How data helped understand Long Covid
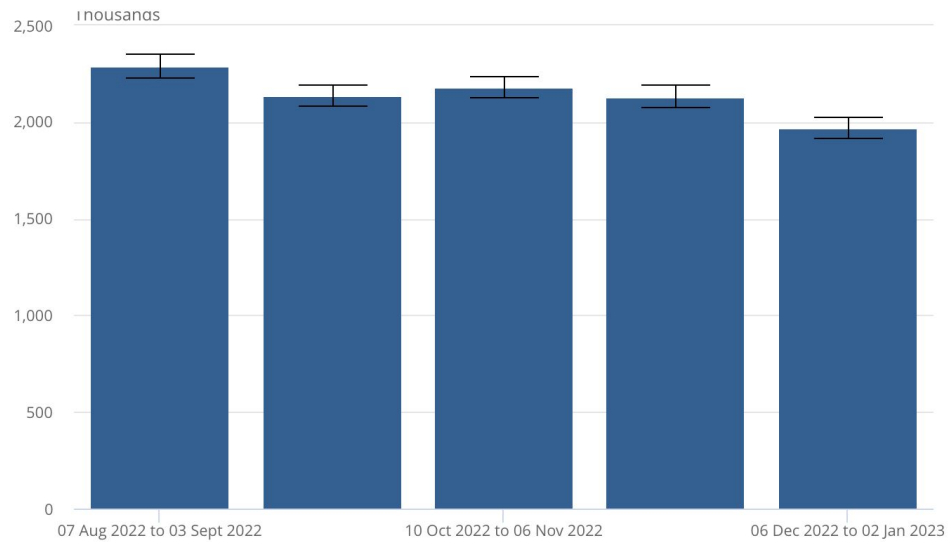
Dr Indra Joshi, Director Health, Research &AI
Palantir Technologies

Long Covid

Figure 1: 2.0 million people were experiencing self-reported long COVID as of 2 January 2023

Estimated number of people living in private households with self-reported long COVID of any duration, UK, four-week periods ending 3 September 2022 to 2 January 2023
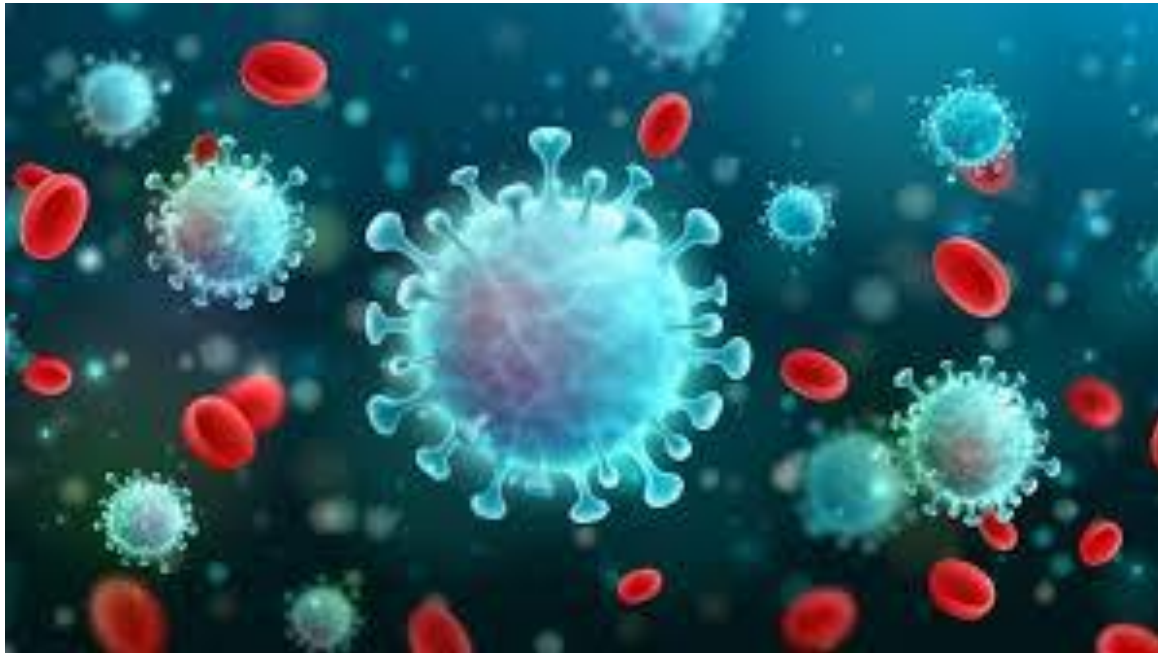

Source: Office for National Statistics – Coronavirus (COVID-19) Infection Survey (CIS)

Fig. 1 The long-term effects of COVID-19
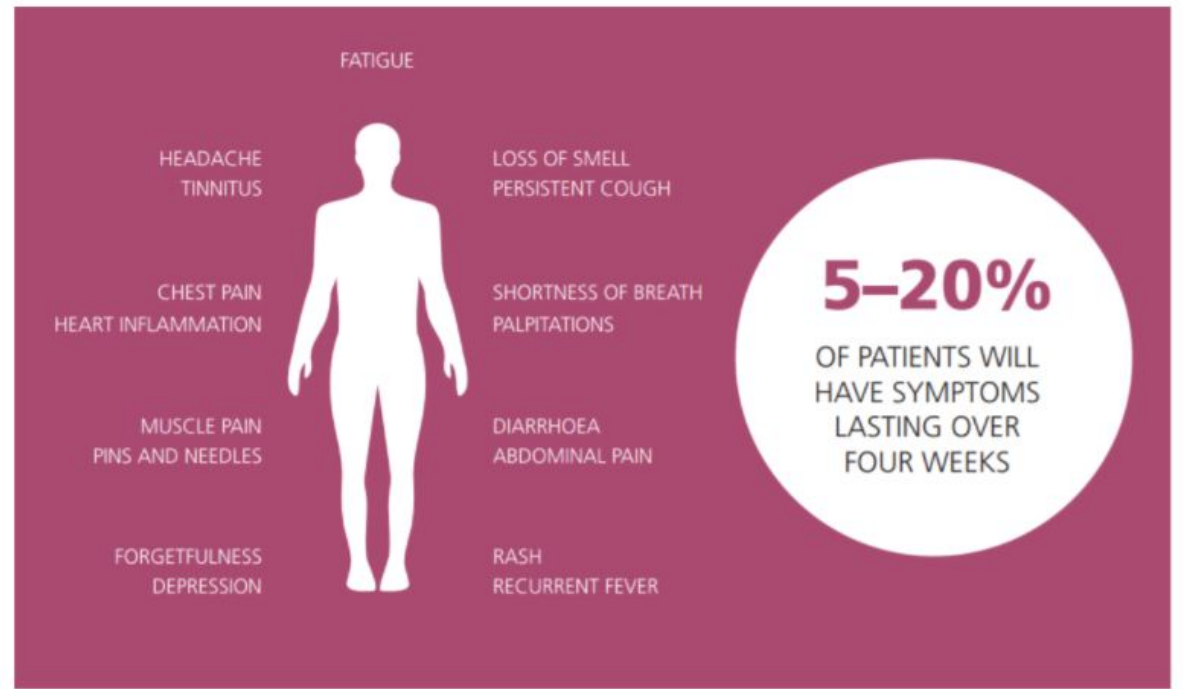
Source: Reproduced with permission from SIMPLECOVID (2020)

Across Europe

At least 17 million people in the WHO European Region experienced long COVID in the first two years of the pandemic; millions may have to live with it for years to come

WHO/Europe urges countries to take post COVID-19 condition seriously by urgently investing in research, recovery, and rehabilitation
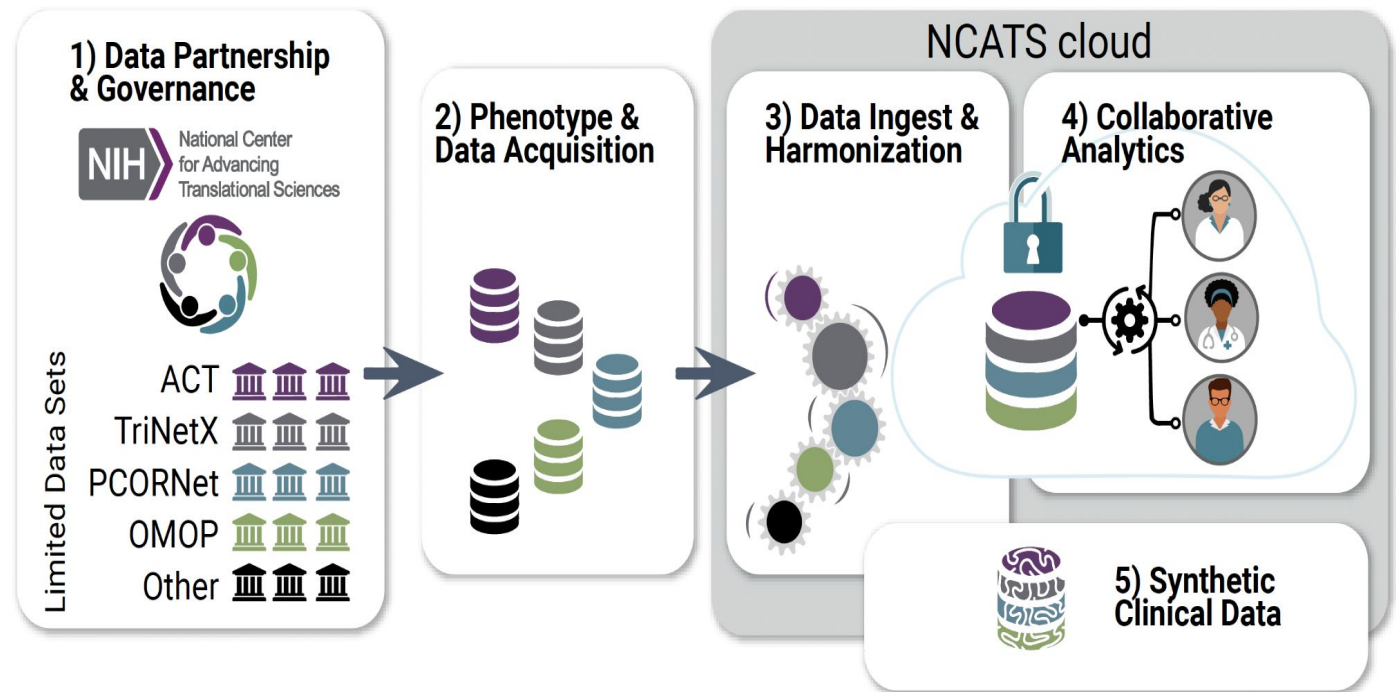
13 September 2022 | Media release | Reading time: 4 min (1072 words)
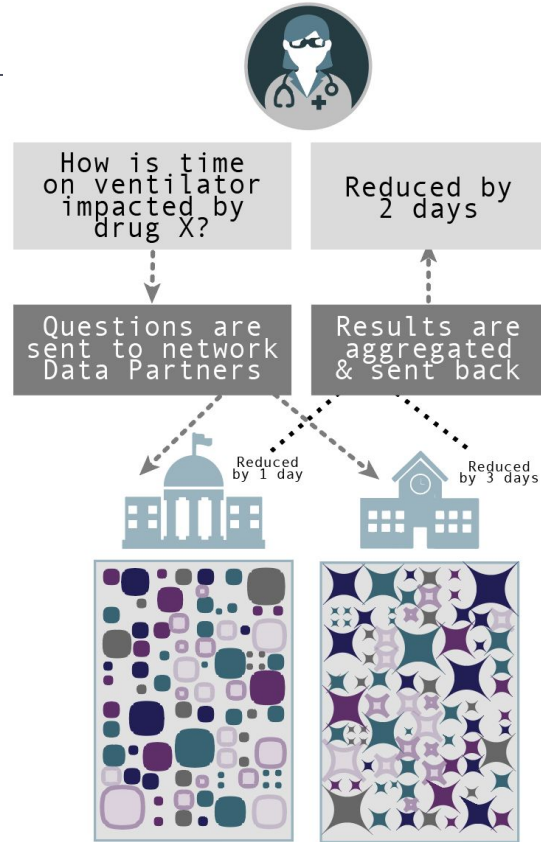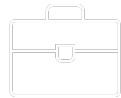
NCATS Goals and Data Environment

Inform data partners about their data quality (highlight good & areas for improvement)

Stop data quality regression & maintain data quality across subsequent payloads

Provide perspective; benchmark set by comparing to other sites

Synergistic federated and centralized approaches

Federation is good for specific queries

How is time on ventilator impacted by drug X?

Reduced by 2 days

Questions are sent to network Data Partners

Results are aggregated & sent back

Reduced by 1 day

Reduced by 3 days

In patients under age 60, which factors are most predictive of severe outcomes?

Collaboratively build, test, and refine algorithmic classifiers

Identify novel associations

Centralized analytics are good for discovery

National COVID Cohort Collaborative

Palantir

Synergies between Centralized and Federated Approaches to Data Quality: A Report from the National COVID Cohort Collaborative (N3C)
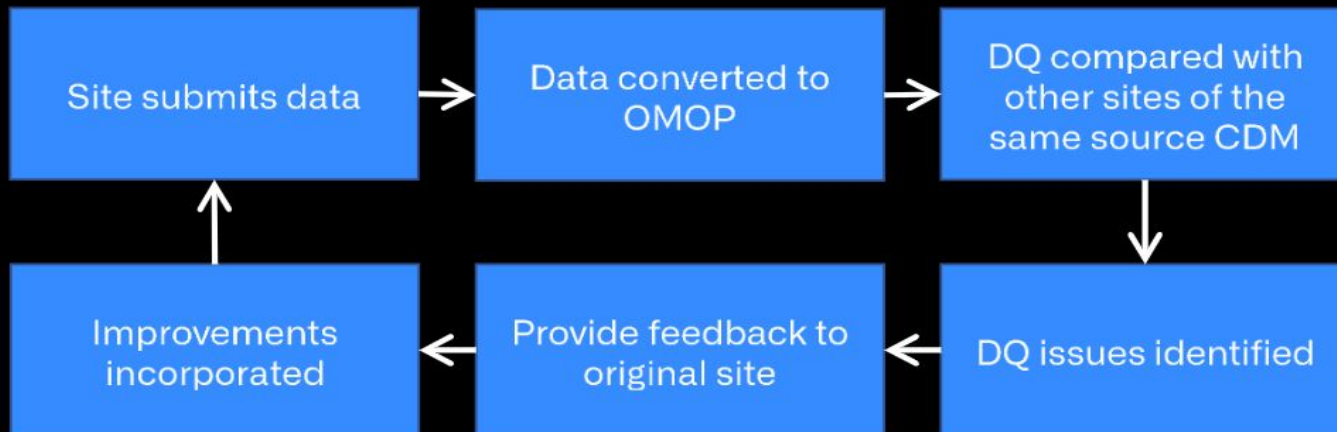
Palantir

" By participating in a consortium like N3C, sites receive routine feedback on their overall quality with tactical information on ways to address local issues. "

N3C's Data Transformation and Feedback Pipeline

Site submits data → Data converted to OMOP → DQ compared with other sites of the same source CDM → DQ issues identified → Provide feedback to original site → Improvements incorporated → (back to Site submits data)

N3C pools data from each partner site and harmonizes all submitted data to the OMOP Common Data Model (CDM). N3C Data Quality (DQ) review involves both automated and manual procedures. In the process:

|	
15	Types of DQ issues discovered
66%	Sites affected by DQ issues
100%	Sites demonstrated DQ improvement after feedback

Ensuring Harmonization

Complete transparency into lineage/provenance of harmonization pipelines for >232 sites (77 DTAs) and >50,000 transforms

Pipeline versioning, deployment, upgrades, and automated data quality checks of new and existing sites

Curators and developers can quickly identify and address issues

Scalability of compute resources; pipelines can be refreshed in <20 mins
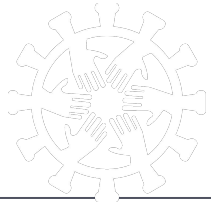
N3C Data Enclave

The N3C Data Enclave represents one of the largest secure collections of harmonized clinical health data in the United States.

Sites: 77

Persons: 18.3 million

COVID+ Cases: 7,217,701

# of Rows: 23.1 billion

Clinical Observations: 2.1 billion

Lab Results: 11.0 billion

Medication Records: 3.6 billion

Procedures: .9 billion

Visits: 1.2 billion

Data as of March 9, 2023

Explore Our Dashboards

Overview	Members	Tracking
Cases	Medications	Diseases

covid.cd2h.org/dashboard

Palantir

Impact of collaborative analytics

RESULTED IN SIGNIFICANT SCHOLARLY PRODUCTIVITY

ATTRIBUTED AT SCALE AND INCENTIVIZED COLLABORATION

TRANSFORMED CARE GUIDELINES

DEVELOPED EVIDENCE-BASED DISEASE DEFINITIONS

DEVELOPED COMPLEX RISK PREDICTION MODELS
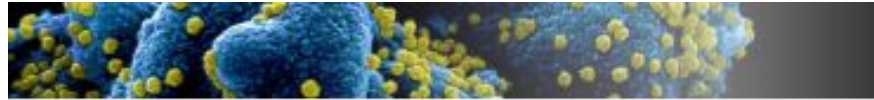
Hepatology Preprint:

Outcomes of SARS-CoV-2 Infection in Patients with Chronic Liver Disease and Cirrhosis: a N3C Study

Authors: Jin Ge, Mark J. Pletcher, Jennifer C Lai, N3C Consortium

medRχiv
THE PREPRINT SERVER FOR HEALTH SCIENCES

N3C Published Paper:

Outcomes of COVID-19 in cancer patients: Report from the National COVID Cohort Collaborative (N3C)

Authored by Noha Sharafeldin, MBBCh, MSc, PhD, et al.

Journal of Clinical Oncology®

Another preprint from Joy Alamgir, et al.

Lack of association between convalescent plasma administration and length of hospital stay: A hospital-day stratified multi-center retrospective cohort study

Preprint available now!

medRχiv
THE PREPRINT SERVER FOR HEALTH SCIENCES

Preprint from Justin Reese, PhD

Association of cyclooxygenase inhibitors with reduced COVID-10 severity identified by graph machine learning followed by retrospective EHR cohort study

Preprint available now!

medRχiv
THE PREPRINT SERVER FOR HEALTH SCIENCES

N3C Cohort Preprint Published

The National COVID Cohort Collaborative: Clinical Characterization and Early Severity Prediction

Access the article on medRxiv, the Preprint Server for Health Sciences

medRχiv
THE PREPRINT SERVER FOR HEALTH SCIENCES

Characteristics, Outcomes, and Severity Risk Factors Associated with SARS-CoV-2 Infection Among Children in the US National COVID Cohort Collaborative

> " The N3C database provides a diverse, granular view of pediatric SARS-CoV-2 infections and allows for novel vital sign and laboratory value trajectory mapping. "

Martin B, DeWitt PE, Russell S, et al. Characteristics, Outcomes, and Severity Risk Factors Associated With SARS-CoV-2 Infection Among Children in the US National COVID Cohort Collaborative. *JAMA Net Open.* 2022;5(2):e2143151. doi:10.1001/jamanetworkopen.2021.43151

Key Severity Predictors

Increased odds ratio for severe disease was observed for:

- Male children relative to female children

- Black/African American children relative to white children

- Children with pediatric complex chronic condition (PCCC) co-morbidities relative to those without.

↳Powered by the Palantir Platform

In-Hospital Vital Sign and Laboratory Value Trajectories

Legend:
- Severe or dead
- Moderate
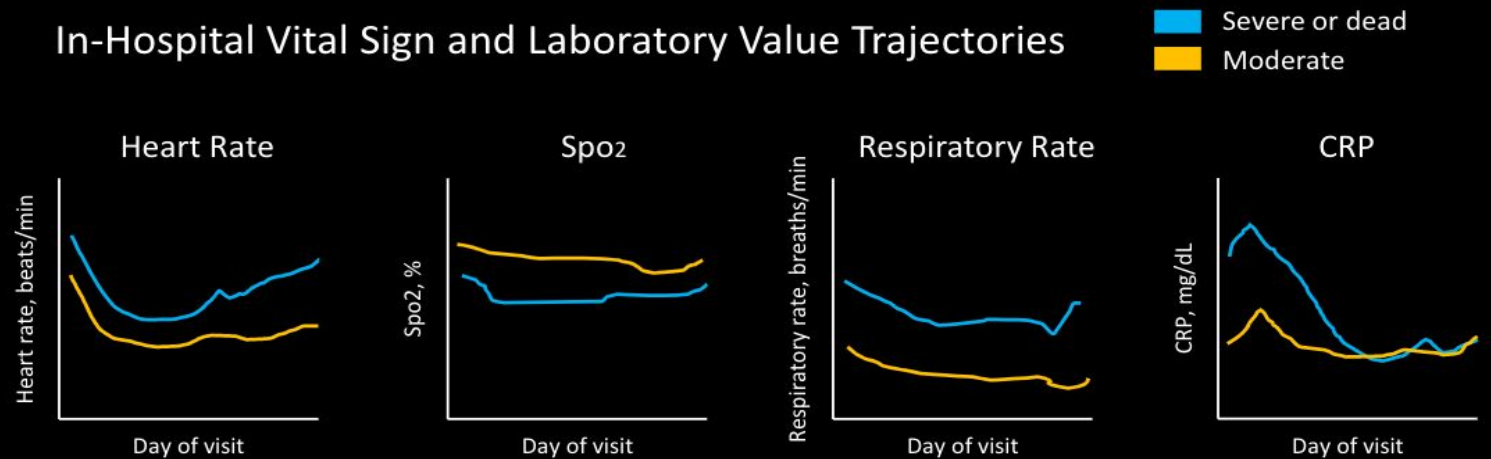
- Compared with the moderate severity subgroup, the severe subgroup had more abnormal initial values for many vital signs.

- Early identification of children likely to progress to a more severe phenotype could be achieved using readily available data from the day of admission

Complete Modeling Lifecyle

01   IDENTIFY THE PROBLEM

EVALUTE DECISION

CHOOSE MODEL

TEST MODEL

02   03   04

TRAIN MODEL

DEPLOY MODEL

MAKE A DECISION

RESOLVE THE PROBLEM   05

Palantir
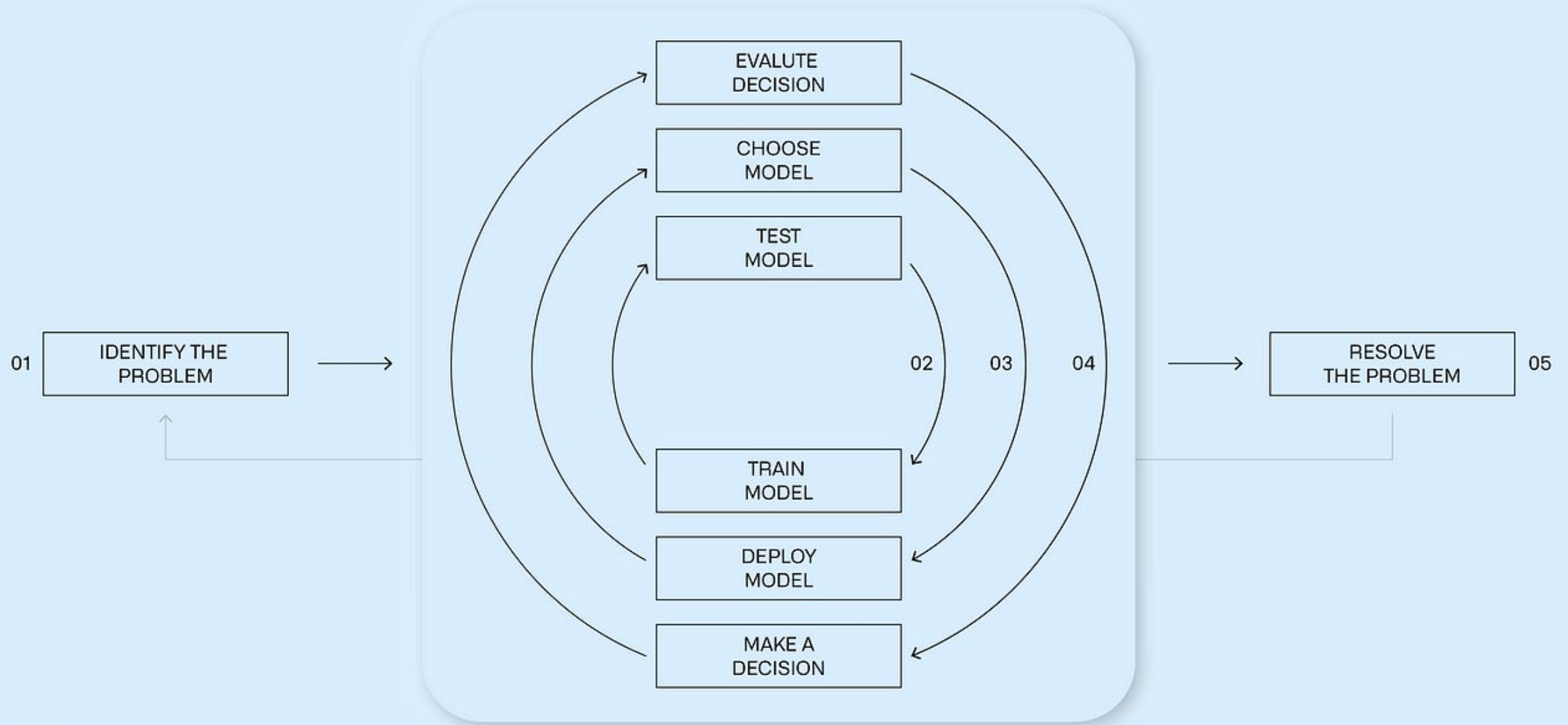
Predicting risk and informing policy

Problem:
- Conflicting research on effects of vaccination on long-covid
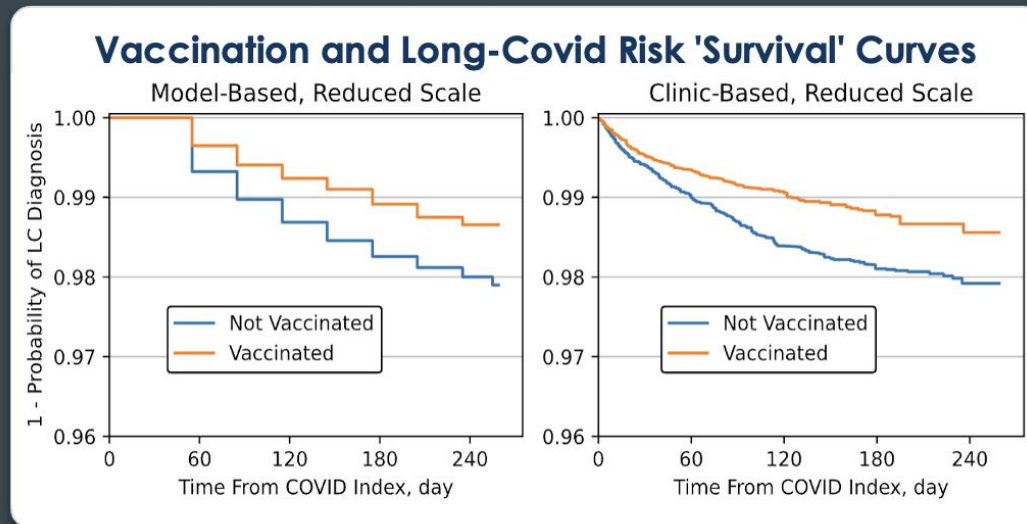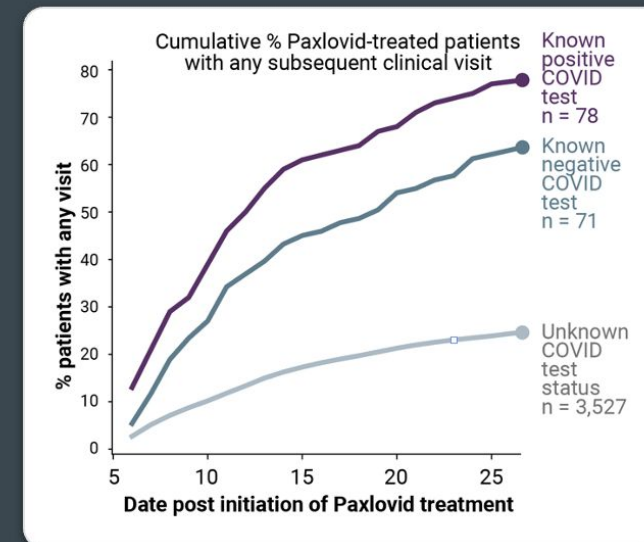- Determining who is vaccinated is challenging in the US

Vaccination and Long-Covid Risk 'Survival' Curves

Model-Based, Reduced Scale / Clinic-Based, Reduced Scale

1 - Probability of LC Diagnosis vs Time From COVID Index, day

Not Vaccinated / Vaccinated

Solution: N3C reconciled vaccination data and demonstrated using multiple methods that vaccination lowers risk of Long-COVID

Problem:
- To plan pandemic response White House needed real-world evidence that Paxlovid was effective

Cumulative % Paxlovid-treated patients with any subsequent clinical visit

Known positive COVID test n = 78
Known negative COVID test n = 71
Unknown COVID test status n = 3,527

% patients with any visit vs Date post initiation of Paxlovid treatment

Solution: N3C showed that few patients require care or are hospitalized post-COVID following Paxlovid treatment

N3C won grand prize in the Dataworks! Competition!
Democratizing access to sensitive clinical data

106 TEAMS

537 PEOPLE

Disciplines represented:
- biochemistry
- clinical research
- genomics
- immunology
- molecular biology
- neuroscience

26+ COUNTRIES

CONGRATS

Palantir

Thank You

www.palantir.com/uk/healthcare