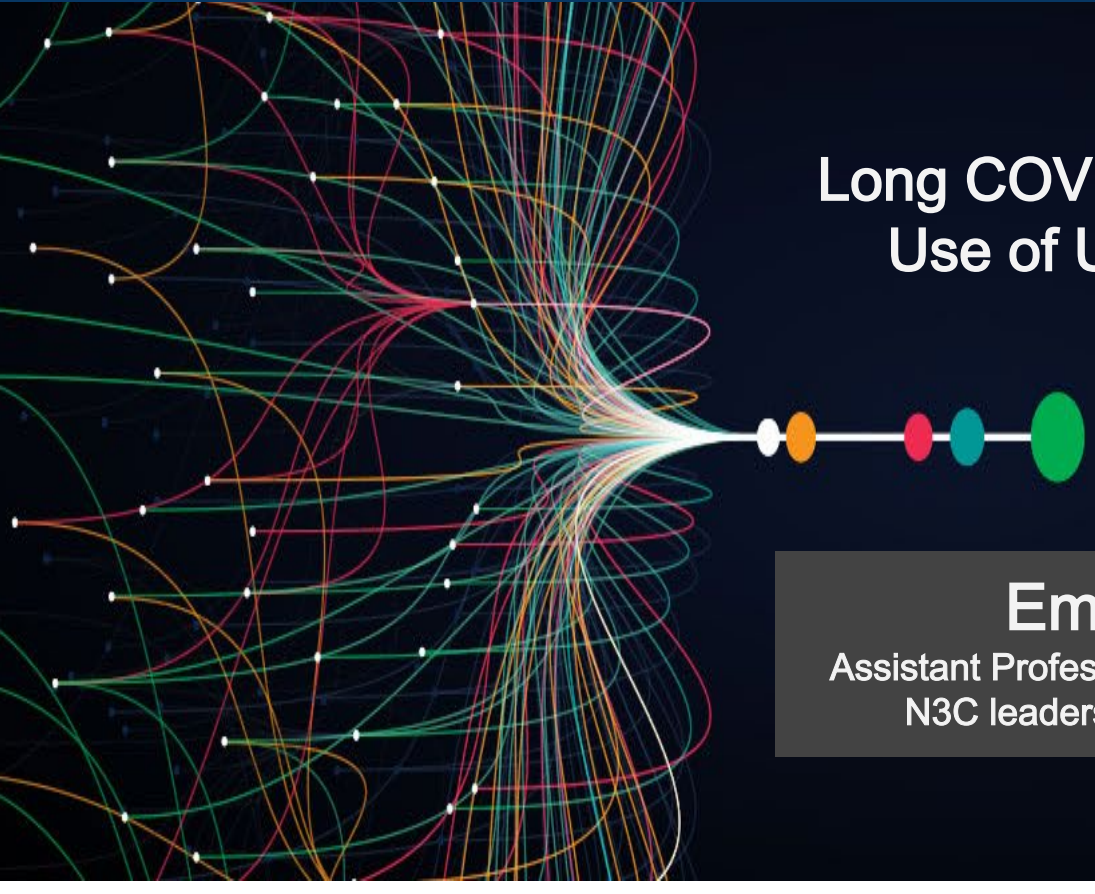
National COVID Cohort Collaborative

A program of NIH's National Center for Advancing Translational Sciences

RECOVER
Researching COVID to Enhance Recovery

An Initiative Funded by the National Institutes of Health



Long COVID through an ICD-10 Lens: Use of U09.9 in the Long COVID definition

Emily Pfaff, PhD, MS

Assistant Professor, UNC Chapel Hill School of Medicine

N3C leadership; multi-PI, RECOVER EHR/RWD

National COVID Cohort Collaborative

A program of NIH's National Center for Advancing Translational Sciences

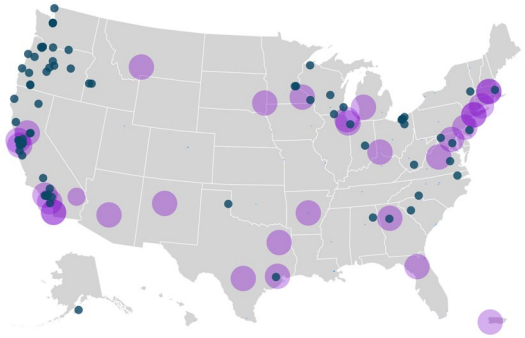
N3C: National COVID Cohort Collaborative
RECOVER program

RECOVER
Researching COVID to Enhance Recovery

An Initiative Funded by the National Institutes of Health

N3C: largest national public HIPAA-limited data set in US history

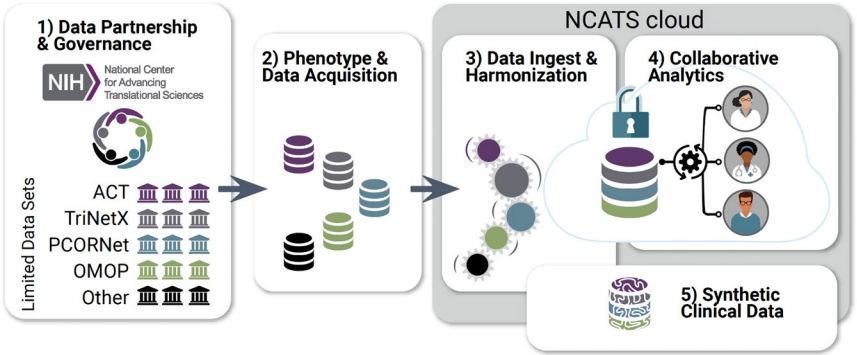
19 M records from >230 institutions



covid.cd2h.org/dashboard

- **Representative pan-US cohort**: race, ethnicity, gender, geography, socio-economic status, health background
- **Harmonized:** overcomes source data heterogeneity
- **Linked:** Patient records, viral variants, vaccine data, CMS, environment, SDoH, etc.
- **Public Health Surveillance:** new variants, comparative effectiveness of drugs

N3C's data-model agnostic, data harmonization and QC pipeline

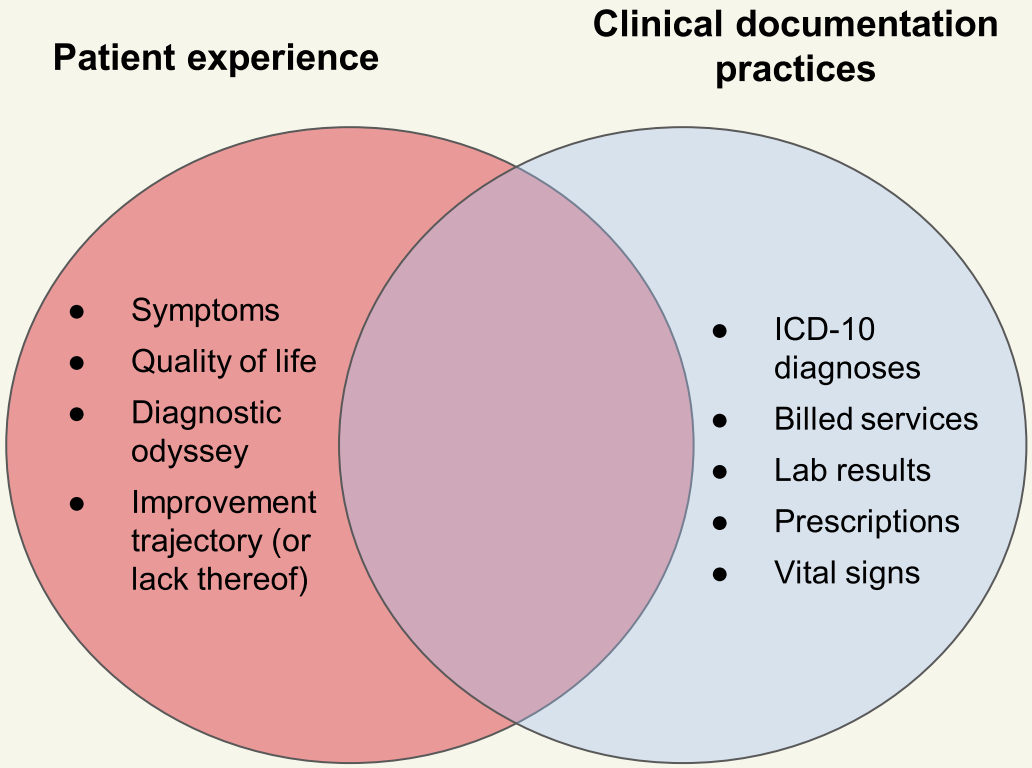


>450 organizations participating, >4,600 users

National COVID Cohort Collaborative

The Challenge of Defining Long COVID in the EHR

RECOVER
Researching COVID to Enhance Recovery
An Initiative Funded by the National Institutes of Health



- Many elements of the WHO Long COVID definition are inconsistently available in the EHR.
- Important Long COVID features (e.g., brain fog, PEM) do not have specific ICD-10 codes.
- EHR data is only available for patients who have access to/seek care.

However, N3C EHR data has the following affordances:

- Large, diverse population
- Data from many (~78) clinical sites
- Longitudinal data (2018 - present)

National COVID Cohort Collaborative

U09.9 as Definition: Advantages

RECOVER
Researching COVID to Enhance Recovery
An Initiative Funded by the National Institutes of Health

- U09.9: ICD-10 code for “Post COVID-19 Condition, Unspecified.” Released 10/1/2021.
- U09.9 gives us *something to grab onto* in EHR data–a platform we can use to go further.
- Characterizing patients who have a U09.9 code can reveal other clinical characteristics that surround that code, including:
  - Other diagnoses/comorbidities
  - Medications
  - Procedures
- We featured this analysis in a recent *BMC Medicine* publication.

National COVID Cohort Collaborative

Co-Occurring Diagnoses Across the Lifespan <21 years of age

RECOVER
Researching COVID to Enhance Recovery

An Initiative Funded by the National Institutes of Health

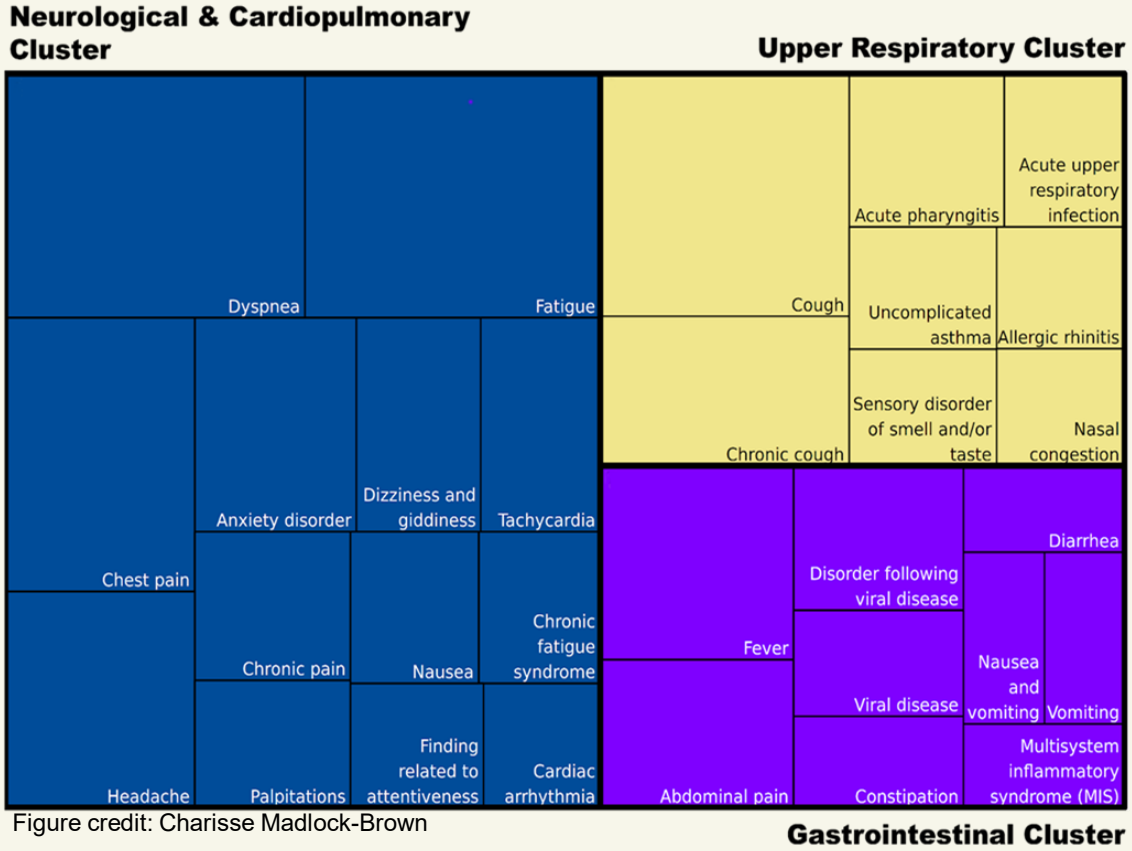


Figure credit: Charisse Madlock-Brown

National COVID Cohort Collaborative

Co-Occurring Diagnoses Across the Lifespan
46-65 years of age

RECOVER
Researching COVID to Enhance Recovery

An Initiative Funded by the National Institutes of Health

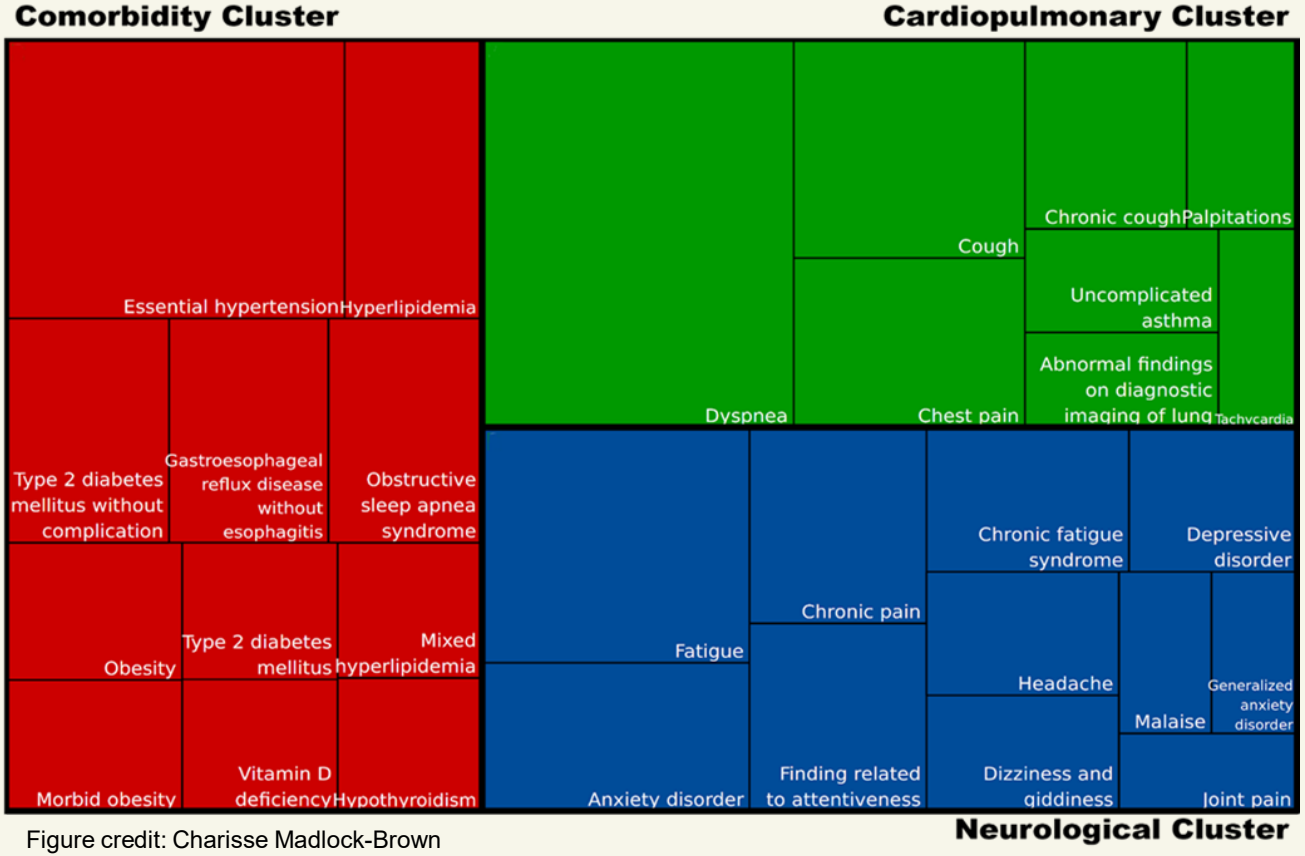


Figure credit: Charisse Madlock-Brown

U09.9 as Definition: Disadvantages

Equity and access issues. The patient group with U09.9 codes is disproportionately White and non-Hispanic, and is significantly less likely to live in areas with higher social deprivation.

Usage issues. There is evidence that U09.9 is frequently "misused," with an unusual number of U09.9 codes that appear to occur on the same day as (or very shortly after) a COVID diagnosis.

Timing and coverage issues. U09.9 came out on October 1, 2021. Vast numbers of Long COVID patients from earlier in the pandemic "missed out" on the code. We cannot be sure that patients without the code should do not have Long COVID.

If we assumed that only patients with a U09.9 code had Long COVID, our data would show a Long COVID prevalence of 1.3% – misleading and far, far too low.

National COVID Cohort Collaborative

Alternative (Still Imperfect) Alternatives

RECOVER
Researching COVID to Enhance Recovery
An Initiative Funded by the National Institutes of Health

1) EHRs for patients diagnosed with PASC*

**U09.9 code or long-covid clinic*

Machine learning

2) Learned patterns of clinical features of PASC

- Dyspnea
- Fatigue
- No vax on record
- New albuterol Rx
- Many outpatient visits
- New corticosteroid Rx
- …

Search EHRs for similar patients

3) Identify previously unknown cases using learned patterns

Pfaff, Girvin, et al. *Lancet Digital Health*; May 16, 2022; https://doi.org/10.1016/S2589-7500(22)00048-6

Late breaking: We're currently revising this model to not require EHR documentation of initial COVID infection, which we do not believe should be a prerequisite for the Long COVID computable phenotype.

Takeaways

- The EHR should not drive the clinical definition of Long COVID.
- Presence or absence of U09.9 in a patient's record should be used as a tool in a larger toolbox, but not relied on too heavily.
- EHR researchers should prepare to adapt the clinical definition to an EHR-ready proxy.
- Machine learning approaches will likely result in a more nuanced EHR definition than any single diagnosis code-based approach.

Thank you!

I would like to acknowledge the scientific contributions of Charisse Madlock-Brown, co-first author of our U09.9 paper, the N3C RECOVER team, representatives from the Patient-Led Research Collaborative, and the N3C community.