**Data Modeling Assignment**

Expected Time: 4 hours

**Background**: A key step in wrangling data is to understand the structure and semantics of the data. Data Modeling is a formal process performed to structure and assemble data. Understanding the underlying model(s) is therefore important for utilizing individual or multiple data sources, and their aggregation and integration. One approach for representing data model is by using Entity Relationship (ER) Diagrams (Review the Entity Relationship Diagramming material provided with this module).

In this assignment, you will use the description below to create ER diagrams. You can use any ER diagramming tool (E.g. Lucidchart, Lucidspark, Enterprise Architect, PowerPoint, Word) for this assignment. Here is a list of some tools: <https://trevor.io/blog/top-7-entity-relationship-diagram-tools/>

**Deliverable:** You will combine the three parts of the deliverable into one document (word or PDF)

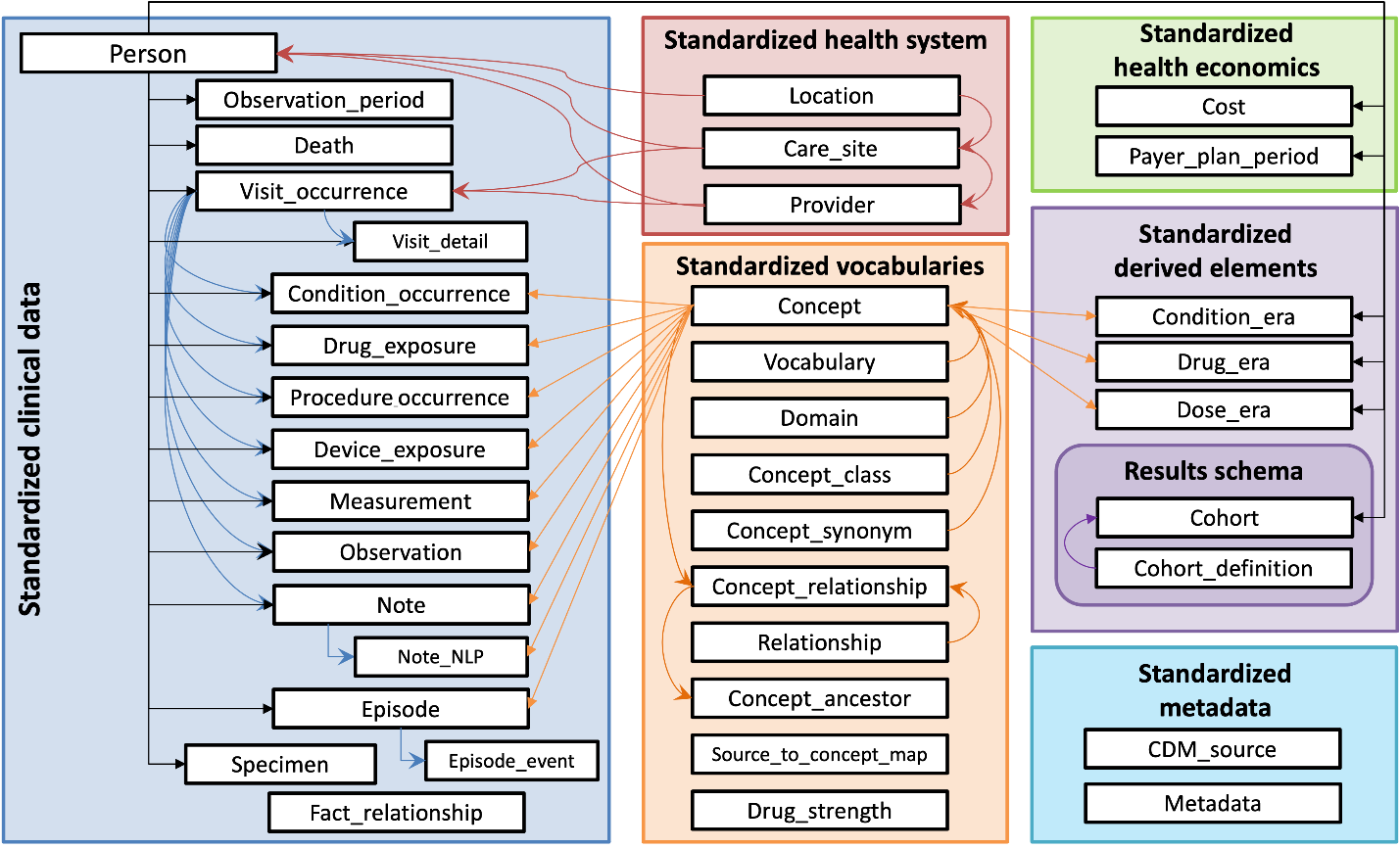
**Part 1:** Mention the name of the tool you use of the ER diagramming. If it is a web based tool provide a link.

**Part 2:** Use the PERSON table information in the below description and create a ER diagram of the Person Entity. Combine the generated diagram into a single document along with Part 1 and 3.

**Part 3:** Use the PERSON, VISIT\_OCCURRENCE, and either the CONDITION\_OCCURRENCE or the DRUG\_EXPOSURE table information to generate a conceptual/relational model that combines the three tables as entities with attributes and their relationship along with any ancillary tables (you don’t need to represent these tables in detail). Combine the generated diagram into a single document along with Part 1 and 2.

**OMOP Common Data Model**

The OMOP common data model (CDM) has become the de facto standard for aggregating clinical data for performing observation health studies [1]. Refer to <http://ohdsi.github.io/CommonDataModel/index.html> for a detailed information on OMOP. The below figure provides a summary of all tables in OMOP CDM. In this assignment you will be using only information for the PERSON, VISIT\_OCCURRENCE, and of from either CONDITION\_OCCURRENCE or DRUG\_EXPOSURE tables [2].



## ****Clinical Data Tables****

### PERSON

**Table Description**

This table serves as the central identity management for all Persons in the database. It contains records that uniquely identify each person or patient, and some demographic information.

**User Guide**

All records in this table are independent Persons.

**ETL Conventions**

All Persons in a database needs one record in this table, unless they fail data quality requirements specified in the ETL. Persons with no Events should have a record nonetheless. If more than one data source contributes Events to the database, Persons must be reconciled, if possible, across the sources to create one single record per Person. The content of the BIRTH\_DATETIME must be equivalent to the content of BIRTH\_DAY, BIRTH\_MONTH and BIRTH\_YEAR.

| **CDM Field** | **User Guide** | **ETL Conventions** | **Datatype** | **Required** | **Primary Key** | **Foreign Key** | **FK Table** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **person\_id** | It is assumed that every person with a different unique identifier is in fact a different person and should be treated independently. | Any person linkage that needs to occur to uniquely identify Persons ought to be done prior to writing this table. This identifier can be the original id from the source data provided if it is an integer, otherwise it can be an autogenerated number. | integer | Yes | Yes | No |  |
| **gender\_concept\_id** | This field is meant to capture the biological sex at birth of the Person. This field should not be used to study gender identity issues. | Use the gender or sex value present in the data under the assumption that it is the biological sex at birth. If the source data captures gender identity it should be stored in the [OBSERVATION](https://ohdsi.github.io/CommonDataModel/cdm531.html#observation) table. [Accepted gender concepts](http://athena.ohdsi.org/search-terms/terms?domain=Gender&standardConcept=Standard&page=1&pageSize=15&query=) | integer | Yes | No | Yes | CONCEPT |
| **year\_of\_birth** | Compute age using year\_of\_birth. | For data sources with date of birth, the year should be extracted. For data sources where the year of birth is not available, the approximate year of birth could be derived based on age group categorization, if available. | integer | Yes | No | No |  |
| **month\_of\_birth** |  | For data sources that provide the precise date of birth, the month should be extracted and stored in this field. | integer | No | No | No |  |
| **day\_of\_birth** |  | For data sources that provide the precise date of birth, the day should be extracted and stored in this field. | integer | No | No | No |  |
| **birth\_datetime** |  | This field is not required but highly encouraged. For data sources that provide the precise datetime of birth, that value should be stored in this field. If birth\_datetime is not provided in the source, use the following logic to infer the date: If day\_of\_birth is null and month\_of\_birth is not null then use the first of the month in that year. If month\_of\_birth is null or if day\_of\_birth AND month\_of\_birth are both null and the person has records during their year of birth then use the date of the earliest record, otherwise use the 15th of June of that year. If time of birth is not given use midnight (00:00:0000). | datetime | No | No | No |  |
| **race\_concept\_id** | This field captures race or ethnic background of the person. | Only use this field if you have information about race or ethnic background. The Vocabulary contains Concepts about the main races and ethnic backgrounds in a hierarchical system. Due to the imprecise nature of human races and ethnic backgrounds, this is not a perfect system. Mixed races are not supported. If a clear race or ethnic background cannot be established, use Concept\_Id 0. [Accepted Race Concepts](http://athena.ohdsi.org/search-terms/terms?domain=Race&standardConcept=Standard&page=1&pageSize=15&query=). | integer | Yes | No | Yes | CONCEPT |
| **ethnicity\_concept\_id** | This field captures Ethnicity as defined by the Office of Management and Budget (OMB) of the US Government: it distinguishes only between “Hispanic” and “Not Hispanic”. Races and ethnic backgrounds are not stored here. | Only use this field if you have US-based data and a source of this information. Do not attempt to infer Ethnicity from the race or ethnic background of the Person. [Accepted ethnicity concepts](http://athena.ohdsi.org/search-terms/terms?domain=Ethnicity&standardConcept=Standard&page=1&pageSize=15&query=) | integer | Yes | No | Yes | CONCEPT |
| **location\_id** | The location refers to the physical address of the person. This field should capture the last known location of the person. | Put the location\_id from the [LOCATION](https://ohdsi.github.io/CommonDataModel/cdm531.html#location) table here that represents the most granular location information for the person. This could represent anything from postal code or parts thereof, state, or county for example. Since many databases contain deidentified data, it is common that the precision of the location is reduced to prevent re-identification. This field should capture the last known location. | integer | No | No | Yes | LOCATION |
| **provider\_id** | The Provider refers to the last known primary care provider (General Practitioner). | Put the provider\_id from the [PROVIDER](https://ohdsi.github.io/CommonDataModel/cdm531.html#provider) table of the last known general practitioner of the person. If there are multiple providers, it is up to the ETL to decide which to put here. | integer | No | No | Yes | PROVIDER |
| **care\_site\_id** | The Care Site refers to where the Provider typically provides the primary care. |  | integer | No | No | Yes | CARE\_SITE |
| **person\_source\_value** | Use this field to link back to persons in the source data. This is typically used for error checking of ETL logic. | Some use cases require the ability to link back to persons in the source data. This field allows for the storing of the person value as it appears in the source. This field is not required but strongly recommended. | varchar(50) | No | No | No |  |
| **gender\_source\_value** | This field is used to store the biological sex of the person from the source data. It is not intended for use in standard analytics but for reference only. | Put the biological sex of the person as it appears in the source data. | varchar(50) | No | No | No |  |
| **gender\_source\_concept\_id** | Due to the small number of options, this tends to be zero. | If the source data codes biological sex in a non-standard vocabulary, store the concept\_id here. | integer | No | No | Yes | CONCEPT |
| **race\_source\_value** | This field is used to store the race of the person from the source data. It is not intended for use in standard analytics but for reference only. | Put the race of the person as it appears in the source data. | varchar(50) | No | No | No |  |
| **race\_source\_concept\_id** | Due to the small number of options, this tends to be zero. | If the source data codes race in an OMOP supported vocabulary store the concept\_id here. | integer | No | No | Yes | CONCEPT |
| **ethnicity\_source\_value** | This field is used to store the ethnicity of the person from the source data. It is not intended for use in standard analytics but for reference only. | If the person has an ethnicity other than the OMB standard of “Hispanic” or “Not Hispanic” store that value from the source data here. | varchar(50) | No | No | No |  |
| **ethnicity\_source\_concept\_id** | Due to the small number of options, this tends to be zero. | If the source data codes ethnicity in an OMOP supported vocabulary, store the concept\_id here. | integer | No | No | Yes | CONCEPT |

### VISIT\_OCCURRENCE

**Table Description**

This table contains Events where Persons engage with the healthcare system for a duration of time. They are often also called “Encounters”. Visits are defined by a configuration of circumstances under which they occur, such as (i) whether the patient comes to a healthcare institution, the other way around, or the interaction is remote, (ii) whether and what kind of trained medical staff is delivering the service during the Visit, and (iii) whether the Visit is transient or for a longer period involving a stay in bed.

**User Guide**

The configuration defining the Visit are described by Concepts in the Visit Domain, which form a hierarchical structure, but rolling up to generally familiar Visits adopted in most healthcare systems worldwide:

* [Inpatient Visit](https://athena.ohdsi.org/search-terms/terms/9201): Person visiting hospital, at a Care Site, in bed, for duration of more than one day, with physicians and other Providers permanently available to deliver service around the clock
* [Emergency Room Visit](https://athena.ohdsi.org/search-terms/terms/9203): Person visiting dedicated healthcare institution for treating emergencies, at a Care Site, within one day, with physicians and Providers permanently available to deliver service around the clock
* [Emergency Room and Inpatient Visit](https://athena.ohdsi.org/search-terms/terms/262): Person visiting ER followed by a subsequent Inpatient Visit, where Emergency department is part of hospital, and transition from the ER to other hospital departments is undefined
* [Non-hospital institution Visit](https://athena.ohdsi.org/search-terms/terms/42898160): Person visiting dedicated institution for reasons of poor health, at a Care Site, long-term or permanently, with no physician but possibly other Providers permanently available to deliver service around the clock
* [Outpatient Visit](https://athena.ohdsi.org/search-terms/terms/9202): Person visiting dedicated ambulatory healthcare institution, at a Care Site, within one day, without bed, with physicians or medical Providers delivering service during Visit
* [Home Visit](https://athena.ohdsi.org/search-terms/terms/581476): Provider visiting Person, without a Care Site, within one day, delivering service
* [Telehealth Visit](https://athena.ohdsi.org/search-terms/terms/5083): Patient engages with Provider through communication media
* [Pharmacy Visit](https://athena.ohdsi.org/search-terms/terms/581458): Person visiting pharmacy for dispensing of Drug, at a Care Site, within one day
* [Laboratory Visit](https://athena.ohdsi.org/search-terms/terms/32036): Patient visiting dedicated institution, at a Care Site, within one day, for the purpose of a Measurement.
* [Ambulance Visit](https://athena.ohdsi.org/search-terms/terms/581478): Person using transportation service for the purpose of initiating one of the other Visits, without a Care Site, within one day, potentially with Providers accompanying the Visit and delivering service
* [Case Management Visit](https://athena.ohdsi.org/search-terms/terms/38004193): Person interacting with healthcare system, without a Care Site, within a day, with no Providers involved, for administrative purposes

The Visit duration, or ‘length of stay’, is defined as VISIT\_END\_DATE - VISIT\_START\_DATE. For all Visits this is <1 day, except Inpatient Visits and Non-hospital institution Visits. The CDM also contains the VISIT\_DETAIL table where additional information about the Visit is stored, for example, transfers between units during an inpatient Visit.

**ETL Conventions**

Visits can be derived easily if the source data contain coding systems for Place of Service or Procedures, like CPT codes for well visits. In those cases, the codes can be looked up and mapped to a Standard Visit Concept. Otherwise, Visit Concepts have to be identified in the ETL process. This table will contain concepts in the Visit domain. These concepts are arranged in a hierarchical structure to facilitate cohort definitions by rolling up to generally familiar Visits adopted in most healthcare systems worldwide. Visits can be adjacent to each other, i.e. the end date of one can be identical with the start date of the other. As a consequence, more than one-day Visits or their descendants can be recorded for the same day. Multi-day visits must not overlap, i.e. share days other than start and end days. It is often the case that some logic should be written for how to define visits and how to assign Visit\_Concept\_Id. For example, in US claims outpatient visits that appear to occur within the time period of an inpatient visit can be rolled into one with the same Visit\_Occurrence\_Id. In EHR data inpatient visits that are within one day of each other may be strung together to create one visit. It will all depend on the source data and how encounter records should be translated to visit occurrences. Providers can be associated with a Visit through the PROVIDER\_ID field, or indirectly through PROCEDURE\_OCCURRENCE records linked both to the VISIT and PROVIDER tables.

| **CDM Field** | **User Guide** | **ETL Conventions** | **Datatype** | **Required** | **Primary Key** | **Foreign Key** | **FK Table** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **visit\_occurrence\_id** | Use this to identify unique interactions between a person and the health care system. This identifier links across the other CDM event tables to associate events with a visit. | This should be populated by creating a unique identifier for each unique interaction between a person and the healthcare system where the person receives a medical good or service over a span of time. | integer | Yes | Yes | No |  |
| **person\_id** |  |  | integer | Yes | No | Yes | PERSON |
| **visit\_concept\_id** | This field contains a concept id representing the kind of visit, like inpatient or outpatient. All concepts in this field should be standard and belong to the Visit domain. | Populate this field based on the kind of visit that took place for the person. For example this could be “Inpatient Visit”, “Outpatient Visit”, “Ambulatory Visit”, etc. This table will contain standard concepts in the Visit domain. These concepts are arranged in a hierarchical structure to facilitate cohort definitions by rolling up to generally familiar Visits adopted in most healthcare systems worldwide. [Accepted Concepts](https://athena.ohdsi.org/search-terms/terms?domain=Visit&standardConcept=Standard&page=1&pageSize=15&query=). | integer | Yes | No | Yes | CONCEPT |
| **visit\_start\_date** | For inpatient visits, the start date is typically the admission date. For outpatient visits the start date and end date will be the same. | When populating VISIT\_START\_DATE, you should think about the patient experience to make decisions on how to define visits. In the case of an inpatient visit this should be the date the patient was admitted to the hospital or institution. In all other cases this should be the date of the patient-provider interaction. | date | Yes | No | No |  |
| **visit\_start\_datetime** |  | If no time is given for the start date of a visit, set it to midnight (00:00:0000). | datetime | No | No | No |  |
| **visit\_end\_date** | For inpatient visits the end date is typically the discharge date. If a Person is still an inpatient in the hospital at the time of the data extract and does not have a visit\_end\_date, then set the visit\_end\_date to the date of the data pull. | Visit end dates are mandatory. If end dates are not provided in the source there are three ways in which to derive them: - Outpatient Visit: visit\_end\_datetime = visit\_start\_datetime - Emergency Room Visit: visit\_end\_datetime = visit\_start\_datetime - Inpatient Visit: Usually there is information about discharge. If not, you should be able to derive the end date from the sudden decline of activity or from the absence of inpatient procedures/drugs. - Non-hospital institution Visits: Particularly for claims data, if end dates are not provided assume the visit is for the duration of month that it occurs. For Inpatient Visits ongoing at the date of ETL, put date of processing the data into visit\_end\_datetime and visit\_type\_concept\_id with 32220 “Still patient” to identify the visit as incomplete. - All other Visits: visit\_end\_datetime = visit\_start\_datetime. If this is a one-day visit the end date should match the start date. | date | Yes | No | No |  |
| **visit\_end\_datetime** | If a Person is still an inpatient in the hospital at the time of the data extract and does not have a visit\_end\_datetime, then set the visit\_end\_datetime to the datetime of the data pull. | If no time is given for the end date of a visit, set it to midnight (00:00:0000). | datetime | No | No | No |  |
| **visit\_type\_concept\_id** | Use this field to understand the provenance of the visit record, or where the record comes from. | Populate this field based on the provenance of the visit record, as in whether it came from an EHR record or billing claim. [Accepted Concepts](https://athena.ohdsi.org/search-terms/terms?domain=Type+Concept&standardConcept=Standard&page=1&pageSize=15&query=). A more detailed explanation of each Type Concept can be found on the [vocabulary wiki](https://github.com/OHDSI/Vocabulary-v5.0/wiki/Vocab.-TYPE_CONCEPT). | Integer | Yes | No | Yes | CONCEPT |
| **provider\_id** | There will only be one provider per visit record and the ETL document should clearly state how they were chosen (attending, admitting, etc.). If there are multiple providers associated with a visit in the source, this can be reflected in the event tables (CONDITION\_OCCURRENCE, PROCEDURE\_OCCURRENCE, etc.) or in the VISIT\_DETAIL table. | If there are multiple providers associated with a visit, you will need to choose which one to put here. The additional providers can be stored in the [VISIT\_DETAIL](https://ohdsi.github.io/CommonDataModel/cdm531.html#visit_detail) table. | integer | No | No | Yes | PROVIDER |
| **care\_site\_id** | This field provides information about the Care Site where the Visit took place. | There should only be one Care Site associated with a Visit. | integer | No | No | Yes | CARE\_SITE |
| **visit\_source\_value** | This field houses the verbatim value from the source data representing the kind of visit that took place (inpatient, outpatient, emergency, etc.) | If there is information about the kind of visit in the source data that value should be stored here. If a visit is an amalgamation of visits from the source then use a hierarchy to choose the visit source value, such as IP -> ER-> OP. This should line up with the logic chosen to determine how visits are created. | varchar(50) | No | No | No |  |
| **visit\_source\_concept\_id** |  | If the visit source value is coded in the source data using an OMOP supported vocabulary put the concept id representing the source value here. | integer | No | No | Yes | CONCEPT |
| **admitted\_from\_concept\_id** | Use this field to determine where the patient was admitted from. This concept is part of the visit domain and can indicate if a patient was admitted to the hospital from a long-term care facility, for example. | If available, map the admitted\_from\_source\_value to a standard concept in the visit domain. [Accepted Concepts](https://athena.ohdsi.org/search-terms/terms?domain=Visit&standardConcept=Standard&page=1&pageSize=15&query=). If a person was admitted from home, set this to 0. | integer | No | No | Yes | CONCEPT |
| **admitted\_from\_source\_value** |  | This information may be called something different in the source data but the field is meant to contain a value indicating where a person was admitted from. Typically this applies only to visits that have a length of stay, like inpatient visits or long-term care visits. | varchar(50) | No | No | No |  |
| **discharged\_to\_concept\_id** | Use this field to determine where the patient was discharged to after a visit. This concept is part of the visit domain and can indicate if a patient was transferred to another hospital or sent to a long-term care facility, for example. It is assumed that a person is discharged to home therefore there is not a standard concept id for “home”. Use concept id = 0 when a person is discharged to home. | If available, map the discharged\_to\_source\_value to a standard concept in the visit domain. [Accepted Concepts](https://athena.ohdsi.org/search-terms/terms?domain=Visit&standardConcept=Standard&page=1&pageSize=15&query=). | integer | No | No | Yes | CONCEPT |
| **discharged\_to\_source\_value** |  | This information may be called something different in the source data but the field is meant to contain a value indicating where a person was discharged to after a visit, as in they went home or were moved to long-term care. Typically this applies only to visits that have a length of stay of a day or more. | varchar(50) | No | No | No |  |
| **preceding\_visit\_occurrence\_id** | Use this field to find the visit that occurred for the person prior to the given visit. There could be a few days or a few years in between. | This field can be used to link a visit immediately preceding the current visit. Note this is not symmetrical, and there is no such thing as a “following\_visit\_id”. | integer | No | No | Yes | VISIT\_OCCURRENCE |

### CONDITION\_OCCURRENCE

**Table Description**

This table contains records of Events of a Person suggesting the presence of a disease or medical condition stated as a diagnosis, a sign, or a symptom, which is either observed by a Provider or reported by the patient.

**User Guide**

Conditions are defined by Concepts from the Condition domain, which form a complex hierarchy. As a result, the same Person with the same disease may have multiple Condition records, which belong to the same hierarchical family. Most Condition records are mapped from diagnostic codes, but recorded signs, symptoms and summary descriptions also contribute to this table. Rule out diagnoses should not be recorded in this table, but in reality their negating nature is not always captured in the source data, and other precautions must be taken when when identifying Persons who should suffer from the recorded Condition. Record all conditions as they exist in the source data. Any decisions about diagnosis/phenotype definitions would be done through cohort specifications. These cohorts can be housed in the [COHORT](https://ohdsi.github.io/CommonDataModel/cdm531.html#payer_plan_period) table. Conditions span a time interval from start to end, but are typically recorded as single snapshot records with no end date. The reason is twofold: (i) At the time of the recording the duration is not known and later not recorded, and (ii) the Persons typically cease interacting with the healthcare system when they feel better, which leads to incomplete capture of resolved Conditions. The [CONDITION\_ERA](https://ohdsi.github.io/CommonDataModel/cdm531.html#condition_era) table addresses this issue. Family history and past diagnoses (‘history of’) are not recorded in this table. Instead, they are listed in the [OBSERVATION](https://ohdsi.github.io/CommonDataModel/cdm531.html#observation) table. Codes written in the process of establishing the diagnosis, such as ‘question of’ of and ‘rule out’, should not represented here. Instead, they should be recorded in the [OBSERVATION](https://ohdsi.github.io/CommonDataModel/cdm531.html#observation) table, if they are used for analyses. However, this information is not always available.

**ETL Conventions**

Source codes and source text fields mapped to Standard Concepts of the Condition Domain have to be recorded here.

| **CDM Field** | **User Guide** | **ETL Conventions** | **Datatype** | **Required** | **Primary Key** | **Foreign Key** | **FK Table** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **condition\_occurrence\_id** | The unique key given to a condition record for a person. Refer to the ETL for how duplicate conditions during the same visit were handled. | Each instance of a condition present in the source data should be assigned this unique key. In some cases, a person can have multiple records of the same condition within the same visit. It is valid to keep these duplicates and assign them individual, unique, CONDITION\_OCCURRENCE\_IDs, though it is up to the ETL how they should be handled. | integer | Yes | Yes | No |  |
| **person\_id** | The PERSON\_ID of the PERSON for whom the condition is recorded. |  | integer | Yes | No | Yes | PERSON |
| **condition\_concept\_id** | The CONDITION\_CONCEPT\_ID field is recommended for primary use in analyses, and must be used for network studies. This is the standard concept mapped from the source value which represents a condition | The CONCEPT\_ID that the CONDITION\_SOURCE\_VALUE maps to. Only records whose source values map to concepts with a domain of “Condition” should go in this table. [Accepted Concepts](https://athena.ohdsi.org/search-terms/terms?domain=Condition&standardConcept=Standard&page=1&pageSize=15&query=). | integer | Yes | No | Yes | CONCEPT |
| **condition\_start\_date** | Use this date to determine the start date of the condition | Most often data sources do not have the idea of a start date for a condition. Rather, if a source only has one date associated with a condition record it is acceptable to use that date for both the CONDITION\_START\_DATE and the CONDITION\_END\_DATE. | date | Yes | No | No |  |
| **condition\_start\_datetime** |  | If a source does not specify datetime the convention is to set the time to midnight (00:00:0000) | datetime | No | No | No |  |
| **condition\_end\_date** | Use this date to determine the end date of the condition | Most often data sources do not have the idea of a start date for a condition. Rather, if a source only has one date associated with a condition record it is acceptable to use that date for both the CONDITION\_START\_DATE and the CONDITION\_END\_DATE. | date | No | No | No |  |
| **condition\_end\_datetime** |  | If a source does not specify datetime the convention is to set the time to midnight (00:00:0000) | datetime | No | No | No |  |
| **condition\_type\_concept\_id** | This field can be used to determine the provenance of the Condition record, as in whether the condition was from an EHR system, insurance claim, registry, or other sources. | Choose the CONDITION\_TYPE\_CONCEPT\_ID that best represents the provenance of the record. [Accepted Concepts](https://athena.ohdsi.org/search-terms/terms?domain=Type+Concept&standardConcept=Standard&page=1&pageSize=15&query=). A more detailed explanation of each Type Concept can be found on the [vocabulary wiki](https://github.com/OHDSI/Vocabulary-v5.0/wiki/Vocab.-TYPE_CONCEPT). | integer | Yes | No | Yes | CONCEPT |
| **condition\_status\_concept\_id** | This concept represents the point during the visit the diagnosis was given (admitting diagnosis, final diagnosis), whether the diagnosis was determined due to laboratory findings, if the diagnosis was exclusionary, or if it was a preliminary diagnosis, among others. | Choose the Concept in the Condition Status domain that best represents the point during the visit when the diagnosis was given. These can include admitting diagnosis, principal diagnosis, and secondary diagnosis. [Accepted Concepts](https://athena.ohdsi.org/search-terms/terms?domain=Condition+Status&standardConcept=Standard&page=1&pageSize=15&query=). | integer | No | No | Yes | CONCEPT |
| **stop\_reason** | The Stop Reason indicates why a Condition is no longer valid with respect to the purpose within the source data. Note that a Stop Reason does not necessarily imply that the condition is no longer occurring. | This information is often not populated in source data and it is a valid etl choice to leave it blank if the information does not exist. | varchar(20) | No | No | No |  |
| **provider\_id** | The provider associated with condition record, e.g. the provider who made the diagnosis or the provider who recorded the symptom. | The ETL may need to make a choice as to which PROVIDER\_ID to put here. Based on what is available this may or may not be different than the provider associated with the overall VISIT\_OCCURRENCE record, for example the admitting vs attending physician on an EHR record. | integer | No | No | Yes | PROVIDER |
| **visit\_occurrence\_id** | The visit during which the condition occurred. | Depending on the structure of the source data, this may have to be determined based on dates. If a CONDITION\_START\_DATE occurs within the start and end date of a Visit it is a valid ETL choice to choose the VISIT\_OCCURRENCE\_ID from the Visit that subsumes it, even if not explicitly stated in the data. While not required, an attempt should be made to locate the VISIT\_OCCURRENCE\_ID of the CONDITION\_OCCURRENCE record. | integer | No | No | Yes | VISIT\_OCCURRENCE |
| **visit\_detail\_id** | The VISIT\_DETAIL record during which the condition occurred. For example, if the person was in the ICU at the time of the diagnosis the VISIT\_OCCURRENCE record would reflect the overall hospital stay and the VISIT\_DETAIL record would reflect the ICU stay during the hospital visit. | Same rules apply as for the VISIT\_OCCURRENCE\_ID. | integer | No | No | Yes | VISIT\_DETAIL |
| **condition\_source\_value** | This field houses the verbatim value from the source data representing the condition that occurred. For example, this could be an ICD10 or Read code. | This code is mapped to a Standard Condition Concept in the Standardized Vocabularies and the original code is stored here for reference. | varchar(50) | No | No | No |  |
| **condition\_source\_concept\_id** | This is the concept representing the condition source value and may not necessarily be standard. This field is discouraged from use in analysis because it is not required to contain Standard Concepts that are used across the OHDSI community, and should only be used when Standard Concepts do not adequately represent the source detail for the Condition necessary for a given analytic use case. Consider using CONDITION\_CONCEPT\_ID instead to enable standardized analytics that can be consistent across the network. | If the CONDITION\_SOURCE\_VALUE is coded in the source data using an OMOP supported vocabulary put the concept id representing the source value here. | integer | No | No | Yes | CONCEPT |
| **condition\_status\_source\_value** | This field houses the verbatim value from the source data representing the condition status. | This information may be called something different in the source data but the field is meant to contain a value indicating when and how a diagnosis was given to a patient. This source value is mapped to a standard concept which is stored in the CONDITION\_STATUS\_CONCEPT\_ID field. | varchar(50) | No | No | No |  |

### DRUG\_EXPOSURE

**Table Description**

This table captures records about the exposure to a Drug ingested or otherwise introduced into the body. A Drug is a biochemical substance formulated in such a way that when administered to a Person it will exert a certain biochemical effect on the metabolism. Drugs include prescription and over-the-counter medicines, vaccines, and large-molecule biologic therapies. Radiological devices ingested or applied locally do not count as Drugs.

**User Guide**

The purpose of records in this table is to indicate an exposure to a certain drug as best as possible. In this context a drug is defined as an active ingredient. Drug Exposures are defined by Concepts from the Drug domain, which form a complex hierarchy. As a result, one DRUG\_SOURCE\_CONCEPT\_ID may map to multiple standard concept ids if it is a combination product. Records in this table represent prescriptions written, prescriptions dispensed, and drugs administered by a provider to name a few. The DRUG\_TYPE\_CONCEPT\_ID can be used to find and filter on these types. This table includes additional information about the drug products, the quantity given, and route of administration.

**ETL Conventions**

Information about quantity and dose is provided in a variety of different ways and it is important for the ETL to provide as much information as possible from the data. Depending on the provenance of the data fields may be captured differently i.e. quantity for drugs administered may have a separate meaning from quantity for prescriptions dispensed. If a patient has multiple records on the same day for the same drug or procedures the ETL should not de-dupe them unless there is probable reason to believe the item is a true data duplicate. Take note on how to handle refills for prescriptions written.

| **CDM Field** | **User Guide** | **ETL Conventions** | **Datatype** | **Required** | **Primary Key** | **Foreign Key** | **FK Table** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **drug\_exposure\_id** | The unique key given to records of drug dispensings or administrations for a person. Refer to the ETL for how duplicate drugs during the same visit were handled. | Each instance of a drug dispensing or administration present in the source data should be assigned this unique key. In some cases, a person can have multiple records of the same drug within the same visit. It is valid to keep these duplicates and assign them individual, unique, DRUG\_EXPOSURE\_IDs, though it is up to the ETL how they should be handled. | integer | Yes | Yes | No |  |
| **person\_id** | The PERSON\_ID of the PERSON for whom the drug dispensing or administration is recorded. This may be a system generated code. |  | integer | Yes | No | Yes | PERSON |
| **drug\_concept\_id** | The DRUG\_CONCEPT\_ID field is recommended for primary use in analyses, and must be used for network studies. This is the standard concept mapped from the source concept id which represents a drug product or molecule otherwise introduced to the body. The drug concepts can have a varying degree of information about drug strength and dose. This information is relevant in the context of quantity and administration information in the subsequent fields plus strength information from the DRUG\_STRENGTH table, provided as part of the standard vocabulary download. | The CONCEPT\_ID that the DRUG\_SOURCE\_VALUE maps to. The concept id should be derived either from mapping from the source concept id or by picking the drug concept representing the most amount of detail you have. Records whose source values map to standard concepts with a domain of Drug should go in this table. When the Drug Source Value of the code cannot be translated into Standard Drug Concept IDs, a Drug exposure entry is stored with only the corresponding SOURCE\_CONCEPT\_ID and DRUG\_SOURCE\_VALUE and a DRUG\_CONCEPT\_ID of 0. The Drug Concept with the most detailed content of information is preferred during the mapping process. These are indicated in the CONCEPT\_CLASS\_ID field of the Concept and are recorded in the following order of precedence: ‘Branded Pack’, ‘Clinical Pack’, ‘Branded Drug’, ‘Clinical Drug’, ‘Branded Drug Component’, ‘Clinical Drug Component’, ‘Branded Drug Form’, ‘Clinical Drug Form’, and only if no other information is available ‘Ingredient’. Note: If only the drug class is known, the DRUG\_CONCEPT\_ID field should contain 0. [Accepted Concepts](https://athena.ohdsi.org/search-terms/terms?domain=Drug&standardConcept=Standard&page=1&pageSize=15&query=). | integer | Yes | No | Yes | CONCEPT |
| **drug\_exposure\_start\_date** | Use this date to determine the start date of the drug record. | Valid entries include a start date of a prescription, the date a prescription was filled, or the date on which a Drug administration was recorded. It is a valid ETL choice to use the date the drug was ordered as the DRUG\_EXPOSURE\_START\_DATE. | date | Yes | No | No |  |
| **drug\_exposure\_start\_datetime** |  | This is not required, though it is in v6. If a source does not specify datetime the convention is to set the time to midnight (00:00:0000) | datetime | No | No | No |  |
| **drug\_exposure\_end\_date** | The DRUG\_EXPOSURE\_END\_DATE denotes the day the drug exposure ended for the patient. | If this information is not explicitly available in the data, infer the end date using the following methods:  1. Start first with duration or days supply using the calculation drug start date + days supply -1 day. 2. Use quantity divided by daily dose that you may obtain from the sig or a source field (or assumed daily dose of 1) for solid, indivisibile, drug products. If quantity represents ingredient amount, quantity divided by daily dose \* concentration (from drug\_strength) drug concept id tells you the dose form. 3. If it is an administration record, set drug end date equal to drug start date. If the record is a written prescription then set end date to start date + 29. If the record is a mail-order prescription set end date to start date + 89. The end date must be equal to or greater than the start date. Ibuprofen 20mg/mL oral solution concept tells us this is oral solution. Calculate duration as quantity (200 example) \* daily dose (5mL) /concentration (20mg/mL) 200\*5/20 = 50 days. [Examples by dose form](https://ohdsi.github.io/CommonDataModel/drug_dose.html) | date | Yes | No | No |  |
| **drug\_exposure\_end\_datetime** |  | This is not required, though it is in v6. If a source does not specify datetime the convention is to set the time to midnight (00:00:0000) | datetime | No | No | No |  |
| **verbatim\_end\_date** | This is the end date of the drug exposure as it appears in the source data, if it is given | Put the end date or discontinuation date as it appears from the source data or leave blank if unavailable. | date | No | No | No |  |
| **drug\_type\_concept\_id** | You can use the TYPE\_CONCEPT\_ID to delineate between prescriptions written vs. prescriptions dispensed vs. medication history vs. patient-reported exposure, etc. | Choose the drug\_type\_concept\_id that best represents the provenance of the record, for example whether it came from a record of a prescription written or physician administered drug. [Accepted Concepts](https://athena.ohdsi.org/search-terms/terms?domain=Type+Concept&standardConcept=Standard&page=1&pageSize=15&query=). A more detailed explanation of each Type Concept can be found on the [vocabulary wiki](https://github.com/OHDSI/Vocabulary-v5.0/wiki/Vocab.-TYPE_CONCEPT). | integer | Yes | No | Yes | CONCEPT |
| **stop\_reason** | The reason a person stopped a medication as it is represented in the source. Reasons include regimen completed, changed, removed, etc. This field will be retired in v6.0. | This information is often not populated in source data and it is a valid etl choice to leave it blank if the information does not exist. | varchar(20) | No | No | No |  |
| **refills** | This is only filled in when the record is coming from a prescription written this field is meant to represent intended refills at time of the prescription. |  | integer | No | No | No |  |
| **quantity** |  | To find the dose form of a drug the RELATIONSHIP table can be used where the relationship\_id is ‘Has dose form’. If liquid, quantity stands for the total amount dispensed or ordered of ingredient in the units given by the drug\_strength table. If the unit from the source data does not align with the unit in the DRUG\_STRENGTH table the quantity should be converted to the correct unit given in DRUG\_STRENGTH. For clinical drugs with fixed dose forms (tablets etc.) the quantity is the number of units/tablets/capsules prescribed or dispensed (can be partial, but then only 1/2 or 1/3, not 0.01). Clinical drugs with divisible dose forms (injections) the quantity is the amount of ingredient the patient got. For example, if the injection is 2mg/mL but the patient got 80mL then quantity is reported as 160. Quantified clinical drugs with divisible dose forms (prefilled syringes), the quantity is the amount of ingredient similar to clinical drugs. Please see [how to calculate drug dose](https://ohdsi.github.io/CommonDataModel/drug_dose.html) for more information. | float | No | No | No |  |
| **days\_supply** |  | Days supply of the drug. This should be the verbatim days\_supply as given on the prescription. If the drug is physician administered use duration end date if given or set to 1 as default if duration is not available. | integer | No | No | No |  |
| **sig** | This is the verbatim instruction for the drug as written by the provider. | Put the written out instructions for the drug as it is verbatim in the source, if available. | varchar(MAX) | No | No | No |  |
| **route\_concept\_id** |  | The standard CONCEPT\_ID that the ROUTE\_SOURCE\_VALUE maps to in the route domain. | integer | No | No | Yes | CONCEPT |
| **lot\_number** |  |  | varchar(50) | No | No | No |  |
| **provider\_id** | The Provider associated with drug record, e.g. the provider who wrote the prescription or the provider who administered the drug. | The ETL may need to make a choice as to which PROVIDER\_ID to put here. Based on what is available this may or may not be different than the provider associated with the overall VISIT\_OCCURRENCE record, for example the ordering vs administering physician on an EHR record. | integer | No | No | Yes | PROVIDER |
| **visit\_occurrence\_id** | The Visit during which the drug was prescribed, administered or dispensed. | To populate this field drug exposures must be explicitly initiated in the visit. | integer | No | No | Yes | VISIT\_OCCURRENCE |
| **visit\_detail\_id** | The VISIT\_DETAIL record during which the drug exposure occurred. For example, if the person was in the ICU at the time of the drug administration the VISIT\_OCCURRENCE record would reflect the overall hospital stay and the VISIT\_DETAIL record would reflect the ICU stay during the hospital visit. | Same rules apply as for the VISIT\_OCCURRENCE\_ID. | integer | No | No | Yes | VISIT\_DETAIL |
| **drug\_source\_value** | This field houses the verbatim value from the source data representing the drug exposure that occurred. For example, this could be an NDC or Gemscript code. | This code is mapped to a Standard Drug Concept in the Standardized Vocabularies and the original code is stored here for reference. | varchar(50) | No | No | No |  |
| **drug\_source\_concept\_id** | This is the concept representing the drug source value and may not necessarily be standard. This field is discouraged from use in analysis because it is not required to contain Standard Concepts that are used across the OHDSI community, and should only be used when Standard Concepts do not adequately represent the source detail for the Drug necessary for a given analytic use case. Consider using DRUG\_CONCEPT\_ID instead to enable standardized analytics that can be consistent across the network. | If the DRUG\_SOURCE\_VALUE is coded in the source data using an OMOP supported vocabulary put the concept id representing the source value here. | integer | No | No | Yes | CONCEPT |
| **route\_source\_value** | This field houses the verbatim value from the source data representing the drug route. | This information may be called something different in the source data but the field is meant to contain a value indicating when and how a drug was given to a patient. This source value is mapped to a standard concept which is stored in the ROUTE\_CONCEPT\_ID field. | varchar(50) | No | No | No |  |
| **dose\_unit\_source\_value** | This field houses the verbatim value from the source data representing the dose unit of the drug given. | This information may be called something different in the source data but the field is meant to contain a value indicating the unit of dosage of drug given to the patient. **This is an older column and will be deprecated in an upcoming version.** | varchar(50) | No | No | No |  |

Reference

[1] <http://ohdsi.github.io/CommonDataModel/index.html>

[2] <http://ohdsi.github.io/CommonDataModel/cdm54.html#Clinical_Data_Tables>