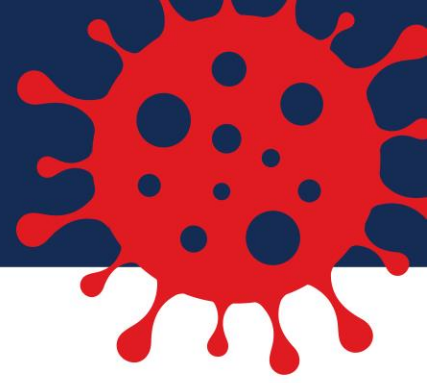


# COVID-19 Patient De-identified Dataset for Research



## 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.)

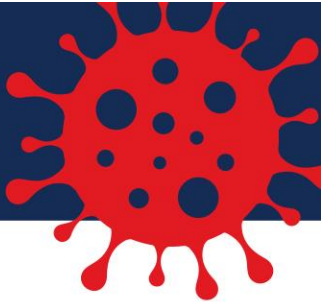
Updated: 11/12/20

Connect with us at:  
[IRBDataRequest@ahc.ufl.edu](mailto:IRBDataRequest@ahc.ufl.edu)

**UF** Clinical and Translational  
Science Institute  
UNIVERSITY of FLORIDA

**UFHealth**  
INFORMATION TECHNOLOGY

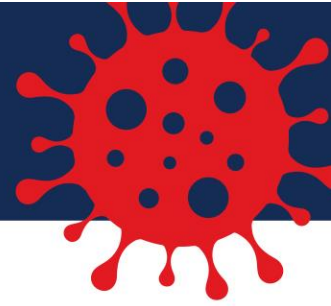
# COVID-19 Patient De-identified Dataset for Research



## CONTENTS

<b>OVERVIEW</b>	<b>3</b>	<b>CONDITION_OCCURRENCE</b>	<b>9</b>
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	<b>HEALTH SYSTEM DATA TABLES</b>	<b>16</b>
		LOCATION, CARE_SITE AND PROVIDER	16
<b>DATASET</b>	<b>4</b>	<b>VOCABULARY TABLES</b>	<b>16</b>
Inclusion Criteria	4	CONCEPT	16
Excluded Data	4	OTHER VOCABULARY TABLES	17
Data De-identification	4		
Identifiers	4	<b>APPENDIX A</b>	<b>18</b>
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
		<b>APPENDIX B</b>	<b>19</b>
<b>DATA DICTIONARY</b>	<b>6</b>	CONDITION_OCCURRENCE	19
<b>CLINICAL DATA TABLES</b>	<b>6</b>	MEASUREMENT	19
PERSON	6	OBSERVATION	21
DEATH	7		
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.

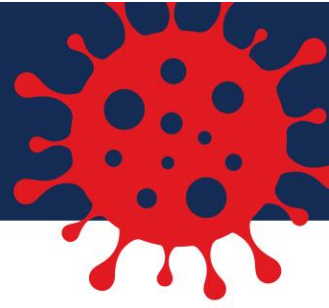
### Operational or Clinical Use of the Dataset

UF Health care teams interested in these data for operational or clinical purposes should contact the IDR team [IRBDataRequest@ahc.ufl.edu](mailto: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](mailto: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.

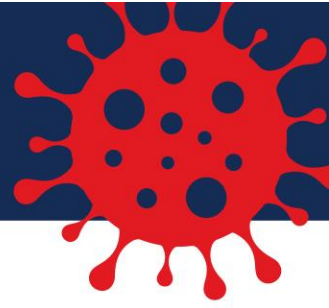
Note that data are provided as they exist in the UF Health clinical information systems (for example, the Epic electronic health record, or EHR) with limited edits or modifications. Researchers analyzing these data should recognize the inherent limitations of electronic clinical data for research. This 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



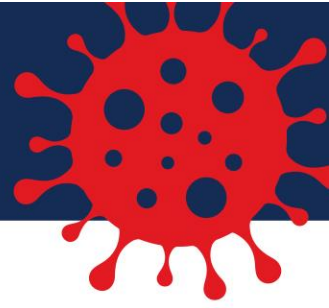
## Versions

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.

# COVID-19 Patient De-identified Dataset for Research



## 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.

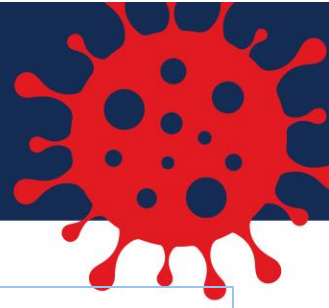
### CLINICAL DATA TABLES

#### PERSON

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

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

# COVID-19 Patient De-identified Dataset for Research



<b>month_of_birth</b>	Month part of birth_datetime	
<b>day_of_birth</b>	Day part of birth_datetime	
<b>location_id ***</b>	Populated with dummy values	
<b>provider_id ***</b>	Populated with dummy values	
<b>care_site_id ***</b>	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.

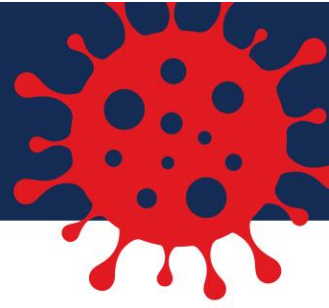
## DEATH

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

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

# COVID-19 Patient De-identified Dataset for Research



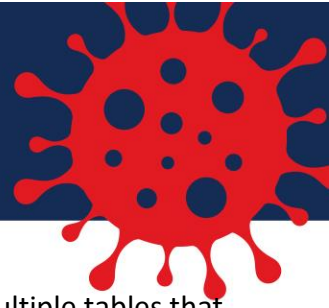
## VISIT\_OCCURRENCE

The *visit\_occurrence* table contains data about each encounter.

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

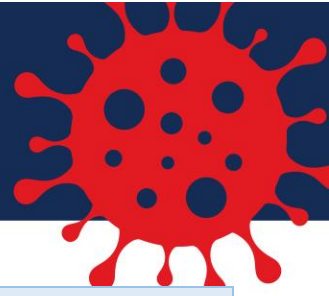
\*\*\*Dummy value is used to comply with requirements of OHDSI tools.

## 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
<b>condition_occurrence_id</b>	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
<b>person_id</b>	See description in person table	
<b>visit_occurrence_id</b>	See description in visit_occurrence table	
<b>visit_detail_id</b>	Not populated	
<b>condition_source_value *</b>	Code denoting the condition	
<b>condition_concept_id</b>	OMOP standard concept_id for specific condition	Derived based on condition_source_value
<b>condition_source_concept_id</b>	OMOP concept_id for specific condition; not necessarily standard concept_id	Derived based on condition_source_value
<b>condition_start_date **</b>	Start date of diagnosis	
<b>condition_start_datetime **</b>	Start date and time of diagnosis	If time is unknown, it is set to midnight
<b>condition_end_date **</b>	End date of diagnosis if available	
<b>condition_end_datetime **</b>	End date and time of diagnosis	If time is unknown, it is set to midnight
<b>condition_type_concept_id</b>	Source of information	Currently set to EHR
<b>stop_reason</b>	Not populated	
<b>condition_status_source_value</b>	Not populated	

# COVID-19 Patient De-identified Dataset for Research



<b>condition_status_concept_id</b>	Not populated	
<b>provider_id ***</b>	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.

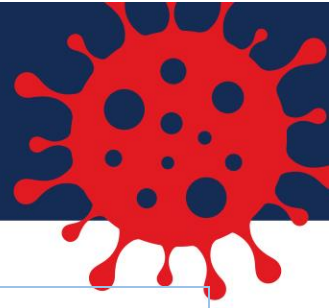
\*\*\*Dummy value is used to comply with requirements of OHDSI tools.

## 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
<b>procedure_occurrence_id</b>	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
<b>person_id</b>	See description in person table	
<b>visit_occurrence_id</b>	See description in visit_occurrence table	
<b>visit_detail_id</b>	Not populated	
<b>procedure_source_value *</b>	Code denoting the procedure	
<b>procedure_concept_id</b>	OMOP standard concept_id for specific procedure	Derived based on procedure_source_value
<b>procedure_source_concept_id</b>	OMOP concept_id for specific procedure; not necessarily standard concept_id	Derived based on procedure_source_value
<b>procedure_date **</b>	Start date of procedure	
<b>procedure_datetime **</b>	Start date and time of procedure	If time is unknown, it is set to midnight
<b>procedure_type_concept_id</b>	Source of information	Currently set to EHR
<b>quantity</b>	Not populated	

# COVID-19 Patient De-identified Dataset for Research



<b>modifier_source_value</b>	Not populated	
<b>modifier_concept_id</b>	Not populated	
<b>provider_id ***</b>	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.

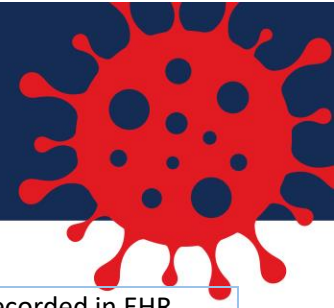
\*\*\*Dummy value is used to comply with requirements of OHDSI tools.

## DRUG\_EXPOSURE

The *drug\_exposure* table contains all medication orders.

Column Name	Description	Notes
<b>drug_exposure_id</b>	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
<b>person_id</b>	See description in person table	
<b>visit_occurrence_id</b>	See description in visit_occurrence table	
<b>visit_detail_is</b>	Not populated	
<b>drug_source_value *</b>	Code denoting the medication	
<b>drug_concept_id</b>	OMOP standard concept_id for specific medication	Derived based on drug_source_value
<b>drug_source_concept_id</b>	OMOP concept_id for specific medication; not necessarily standard concept_id	Derived based on drug_source_value
<b>drug_exposure_start_date **</b>	Start date for taking medication	If start date is not recorded in EHR, order date is used
<b>drug_exposure_start_datetime **</b>	Start date and time for taking medication	If time is unknown, it is set to midnight
<b>drug_exposure_end_date **</b>	End date for taking medication	If end date is not recorded in EHR, it is set to start date
<b>drug_exposure_end_datetime **</b>	End date and time for taking medication	If time is unknown, it is set to midnight

# COVID-19 Patient De-identified Dataset for Research



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

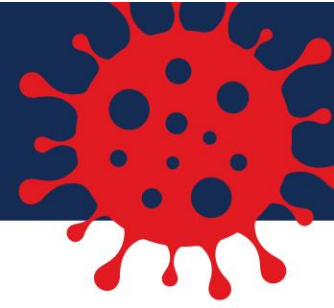
\*RxNorm code.

\*\*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

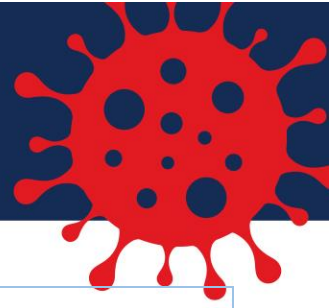


## MEASUREMENT

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
<b>measurement_id</b>	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
<b>person_id</b>	See description in person table	
<b>visit_occurrence_id</b>	See description in visit_occurrence table	
<b>visit_detail_id</b>	Not populated	
<b>measurement_source_value *</b>	Code denoting the measurement	
<b>measurement_concept_id</b>	OMOP standard concept_id for specific measurement	Derived based on measurement_source_value
<b>measurement_source_concept_id</b>	OMOP concept_id for specific measurement; not necessarily standard concept_id	Derived based on measurement_source_value
<b>measurement_date **</b>	Date of measurement	
<b>measurement_datetime **</b>	Date and time of measurement	
<b>measurement_time</b>	Time part of measurement_datetime	
<b>measurement_type_concept_id</b>	Source of information	Currently set to EHR
<b>value_source_value</b>	Result of measurement	
<b>value_as_number</b>	Result of measurement if the result is numeric	
<b>value_as_concept_id</b>	Not populated	
<b>operator_concept_id</b>	Not populated	
<b>range_low</b>	The lower bound of normal range of the measurement	Only populated for a subset of labs
<b>range_high</b>	The upper bound of normal range of the measurement	Only populated for a subset of labs

# COVID-19 Patient De-identified Dataset for Research



<b>unit_source_value</b>	The unit of measurement result	
<b>unit_concept_id</b>	OMOP standard concept_id for specific unit	Derived based on unit_source_value
<b>provider_id ***</b>	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.

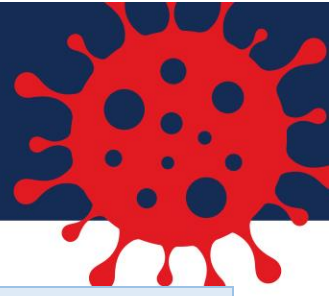
\*\*\*Dummy value is used to comply with requirements of OHDSI tools.

## 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
<b>observation_id</b>	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
<b>person_id</b>	See description in person table	
<b>visit_occurrence_id</b>	See description in visit_occurrence table	
<b>visit_detail_id</b>	Not populated	
<b>observation_source_value *</b>	Code denoting the observation	
<b>observation_concept_id</b>	OMOP standard concept_id for specific observation	Derived based on observation_source_value
<b>observation_source_concept_id</b>	OMOP concept_id for specific observation; not necessarily standard concept_id	Derived based on observation_source_value
<b>observation_date **</b>	Date of observation	
<b>observation_datetime **</b>	Date and time of observation	If time is unknown, it is set to midnight
<b>observation_type_concept_id</b>	Source of information	Currently set to EHR

# COVID-19 Patient De-identified Dataset for Research



<b>value_as_number ***</b>	Result of observation if numeric	
<b>value_as_string ***</b>	Result of observation if non-numeric	
<b>value_as_concept_id</b>	Not populated	
<b>qualifier_source_value</b>	Not populated	
<b>qualifier_concept_id</b>	Not populated	
<b>unit_source_value</b>	Not populated	
<b>unit_concept_id</b>	Not populated	
<b>provider_id ****</b>	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.

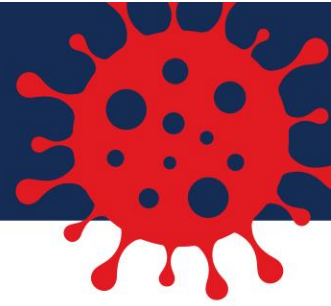
\*\*\*Each observation will have populated value\_as\_number or value\_as\_string column.

\*\*\*\*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>.

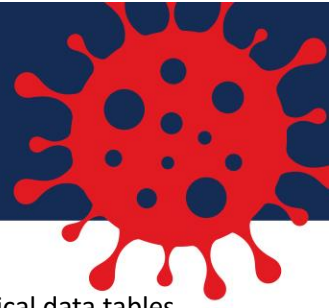
## CONCEPT

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

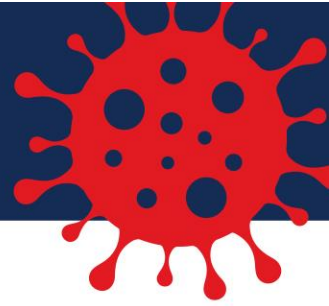


- **standard\_concept:** OMOP model requires use of standard concepts in certain columns of clinical data tables. These concepts are marked with an 'S' in standard\_concept column. If a concept of interest is not marked as standard, concept\_relationship table can be used to find the relevant standard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

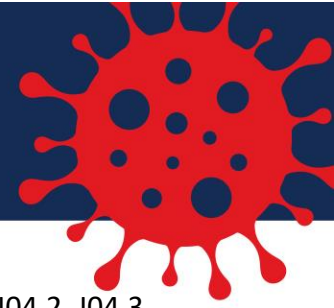
Any medication order that contains word 'Remdesivir' in its name.

### 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

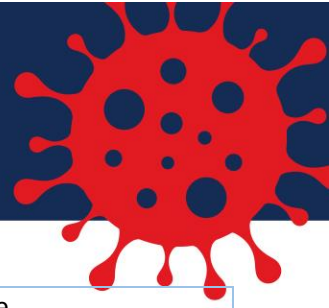
condition_source_value	Description
ICD9 or ICD10 code *	Diagnostic ICD9 or ICD10 code
Mechanical vent Y/N	Denotes whether mechanical ventilator was used during the encounter

\*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

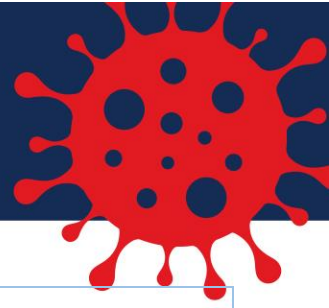
# COVID-19 Patient De-identified Dataset for Research



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



# COVID-19 Patient De-identified Dataset for Research



<b>TEMPERATURE</b>	Temperature
<b>TIDAL VOLUME</b>	Tidal volume
<b>TIDAL VOLUME EXHALED</b>	Tidal volume exhaled
<b>VENT MODE – Adult</b>	Mode of ventilator used
<b>VENT MODE – Peds</b>	Mode of ventilator used
<b>WEIGHT</b>	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
<b>ICD code *</b>	ICD code for diagnosis or procedure
<b>ICU stay Y/N **</b>	Whether the patient was admitted into an ICU during the encounter
<b>LDA – intubation start and end times</b>	Start or end of the intubation
<b>Payer</b>	All historic values for insurance type
<b>SMOKING STATUS</b>	Smoking status as recorded in EHR
<b>VENT START - Adult</b>	Date and time when mechanical ventilation started
<b>VENT START - Peds</b>	Date and time when mechanical ventilation started
<b>VENT STOP - Adult</b>	Date and time when mechanical ventilation stopped
<b>VENT STOP - Peds</b>	Date and time when mechanical ventilation stopped
<b>Zipcode***</b>	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.