Tribal Consultation Informational Webinar: National COVID Cohort Collaborative (N3C)

Joni L. Rutter, Ph.D.

Acting Director, National Center for Advancing Translational Sciences
Purpose of Webinar

• Describe the NCATS N3C electronic health records (EHR) data resource, why it was created, and how it can be used
• Describe how N3C handles AI/AN data while awaiting a Tribal Consultation.
• Start a dialogue for considering a path forward for use of tribal data for public health research on COVID-19, and to work together on policies and governance on the use of this data, including methods for ongoing Tribal Consultation and governance
• Understand Tribes’ perspectives on benefits/risks of N3C
  • Consider how N3C can directly benefit Tribal communities
  • Consider both individual and community privacy concerns
To catalyze the generation of innovative methods and technologies that will enhance the development, testing and implementation of diagnostics and therapeutics across a wide range of human diseases and conditions.
When the pandemic began, there were so many questions and few answers...

Which **comorbidities** (e.g., diabetes, heart disease, cancer, hypertension) increase the risk for severe COVID-19?

What are the **co-morbidities**?

What are COVID-19 symptoms?

What are best treatments for COVID-19?

How does it **spread**?

Why are symptoms so **variable**?
What types of information can be found in an Electronic Health Record (EHR)?

- Medications
- Medical History
- Immunization Dates
- Diagnoses
- Allergies
- Billing Data
- Patient Demographics
- Radiology Data
- Lab Results
- Vital Signs
Translational Research Need: Making EHRs Interoperable and Accessible for Research

- In the U.S., we do not have a standard process to collect and manage electronic health record (EHR) data
- No standard way to use patient EHRs for research or help make or inform public health decisions using near real-time data
What is the NCATS National COVID Cohort Collaborative (N3C)?

- The N3C is collecting data from EHRs, harmonizing them into a single format and making them available to researchers in the N3C research platform, the N3C Data Enclave. The data are not consented.
- Researchers cannot take data out of the Data Enclave. It is a secure platform for the data, and it provides tools that let researchers collaborate within it.
  - Data are refreshed on a regular basis to capture COVID-19 as it evolves, including the introduction of viral variants.
Harnessing resources of the Clinical and Translational Sciences Awards (CTSA) Program institutions, CTRs in IDeA States, and the Center for Data to Health (CD2H), the goals are to:

- Generate a secure, national resource of electronic medical record data from COVID-19 tested patients
- Make available real-world clinical data for speeding COVID-19 research and improving patient care
NIH Clinical Research Networks and their partners: Leveraging Real World EHR Data to Research COVID-19

Sites: 67

- Persons: 9.5 million
- COVID+ Cases: 3,312,134
- Total Number of Rows: 10.6 billion
- Clinical Observations: 957.2 million
- Lab Results: 5.1 billion
- Medication Records: 1.6 billion
- Procedures: 511.9 million
- Visits: 506.5 million

https://covid.cd2h.org/dashboard/

N3C Data Enclave Statistics as of December 2, 2021
Why centralize COVID-related data?

Local information

- Is drug X beneficial to covid-19 patients?
- Does Disease Y impair course?
- Does an income > $50,000 per year improve outcomes?

Nationwide information, e.g., N3C

- Is drug X beneficial to covid-19 patients?
- Does Disease Y impair course?
- Does an income > $50,000 per year improve outcomes?

What drugs help covid-19 patients, and which hinder?
What Diagnoses impact outcome?
What Social Determinants impact course and outcome?
# Four Pillars of Data Protections

## Regulatory & Policy
- Data-Contributing Sites abide by the HIPAA Privacy Rule
- N3C research is subject to the Federal Policy for the Projection of Human Subjects in research ('Common Rule')
- Data are provided as HIPAA-defined Limited Data Set
- NIH IRB oversight & waiver of consent
- For COVID-19 related research only
- No genomic data
- No emergency public health authorities were used to obtain the data under these conditions.

## Privacy Measures
- Certificate of Confidentiality
- Data stays within the Enclave: No download or capture of raw data
- Privacy Impact Assessment
- Review of project requests by Data Access Committee
- Additional Tribal data privacy measures (while seeking a consultation with Tribal Nations)

## Security Testing and Monitoring
- Federal Government Compliant Enclave managed by NCATS
- Meets government security controls for cloud security and privacy
- Data encryption in transit and at rest, without exception
- Scheduled penetration testing
- Active monitoring and logging by NIH and HHS
- Auditing of activities in the N3C Enclave

## Researcher Responsibilities
- A User’s organization signs a Data Use Agreement with NCATS for terms and conditions of use
- Users Adhere to a Code of Conduct
- Required NIH IT Security Training
- Required Human Subjects’ Protections Training
- Follow Community Guiding Principles
Additional privacy measures for AI/AN data, while we seek Consultation

Data Characteristics
- Self-identified or provider identified information
- OMB Race Categories
- No tribal affiliation
- Existing data only; no connection to patients directly

Current Strategy Until Consultation
- AI/AN is concealed in “other” category
- 2,659 zip codes overlapping Tribal lands are not available, to block any inference of tribal affiliation
- Tribal affiliation is NOT provided to N3C by the contributing sites
Prioritizing Tribal Perspectives

• Working with THRO
  • Monthly Meetings since September 2020

• Conversations with external experts
  • Johns Hopkins Center for American Indian Health (August 2020)
  • Southcentral Foundation (September 2020)
  • Tribal Epidemiology Center Directors (September & November 2020)

• Learning from NIH COVID-19 Tribal Consultations
  • May 2020 – NIH Tribal Consultation on COVID-19 Research
  • July 2021 – NIH Tribal Consultation on Rapid Acceleration of Diagnostics (RADx) Tribal Data Repository
  • April 2019 – NIH Tribal Consultation on Draft Policy for Data Management and Sharing
What research does N3C enable?

N3C’s Utility
• Largest EHR research resource
• Longitudinal data
• Variant Waves
• Long COVID

• Characterizing the short- and long-term effects of COVID-19
• Understanding the social determinants of health (SDoH) and outcomes for SARS-CoV-2 infected patients
• Identifying effective drugs, repurposed as a treatment for COVID-19
• Predicting who might have severe outcomes if they have COVID-19
• Observing and tracking the progress of COVID-19 over time and geographically
N3C is contributing to our understanding of COVID-19

Publications
- Description of N3C
- Health equity
- Diabetes
- Cancer
- Machine Learning
- COVID meds

Pre-prints:
- HIV COVID risk
- Rural population mortality rates
- Long COVID
- others
N3C Registration/Training
https://covid.cd2h.org/Tutorials

Registration for Documents, Meetings & the N3C Data Enclave
Requires Authentication

Enclave Checklist

Training Office Hours:
Tuesdays & Thursdays at 10-11 am PT/1-2 pm ET
Registration Required at this link

Additional Training Tutorials available in the Enclave

Confidential - For Internal Use Only
Seeking input from Tribal Nations

• N3C receives AI/AN data from contributing health sites
  • AI/AN demographic data
  • Zip codes that overlap with tribal land

• AI/AN data are obscured
  • AI/AN → “other”
  • Zip codes → Not available (00000)

We want your opinions + ideas!
### Examples of Potential ideas for managing AI/AN data within N3C

<table>
<thead>
<tr>
<th>Ideas to consider</th>
<th>Impact of this idea…</th>
</tr>
</thead>
<tbody>
<tr>
<td>Should the data continue to be obscured?</td>
<td>Demographics and zip codes remain unavailable for research purposes</td>
</tr>
<tr>
<td>Should we create a separate enclave for the AI/AN data?</td>
<td>Researchers with AI/AN data approvals would be able to access unobscured data; Tribal representatives determine process and level of involvement</td>
</tr>
<tr>
<td>Should AI/AN data be made available in the same way as other demographic categories?</td>
<td>Visibility by all researchers in the Enclave; greatest impact of data</td>
</tr>
</tbody>
</table>
Seeking Input (pre-consultation questions)

• Should AI/AN data be unobscured and made available for COVID-19 research?
  • How would Tribal communities want to be involved in governance, access, use, etc.?

• What steps would the N3C need to take to consider making the data available?
  • Which aspects of governance would be particularly critical for this resource?
  • How can ongoing partnerships with Tribal communities and researchers be instituted and maintained?

• If Tribal nations identify benefits to using the N3C resource, what outreach is needed to better engage AI/AN researchers?
NCATS

COLLABORATE. INNOVATE. ACCELERATE.

ncats.nih.gov  @ncats_nih_gov  @ncats.nih.gov  NIH-NCATS

To learn more about N3C, visit us at https://ncats.nih.gov/n3c