National COVID Cohort Collaborative (N3C)

The National Covid Cohort Collaborative (N3C) (link: https://covid.cd2h.org/about) is an initiative launched by NIH/NCATS in response to the COVID-19 pandemic. Its main aim is to accelerate scientific discovery and provide a tool for collaboration on COVID-19 research by creating an aggregated and harmonized repository of electronic health record (EHR) data from COVID-19 patients across the country.   This data enclave represents one of the largest secure collections of harmonized clinical health data in the United States. Importantly, while data submitted to the enclave represents four different data models (i.e., OMOP, PCORNet, ACT, and TriNetX), the underlying data model used to aggregate, standardize, and harmonize the entire N3C enclave data set is the open-source internationally supported OMOP common data model. To date, N3C contains EHR data on over 18.4 million people of which there are almost 7.3 million COVID-19 positive patients, with over 23.4 billion rows of clinical data.

The N3C patient cohort objective is defined as lab-confirmed, suspected, and possible cases of COVID-19 who are then demographically matched (on age group, sex, race and ethnicity) to patients who have tested negative/equivocal for COVID-19, at a ratio of 1:2 (cases:control). In addition to EHR derived data, N3C records are being linked, via privacy-preserving record linkage (PPRL), to other data sources (e.g., Medicare, Medicaid, All of Us) to enrich the EHR derived data and to de-duplicate cohort counts. The N3C library also has more than 30 external data sets ranging from mortality data to viral variance data to environment data. N3C Dashboards and a Data Overview can be explored here: https://covid.cd2h.org/dashboard/

Around the country, more than 76 sites are submitting data to the N3C, over 450 research projects are using N3C data, and more than 100 published manuscripts have used N3C data. The MetroHealth System contributes data to N3C, and University Hospitals will do so by mid-2023.

Request for researcher access to the N3C data sets can come in 3 different forms - limited data sets, de-identified data sets, and synthetic data sets.   Researchers in the CTSC community at Metro, CWRU, and UH can propose studies that utilize the N3C data enclave by registering for access and completing a Data Use Request (DUR). In addition, local IRB approval will be required for studies proposing access to de-identified and Limited Data Set access prior to submission of the DUR to N3C. A link where researchers can get started is here: https://covid.cd2h.org/for-researchers, however, CTSC informatics representatives will be able to provide guidance on all these steps. Please email David Kaelber dkaelber@metrohealth.org or Mark Beno mark.beno@case.edu for further information and guidance on getting started. Finally, the CTSC informatics will be hosting an N3C workshop to be scheduled shortly after University Hospitals begins submitting data to the N3C enclave. Details on this will be provided in an upcoming newsletter and also emailed out to CTSC members.