# CURE ID Final Report for ASPE
# September 30, 2024

## 1. Overview and Objectives

Real-world data generated during routine clinical care provides a promising source of information on potentially useful drugs for conditions which lack adequate treatment. COVID-19 has highlighted the need to learn from healthcare provider's treatment experiences. The ASPE funded CURE ID EHR and registry expansion was designed to leverage the existing CURE ID platform to expand the number of cases included and provide open access to data that is currently largely inaccessible to patients, policymakers, healthcare providers, and researchers.

Through the CURE ID expanded platform, healthcare providers, researchers, patients and their loved ones, policymakers and others interested in patient-centered outcomes research are able to have open access to comprehensive (de-identified) case reports on each patient treated for COVID-19 in participating hospital systems, including medications used and treatment outcomes such as recovery, deterioration, hospitalization, and death.

Expanding data capacity: The existing CURE ID platform is a clinical registry facilitating the study of the comparative safety and effectiveness of treatments for infectious diseases. Through the ASPE-supported expansion using automated extraction technologies, data was combined from disparate sources around the world (EHRs, other registries, clinician-submitted, and published cases), increasing capacity for patient-centered outcomes research. Partnerships with professional societies, academic medical centers, and non-profits helped to facilitate this data expansion. By focusing on a limited, high-value subset of clinical data available in EHRs and registries, automated extraction and aggregation of data in a common format from these varied sources became feasible.

Goal: The primary goal of this program was to identify promising COVID-19 treatments from the armamentarium of existing FDA-approved therapeutic products by expanding the existing infrastructure to enable collection from two additional sources into CURE ID: (1) EHRs and (2) clinical disease registries. This information was collected with the goal of helping FDA and NIH to work with sponsors, drug developers, policymakers, and other stakeholders through CDRC (the public-private partnership supporting CURE ID), to develop the scientific data needed to support new treatment indications for COVID-19 and other unmet medical needs. The program also aimed to address the needs of other audiences including healthcare providers, researchers, and patients facing diseases with no adequate therapy. The data provided by the program gives healthcare providers, patients, researchers, and policymakers insight into the use of existing drugs through the implementation of visualization tools within CURE ID.

The objectives of our proposal were to:

1. Develop partnerships with clinical consultants, technical consultants, and data providers to capture the most critical treatment data from EHRs and registries;

2. Build the infrastructure, technology, and methodology needed to extract and aggregate targeted clinical data from many global sources (EHRs, registries, and clinician-submitted and published cases);
3. Customize the CURE ID data fields to accommodate COVID-19-specific information available in EHRs;
4. Make large quantities of de-identified patient-level data on COVID-19 treatment from many different sources rapidly and openly available;
5. Expand clinician engagement and creation of global treatment networks;
6. Increase patient involvement in the platform;
7. Identify promising drugs, drug combinations or treatment regimens for COVID-19 and other diseases with inadequate therapy from the data in CURE ID;
8. Ensure the sustainability and use of the CURE ID Platform's EHR/registry expansion beyond COVID-19. This also includes open dissemination of the work products of the platform, such as providing open-source access to the code, so that it can be used by other interested parties, as well as sharing of the methodological learnings and the tools developed for automated extraction.

The key deliverables of the project were:

1. Four primary institutional partnerships providing drug repurposing data from more than 300 sites, organized through a partnership coordinating center managed by the CDRC;
2. Creation of an automated data extraction tool to capture targeted fields from diverse EHRs and registries globally;
3. Development of a data coordinating center for transformation and harmonization of the data;
4. An open platform for visualizing COVID-19 treatment data for tens of thousands of patients;
5. Expansion in the numbers of COVID-19 cases captured by the program (estimated to include ~250,000 cases by the end of the two-year grant period – later adjusted to an estimated 125,000-150,000 cases);
6. Increased number of healthcare providers and researchers visiting CURE ID and using the data;
7. Identification of drugs from the CURE ID database that appear promising for the treatment of COVID-19 and other diseases with inadequate therapies;
8. Use of the data generated in CURE ID to design clinical trials and drug development programs that may result in clinical guidelines or updated drug labeling.

## 2. Background - Problems Addressed

Background and Premise: The rapid emergence and exponential spread of COVID-19 had not given industry, governments, or research institutes time to develop de novo treatments. The treatment options available now are existing drugs that were originally developed for other indications. The medical community has scrambled to find potential treatments based on disease

manifestations, virologic characteristics, high-throughput screens, animal models and in-vitro and in-silico tests. Numerous treatments have been tried alone and in combination, including drugs with antiviral activity, anti-inflammatories, immune modulators, steroids, convalescent serum, anticoagulants, and others. Published research has resulted in limited identification of effective drugs for COVID-19 and the infected patients included in publications represent only a tiny subset of the collective global clinical experience.

Without a broad standardized platform to gather the experiences with different drugs, this chaotic treatment landscape will remain uninterpretable. Funding is needed to create such a platform and increase its size, while developing partnerships to obtain data and methods to automate the data extraction, to ensure its sustainability.

We have developed a mobile app and website to capture critical data fields in order to evaluate the major outcomes of treatment. This data is captured in an anonymized fashion as part of routine clinical care, not research, meaning patient consent is not required and privacy concerns are minimized.

The platform currently relies on the individual submission of case reports by healthcare providers (manually) and does not leverage the data in electronic health records. Relying on healthcare providers to enter cases in this context is infeasible due to demands on their time and large volumes of patients. EHRs and registries provide a large source of available data that can be probed electronically without requiring additional effort by healthcare providers, serving as a powerful resource to identify potential treatments for COVID-19, pending the development of new drugs.

Overall Strategic Vision: The overall strategic vision is to exploit prevalent information technology utilized in healthcare systems (i.e., EHRs and electronic registries) to access the global clinical practice environment. This clinical information will accelerate the identification of drugs with potential value for COVID-19 and other diseases without adequate therapies. The scope of diseases where this information will be important will include emerging infectious diseases, outbreaks, bioterrorist attacks, drug-resistant infections and neglected infectious diseases. The interactive features of the program will build international communities with expertise in the management of COVID-19 and other diseases. The public sharing of all data in the program will facilitate global efforts to find effective treatments. In addition, the public-private partnership will assist FDA and NIH in using hypotheses derived from this information to develop the scientific data needed to support approval for these indications.

## 3. **Methodology**

To achieve the goals of this program, a network of sites was developed (leveraging the Society of Critical Care Medicine's VIRUS Registry and other health system partners) to contribute electronic health record data on COVID-19 patients. This data was mapped to a common data model (OMOP) with the help of a suite of tools and guidance provided by experts. The de-identified cases were then shared in this common format with the Society of Critical Care Medicine (SCCM), who made the large dataset available for research purposes. The data was

further shared with the Infectious Diseases Data Observatory (IDDO) at Oxford University in the UK, who mapped the data from OMOP into CDISC SDTM (the data model typically used for clinical trial data). This enables it to be aggregated with other data and to be accessible to a different set of researchers and a more global community. Finally, the data was shared from IDDO to NCATS/NIH where a small dataset was made completely openly accessible via the CURE ID platform. Ultimately 130,000 COVID-19 cases were aggregated across 15 major health systems (102 hospitals) and made openly accessible with patient-level data on CURE ID. A data visualization tool was also developed to help CURE ID users explore the data: (https://cure.ncats.io/explore/ehr/1589).

## 4. Accomplishments by Final Deliverables

| End Product | Target Audience | How it is Used | Location of Product |
|---|---|---|---|
| CURE ID Database with 130,000 EHR cases of COVID-19 patients in common format | Healthcare providers, patients, general public, PCOR researchers, policymakers, drug developers | Used to explore treatments used for COVID-19 patients | https://cure.ncats.io/explore/ehr/1589 |
| Data Dictionary | PCOR Researchers | Provides guidance on the data elements included in the EHR extracted data, how they were defined, and the types of values | https://cure.ncats.io/assets/data-dictionary.pdf |
| Researcher's Guide | PCOR Researchers | Provides a summary of data processing and pathways, data sources and characterization, data access, data security, data quality assurance, and frequently asked questions for researchers to use to inform their use and exploration of the data in CURE ID | https://cure.ncats.io/assets/researchers-guide.pdf |
| Report on Outcomes of Importance to Patients | PCOR Researchers | May inform researchers about what measures are important to patients in evaluating their treatment outcomes, beyond death and length of | https://cure.ncats.io/assets/outcome-measures-of-importance-to-patients.pdf |

| | | | |
|---|---|---|---|
| | | hospitalization, that may be available in EHRs | |
| Publication 1: Identification, Verification, and Use of Respiratory Support FlowSheet Data for Observational Health Research Using the OMOP CDM | Researchers and scientists pursuing the use of OMOP and working with RWD | Paper may be used as a reference tool and knowledge generation | Journal of the American Medical Informatics Association (JAMIA) -<br>Status: publication has been cleared by FDA pending review of one summary table and is ready for submission |
| Publication 2: Impact of Immunomodulation Strategies on Critically Ill COVID-19 Patients Outcomes: A Comprehensive Retrospective Cohort Study | Researchers and scientists pursuing the use of OMOP and working with RWD | Paper may be used as a reference tool and knowledge generation | Primary: Journal of Intensive Care Medicine or<br>Secondary: American Journal of Respiratory and Critical Care Medicine<br><br>Status: development of this publication has been delayed as academic investigators leading it have been moving slowly; a complete draft is expected by mid-October and hoping to clear by late November |
| Publication 3: Real World Evidence in Action: Evaluating Dexamethasone's Effectiveness for COVID 19 | Researchers and scientists | Paper may be used as a reference tool and knowledge generation | Primary: JAMA<br>Secondary: Journal of Intensive Care Medicine or American Journal of Respiratory and Critical Care Medicine |

| | | | |
|---|---|---|---|
| through Target Trial Emulation | | | Status: Currently undergoing FDA review, expect to be submitted for publication by mid-October |
| Publication 4: Unveiling Sub-Populations in Critical Care Settings: A Real Word Data Approach in COVID 19 | Researchers and scientists | Paper may be used as a reference tool and knowledge generation | International Journal of Environmental Research and Public Health<br><br>Status: Currently undergoing FDA review, expect to be submitted for publication by mid-October |
| Software: Data Extraction Toolkit/Guidance Documents | PCOR Researchers | This open-source OHDSI tool stack can be used by researchers to facilitate conversion of proprietary EHR data to OMOP. Source code is available on Github. | https://github.com/OHDSI/CureIdRegistry |
| Data Brief | PCOR Researchers | This document will describe the data that researchers could access through the platform (e.g., description of the case reports published during the first year and subsequently at the end of the project, description of the SDOH and patient-reported data included in the database). | https://cure.ncats.io/assets/data-brief.pdf |
| Report on efficacy signals in CURE ID | Policymakers, PCOR Researchers, Healthcare providers, Patients and Family | The conference proceedings (508-compliant) of the meetings and workshops on data need discussions to confirm signals of drug efficacy generated in | See attached pdf (will be posted on CURE ID website under the EHR Resources in the next release). |

| | | | |
|---|---|---|---|
| | | CURE ID and recommended actions (including studies) to support guidance development on drug labeling. | |
| Resource Webpage on CURE ID | PCOR Researchers | The webpage will house information and products related to the project. This will include a Project brief (2-page document that is 508 compliant and cleared by the implementing agency) that can be used to orient people about what the project is trying to achieve and will be updated periodically to include accomplishments and availability of resources | https://cure.ncats.io/about (bottom of page under Related Links Section – planned to be moved to a separate page in the site in next deployment under new Resources section) |
| Patient Portal Added to CURE ID | Patients and Care partners | A module that has been added to the current CURE ID functionality that allows patients and their care partners to submit data to the platform directly. | Available throughout the CURE ID website and mobile app and for each disease the option to fill in a patient/care partner report. |
| Conference 1: CDRC Virtual Annual Meeting | CDRC Members and Stakeholders | Dissemination of CURE ID information | https://c-path.org/cdrc-accomplishments/ |
| Conference 2: SCCM Annual Meeting 2023 | Critical Care researchers and medical professionals, CDRC partners, RWD professionals | Dissemination of CURE ID information | https://c-path.org/cdrc-accomplishments/ |
| Conference 3: CDRC Annual Meeting 2023 | CDRC Members and Stakeholders | Dissemination of CURE ID information | https://c-path.org/cdrc-accomplishments/ |
| Webinar 1: CURE ID SCCM | SCCM and CDRC Members and Stakeholders | Dissemination of CURE ID information | https://sccmmedia.sccm.org/video/Webcast/Discovery/DIS |

| | | | |
|---|---|---|---|
| | | | CWEB22A.asp?bk=ok2 |
| Webinar 2: Q&A FDA Podcast | FDA and Healthcare professionals | Dissemination of CURE ID information | https://www.fda.gov/media/170476/download |
| Data Visualization Tool on CURE ID | PCOR Researchers, Healthcare professionals, Patients, Policymakers | This module that will be added to the current CURE ID functionality will allow users to view COVID-19 case reports and visualize aggregated data. Updates will be made to web and mobile application. | https://cure.ncats.io/explore/ehr/1589 |
| Final Project Report | | | |

## 5. Lessons Learned and considerations for future work

**Timing and Need for Underlying Infrastructure to Be Supported:**

- Time is of the essence. The faster you can move in the face of a pandemic, the more impact the work will have.
- Importance of having infrastructure in place ahead of pandemic (e.g., DUAs/DTAs, IRBs, funding mechanisms, contracts, etc.) -- it took so long to put all of these in place that by the time we got the data it's value had decreased substantially.
- This effort began in June 2021, more than a year after the pandemic started. By this time, many groups (e.g. N3C in June 2020) had already launched and pragmatic trial data had already been published (RECOVERY July 2020). The effort was initiated too late and we were playing catch up with groups who were better funded, more agile, and already had more infrastructure in place to build upon.
- Things inevitably take 3-10 times longer than you expect them to, so build that in and ensure that early deliverables are actually achievable in the specified timeframe. In this case, although the project was appropriately funded, it was a 5 year grant packaged in a two year deliverable. Although we were able to meet most deliverables, they were completed in year three and the breadth and quantity was reduced to meet realistic expectations.
- There was a lack of documentation for sites to implement the tools. This required the team to develop checklists, guides, and SOPs to help sites implement. However, this was not accounted for in the action items and needed to be retrofitted despite already tight timelines.

**Funding (including what it could help address and what it couldn't):**

- It is difficult and sometimes impossible to compete with other priorities of sites, particularly during COVID surges, as their staff's attention must go to patient care and operation of systems to support this. No amount of funding could overcome the limited bandwidth of existing staff. In many cases it was just not possible to hire skilled staff due to a shortage in the workforce.
- The amount of funding per site was not commensurate with the amount of effort ultimately required. However, in some cases even increased funding was not able to move projects as the informatics team on site were being tapped out and there was a lack of skilled workers to pick up the pace.
- Longer-term, sustained funding is necessary to have continued engagement from sites.

**Technical Expertise and Insight into the Development Stage of the Tool:**

- Having greater insight into how advanced the EDGE tool was early on would have been helpful, it was not nearly as developed as initially thought based on the information provided by JHU. As a result, significant effort was required to program and develop the tools and sites were not able to begin implementing the platform for a considerable period of time while it was undergoing further development.
- The lack of a data dashboard to perform data quality checks required data to be vetted though C-Path which came with added costs and time.

**Other Partners:**

- Consensus-building and getting feedback from many stakeholders is valuable, but also can be extremely time-consuming. Using a smaller group of stakeholders to make initial decisions and then seeking input on those decisions from a larger group within a pre-specified time frame might have been more efficient.
- A partner who participated peripherally but could have made more impact was the OHDSI community. We tried to engage them on several occasions but never got the kind of intense collaboration that would have been helpful to move the project and the field forward.
- The quality of the partners really matters, both in terms of their technical capabilities, and how easy they are to work with.
- Having existing IAAs and cooperative agreements in place to move the money from FDA to NCATS and from FDA to C-Path were critical. Then C-Path's flexibility in contracting with PIPs was essential.
- Finding the right partners at a site can be challenging and may involve the need to identify multiple people with different expertise – e.g., clinical PI versus informaticist.

- Finding champions at sites provided the best opportunity for moving projects forward.
- Ensuring that there is a strong shared belief in the mission of the project is important. It's fine for groups to get ancillary benefits or to have their own motivations that align with the project more broadly, but there should be a core belief in and support of the central endeavor. It also must be clear that the deliverables for the grant are determined by the success of the central project.

**Strategic Decisions on Scope and Methods:**

- More complex data flows add significant delays to receipt of data and the benefits of having the data in different places and different formats should be weighed against the speed with which you can get the data. In retrospect, having EHR data sent from SCCM to CURE ID and having SCCM share the data with IDDO after, would have been better than having IDDO send the data to CURE ID. This would have made the process faster (no need for us to wait for curation to CDISC) and more sites would have agreed to the DUA/DTA that needed to be negotiated and put in place. Building in flexibility to agreements in terms of data flow could be helpful, to enable such changes to the flow of information, even if it ultimately ends up in the same places, as delays and road-blocks become apparent over time.
- There are advantages to doing the work yourself/with your immediate team, versus having sites do the work. However, having the sites do the work builds their capacity, which was the intent of developing these tools at the sites.
- Inclusion of CERNER sites (with the goal of demonstrating that the tool could be system agnostic – e.g., not only work with Epic) proved problematic and required additional consultants and costs to get the tool over the finish line by rewriting code specific for the sites to implement the tool.
- Lack of clarity around scalability regarding what data is mapped from the EHR and what was not, as well as the inability to include unstructured data elements (clinician notes, pathology reports, radiology reports…) was a major limitation.
- Limited interest from sites to explore opportunities beyond COVID, a required deliverable for ASPE was a challenge. Although FDA proposed meningitis, there was no regulatory roadmap developed to help guide the effort. Greater interest was found in the sites to pursue something like sepsis, given it's substantial public health impact and cost to institutions, however, the timeline was too tight to fully flush this out and the lack of dedicated funding over a longer period proved to be a substantial obstacle.

**Patient Involvement:**

- The needs of patients and those with lived experience when it comes to engagement and contracting needs to be better thought through ahead of time and support should be provided (e.g., what are the expectations for how they will bill or invoice, what

participation do they need to record, what documentation will they need to provide for payment, etc.).

- We struggled to engage patients in the acute COVID space, likely because by the time we got to the point where patient input was relevant to the specific deliverables being worked on, pandemic fatigue had set in in full force. We were, however, very successful in engaging the patient community with Long COVID and other infection-associated chronic conditions. This ended up being one of the most fruitful parts of the project. They participated with passion and enormous insight in the patient advisory group (aided significantly by the modest funding from ASPE to support their time and effort) and helped provide extensive feedback on expanding the platform to be more accessible to patients and care partners both in exploring the data and in developing case report forms with language appropriate for patients that could be entered directly.
- The support from ASPE helped to launch us into the space of patient-reported cases as we thought more about the data and outcomes that would be most patient-centered. We expanded the whole CURE ID platform to enable patient and care-partner submitted cases to be added for CURE ID for any disease. In addition, we focused significantly on a couple of new therapeutic areas where we felt patients and care partners would be uniquely positioned to participate (e.g., Long COVID, RASopathies, Sarcomas, etc.).
- We also have focused attention on how the data available can be most useful for patients, care partners, and clinicians to use through the Explore feature and ways that we can share the findings more broadly through webinars, social media, etc.
- We have increased our registered user base from ~1800 registered users to >3300, the majority of whom we believe are patients, primarily interested in Long COVID. Non-registered users can view the data and we have been reaching up to 3800 users a month to the CURE ID platform, whereas previously it averaged 200-400 users a month.

**Opportunity for use of the EDGE Tool for Diseases Beyond COVID-19:**

- While there is significant potential for the EDGE tool to be utilized for diseases beyond COVID-19, challenges exist to its implementation.
- The CURE ID EDGE Tool would be easiest to adapt to conditions that are similar to COVID-19 (particularly COVID-19 as it presented early on in the pandemic) --
    - An illness requiring hospitalization, typically over a single hospital stay
    - Has a clear means of diagnosis that is captured routinely and consistently in the EHR (e.g., a single lab measure like a positive PCR)
    - Has a hard outcome, preferably mortality (and a relatively high fatality rate); hospital length of stay may be a reasonable substitute in some situations, but it is not well-suited for use in a pandemic where length of hospitalization is heavily tied to resource constraints and the influx of new patients.
    - The variables of greatest clinical significance are structured variables in the EHR that are easily accessible

        - This becomes a substantial challenge for many infectious diseases that were explored as the diagnosis often requires send-outs to an external lab, therefore culture reports may only be accessible as scanned PDFs, which are both unstructured and may not be immediately machine-readable.
    - Illnesses requiring admittance to the ICU are conducive for inclusion in this platform as often the frequent lab and physiologic measures can be used for analysis or as measures of confounders, etc..
    - Sepsis or acute respiratory distress syndrome (ARDS) were identified as the conditions requiring the least change from the existing tool (e.g., for sepsis, it was believed that three additional variables would be needed on top of the existing COVID-19 variables). While sepsis is of interest to the academic partners, it was not thought to be an example where there was likely to be any sort of regulatory path forward. The condition is extremely heterogenous and several decades of RCTs for sepsis therapeutics have been unsuccessful.
    - Meningitis was of much greater interest from a regulatory perspective, as there are few approved therapeutics for many of the rarer causes of meningitis, and in some circumstances it would not be possible to conduct an RCT for ethical or practical reasons, making the potential value of the observational data in CURE ID more substantial. However, engagement with several meningitis experts found that they thought a large number of additional variables would be required, including many that are not structured in the EHR or would be difficult to access (like organism from culture reports).
    - Ultimately, it was felt that the sweet spot for use of the EDGE tool would probably be another outbreak causing significant public health concern (although it wouldn't have to be respiratory, it could be of meningitis, encephalitis, or any number of other presentations). This situation would result in the variables of greatest importance being reported in a structured way, relatively quickly, and would excite the necessary partners to come to the table to engage.
    - The need for a platform that was "kept warm" during non-outbreak times, was frequently mentioned. This is where using the existing tool for sepsis or ARDS might make sense, so that it could be rapidly pivoted to the cause of a future outbreak. This way the agreements, technical infrastructure, etc., would all be in place and could be scaled up quickly. However, there are substantial costs associated with ongoing support of this kind of platform, and currently the resources are not available to maintain it.