visits his Primary Care Physician (PCP) and who is diagnosed leading to a referral to a specialist. One week later, the Person visits the specialist, who confirms the PCP's diagnosis and provides the appropriate treatment to resolve the condition. These two independent doctor visits should be aggregated into one Condition Era. Just as with Drug Eras, Persistence Windows can be applied for periods of time between the end date of the last and the start date of the following occurrence. OMOP uses Persistence Windows of 30 days.

This model with a standard length Persistence Window generally fits well for acute conditions, but may be less robust for chronic conditions. For example, chronic conditions that do not require regular follow-up may be broken up into multiple Condition Eras because of the absence of data in long periods between visits that exceed the standard persistence. Because the Persistence Window is small, it is likely that multiple visits will be captured in rapid succession for the same condition; however, it is unlikely that infrequent visits for chronic conditions (e.g. a Person with Rheumatoid Arthritis who visits his rheumatologist every three months) will be captured. However, the small window also reduces the likelihood that independent events will be falsely classified as the same Condition Era.

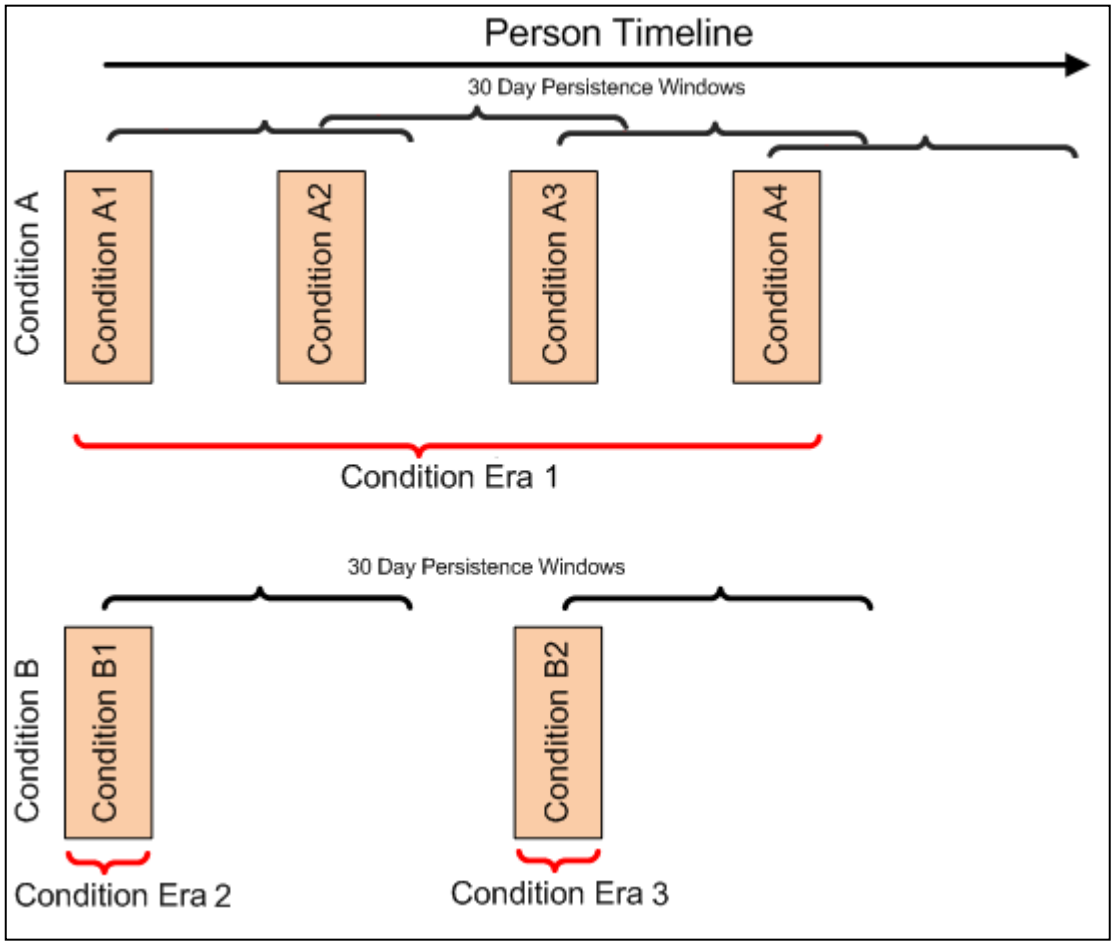
| Field | Required | Type Precision | Standard | Description |
|---|---|---|---|---|
| condition_era_id | Yes | integer | | A system-generated unique identifier for each condition era. |
| person_id | Yes | integer | | A foreign key identifier to the person who is experiencing the condition during the condition era. The demographic details of that person are stored in the person table. |
| condition_concept_id | Yes | integer | SNOMED | A foreign key that refers to a standard condition concept identifier in the vocabulary. |
| condition_era_start_date | Yes | 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 | Yes | 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 | Yes | integer | OMOP | A foreign key to the predefined concept identifier in the vocabulary reflecting the parameters used to construct the condition era. |
| condition_occurrence_count | No | number(4) | | The number of individual condition occurrences used to construct the condition era. |

### 3.5.1 BUSINESS RULES

A Condition Era represents the span of time for which a Person has an episode of care for a given condition. An example is illustrated graphically in Diagram 5: Condition Era Examples. A Person who has been diagnosed with Condition A and Condition B, with Condition A four times (A1, A2, A3, A4), and with Condition B twice (B1, B2).

Diagram 5: Condition Era Examples




To define condition persistence for Condition A, the timing of successive diagnoses is considered. Here, A2 is within the Persistence Window of A1. Similarly, A3 is within the Persistence Window of A2, and A4 is within the Persistence Window of A3. Thus, the four diagnoses of Condition A are consolidated into Condition Era 1, with the start date equal to the diagnosis date for A1, and the end date equal to the diagnosis date for A4.

With Condition B, significant time has elapsed between diagnoses B1 and B2. Therefore, it cannot be assumed that there is dependence between the diagnoses as the time exceeds the Persistence Window for B1. Therefore two distinct Condition Eras are defined, one each that corresponds to B1 and B2.

Note, that for Eras built using 30 day-Persistence Windows no additional 30 days is being added at the end of the last Condition Occurrence. That means, that Condition-free times within an Era is treated as continual Condition, while in the time following the Era no extra Condition is assumed.

### 3.5.2 EXAMPLE OF A LOADED TABLE

Consider the following fictitious example from the Condition Occurrence table.

| Field | Example 1 | Example 2 | Example 3 |
|---|---|---|---|
| condition_occurrence_id | 3003 | 3004 | 3005 |
| person_id | 127260 | 127260 | 127260 |
| condition_concept_id | 4197598 | 4197598 | 4197598 |
| condition_start_date | 30-May-03 | 29-Jul-03 | 23-Aug-03 |
| condition_end_date | 30-May-03 | | |
| condition_type_concept_id | 31967 | 31967 | 31967 |
| stop_reason | Unknown | Resolved | Resolved |
| associated_provider_id | integer foreign key | integer foreign key | integer foreign key |
| visit_occurrence_id | | | |
| condition_source_value | 787.02 | 787.02 | 787.02 |

The above example uses the following Condition Concept from the Vocabulary that is equivalent to ICD9 787.02 "Nausea alone", which can be looked up in the Source To Concept Map table in the Vocabulary.

| Concept ID | Concept Description |
|---|---|
| 31967 | Nausea |

The Condition Occurrence Type Concept in the Vocabulary is determined based on a 30-day Persistence Window, as follows:

| Concept ID | Concept Description |
|---|---|
| 38000247 | Condition era - 30 days persistence window |

The resulting sample representation of the above data in the Condition Era table looks as follows:

| Field | Example 1 | Example 2 |
|---|---|---|
| condition_era_id | 4197598 | 31967 |
| person_id | 127260 | 127260 |
| condition_concept_id | 31967 | 31967 |
| condition_era_start_date | 30-May-03 | 29-Jul-03 |
| condition_era_end_date | 30-May-03 | 23-Aug-03 |
| condition_type_concept_id | 38000247 | 38000247 |
| condition_occurrence_count | 1 | 2 |

The Eras are not aggregated into a higher-level class. Therefore, the Condition Concept ID is unchanged. The first Example of the hypothetical Condition Occurrence gives rise to Era Example 1. The Occurrence Examples 2 and 3 are combined to Era Example 2, as the Persistence Window makes them overlap (29-Jul-03 + 30 days > 23-Aug-03). Note that the Condition End Date is assumed identical to the Condition Start Date for conditions, which is different for some Drug Occurrence types.

### 3.5.3 CONDITION OCCURRENCE TYPES

Condition Occurrence Types record from what source the Condition Occurrence was captured, whether the condition (diagnosis) was primary or secondary and the relative positioning within a Person's condition record. It also defines the parameters of Condition Era construction.

Condition Occurrence Types are stored as Concepts in the Vocabulary. They have been generated based on Thomson and GE References. Different databases might need additional concepts to reflect information that is not currently captured. In that case, OMOP will be providing new Condition Occurrence Types upon request.

The following Condition Occurrence Types are currently part of the Vocabulary:

| Concept ID | Concept Name | Concept Level | Concept Class | Vocabulary ID |
|---|---|---|---|---|
| 38000183 | Inpatient detail – primary | 1 | Condition Occurrence Type | 37 |
| 38000184 | Inpatient detail - 1st position | 1 | Condition Occurrence Type | 37 |
| 38000185 | Inpatient detail - 2nd position | 1 | Condition Occurrence Type | 37 |
| 38000186 | Inpatient detail - 3rd position | 1 | Condition Occurrence Type | 37 |
| 38000187 | Inpatient detail - 4th position | 1 | Condition Occurrence Type | 37 |
| 38000188 | Inpatient detail - 5th position | 1 | Condition Occurrence Type | 37 |
| 38000189 | Inpatient detail - 6th position | 1 | Condition Occurrence Type | 37 |
| 38000190 | Inpatient detail - 7th position | 1 | Condition Occurrence Type | 37 |
| 38000191 | Inpatient detail - 8th position | 1 | Condition Occurrence Type | 37 |
| 38000192 | Inpatient detail - 9th position | 1 | Condition Occurrence Type | 37 |
| 38000193 | Inpatient detail - 10th position | 1 | Condition Occurrence Type | 37 |
| 38000194 | Inpatient detail - 11th position | 1 | Condition Occurrence Type | 37 |
| 38000195 | Inpatient detail - 12th position | 1 | Condition Occurrence Type | 37 |
| 38000196 | Inpatient detail - 13th position | 1 | Condition Occurrence Type | 37 |
| 38000197 | Inpatient detail - 14th position | 1 | Condition Occurrence Type | 37 |
| 38000198 | Inpatient detail - 15th position | 1 | Condition Occurrence Type | 37 |
| 38000199 | Inpatient header - primary | 1 | Condition Occurrence Type | 37 |
| 38000200 | Inpatient header – 1st position | 1 | Condition Occurrence Type | 37 |
| 38000201 | Inpatient header – 2nd position | 1 | Condition Occurrence Type | 37 |
| 38000202 | Inpatient header – 3rd position | 1 | Condition Occurrence Type | 37 |
| 38000203 | Inpatient header - 4th position | 1 | Condition Occurrence Type | 37 |
| 38000204 | Inpatient header - 5th position | 1 | Condition Occurrence Type | 37 |
| 38000205 | Inpatient header - 6th position | 1 | Condition Occurrence Type | 37 |
| 38000206 | Inpatient header - 7th position | 1 | Condition Occurrence Type | 37 |
| 38000207 | Inpatient header - 8th position | 1 | Condition Occurrence Type | 37 |
| 38000208 | Inpatient header - 9th position | 1 | Condition Occurrence Type | 37 |
| 38000209 | Inpatient header - 10th position | 1 | Condition Occurrence Type | 37 |
| 38000210 | Inpatient header - 11th position | 1 | Condition Occurrence Type | 37 |
| 38000211 | Inpatient header - 12th position | 1 | Condition Occurrence Type | 37 |
| 38000212 | Inpatient header - 13th position | 1 | Condition Occurrence Type | 37 |
| 38000213 | Inpatient header - 14th position | 1 | Condition Occurrence Type | 37 |
| 38000214 | Inpatient header - 15th position | 1 | Condition Occurrence Type | 37 |
| 38000215 | Outpatient detail - 1st position | 1 | Condition Occurrence Type | 37 |
| 38000216 | Outpatient detail - 2nd position | 1 | Condition Occurrence Type | 37 |
| 38000217 | Outpatient detail - 3rd position | 1 | Condition Occurrence Type | 37 |

| Concept ID | Concept Name | Concept Level | Concept Class | Vocabulary ID |
|---|---|---|---|---|
| 38000218 | Outpatient detail - 4th position | 1 | Condition Occurrence Type | 37 |
| 38000219 | Outpatient detail - 5th position | 1 | Condition Occurrence Type | 37 |
| 38000220 | Outpatient detail - 6th position | 1 | Condition Occurrence Type | 37 |
| 38000221 | Outpatient detail - 7th position | 1 | Condition Occurrence Type | 37 |
| 38000222 | Outpatient detail - 8th position | 1 | Condition Occurrence Type | 37 |
| 38000223 | Outpatient detail - 9th position | 1 | Condition Occurrence Type | 37 |
| 38000224 | Outpatient detail - 10th position | 1 | Condition Occurrence Type | 37 |
| 38000225 | Outpatient detail - 11th position | 1 | Condition Occurrence Type | 37 |
| 38000226 | Outpatient detail - 12th position | 1 | Condition Occurrence Type | 37 |
| 38000227 | Outpatient detail - 13th position | 1 | Condition Occurrence Type | 37 |
| 38000228 | Outpatient detail - 14th position | 1 | Condition Occurrence Type | 37 |
| 38000229 | Outpatient detail - 15th position | 1 | Condition Occurrence Type | 37 |
| 38000230 | Outpatient header - 1st position | 1 | Condition Occurrence Type | 37 |
| 38000231 | Outpatient header - 2nd position | 1 | Condition Occurrence Type | 37 |
| 38000232 | Outpatient header - 3rd position | 1 | Condition Occurrence Type | 37 |
| 38000233 | Outpatient header - 4th position | 1 | Condition Occurrence Type | 37 |
| 38000234 | Outpatient header - 5th position | 1 | Condition Occurrence Type | 37 |
| 38000235 | Outpatient header - 6th position | 1 | Condition Occurrence Type | 37 |
| 38000236 | Outpatient header - 7th position | 1 | Condition Occurrence Type | 37 |
| 38000237 | Outpatient header - 8th position | 1 | Condition Occurrence Type | 37 |
| 38000238 | Outpatient header - 9th position | 1 | Condition Occurrence Type | 37 |
| 38000239 | Outpatient header - 10th position | 1 | Condition Occurrence Type | 37 |
| 38000240 | Outpatient header - 11th position | 1 | Condition Occurrence Type | 37 |
| 38000241 | Outpatient header - 12th position | 1 | Condition Occurrence Type | 37 |
| 38000242 | Outpatient header - 13th position | 1 | Condition Occurrence Type | 37 |
| 38000243 | Outpatient header - 14th position | 1 | Condition Occurrence Type | 37 |
| 38000244 | Outpatient header - 15th position | 1 | Condition Occurrence Type | 37 |
| 38000245 | EHR problem list | 1 | Condition Occurrence Type | 37 |
| 38000246 | Condition era - 0 days persistence window | 1 | Condition Occurrence Type | 37 |
| 38000247 | Condition era - 30 days persistence window | 1 | Condition Occurrence Type | 37 |

## 3.6 Visit Occurrence

The Visit Occurrence table contains all Person visits to health care providers, including inpatient, outpatient, and ER visits. A Visit is an encounter for a patient at a point of care for a duration of time. There could be several Providers involved in the patient's care during the Visit. Visits are recorded in various data sources in different forms with varying levels of standardization. For example:

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

| Field | Required | Type Precision | Standard | Description |
|---|---|---|---|---|
| visit_occurrence_id | Yes | integer | | A system-generated 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_start_date | Yes | date | | The start date of the visit. |
| 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. |
| place_of_service_concept_id | Yes | integer | OMOP CMS | A foreign key that refers to a place of service concept identifier in the vocabulary. |
| care_site_id | No | integer | | A foreign key to the care site in the care site table that was visited. |
| place_of_service_source_value | No | string(50) | | The source code used to reflect the type or source of the visit in the source data. Valid entries include office visits, hospital admissions, etc. These source codes can also be type-of service codes and activity type codes. |

### 3.6.1 BUSINESS RULES

A Visit Occurrence is recorded for each visit to a healthcare facility. Each visit is standardized by assigning a corresponding Concept Identifier based on the type of facility visited and the type of services rendered. As a minimum, the Person ID, Place Of Service Concept ID, Start and End Date need to be available for a valid record.

### 3.6.2 EXAMPLE OF A LOADED TABLE

Consider the following example visit data for Person 127260:

- 03-May-03: Hospital admission to the MGH Hemodialysis Unit, discharged 04-May-03 (Place of Service code 21)
- 29-Jul-03: Outpatient Dialysis visit to the office of the Massachusetts General Renal Associates (Place of Service code 22)

The following Concepts correspond to the meanings of the types of visits that were indicated in the source data.

| Concept ID | Concept Description |
|---|---|
| 8715 | Hospital Admission |
| 8614 | Outpatient Visit |

The data above is represented in the Visit Occurrence table as follows.

| Field | Example 1 | Example 2 |
|---|---|---|
| visit_occurrence_id | 5003 | 5004 |
| person_id | 127260 | 127260 |
| visit_start_date | 03-May-03 | 29-Jul-03 |
| visit_end_date | 04-May-03 | 29-Jul-03 |
| place_of_service_concept_id | 8715 | 8614 |
| care_site_id | integer foreign key | |
| place_of_service_source_value | 21 | 22 |

## 3.7 Procedure Occurrence

Procedure occurrences record individual instances of procedures performed on Persons extracted from the source data. Procedures are present in various data sources in different forms with varying levels of standardization. For example:

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

| Field | Required | Type Precision | Standard | 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 | CPT-4 HCPCS ICD-9-Proc, ICD-9-CM, LOINC | A foreign key that refers to a standard procedure concept identifier in the vocabulary. |
| procedure_date | Yes | date | | The date on which the procedure was performed. |
| procedure_type_concept_id | Yes | integer | OMOP | A foreign key to the predefined concept identifier in the vocabulary reflecting the type of the procedure. |
| associated_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. |
| relevant_condition_concept_id | No | integer | SNOMED | A foreign key to the predefined concept identifier in the vocabulary reflecting the condition that was the cause for initiation of the procedure. Note that this is not a direct reference to a specific condition record in the condition table, but rather a condition concept in the vocabulary. |

| Field | Required | Type Precision | Standard | Description |
|---|---|---|---|---|
| procedure_source_value | No | string(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 vocabulary and the original code is, stored here for reference. Procedure source codes are typically ICD-9-Proc, CPT-4 or HCPCS codes. |

### 3.7.1 BUSINESS RULES

Procedure Occurrences are recorded for each procedure performed on a Person. Each procedure is standardized by assigning a Concept code corresponding to the definition of the procedure code and code type used.

If possible, the visit in which the procedure was performed is recorded through a reference to the visit table.

The Relevant Condition Concept is defined as the condition that is associated with the Procedure. This can be the indication, or the condition to be diagnosed or ruled out. This information is not typically available.

As a minimum, the Person ID, the Procedure Concept ID, the Procedure Type Concept ID and the Date need to be available for a valid record.

### 3.7.2 EXAMPLE OF A LOADED TABLE

Consider the following hypothetical procedures carried out for Person 127260:

- 03-May-03: CPT 71020 "Chest X-Ray", to diagnose or rule out a pneumonia
- 29-Jul-03: CPT 93925 "Lower Extremity Arterial Duplex, Bilateral" to rule out a peripheral arterial occlusive disease of the legs
- 15-Sep-03: LOINC 11118-7 "Plasma cells/100 cells in Bone marrow by Microscopy", recorded as a test result
- 15-Dec-04: ICD-9-CM V42.82 "Peripheral stem cells replaced by transplant", recorded as a diagnosis.

The following Concepts correspond to the definition of the procedures that were captured in the source data and mapped to Standard Vocabulary Concepts:

| Concept ID | Concept Description |
|---|---|
| 2211361 | Chest X-Ray |
| 2313985 | Lower Extremity Arterial Duplex Bilateral |
| 2008321 | Bone marrow aspiration procedure |
| 4242257 | Allogeneic bone marrow transplantation |

The corresponding Procedure Occurrence Type Concepts as extracted from the EHR Order list are as follows:

| Concept ID | Concept Description |
|---|---|
| 38000275 | EHR order list entry |
| 38003621 | Procedure recorded as lab test |
| 38003622 | Procedure recorded as diagnostic code |

The Condition Concepts that represent the indications for the procedures are:

| Concept ID | Concept Description |
|---|---|
| 255848 | Pneumonia |
| 434961 | Occlusion of lower limb artery |
| 437233 | Multiple myeloma |

The above data is represented in the Procedure Occurrence table as follows:

| Field | Example 1 | Example 2 | Example 3 | Example 4 |
|---|---|---|---|---|
| procedure_occurrence_id | 5003 | 5004 | 10545 | 10553 |
| person_id | 127260 | 127260 | 127260 | 127260 |
| procedure_concept_id | 2211361 | 2313985 | 2008321 | 4242257 |
| procedure_date | 03-May-03 | 29-Jul-03 | 15-Sep-03 | 15-Dec-04 |
| procedure_type_concept_id | 38000275 | 38000275 | 38003621 | 38003622 |
| associated_provider_id | integer foreign key | integer foreign key | integer foreign key | integer foreign key |
| visit_occurrence_id | integer foreign key | integer foreign key | integer foreign key | integer foreign key |
| relevant_condition_concept_id | 255848 | 434961 | 437233 | 437233 |
| procedure_source_value | 71020 | 93925 | 11118-7 | V42.82 |

Note that the Relevant Condition Concept ID is not a foreign key to an actual Condition Occurrence record, but to a Condition Concept in the Vocabulary. It remains up to the ETL to ensure there is such a record in the table. This is in contrast to the Provider or Visit Occurrence ID, which link directly to the corresponding Visit Occurrence where the Procedure was carried out.

### 3.7.3 PROCEDURE TYPES

The Procedure Types define from which the Procedure Occurrence is drawn or inferred, and indicates whether a Procedure was primary or secondary and their relative positioning within a Procedure record.

Procedure Occurrence Types have been populated based on Thomson and GE References. This table can be amended with records to represent proprietary data upon request. Different databases might need additional concepts to reflect information that is not currently captured.

The following Procedure Occurrence Types are currently listed in the Vocabulary:

| Concept ID | Concept Name | Concept Level | Concept Class | Vocabulary ID |
|---|---|---|---|---|
| 38000248 | Inpatient detail - primary position | 1 | Procedure Occurrence Type | 38 |
| 38000249 | Inpatient detail - 1st position | 1 | Procedure Occurrence Type | 38 |
| 38000250 | Inpatient header - primary position | 1 | Procedure Occurrence Type | 38 |
| 38000251 | Inpatient header - 1st position | 1 | Procedure Occurrence Type | 38 |
| 38000252 | Inpatient header - 2nd position | 1 | Procedure Occurrence Type | 38 |
| 38000253 | Inpatient header - 3rd position | 1 | Procedure Occurrence Type | 38 |
| 38000254 | Inpatient header - 4th position | 1 | Procedure Occurrence Type | 38 |
| 38000255 | Inpatient header - 5th position | 1 | Procedure Occurrence Type | 38 |
| 38000256 | Inpatient header - 6th position | 1 | Procedure Occurrence Type | 38 |
| 38000257 | Inpatient header - 7th position | 1 | Procedure Occurrence Type | 38 |
| 38000258 | Inpatient header - 8th position | 1 | Procedure Occurrence Type | 38 |

| Concept ID | Concept Name | Concept Level | Concept Class | Vocabulary ID |
|---|---|---|---|---|
| 38000259 | Inpatient header - 9th position | 1 | Procedure Occurrence Type | 38 |
| 38000260 | Inpatient header - 10th position | 1 | Procedure Occurrence Type | 38 |
| 38000261 | Inpatient header - 11th position | 1 | Procedure Occurrence Type | 38 |
| 38000262 | Inpatient header - 12th position | 1 | Procedure Occurrence Type | 38 |
| 38000263 | Inpatient header - 13th position | 1 | Procedure Occurrence Type | 38 |
| 38000264 | Inpatient header - 14th position | 1 | Procedure Occurrence Type | 38 |
| 38000265 | Inpatient header - 15th position | 1 | Procedure Occurrence Type | 38 |
| 38000266 | Outpatient detail - primary position | 1 | Procedure Occurrence Type | 38 |
| 38000267 | Outpatient detail - 1st position | 1 | Procedure Occurrence Type | 38 |
| 38000268 | Outpatient header - primary position | 1 | Procedure Occurrence Type | 38 |
| 38000269 | Outpatient header - 1st position | 1 | Procedure Occurrence Type | 38 |
| 38000270 | Outpatient header - 2nd position | 1 | Procedure Occurrence Type | 38 |
| 38000271 | Outpatient header - 3rd position | 1 | Procedure Occurrence Type | 38 |
| 38000272 | Outpatient header - 4th position | 1 | Procedure Occurrence Type | 38 |
| 38000273 | Outpatient header - 5th position | 1 | Procedure Occurrence Type | 38 |
| 38000274 | Outpatient header - 6th position | 1 | Procedure Occurrence Type | 38 |
| 38000275 | EHR order list entry | 1 | Procedure Occurrence Type | 38 |
| 38003621 | Procedure recorded as lab test | 1 | Procedure Occurrence Type | 38 |
| 38003622 | Procedure recorded as diagnostic code | 1 | Procedure Occurrence Type | 38 |

## 3.8 Observation

The Observation table contains all general observations from the following categories:

- Lab observations (i.e., test results) from Medical Claims
- Lab and other observations from Electronic Health Records
- Chief complaints as captured in Electronic Health Records
- General clinical findings, signs and symptoms
- Radiology and pathology reports
- General catch-all categories from various data sources that cannot be otherwise categorized within the entities provided (Drug, Condition, Procedure)

| Field | Required | Type Precision | Standard | Description |
|---|---|---|---|---|
| observation_id | Yes | integer | | A system-generated 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 | LOINC SNOMED | A foreign key to the standard observation concept identifier in the vocabulary. |
| observation_date | Yes | date | | The date of the observation. |
| observation_time | No | time | | The time of the observation. |

| Field | Required | Type Precision | Standard | Description |
|---|---|---|---|---|
| value_as_number | No | number(14,3) | | 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 | string(60) | | The observation result stored as a string. This is applicable to observations where the result is expressed as verbatim text, such as in radiology or pathology reports. |
| 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 vocabulary (e.g., positive/negative, present/absent, low/high, etc.). |
| unit_concept_id | No | integer | UCUM | A foreign key to a standard concept identifier of measurement units in the vocabulary. |
| range_low | No | number(14,3) | | The lower limit of the normal range of the observation. It is not applicable if the observation results are non-numeric or categorical, and must be in the same units of measure as the observation value. |
| range_high | No | number(14,3) | | The upper limit of the normal range of the observation. It is not applicable if the observation results are non-numeric or categorical, and must be in the same units of measure as the observation value. |
| observation_type_concept_id | Yes | integer | OMOP | A foreign key to the predefined concept identifier in the vocabulary reflecting the type of the observation. |
| associated_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. |
| relevant_condition_concept_id | No | integer | SNOMED | A foreign key to the predefined concept identifier in the vocabulary reflecting the condition that was associated with the observation. Note that this is not a direct reference to a specific condition record in the condition table, but rather a condition concept in the vocabulary. |
| observation_source_value | No | string(50) | | The observation code as it appears in the source data. This code is mapped to a standard concept in the vocabulary and the original code is, stored here for reference. |

| Field | Required | Type Precision | Standard | Description |
|---|---|---|---|---|
| unit_source_value | No | string(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 vocabulary and the original code is, stored here for reference. |

### 3.8.1 BUSINESS RULES

The approach to extraction and representation of Observation data are based on the individual data source, but the following guidelines are common to all data sources.

- Source attribute values mapped to Observations are checked for standardization. If the source attribute values are available as national or international standard codes (e.g. LOINC codes) then they are mapped to standard Concept Identifiers in the Vocabulary.
- If the source data are not coded to a national or international standard then a finite listing of attribute values is created and mapped to standard Observation Concepts in the Vocabulary.
- The type of result recorded for the Observation is important for further processing of the Observation data. Knowledge of whether an Observation result is captured as a numeric value (with the range of values considered normal), standard Concept code or non-standard text will determine the handling of the Observation data.
- An Observation Type is assigned based on the type of source data from which the Observation is extracted and type of result expected. Observation Types are standard Concepts in the Vocabulary.
- Each Observation for every Person, along with its matching standard Concept Identifier from the Vocabulary, is extracted from the source data along with the Person identifier. Related attributes including date of the Observation, type of observation, type of result, result as a number/text/Concept Identifier, and reference range for numeric results are also extracted.
- If possible, the visit in which the observation was made is recorded through a reference to the visit table.
- The Relevant Condition Concept is defined as the condition that is associated with the Observation. This can be the underlying condition, or the condition to be diagnosed or ruled out. This information is not typically available.
- As a minimum, the Person ID, Observation Concept ID, Observation Date and Type need to be available for a valid record.

### 3.8.2 EXAMPLE OF A LOADED TABLE

Consider the following example Observation data extracted from Lab Claim Supplement data and Observations for Person 127260:

- 03-MAY-2003: LDL cholesterol serum test
- 29-JUL-2003: White blood count, to diagnose a bacterial infection of the upper respiratory tract
- 23-AUG-2003: Assessing smoking status

The following Concept codes correspond to the observations that were captured in the source data.

| Concept ID | LOINC Code | Concept Description |
|---|---|---|
| 3028288 | 13457-7 | Lipid Panel – LDL Check |
| 3000905 | 6690-2 | White Blood Count Check |
| 4275495 | 4766 | Smoking Status |

The following Concepts correspond to the meanings of the units of measure associated with the lab observations in the source data:

| Concept ID | Concept Code | Concept Description |
|---|---|---|
| 8840 | mg/dL | mg/dL |
| 8784 | {cells}/uL | Cells per microliter |

The Concept Types for these Observations are:

| Concept ID | Concept Description |
|---|---|
| 38000277 | Observation numeric result |
| 38000278 | Observation text |

Upper respiratory tract infection is represented by the following concept:

| Concept ID | Concept Description |
|---|---|
| 4207185 | Bacterial upper respiratory infection |

The above data is represented in the Observation table as follows:

| Field | Example 1 | Example 2 | Example 3 |
|---|---|---|---|
| observation_id | 3462346 | 3462347 | 3462348 |
| person_id | 127260 | 127260 | 127260 |
| observation_concept_id | 3028288 | 3000905 | 4275495 |
| observation_date | 03-May-03 | 29-Jul-03 | 23-Aug-03 |
| observation_time | | | |
| value_as_number | 124 | 6000 | |
| value_as_string | | | PASSIVE SMOKER |
| value_as_concept_id | | | |
| unit_concept_id | 8840 | 8784 | |
| range_low | 0 | 4500 | |
| range_high | 130 | 10000 | |
| observation_type_concept_id | 38000277 | 38000277 | 38000278 |
| associated_provider_id | integer foreign key | integer foreign key | integer foreign key |
| visit_occurrence_id | integer foreign key | integer foreign key | integer foreign key |
| relevant_condition_concept_id | | 4207185 | |
| observation_source_value | 13457-7 | 6690-2 | 4766 |
| unit_source_value | mg/dL | {cells}/uL | |

Note that the Relevant Condition Concept ID is not a foreign key to an actual Condition Occurrence record, but to a Condition Concept in the Vocabulary. It remains up to the ETL to ensure there is such a record in the table. This is in contrast to the Visit Occurrence ID, which links directly to the corresponding Visit Occurrence where the Observation was made.

### 3.8.3 OBSERVATION TYPES

Assignment of an Observation Type is essential to determine the nature of the source data, the level of standardization and coding, as well as the type of result recorded for the Observation. The Observation Types can include the following:

- Chief Complaint
- Observation recorded from Electronic Health Records
- Lab Result
- Problem List from Electronic Health Records
- Observation recorded from Electronic Health Records with text results

Observation Types are Concepts in the Vocabulary and have been populated based on Thomson and GE References. Different databases might need additional concepts to reflect information that is not currently captured, and can be added by OMOP upon request.

## 3.9 Observation Period

The Observation Period table is designed capture the time intervals in which data are being recorded for the Person. An Observation Period is the span of time when a Person is expected to have the potential of Drug and Condition information recorded. For claims data, observation periods are equivalent to enrollment periods to a plan.

Analytical methods use Observation Period records to distinguish periods with no observed records from periods where data are systematically not captured, such as a person not having insurance coverage..

| Field | Required | Type Precision | Standard | Description |
|---|---|---|---|---|
| observation_period_id | Yes | integer | | A system-generated 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. |

### 3.9.1 BUSINESS RULES

Tracking the Observation Period of a Person requires unique handling for each raw data source from which Person data are extracted:

- Only active periods should be recorded into Observation Periods. Records indicating Persons as inactive or deceased must not contribute to Observation Periods.

- For data sources in which the status of a Person for each calendar month or year is recorded as a separate entry, even if there are no changes, a single consolidated Person Status entry is recorded in the CDM. A Persistence Window can be applied if it is known that the source plan allows enrollment windows for their clients.

### 3.9.2 EXAMPLE OF A LOADED TABLE

Consider the following example data extracted from a claims data source for Person 127260:

- Enrollment from 01-Jan-03 through 30-Sep-03, medical and drug benefit included
- Enrollment from 15-Oct-03 through 01-Jan-04, only medical benefit included
- Enrollment from 15-Apr-04 through 31-Dec-04, medical and drug benefit included

The following Concept codes correspond to the meanings of the Person Status values that were present in the source data.

The above data are consolidated and represented in the OBSERVATION_PERIOD table as follows:

| Field | Example 1 | Example 2 |
|---|---|---|
| observation_period_id | 80001 | 80003 |
| person_id | 127260 | 127260 |
| observation_period_start_date | 01-Jan-03 | 15-Apr-04 |
| observation_period_end_date | 01-Jan-04 | 31-Dec-04 |

For the first record, the lag time between the enrollment ending in September and the new period starting in October is covered by something like a Persistence Window. It is up to the ETL design to decide, how long such a re-enrollment period should last. Generally, a 30 day period is a plausible size.

Note that in this CDM version medical and drug benefit are not recorded in the Observation Period any longer. If such detailed Plan Design information is available it will be recorded in the Payer Plan Family table.

## 3.10 Death

The Death table is designed to capture the time when a Person is deceased and causes of death. Depending on the source, this information can be derived from a variety of information:

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

Note: Data sources might contain multiple records of death at different dates. It is the task of the ETL to pick the most plausible or most accurate records to be stored to this table.

| Field | Required | Type Precision | Standard | 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 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. |

| Field | Required | Type Precision | Standard | Description |
|---|---|---|---|---|
| death_type_concept_id | Yes | integer | OMOP | A foreign key referring to the predefined concept identifier in the vocabulary reflecting how the death was represented in the source data. |
| cause_of_death_concept_id | No | integer | SNOMED | A foreign key referring to a standard concept identifier in the vocabulary for conditions. |
| cause_of_death_source_value | No | string(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 vocabulary and the original code is, stored here for reference. |

### 3.10.1 BUSINESS RULES

Each Person may have more than one record in the Death table if there is available information for multiple contributing causes of death. If the Death Date cannot be precisely determined from the data, the best approximation should be used. The Cause of Death Concept ID is a reference to a Condition Concept ID in the Vocabulary that is recorded in the primary cause of death on the death certificates. At a minimum, the Person ID, Death Date and Death Type are required for a valid record.

### 3.10.2 EXAMPLE OF A LOADED TABLE

Death can be recorded from a variety of sources, which have a different level of precision. Death Certificates in the National Death Index are, but Status codes in the enrollment file of claims data may not be. It is the task of the ETL to make the best call. Consider the following cases:

- Person 127260: The enrollment status in the claims file has the status "Deceased" for an enrollment period starting 1/1/05 and ending 3/31/05
- Person 127261: A medical claim discharge status "Died" dated 2/1/05
- Person 127262: A medical claim contains diagnostic code 761.6 "Maternal death affecting fetus or newborn", dated 3/1/05
- Person 127263: A medical claim contains DRG code 385 "Neonates, Died or Transferred To Another Acute Care Facility", dated 4/1/05
- Person 127264: The EHR record contains patient status "Deceased", dated May 2005
- National Death Index record contains a record that can be linked to Person 127265, with the cause of death 410 "Myocardial Infarction", dated 6/1/05

The following Concept IDs correspond to the above data:

| Concept ID | Concept Description | Concept Class |
|---|---|---|
| 38003565 | Payer enrollment status "Deceased" | Death Type |
| 38003566 | Medical claim discharge status "Died" | Death Type |
| 38003567 | Medical claim diagnostic code indicating death | Death Type |
| 38003568 | Medical claim DRG code indicating death | Death Type |
| 38003569 | EHR record patient status "Deceased" | Death Type |
| 38003570 | Death Certificate immediate cause | Death Type |

The corresponding Death records might look like in the following table. Note that the date for Example 1 the date was "rounded" down to the beginning of the period, and in Example 6 to the first of the month.

| Field | Example 1 | Example 2 | Example 3 | Example 4 | Example 5 | Example 6 |
|---|---|---|---|---|---|---|
| person_id | 127260 | 127261 | 127262 | 127263 | 127264 | 127265 |
| death_date | 1-Jan-2005 | 1-Feb-2005 | 1-Mar-2005 | 1-Apr-2005 | 1-May-2005 | 1-Jun-2005 |
| death_type_concept_id | 38003565 | 38003566 | 38003567 | 38003568 | 38003569 | 38003570 |
| cause_of_death_concept_id | | | 4192513 | | | 448759 |
| cause_of_death_source_value | | | 761.6 | | | 410 |

### 3.10.3 DEATH TYPES

Assignment of a Death Type is essential to determine the nature of the source data, the level of standardization and coding, as well as the type of result recorded for the Death record. The Death Types include the following:

- Claims: Enrollment status "Deceased"
- Medical claims: Discharge status "Died"
- Medical claims: Condition in one of the various diagnose fields containing codes indicating death
- Medical claims: DRG code indicating death
- EHR records: Patient status "Deceased"
- Standard Death Certificate Record: A record that can be linked to a Person in the Person table, with a single primary or several causes of death. On the Death Certificates, immediate cause, contributory cause, underlying cause can be separately indicated.

Death Types are Concepts in the Vocabulary and have been populated based on the current experience with a variety of data sources. Different databases might need additional concepts to reflect information that is not currently captured, and can be added by OMOP upon request.

## 3.11 Drug Cost

The Drug Cost table captures the cost of a Drug Exposure. The information about the cost is defined in the following components:

- The various amounts of money paid for the Drug
- The various costs of the Drug

In addition, a reference to the health plan information in the Payer Plan Period table is stored in the record that is responsible for the determination of the cost as well as some of the payments.

| Field | Required | Type Precision | Standard | Description |
|---|---|---|---|---|
| drug_cost_id | Yes | integer | | A system-generated unique identifier for each drug cost record. |
| drug_exposure_id | Yes | integer | | A foreign key identifier to the drug record for which cost data are recorded. |
| paid_copay | No | number(8,2) | | The amount paid by the person as a fixed contribution to the expenses. Copay does not contribute to the out of pocket expenses. |
| paid_coinsurance | No | number(8,2) | | The amount paid by the person as a joint assumption of risk. Typically, this is a percentage of the expenses defined by the payer plan (policy) after the person's deductible is exceeded. |
| paid_toward_deductible | No | number(8,2) | | The amount paid by the person that is counted toward the deductible defined by the payer plan (policy). |
| paid_by_payer | No | number(8,2) | | The amount paid by the payer (insurer). If there is more than one payer, several drug_cost records indicate that fact. |
| paid_by_coordination_benefits | No | number(8,2) | | The amount paid by a secondary payer through the coordination of benefits. |
| total_out_of_pocket | No | number(8,2) | | The total amount paid by the person as a share of the expenses, excluding the copay. |
| total_paid | No | number(8,2) | | The total amount paid for the expenses of drug exposure. |
| ingredient_cost | No | number(8,2) | | The portion of the drug expenses due to the cost charged by the manufacturer for the drug, typically a percentage of the Average Wholesale Price. |
| dispensing_fee | No | number(8,2) | | The portion of the drug expenses due to the dispensing fee charged by the pharmacy, typically a fixed amount. |
| average_wholesale_price | No | number(8,2) | | List price of a drug set by the manufacturer. |
| payer_plan_period_id | No | integer | | A foreign key to the payer_plan_period table, where the details of the payer, plan and family are stored. |

### 3.11.1 BUSINESS RULES

Each Drug Exposure may have any number of corresponding records in the Drug Cost table, but typically it is none (no cost data recorded) or one. They are linked directly through the Drug Exposure ID field.

The amounts paid are:

- Copay – a fixed amount to be paid by the Person
- Coinsurance – a relative amount of the total paid by the Person
- Deductible – an amount of money paid by the Person before the Payer starts contributing
- Primary Payer – the amount the primary Payer pays towards the total
- Coordination of Benefits – the amount a secondary Payer or Family Plan pays towards the total
- Out of Pocket = Copay + Coinsurance + Deductible
- Total – the total amount paid for the Drug Exposure

The costs are:

- Ingredient Cost – the amount charged by the wholesale distributor or manufacturer
- Dispensing Fee – the amount charged by the pharmacy
- Sales Tax. This is usually very small and typically not provided by most source data, and therefore not included in the CDM

The amount paid should equal the cost, so Copay + Coinsurance + Deductible + Primary Payer + Coordination Of Benefits = Total Paid = Ingredient Cost + Dispensing Fee. In reality, this is not always reflected in the source data. It is up to the ETL to determine how to deal with quality problems in the data.

The Average Wholesale Price is the list price of the drug, but not the price actually charged or paid.

Finally, the health plan of the Person that is determined by these numbers is referred to through the Payer Plan Period ID (see below).

### 3.11.2 EXAMPLE OF A LOADED TABLE

The select three example drug claims are represented as following:

- Example 1: 90 oral tablets Labetalol 300 mg, copay $4.98, coinsurance $5.02, no deductible, insurance $40.15, no cob, ingredient cost $48.25, dispensing fee $1.90, AWP $85.40 (patient pays $10 out of pocket of the $50.15 total price which is discounted from the AWP), payer, plan and family id recorded in Payer Plan Period.
- Example 2: 30 oral tablets Furosemide 20 mg, paid toward deductible $5.99, no cob, ingredient cost $3.84, dispensing fee $2.15, AWP $4.53 (since deductible is not met, patient pays full price that is higher than AWP)
- Example 3: 30 oral tablets Zoloft (Sertraline) 100 mg, insurance $ 21.51, cob $ 48.89 (patient's insurance and cob share the total price of $70.40).

| Field | Example 1 | Example 2 | Example 3 |
|---|---|---|---|
| drug_cost_id | 349934952 | 349934955 | 349934503 |
| drug_exposure_id | integer foreign key | integer foreign key | integer foreign key |
| paid_copay | 4.98 | 0 | 0 |
| paid_coinsurance | 5.02 | 0 | 0 |
| paid_toward_deductible | 0 | 5.99 | 0 |
| paid_by_payer | 40.15 | 0 | 21.51 |
| paid_by_coordination_benefits | 0 | 0 | 48.89 |
| total_out_of_pocket | 10 | 5.99 | 0 |

| Field | Example 1 | Example 2 | Example 3 |
|---|---|---|---|
| total_paid | 50.15 | 5.99 | 70.4 |
| ingredient_cost | 48.25 | 3.84 | |
| dispensing_fee | 1.9 | 2.15 | |
| average_wholesale_price | 85.4 | 4.53 | |
| payer_plan_period_id | integer foreign key | | |

## 3.12 Procedure Cost

The Procedure Cost table captures the cost of a Procedure performed on a Person. The information about the cost is only derived from the amounts paid for the Procedure. This is in contrast to the Drug Cost data which also contain information about the cost.

In addition, Disease Reference Groups, Ambulatory Payment Classifications and Revenue codes are captured. Finally, a reference to the health plan information in the Payer Plan Period table is stored in the record.

| Field | Required | Type Precision | Standard | Description |
|---|---|---|---|---|
| procedure_cost_id | Yes | integer | | A system-generated unique identifier for each procedure cost record. |
| procedure_occurrence_id | Yes | integer | | A foreign key identifier to the procedure record for which cost data are recorded. |
| paid_copay | No | number(8,2) | | The amount paid by the person as a fixed contribution to the expenses. Copay does not contribute to the out_of_pocket expenses. |
| paid_coinsurance | No | number(8,2) | | The amount paid by the person as a joint assumption of risk. Typically, this is a percentage of the expenses defined by the payer plan (policy) after the person's deductible is exceeded. |
| paid_toward_deductible | No | number(8,2) | | The amount paid by the person that is counted toward the deductible defined by the payer plan (policy). |
| paid_by_payer | No | number(8,2) | | The amount paid by the payer (insurer). If there is more than one payer, several procedure_cost records indicate that fact. |
| paid_by_coordination_benefits | No | number(8,2) | | The amount paid by a secondary payer through the coordination of benefits. |
| total_out_of_pocket | No | number(8,2) | | The total amount paid by the person as a share of the expenses, excluding the copay. |
| total_paid | No | number(8,2) | | The total amount paid for the expenses of the procedure. |
| disease_class_concept_id | No | integer | DRG, APC | A foreign key referring to a standard concept identifier in the vocabulary for disease classes, such as DRGs and APCs. |

| Field | Required | Type Precision | Standard | Description |
|---|---|---|---|---|
| revenue_code_concept_id | No | integer | HCFA | A foreign key referring to a standard concept identifier in the vocabulary for revenue codes. |
| payer_plan_period_id | No | integer | | A foreign key to the payer_plan_period table, where the details of the payer, plan and family are stored. |
| disease_class_source_value | No | string(50) | | he source code for the disease class as it appears in the source data, stored here for reference. |
| revenue_code_source_value | No | string(50) | | The source code for the revenue code as it appears in the source data, stored here for reference. |

### 3.12.1 BUSINESS RULES

Each Procedure Occurrence may have any number of corresponding records in the Procedure Cost table, but typically it is none (cost data not captured) or one (one payment per Procedure). They are linked directly through the Procedure Occurrence ID field.

The amounts paid are:

- Copay – a fixed amount to be paid by the Person
- Coinsurance – a relative amount of the total paid by the Person
- Deductible – an amount of money paid by the Person before the Payer starts contributing
- Primary Payer – the amount the primary Payer pays towards the total
- Coordination of Benefits – the amount a secondary Payer or Family Plan pays towards the total
- Out of Pocket = Copay + Coinsurance + Deductible
- Total – the total amount paid for the procedure

The amounts in various payment components should equal the total, so Copay + Coinsurance + Deductible + Primary Payer + COB = Total Paid. In reality, this is not always reflected in the source data. It is up to the ETL to determine how to deal with quality problems in the data.

There are important indicators for the amount paid that are determined through the health plan design:

- DRG – Diagnosis-related Group for hospital inpatients
- APC – Ambulatory Payment Classification for hospital outpatients
- Revenue Codes – determining what service within a provider is charging for the service

All these data are captured as Source Values and Concept IDs referring to the Vocabulary.

Finally, the health plan of the Person that is determined by these numbers is referred to through the Payer Plan Period ID (see below).

### 3.12.2 EXAMPLE OF A LOADED TABLE

The select three example drug claims are represented as following:

- Example 1: Outpatient, biopsy of orbital lesion, copay $20, coinsurance $48, no deductible, insurance $92, no coordination of benefits, total of $160
- Example 2: Outpatient, myocardial perfusion imaging (SPECT), no patient pay, insurance $207.76, coordination of benefits $394.24, total of $602.00

- Example 3: Inpatient, total abdominal hysterectomy, coinsurance $1,299.97, deductible $300, insurance $3,033.29, no coordination of benefits, total of $4,633.26, DRG 369 (Menstrual and Other Female Reproductive System Disorders), Revenue Code 360 (Operating Room Services - General Classification)
- Example 4: Inpatient, total abdominal hysterectomy, no copay or coinsurance, total of $2,671.14 paid through deductible, DRG 359 (Uterine and Adnexa Procedure For Non-Malignancy without Complications, Comorbidities), Revenue Code 0360 (Operating Room Services - General Classification)

| Field | Example 1 | Example 2 | Example 3 | Example 4 |
|---|---|---|---|---|
| procedure_cost_id | integer foreign key | integer foreign key | integer foreign key | integer foreign key |
| procedure_occurrence_id | integer foreign key | integer foreign key | integer foreign key | integer foreign key |
| paid_copay | 20 | 0 | 0 | 0 |
| paid_coinsurance | 48 | 0 | 1299.97 | 0 |
| paid_toward_deductible | 0 | 0 | 300 | 2671.14 |
| paid_by_payer | 92 | 207.76 | 3033.29 | 0 |
| paid_by_coordination_benefits | 0 | 394.24 | 0 | 0 |
| total_out_of_pocket | 48 | 0 | 1599.97 | 2671.14 |
| total_paid | 160 | 602 | 4633.26 | 2671.14 |
| disease_class_concept_id | | | 38000667 | 38000657 |
| revenue_code_concept_id | | | 38003208 | 38003208 |
| payer_plan_period_id | integer foreign key | | integer foreign key | integer foreign key |
| disease_class_source_value | | | 369 | 359 |
| revenue_code_source_value | | | 0360 | 0360 |

## 3.13 Location

The Location table represents a generic way to capture physical location or address information. Each address or Location must be only present once in the table. Locations are used to define the addresses for Persons, Care Sites and Organizations. Locations do not contain names; 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 or Organization (the Person table does not contain name information).

| Field | Required | Type Precision | Standard | Description |
|---|---|---|---|---|
| location_id | Yes | integer | | A system-generated unique identifier for each geographic location. |
| address_1 | No | string(50) | | The address field 1, typically used for the street address, as it appears in the source data. |
| address_2 | No | string(50) | | The address field 2, typically used for additional detail such as buildings, suites, floors, as it appears in the source data. |
| city | No | string(50) | | The city field as it appears in the source data. |
| state | No | string(2) | | The state field as it appears in the source data. |
| zip | No | string(9) | | The zip code. For US addresses, valid zip codes can be 3, 5 or 9 digits long, depending on the source data. |

| Field | Required | Type Precision | Standard | Description |
|---|---|---|---|---|
| county | No | string(20) | | The county. The county information is necessary because not all zip codes fall into one and the same county. |
| location_source_value | No | string(50) | | The verbatim information that is used to uniquely identify the location as it appears in the source data. |

### 3.13.1 BUSINESS RULES

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 will expect and utilize only the first 3 digits. For international addresses, different rules apply.

### 3.13.2 EXAMPLE OF A LOADED TABLE

The examples show the addresses of the following records:

- Massachusetts General Hospital in Boston, MA
- Guardian Urgent Care in Denver, CO
- Internal Medicine Physician Office in Carmel, IN
- 3-letter zip code of Person 7723462

| Field | Example 1 | Example 2 | Example 3 | Example 4 |
|---|---|---|---|---|
| location_id | 36433 | 36434 | 36435 | 7723462 |
| address_1 | 55 Fruit Street | 1 Broadway | 11911 N Meridian St | |
| address_2 | | Building A, Suite 100 | Suite 110 | |
| city | Boston | Denver | Carmel | |
| state | MA | CO | IN | |
| zip | 2114 | 80203 | 46032 | 993 |
| county | Suffolk | Denver | Hamilton | Benton |
| location_source_value | | | | 993XX |

## 3.14 Provider

The Provider table contains a list of uniquely identified health care providers (physicians).

| Field | Required | Type Precision | Standard | Description |
|---|---|---|---|---|
| provider_id | Yes | integer | | A system-generated unique identifier for each provider. |
| npi | No | string(20) | | The National Provider Identifier (NPI) of the provider. |
| dea | No | string(20) | | The Drug Enforcement Administration (DEA) number of the provider. |
| specialty_concept_id | No | integer | CDC | A foreign key to a standard provider's specialty concept identifier in the vocabulary. |

| Field | Required | Type Precision | Standard | Description |
|---|---|---|---|---|
| care_site_id | No | integer | | A foreign key to the main care site where the provider is practicing. |
| provider_source_value | Yes | string(50) | | The identifier used for the provider in the source data, stored here for reference. |
| specialty_source_value | No | string(50) | | The source code for the provider specialty as it appears in the source data, stored here for reference. |

### 3.14.1 BUSINESS RULES

Providers are defined in public databases through their NPI number or DEA number. However, at a minimum for the purpose of the CDM, only the Provider Source Value is required. It is up to the ETL to ensure that providers are not duplicated in the table. The Specialty Concept ID are references to the Vocabulary, containing CDC-derived Concepts for Specialties.

There is no defined relationship between Providers and Organizations or Care Sites.

### 3.14.2 EXAMPLE OF A LOADED TABLE

The following three shown example providers are represented as following:

- Family medicine practitioner, NPI# 12345678, Provider ID 9238475, practicing at a specified care site
- Nurse Practitioner, DEA# BB6548987, Provider ID Fhkla8902
- Midwife, Provider ID 892slknv

The specialty codes for Family Practice are 38004453, Nurse Practitioner 38004487 and Midwife 38004482.

| Field | Example 1 | Example 2 | Example 3 |
|---|---|---|---|
| provider_id | 1 | 5436 | 548 |
| npi | 12345678 | | |
| dea | | BB6548987 | |
| specialty_concept_id | 38004453 | 38004487 | 38004482 |
| care_site_id | integer foreign key | | |
| provider_source_value | 9238475 | Fhkla8902 | 892slknv |
| specialty_source_value | Family Medicine | NP | Midwife |

## 3.15 Organization

The Organization table contains a list of uniquely identified health care organizations (hospitals, clinics, practices, etc.).

| Field | Required | Type Precision | Standard | Description |
|---|---|---|---|---|
| organization_id | Yes | integer | | A system-generated 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. |
| place_of_service_concept_id | No | integer | CMS | A foreign key that refers to a place of service concept identifier in the vocabulary. |
| 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. |
| organization_source_value | Yes | string(50) | | The identifier for the organization in the source data, stored here for reference. |
| place_of_service_source_value | No | string(50) | | The source code for the place of service as it appears in the source data, stored here for reference. |

### 3.15.1 Business Rules

An Organization is an administrative health care entity that consists of one or more Care Sites. This is the highest level of the health care hierarchy. Most Organizations contain multiple Care Sites, but each Care Site belongs to one Organization. Location information about the organization is stored in the location table and referenced by the foreign key “Organization Location ID”. For the purposes of the OMOP CDM, organizations typically share a single EHR database. As a minimum, the Organization Source Value is needed for a valid record. In addition, if available the address (Location ID reference to the Location table) and the Place Of Service Concept ID (reference to the Vocabulary) can be provided.

### 3.15.2 Example of a Loaded Table

The example shows records for the following Organizations in the source data:

- Massachusetts General Hospital (MGH) in Boston, MA, Academic Medical Center with in and outpatient departments
- Guardian Urgent Care in Denver, CO, Medical walk-in clinic
- General or Family Practice "Internal Medicine Of Carmel" in Carmel, IN

| Field | Example 1 | Example 2 | Example 3 | Example 4 |
|---|---|---|---|---|
| organization_id | 1 | 2 | 3 | 4 |
| place_of_service_concept_id | 8717 | 8756 | 8782 | 8940 |
| location_id | integer foreign key | integer foreign key | integer foreign key | integer foreign key |
| organization_source_value | MGH | MGH | Guardian Urgent Care | Internal Medicine of Carmel |

| Field | Example 1 | Example 2 | Example 3 | Example 4 |
|---|---|---|---|---|
| place_of_service_source_value | Hospital | Hospital Outpatient Dept | Urgent Care Facility | Family Practice |

## 3.16 Care Site

The Care Site table contains a list of uniquely identified points of care, or an individual clinical location within an organization. Each care site belongs to one organization. There might be more than one Care Site in a Location (address).

| Field | Required | Type Precision | Standard | Description |
|---|---|---|---|---|
| care_site_id | Yes | integer | | A system-generated unique identifier for each care site. A care site is the place where the provider delivered the healthcare to the person. |
| location_id | No | integer | | A foreign key to the geographic location in the location table, where the detailed address information is stored. |
| organization_id | No | integer | | A foreign key to the organization in the organization table, where the detailed information is stored. |
| place_of_service_concept_id | No | integer | CMS | A foreign key to the predefined concept identifier in the vocabulary reflecting the place of service. |
| care_site_source_value | No | string(50) | | The identifier for the care site as it appears in the source data, stored here for reference. |
| place_of_service_source_value | No | string(50) | | The source code for the place of service as it appears in the source data, stored here for reference. |

### 3.16.1 Business Rules

A Care Site is a physical point of care where Persons are seen by Providers for providing health care. There is no useful minimum a Care Site record becomes valid, because dependent on the source a variety of data may or may not be available. However, either the address (Location ID reference to the Location table), the Place of Service Concept ID (reference to the Vocabulary) or the Organization ID that the Care Site belongs to should be provided.

There is a one-to-many relationship between Organizations and Care Sites.

### 3.16.2 Example of a Loaded Table

The example shows records for the following Care Sites that are part of Massachusetts General Hospital:

- Cancer Center
- Digestive Healthcare Center
- Transplant Center

| Field | Example 1 | Example 2 | Example 3 |
|---|---|---|---|
| care_site_id | 234 | 4237 | 2381 |
| location_id | integer foreign key | integer foreign key | integer foreign key |
| organization_id | integer foreign key | integer foreign key | integer foreign key |
| place_of_service_concept_id | 8870 | 8717 | 8717 |
| care_site_source_value | Massachusetts General Hospital Department of Emergency | MGH Digestive Health Center | MGH Transplant Center |
| place_of_service_source_value | ER | Hospital Department | Hospital Department |

## 3.17 Payer Plan Period

Each Person receiving health care and covered by a health benefits is subject to a Plan defined by the Payer for the Person or his Family. For a given benefit policy, there may be one or more Plans that are active for certain periods of time (e.g. before and after the Deductible is reached), determining the cost of Drug Exposure and Procedure Occurrence.

| Field | Required | Type Precision | Standard | Description |
|---|---|---|---|---|
| payer_plan_period_id | Yes | integer | | A system-generated identifier for each unique combination of payer, plan, family code and time span. |
| person_id | Yes | integer | | A foreign key identifier to the person covered by the payer. The demographic details of that person are stored in the person table. |
| payer_plan_period_start_date | Yes | date | | The start date of the payer plan period. |
| payer_plan_period_end_date | Yes | date | | The end date of the payer plan period. |
| payer_source_value | No | string(50) | | The source code for the payer as it appears in the source data. |
| plan_source_value | No | string(50) | | The source code for the person's coverage plan as it appears in the source data. |
| family_source_value | No | string(50) | | The source code for the person's family as it appears in the source data. |

### 3.17.1 BUSINESS RULES

Different Payers have different designs for their health benefit Plans. The Payer Plan Period table does not capture all details of the plan design or the relationship between the Plan and the cost of healthcare. However, it allows identifying the unique combination of Payer (insurer), Plan (determining health care benefits and limits) and Family membership for each Person. Usually, depending on health care utilization a Person may have one or many subsequent Plans during coverage by a single Payer. The tables captures the period a plan is active (Start and End Date), the Payer Source Value (name or ID of the Payer), the Plan Source Value (name or ID of the Plan) and Family Source Value (ID of the family, which may consist of just one family member – the Person).

## 3.18 Cohort

The Cohort table contains records who share a particular feature during a particular time span (eg, cohort of patients, visits, providers). Cohorts can be defined as group of entities exposed to a common circumstance. For example, Health Outcome of Interest (HOI) cohorts define a group of Persons with the same Condition as defined by the HOI definition. Each person can have one or many records, meaning, they can be part of the cohort one or many times.

Likewise, Providers can form cohorts that share a common feature, like the availability of a certain diagnostic or treatment facility. Finally, visits can be combined to cohorts if they again share a common feature. Cohorts can be derived from the observational data (like HOI occurrences), or they can be applied in the framework of a study (like for testing the efficacy of certain Provider quality programs). Note that records in Cohort do not create a protocol definition, they represent the elements that belong to such definition.

| Field | Required | Type Precision | Standard | Description |
|---|---|---|---|---|
| cohort_id | Yes | integer | | A system-generated unique identifier for each cohort record. |
| cohort_concept_id | Yes | integer | | A foreign key to a standard cohort concept identifier in the vocabulary. Cohort concepts identify the cohorts: whether they are defined through persons, providers or visits, or any combination thereof. |
| cohort_start_date | Yes | date | | The date when the cohort definition criteria for the person, provider or visit first match. |
| cohort_end_date | No | date | | The date when the cohort definition criteria for the person, provider or visit no longer match or the cohort membership was terminated. |
| subject_id | Yes | integer | | A foreign key to the subject in the cohort. These could be referring to records in the Person, Provider, Visit Occurrence table. |
| stop_reason | No | string(20) | | The reason for the end of a cohort membership other than defined by the cohort definition criteria as it appears in the source data. |

### 3.18.1 Business Rules

The core of a Cohort is the definition of the unifying definition or feature of the Cohort. This is captured in the Cohort Concept ID. Cohorts must have a Start Date, might have an End Date, and should contain a Subject ID. To define whether the cohort was terminated in a planned or unplanned fashion, the Stop Reason can be used.

In the case of Health Outcome of Interest (HOI) for example, the Cohort Concept ID defines the HOI (e.g. "Acute Liver Failure 1", and the Subject ID are the Persons who are defined as having the HOI.

### 3.18.2 Example of a Loaded Table

For the Health Outcome of Interest "Acute Renal Failure 1" is defined as a set of ICD-9-CM diagnosis codes (see OMOP website for details of these definitions). Any Person with a record in the Condition table that corresponds to this definition is a member of the cohort, e.g.:

- Patient 32022001, 14-Jun-07, Condition Concept ID 197320 "Acute renal failure syndrome", recorded in Inpatient Detail 2

- Patient 31938902, 2-Aug-06, Condition Concept ID 197320 "Acute renal failure syndrome", recorded in Outpatient Header 1 and Outpatient Detail 1
- Patient 1398201, 18-Oct-05, Condition Concept ID 197320 "Acute renal failure syndrome", recorded in Primary Inpatient Detail, Primary Inpatient Header and Inpatient Header 1

The cohort Concept ID in the Concept table for "Acute Renal Failure 1" is 500000401. Since this is a Person-based cohort, Person ID is used for Subject ID.

| Field | Example 1 | Example 2 | Example 3 |
|---|---|---|---|
| cohort_id | 50026078 | 50028491 | 50030900 |
| cohort_concept_id | 500000401 | 500000401 | 500000401 |
| cohort_start_date | 14-Jun-07 | 2-Aug-06 | 18-Oct-05 |
| cohort_end_date | 14-Jun-07 | 2-Aug-06 | 18-Oct-05 |
| subject_id | 32022001 | 31938902 | 1398201 |
| stop_reason | | | |

# 4 Vocabulary Logical Data Model

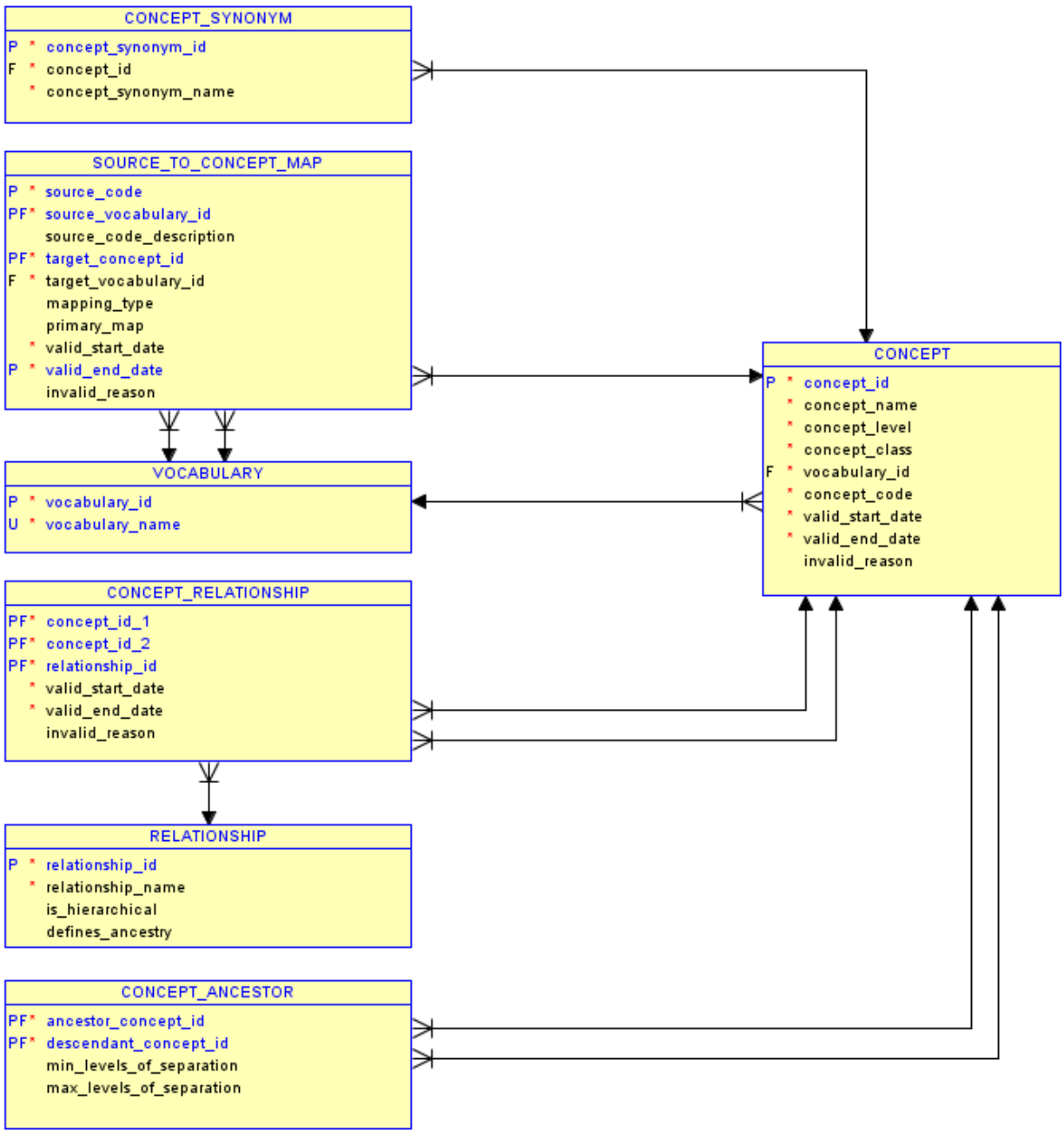
A number of assumptions were made for the design of the Vocabulary tables:

- There is one design which will accommodate all different source terminologies and classifications.
- Source terminologies might get loaded exhaustively, but only domains that are relevant for outcome research are declared Standard Vocabulary and receive a Concept Level. For example, Clinical Drug is a Concept of the Standard Vocabulary, but Clinical Drug Form is not.
- Source Vocabulary data need to be transformed prior to loading CDM data, as they are used to map Source Values to Concepts.
- Source Concepts are preserved, but the Source relationships might be adapted somewhat. For example, an inferred relationship between Clinical Drug and Ingredient is added that is constructed from the relationships between Clinical Drug and Clinical Drug Form, and Clinical Drug Form and Ingredient.
- Concept hierarchy contains both individual relationships as well as Ancestor and Descendant relationships to simplify building of Drug and Condition Eras.

The advantage of this approach lies in the preservation of source relationships without adherence to the multiple different source data structures, a simple design for standardized access, and the optimization of performance for OMOP analysis. The navigation does not require any knowledge of source vocabulary. Finally, the approach is scalable and future vocabularies can be integrated easily. On the other hand, extensive transformation of source data to the Vocabulary is required and not every source data structure and original source hierarchy can be retained.

Diagram 6: Entity-relationship diagram of the Vocabulary tables and their relationships



## 4.1 Concept

In the context of the Vocabulary, a Concept is a basic unit of medical information that is identified by a unique static identifier. Concepts can represent broad categories (like “Cardiovascular disease”), detailed clinical elements (”Myocardial infarction of the anterolateral wall”) or characteristics and relationships that define Concepts at various levels of detail (severity of a disease, associated morphology, etc.).

Records in the Concept tables are derived from standard national or international vocabularies such as SNOMED-CT, NDF-RT, and MedDRA, or custom Concepts defined to cover various aspects of observational data analysis. The detailed description of all source vocabularies, their implementation, the definitions of the relationships, the choice of hierarchical relationships to define ancestry between concepts as well as the mapping from non-standard vocabularies into the standard vocabularies is described in a separate specification document, the OMOP Standard Vocabulary Specification.

| Field | Required | Type Precision | Description |
|---|---|---|---|
| concept_id | Yes | integer | A system-generated unique identifier for each concept across all domains. |
| concept_name | Yes | string(256) | An unambiguous, meaningful and descriptive name for the concept. |
| concept_level | Yes | integer | The level of hierarchy associated with the concept. Different concept levels are assigned to concepts to depict their seniority in a clearly defined hierarchy, such as drugs, conditions, etc. A concept level of 0 is assigned to concepts that are not part of a standard vocabulary, but are part of the vocabulary for reference purposes (e.g. drug form). |
| concept_class | Yes | string(60) | The category or class of the concept along both the hierarchical tree as well as different domains within a vocabulary. Examples are “Clinical Drug”, “Ingredient”, “Clinical Finding” etc. |
| vocabulary_id | Yes | integer | A foreign key to the vocabulary table indicating from which source the concept has been adapted. |
| concept_code | Yes | string(20) | The concept code represents the identifier of the concept in the source data it originates from, such as SNOMED-CT concept IDs, RxNorm RXCUIs etc. Note that concept codes are not unique across vocabularies. |
| valid_start_date | Yes | date | The date when the concept was first recorded. |
| valid_end_date | Yes | date | The date when the concept became invalid because it was deleted or superseded (updated) by a new concept. The default value is 31-Dec-2099. |
| invalid_reason | No | string(1) | Reason the concept was invalidated. Possible values are D (deleted), U (replaced with an update) or NULL when valid_end_date has the default value. |

### 4.1.1 BUSINESS RULES

Concepts in the Common Data Model are derived from a number of public or commercial terminologies such as SNOMED-CT and MedDRA, or custom generated to standardize aspects of observational data. Both standard and custom Concepts are integrated based on the following rules:

- All Concepts are maintained centrally by OMOP. Additional concepts can be added, as needed, upon request.
- For all Concepts, whether they are custom generated or adopted from published terminologies, a unique numeric identifier is assigned and used as the key to link all observational data to the corresponding Concept reference data.

- A descriptive name for each Concept is stored as the Concept name as part of the Concept table. Additional names and descriptions for the Concept are stored as Synonyms in the Concept Synonym table.
- For standard Concepts inherited from published terminologies, the source Concept Code is retained as part of the Concept reference data and used to reference the source vocabulary.
- All logical data elements associated with the various CDM tables, usually called Types, including defining characteristics, qualifying attributes etc. are also stored as Concepts. Since they are generated by OMOP, their Source Code is omitted.
- The lifespan of concepts is recorded through their Valid Start Date, Valid End Date and the Reason for Invalidation. This allows to concepts to correctly reflect at which point in time were in active clinical use. For example, drugs that are taken off the market might be dropped by the terminology vendor. However, since observational data are valid with respect to the time they are recorded, it is key for a Vocabulary to provide even obsolete codes and maintain their relationships to other Concepts and Classifications.

### 4.1.2 EXAMPLE OF A LOADED TABLE

Each concept, whether external or generated by OMOP, has a single record, for example:

| Field | Example |
|---|---|
| concept_id | 19055132 |
| concept_name | ADIPHENINE HCL |
| concept_level | 2 |
| concept_class | Ingredient |
| vocabulary_id | 8 |
| concept_code | 235433 |
| valid_start_date | 16-Jun-05 |
| valid_end_date | 06-Jul-09 |
| invalid_reason | D |

For a detailed discussion of the Concepts see the separate Standard Vocabulary Specifications Document.

## 4.2 Concept Synonym

The Synonym table is used to store all alternate names and descriptions for a Concept. Each Synonym is assigned its own unique identifier and contains the text of a description and the identifier of the Concept that it represents.

Each Concept may include zero, one, or more Synonyms in the Synonym table. As an example, for a SNOMED-CT Concept, if the fully specified name is stored as the Concept name in the Concept table, then the Preferred Term and Synonyms associated with the Concept are stored in the Synonym table. Synonyms are used to express Descriptions that may denote the same basic Concept but are expressed in different terms. Only synonyms that are active and current are stored in the Synonym table. Tracking synonym/description history and mapping of obsolete synonyms to current Concepts/Synonyms is out of scope for observational analysis. Synonyms entities are stored in the Concept_Synonym table.

| Field | Required | Type Precision | Description |
|---|---|---|---|
| concept_synonym_id | Yes | integer | A system-generated unique identifier for each concept synonym. |
| concept_id | Yes | integer | A foreign key to the concept in the concept table. |
| concept_synonym_name | Yes | string(1000) | The alternative name for the concept. |

### 4.2.1 EXAMPLE OF A LOADED TABLE

The following example illustrates the representation of alternative naming associated with a SNOMED-CT Concept “Chronic Atrial Fibrillation’. The Concept can be described in the following ways:

- Concept Name: Chronic atrial fibrillation
- Alternative Name: Chronic atrial fibrillation (disorder)

| Concept Synonym ID | Concept ID | Concept Synonym Name |
|---|---|---|
| 3409069 | 4141360 | Chronic atrial fibrillation (disorder) |

For a detailed discussion of the Concepts see the separate Standard Vocabulary Specifications Document.

## 4.3 Concept Relationship

The Concept Relationship table stores the relationship between two Concepts. The relationships described in the CDM are directional and are intended to include a Source Concept and a Target Concept with an explicit relationship from Source to Target.

Concept Relationship includes many different types of relationships between Concepts. The type of relationships is defined in the Relationship table. Generally, there are hierarchical (parent-child) and non-hierarchical relationships. The parent-child Concepts hierarchy in the Standard Vocabulary is of special importance to OMOP, as it allows researchers to query a CDM instance for classes of Concepts without needing to know the underlying subclasses. For example, a researcher will be able to query a CDM instance for all drugs within a specific therapeutic class without needing to know the specific Concepts codes of each drug within the class, and will be able to query a CDM instance for a particular medical condition without the necessity to know the individual indicators (i.e., diagnoses) of that condition. The researcher would search or browse the Vocabulary to find the class of Concepts on which to query or analyze, then transfer the appropriate Concept Code of that class from the Vocabulary to the query and analysis tool.

| Field | Required | Type Precision | Description |
|---|---|---|---|
| concept_id_1 | Yes | Integer | A foreign key to the concept in the concept table associated with the relationship. Relationships are directional, and this field represents the source concept designation. |
| concept_id_2 | Yes | Integer | A foreign key to the concept in the concept table associated with the relationship. Relationships are directional, and this field represents the destination concept designation. |
| relationship_id | Yes | string(3) | The type of relationship as defined in the relationship table. |
| valid_start_date | Yes | Date | The date when the instance of the relationship is first recorded. |
| valid_end_date | Yes | Date | The date when the relationship became invalid because it was deleted or superseded (updated) by a new relationship. Default value is 31-Dec-2099. |
| invalid_reason | No | string(1) | Reason the relationship was invalidated. Possible values are D (deleted), U (replaced with an update) or NULL when valid_end_date has the default value. |

All Relationships are symmetrical, ie. they exist twice for the convenience of the searcher, who doesn't have to know the directionality of a relationship – all Relationships are bidirectional. For example, an "isa" relationship (144 in the Relationship table) between the child Concept in concept_id_1 and the parent

Concept in concept_id_2 will be also exist as a "subsumes" Relationship between the parent in concept_id_1 and the child in concept_id_2.

### 4.3.1 EXAMPLE OF A LOADED TABLE

The following examples illustrate the representation of Concept Relationships in the CDM.

The Concept “Chronic Atrial Fibrillation” includes a hierarchic subtype relationship with “Atrial Fibrillation” and an attribute relationship with a defining characteristic “Severity” Concept. The relationship is represented as follows:

- Concept 4141360 (Chronic atrial fibrillation (disorder)) is related to Concept 313217 (Atrial Fibrillation) through Relationship ID 10 (“Isa”) and Concept 313217 is connected to 4141360 through the 144 ("Subsumes") relationship.
- Concept 4141360 (Chronic atrial fibrillation (disorder)) is related to Concept 4153899 (Severities) through Relationship ID 34 (has severity) and with reverse direction through Relationship ID 168 (severity of)

The resulting records will look as follows:

| Concept ID 1 | Concept ID 2 | Relationship ID | Valid Start Date | Valid End Date | Invalid Reason |
|---|---|---|---|---|---|
| 4141360 | 313217 | 10 | 01-Jan-80 | 31-Dec-99 | |
| 4141360 | 4153899 | 34 | 01-Jan-80 | 31-Dec-99 | |
| 313217 | 4141360 | 144 | 01-Jan-80 | 31-Dec-99 | |
| 4153899 | 4141360 | 168 | 01-Jan-80 | 31-Dec-99 | |

## 4.4 Concept Ancestor

Concept Ancestor table is a custom table designed to simplify observational analysis by consolidating the hierarchical relationship between various Concepts. Parent-child relationships between Concepts are stored in the Concept Relationship table. However, it is stored in a form that is hard to navigate due to the interlocking nature of the relationships and the multiplicity of parent-child relationships for many Concepts.

The Ancestor-Descendant relationship captures hierarchical relationships between Ancestor and any Descendant Concepts, along with indicators for the shortest and longest navigation path (Maximum and Minimum Levels of Separation) between them.

The Ancestor relationship is primarily targeted at observational analysis that would involve:

- Rollup of lower level Concepts into higher-level aggregation Concepts.
- Collection of all lower level Concepts in the hierarchy that follow from a high level Concept.

| Field | Required | Type Precision | Description |
|---|---|---|---|
| ancestor_concept_id | Yes | integer | A foreign key to the concept code in the concept table for the higher-level concept that forms the ancestor in the relationship. |
| descendant_concept_id | Yes | integer | A foreign key to the concept code in the concept table for the lower-level concept that forms the descendant in the relationship. |

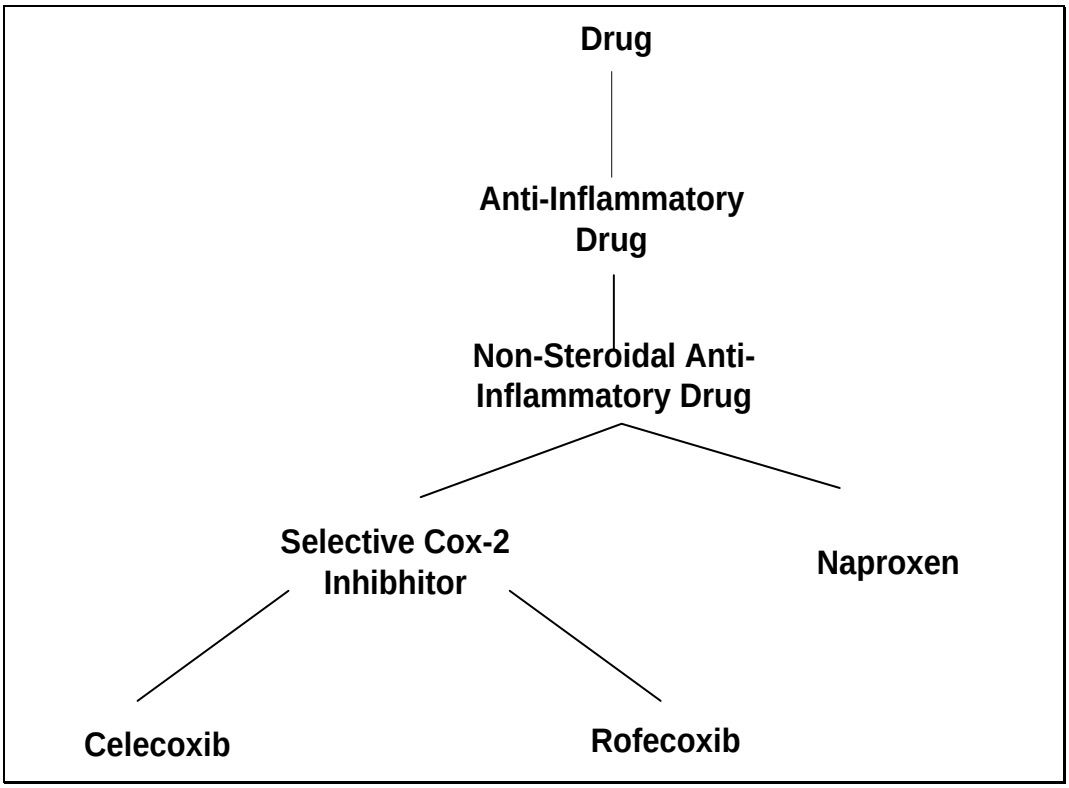
 

| Field | Required | Type Precision | Description |
|---|---|---|---|
| min_levels_of_separation | No | integer | The minimum separation in number of levels of hierarchy between ancestor and descendant concepts. This is an optional attribute that is used to simplify hierarchic analysis. |
| max_levels_of_separation | No | integer | The maximum separation in number of levels of hierarchy between ancestor and descendant concepts. This is an optional attribute that is used to simplify hierarchic analysis. |

### 4.4.1 EXAMPLE OF A LOADED TABLE

Diagram 7 illustrates in the CDM the representation of the Ancestor – Descendant relationships between Concepts. The Non-steroidal anti-inflammatory drugs (NSAID) hierarchy is represented in the following diagram based on hypothetical drug ontology:

Diagram 7: Representation of the Ancestor




Based on the ontology described in the Ancestor diagram, two of the Ancestor – Descendant relationships could be captured as follows:

Relationship 1:

- Ancestor Concept: "Non Steroidal Anti-inflammatory drugs"
- Ancestor Concept ID: 16403005
- Descendant Concept: "Naproxen"
- Descendant Concept ID: 4186860
- Maximum Levels of separation: 1
- Minimum Levels of separation: 1

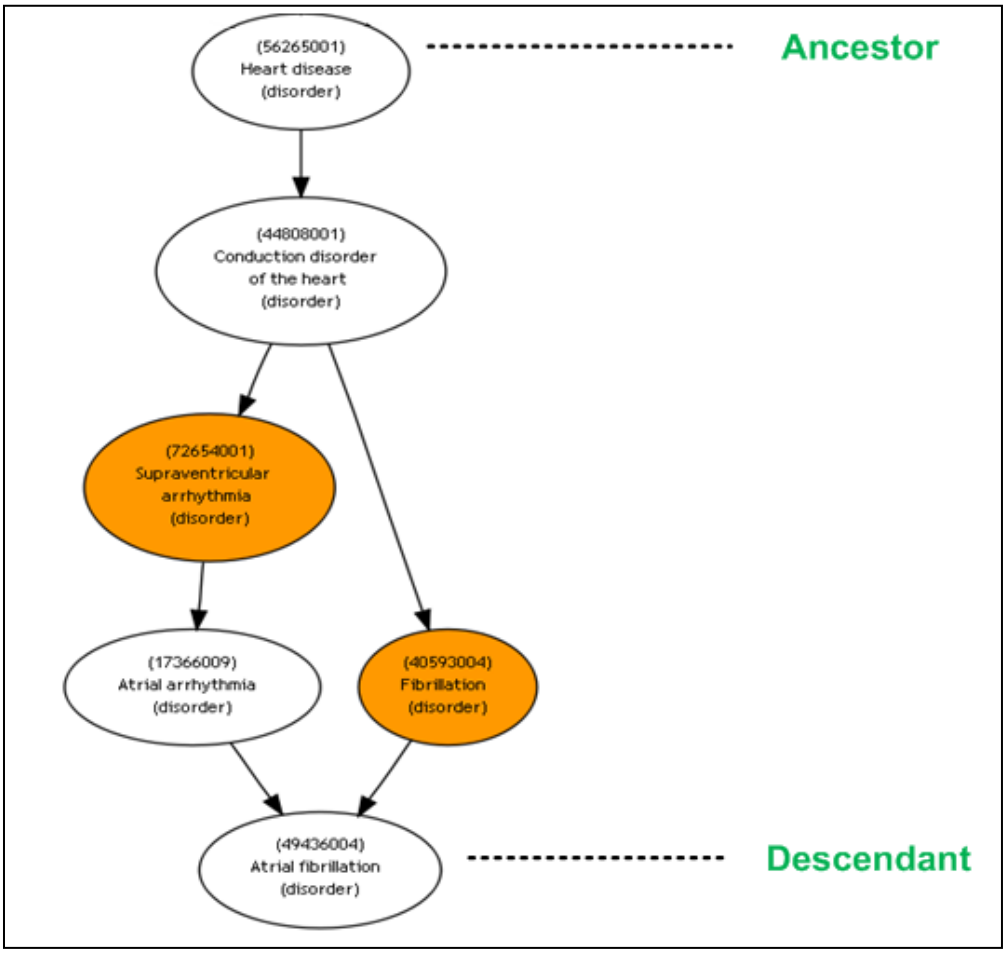
Relationship 2:

- Ancestor Concept: "Non Steroidal Anti-inflammatory drugs"
- Ancestor Concept ID: 16403005
- Descendant Concept: "Celecoxib"
- Descendant Concept ID: 4021058
- Ancestor Level to Root: 2
- Ancestor Maximum Levels to Leaf: 2

Consolidated, the data represented in the CDM as follows:

| Concept Ancestor Map ID | Ancestor Concept ID | Descendant Concept ID | Max Levels of Separation | Min Levels of Separation |
|---|---|---|---|---|
| 1 | 16403005 | 4186860 | 1 | 1 |
| 2 | 16403005 | 4021058 | 2 | 2 |

Diagram 8: Ancestor to Descendant displays an example of a partial hierarchy, where navigating the hierarchy from the Ancestor to the Descendant involves multiple paths.



Details of the examples are as follows:

- Ancestor: SNOMED-CT Concept 56265001 “Heart disease (disorder)”
- Ancestor Concept ID: 321588
- Descendant: SNOMED-CT Concept 49436004 “Atrial Fibrillation (disorder)”
- Descendant Concept ID: 4344544
- Min Levels of Separation: 3
- Max Levels of Separation: 2

| Ancestor Concept ID | Descendant Concept ID | Max Levels of Separation | Min Levels of Separation |
|---|---|---|---|
| 321588 | 4344544 | 4 | 3 |

Variance in minimum and maximum levels is an indicator of the complex traversal path that offers alternate navigation paths.

## 4.5 Vocabulary

The Vocabulary table includes a list of the standard vocabularies collected from various sources or created de novo by the OMOP community. This reference table is populated with a single record for each Vocabulary source and includes a descriptive name for the Vocabulary.

| Field | Required | Type Precision | Description |
|---|---|---|---|
| vocabulary_id | Yes | integer | A unique identifier for each vocabulary. |
| vocabulary_name | Yes | string(256) | The name describing the vocabulary, for example "SNOMED-CT", "ICD-9", "Visit", etc. |

### 4.5.1 Example of a Loaded Table

For a complete listing of this reference table see separate Standard Vocabulary Specifications Document.

## 4.6 Source To Concept Map

Observational data elements that need to be standardized into Concepts require mapping tables to translate source codes into standard Concept Identifiers. The mapping entity serves as a lookup table that stores a cross-reference between source codes and standard Concept Identifiers. Concept mappings need to be created for both:

- Mapping of Concept Codes of a widely adopted standard to other standard Concept Identifiers, such as mapping of ICD-9-CM diagnosis codes to SNOMED-CT Clinical finding Concepts or NDC codes to RxNorm Drug Concepts.
- Mapping of source specific or original codes to standard Concepts wherever possible, such as mapping of source specific problem strings to SNOMED-CT clinical findings or source specific Gender codes to HL7 Sex Concepts.

The mapping table serves the critical function of inferring standard Concepts from source data created with various objectives in mind. The ETL logic that loads the CDM instance from the source dataset looks-up Concepts from the Vocabulary as it transforms (i.e., reformats) the source and Concepts data to

conform to the CDM table structures. The look-up process entails matching the inbound source data (vocabulary and source code) to Target Concept Identifier values.

Source To Concept Map records are generally unique for each Source Code, Vocabulary and Mapping Type. However, sometimes it is impossible to create a single unique mapping for the trio combination. In those cases, there is more than one record in the table, but one record is marked as primary ("Y" in the Primary Map field).

| Field | Required | Type Precision | Description |
|---|---|---|---|
| source_code | Yes | string(20) | The source code being translated into a standard concept. |
| source_vocabulary_id | Yes | integer | A foreign key to the vocabulary table defining the vocabulary of the source code that is being mapped to the standard vocabulary. |
| source_code_description | No | string(256) | An optional description for the source code. This is included as a convenience to compare the description of the source code to the name of the concept. |
| target_concept_id | Yes | integer | A foreign key to the concept to which the source code is being mapped. |
| target_vocabulary_id | Yes | integer | A foreign key to the vocabulary table defining the vocabulary of the target concept. |
| mapping_type | No | string(20) | A string identifying the observational data element being translated. Examples include 'DRUG', 'CONDITION', 'PROCEDURE', 'PROCEDURE DRUG' etc. It is important to pick the appropriate mapping record when the same source code is being mapped to different concepts in different contexts. As an example a procedure code for drug administration can be mapped to a procedure concept or a drug concept or both. |
| primary_map | No | string(1) | A boolean value identifying the primary mapping relationship for those sets where the source_code, the source_vocabulary_id and the mapping type is identical (one-to-many mappings). The ETL will only consider the primary map. Permitted values are Y and NULL. |
| valid_start_date | Yes | date | The date when the mapping instance was first recorded. |
| valid_end_date | Yes | date | The date when the mapping instance became invalid because it was deleted or superseded (updated) by a new relationship. Default value is 31-Dec-2099. |
| invalid_reason | No | string(1) | Reason the mapping instance was invalidated. Possible values are D (deleted), U (replaced with an update) or NULL when valid_end_date has the default value. |

### 4.6.1 BUSINESS RULES

The Source To Concept Map is the only place where codes are stored that are not part of the Standard Vocabulary. For details, look into the Standard Vocabulary Specification Document. But briefly, for each domain there is one set of codes that is used as a standard for the CDM. All other codes are mapped to this standard using this table. There are several rules:

- Source codes are not unique across vocabularies. A source code should always be looked up in combination with the vocabulary.
- Vocabularies are defined in the Vocabulary table.

- Some vocabularies re-use codes over time (e.g. NDC 74227414 was used from 23 to 30-May-2002 for Vicoprofen 200-75 mg, and since 14-Aug-2002 for Vicodon HP 10-660 mg). Therefore, it is necessary to also check the Valid Start and Valid End Date for each record.
- An attempt is made to capture all source codes of a vocabulary and provide a comprehensive set. This cannot be guaranteed, since some vocabularies don't have an authoritative source (e.g. NDC).
- If a source code has no mapping to a Standard Vocabulary, the Target Concept ID as well as the vocabulary is set to 0.
- The Mapping Type indicates what meaning of a Concept is being mapped. For example, ICD-9-CM codes can be interpreted as a diagnosis (Mapping Type "CONDITION") or a procedure (Mapping Type "PROCEDURE"). Currently, the following Mapping Types are used:
    - ACTIVITY TYPE
    - CONDITION
    - DISCHARGE STATUS
    - DRUG
    - GENDER
    - OBSERVATION
    - PATIENT STATUS
    - PLACE OF SERVICE
    - PROCEDURE
    - PROCEDURE DRUG
    - PROVIDER SPECIALTY
    - RACE
    - RESULT CATEGORY
    - UOM (Unit of Measure)

- In cases where an unambiguous mapping cannot be established, the table contains two or more records to different Target Concept IDs. For example, ICD-9-CM 112.1 "Candidiasis of vulva and vagina" is mapped to Standard Vocabulary 198363 "Candidiasis of vagina" and 444106 "Candidiasis of vulva". However, only the one carrying the main meaning of the Source Code has the Primary Map flag set to "Y".

### 4.6.2 EXAMPLE OF A LOADED TABLE

The following are several examples of sample records in the Source to Concept mapping table.

1. Mapping of national standard identifiers to standard Concepts. In this case the mapping of ICD-9-CM (vocabulary 2) diagnosis code 140. 0 for “Malignant Neoplasm of Upper lip, vermilion border” is mapped to SNOMED-CT (vocabulary1) Concept 187602007 “Malignant neoplasm of upper lip, vermilion border NOS (disorder)” (Concept ID 4093013).

The Source to Concept mapping can be represented as follows:

| Source Vocabulary ID | Source Code | Target Concept ID | Target Vocabulary ID | Mapping Type | Primary Map |
|---|---|---|---|---|---|
| 2 | 140.0 | 4093013 | 1 | CONDITION | |

2. Mapping of source specific identifiers to standard Concepts. In this case a source specific GE – EHR (vocabulary 51) Observation key 5334 with description of “hemoglobin, blood” is being mapped to LOINC (vocabulary 6) Concept 718-7 with description of “Hemoglobin [Mass/volume] in Blood” (Concept ID: 3000963).

| Source Vocabulary ID | Source Code | Target Concept ID | Target Vocabulary ID | Mapping Type | Primary Map |
|---|---|---|---|---|---|
| 51 | 5334 | 3000963 | 6 | OBSERVATION | |

## 4.7 Relationship

Concept Relationships are defined through the Relationship table. The Relationship codes are adopted from the various source vocabularies the Standard Vocabulary is derived from. Some of the relationships are hierarchical and define ancestry (see above), and others are preserved for the convenience of the researcher from their original source:

- Hierarchical Relationships, which are used to define a hierarchical tree between Concepts. For example, "has_ingredient" is a Relationship between Clinical Drugs and Ingredients, and all Ingredients can be assumed as the "parental" hierarchical Concepts for the Drugs they are included in. All "isa" Relationships are hierarchical. Hierarchical Relationships can be between Concepts that are adopted from the same Vocabulary source or between Concepts adopted from difference Vocabulary sources. Only Hierarchical Relationships are used to build the Concept Ancestor relationships.
- Non-hierarchical Relationships are all remaining non- inclusive relationships, for example between Clinical Drugs and Branded Drugs. These Relationships are not utilized to create Ancestor relationships.

| Field | Required | Type Precision | Description |
|---|---|---|---|
| relationship_id | Yes | integer | The type of relationship captured by the relationship record. |
| relationship_name | Yes | string(256) | The text that describes the relationship type. |
| is_hierarchical | No | string(1) | Defines whether a relationship defines concepts into classes or hierarchies. Values are Y for hierarchical relationship or NULL if not. |
| defines_ancestry | No | string(1) | Defines whether a hierarchical relationship contributes to the concept_ancestor table. These are subsets of the hierarchical relationships. Valid values are Y or NULL. |

# 5 Glossary of Terms

| Term | Abbr. | Description |
|---|---|---|
| Ancestor | | The higher level Concept in a hierarchical relationship. Note that ancestors and descendants can be many levels apart from each other. |
| Ambulatory Payment Classification | APC | The Ambulatory Payment Classifications is used as a method of paying for outpatient services for the Medicare program, which is analogous to the DRGs for inpatient services. |
| Average Wholesale Price | AWP | The price manufacturers set for prescription drugs to be purchased at the wholesale level to pharmacies and healthcare provider. |
| Centers for Disease Control and Prevention | CDC | The Centers for Disease Control and Prevention are a federal agency under the Department of Health and Human Services. It works to protect public health and safety by providing information to enhance health decisions. |
| Common Data Model | CDM | The CDM intends to facilitate observational analyses of disparate healthcare databases. The CDM defines table structures for each of the data entities (e.g., Persons, Visit Occurrence, Drug Exposure, Condition Occurrence, Observation, Procedure-Occurrence, etc.). It includes all observational data elements that are relevant to identifying exposure to various treatments and defining condition occurrence. The CDM includes both the vocabulary of terms and the entity domain tables. |
| Concept | | A concept is the basic unit of information. Concepts may be grouped into a given domain. A concept is a unique term that has a unique and static identifier/name, belongs to a Namespace, and may exist in relation to other concepts. The vertical relationships consist of "is a" statements that form a logical hierarchy. In general, concepts above a given concept are referred to as ancestors and those below as descendants. |
| Conceptual Data Model | | A Conceptual Data Model is a map of concepts and their relationships. This describes the semantics of an organization and represents a series of assertions about its nature. Specifically, it describes the things of significance to an organization (entity classes), about which it is inclined to collect information, and characteristics of (attributes) and associations between pairs of those things of significance (relationships). |
| Condition | | A condition is an observation of a disease, such as diagnosis of a heart condition. |
| Condition Era (domain) | | A Condition Era entity consist of individual records of a Condition Occurrences that serve as indicators for the presence of a Person's Condition, and are stored in the Condition Era table. Combining individual Condition Occurrences into a single Condition Era serves two purposes: 1) aggregation of chronic conditions that require continuous ongoing care that refers to the same underlying illness; and 2) aggregation of multiple, closely timed events whether either condition is acute or chronic. |

| Term | Abbr. | Description |
|---|---|---|
| Condition Occurrence (domain) | | Condition Occurrences record individual instances of a Person's Conditions (i.e., diagnoses) extracted from source data. Conditions are recorded in various data sources in different forms with varying levels of standardization, and are stored in the Condition Occurrence table. |
| Current Procedural Terminology, 4th edition | CPT-4 | A terminology that is maintained by the American Medical Association (AMA). It is used by hospitals for Medicare hospital outpatient and by physician for outpatient services. |
| Data mapping | | It is the data element mappings between two distinct data models, terminologies, or concepts. Data mapping is the process of creating data element mappings between two distinct data models. Data mapping is used as a first step for a wide variety of data integration tasks. |
| Demographics | | Demographics refer to selected population characteristics. Demographics may include data such as race, age, sex, date of birth, location, etc. |
| Descendant | | The lower level Concept in a hierarchical relationship. Note that ancestors and descendants can be many levels apart from each other. |
| Design Principle | | An organized arrangement of one or more elements or principles for a purpose. It identifies core principles and best practices to assist developers to produce software. Thoroughly understanding the goals of stakeholders and designing systems with those goals in mind are the best approaches to successfully deliver results. |
| Diagnosis-Related Group | DRG | The Diagnosis-related Groups is used as a method of paying for inpatient services for the Medicare program, which is analogous to the APCs for outpatient services. |
| Domain | | A data domain refers to all the unique values which a data element may contain. For example, a database table that has information about people, with one record per person, might have a "gender" column. This gender column might be declared as a string data type, and allowed to have one of two known code values: "M" for male, "F" for female -- and NULL for records where gender is unknown or not applicable (or arguably "U" for unknown as a sentinel value). The data domain for the gender column is: "M", "F".<br><br>In database technology, domain refers to the description of an attribute's allowed values. The physical description is a set of values the attribute can have, and the semantic, or logical, description is the meaning of the attribute. |
| Drug | | In pharmacology, a drug as "a chemical substance used in the treatment, cure, prevention, or diagnosis of disease or used to otherwise enhance physical or mental well-being." Drugs may be prescribed for a limited duration, or on a regular basis for chronic disorders. |

| Term | Abbr. | Description |
|---|---|---|
| Electronic Health Record | EHR | Electronic health record refers to an individual person's medical record in digital format. It may be made up of electronic medical records from many locations and/or sources. The EHR is a longitudinal electronic record of person health information generated by one or more encounters in any care delivery setting. Included in this information are person demographics, progress notes, problems, medications, vital signs, past medical history, immunizations, laboratory data and radiology reports. The EHR has the ability to generate a complete record of a clinical person encounter - as well as supporting other care-related activities directly or indirectly via interface - including evidence-based decision support, quality management, and outcomes reporting. |
| Electronic Medical Record | EMR | An electronic medical record is a computerized legal medical record created in an organization that delivers care, such as a hospital or outpatient setting. Electronic medical records tend to be a part of a local stand-alone health information system that allows storage, retrieval and manipulation of records. This document will reference EHR moving forward even if specific data source might internally use EMR definition. |
| Extract Transform Load | ETL | Process of getting data out of one data store (Extract), modifying it (Transform), and inserting it into a different data store (Load). |
| Generic Product Information | GPI | A proprietary unique identifier for a drug used by the commercial Medi-Span® formulary database. |
| Healthcare Common Procedure Coding System | HCPCS | HCPCS Level I codes are managed by the AMA (licensing fees apply). The HCPCS Level II codes are managed by CMS (Centers for Medicare & Medicaid Services). The Level II codes includes: alphanumeric HCPCS procedure and modifier codes, their long and short descriptions, and applicable Medicare administrative, coverage, and pricing data. These codes are used for Medicare outpatient services. |
| Health Insurance claims | | An insurance claim is the actual application for benefits provided by an insurance company. Policyholders must first file an insurance claim before any money can be disbursed. Computerized health insurance claims databases are maintained largely for billing and administrative purposes. Unlike studies with primary data collection, claims data are not collected to meet specific research objectives. Nevertheless, these databases are useful for describing health care utilization, patterns of care, disease prevalence, drug and disease outcomes, and cost of care. |
| Health Insurance Portability and Accountability Act | HIPAA | A federal law that was designed to allow portability of health insurance between jobs. In addition, it required the creation of a federal law to protect personally identifiable health information; if that did not occur by a specific date (which it did not), HIPAA directed the Department of Health and Human Services (DHHS) to issue federal regulations with the same purpose. DHHS has issued HIPAA privacy regulations (the HIPAA Privacy Rule) as well as other regulations under HIPAA. |

| Term | Abbr. | Description |
|---|---|---|
| Health Level Seven | HL7 | HL7 is a global not-for-profit, ANSI-accredited standards developing organization dedicated to providing a comprehensive framework and related standards for the exchange, integration, sharing, and retrieval of electronic health information that supports clinical practice and the management, delivery and evaluation of health services. HL7 specifications primarily draw upon codes and vocabularies from a variety of sources. |
| Health Outcomes of Interest | HOI | May be defined by clinical events (e.g., drugs, conditions, observations, procedures, etc.) in predefined temporal relationships. |
| International Classification of Disease, 9th Revision, Clinical Modifications | ICD-9-CM | The official system of assigning codes to diagnoses and procedures associated with hospital utilization in the United States. |
| Logical Data Model | | Logical data models are graphical representation of the business requirements. They describe the things of importance to an organization and how they relate to one another, as well as business definitions and examples. The logical data model can be validated and approved by a business representative, and can be the basis of physical database design. |
| Logical Observation Identifiers Names and Codes | LOINC | Universal code names and identifiers to medical terminology related to the Electronic Health Record and assists in the electronic exchange and gathering of clinical results (such as laboratory tests, clinical observations, outcomes management and research). |
| Medical Dictionary for Regulatory Activities | MedDRA | MedDRA is a clinically-validated international medical terminology used by regulatory authorities and the regulated biopharmaceutical industry. The terminology is used through the entire regulatory process, from pre-marketing to post-marketing, and for data entry, retrieval, evaluation, and presentation. |
| National Drug Codes | NDC | Unique identifiers assigned to individual drugs. NDCs are used primarily as an inventory code and for prescriptions. |
| National Drug File - Reference Terminology | NDF-RT | A nonproprietary drug reference terminology that includes drug knowledge and classifies drugs, most notably by mechanism of action and physiologic effect. |
| Observation | | An observation represents a conclusion reached after examination or investigation (i.e., something that has been found). It may be delivered as a statement or document containing an authoritative decision or conclusion. |
| Observational Medical Data Simulator | OSIM | A program that constructs CDM simulation dataset that mimic real observational data sources. The simulated datasets can be used to perform statistical evaluations of the analytical methods offered to identify drug-outcome associations. |

| Term | Abbr. | Description |
|---|---|---|
| Observational Medical Outcomes Partnership | OMOP | A public-private partnership designed to protect human health by improving the monitoring of drugs for safety and effectiveness. |
| Primary Care Physician | PCP | A physician designated as responsible to provide specific care to a patient, including evaluation and treatment as well as referral to specialists. |
| Protected Health Information | PHI | Protected health information under HIPAA includes any individually identifiable health information. Identifiable refers not only to data that is explicitly linked to a particular individual (that's identified information). It also includes health information with data items which reasonably could be expected to allow individual identification. De-indentified information is that from which all potentially identifying information has been removed. |
| RxNorm | | A standardized nomenclature for clinical drugs and drug delivery devices is produced by the National Library of Medicine. In RxNorm, the name of a clinical drug combines its ingredients, strengths, and/or form.<br><br>RxNorm provides normalized names for clinical drugs and links its names to many of the drug vocabularies commonly used in pharmacy management and drug interaction software, including those of First DataBank, Micromedix, MediSpan, Gold Standard Alchemy, and Multum. By providing links between these vocabularies, RxNorm can mediate messages between systems not using the same software and vocabulary. |
| Systematized Nomenclature of Medicine - Clinical Terms | SNOMED-CT® | SNOMED-CT is one of a suite of designated standards for use in U.S. Federal Government systems for the electronic exchange of clinical health information, and is also a required standard in interoperability specifications of the U.S. Healthcare Information Technology Standards Panel. SNOMED-CT is also being implemented internationally as a standard within other IHTSDO Member countries. |
| Terminology | | Technical or special terms used in a business or special subject area. |
| Vocabulary | | A computerized list (as of items of data or words) used for reference (as for information retrieval or word processing). |