Last Updated: June 14, 2022

National COVID Cohort Collaborative

N3C Short Course Pilot

Tuesday, June 14, 2022 to Friday, June 17, 2022
2:00pm-5:00pm ET (Zoom)

Organizers/Instructors*

- Jerrod Anazalone, MS (University of Nebraska) *
- Will Beasley, PhD (University of Oklahoma) *
- Liz Chen, PhD (Brown University)
- Karen Crowley, MS, PhD (Brown University)
- Harold Lehmann, MD, PhD (Johns Hopkins University) *
- Johanna Loomba, ME (University of Virginia)
- Mary Helen Mays, PhD (University of Puerto Rico)
- Shawn O'Neil, PhD, MS (University of Colorado) *
- Andrea Zhou, ME (University of Virginia) *

Participants

Faculty, staff, and students from: Brown University, Johns Hopkins University, Oklahoma University Health Sciences Center, University of Nebraska Medical Center, University of Puerto Rico, and University of Virginia.

Description

The National COVID Cohort Collaborative (N3C) is an open science community focused on analyzing patient-level data from many clinical centers to reveal patterns in COVID-19 patients. Over 10 billion rows of data are stored in the secure N3C Data Enclave in the OMOP common data model for Electronic Health Records. The Enclave (Palantir Foundry) supports analysis with SQL, R, and Python for conducting statistical and machine learning analyses, backed by the distributed-computing framework Apache Spark.

This short course introduces researchers (students, staff, and faculty) to N3C and will include hands-on experiences in computing, exploring and analyzing EHR data, and good research practices in the context of N3C. The knowledge and skills obtained through this course will prepare participants for developing their own N3C studies and understanding secondary use of de-identified EHR data in general.

Overview

Session	Name	Topics
Prerequisites (Asynchronous)	**Computing Setup** (by Tuesday, June 7) **Computing Skills** (by Tuesday, June 14)	• N3C Account • Analytic Tools: SQL; Python or R
	Background Knowledge (by Tuesday, June 14*) ** Materials (readings, videos, etc.) can be reviewed before, during, or after the course*	• N3C Primer • EHR Primer • Observational Research
Session 1 (Synchronous)	**Study Design Overview & Working in N3C** Tuesday, June 14 2:00-5:00pm ET	• Study Design & Planning • Foundry & Knowledge Store • Concept Sets & Code Workbooks
Session 2 (Synchronous)	**Office Hours & Open Discussion (optional)** Wednesday, June 15 2:00-4:00pm ET	• Student and Project Introductions • Troubleshooting • Topic Discussion
Session 3 (Synchronous)	**Computable Phenotyping & Data Extraction** Thursday, June 16 2:00-5:00pm ET	• Team Science & Collaboration • Analytic Pipeline • Computable Phenotypes & Data Extraction
Session 4 (Synchronous)	**Data Analysis & Results Reporting** Friday, June 17 2:00-5:00pm ET	• Data Quality & Validation • Data Analysis & Visualization • Dissemination & Reproducibility

Last Updated: June 14, 2022

!!! Important Reminders !!!

- **N3C Data Enclave Institutional Data Use Agreement (Confidentiality and Data Security)**
  - **Users agree to not attempt to re-identify or contact any individuals** who are the subjects of the Data or any known living relatives unless required by law to maintain public health and safety.
  - **Users agree not to attempt to use the Data to identify or contact any Data Contributors or healthcare providers** unless such identification is needed for Data preparation and management purposes and only at the request of NCATS or required by law to maintain public health and safety.
  - **Users agree to not photograph, create screenshots, nor download data** viewed on the NCATS N3C Data Enclave. Furthermore, Users' access to the NCATS N3C Data Enclave will be terminated after 1 year. Users' access to their analyses and findings on the NCATS N3C Data Enclave may continue with DAC approval of DUR renewal.
- **N3C Code of Conduct**
  - **Make no attempt to identify** institutions, communities or identifiable populations associated with particular N3C data; to re-identify or contact individuals, their relatives or relevant groups from whom data represented within the N3C Data Enclave were collected; or to generate information that could allow individual identities to be readily ascertained.
  - **Maintain the confidentiality of the data** accessible within the enclave and not distribute, provide access to, attempt to download or otherwise capture views of N3C data for use or disclosure to any entity or individual beyond those specified in the approved DUR.
- **No** recording, hand copying and pasting, photographs/screenshots (or emailing of screenshots), verbal description of data, or sharing of access to data
  - **Everything should stay within the N3C Data Enclave (we can look at any work in there)**
- Only share screen with those with same level access DUR (e.g., Level 2 de-identified) for in person or virtual (e.g., Zoom) meetings
- Be sure to lock your computer whenever you step away, even for a short period of time.
- **If not sure about something, ask the course organizers/instructors first!**

Materials and Communication

- Materials will be distributed through the course GitHub repo: https://github.com/National-COVID-Cohort-Collaborative/short-course-2022-june

- Communication will be done using the `#n3c-short-course` Slack channel in the CD2H workspace:
  - Use for troubleshooting and questions during class (rather than Zoom chat)
    - **Do not take or send photographs/screenshots or copy data (everything should stay within the N3C Data Enclave - we can look at any work in there)**
  - Use for questions outside of class (avoid direct messages)

Last Updated: June 14, 2022

Detailed Schedule

Prerequisites

N3C Registration	• N3C account registration: https://covid.cd2h.org/onboarding
	• Data enclave and access requirements: https://covid.cd2h.org/enclave-checklist
	  ○ *Notes: These can take some time to complete; disregard section about Data Use Request (DUR)*
	  ○ NIH Information Security and Information Management Training
	  ○ Human Subjects Research Protection Training (e.g., CITI)
	  ○ ORCID iD

DUR Request	• Once your N3C account has been created, please join the DUR (Data Use Request) for the project ("Malnutrition and COVID-19 Outcomes") that will be used for the course using the following link: https://unite.nih.gov/workspace/slate/documents/collaborator-data-use-request?DUR=DUR-8B67B2D • Both an N3C account and this DUR access are needed starting on the first day of the course. These can take some time to approve, so please complete asap.

Slack Channel Access	• Join the CD2H Slack Workspace: https://cd2h.slack.com/join/signup#/domain-signup
	  ○ Once you join, we will add you to the #n3c-short-course Slack channel (this will be used for communication before, during, and after the course).
	  ○ If you could then send us a message in the channel confirming DUR approval (should receive an email), that would be great!

DUR Workspace Access	• Navigate to the workspace for the "Malnutrition and COVID-19 Outcomes" project: https://unite.nih.gov/workspace/compass/view/ri.compass.main.folder.2091276a-fa02-4612-b939-6952082ecfa2
	• If your access **has** been granted
	  ○ You should see a workspace with a list of folders in the middle and are all set!
	  ○ In the Slack channel (`n3c-short-course`), please confirm by sending us the message: "DUR Workspace Access = Yes"
	• If your access **has not** yet been granted
	  ○ It will say "This project is private. You must have access to view or edit it. Request access."
	  ○ In the Slack channel (`n3c-short-course`), please send us the message: "DUR Workspace Access = No"

Background Materials	Review the following background materials
	• Required
	  ○ N3C Orientation Session A: https://covid.cd2h.org/training
	  ○ Introduction to use of Structured Medical Record Data: https://www.youtube.com/watch?v=qrNPjBBpngE
	• Supplemental Materials

	○ Haendel MA, Chute CG, Bennett TD, et al. The National COVID Cohort Collaborative (N3C): Rationale, design, infrastructure, and deployment. J Am Med Inform Assoc. 2021;28(3):427-443. doi:10.1093/jamia/ocaa196 ○ Bennett TD, Moffitt RA, Hajagos JG, et al. Clinical Characterization and Prediction of Clinical Severity of SARS-CoV-2 Infection Among US Adults Using Data From the US National COVID Cohort Collaborative. JAMA Netw Open. 2021;4(7):e2116901. Published 2021 Jul 1. doi:10.1001/jamanetworkopen.2021.16901 ○ Ehrenstein V, Kharrazi H, Lehmann H, et al. Obtaining Data From Electronic Health Records. In: Gliklich RE, Leavy MB, Dreyer NA, editors. Tools and Technologies for Registry Interoperability, Registries for Evaluating Patient Outcomes: A User's Guide, 3rd Edition, Addendum 2 [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2019 Oct. Chapter 4 Obtaining Data from Electronic Health Records. Available from: https://www.ncbi.nlm.nih.gov/books/NBK551878/ ● Learn basic analytic tools (recommended if new to SQL, Python, or R) ○ https://github.com/National-COVID-Cohort-Collaborative/short-course-2022-june/blob/main/background/analytical-tools.md

Session 1: Study Design Overview & Working in N3C

Tuesday, June 14 2:00-5:00pm ET

Agenda	● Introduction & DUR Access Test → **Liz/Jerrod** ● Study Design & Project Planning → **Harold** ● N3C Navigation (Part 1) → **Shawn** ○ Foundry ○ Directory Setup ● Break ● N3C Navigation (Part 2) → **Shawn** ○ Concept Set Curation ○ Code Workbooks ○ Knowledge Store ● Wrap Up → **Liz**
Materials	● Class materials can be found in the course GitHub ○ https://github.com/National-COVID-Cohort-Collaborative/short-course-2022-june ● "Malnutrition and COVID-19 Outcomes" Project Workspace ○ https://unite.nih.gov/workspace/compass/view/ri.compass.main.folder.2091276a-fa02-4612-b939-6952082ecfa2
Assignment **Due Thursday, June 16, 2022**	● Take Home Exercise ○ https://github.com/National-COVID-Cohort-Collaborative/short-course-2022-june ● Logic Liaison (LL) Templates ○ Watch recording from last Friday's N3C Enclave Users Group on Logic Liaison Updates: New Data Quality and Patient Fact Templates

	○ Review documentation (README) for ■ LL Template for L2 and L3 Fact Tables: All Patients ■ L2 and L3 Fact Tables: COVID-19 Diagnosed or Lab Confirmed Patients ● Malnutrition Study ○ Review Ponce J, Anzalone AJ, Bailey K, Sayles H, Timmerman M, Jackson M, McClay J, Hanson C; National COVID Cohort Collaborative (N3C) Consortium. The Impact of Malnutrition on Clinical Outcomes in Patients Diagnosed with COVID-19. JPEN J Parenter Enteral Nutr. 2022 Jun 7. doi: 10.1002/jpen.2418. Epub ahead of print. PMID: 35672915. ● Syntactic and Semantic Standards ○ Intro to OMOP and Contour (review first 19 minutes for OMOP) ○ OMOP Vocabulary 101 ● Supplementary Materials ○ Book of OHDSI Chapter 4 (The Common Data Model) ○ Book of OHDSI Chapter 5 (Standardized Vocabularies)

Session 2

Agenda	● Brief Introductions ● Office Hours ● Open Discussion (Session 1 Review & Background Discussion) ● Special Topics (from Session 1)
Materials	● N/A
Assignment	● See Session 1