**DECOVID Data Dictionary**

This document presents descriptions and important notes for key data tables that form the DECOVID database.

|  |
| --- |
| **CLINICAL DATA TABLES** |

**PERSON**

The PERSON table contains records that uniquely identify each patient in the source data who is time at-risk to have clinical observations recorded within the source systems. The PERSON table includes the following fields:

|  |  |
| --- | --- |
| **Field** | **Description** |
| person\_id | A unique identifier for each person for this database. Ideally serially generated numbers. |
| gender\_concept\_id | A foreign key that refers to an identifier in the CONCEPT table for the unique gender of the person. See Gender standard mappings below. |
| year\_of\_birth | 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 | Not used. |
| day\_of\_birth | Not used. |
| birth\_datetime | Not used. |
| race\_concept\_id | A foreign key that refers to an identifier in the CONCEPT table for the unique race of the person. |
| ethnicity\_concept\_id | Not used. |
| location\_id | Not used |
| provider\_id | Not used. |
| care\_site\_id | Not used. |
| person\_source\_value | Not used. |
| gender\_source\_value | Not used. |
| gender\_source\_concept\_id | Not used. |
| race\_source\_value | Not used. |
| race\_source\_concept\_id | Not used. |
| ethnicity\_source\_value | Not used. |
| ethnicity\_source\_concept\_id | Not used. |
| last\_updated\_datetime | Initially this is the time the row was written to the OMOP standardised database. If the row is subsequently updated, it then becomes the most recent update time. |
| deleted\_datetime | Null initially. Set to the time when the row was marked for deletion. A value in this column sent to the central data source will result in this row being deleted (by person\_id) from the combined dataset. Deletions only need to be sent once. |

**Notes:**

* The Trust can be derived by taking the last digit of the person\_id field, where 4 is for UHB and 6 is for UCLH. This applies to person\_id in any DECOVID data table.
* The ethnicity of patients is recorded in the race\_concept\_id column, according to the 18 ethnic groups used in the 2011 England and Wales Census [1].
* The sex at birth of patients is recorded in the gender\_concept\_id field. This field is mapped to standard mapping values of male, female or other/unknown, which includes cases where the sex at birth of a patient is withheld, or not asked/missing. Note, the name of this field is the naming convention used in PERSON table of the OMOP CDM, which is why it the name has not been revised to sex.
* In rare cases, a patient might arrive in hospital unconscious and without any form of identification. They might get added to the system with an estimated date of birth (often 01/01/1970 as a placeholder).
* The PERSON table only includes year\_of\_birth, so July 2 was assumed to be the month and day of birth, respectively, when calculating age.

**DEATH**

The DEATH table contains the clinical event for how and when a person dies, including both in-hospital and out-of-hospital deaths (from the NHS Spine). The DEATH table contains the follow fields:

|  |  |
| --- | --- |
| **Field** | **Description** |
| person\_id | A foreign key identifier to the deceased person. The demographic details of that person are stored in the PERSON table. |
| death\_date | The UTC date of death (e.g. possibly from the NHS Spine). |
| death\_datetime | Not used. |
| death\_type\_concept\_id | Not used. |
| cause\_concept\_id | Not used. |
| cause\_source\_value | Not used. |
| cause\_source\_concept\_id | Not used. |
| last\_updated\_datetime | Initially this is the time the row was written to the OMOP standardised database. If the row is subsequently updated, it then becomes the most recent update time. |
| deleted\_datetime | Null initially. Set to the time when the row was marked for deletion. A value in this column sent to the central data source will result in this row being deleted (by person\_id) from the combined dataset. Deletions only need to be sent once. |

**Notes:**

* Includes both in-hospital and out-of-hospital deaths (extracted from the NHS Spine up until 31st March 2021). Cause of death is not included, and date of death rather than the precise time of death is recorded.
* Cause of death is not currently captured in DECOVID, since it generally isn't well represented in structured health records.

**VISIT\_OCCURRENCE**

The VISIT\_OCCURRENCE table contains records on the spans of time describing a Person's individual episodes of care/visit. The VISIT\_OCCURRENCE table contains the follow fields:

|  |  |
| --- | --- |
| **Field** | **Description** |
| visit\_occurrence\_id | A unique identifier for each Person's visit or encounter at a healthcare provider. |
| person\_id | A foreign key identifier to the Person for whom the visit is recorded. The demographic details of that Person are stored in the PERSON table. |
| visit\_concept\_id | A foreign key that refers to a visit Concept identifier in the Standardized Vocabularies. See mappings below. |
| visit\_start\_date | The UTC start date of the visit. |
| visit\_start\_datetime | The UTC date and time of the admission. This should be the earliest presentation time at the hospital (potentially before getting assigned a bed), since all observations should be bounded by the start and end datetime. |
| visit\_end\_date | The UTC end date. Null means the patient has not been discharged. |
| visit\_end\_datetime | The UTC end date and time. Null means the patient has not been discharged. |
| visit\_type\_concept\_id | Not used. |
| provider\_id | Not used. |
| care\_site\_id | Not used. |
| visit\_source\_value | Visit / episode / contact identifier in the local site system. |
| visit\_source\_concept\_id | Not used. |
| admitting\_source\_concept\_id | A foreign key to the predefined concept in the Place of Service Vocabulary reflecting the admitting source for a visit. Same possible values as discharge\_to\_concept\_id. |
| admitting\_source\_value | Not used. |
| discharge\_to\_concept\_id | A foreign key to the predefined concept in the Place of Service Vocabulary reflecting the discharge disposition for a visit. See table below. Where a visit\_end\_datetime exists, NULL means that they left to an unknown location alive. |
| discharge\_to\_source\_value | Not used. |
| preceding\_visit\_occurrence\_id | A foreign key to the VISIT\_OCCURRENCE table of the visit immediately preceding this visit. |
| last\_updated\_datetime | Initially this is the time the row was written to the OMOP standardised database. If the row is subsequently updated, it then becomes the most recent update time. |
| deleted\_datetime | Null initially. Set to the time when the row was marked for deletion. A value in this column sent to the central data source will result in this row being deleted (by person\_id) from the combined dataset. Deletions only need to be sent once. |

**Notes:**

* The Trust can be derived by taking the last digit of the visit\_occurrence\_id, where 4 is for UHB and 6 is for UCLH. This applies to visit\_occurrence\_id in any DECOVID data table.

The admitting source and discharge to concepts are recorded according to the NHS Data Model and Dictionary (DM+D) [3, 4]. The following tables present the count of visit\_occurrence\_ids for each admitting source and discharge to concepts by hospital Trust (cells <10 have been suppressed to display “<10”). Note, UHB records include NULLs for admitting source and discharge to concepts, however, these should be “0” and correspond to “No matching concept”, which include records where a patient was admitted from, or discharged to, a source that was not recorded/missing or not part of NHS’s standard admission sources/discharge destinations [3, 4].

|  |  |  |  |
| --- | --- | --- | --- |
| **admitting\_source\_concept\_name** | **admitting\_source\_concept\_id** | **UCLH** | **UHB** |
| Adult Living Care Facility | 8882 | 42 | 904 |
| Assisted Living Facility | 8615 | 13 | <=10 |
| Home | 8536 | 45405 | 59924 |
| Hospice | 8546 | <=10 | <=10 |
| Inpatient Hospital | 8717 | 2306 | 3171 |
| Inpatient Psychiatric Facility | 8971 | 29 | 17 |
| No matching concept | 0 | 70440 | <=10 |
| Prison / Correctional Facility | 38003619 | 24 | <=10 |

|  |  |  |  |
| --- | --- | --- | --- |
| **discharge\_to\_concept\_name** | **discharge\_to\_concept\_id** | **UCLH** | **UHB** |
| Adult Living Care Facility | 8882 | 240 | 1688 |
| Assisted Living Facility | 8615 | 123 | 831 |
| Home | 8536 | 108176 | 115093 |
| Hospice | 8546 | 77 | 84 |
| Inpatient Hospital | 8717 | 3309 | 4118 |
| Inpatient Psychiatric Facility | 8971 | 420 | 137 |
| Intermediate Mental Care Facility | 8951 | 18 | 13 |
| No matching concept | 0 | 4499 | <=10 |
| Patient died | 4216643 | 1103 | 2649 |
| Prison / Correctional Facility | 38003619 | 296 | 186 |

**VISIT\_DETAIL**

The VISIT\_DETAILtable contains records on clinically meaningful movements of a patient within each record of the parent VISIT\_OCCURRENCE table. Each row represents a movement between geographically separate care sites (e.g. patient transferred from an Adult Inpatient Ward to Adult Inpatient Ward (geographically new ward) or between care sites within a hospital (e.g., patient transferred from A & E Majors to ICU within a hospital). The VISIT\_DETAIL table contains the following fields:

|  |  |
| --- | --- |
| **Field** | **Description** |
| visit\_detail\_id | A unique identifier for each Person's visit or encounter at a healthcare provider. |
| person\_id | 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\_detail\_concept\_id | A foreign key that refers to a visit Concept identifier in the Standardized Vocabularies. |
| visit\_detail\_start\_date | The UTC start date of the visit. |
| visit\_detail\_start\_datetime | The date and time of the visit started. |
| visit\_detail\_end\_date | The end date of the visit. If this is a one-day visit the end date should match the start date. NULL if the patient is still there. |
| visit\_detail\_end\_datetime | The UTC date and time of the visit end. NULL if the patient is still there. |
| visit\_detail\_type\_concept\_id | Not used. |
| provider\_id | Not used. |
| care\_site\_id | A foreign key to the care site in the CARE\_SITE table that was visited. |
| visit\_detail\_source\_value | The source code for the visit as it appears in the source data (e.g. a transfer identifier). |
| visit\_detail\_source\_concept\_id | Not used. |
| admitting\_source\_value | Not used. |
| admitting\_source\_concept\_id | Not used. |
| discharge\_to\_source\_value | Not used. |
| discharge\_to\_concept\_id | Not used. |
| preceding\_visit\_detail\_id | A foreign key to the VISIT\_DETAIL table of the visit immediately preceding this visit |
| visit\_detail\_parent\_id | Not used. |
| visit\_occurrence\_id | A foreign key that refers to the record in the VISIT\_OCCURRENCE table. This is a required field, because for every VISIT\_DETAIL is a child of VISIT\_OCCURRENCE and cannot exist without a corresponding parent record in VISIT\_OCCURRENCE. |
| last\_updated\_datetime | Initially this is the time the row was written to the OMOP standardised database. If the row is subsequently updated, it then becomes the most recent update time. |
| deleted\_datetime | Null initially. Set to the time when the row was marked for deletion. A value in this column sent to the central data source will result in this row being deleted (by person\_id) from the combined dataset. Deletions only need to be sent once. |

**Notes:**

* The Trust can be derived by taking the last digit of the visit\_detail\_id, where 4 is for UHB and 6 is for UCLH. This applies to visit\_detail\_id in any DECOVID data table; same applies to visit\_occurrence\_id in this table.

The following table presents the count of visit\_detail\_ids for each geographically separate care site within the hospital Trusts (note, patients can be counted more than once, as this summary is based on visit\_detail\_id). The care\_site\_id field in the VISIT\_DETAIL table allows for the analysis of clinically meaningful movements (i.e. between or within wards or departments of a hospital Trust) of a patient within each record of the parent VISIT\_OCCURRENCE table (cells <10 have been suppressed to display “<10”).

|  |  |  |  |
| --- | --- | --- | --- |
| **care\_site\_id** | **care\_site\_name** | **UCLH** | **UHB** |
| 1 | High Dependency Unit (Level 2) | <=10 | 2654 |
| 2 | Intensive Care Unit (Level 3) | 4495 | <=10 |
| 3 | Mixed Critical Care Unit (Level 2/3) | <=10 | 8088 |
| 4 | Theatre Recovery | <=10 | <=10 |
| 5 | Operating Theatre | 13851 | <=10 |
| 7 | Adult Inpatient Ward | 53824 | 85206 |
| 8 | Paediatric Inpatient Ward | 324 | <=10 |
| 9 | Accident & Emergency Department | 84823 | 63543 |
| 10 | A&E Majors | <=10 | 25928 |
| 11 | A&E Resus | <=10 | 8792 |
| 12 | Urgent Care (Minors) | <=10 | 18923 |
| 13 | Out Patient Department | 35 | 2089 |
| 14 | Endoscopy or other procedure room | 912 | 166 |
| 17 | In Patient Areas | 3541 | 16888 |
| 20 | Escalation High Dependency Unit (Level 2) | 2738 | 6812 |
| 21 | Escalation Intensive Care Unit (Level 3) | 321 | <=10 |

**CONDITION\_OCCURRENCE**

The CONDITION\_OCCURRENCEtable contains records on the presence of a disease or medical condition stated as a diagnosis, and a sign or a symptom, which is either observed by a provider or reported by the patient. The CONDITION\_OCCURRENCE table contains the follow fields:

|  |  |
| --- | --- |
| **Field** | **Description** |
| condition\_occurrence\_id | A unique identifier for each Condition Occurrence event. |
| person\_id | A foreign key identifier to the Person who is experiencing the condition. The demographic details of that Person are stored in the PERSON table. |
| condition\_concept\_id | A foreign key that refers to the Concept identifier representing the condition. |
| condition\_start\_date | The definition of the following date/datetime fields varies by condition type. Please also refer to Condition Type below. For Past medical history and Problem list condition types, this is the date corresponding to the condition\_start\_datetime. For Encounter diagnosis, this is the visit-level date associated with the diagnosis entry, either the appointment date or the admission date. Similarly for Chief complaint. For Billing diagnosis, this is the start date of the Consultant Episode the coded account is associated with. In all cases, the date should be UTC transformed. |
| condition\_start\_datetime | This field is only populated for conditions of type Past medical history or Problem list. In those cases, this represents the UTC transformed user entry datetime. The back-entered, physiologically relevant date was not used, even if it was available. |
| condition\_end\_date | For Past medical history and Problem list condition types, this is the UTC transformed date corresponding to the condition\_end\_datetime. For Encounter diagnosis, Billing diagnosis or Chief complaint, this should be left null as a visit level coding doesn't meaningfully end. |
| condition\_end\_datetime | This field is only populated for conditions of type Past medical history or Problem list. In those cases this represents either the resolution date and time or the deletion date and time (UTC transformed). The rationale between including deletion dates here is that users often use deletion to remove a problem from the problem list that has in fact been resolved. |
| condition\_type\_concept\_id | The Concept that represents the type of diagnosis this is (e.g billing, self-reported, problem list). See Condition Type below. |
| stop\_reason | Not used. |
| provider\_id | Not used. |
| visit\_occurrence\_id | This field is only populated for condition types where the diagnosis was entered in a manner that directly attributes it to a visit\_occurrence\_id in our subset (e.g. Billing diagnosis, Encounter diagnosis, Chief complaint). A foreign key to the visit in the VISIT\_OCCURRENCE table during which the Condition was determined (diagnosed). |
| visit\_detail\_id | Not used. Visit details in DECOVID relate to ward stays where a visit-level diagnoses such as Billing diagnoses related to Consultant Episodes. There is no simple mapping between the two |
| condition\_source\_value | Not used. |
| condition\_source\_concept\_id | Not used. |
| condition\_status\_source\_value | Not used. |
| condition\_status\_concept\_id | The status of the diagnosis. See condition\_status\_source\_value. |
| last\_updated\_datetime | Initially this is the time the row was written to the OMOP standardised database. If the row is subsequently updated, it then becomes the most recent update time. |
| deleted\_datetime | Null initially. Set to the time when the row was marked for deletion. A value in this column sent to the central data source will result in this row being deleted (by person\_id) from the combined dataset. Deletions only need to be sent once. |

The concepts in this table were mapped from vocabularies including diagnosis standards. The conditions in this table (i.e. condition\_concept\_id) are predominantly mapped from ICD-10 codes. The following table presents a tabulation of the count of condition\_occurrence\_ids, or data records, for each condition vocabulary by Trust (cells <10 have been suppressed to display “<10”):

| **Vocabulary** | **UCLH** | **UHB** |
| --- | --- | --- |
| CIEL | <=10 | <=10 |
| ICD10 | 768253 | 965308 |
| ICD10CM | <=10 | 10347 |
| ICD10CN | <=10 | 166 |
| KCD7 | <=10 | 7871 |
| Nebraska Lexicon | <=10 | <=10 |
| None | 14615 | <=10 |
| OXMIS | <=10 | 2836 |
| Read | <=10 | 582 |
| SMQ | <=10 | <=10 |
| SNOMED | 622142 | 244182 |

**Notes:**

* The Trust can be derived by taking the last digit of the visit\_detail\_id, person\_id and/or visit\_occurrence\_id where 4 is for UHB and 6 is for UCLH. Not all conditions have a related visit\_occurrence\_id/visit\_detail\_id.
* Condition records are typically inferred from diagnostic codes recorded in the source data. Such coding systems, like ICD-9-CM, ICD-10-CM, Read etc., provide a comprehensive coverage of conditions. However, if the diagnostic code in the source does not define a condition, but rather an observation or a procedure, then such information is not stored in the CONDITION\_OCCURRENCE table, but in the respective tables instead.

**MEASUREMENT**

The MEASUREMENTtable contains records of measurements, i.e. structured values (numerical or categorical) obtained about a Person (e.g. Body weight, height and temperature) or Person's sample (e.g C reactive protein [Mass/volume] in Serum or Plasma, and Bilirubin.total [Moles/volume] in Blood). The concepts of this table were primarily mapped from SNOMED and LOINC codes. There are 135 distinct clinical measurement types.

|  |  |
| --- | --- |
| **Field** | **Description** |
| measurement\_id | A unique identifier for each Measurement. |
| person\_id | A foreign key identifier to the Person about whom the measurement was recorded. The demographic details of that Person are stored in the PERSON table. |
| measurement\_concept\_id | A foreign key to the standard measurement concept identifier in the Standardized Vocabularies. |
| measurement\_date | The UTC date of the Measurement. |
| measurement\_datetime | The result / EPR entry UTC datetime of the Measurement |
| measurement\_time | Not used. |
| measurement\_type\_concept\_id | Not used. |
| operator\_concept\_id | A foreign key identifier to the predefined Concept in the Standardized Vocabularies reflecting the mathematical operator that is applied to the value\_as\_number. See operators. |
| value\_as\_number | A Measurement result, expressed as a numeric value. |
| value\_as\_concept\_id | A foreign key to a Measurement result represented as a Concept from the Standardized Vocabularies (e.g., positive/negative, present/absent, low/high, etc.). |
| unit\_concept\_id | A foreign key to a Standard Concept ID of Measurement Units in the Standardized Vocabularies. |
| range\_low | The lower limit of the normal range of the Measurement. Assumed to be of the same unit of measure as the Measurement value. |
| range\_high | The upper limit of the normal range of the Measurement. Assumed to be of the same unit of measure as the Measurement value. |
| provider\_id | Not used. |
| visit\_occurrence\_id | A foreign key to the Visit in the VISIT\_OCCURRENCE table during which the Measurement was recorded. |
| visit\_detail\_id | A foreign key to the Visit Detail in the VISIT\_DETAIL table during which the Measurement was recorded. |
| measurement\_source\_value | Not used. |
| measurement\_source\_concept\_id | Not used. |
| unit\_source\_value | Not used. |
| value\_source\_value | Not used. |
| last\_updated\_datetime | Initially this is the time the row was written to the OMOP standardised database. If the row is subsequently updated, it then becomes the most recent update time. |
| deleted\_datetime | Null initially. Set to the time when the row was marked for deletion. A value in this column sent to the central data source will result in this row being deleted (by person\_id) from the combined dataset. Deletions only need to be sent once. |

Notes:

* Duplicates of measurements sometimes exist in the measurement table, and can be identified by the same visit\_occurrence\_id, person\_id, measurement\_date(time), measurement\_concept\_id, value\_as\_number.
* The Trust can be derived by taking the last digit of the visit\_detail\_id, person\_id and/or visit\_occurrence\_id where 4 is for UHB and 6 is for UCLH. Not all measurements have a related visit\_occurrence\_id/visit\_detail\_id.
* Not all measurement types are recorded at both Trusts. Please refer to the concept\_id tabulations by Trust here: https://docs.google.com/spreadsheets/d/1CGOfL0RwUgSVF8LoACeAXpJLuS-JvQIm/edit?usp=sharing&ouid=113881954064126371771&rtpof=true&sd=true.

**SPECIMEN**

The SPECIMEN table contains records identifying biological samples from a person (e.g. Cerebrospinal fluid sample and Plasma specimen or serum specimen or whole blood specimen). This table has been extended in DECOVID to also include "physiological samples” for example, a panel of vital sign measurements collected from a Person.

| **Field** | **Description** |
| --- | --- |
| specimen\_id | A unique identifier for each specimen. |
| person\_id | A foreign key identifier to the Person for whom the Specimen is recorded. |
| specimen\_concept\_id | A foreign key referring to a Standard Concept identifier in the Standardized Vocabularies for the Specimen (arterial blood, urine, etc). |
| specimen\_type\_concept\_id | Not used. |
| specimen\_date | The UTC date the specimen was obtained from the Person. |
| specimen\_datetime | The UTC date and time when the Specimen was obtained from the person. |
| quantity | Not used. |
| unit\_concept\_id | Not used. |
| anatomic\_site\_concept\_id | A foreign key to a Standard Concept identifier for the anatomic location of specimen collection. |
| disease\_status\_concept\_id | Not used. |
| specimen\_source\_id | Not used. |
| specimen\_source\_value | Not used. |
| unit\_source\_value | Not used. |
| anatomic\_site\_source\_value | Not used. |
| disease\_status\_source\_value | Not used. |
| last\_updated\_datetime | Initially this is the time the row was written to the OMOP standardised database. If the row is subsequently updated, it then becomes the most recent update time. |
| deleted\_datetime | Null initially. Set to the time when the row was marked for deletion. A value in this column sent to the central data source will result in this row being deleted (by person\_id) from the combined dataset. Deletions only need to be sent once. |

**Notes:**

* Any sample (blood, sputum, urine, tissue etc.) is stored in this table with datetimes referencing the time the sample was taken.
* UHB has a large number of specimens with a NULL anatomic\_site\_source\_value.

**DRUG\_EXPOSURE**

The DRUG\_EXPOSUREtable contains records about the utilisation of a drug when ingested or otherwise introduced into the body. The concepts used in this table were primarily mapped from SNOMED codes.

| **Field** | **Description** |
| --- | --- |
| drug\_exposure\_id | A system-generated unique identifier for each Drug utilisation event. |
| person\_id | 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 | A foreign key that refers to a Standard Concept identifier in the Standardized Vocabularies for the Drug concept. Preferably at the RxNorm Ingredient level |
| drug\_exposure\_start\_date | The UTC Start date for the current instance of Drug utilisation. |
| drug\_exposure\_start\_datetime | The UTC start date and time for the current instance of Drug utilisation. |
| drug\_exposure\_end\_date | The UTC End date for the current instance of Drug utilisation. This is populated in the event that a datetime is not available (e.g. for home administrations) |
| drug\_exposure\_end\_datetime | The UTC end date and time for the current instance of Drug utilisation. It is NULL if the administration is not complete. |
| verbatim\_end\_date | Not used. |
| drug\_type\_concept\_id | A foreign key to the predefined Concept identifier of the "type concept" domain in the Standardized Vocabularies reflecting the type of Drug Exposure recorded. See the **Drug Type** tablebelow. |
| stop\_reason | Not used. |
| refills | Not used. |
| quantity | Records "dose" or "infusion rate" depending upon case use. |
| dose\_unit\_concept\_id | A standardised Athena mapping to the units that quality data in the "quantity" field. This information is used to differentiate "dose" from "rate". Note: this field has been added specially for DECOVID. |
| days\_supply | Not used. |
| sig | Not used. |
| route\_concept\_id | A foreign key to a predefined concept in the Standardized Vocabularies reflecting the route of administration. |
| lot\_number | Not used. |
| provider\_id | Not used. |
| visit\_occurrence\_id | A foreign key to the Visit in the VISIT\_OCCURRENCE table during which the Drug Exposure was initiated. |
| visit\_detail\_id | A foreign key to the Visit Detail in the VISIT\_DETAIL table during which the Drug Exposure was initiated. |
| drug\_source\_value | Not used. |
| drug\_source\_concept\_id | Not used. |
| route\_source\_value | Not used. |
| dose\_unit\_source\_value | Not used. |
| last\_updated\_datetime | Initially this is the time the row was written to the OMOP standardised database. If the row is subsequently updated, it then becomes the most recent update time. |
| deleted\_datetime | Null initially. Set to the time when the row was marked for deletion. A value in this column sent to the central data source will result in this row being deleted (by person\_id) from the combined dataset. Deletions only need to be sent once. |

**Drug Type**

The concepts in this table were mapped from vocabularies including drug standards. The drug exposure records in this table are predominantly mapped from DM+D codes (i.e. drug\_concept\_id). The following table presents the count of drug\_exposure\_ids, or data records in the DRUG\_EXPOSURE table, for drug vocabularies by Trust (cells <10 have been suppressed to display “<10”):

|  |  |  |
| --- | --- | --- |
| **Vocabulary** | **UCLH** | **UHB** |
| dm+d | 5363282 | 9575852 |
| None | 139493 | <=10 |
| SNOMED | 50018 | 64590 |

In addition, the drug vocabularies primary map to a number of concept classes that may be of interest:

|  |  |  |
| --- | --- | --- |
| **Concept Class** | **UCLH** | **UHB** |
| AMP | 292054 | 151757 |
| Clinical Drug | <=10 | 485 |
| Ingredient | <=10 | 837098 |
| Pharma/Biol Product | 50018 | 64105 |
| Undefined | 139493 | <=10 |
| VMP | 5071228 | 1387037 |

**Notes:**

* The Trust can be derived by taking the last digit of the visit\_detail\_id, person\_id and/or visit\_occurrence\_id where 4 is for UHB and 6 is for UCLH. Not all drug exposures have a related visit\_occurrence\_id/visit\_detail\_id.
* All drug administrations refer to inpatient administrations where we are able to identify the formulation, and time of specific delivery.
* UHB implemented a system of “panels” of measurements, where measurements (even those which did not come from a physical specimen) were grouped by virtual specimens. This allows grouping of, for example, simultaneous NEWS2 component measurements. It also allows recording of a sample time (as opposed to the time a measurement was made available in the EHR, which is what the measurement time reflects).
* Dosage data for drugs at UHB may be of variable quality, because UHB maps most of its drug exposures to VTM that does not encode dosage in the drug name as done for VMP primarily used at UCLH, and the OMOP quantity field only permits one quantity type to describe the dosage, which in the case of drug infusions is reserved for the rate of infusion, so it is not possible to determine the concentration of drug infusions at UHB. For example, “Glucose” may be recorded as administered at 100mL/h at UHB with no information on the concentration of glucose being administered.
* The OMOP standardised database uses Athena the Standardized Vocabularies can be viewed on the OHDSI Athena online browser (<https://athena.ohdsi.org/>). For the DRUG\_EXPOSURE table, it stores a one-to-one link between most DM+D entries and their RxNorm counterparts. The logic to query, for example, all drugs that contain a particular ingredient would be:
  + Find the mapping for the drug in DM+D to RxNorm in the CONCEPT\_RELATIONSHIP table
  + Left join to the CONCEPT\_ANCESTOR table
  + Select all descendants of the concept
  + Left join back to the CONCEPT\_RELATIONSHIP table
  + Left join to CONCEPT table
  + filter out all those that do not have a vocabulary\_id that is “DM+D”

**PROCEDURE\_OCCURRENCE**

In DECOVID, the PROCEDURE\_OCCURRENCE table contains only records on the insertion and removal of endotracheal and tracheostomy tubes. The concepts used in this table were all mapped from SNOMED codes.

| **Field** | **Description** |
| --- | --- |
| procedure\_occurrence\_id | A system-generated unique identifier for each Procedure Occurrence. |
| person\_id | 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 | A foreign key that refers to a standard procedure Concept identifier in the Standardized Vocabularies. |
| procedure\_date | The date on which the Procedure was performed. |
| procedure\_datetime | The date and time on which the Procedure was performed. |
| procedure\_type\_concept\_id | A foreign key to the predefined Concept identifier in the Standardized Vocabularies reflecting the type of source data from which the procedure record is derived. |
| modifier\_concept\_id | A foreign key to a Standard Concept identifier for a modifier to the Procedure (e.g. bilateral) |
| Quantity | Not used. |
| provider\_id | Not used. |
| visit\_occurrence\_id | A foreign key to the Visit in the VISIT\_OCCURRENCE table during which the Procedure was carried out. |
| visit\_detail\_id | A foreign key to the Visit Detail in the VISIT\_DETAIL table during which the Procedure was carried out. |
| procedure\_source\_value | The source code for the Procedure as it appears in the source data. This code is mapped to a standard procedure Concept in the Standardized Vocabularies and the original code is stored in this field for reference. Procedure source codes are typically ICD-9-Proc, CPT-4, HCPCS or OPCS-4 codes. |
| procedure\_source\_concept\_id | A foreign key to a Procedure Concept that refers to the code used in the source. |
| modifier\_source\_value | The source code for the qualifier as it appears in the source data. |
| last\_updated\_datetime | Initially this is the time the row was written to the OMOP standardised database. If the row is subsequently updated, it then becomes the most recent update time. |
| deleted\_datetime | Null initially. Set to the time when the row was marked for deletion. A value in this column sent to the central data source will result in this row being deleted (by person\_id) from the combined dataset. Deletions only need to be sent once. |

**Notes:**

* The procedure\_occurrence table only contains four procedures, namely, the insertion and removal of endotracheal and tracheostomy tubes.

**FACT\_RELATIONSHIP**

The FACT\_RELATIONSHIP table contains records (i.e. facts) that belong to OMOP-CDM domains (e.g. Measurement) and their relationship(s) with other records from any of the OMOP-CDM data tables that may belong to the same OMOP-CDM domain or a different OMOP-CDM domain.

| **Field** | **Description** |
| --- | --- |
| domain\_concept\_id\_1 | The concept representing the domain of fact one, from which the corresponding table can be inferred. |
| fact\_id\_1 | The unique identifier in the table corresponding to the domain of fact one. |
| domain\_concept\_id\_2 | The concept representing the domain of fact two, from which the corresponding table can be inferred. |
| fact\_id\_2 | The unique identifier in the table corresponding to the domain of fact two. |
| relationship\_concept\_id | A foreign key to a Standard Concept ID of relationship in the Standardized Vocabularies. |
| last\_updated\_datetime | Initially this is the time the row was written to the OMOP standardised database. If the row is subsequently updated, it then becomes the most recent update time. |
| deleted\_datetime | Null initially. Set to the time when the row was marked for deletion. A value in this column sent to the central data source will result in this row being deleted (by person\_id) from the combined dataset. Deletions only need to be sent once. |

**Notes:**

* The FACT\_RELATIONSHIP table links two otherwise separately stored facts. For example:
  + There may be a row in the MEASUREMENT table that is the results of a laboratory sample.
  + There may be a row in the SPECIMEN table that has specimen\_concept\_id as something which means " Swab from nasal sinus" in which the measurement was based on.
  + The FACT\_RELATIONSHIP would include two rows for each direction of the link between the two rows above. The relationship types may be something along the lines of "metadata of" and "metadata for".

|  |
| --- |
| **STANDARDIZED VOCABULARIES DATA TABLES** |

**CONCEPT**

The CONCEPT table stores records representing standardized vocabularies national or international vocabularies. Concepts can represent broad categories (e.g. 'Cardiovascular disease'), detailed clinical elements ('Myocardial infarction of the anterolateral wall') or modifying characteristics and attributes that define Concepts at various levels of detail (severity of a disease, associated morphology, etc.).

| **Field** | **Description** |
| --- | --- |
| concept\_id | A unique identifier for each Concept across all domains. |
| concept\_name | An unambiguous, meaningful and descriptive name for the Concept. |
| domain\_id | A foreign key to the DOMAIN table the Concept belongs to. |
| vocabulary\_id | A foreign key to the VOCABULARY table indicating from which source the Concept has been adapted. |
| concept\_class\_id | The attribute or concept class of the Concept. Examples are 'Clinical Drug', 'Ingredient', 'Clinical Finding' etc. |
| standard\_concept | This flag determines where a Concept is a Standard Concept, i.e. is used in the data, a Classification Concept, or a non-standard Source Concept. The allowable values are 'S' (Standard Concept) and 'C' (Classification Concept), otherwise the content is NULL. |
| concept\_code | The concept code represents the identifier of the Concept in the source vocabulary, such as SNOMED-CT concept IDs, RxNorm RXCUIs etc. Note that concept codes are not unique across vocabularies. |
| valid\_start\_date | The date when the Concept was first recorded. The default value is 1-Jan-1970, meaning, the Concept has no (known) date of inception. |
| valid\_end\_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, meaning, the Concept is valid until it becomes deprecated. |
| invalid\_reason | 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. |

**VOCABULARY**

The VOCABULARY table includes a list of the Vocabularies collected from various sources or created de novo by the OMOP community.

| **Field** | **Description** |
| --- | --- |
| vocabulary\_id | A unique identifier for each Vocabulary, such as ICD9CM, SNOMED, Visit. |
| vocabulary\_name | The name describing the vocabulary, for example "International Classification of Diseases, Ninth Revision, Clinical Modification, Volume 1 and 2 (NCHS)" etc. |
| vocabulary\_reference | External reference to documentation or available download for the vocabulary. |
| vocabulary\_version | Version of the Vocabulary as indicated in the source. |
| vocabulary\_concept\_id | A foreign key that refers to a standard concept identifier in the CONCEPT table for the Standardised/Unstandardised Vocabulary that the record belongs to. |

**DOMAIN**

The DOMAIN table includes a list of OMOP-defined Domains the Concepts of the Standardized Vocabularies can belong to.

| **Field** | **Description** |
| --- | --- |
| domain\_id | A unique key for each domain. |
| domain\_name | The name describing the Domain, e.g. "Condition", "Procedure", "Measurement" etc. |
| domain\_concept\_id | A foreign key that refers to an identifier in the CONCEPT table for the unique Domain Concept the Domain record belongs to. |

**CONCEPT\_CLASS**

The CONCEPT\_CLASS table is a reference table, which includes a list of the classifications used to differentiate Concepts within a given Vocabulary.

| **Field** | **Description** |
| --- | --- |
| concept\_class\_id | A unique key for each class. |
| concept\_class\_name | The name describing the Concept Class, e.g. "Clinical Finding", "Ingredient", etc. |
| concept\_class\_concept\_id | A foreign key that refers to an identifier in the CONCEPT table for the unique Concept Class the record belongs to. |

**CONCEPT\_RELATIONSHIP**

The CONCEPT\_RELATIONSHIP table contains records that define direct relationships between any two Concepts and the nature or type of the relationship. Each type of a relationship is defined in the RELATIONSHIP table.

| **Field** | **Description** |
| --- | --- |
| concept\_id\_1 | A foreign key to a Concept in the CONCEPT table associated with the relationship. Relationships are directional, and this field represents the source concept designation. |
| concept\_id\_2 | A foreign key to a Concept in the CONCEPT table associated with the relationship. Relationships are directional, and this field represents the destination concept designation. |
| relationship\_id | A unique identifier to the type or nature of the Relationship as defined in the RELATIONSHIP table. |
| valid\_start\_date | The date when the instance of the Concept Relationship is first recorded. |
| valid\_end\_date | The date when the concept\_relationship field became invalid because it was deleted or superseded (updated) by a new relationship. Default value is 31-Dec-2099. |
| invalid\_reason | 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. |

**RELATIONSHIP**

The RELATIONSHIP table provides a reference list of all types of relationships that can be used to associate any two concepts in the CONCEPT\_RELATIONSHIP table.

| **Field** | **Description** |
| --- | --- |
| relationship\_id | The type of relationship captured by the relationship record. |
| relationship\_name | The text that describes the relationship type. |
| is\_hierarchical | Defines whether a relationship defines concepts into classes or hierarchies. Values are 1 for hierarchical relationship or 0 if not. |
| defines\_ancestry | Defines whether a hierarchical relationship contributes to the CONCEPT\_ANCESTOR table. These are subsets of the hierarchical relationships. Valid values are 1 or 0. |
| reverse\_relationship\_id | The identifier for the relationship used to define the reverse relationship between two concepts. |
| relationship\_concept\_id | A foreign key that refers to an identifier in the CONCEPT table for the unique relationship concept. |

**CONCEPT\_SYNONYM**

The CONCEPT\_SYNONYM table is used to store alternate names and descriptions for Concepts.

| **Field** | **Description** |
| --- | --- |
| concept\_id | A foreign key to the Concept in the CONCEPT table. |
| concept\_synonym\_name | The alternative name for the Concept. |
| language\_concept\_id | A foreign key to a Concept representing the language. |

**CONCEPT\_ANCESTOR**

The CONCEPT\_ANCESTOR table is designed to simplify observational analysis by providing the complete hierarchical relationships between Concepts. Only direct parent-child relationships between Concepts are stored in the CONCEPT\_RELATIONSHIP table.

| **Field** | **Description** |
| --- | --- |
| ancestor\_concept\_id | A foreign key to the concept in the CONCEPT table for the higher-level concept that forms the ancestor in the relationship. |
| descendant\_concept\_id | A foreign key to the concept in the CONCEPT table for the lower-level concept that forms the descendant in the relationship. |
| min\_levels\_of\_separation | The minimum separation in number of levels of hierarchy between ancestor and descendant concepts. This is an attribute that is used to simplify hierarchic analysis. |
| max\_levels\_of\_separation | The maximum separation in number of levels of hierarchy between ancestor and descendant concepts. This is an attribute that is used to simplify hierarchic analysis. |

|  |
| --- |
| **STANDARDIZED HEALTH SYSTEM DATA TABLES** |

**LOCATION**

The LOCATION table captures the physical location or address information of patients in the PERSON table.

| **Field** | **Description** |
| --- | --- |
| location\_id | A unique identifier for each geographic location. |
| address\_1 | Not used. |
| address\_2 | Not used. |
| city | Not used. |
| state | Not used. |
| zip | In DECOVID this is the LSOA |
| county | Not used. |
| location\_source\_value | The verbatim information that is used to uniquely identify the location as it appears in the source data. |

**Notes:**

* The place of residency is stored in the location\_id column, where each patient’s LSOA is listed in place of their postcode (i.e. the zip field). The location\_id field can be linked to the PERSON table.

**CARE\_SITE**

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

| **Field** | **Description** |
| --- | --- |
| care\_site\_id | A unique identifier for each Care Site. |
| care\_site\_name | The description of the Care Site care level (e.g. ED, ICU, Ward). |
| place\_of\_service\_concept\_id | Not used. |
| location\_id | Not used. |
| care\_site\_source\_value | Not used. |
| place\_of\_service\_source\_value | Not used. |

**Notes:**

* care\_site\_id 13 and care\_site\_id 19 both have care\_site\_name "Out Patient Department", and only care\_site\_id 13 is used in DECOVID.

**REFERENCES**

[1] Office for National Statistics. Ethnic group, national identity and religion.

<https://www.ons.gov.uk/methodology/classificationsandstandards/measuringequality/ethnicgroupnationalidentityandreligion>

[2] UK Data Service. 2011 Census Geography boundaries (Lower Layer Super Output Areas and Data Zones). <https://statistics.ukdataservice.ac.uk/dataset/2011-census-geography-boundaries-lower-layer-super-output-areas-and-data-zones>

[3] NHS Digital. NHS Data Model and Dictionary, Admission Source. <https://www.datadictionary.nhs.uk/attributes/admission_source.html?hl=admission%2Csource>

[4] NHS Digital. NHS Data Model and Dictionary, Destination of Discharge. <https://www.datadictionary.nhs.uk/attributes/destination_of_discharge.html>