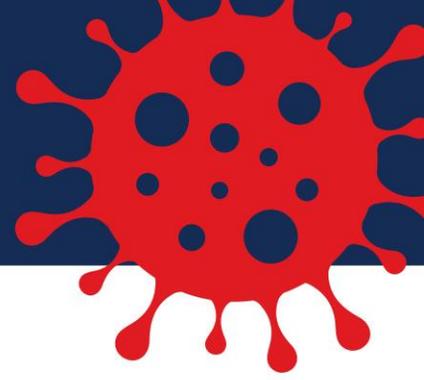


COVID-19 Patient De-identified Dataset for Research



DATA GUIDE

A program of the UF Clinical and Translational Science Institute and UF Health Information Technology, the [UF Integrated Data Repository](#), or IDR, consists of a secure, clinical data warehouse that aggregates figures from the university's various clinical and administrative information systems, including the Epic electronic health record system. The IDR contains more than 1 billion observational facts pertaining to more than 1 million patients.

The IDR now includes data on every patient who has presented at UF Health with COVID-19-like symptoms and/or who has been tested for the SARS-COV-2 coronavirus since January 2020. This document supports use of the [COVID-19 De-identified Patient Dataset](#) (please use that link for more details, including the COVID-19 Patient Dataset Fact Sheet and Instructions.)

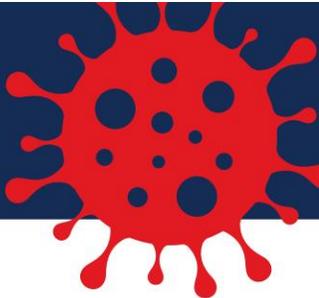
Updated: 11/12/20

Connect with us at:
IRBDataRequest@ahc.ufl.edu

UF Clinical and Translational
Science Institute
UNIVERSITY of FLORIDA

UFHealth
INFORMATION TECHNOLOGY

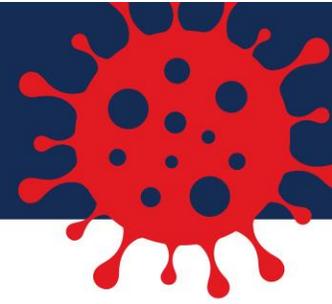
COVID-19 Patient De-identified Dataset for Research



CONTENTS

OVERVIEW	3	CONDITION_OCCURRENCE	9
IRB Approval	3	PROCEDURE_OCCURRENCE	10
Access to the Dataset	3	DRUG_EXPOSURE	11
Data Use Agreement	3	MEASUREMENT	13
Acknowledgement	3	OBSERVATION	14
Operational or Clinical Use of the Dataset	3	OBSERVATION_PERIOD	15
Questions and Comments	3	HEALTH SYSTEM DATA TABLES	16
		LOCATION, CARE_SITE AND PROVIDER	16
DATASET	4	VOCABULARY TABLES	16
Inclusion Criteria	4	CONCEPT	16
Excluded Data	4	OTHER VOCABULARY TABLES	17
Data De-identification	4		
Identifiers	4	APPENDIX A	18
Dates	4	COVID-19 LAB	18
Notes on Data Quality and Limitations	4	PROCEDURE CODES FOR COVID-19 LAB ORDER	18
OMOP Data Format	4	MEDICATIONS	18
Versions	5	DIAGNOSES	18
DATA DICTIONARY	6	APPENDIX B	19
CLINICAL DATA TABLES	6	CONDITION_OCCURRENCE	19
PERSON	6	MEASUREMENT	19
DEATH	7	OBSERVATION	21
VISIT_OCCURRENCE	8		

COVID-19 Patient De-identified Dataset for Research



OVERVIEW

IRB Approval

The COVID-19 De-identified Dataset adheres to a UF Institutional Review Board-approved protocol that releases the data for research purposes. A researcher who publishes based solely on this dataset is not required to submit a separate IRB protocol for approval.

Access to the Dataset

To access the dataset, fill out the [online data access form](#). Enter your information in PI fields. Fill out one form for each person on your research team. Select 'COVID-19 Dataset (Deidentified)' as request type. Provisioning is typically completed within three business days of the request. Once provisioned, you can follow our [COVID-19 Patient Dataset Instructions](#) for how to access the data.

Data Use Agreement

- These data are approved for use by the UF and UF Health community only. These data should not be shared outside the organization.
- These data are approved only for research purposes and should not be used for non-research purposes.
- Research publications based on these data should publish only aggregate analyses.
- Users agree not to attempt to re-identify any patients whose data are included.

Acknowledgement

Research publications using these data should acknowledge the UF Integrated Data Repository, the UF Clinical and Translational Science Institute and UF Health IT Services.

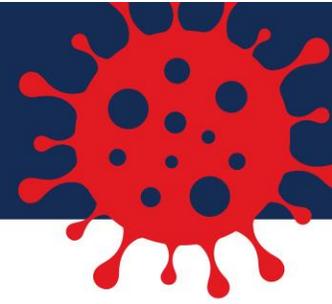
Operational or Clinical Use of the Dataset

UF Health care teams interested in these data for operational or clinical purposes should contact the IDR team IRBDataRequest@ahc.ufl.edu for additional guidance.

Questions and Comments

The IDR team welcomes feedback and collaboration from the UF and UF Health research community. We welcome ideas for how we can grow and improve the COVID-19 De-identified Dataset to maximally support high-impact research. For questions and comments, please email us at IRBDataRequest@ahc.ufl.edu.

COVID-19 Patient De-identified Dataset for Research



DATASET

Inclusion Criteria

The current dataset includes all patients who were tested for COVID-19 as well as patients with other COVID-19-related symptoms (such as respiratory illness, cough, fever, etc.) served at any UF Health site (Gainesville, Jacksonville or Central Florida). For the detailed list of inclusion criteria, see Appendix A.

Excluded Data

Laboratory tests performed through the UF Health Screen, Test, and Protect program are excluded from the data.

Data De-identification

To protect patients' privacy, all protected health information is de-identified:

Identifiers

Each patient is assigned a unique random `person_id`. The `person_id` for a patient remains consistent in all files.

Each encounter is assigned a unique random `visit_id`. The `visit_id` for a specific encounter remains consistent in all files.

Dates

For de-identification and confidentiality purposes, all dates are shifted forward or backwards for patients.

For an individual patient, each date along their care path is shifted the same amount, which allows examining time-series data for a single patient. Note that since each patient has a different date

shift, time-series analysis of the entire population is not possible.

The date of birth is set to January 1, 1800 for patients who are at least 90 years old.

Notes on Data Quality and Limitations

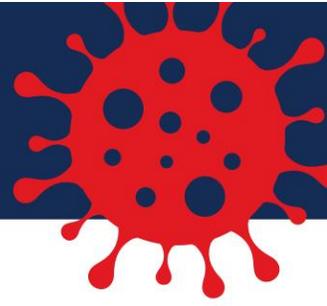
There is a lack of coding standardization for COVID-19 thus far in real-life health system practice, thus [phenotypes](#) may be unreliable.

Note that data are provided as they exist in the UF Health clinical information systems (for example, the Epic electronic health record, or EHR) with limited edits or modifications. Researchers analyzing these data should recognize the inherent limitations of electronic clinical data for research. This includes the fact that UF Health patient records will not include comprehensive history of care because many patients receive care at multiple health care organizations.

OMOP Data Format

Data is provided in [OMOP Common Data Model](#) format, managed by [OHDSI](#) community, which is the model most-used internationally for standard representation of EHR data for research purposes. Only selected tables from the OMOP model are populated. Note that additional tables might be added in future versions. Each table is provided as a .csv file.

COVID-19 Patient De-identified Dataset for Research

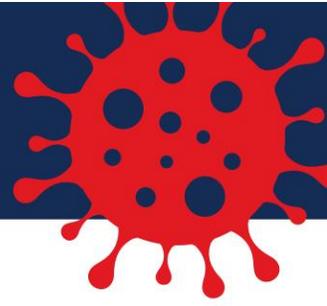


Versions

A new version of the dataset is released regularly.

The person_id and visit_id de-identifications are not necessarily back-compatible with versions earlier than version 5.0, but they will be forward-compatible. This means that if you used one of the versions before 5.0 and identified a person_id and/or a visit_id of interest, they might not remain the same in the current version. However, the person_id and visit_id for a specific patient/encounter will remain the same in all future versions starting with version 5.0.

Version 5.0 includes several modifications from previous versions to make the data fully compliant with OMOP version 5.3.1. These modifications will remain consistent in future versions.



DATA DICTIONARY

Data are provided in OMOP Common Data Model version 5.3.1 format. Each .csv file corresponds to an OMOP table. For research analysis using custom scripts (R, Python, SAS, etc.), at minimum, you will be using some of the clinical data tables and the concept table (see the section on the concept table to learn how to use it). Other tables that are provided in this dataset are necessary for proper use of OHDSI tools, such as Atlas. However, you may find these other tables useful for processing data using custom scripts as well. The IDR team does not currently provide support for any of OHDSI tools, but you may install them on your own and load this dataset to use them. Note that some columns in the tables are not populated or are populated with dummy values due to de-identification requirements or non-existence of data. Columns populated with dummy data are clearly indicated in the descriptions below.

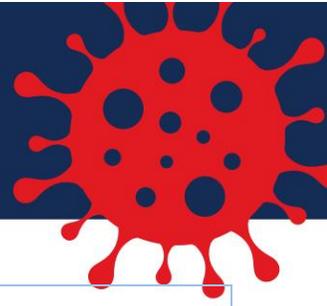
CLINICAL DATA TABLES

PERSON

The *person* table contains demographical information for each patient in the cohort.

Column Name	Description	Notes
person_id *	A unique identifier for each patient	
person_source_value	Not populated	
gender_source_value	Sex as reported at birth	
gender_concept_id	OMOP standard concept_id for specific gender	Derived based on gender_source_value
gender_source_concept_id	Not populated	
race_source_value	Race as reported in EHR	
race_concept_id	OMOP standard concept for specific race	Derived based on race_source_value
race_source_concept_id	Not populated	
ethnicity_source_value	Ethnicity as reported in EHR	
ethnicity_concept_id	OMOP standard concept_id for specific ethnicity	Derived based on ethnicity_source_value
ethnicity_source_concept_id	Not populated	
birth_datetime **	Date and time of birth	If time is unknown, it is set to midnight
year_of_birth	Year part of birth_datetime	

COVID-19 Patient De-identified Dataset for Research



month_of_birth	Month part of birth_datetime	
day_of_birth	Day part of birth_datetime	
location_id ***	Populated with dummy values	
provider_id ***	Populated with dummy values	
care_site_id ***	Populated with dummy values	

*person_id is used in each table in the dataset. It can be used to link data from multiple tables that is related to the same patient.

**Value is shifted to comply with de-identification requirements.

***Dummy value is used to comply with requirements of OHDSI tools.

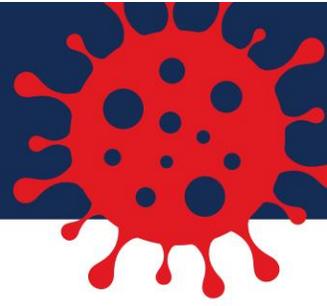
DEATH

The *death* table contains the death date for each patient that is known to be deceased. Data is collected from both the Epic Electronic Health Record, or EHR, system and Social Security Death Index. If death dates differ across the two sources, the death date as recorded in the Epic EHR is used.

Column Name	Description	Notes
person_id	See description in person table	
death_date *	Date of death	
death_datetime *	Date and time of death	If time is unknown, it is set to midnight
death_type_concept_id	Source of information	Currently set to EHR
cause_concept_id	Not populated	
cause_source_value	Not populated	
cause_source_concept_id	Not populated	

*Value is shifted to comply with de-identification requirements.

COVID-19 Patient De-identified Dataset for Research



VISIT_OCCURRENCE

The *visit_occurrence* table contains data about each encounter.

Column Name	Description	Notes
visit_occurrence_id *	A unique identifier for each encounter	
person_id	See description in person table	
preceding_visit_occurrence_id	Not populated	
visit_source_value	Type of visit (e.g., inpatient, outpatient)	
visit_concept_id	OMOP standard concept_id for specific visit type	Derived from visit_source_value
visit_source_concept_id	Not populated	
visit_start_date **	Start date of encounter	Admit date for hospital visit, appointment date for clinic visit
visit_start_datetime **	Start date and time of encounter	If time is unknown, it is set to midnight
visit_end_date **	End date of encounter	Discharge date for hospital visit, appointment date for clinic visit
visit_end_datetime **	End date and time of encounter	If time is unknown, it is set to midnight
visit_type_concept_id	Source of information	Currently set to EHR
admitting_source_value	Where the patient was admitted from (e.g., referral, emergency room, from another hospital)	
admitting_source_concept_id	OMOP standard concept_id for specific admitting source	Derived from admitting_source_value
discharge_to_source_value	Where the patient was discharged following the visit (e.g., home, alternate health care facility)	
discharge_to_concept_id	OMOP standard concept_id for specific discharge location	Derived from discharge_to_source_value
provider_id ***	Populated with dummy values	
care_site_id ***	Populated with dummy values	

COVID-19 Patient De-identified Dataset for Research



*visit_occurrence_id is used in many tables in the dataset. It can be used to link data from multiple tables that is related to the same encounter.

**Value is shifted to comply with de-identification requirements.

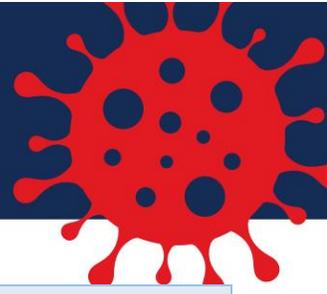
***Dummy value is used to comply with requirements of OHDSI tools.

CONDITION_OCCURRENCE

The *condition_occurrence* table contains most conditions and diagnoses as coded in hospital billing, professional billing, and the problem list. Note that some concepts that are coded with diagnostic ICD codes are classified as measurements, procedures, or observations in OMOP vocabulary and will therefore be found in those tables. Additionally, dependence on mechanical ventilation is found in this *condition_occurrence* table even though it is not coded with an ICD diagnostic code.

Column Name	Description	Notes
condition_occurrence_id	A unique identifier for the table	The field is important for proper function of OHDSI tools (e.g., Atlas), but not necessarily for custom scripts
person_id	See description in person table	
visit_occurrence_id	See description in visit_occurrence table	
visit_detail_id	Not populated	
condition_source_value *	Code denoting the condition	
condition_concept_id	OMOP standard concept_id for specific condition	Derived based on condition_source_value
condition_source_concept_id	OMOP concept_id for specific condition; not necessarily standard concept_id	Derived based on condition_source_value
condition_start_date **	Start date of diagnosis	
condition_start_datetime **	Start date and time of diagnosis	If time is unknown, it is set to midnight
condition_end_date **	End date of diagnosis if available	
condition_end_datetime **	End date and time of diagnosis	If time is unknown, it is set to midnight
condition_type_concept_id	Source of information	Currently set to EHR
stop_reason	Not populated	
condition_status_source_value	Not populated	

COVID-19 Patient De-identified Dataset for Research



condition_status_concept_id	Not populated	
provider_id ***	Populated with dummy values	

*Diagnosis ICD9 or ICD10 code or description of the condition if not coded with a diagnostic ICD code. For the list of possible non-ICD values, see Appendix B.

**Value is shifted to comply with de-identification requirements.

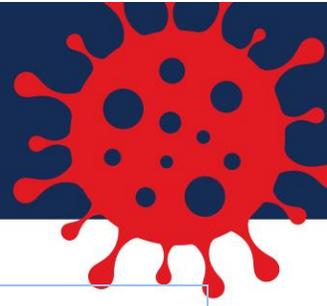
***Dummy value is used to comply with requirements of OHDSI tools.

PROCEDURE_OCCURRENCE

The *procedure_occurrence* table contains most procedures as coded in hospital billing and professional billing. Note that some concepts that are coded with CPT codes are classified as measurements in OMOP vocabulary and will therefore be found in the measurement table. Additionally, some concepts that are coded with diagnostic ICD codes are classified as procedures and are therefore found in this procedure_occurrence table.

Column Name	Description	Notes
procedure_occurrence_id	A unique identifier for the table	The field is important for proper function of OHDSI tools (e.g., Atlas), but not necessarily for custom scripts
person_id	See description in person table	
visit_occurrence_id	See description in visit_occurrence table	
visit_detail_id	Not populated	
procedure_source_value *	Code denoting the procedure	
procedure_concept_id	OMOP standard concept_id for specific procedure	Derived based on procedure_source_value
procedure_source_concept_id	OMOP concept_id for specific procedure; not necessarily standard concept_id	Derived based on procedure_source_value
procedure_date **	Start date of procedure	
procedure_datetime **	Start date and time of procedure	If time is unknown, it is set to midnight
procedure_type_concept_id	Source of information	Currently set to EHR
quantity	Not populated	

COVID-19 Patient De-identified Dataset for Research



modifier_source_value	Not populated	
modifier_concept_id	Not populated	
provider_id ***	Populated with dummy values	

*Procedure ICD9, ICD10, CPT or HCPCS code or Diagnostic ICD9 or ICD10 code.

**Value is shifted to comply with de-identification requirements.

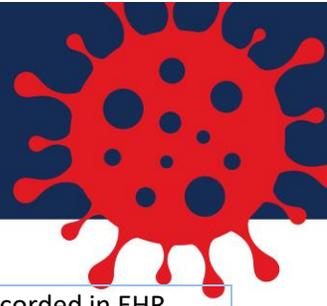
***Dummy value is used to comply with requirements of OHDSI tools.

DRUG_EXPOSURE

The *drug_exposure* table contains all medication orders.

Column Name	Description	Notes
drug_exposure_id	A unique identifier for the table	The field is important for proper function of OHDSI tools (e.g., Atlas), but not necessarily for custom scripts
person_id	See description in person table	
visit_occurrence_id	See description in visit_occurrence table	
visit_detail_is	Not populated	
drug_source_value *	Code denoting the medication	
drug_concept_id	OMOP standard concept_id for specific medication	Derived based on drug_source_value
drug_source_concept_id	OMOP concept_id for specific medication; not necessarily standard concept_id	Derived based on drug_source_value
drug_exposure_start_date **	Start date for taking medication	If start date is not recorded in EHR, order date is used
drug_exposure_start_datetime **	Start date and time for taking medication	If time is unknown, it is set to midnight
drug_exposure_end_date **	End date for taking medication	If end date is not recorded in EHR, it is set to start date
drug_exposure_end_datetime **	End date and time for taking medication	If time is unknown, it is set to midnight

COVID-19 Patient De-identified Dataset for Research



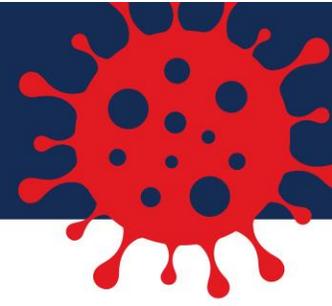
verbatim_end_date **	End date for taking medication as recorded in EHR	If end date is not recorded in EHR, this column remains blank
drug_type_concept_id	Source of information	Currently set to EHR
refills	The number of refills after the initial prescription	
quantity	The drug quantity as recorded in the original prescription or dispensing record	
days_supply	Not populated	
sig	The directions on the drug prescription as recorded in the original prescription (and printed on the container) or dispensing record	
route_source_value	Med route as specified in the EHR (e.g., epidural, intravenous, inhalation, injection, etc.)	
route_concept_id	OMOP standard concept_id for specific medication route	Derived based on route_source_value
dose_unit_source_value	Dose unit as specified in the EHR (e.g., bottle, capsule, etc.)	
stop_reason	Not populated	
lot_number	Not populated	
provider_id	Populated with dummy values	

*RxNorm code.

**Value is shifted to comply with de-identification requirements.

***Dummy value is used to comply with requirements of OHDSI tools.

COVID-19 Patient De-identified Dataset for Research

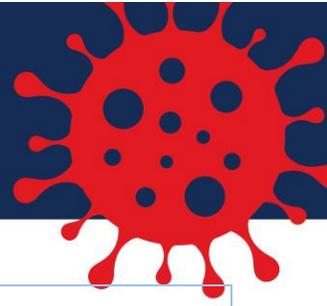


MEASUREMENT

The *measurement* table contains various measurement collected during all encounters. It contains measurements including but not limited to lab results, vital measurements, and flowsheet values. Note that some diagnoses and procedures are classified as measurements in the OMOP vocabulary, so they will appear in this measurement table.

Column Name	Description	Notes
measurement_id	A unique identifier for the table	The field is important for proper function of OHDSI tools (e.g., Atlas), but not necessarily for custom scripts
person_id	See description in person table	
visit_occurrence_id	See description in visit_occurrence table	
visit_detail_id	Not populated	
measurement_source_value *	Code denoting the measurement	
measurement_concept_id	OMOP standard concept_id for specific measurement	Derived based on measurement_source_value
measurement_source_concept_id	OMOP concept_id for specific measurement; not necessarily standard concept_id	Derived based on measurement_source_value
measurement_date **	Date of measurement	
measurement_datetime **	Date and time of measurement	
measurement_time	Time part of measurement_datetime	
measurement_type_concept_id	Source of information	Currently set to EHR
value_source_value	Result of measurement	
value_as_number	Result of measurement if the result is numeric	
value_as_concept_id	Not populated	
operator_concept_id	Not populated	
range_low	The lower bound of normal range of the measurement	Only populated for a subset of labs
range_high	The upper bound of normal range of the measurement	Only populated for a subset of labs

COVID-19 Patient De-identified Dataset for Research



unit_source_value	The unit of measurement result	
unit_concept_id	OMOP standard concept_id for specific unit	Derived based on unit_source_value
provider_id ***	Populated with dummy values	

*LOINC code for lab, CPT code for procedure, ICD9 or ICD10 code for diagnosis or description of the measurement. For the list of possible values, see Appendix B.

**Value is shifted to comply with de-identification requirements.

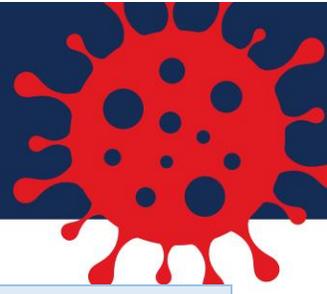
***Dummy value is used to comply with requirements of OHDSI tools.

OBSERVATION

The *observation* table contains concepts that are classified as observations by OMOP vocabulary. Its features include smoking history, three-digit zip code, whether the patient was treated in an ICU during an encounter, start and stop times of mechanical ventilation, and the patient's insurance type among others.

Column Name	Description	Notes
observation_id	A unique identifier for the table	The field is important for proper function of OHDSI tools (e.g., Atlas), but not necessarily for custom scripts
person_id	See description in person table	
visit_occurrence_id	See description in visit_occurrence table	
visit_detail_id	Not populated	
observation_source_value *	Code denoting the observation	
observation_concept_id	OMOP standard concept_id for specific observation	Derived based on observation_source_value
observation_source_concept_id	OMOP concept_id for specific observation; not necessarily standard concept_id	Derived based on observation_source_value
observation_date **	Date of observation	
observation_datetime **	Date and time of observation	If time is unknown, it is set to midnight
observation_type_concept_id	Source of information	Currently set to EHR

COVID-19 Patient De-identified Dataset for Research



value_as_number ***	Result of observation if numeric	
value_as_string ***	Result of observation if non-numeric	
value_as_concept_id	Not populated	
qualifier_source_value	Not populated	
qualifier_concept_id	Not populated	
unit_source_value	Not populated	
unit_concept_id	Not populated	
provider_id ****	Populated with dummy values	

*Diagnostic ICD9 or ICD10 code or description of the observation. For the list of possible values, see Appendix B.

**Value is shifted to comply with de-identification requirements.

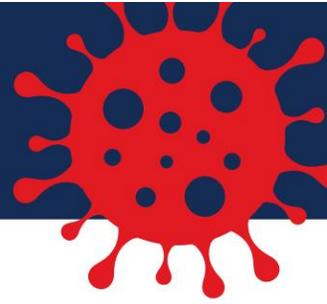
***Each observation will have populated value_as_number or value_as_string column.

****Dummy value is used to comply with requirements of OHDSI tools.

OBSERVATION_PERIOD

The *observation_period* table provides start and end dates for the available records for each person. In this dataset, we include data from January 1, 2012 until the date of data release, so the observation period is the same for each patient. This table is needed for proper function of OHDSI tools (e.g., Atlas).

COVID-19 Patient De-identified Dataset for Research



HEALTH SYSTEM DATA TABLES

LOCATION, CARE_SITE and PROVIDER

These tables are not populated with real values for this dataset. However, in order for OHDSI tools, such as Atlas, to load data properly, these tables are necessary. Thus, each table is populated with minimum necessary dummy values to enable use of OHDSI tools.

VOCABULARY TABLES

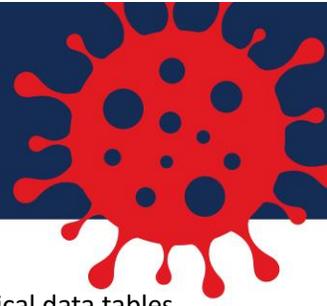
Most vocabulary tables are only needed if you plan to load data into tools provided by OHDSI community for data analysis, such as Atlas. If you are writing custom scripts using R, Python, SAS or other data analysis tools, you will need to use the concept table, but you may find other tables useful as well. Vocabulary tables are downloaded from OHDSI community vocabulary webpage <https://athena.ohdsi.org/vocabulary/list>.

CONCEPT

The concept table contains all concepts that can be represented using OMOP common data model. Note that this dataset contains only a subset of all clinical concepts, so not all concepts from the concept table are represented in this dataset. The following columns from the concept table enable locating concepts and their meaning in clinical data tables:

- **concept_id:** As described in clinical data tables section, most tables contain at least one column that contains OMOP standard code. For example, in condition_occurrence table, this is condition_concept_id column. This is the column that is used to link a clinical data table with the concept table.
- **concept_code:** This column contains the code used for the concept. For example, it could be a specific diagnostic ICD10 code (e.g., E08).
- **concept_name:** This column contains description of the concept. It is a more human understandable description than concept_code. For example, it is a name of a disease (e.g., diabetes mellitus).
- **domain_id:** This column specifies the OMOP classification of the concept. In most cases, it specifies which clinical data table contains the concept that you are interested in. For example, if the domain_id is Condition, the concept is found in condition_occurrence table.
- **vocabulary_id:** This column specifies the type of coding that is used for the concept. For example, if vocabulary_id is ICD10CM, it means that the code in concept_code column is a diagnostic ICD10 code. Note that even if vocabulary_id is ICD10CM code, it does not mean that the concept is found in condition table; instead, domain_id dictates which table contains the concept.

COVID-19 Patient De-identified Dataset for Research

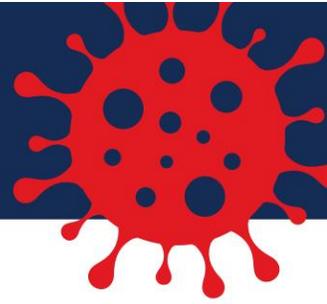


- **standard_concept:** OMOP model requires use of standard concepts in certain columns of clinical data tables. These concepts are marked with an 'S' in standard_concept column. If a concept of interest is not marked as standard, concept_relationship table can be used to find the relevant standard code.

OTHER VOCABULARY TABLES

Other vocabulary tables included in the dataset are: domain, vocabulary, concept_class, concept_relationship, concept_synonym, concept_ancestor, relationship, and drug_strength.

COVID-19 Patient De-identified Dataset for Research



APPENDIX A

Patients who have one or more of any of the following laboratory, procedure, medication, or diagnosis codes since January 1, 2020 are included in the dataset.

COVID-19 Lab

We use internal lab codes to identify these labs. However, COVID-19 laboratory results in the dataset can be identified by LOINC codes 94500-6 and 94309-2.

Procedure Codes for COVID-19 Lab Order

CPT code: 87635

HCPCS codes: U0001, U0002

Medications

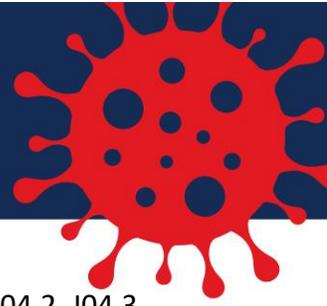
Any medication order that contains word 'Remdesivir' in its name.

Diagnoses

Includes coronavirus, COVID-19, flu, pneumonia, ARDS, and other non-COVID-19 diagnosis that may have been coded for patients with COVID-19, especially in early 2020.

041, 041.5, 041.81, 079.1, 079.2, 079.3, 079.6, 079.89, 466, 466.11, 466.19, 490, 491, 491.1, 491.8, 491.9, 506, 519.8, J20, J20.0, J20.1, J20.2, J20.3, J20.4, J20.5, J20.6, J20.7, J20.8, J20.9, J21, J21.0, J21.1, J21.8, J21.9, J22, J40, J41, J41.0, J41.1, J41.8, J42, J44.0, J47.0, J68.0, J84.115, J98.8, U07.1, B34.2, B97.2, B97.21, B97.29, J12.81, U04, U04.9, Z20.828, B34.9, 786.2, R05, 786.05, R06.02, R50.2, R50.84, R50.9, R56.00, 780.60, 780.61, R50, R50.8, R50.81, J09, J09, J09.01, J09.010, J09.018, J09.019, J09.02, J09.03, J09.090, J09.091, J09.092, J09.098, J09.11, J09.110, J09.118, J09.119, J09.12, J09.13, J09.190, J09.191, J09.192, J09.198, J09.X, J09.X1, J09.X2, J09.X3, J09.X9, J10, J10, J10.0, J10.0, J10.00, J10.01, J10.08, J10.1, J10.1, J10.2, J10.8, J10.8, J10.81, J10.82, J10.83, J10.89, J11, J11, J11.0, J11.0, J11.00, J11.08, J11.1, J11.1, J11.2, J11.8, J11.8, J11.81, J11.82, J11.83, J11.89, A01.03, A02.22, A37.01, A37.11, A37.81, A37.91, A54.84, B01.2, B05.2, B06.81, B20.6, B77.81, J12, J12.2, J12.3, J12.8, J12.81, J13, J14, J14, J15, J15.0, J15.1, J15.2, J15.20, J15.21, J15.211, J15.212, J15.29, J15.3, J15.4, J15.5, J15.6, J15.7, J15.8, J15.8, J15.9, J15.9, J16, J16.0, J16.8, J17, J17.0, J17.1, J17.2, J17.3, J17.8, J18, J18, J18.0, J18.0, J18.1, J18.1, J18.2, J18.2, J18.9, J18.9, J84.11, J84.111, J84.116, J84.117, J84.2, J85.1, J95.851, 480.0, 480.0, 480.1, 480.8, 480.9, 486, 487.0, J12, J12.0, J12.1, J12.8, J12.89, J12.9, J18.8, P23.0, J95.82, J95.821, J95.822, J96, J96.0, J96.00, J96.01, J96.02, J96.1, J96.10, J96.11, J96.12, J96.2, J96.20, J96.21, J96.22, J96.9, J96.90, J96.91, J96.92, P22, P22.0, P22.8, P22.9, P28.11, P28.5, P28.81, R06.03, R09.2, 518.82, J80, R06.03, B33.8, B34.8, B97.89, J00, J01.0, J01.00, J01.01, J01.1, J01.10, J01.11, J01.2, J01.20, J01.3, J01.30, J01.31, J01.4, J01.40, J01.41, J01.8, J01.80, J01.81, J01.9, J01.90, J01.91, J02, J02.0, J02.21, J02.8, J02.9, J03,

COVID-19 Patient De-identified Dataset for Research



J03.0, J03.00, J03.01, J03.8, J03.80, J03.81, J03.9, J03.90, J03.91, J04.0, J04.1, J04.10, J04.11, J04.2, J04.3, J04.30, J04.31, J05.0, J05.0, J05.1, J05.10, J05.11, J06.0, J06.9, J39.8

APPENDIX B

This section provides list and description of possible source values in each table.

CONDITION_OCCURRENCE

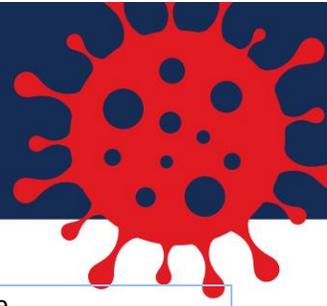
condition_source_value	Description
ICD9 or ICD10 code *	Diagnostic ICD9 or ICD10 code
Mechanical vent Y/N	Denotes whether mechanical ventilator was used during the encounter

*The field will contain the actual ICD9 or ICD10 code as recorded in EHR.

MEASUREMENT

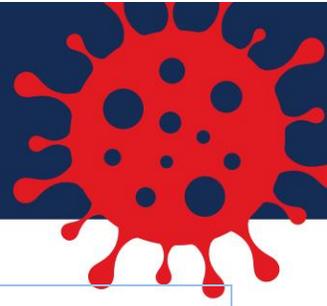
measurement_source_value	Description
LOINC code *	LOINC code for lab
CPT code **	CPT code for procedure
ICD code ***	ICD code for diagnosis or procedure
BP – Art Line DBP	Diastolic blood pressure
BP – Art Line SBP	Systolic blood pressure
ETCO2	End-tidal carbon dioxide
ETCO2 – Oral/Nasal	End-tidal carbon dioxide
FIO2	Fracture of inspired oxygen
GCS SCORE – Adult	Glasgow Coma Scale
GCS SCORE – Peds	Glasgow Coma Scale
HEART RATE	Heart rate
HEIGHT	Height
LDA – intubation tube type	Type of intubation tube

COVID-19 Patient De-identified Dataset for Research



MAP – Art Line	Mean arterial pressure – arterial line
MAP – Central Venous	Mean arterial pressure – central venous
MAP – Cuff	Mean arterial pressure – cuff
MAP – Non Invasive	Mean arterial pressure – non-invasive
MAP – Pulmonary	Mean arterial pressure – pulmonary
O2 FLOW RATE – L/MIN	Oxygen flow rate L/min
O2 FLOW RATE – mL/MIN	Oxygen flow rate mL/min
PAIN SCALE – Jax	Pain scale – JAX measure
PAIN SCALE – Peds Wong-Baker	Pain scale – pediatric measure
PAIN SCALE – UF DVPRS	Pain scale – GNV measure
PEEP	Positive End Expiratory Pressure
PIP	Peak Inspiratory Pressure
QTCB	QTc intervals
RESP DEVICE	Type of ventilator device used
RESP RATE	Respiratory rate
RESP RATE – Adult Mech	Breath rate mechanical -on ventilator
RESP RATE – Adult Spont	Breath rate spontaneous
RESP RATE – Peds Mech	Breath rate mechanical -on ventilator
RESP RATE – Peds Spont	Breath rate spontaneous
ROTHMAN SCORE	Rothman Index
SOFA_SCORE	Sequential Organ Failure Assessment score
SOFA – CARDIOVASCULAR	SOFA – cardiovascular component
SOFA – CNS	SOFA – CSN component
SOFA – COAGULATION	SOFA – coagulation component
SOFA - LIVER	SOFA – liver component
SOFA – RENAL	SOFA – renal component
SOFA – RESPIRATION	SOFA – respiratory component
SPO2	Oxygen saturation

COVID-19 Patient De-identified Dataset for Research



TEMPERATURE	Temperature
TIDAL VOLUME	Tidal volume
TIDAL VOLUME EXHALED	Tidal volume exhaled
VENT MODE – Adult	Mode of ventilator used
VENT MODE – Peds	Mode of ventilator used
WEIGHT	Weight

*The field will contain the actual LOINC code as recorded in EHR.

**The field will contain the actual CPT code as recorded in EHR.

***The field will contain the actual ICD code as recorded in EHR.

OBSERVATION

observation_source_value	Description
ICD code *	ICD code for diagnosis or procedure
ICU stay Y/N **	Whether the patient was admitted into an ICU during the encounter
LDA – intubation start and end times	Start or end of the intubation
Payer	All historic values for insurance type
SMOKING STATUS	Smoking status as recorded in EHR
VENT START - Adult	Date and time when mechanical ventilation started
VENT START - Peds	Date and time when mechanical ventilation started
VENT STOP - Adult	Date and time when mechanical ventilation stopped
VENT STOP - Peds	Date and time when mechanical ventilation stopped
Zipcode***	The first three digits of the zip code where the patient lived at the time of recording

*The field will contain the actual ICD code as recorded in EHR.

**value_as_number column contains length of stay in ICU during the encounter.

***value 1900-01-01 means that the date is missing.