Update on the All of Us Research Program

September 17, 2021

Josh Denny, MD, MS
Chief Executive Officer
All of Us Research Program

@AllofUsCEO

National Institutes of Health
Nurture relationships
with one million or more participant partners, from all walks of life, for decades

Our mission
To accelerate health research and medical breakthroughs that enable individualized prevention, treatment, and care for all of us

Deliver one of the largest, richest biomedical dataset that is secure and easy to access

Catalyze a robust ecosystem of researchers and funders hungry to use and support it
The *All of Us* Research Program: An Innovative Research Effort

- Diversity at the scale of 1 million people or more
- **Focus on participants as partners**, with return of value as a priority
- **Longitudinal design**, ability to recontact participants
- **Multiple data types**: EHR, surveys, baseline physical measurements, biospecimens, genomics, and more
- **National, open resource for all**: broadly accessible to all researchers with open-source software & tools
- **Security and privacy safeguards** for all participants
Status of the *All of Us* Research Program (as of September 6, 2021)

- **406,000+** Participants
- **253,000+** Electronic Health Records
- **295,000+** Participants who have completed initial steps of the program
- **310,000+** Biosamples

COVID-19 in-person enrollment pause

In-person enrollment beginning to restart

researchallofus.org
Status of the All of Us Research Program (as of September 6, 2021)

Race and Ethnicity

- White: 47.3%
- Black, African American, or African: 21.4%
- Hispanic, Latino, or Spanish: 17.1%
- Asian: 6.5%
- More than one race/ethnicity: 2.9%
- Other: 2.9%
- Prefer not to say: 0.7%

Over 80% of All of Us participants are underrepresented in biomedical research.

Age

- 18–29: 10.7%
- 30–39: 15.0%
- 40–49: 14.4%
- 50–59: 19.9%
- 60–69: 21.4%
- 70–79: 14.1%
- 80–89: 3.9%
- 89+: 0.5%

researchallofus.org
Data Collected from *All of Us* Participants

- Consent and Electronic Health Records
- Participant Surveys
- Physical Measurements
- Biosamples
- Mobile/Wearable Tech
Data Collected from *All of Us* Participants: Surveys

Current surveys focused on:
- Demographics and Lifestyle
- Personal and Family Medical History
- Healthcare Access
- **COVID Participant Experience (COPE)**

Next in queue:
- Social Determinants of Health

<table>
<thead>
<tr>
<th>Consent and Electronic Health Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant Surveys</td>
</tr>
<tr>
<td>Physical Measurements</td>
</tr>
<tr>
<td>Biosamples</td>
</tr>
<tr>
<td>Mobile/Wearable Tech</td>
</tr>
</tbody>
</table>
Data Collected from *All of Us* Participants: Biosamples

- Blood
  - DNA
  - RNA
  - cfDNA
  - Serum
  - Plasma
- Saliva kits (if not blood)
- Urine

Consent and Electronic Health Records
Participant Surveys
Physical Measurements
Biosamples
Mobile/Wearable Tech
COVID-19 Response - Serology Study
COVID-19 Serology Study

• Tested 24,079 blood specimens collected between January 2 and March 18, 2020

• Tested spike and nucleocapsid antibodies on multiple platforms

SARS-CoV-2 antibodies identified prior to the first recognized cases in 5 U.S. states
COVID-19 Serology Study Published in *Clinical Infectious Diseases* (June 15, 2021)

9 cases in 5 states (IL, MA, WI, PA, MS) in 24,079 tested samples
7 earlier than known cases in those states
Earliest case January 7, 2020
Press Coverage of the *All of Us* COVID-19 Serology Study

**The Washington Post**

NIH study suggests coronavirus may have been in U.S. as early as December 2019

**The Wall Street Journal**

U.S. Covid-19 Ranged From Illinois to Massachusetts Before States Reported First Cases

Blood samples show people in five U.S. states were infected early, including as late as December 2019.

**CNN**

NIH researchers find more evidence Covid was circulating in the US in December 2019

By Maggie Fox, CNN

Updated 1:57 PM ET, Tue June 15, 2021

**AP**

More evidence suggests COVID-19 was in US by Christmas 2019

By MIKE STOBBE

June 15, 2021
Making Research Reproducible: Any Published *All of Us* Analysis Can Be Accessed by Any Other *All of Us* Researcher

Reference:

**How to Reproduce the All of Us SARS-CoV-2 Antibody Study**

Author: Aymane Kouame
Contributors: Francis Reisimbrazly, Adrienne Roman

**Objectives**

We recommend that users read this notebook to learn how to reproduce the Antibodies to SARS-CoV-2 in All of Us Research Program Participants, January 2- March 18, 2020 study published on June 15, 2021 in *Clinical Infectious Diseases*. For the best experience, we recommend opening this notebook in Playground Mode.

What should you expect? This notebook will give an overview of the Antibodies to SARS-CoV-2 in All of Us Research Program Participants, January 2-March 18, 2020 study in the current Curated Data Repository (CDR) and how to reproduce it.

This tutorial is divided into the following sections:
1. **Setup**: How to set up this notebook and install and import all necessary software packages.
2. **Materials and Methods**: What are the characteristics of the study population and what are the methods used in study.
3. **Data Extraction**: How to query the relevant CDR tables to extract the study data.
4. **Results**: How to reproduce the results of the study.

**Notebook URL**: [https://bit.ly/3h7GnEF](https://bit.ly/3h7GnEF)

Any *All of Us* researcher can review and reuse the exact data and analyses used in the paper
All of Us Research Program Core Values

Return of information

- Participation in **open** to all.
- Participants reflect the rich **diversity** of the U.S.
- Participants are **partners**.
- Trust will be earned through **transparency**.
- Participants have **access** to their information.
- Data will be accessed **broadly** for research purposes.
- **Security and privacy** will be of highest importance.
- The program will be a catalyst for **positive change** in research.
We Returned Serology Results to Participants

Participants can have a discussion with infectious disease expert if desired.
Current Data and Applications
**Researcher Data Access**

- **Public Tier** (Available Now)
  - Summary Statistics Aggregate Counts

- **Registered Tier** (Available Now)
  - Surveys, EHRs, Physical Measurements
  - Exceeds HIPAA Safe Harbor Standards

- **Controlled Tier** (Available in the future)
  - No obvious PII
  - Genomics, Clinical Narrative data, Data Linkages, Other Data Types

- **Future ancillary studies**
  - Could recontact participants, use biospecimens, issue new surveys or DHT, enroll in clinical trials

**Data Browser**
- databrowser.researchallofus.org
  - (public, no login required)

**Researcher Workbench**
- ResearchAllofUs.org/Apply/
  - (Winter 21/22)

**Data Curation**
- Raw Data Repository
  - Data Harmonization
  - Privacy
  - Methodology
  - QA/CC

**Future**
Summary statistics of participant data

- **EHR Data** (Conditions, Drug Exposures, Lab & Measurements, Procedures)
- **Survey Questions** (including COVID-19 surveys)
- **Physical Measurements**
- **Open access** (no login required)
All of Us Research Hub: Data Browser – EHR Conditions

Search Across Data Types

Type 2 Diabetes

Data includes 316,760 participants and is current as of 10/1/2020.

EHR Domains:

Conditions

119 matching medical concepts
192,000 participants in this domain

View Results

Survey Questions:

No Survey Results. Please type in a new search term.

Physical Measurements and Wearables:

No Program Physical Measurement Results. Please type in a new search term.
Researcher Data Access

- **Public Tier** *(Available Now)*
  - Summary Statistics Aggregate Counts

- **Registered Tier** *(Available Now)*
  - Surveys, EHRs, Physical Measurements
  - Exceeds HIPAA Safe Harbor Standards

- **Controlled Tier** *(Available in the future)*
  - No obvious PII
  - Genomics, Clinical Narrative data, Data Linkages, Other Data Types

Future ancillary studies *(future)*
- Could recontact participants, use biospecimens, issue new surveys or DHT, enroll in clinical trials

- **Researcher Workbench** *(Winter 21/22)*
  - ResearchAllofUs.org/Apply/

- **Data Browser** *(now)*
  - databrowser.researchallofus.org
    - (public, no login required)

- **Data Curation**
  - Raw Data Repository
  - Data Harmonization
  - Privacy
  - Methodology
  - QA/CC

Individual Biospecimen and Participant Data *(Available in the future)*
All of Us Researcher Workbench: Access to Row-Level Data for Analysis

Researcher Workbench Beta
Launched on May 27, 2020

- Cloud based central resource
- Personally-identified information is removed

**Passport access model** - just create, describe your workspace, and get to work! No separate IRB approval needed.

- During beta phase, access requires eRA commons ID and limited to US nonprofits

ResearchAllofUs.org
Researcher Workbench (Launched May 2020): By the Numbers

Data Available on the Researcher Workbench

- 329,000+ Participants
- 267,600+ Physical Measurements
- 214,200+ EHRs
- 329,000+ Surveys (100,000 COPE Surveys)
- 11,600+ Fitbit Records
Researcher Workbench (Launched May 2020): By the Numbers

Research on the Researcher Workbench

1000+ Registered Researchers
660+ Active Projects
15+ Publications using All of Us data

Institutional Agreements

More than 240 registered institutions

Over 24% are Historically Black Colleges and Universities, Hispanic Serving Institutions, or Non-Profits
Participant EHRs and Fitbit Provide Longitudinal Data

- 315k participants in dataset
- 53 million conditions
- 20 million procedures
- 42 million survey answers
- 200 million labs and measurements
- 42 million drug exposures
- 15 billion Fitbit data points

Program Launch
Current CDR
March 2021, Dr. Patrick Wu Defends His PhD Dissertation: “Repurposing drugs using gene expression signatures and EHR data”

gene expression data identifies drugs that might lower cholesterol

found 69 candidate drugs

study those drugs in the Electronic Health Record data at Vanderbilt

This took ~3 years.

A question from the audience: “You did this at one site. What about replication?”
One Week Later: He has applied for the first time to use *All of Us*, completes his onboarding, and has replicated his results in *All of Us*...
What’s coming next: Controlled Tier Researcher Data Access

- **Public Tier** (Available Now)
  - Summary Statistics Aggregate Counts

- **Registered Tier** (Available Now)
  - Surveys, EHRs, Physical Measurements
  - Exceeds HIPAA Safe Harbor Standards

- **Controlled Tier** (Available in the future)
  - No obvious PII, Genomics, Clinical Narrative data, Data Linkages, Other Data Types

- **Future ancillary studies**
  - Could recontact participants, use biospecimens, issue new surveys or DHT, enroll in clinical trials

**Data Browser**
databrowser.researchallofus.org
(public, no login required)

**Researcher Workbench**
ResearchAllofUs.org/Apply/
First Genomics Data Release is coming Winter 2021/2022

- Expected 90,000 WGS + 130,000 arrays
  - >40% non-White
- More COVID-19 data
- More detailed demographic & EHR data
First Ancillary Study: Nutrition for Precision Health

**Goal:** Enroll 10,000 *All of Us* participants to study to develop algorithms to predict individual responses to foods and dietary patterns

- Build on *All of Us* to add a comprehensive set of microbiome, physiological, metabolic, behavioral, cognitive, contextual, survey, and environmental data
- In large and diverse population of participants
**All of Us Roadmap**

**May 2019**
- Public Tier Data: Data Browser Launch

**May 2020**
- Launch of Researcher Workbench: Public Beta Launch
- Data Browser Launch
- Updated data releases

**December 2020**
- COVID-19 Serology return of results
- Return of Genetic Results

**Q2 2021**
- Health Related Genetic Results (ACMG & Pharmacogenomics)
- Return of Genetic Results

**Winter 2021/2022**
- Controlled Tier data with initial genomics release and expanded COVID-19 data
- Return of Genetic Results
- Ancestry and Traits Results

**Future:**
- Ancillary studies, More Data & Data Linkages, Participant Recontact, & Results Return

**We Are Here**
- Expected 2022

**September 2021**
- Registered Tier data refresh (More participants, COPE surveys, Fitbit data)

**November 2021**
- Expected 2022

**December 2020**
- Return of Genetic Results

**May 2018**
- Enrollment launched