NIH launches analytics platform to harness COVID-19 patient data to speed treatments

June 16 2020

The National COVID Cohort Collaborative aims to transform clinical information from across the country into knowledge urgently needed to study COVID-19, including health risk factors that indicate better or worse outcomes of the disease, and identify potentially effective treatments. Credit: National Center for Translational Sciences

The National Institutes of Health has launched a centralized, secure enclave to store and study vast amounts of medical record data from people diagnosed with coronavirus disease across the country. It is part

of an effort, called the National COVID Cohort Collaborative (N3C), to help scientists analyze these data to understand the disease and develop treatments. This effort aims to transform clinical information into knowledge urgently needed to study COVID-19, including health risk factors that indicate better or worse outcomes of the disease, and identify potentially effective treatments.

The N3C is funded by the National Center for Advancing Translational Sciences (NCATS), part of NIH. The initiative will create an analytics platform to systematically collect clinical, laboratory and diagnostic data from health care provider organizations nationwide. It will then harmonize the aggregated information into a standard format and make it available rapidly for researchers and health care providers to accelerate COVID-19 research and provide information that may improve clinical care. A demonstration of the platform can be viewed at ncats.nih.gov/n3c.

Having access to a centralized enclave of this magnitude will help researchers and health care providers answer clinically important questions they previously could not, such as, "Can we predict who might need dialysis because of kidney failure?" or "Who might need to be on a ventilator because of lung failure?" and "Are there different patient responses to coronavirus infection that require distinct therapies?"

"NCATS initially supported the development of this innovative collaborative technology platform to speed the process of understanding the course of diseases, and identifying interventions to effectively treat them," said NCATS Director Christopher P. Austin, M.D. "This platform was deployed to stand up this important COVID-19 effort in a matter of weeks, and we anticipate that it will serve as the foundation for addressing future public health emergencies."

Data access will be open to all approved users, regardless of whether

they contribute data. The data are being provided to NCATS as a Limited Data Set (LDS) that retains only two of 18 HIPAA-defined elements: health care provider zip code and dates of service.

NCATS, which is serving as stewards of the data, is taking multiple security and privacy measures. For example, NCATS oversees the use of N3C through user registration, federated login, data use agreements with institutions and data use requests with users. The data reside and remain in NCATS' secure, cloud-based database certified through the Federal Risk and Authorization Management Program, or FedRAMP, which provides standardized assessment, authorization, and continuous monitoring for cloud products and services ensuring the validity of the data while protecting patient privacy. Approved users must analyze data within the platform. In addition, the N3C data will be used only for COVID-19 research purposes, including clinical and translational research and public health surveillance.

The information available via the N3C enclave will be rich in scope and scale. There currently are 35 collaborating sites across the country and the platform contains diverse data from individuals tested for COVID-19. A key component is the harmonization of data, which translates the different ways that contributing hospitals store patient data into a single, common format to enable combined 'apples to apples' analyses. Contributing sites add demographics, symptoms, medications, lab test results, and outcomes data regularly over a five-year period, enabling both the immediate and long-term study of the impact of COVID-19 on health outcomes.

The platform is built to enable machine learning approaches and rigorous statistical analyses, identifying connections and patterns more quickly than can be done through traditional methodologies. These advanced analytics approaches require large, robust datasets to generate statistically valid results and can lead to the simultaneous exploration of

multiple questions—and the revealing of likely answers—on a powerful scale.

"The exciting transformation this platform represents is in providing an environment where data and the power of the analytics can be used by researchers and clinicians to quickly examine and answer new COVID-19 hypotheses," said Warren A. Kibbe, Ph.D., chief of Translational Biomedical Informatics in the Department of Biostatistics and Bioinformatics and chief data officer for the Duke Cancer Institute.

The N3C harnesses the extensive resources of the NCATS-funded Clinical and Translational Sciences Awards (CTSA) Program and its Center for Data to Health (CD2H). "By leveraging our collective data resources, unparalleled analytics expertise, and medical insights from expert clinicians, we can catalyze discoveries that address this pandemic that none of us could enable alone," said Melissa Haendel, Ph.D., director of CD2H at the Oregon Health & Science University School of Medicine and Director of Translational Data Science at Oregon State University.

Provided by National Institutes of Health