

# NCATS

COLLABORATE. INNOVATE. ACCELERATE.

## N3C

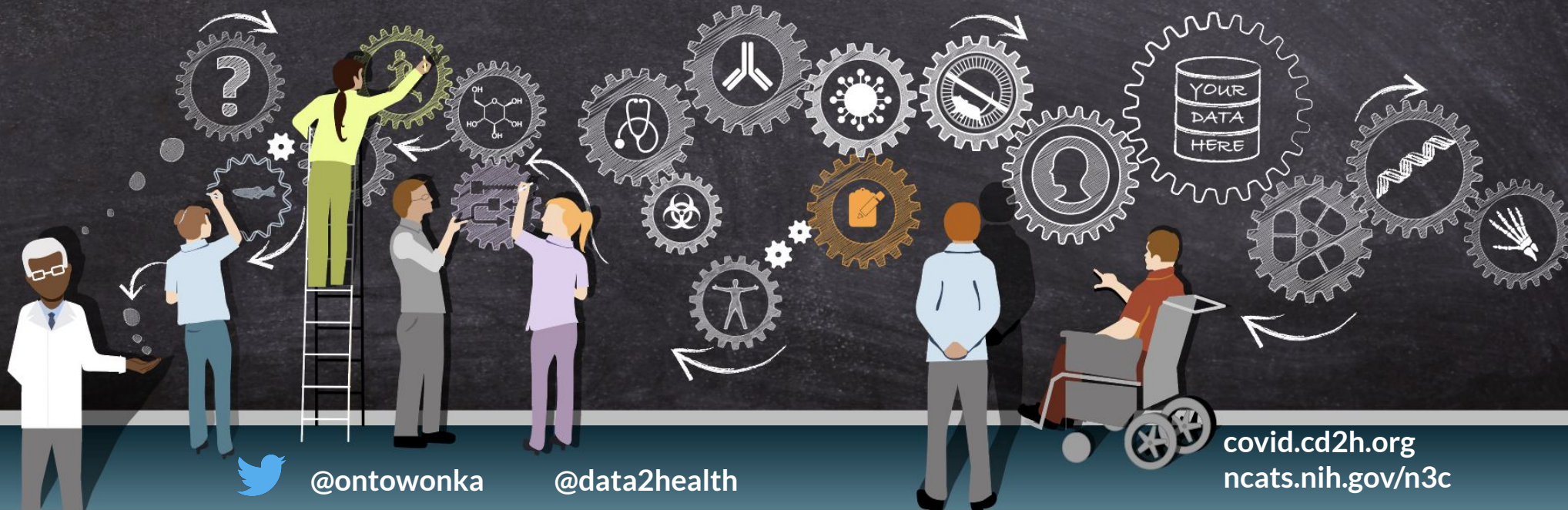


**NIH** National Center  
for Advancing  
Translational Sciences

# National Covid Cohort Collaborative update

NCATS CTSA Steering Committee

January 11th, 2021

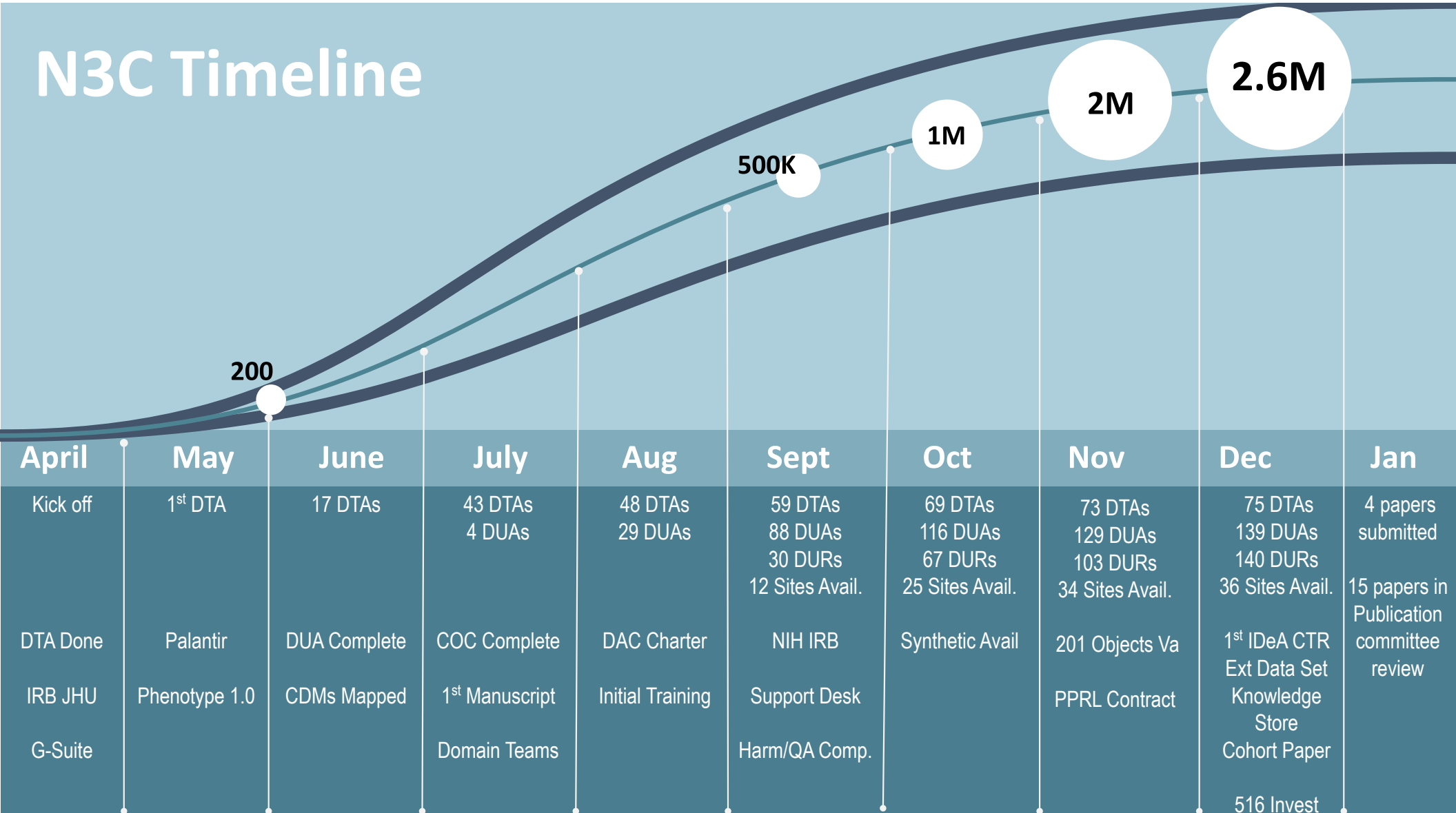


@ontowonka

@data2health

[covid.cd2h.org](https://covid.cd2h.org)  
[ncats.nih.gov/n3c](https://ncats.nih.gov/n3c)

# N3C Timeline





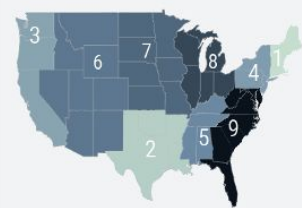
## N3C Enclave Data Stats

**Over Two Billion Rows**  
**2.5 Million Patients**

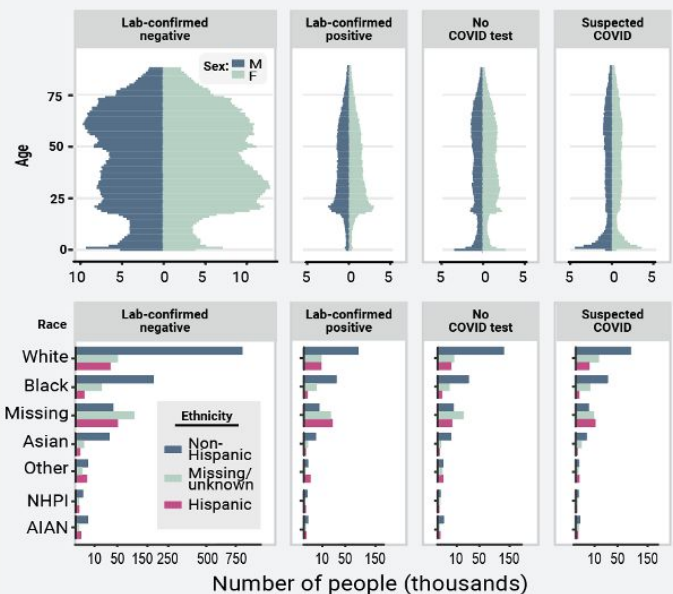
COVID-19 Positive Patients*	Total Patients	Sites	Total Rows	
413,228	2,509,112	36	2.6b	
Procedures	Lab Results	Drug Exposures	Visits	Observations
185.6m	1.1b	472.9m	134.1m	185.5m

### A. Geographic Distribution of N3C Cohort

Region Name	#	Cumulative In N3C Cohort	Monthly COVID+ Trend (Mar-Nov)
New England	1	5,560	
West South Central	2	23,003	
Pacific	3	66,974	
Middle Atlantic	4	75,208	
East South Central	5	95,746	
Mountain	6	193,343	
West North Central	7	277,347	
East North Central	8	452,921	
South Atlantic	9	736,424	



### B. Age, Sex, Race, and Ethnicity Distributions of N3C Cohort



### C. Comorbidity Distribution of COVID+ in N3C Cohort

	Mild		Hospitalized		All COVID positive	Hospitalized
	outpatient (no ED)	ED	Moderate	Severe		
N	121,078	21,018	25,907	2,790	174,568	32,472
Diabetes Mellitus	7.1%	10.9%	24.5%	27.4%	11.0%	25.9%
Renal Disease	2.3%	3.5%	13.6%	13.3%	4.8%	15.0%
Congestive Heart Failure	1.6%	2.5%	11.2%	11.1%	3.7%	12.4%
Chronic Pulmonary Disease	7.0%	10.4%	16.6%	14.0%	9.2%	17.0%
Peripheral Vascular Disease	2.6%	4.0%	10.7%	8.7%	4.4%	11.5%
Stroke	1.7%	2.8%	8.9%	8.4%	3.3%	9.8%
Cancer	3.0%	3.4%	9.3%	6.7%	4.3%	9.8%
Dementia	0.5%	0.6%	4.0%	2.7%	1.3%	5.0%
Myocardial Infarction	0.9%	1.6%	5.4%	6.1%	1.9%	6.1%
Liver Disease	2.0%	3.2%	6.4%	6.7%	3.0%	6.7%
Rheumatologic Disease	1.8%	2.3%	4.0%	3.4%	2.3%	4.0%
Hemiplegia or Paraplegia	0.2%	0.3%	1.8%	1.9%	0.6%	2.0%
Peptic Ulcer Disease	0.4%	0.5%	1.3%	1.3%	0.6%	1.5%

## Cohort Characterization manuscript

### Predicting Clinical Severity using machine learning (64 input variables)

Variable	Random Forest	XG Boost	Logistic Regression			Mean
			None	L1	L2	
pH	0	0	1	1	1	0.6
Age at visit start (years)	3	4	0	0	0	1.4
Respiratory rate	7	3	2	2	2	3.2
Oxygen saturation (SpO2)	2	2	5	4	3	3.2
Systolic blood pressure (SBP)	9	8	4	3	5	5.8
Blood urea nitrogen (BUN)	1	1	11	11	10	6.8
Albumin	21	6	6	5	4	8.4
Lactate	18	9	7	7	7	9.6
C-reactive protein (CRP)	16	11	8	8	8	10.2
Aspartate aminotransferase (AST)	6	5	12	22	18	12.6
Absolute neutrophil count	12	25	9	9	9	12.8
Ethnicity = missing or unknown	45	7	3	6	6	13.4
Glucose	5	18	16	15	14	13.6
Platelet count	10	22	14	13	13	14.4
Diastolic blood pressure (DBP)	13	15	21	17	16	16.4
B-type natriuretic peptide (BNP)	32	12	18	14	15	18.2
Sodium	15	27	15	18	17	18.4
Troponin	14	10	30	21	23	19.6
Erythrocyte sedimentation rate (ESR)	36	30	13	12	12	20.6
Sex = male	38	36	10	10	11	21
Body weight	19	49	19	16	19	24.4
Hemoglobin	17	51	24	20	21	26.6
Charlson Diabetes mellitus	37	24	25	27	24	27.4

The most powerful predictors are patient age and widely available vital sign and laboratory values.



National  
COVID  
Cohort  
Collaborative

# N3C team Science within & across institutions

## CTSA

### Key functions can nucleate projects:

- Education & training
- Biostatistics
- Study design
- Evaluation
- Informatics
- Clinical expertise
- Innovation & commercialization
- Community & partnerships



**Roles**

**Ingredients (Methods, datasets, instruments)**

**Scientific questions**

## N3C Domain Team Expertise:

- Enclave technology
- Data model (OMOP)
- Terminologies
- Data quality
- Codesets, variables, phenotype
- Using/parsing N3C data
- Workflows, methods, algorithms

<https://covid.cd2h.org/domain-teams>



National  
COVID  
Cohort  
Collaborative

# A Tsunami of Acute Kidney Injury

## The New York Times

### *An Overlooked, Possibly Fatal Coronavirus Crisis: A Dire Need for Kidney Dialysis*

Ventilators aren't the only machines in intensive care units that are in short supply. Doctors have been confronting an unexpected rise in patients with failing kidneys.



Miriam Figueroa, a dialysis nurse at the Brooklyn Hospital Center, tended to a patient in the intensive care unit on Thursday. Victor J. Blue for The New York Times

April 2020

### *A Life and Death Battle: 4 Days of Kidney Failure but No Dialysis*

At the peak of the pandemic in New York, a longtime city employee joined a crush of patients in desperate need of treatment.



May 2020



National  
COVID  
Cohort  
Collaborative

# Kidney Domain Team



NATIONAL CENTER  
FOR DATA TO HEALTH



**Stony  
Brook**

## PREMISE:

Kidney Injury contributes importantly to mortality and leads to long term mortality.



**Sandeep  
Mallipatu**

## QUESTIONS:

What factors predict kidney injury early?  
Are there regional differences in AKI incidence?  
Is AKI incidence in COVID-19 decreasing?  
What are the predictors of ESKD after AKI?



**Richard Moffit**

**Chirag  
Parikh**



**Johns  
Hopkins**

**Stephanie  
Hong**



**Farrukh  
Koraishy**

**Faifan Liu**



**David H.  
Ellison**



**J. Brian  
Byrd**



**Ivonne  
Schulman**



National Institute of  
Diabetes and Digestive  
and Kidney Diseases

<https://covid.cd2h.org/kidney/>



# N3C Upcoming Events

Monday, Jan 11 <sup>th</sup>	9-11am PT/12-2pm ET	<a href="#">N3C Data Harmonization Symposium</a>
Tuesday, Jan 12 <sup>th</sup>	8am PT/11am ET	N3C Pilot Orientation Part A
Friday, Jan 15 <sup>th</sup>	11-2pm PT/ 2-4 ET	<a href="#">N3C Nurses &amp; Nurse Scientists Webinar</a>
Tuesday, Jan 19 <sup>th</sup>	9am PT/12pm ET	<a href="#">N3C Community Symposium &amp; N3C Open House Week</a> Jan 19th - 25th
Tuesday, Jan 19 <sup>th</sup>	8am PT/11am ET	N3C Pilot Orientation Part B
Monday, Feb 1st	9am PT/12pm ET	<a href="#">Trial Innovation Network (TIN) Collaboration Webinar</a> - <i>Using N3C Data to Inform your Protocol Development</i>

<https://covid.cd2h.org/n3c-calendar>

# Benefits to CTSA (and other) Organizations

- Access to large scale COVID-19 data from across the nation
- Pilot data for grant proposals
- Opportunities for KL2 and TL1 and other scholars
- Team science opportunities for new questions and access to Teams, statistics, machine learning (ML), informatics expertise
- Learn ML analytics, NLP methods & access to tools, software, additional datasets







# Help, Support & Training

**N3C Support Desk page:** [https://covid.cd2h.org/help\\_desk](https://covid.cd2h.org/help_desk)

## Support Desk

View Edit Delete Layout Revisions

If you need assistance with the N3C Data Enclave, the following options are available:

-  Attend Support Desk office hours on **Tuesdays & Thursdays at 10-11 am PT/1-2 pm ET** [Register here](#).
-  [Submit a Support Request](#).
-  Consult the [Frequently Asked Questions \(FAQs\)](#).
-  A quick start tour is available inside the Enclave.

Orientation will be scheduled in January/ Tutorials available in the Enclave

**NCATS**

COLLABORATE. INNOVATE. ACCELERATE.

Numerous  
Projects

Numerous  
People

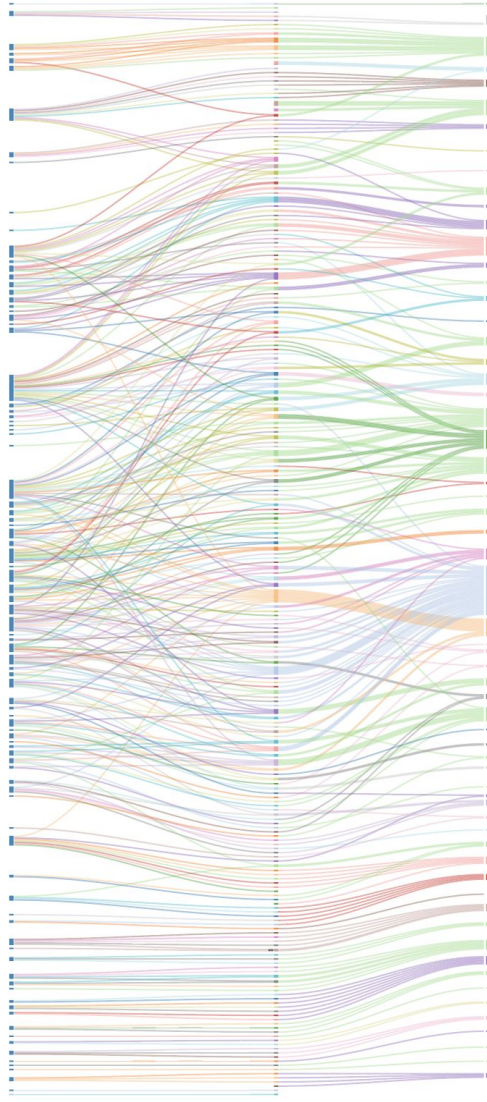
Numerous  
Institutions

Expertise &  
Resources  
Connected

**140 Projects**

diium glucose co-transporter 2 inhibitor (SGLT2) use and COVID-19 outcomes: A national retr  
 COVID protein primary and sub-structures to interrupt protein activity followed by epidemiologica  
 and in-vivo v  
 antibody response to SARS-CoV-2 in people with multiple sclerosis treated with B cell depleting t  
 Alcohol Use and Respiratory Outcomes in COVID-19 Ir  
 Smoking and COVID-19 Outcomes in U.S. Adults (S  
 Validate a Machine Learning Model to Predict Decompensation in Patients with C  
 Workflow Construction with Synthetic Data  
 Understanding non-invasive ventilation treatment failures in C  
 N3C Cohort Charact  
 COVID-19 in individuals with Down's Si  
 Examining Associations between Vitamin D Status and COVID-19 Test  
 Identification of Novel COVID-19 Subphenotypes Using Temperature Tra  
 Identification of Novel COVID-19 Subphenotypes Using Temperature Tra  
 An evaluation of Direct Acting Anticoagulants and Dexamethasone in Patients with COVID-19 Ir  
 N3C Diabetes and Obesity Domain Team level 2 request  
 [N3C Operational] Data Ingestion and Harm  
 Studying COVID-19 Remission, Recrudescence, Recurrence, and Re  
 [N3C Operational] Implementation of Syntegra Synthetic Data Gr  
 Impact of medication on outcomes for diabetic  
 Monocytopenia in C  
 Acute Kidney Injury in Pediatric COVID-19  
 Using Machine Learning to differentiate COVID-19 infection with sea  
 COVID-19 and Percutaneous Catheterization Interventions: Modeling Risk of Serious Adverse O  
 COVID-19 infection and mortality among individuals with  
 clinical variables to predict needs for invasive mechanical ventilation and mortality in COVID-19  
 Continued In-Hospital ACEi and ARB Use in Hypertensive COVID-19  
 AI-Assisted Assessment, Tracking, and Reporting of COVID-19 Severity on C  
 19 infection in patients with pulmonary non-tuberculous Mycobacterium (NTM) infection: A coh  
 Disparities in C  
 [N3C Operational] Collaborative/  
 rsarial Network (GAN) based synthetic EHR data generation methods for enabling COVID-19  
 Outcomes of Acute Pulmonary Embolism in C  
 [N3C Operational] Synthetic Data Validation Us  
 N3C Pregnancy Task Team: COVID-19 Incidence, Treatment, and Outcomes in Pregnant  
 Oution of Race and other Social Determinants of Health to Disparities in Patient Outcomes C  
 Hospitalization Rates Feb-April vs May  
 The Effect of COVID-19 Stay-At-Home Orders on Hemoglobin A1c Levels in Diabetic  
 Developing Dynamic Graphs for Prediction and Classification of COVID-19 Patient Tra  
 Occurrence of neurological diseases in hospitalized patients with COVID-19  
 COVID-19 and Air  
 ing health system in the COVID-19 pandemic: analysis of patient data to assist clinical decisor  
 Protective Effects of Medications Against SARS-CoV-2 Infection in Patients with Existing Pres  
 (1-RA) and sodium glucose co-transporter 2 inhibitor (SGLT2) use and COVID-19 outcomes: A  
 retrospective coh  
 Multiorgan Dysfunction Syndrome and Complex Clinical Ti  
 Data-driven Investigation of Health Disparities in COVID-19 Outcomes: A Focus on Behavior  
 Investigating the impact of biases on the performance of analytical models for COVID-19 retail  
 Factors affecting risk of poor outcomes for COVID-19 for patients with  
 us Process: Varying Underlying Pathophysiological Mechanisms at Different Time Points and S  
 Orthopaedic Surgery in the National COVID Cohort Collaborat  
 Modeling of COVID-19 Risks for Back-to-Work P  
 COVID-19 impact on pregnancy and fetal v  
 Treatment associated with lower mortality in hospitalized COVID-19  
 Stage Aware Prediction of COVID-19 Disease Pro  
 Stroke and COVID Population: An N3C and UVA OMOP  
 Association of Dupilumab with Protection from COVID-19 Respirator  
 Cardiovascular Complications of C  
 social determinants of health with regards to differential COVID-19 incidence across the Unite  
 Use, Safety and Effectiveness of Therapies to Treat C  
 N  
 he severity of COVID-19 infections of individual patients by applying artificial intelligence metho  
 he severity of COVID-19 infections of individual patients by applying artificial intelligence metho  
 N3C ImmunoSuppressed/Compromised Task Team: The Impact of COVID-19 on the ISC Pr  
 Epidemiology of Acute and Chronic Kidney Injury Associated SARS-CoV-2  
 Analysis of Time Windows as Determining Factors for COVID-19 Patient O  
 mporal Lab Value Changes and Medications as Predictors of Health Outcomes for COVID-19+  
 [N3C Operational] Phenotype and Data Acquisition Team Or  
 Retrospective identification of medications to treat C  
 Adaptive and interpretable machine learning to predict COVID-19 trajectory and  
 Case-Control Studies of Medications and Their Possible Associations with COVID-19  
 HIV and COVID-19: Effect of Shelter-In-Place Orders on Virologic Sup  
 Impacts of COVID -19 in Older Adults (Elder Impact Domain  
 Impacts of COVID -19 in Older Adults (Elder Impact  
 Clinical Outcomes for Drug Repurposing Candidates Discovered in BSL3 SARS-CoV-2 Infect  
 Study Phenotype of COVID-19 in Patients with Hemophilia: A National Coh  
 2 infection in African Americans: AI-based Novel Approach of Prognostic and Risk Stratifier  
 COVID-19 Synthetic Data V  
 Clinical Characterization of Critically-ill COVID-19 Patients based on Ventilation  
 Study effects of COVID-19 Infection on cancer  
 wide association study of COVID-19 outcomes using electronic health records and informatic r  
 Predicting organ failure in patients with COVID-19

**1231/561 People**



**313 Institutions**

- Southern Illinois University School Of Medicine
- Washington University In St. Louis
- University Of Virginia
- Oregon State University
- Cornell University
- University Of Rochester
- The University Of Texas At Austin
- George Washington University
- National Institute Of Allergy And Infectious Diseases
- University Of Washington
- The University Of Tennessee Health Science Center
- The University Of Iowa
- Citizen Scientist
- The University Of Texas Medical Branch At Galveston
- University Of Kentucky
- Duke University
- Medical College Of Wisconsin
- University Of Colorado Denver I Anschutz Medical Campus
- Virginia Commonwealth University
- Oregon Health & Science University
- Palila Software
- University Of Alabama At Birmingham
- Stony Brook University
- University Of North Carolina At Chapel Hill
- Mayo Clinic Rochester
- University Of Pennsylvania
- University Of Minnesota
- Novo Nordisk
- Rutgers, The State University Of New Jersey
- Johns Hopkins University
- Indiana University
- National Institute Of Diabetes And Digestive And Kidney Diseases
- National Center For Advancing Translational Sciences
- National Cancer Institute
- Albert Einstein College Of Medicine
- Sage Bionetworks
- University Of Nebraska Medical Center
- Auburn University
- University Of Massachusetts Medical School Worcester
- Brown University
- University Of Arizona
- Wake Forest University Health Sciences
- All Science
- National Library Of Medicine
- Vanderbilt University Medical Center
- University Of Pittsburgh At Pittsburgh
- Colorado State University
- Emory University
- Syntegra
- Mount Sinai Genomics Inc
- University Of Oklahoma Health Sciences Center, Board Of Regents
- University Of Illinois At Chicago
- The University Of Chicago
- University Of Texas Health Science Center At San Antonio
- University Of Houston
- Northwestern University At Chicago
- National Human Genome Research Institute
- The Jackson Laboratory
- University Of California, Davis
- University Of Missouri
- The University Of Michigan At Ann Arbor