Data Brief

Summary

CURE ID is a joint initiative between the US Food and Drug Administration (FDA) and the National Center for Advancing Translational Sciences (NCATS), a part of the National Institutes of Health (NIH). This online platform and mobile app enables clinicians and patients to share their real-world experiences treating and being treated with repurposed drugs. Users can report their repurposing experiences through the CURE ID case report form (CRF) from their computers, smartphones, or mobile devices. The platform promotes information sharing between healthcare professionals, researchers, pharmacists, and patients or caregivers to better inform treatment when there is a lack of adequate approved therapies. CURE ID aims to forge an efficient pathway to narrow the potential drug candidates for repurposing and prompt the development of targeted investigation of the most promising treatment options.

During the COVID-19 pandemic CURE ID, in partnership with the Cure Drug Repurposing Collaboratory[1] (CDRC), aimed to build a repository of the repurposed drugs used as COVID-19 treatment(s) to better understand the health outcomes with these treatments. Healthcare providers directly submitted their case reports to CURE ID and cases from the published literature were also added manually. However, the main source of cases for the repository were cases extracted from electronic health records (EHRs) of participating healthcare institutions and disease registries such as the Society of Critical Care Medicine (SCCM) Discovery Viral Infection and Respiratory Illness Universal Study (VIRUS). The automated extraction tool—the **EDGE Tool**[2]—was developed to extract data from different EHRs and registry systems, to convert them into the CURE ID format. This was made possible through partnerships with SCCM, Mayo Clinic, the Infectious Diseases Data Observatory (IDDO), Johns Hopkins University School of Medicine (JHU), and Emory School of Medicine.

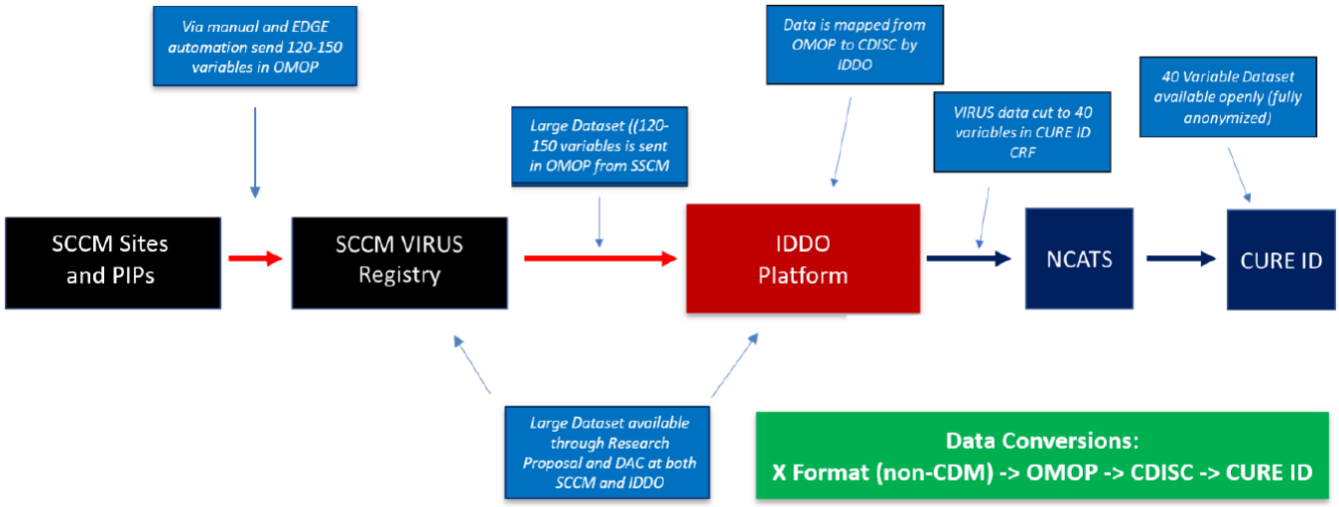
The project aimed to capture all cases of acute COVID-19 that required hospitalization. All inpatients at participating partner healthcare facilities beginning March 2020, with confirmed positive COVID-19 test within 14 days of admission were included. Outpatient encounters were not systematically captured. Patients with primary admission diagnoses related to trauma or surgery were excluded. Their COVID test was considered an ancillary finding rather than the cause of their hospitalization. Shift and truncate (SANT) de-identification was used to remove true calendar date information while still preserving temporal relationships. The displayed dates may be as early as 6 months before March 2020. All data collected in this project are completely de-identified and exclusively for the purpose of research.

The expanded CURE ID repository will ultimately house approximately 130,000 COVID-19 case reports. This large collection of cases from different sources of real-world data (registries, EHRs, clinician generated data, etc.) will help identify signals of potentially safe and effective off-label use of approved drugs. This data brief will highlight the data process and outline the summary statistics for the cohort population. The current version is based on the initial 12,129 cases that have made their way through the entire process to CURE ID. This data brief will be updated in Summer 2024 based on the newly acquired data.

Summary of Data Processing and Pathway

For participating institutions (sites) that do not use a common data model, the EDGE tool maps the data from EHRs to Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) and automates the extraction of the required data fields into the CURE ID case report form. Most sites in this category are existing SCCM VIRUS Registry sites. Once data has gone through the de-identification process, OMOP transformation, and data quality assessment, it is submitted to the Mayo Clinic or SCCM. This data, consisting of 150 variables, then flows to SCCM VIRUS for site de-identification variables for quality check, and then on to IDDO. IDDO converts the data from OMOP to Clinical Data Interchange Standards Consortium Study Data Tabulation Model (CDISC SDTM), so that it can be aggregated into the COVID-19 data platform at IDDO, which includes data from the International Severe Acute Respiratory and Emerging Infection Consortium (ISARIC). IDDO makes a subset of this data (40 variables) available to NCATS after having converted it to the CURE ID format. NCATS then incorporates the data into the CURE ID database. NCATS is also responsible for continued collection of cases manually extracted from the published literature, and those submitted by healthcare providers. See Figure 1.

Figure 1. Key Architecture Components Data Sources and Formats



Summary Statistics

Total Cases
12,129

Male	6,072	50.1%
Female	6,055	49.9%
Unknown	2	0.0%

Country	Count	Percent
USA	12,129	100.0%

Race	Count	Percent
White	7,857	64.8%
Black	3,546	29.2%
Other/Unknown	726	6.0%

Ethnicity	Count	Percent
Hispanic	748	6.2%
Non-Hispanic	11,381	93.8%

	Mean	Median	Min	Max	25%	75%
Age (in years)	60	64	0	99	48	76

Lab Measurement	Mean	Median	Min	Max	25%	75%
Leukocyte Count (K/uL)	10.7	9.2	0	305.7	6.3	13.1
Serum Creatinine (mg/dL)	1.7	0.98	0.2	37	0.75	1.65
Total Bilirubin (mg/dL)	0.74	0.5	0.1	36.4	0.4	0.8
ALT (IU/L)	59.4	31	3	12,403	19	55
AST (IU/L)	69.4	34	2	9,130	23	55
Respiratory Rate	24.3	24	4	52	19	30
Oxygen Saturation	94.5	95	30	100	92	98
LDH (IU/L)	439.3	372	72	7015	275	517
PT INR	1.8	1.3	0.7	17.9	1.1	2
PT (s)	19.9	15.9	10.2	120	14	22.6
Inhaled Oxygen Flow Rate (L/min)	20.7	15	0	101.2	4	40

Heart Rate	88.9	88	1	455	74	103
Body Temp. (°F)	97.6	98.6	19.1	106	97.9	99.5
Systolic Blood Pressure	122.1	120	39	266	106	136
Diastolic Blood Pressure	68.8	68	30	232	59	77

Co-morbidity	Count	Percent*
Hypertension	6,211	70.6%
Diabetes Mellitus	3,773	42.9%
Cardiovascular Disease	1,573	17.9%
COPD	1,401	15.9%
Other Chronic Lung Disease	1,275	14.5%

*Note that 8,792 (72.4%) of patients had a recorded associated co-morbidity

Medication	Count	Percent*
Dexamethasone	7,977	65.9%
Remdesivir	4,155	34.3%
Azithromycin	1,612	13.3%
Baricitinib	602	5.0%
Tocilizumab	176	1.5%

*Note that 12,105 (99.8%) of patients had a recorded medication received during their hospitalization

28-day Mortality	Count	Percent
Died	2,305	19.0%
Did Not Die	6,055	81.0%

	Mean	**Median**	**Min**	**Max**	**25%**	**75%**
LOS (days)	2.17	0	0	606	0	2

Oxygen Support Device	Count	Percent*
Oxygen Nasal Cannula	2,593	63.6%
High Flow Oxygen Nasal Cannula	931	22.8%
Intensive-care Ventilator	612	15.0%
Oxygen Mask	353	8.7%
Nasal Application Device Use with Positive Airway Pressure Device	205	5.0%

*Note that 4,078 (33.6%) of patients were on some type of oxygenation device

Visualization Description

The Exploratory Visualization Tool (EVT), accessible through the CURE ID app or website, is a user friendly interface designed to empower researchers, providers, and patients alike in exploring and gaining insights from EHR data related to patients hospitalized with acute COVID-19. This tool harnesses the wealth of information stored in EHRs to facilitate data-driven exploration of COVID-19 cases.

Bar charts can be used to visualize the number of COVID-19 patients with select demographic characteristics, comorbid conditions, or those receiving select drugs, and allows for comparison between groups.

Although the EVT displays real-world data, robust data security measures, such as de -identification techniques, have been employed to ensure compliance with privacy regulations and protect patient confidentiality.

Possible use cases for this tool include assisting in the design of clinical trials specific to COVID-19 by providing insights into patient populations and drug use, highlighting potential research gaps. Researchers may also use this tool to investigate healthcare disparities among COVID-19 patients by exploring data filtered on demographic factors, helping identify areas where healthcare access and outcomes may be unequal.

In summary, the EVT offers an innovative and user-friendly experience for researchers, patients, and providers to explore EHR data for patients hospitalized with COVID-19.

Appendix

1. CDRC is a public private partnership between the FDA, NIH NCATS and the Critical Path Institute (C-Path). The CDRC also includes four primary institutional partners: the Society for Critical Care Medicine (SCCM), Johns Hopkins University, the Infectious Disease Data Observatory (IDDO) at Oxford University, and Emory University.
2. The EDGE Tool is a series of resources to expedite the implementation of the OMOP (a data quality dashboard, and a platform for building exportable cohort definitions) data model's extraction, transformation, and loading (ETL) process. The EDGE Tool can assist in extracting data from discrete fields or those with defined ontologies (e.g., drop down menus, type ahead fields, or fields restricted to integers). Specific examples of discrete data include flowsheet rows in Epic (e.g., vital signs, nursing assessments), measurements (e.g., laboratory results) and certain past medical history fields and assessment forms. The EDGE Tool is not currently capable of extracting data from unstructured fields such as notes or reports (e.g., imaging results, history, and physical). The EDGE tool was developed by JHU to automate the extraction of data from different electronic health records (EHRs) and convert it into the OMOP format. JHU deployed the EDGE tool within a cohort of recruited sites through the SCCM VIRUS Registry and Emory and supported the institutional partners in this process.
3. A concept is a specific type of data captured in the EHR. In a common data model, concept IDs help standardize the process of how data is obtained, captured, and stored. It is a clear definition outlining acceptable methods or devices for capturing the measurements. For example, in OMOP, these different ways of measuring, capturing, and recording the patient's pulse are stored in concept IDs.