RECOVER Adult PCORnet EHR Cohort

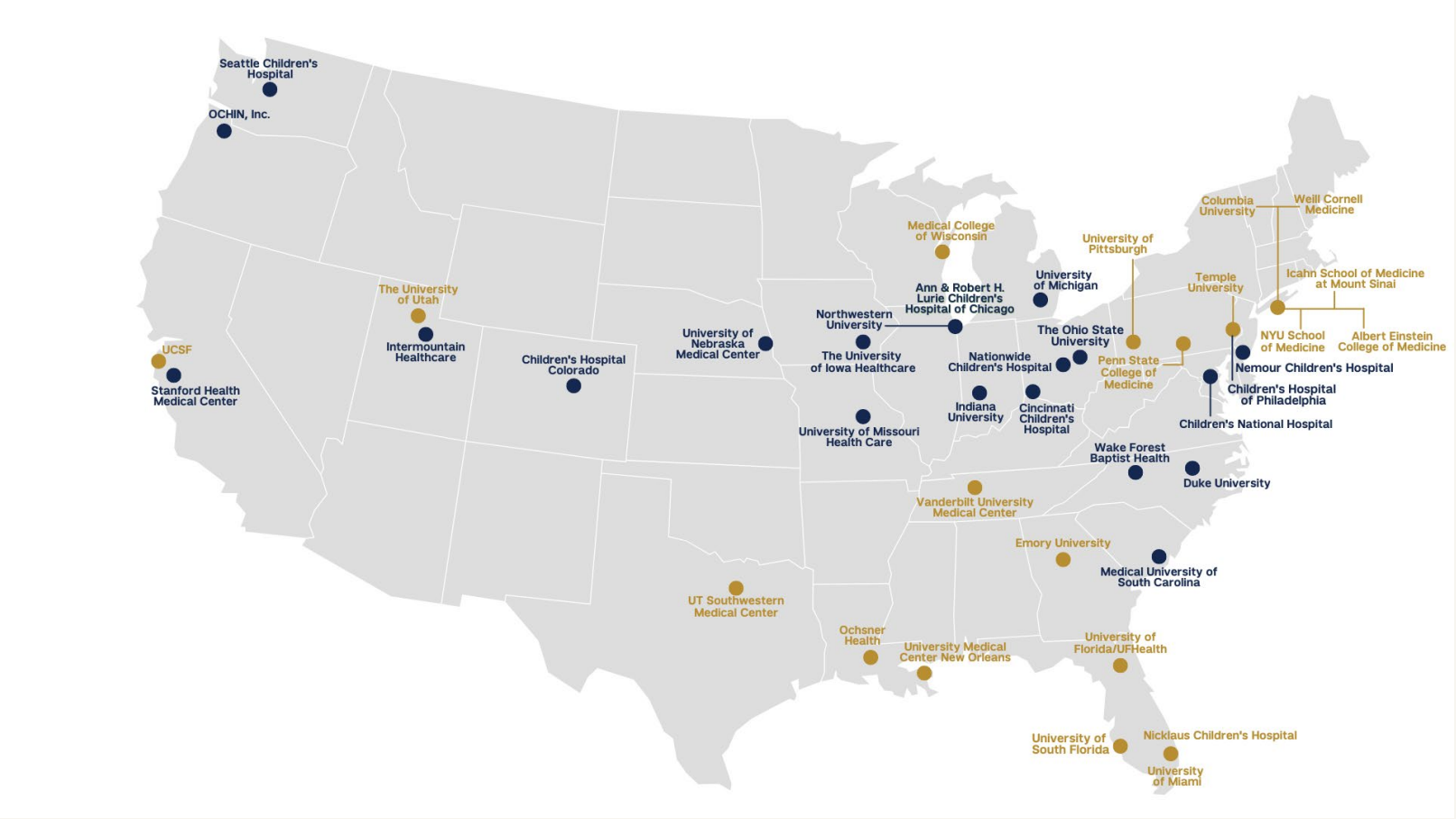
Defining PASC based on EHR Data

Mark Weiner, MD, FACMI
Weill Cornell Medicine

RECOVER
Researching COVID to Enhance Recovery

An Initiative Funded by the National Institutes of Health

PCORnet-based RECOVER sites providing EHR data



PCORNet CRNs
ADVANCE
CAPRiCORN
GPC
INSIGHT
OneFlorida
PaTH
PEDSnet
REACHnet
STAR

Blue sites = CHOP coordinates
Yellow sites = Weill Cornell coordinates

recoverCOVID.org

Challenges in recognizing PASC

- As a new condition there is no gold standard definition of the Post-Acute Sequelae of COVID

- However, it is clearly an amalgam of a diverse set of conditions affecting a variety of organ systems that *may* include the following
  - Cardiovascular – myocardial infarction, rhythm abnormalities, thrombotic, embolic
  - Pulmonary – obstructive, restrictive, inflammatory
  - GI – diarrhea, constipation, nausea
  - Renal – Renal dysfunction, electrolyte disturbances
  - Musculoskeletal – joint and muscle pain with different degrees of weakness and inflammation
  - Neurological – cerebrovascular disease, fatigue, cognitive issues, dizziness
  - Hematological – anemia, --penias (low counts), --cytoses (high counts)
  - Endocrine – Diabetes, thyroid disease,
  - Psychological – depression, anxiety

Challenges in recognizing PASC

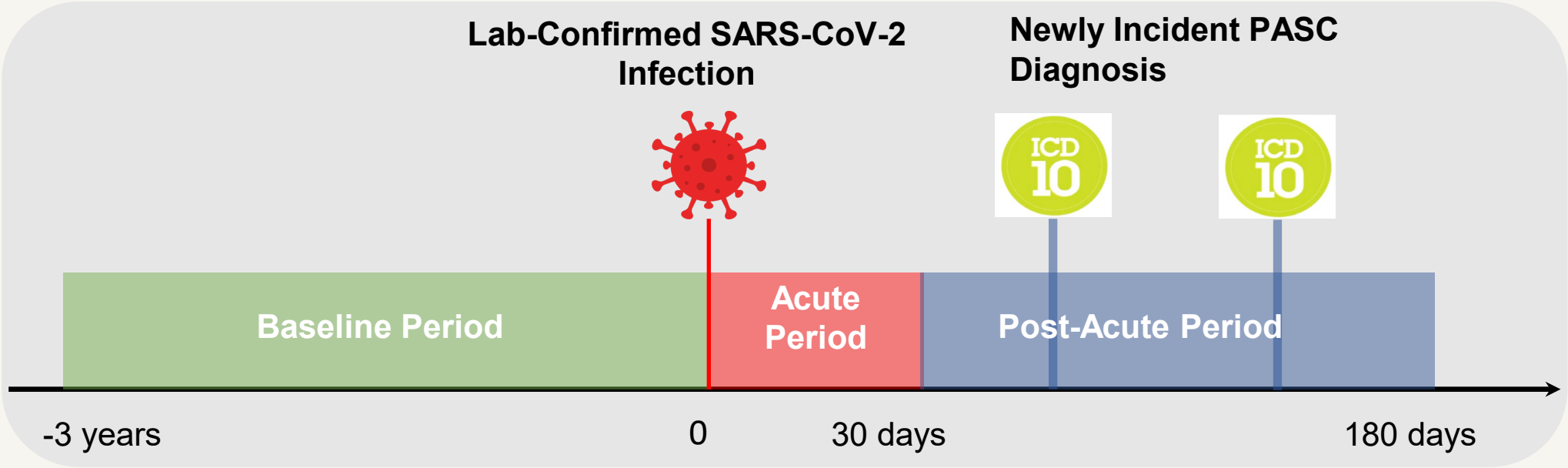
- For a syndrome with no gold standard definition, and many possible associated conditions, how do you recognize clinical correlates for that syndrome?

- Find concordant diagnoses and other characteristics of patients where a clinician has made the association with Long COVID
  - Explore characteristics of patients referred to a Long COVID clinic
  - Explore characteristics of patients who are assigned a Long COVID Diagnosis
  - Dependency on recognition of association of symptoms as Long COVID

- Explore all incident diagnoses among patients with COVID
  - Explore clusters of incident diagnoses that appears after a COVID infection
  - Find a set of incident diagnoses that appear with a higher hazard among a cohort of patients with COVID than without COVID

PASC Subphenotypes

- Overarching approach



- Ensure absence of condition in baseline period before COVID
- Explore for new presence of condition in 30-180 days post COVID
- Find incident conditions that "travel together" either by formal cluster analysis or by similar organ system involvement
- We recognize that this approach that depends on incident diagnoses may miss exacerbations of existing illnesses that may have been stable in the absence of a patient's COVID infection
  - Need to look separately at exacerbations of existing conditions

Machine Learning for Identifying Data-Driven Sub-phenotypes

Objectives

- Utilize topic modeling to identify sub-phenotypes of PASC

Methods

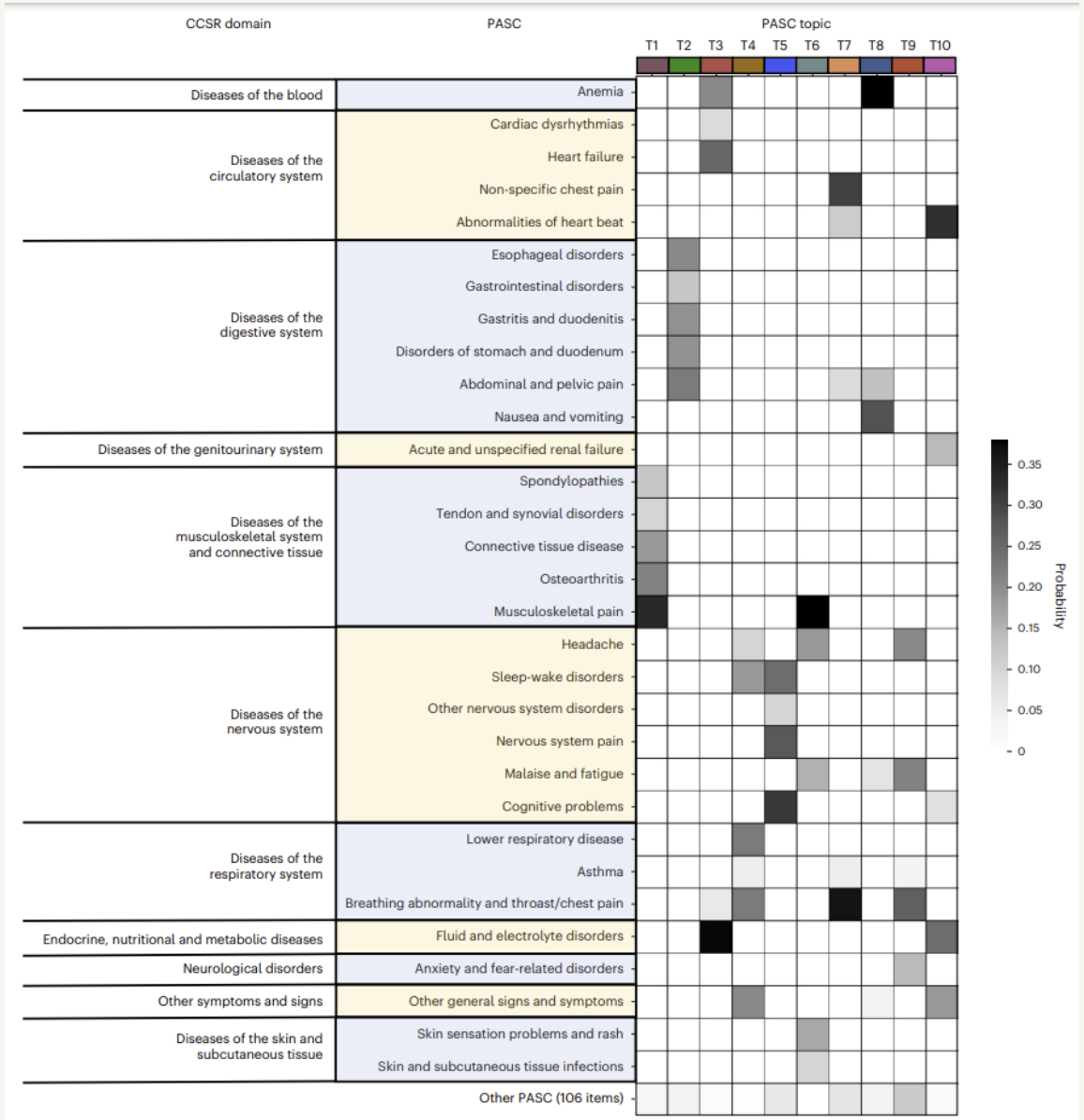
- Mapped 137 newly incident diagnoses into 10 topics based on co-occurrence patterns in 34,605 patients
- Analyzed clustering of topics in patients to demonstrate four sub-phenotypes

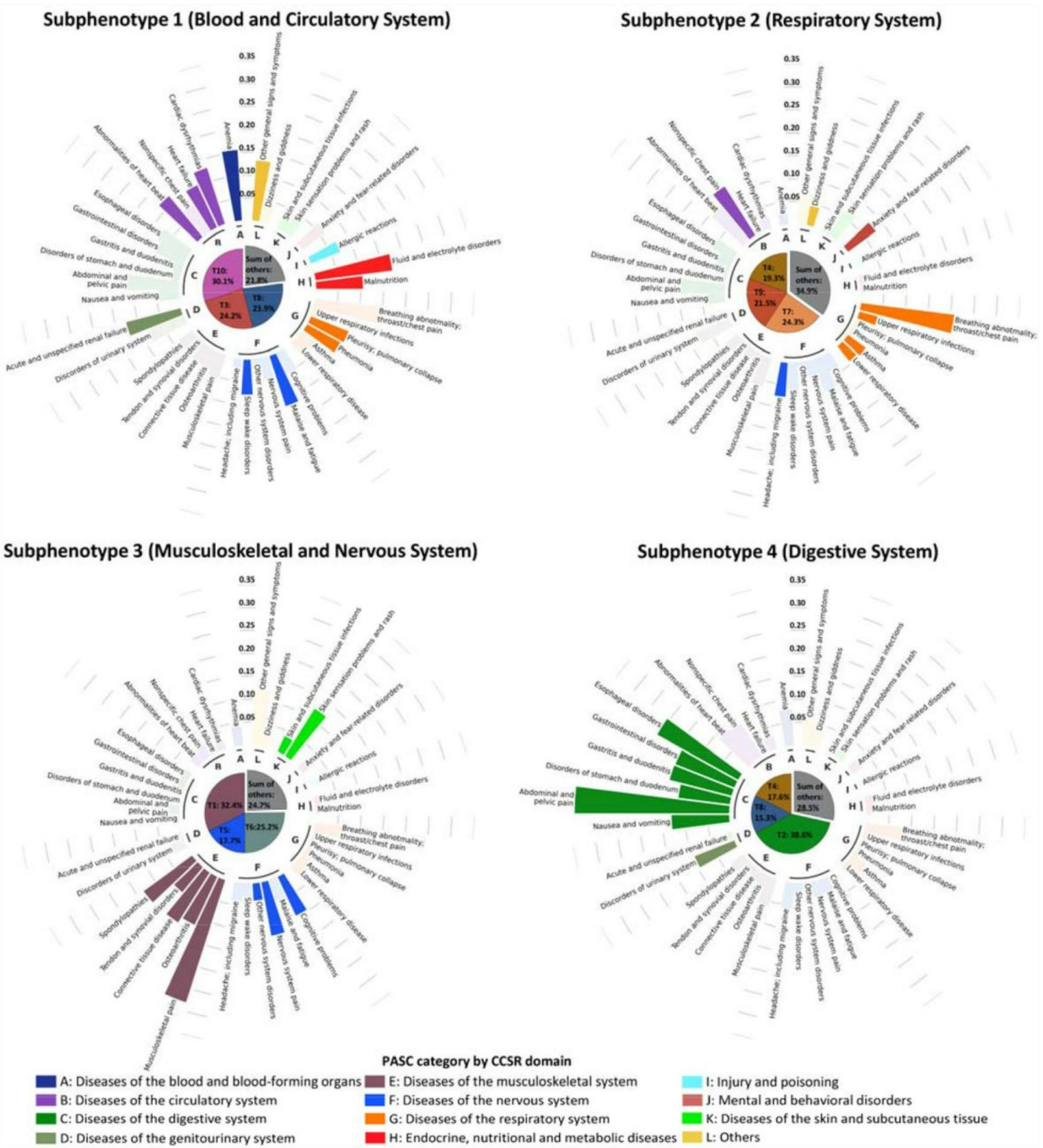
Results

- Identified 10 topics across the following systems: blood, circulatory, digestive, genitourinary, musculoskeletal & connective tissue, nervous, respiratory, endocrine & nutritional & metabolic, neurological disorders, skin & subcutaneous tissues, and other
- Four sub-phenotypes characterized by:
  - **Cardiac and renal** (median age 65, 51% female, higher severity in acute phase)
  - **Respiratory conditions, sleep disorders & anxiety** (median age 51, 63% female, lowest rates of hospitalization)
  - **Musculoskeletal and nervous system** (median age 57, 61% female)
  - **Digestive system** (median age 54, 62% female, lowest rates of ICU care)

Zhang, H., Zang, C., Xu, Z. *et al.* Data-driven identification of post-acute SARS-CoV-2 infection subphenotypes. *Nat Med* **29**, 226–235 (2023). https://doi.org/10.1038/s41591-022-02116-3

PASC Subphenotypes: Adults





High-throughput causal inference pipeline to identify PASC diagnoses

Objectives

- Define PASC on the basis of diagnostic categories with elevated hazard ratios in COVID vs non-COVID cohorts

Methods

- Define cohorts of patients with and without COVID on the basis of PCR/Antigen testing results
- Apply inverse propensity treatment re-weighting (IPTW) to account for confounding in the two cohorts
- Assess the hazard ratio of diagnostic categories across the two cohorts

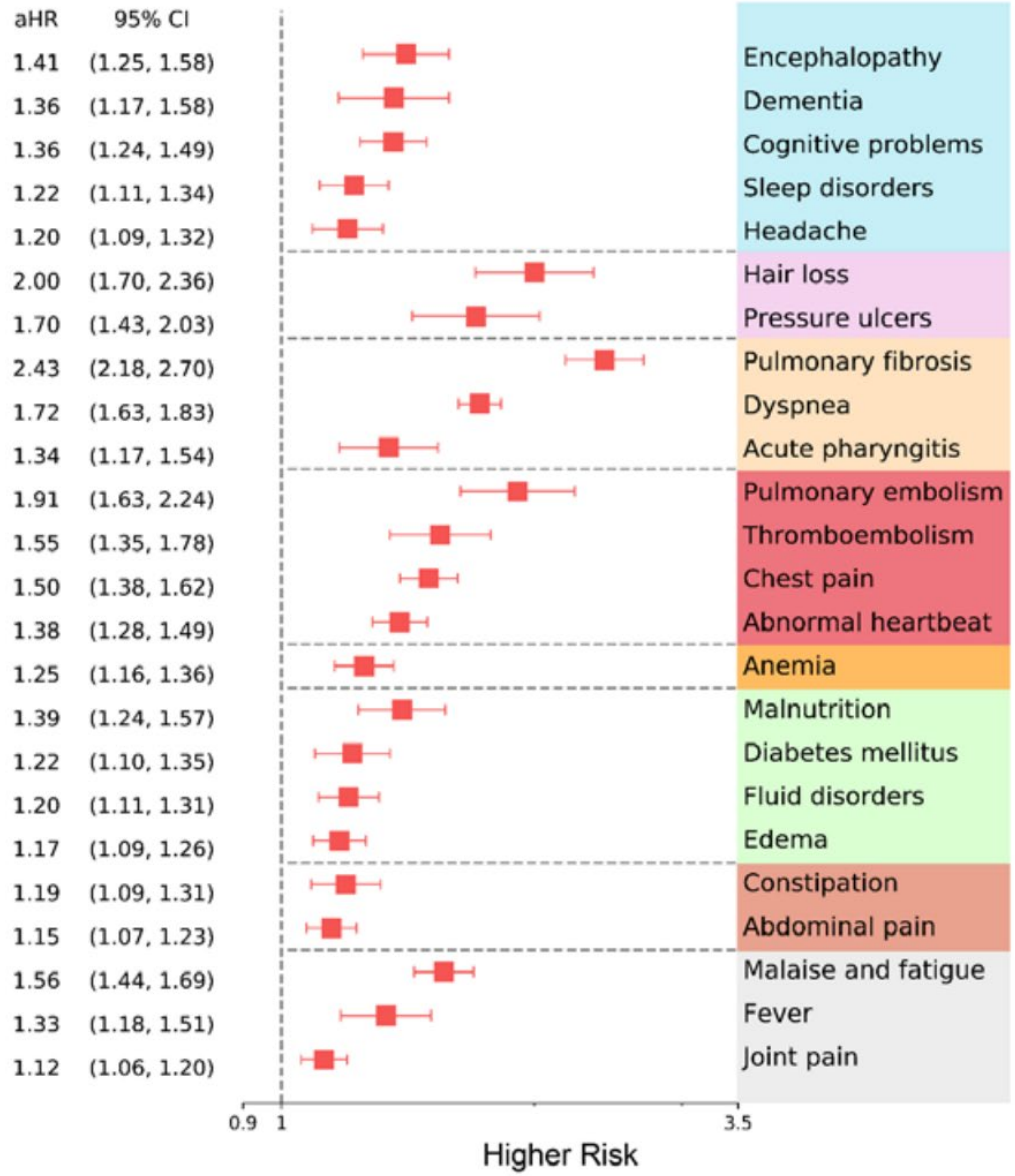
Results

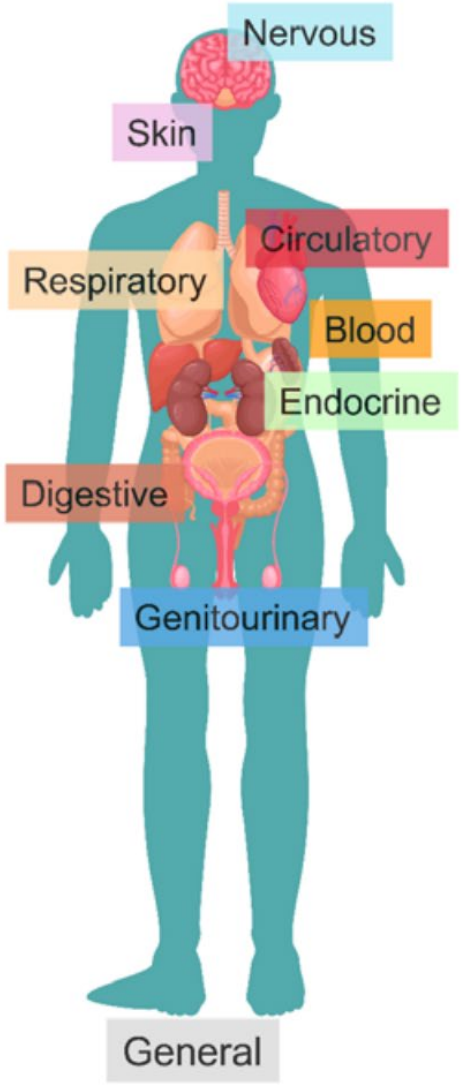
- Selected diagnostic categories across several organ systems including Neurological, respiratory musculoskeletal, cardiovascular, gastrointestinal, hematological, dermatological, endocrine, and renal were associated with a higher hazard ratio in COVID patients than non-COVID, suggesting that these conditions defined PASC

Zang, C., Zhang, Y., Xu, J. *et al.* Data-driven analysis to understand long COVID using electronic health records from the RECOVER initiative. *Nat Commun* **14**, 1948 (2023). https://doi.org/10.1038/s41467-023-37653-z

Clinical Spectrum: Adults

Potential incident PASC conditions from the PCORnet INSIGHT cohort, 3/2020 - 11/2021



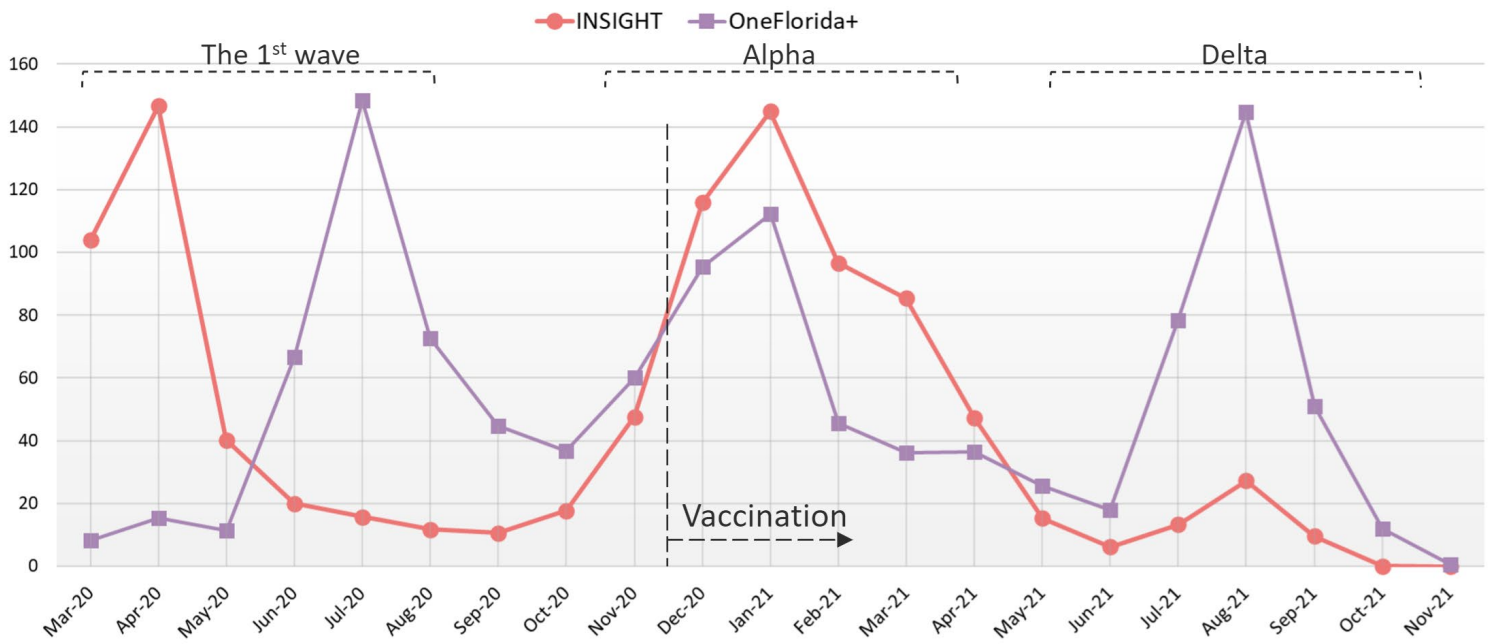


recoverCOVID.org

Impact of Variants

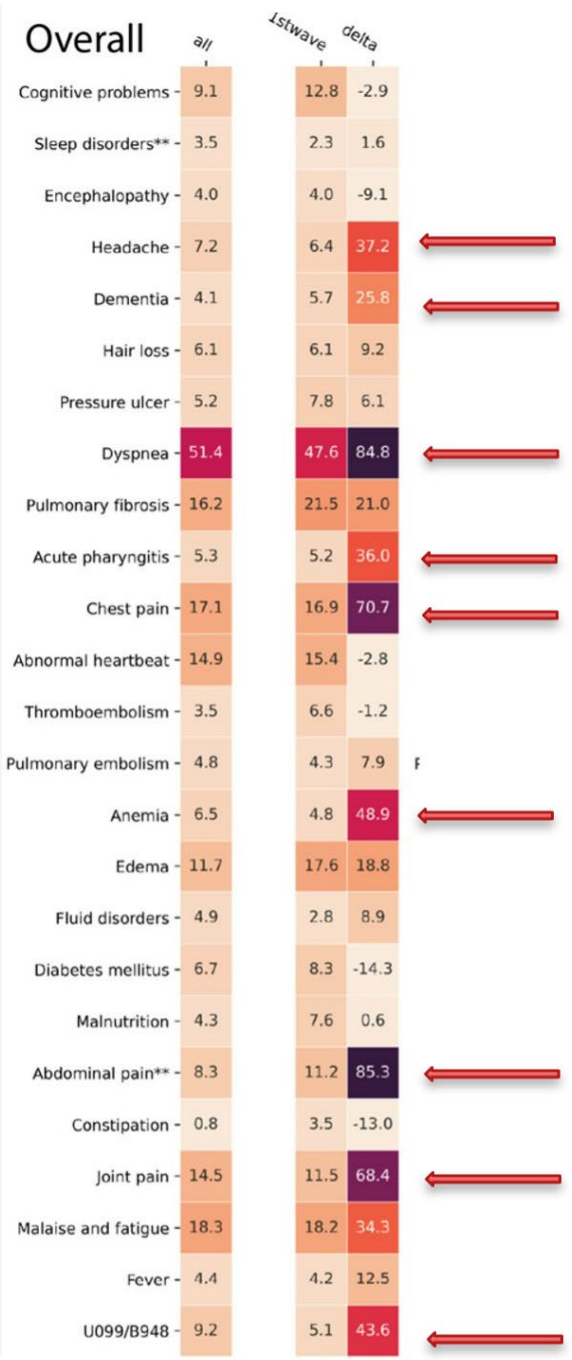
- Overall incidence of PASC conditions was greater for ancestral strain variant infections (52%) than Delta (45%)
- Relative distribution of PASC conditions after each variant was distinct, with dyspnea most common after the ancestral strain and abdominal pain, dyspnea, chest pain and joint pain most common after Delta)

Lab-confirmed New SARS-CoV-2 Cases per 10,000 patients, March 2020 to November 2021



	Mar-20	Apr-20	May-20	Jun-20	Jul-20	Aug-20	Sep-20	Oct-20	Nov-20	Dec-20	Jan-21	Feb-21	Mar-21	Apr-21	May-21	Jun-21	Jul-21	Aug-21	Sep-21	Oct-21	Nov-21
INSIGHT	103.9	146.7	40.2	20.0	15.8	11.7	10.6	17.8	47.6	115.9	144.9	96.6	85.4	47.2	15.4	6.2	13.3	27.2	9.6	0.1	0.0
OneFlorida+	8.4	15.5	11.3	66.7	148.5	72.6	44.7	36.9	60.2	95.5	112.4	45.7	36.1	36.6	25.5	18.0	78.3	144.7	50.9	11.8	0.5

Excess Burden



https://www.medrxiv.org/content/10.1101/2023.02.15.23286012v1

Limitations and implications for applications

- Data sources from EHRs
  - Patients need to present for care, and providers need to record diagnoses
  - COVID testing early on was limited to those who were very ill
  - COVID testing later on was not necessarily recorded given home testing
    - Anchoring with respect to timing of initial COVID infection is difficult
    - Status as initial versus recurrent COVID infection (and timing) may be uncertain
- Data is from pre-omicron era
  - Can repeat analysis to include omicron, but so many people have had COVID by now that COVID-negative label is less certain
    - Fewer conditions associated with significant hazard ratio, perhaps because many in the COVID negative cohort have had COVID, though bias toward null makes the fewer significant findings more notable
- Confounding secular trends
  - Varying peaks and valleys in different parts of the country at different times
  - Variable uptake of vaccines
  - Variable uptake of treatments
- Value of our PASC definition is better for population trend analysis –impact of vaccines, risk assessment, medication usage on outcomes – more than labelling of individuals as having PASC or not

Special Thanks

Study Leadership

Rainu Kaushal
Chris Forrest
Charles Bailey
Elizabeth Ann Shenkman
Russell Rothman

Scientific Leadership

Thomas Carton
Yongkang Zhang
Fei Wang
Grace Lee
Jason Block

Science and Analytics Team

Kristin Lyman
Beth Nauman
Chengxi Zang
Jiang Bian
Dhruv Khullar
Anna Starikoysky
Colby Lewis
Hao Zhang
Yu Hou
Zehra Abedi
Rosie Ferris
Michael Koropsak
Yonghui Wu
Suchitra Rao
Edward Schenck
Anna Nordvig
Wei-Qi Wei

AWS and Data Team

Hanieh Razzaghi
Dmitry Morozyuk
John Ruffing
Anthony DiFazio
Fabrice Garzulino
Carlos Castilla
Sajjad Abedian
Rimsha Azhar
Thomas Campion
Susan Hague
Evan Sholle
Allen Ladd
Brian Whelan
John Knab
Brian Tschinkel
Justin Barber
Sumanth Reddy Thunga

Project Management

Andrea Cohen
Alexandra LaMar
Rosie Ferris
Stephen Hunt
Wilson Jacobs
Jill McDonald
Kelly Clark
Nicole Marchesani
Miranda Higginbotham