

New COVID-19 Data Availability under Health Data for Action Program

Background

The purpose of Health Data for Action (HD4A), a program of the Robert Wood Johnson Foundation (RWJF) administered by AcademyHealth, is to connect researchers with high value data to further discoveries that will advance the Foundation's mission to achieve health equity. Since the pandemic, RWJF has been working to more closely link the purpose and potential of Health Data for Action with the data needs generated by COVID-19. The scheduled 2020 HD4A call for proposals was postponed to enable RWJF to prioritize making COVID-19-related data broadly available as quickly as possible. In recent months, RWJF has worked with data providers that have timely data and sought ways to make that data more readily available.

New COVID-19 Data Availability

The first new data availability is now available under the COVID-19 prioritization of Health Data for Action. athenahealth, a provider of electronic health record (EHR) and claims software to outpatient provider organizations spanning all 50 states and a longtime HD4A partner, is making de-identified data available that can improve understanding of visits to clinicians related to COVID-19. To help the research community better understand the impact of COVID-19, athenahealth is making available two de-identified data sources, which can be linked. The first contains information about patients, including demographic and insurance characteristics, and co-morbidities, along with a unique patient ID. These data can be matched to a second file, which contains information on COVID-19 related outpatient visits, including symptoms, testing orders, and results. These data can help researchers better understand important issues related to care-seeking and case identification, for example how race, ethnicity, co-morbidities, or other factors may be related to the presence of symptoms, receipt of testing, and test results.

A Look into the Data

As of August 10, 2020, providers on the athenahealth network have tested more than a million patients. athenahealth is able to report large racial and ethnic disparities in positive test results. For example, during the week of July 6, 2020, in Texas 34% of Hispanics and 26% of non-Hispanic Blacks tested positive for COVID-19, as compared with 18% of non-Hispanic Whites. These data will be updated weekly and summaries including a map can be seen on an interactive [dashboard](#).

How to Get the Data

In an effort to speed research related to COVID-19, these data are being made available to researchers immediately (pending proposal selection and a signed data use agreement). The data are being made available exclusively for non-commercial use, and while RWJF and athenahealth is not engaging in a process of formally reviewing proposals, if necessary RWJF and athenahealth will prioritize requests that are most closely related to improving our collective response to COVID-19. athenahealth will review requests and execute data use agreements with researchers. A more detailed description of the data is available below.

Contact Us

Please email research@athenahealth.com with questions, and/or to apply for data access with a 1-page, single spaced research proposal briefly covering:

- Title, authors, institutional affiliation(s)
- Background/significance
- Research question(s)
- Methods
- Summary of computing and data security infrastructure available through your organization

Applications will be accepted on a rolling basis but should be submitted no later than **October 15, 2020**.

athenahealth COVID-19 research dataset

Overview

athenahealth is a leading provider of cloud-based software for ambulatory providers nationwide, offering medical record, revenue cycle, patient engagement, care coordination, and population health services. athenahealth providers are largely based in ambulatory care settings, at organizations of all sizes, and in both urban and rural communities.

athenahealth collects claims and EHR data across all 50 U.S. states today. Patients visiting athenahealth providers are broadly representative of the nation's outpatient visits when compared to national benchmarks provided by the National Ambulatory Medical Care Survey (NAMCS).

athenahealth plans to share with selected researchers and public sector partners two de-identified, cross-linked datasets addressing outpatient care activities related to COVID-19, based on visits to clinicians using the athenaClinicals EHR system.

The first dataset establishes a static cohort of patients on the athenahealth network, whose COVID-19-related activity can then be tracked over time. It includes detailed information on patient medication history, demographics, and chronic condition status. This will be shared with researchers as a one-time transfer.

The second dataset provides dynamic, longitudinal information on COVID-19-related activity (symptom diagnoses, lab tests and results, and COVID-19 diagnoses). This information will be updated weekly, and will be shared with researchers as an ongoing feed. As all identifiers will be randomized to protect privacy, the dataset will not be linkable to outside data sources other than through high-level geography and time (state & week).

The two datasets described below are aimed at supporting research across a wide variety of study areas related to COVID-19. The datasets will be made accessible to researchers as specified in a signed Data Use Agreement (DUA). Details in the data dictionary below may change based on statistical expert de-identification review.

A wide range of research ideas are welcome, but suggested topics include:

- Demographic patterns in testing rates and results by race, ethnicity, gender, and age
- Geographic trends in COVID-19 cases
- Patient and clinical characteristics associated with the development of COVID-19 (e.g., comorbidities, demographics, symptoms)
- Predictive modeling for COVID outbreaks

To apply for access to the data, please email research@athenahealth.com with a 1-page, single-spaced research proposal briefly covering:

- Title, authors, institutional affiliation(s)
- Background/significance
- Research question(s)
- Methods
- Summary of computing and data security infrastructure available through your organization

Applications will be accepted on a rolling basis but should be submitted no later than **October 15, 2020**.

Table 1: Patient Cohort View

- Timeframe: Static cohort of patients
- Sample: More than 15 million patients who visited a primary care clinician using athenaClinicals EHR software between Mar 1, 2018 and Feb 29, 2020

Variable/Category	Levels (where applicable)	Notes
U.S. state	[50 U.S. states + DC]	
Randomized patient ID		Can be cross-linked to Table #2
Patient gender	Male Female	
Patient age	<18 18-34 35-44 45-54 65+	
Patient race/ethnicity	Black or African American Hispanic/Latino Other White NULL	
Patient insurance type	Commercial Medicaid Medicare Other Self-Pay	
Patient comorbidities	0 1 2+	Conditions include: Type 2 diabetes, cancer, heart conditions, immunosuppressed with organ transplant, obesity, and sickle cell anemia, COPD, and chronic kidney disease

Table 2: COVID-19 Event View

- Timeframe: Updates weekly beginning on Mar 1, 2020
- Sample: As of mid-Aug, more than 5 million patients with a COVID-related event (lab test, COVID diagnosis, or COVID symptom diagnosis) recorded in athenaClinicals EHR software since Mar 1, 2020

Variable/Category	Levels (where applicable)	Notes
Randomized COVID-19 event ID		
U.S. state	[50 U.S. states + DC]	
Week		Week in which COVID-related event took place
Lab test types: PCR Antigen Serology	1/0	Binary indicator for each lab test type performed
Lab result values: PCR = positive PCR = negative Antigen = positive Antigen = negative Serology = positive Serology = negative	1/0	Binary indicator for each lab result value; a lab test without either positive or negative result value indicates pending or unknown result
COVID diagnosis flag	1/0	Binary indicator for COVID diagnosis
COVID symptoms: Dyspnea/shortness of breath Fever/chills Sore throat Cough Loss of taste/smell Nausea/vomiting Nasal congestion Diarrhea Headache General/unspecified ILI Suspected/exposure to COVID	1/0	Binary indicator for each COVID-19 symptom documented on a claim
Randomized patient ID		Can be cross-linked to Table #1
Patient gender	Male Female	
Patient age	<18 18-34 35-44 45-54 65+	
Patient race/ethnicity	Black or African American Hispanic/Latino Other White NULL	
Patient insurance type	Commercial Medicaid Medicare Other Self-Pay	