Progress, Challenges, and Opportunities in Implementing a Nationwide, Large-scale Digital Research Platform for Precision Medicine

April 11-12, 2023

Changing the Future of Public Health and Biomedical Research through Cloud-Based Computing: Lessons From the NIH *All of Us* Research Program

Romuladus Azuine, DrPH, MPH, RN (NIH)
Nakia Mack, MPH, MS (NIH)
Lew Berman, PhD (NIH)
Chris Lunt (NIH)

Opportunities and Challenges in Implementing a Large-Scale Participant Digital Research Data Collection

James McClain, Ph.D (NIH)
Izzy Seo, MPH (NIH)
Mark Begale (Vibrent Health)

Working with NIH *AoU* Researcher Workbench: A Guided Tour

Gage Rion, ME (Vanderbilt University Medical Center)
Danielle Wilfong, M.S., Ed.M. (Vanderbilt University Medical Center)
Samantha Lee Stewart, M.S.(Vanderbilt University Medical Center)

NIH National Institutes of Health

All of Us RESEARCH PROGRAM

Changing the Future of Public Health and Biomedical Research through Cloud-Based Computing: Lessons From the NIH *All of Us* Research Program

April 11-12, 2023

Romuladus Azuine, DrPH, MPH, RN (NIH)
Nakia Mack, MPH, MS (NIH)
Lew Berman, PhD (NIH)
Chris Lunt (NIH)

NIH National Institutes of Health

What is the *All of Us* Research Program?

About the *All of Us* Research Program

WHAT
is the *All of Us* Research Program?

It is one of the largest, richest, most diverse biomedical datasets of its kind

The *All of Us* Research Program will nurture partnerships for decades with **at least one million people who reflect the diversity of the United States** to help build one of the largest, most diverse health datasets in history.

By engaging the participation of people and communities who have been left out of medical research in the past and making their data available to diverse researchers across a wide range of settings, the program seeks to **accelerate health research and medical breakthroughs, enabling individualized prevention, treatment, and care for all of us.**

WHY
it matters?

It combines biological factors and social determinants on a large, inclusive scale

Research takes data, and lots of it, from diverse populations and spanning many different types: biological, environmental, behavioral, genomic, and sociodemographic. But this data is not easy to capture in one place.

The program **captures a wide variety of data types** — survey responses, physical measurements, biosamples, EHRs, genomics, and data from wearables — enabling the acceleration of research breakthroughs. Its longitudinal dataset **follows participants** as they move, age, get sick, and try treatments.

WHO
benefits?

It equips researchers to make discoveries that will enable more precise approaches to care, informing providers' recommendations and individuals' choices

Participants represent a broad cross-section of the U.S. population (413K* who have completed the initial steps of the program*)

Researchers come from diverse settings and career stages (>4,700 registered researchers across >490 research institutions around the United States, and counting!**)

HOW
do I access it?

It is available to researchers with login.gov credentials and institutional sign-off across a wide range of settings

The program's expansive dataset is housed on a secure, cloud-based platform that **registered researchers can easily access with just a computer and internet connection.**

Data are curated into research datasets with tiered access, including Public, Registered, and Controlled Tiers.

See more about access requirements on slide 17

*Source: Data Snapshots, as of March 2023

**Source: DRC Workbench Executive Report data, March 2023

One of the largest, richest, most diverse biomedical datasets of its kind

Inviting at least

1 Million people from across the United States

Data available from **372,000+** participants

80% are from communities underrepresented in biomedical research

45% are from racial and ethnic minority groups

Data as of June 2022

Enables research discoveries that drive more precise approaches to care

Engages people & communities who have been left out of medical research in the past

Combines biological factors and social determinants on a large, inclusive scale

Easily accessible to any researcher with a secure internet connection and data use agreement

Follows participants as they move, age, and grow

By securely capturing a range of different data types

Surveys

(from 372,000+ participants)
Including lifestyle, access to care, medical history, and **data from nearly 100,000 participants on their experiences during the COVID-19 Pandemic**

Electronic Health Records

(from 258,000+ participants)
Standardized to OMOP common data model

Physical Measurements

(from 306,000+ participants)
Blood pressure, heart rate, BMI and more

Wearables

(from 12,800+ participants)
Physical activity and heart rate from participants who have connected their Fitbit devices

Genomics

(Nearly 100,000 whole genome sequences) and 165,200 shorter DNA samples known as genotyping arrays
Only available via the Controlled Tier

Biosamples

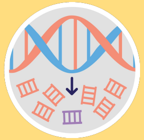
Additional data from biosamples to come

Data as of June 2022

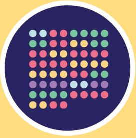
Including the first batch of genomic data

Genomics

Only available via the Controlled Tier



98,600
Whole Genome Sequences



165,200
Genotyping Arrays

593,500,000+
Unique Variants

Represented and searchable through public genomic data browser

Genomics Analysis Tools

Hail and PLINK in addition to R, Python, and Jupyter Notebooks

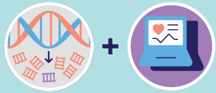
Genomic Data is Paired with Rich Phenotypic Data

81,000+

Have Whole Genome Sequences + Electronic Health Records + Physical Measurements + Survey Responses

98,000+

Have Whole Genome Sequences + Physical Measurements + Survey Responses



81,000+

Have Whole Genome Sequences + Electronic Health Records

3,300+

Have Whole Genome Sequences + Fitbit Records

Representing >25% of all participant Fitbit records

And data from the recurring COVID-19 Participant Experience (COPE) survey

COPE Survey Data Currently Available in the Researcher Workbench

More than **100,300**

participants responded to

6

one or more of the **six** COPE surveys administered between May 2020 and March 2021

*45,547 participants responded to three or more surveys.

Topics Covered

- **Social Distancing Experiences**
- **COVID-19 Related Impact**
- **Mood**
- **Substance Use**
- **COVID-19 Related Symptoms**
- **General Well-Being**
- **Stress**
- **Resilience**
- **COVID-19 Related Testing**
- **Social Support**
- **Physical Activity**
- **Discrimination**
- **COVID-19 Related Treatment**
- **Anxiety**
- **Loneliness**
- **Vaccine Perceptions**

Mental Health

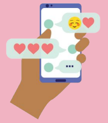
The COPE data represents the biggest infusion of mental health data into the Researcher Workbench so far. Insights include:

62%

felt bothered by sleeping problems

53%

felt nervous and anxious



94%

have someone to love and make them feel wanted

95%

have someone to have a good time with

Embedded in the survey were resources for participants struggling with emotional distress, including suicide prevention tools.

*Participants could select multiple symptoms.

And making them accessible to researchers across stages and settings

Our Researchers

4,700+

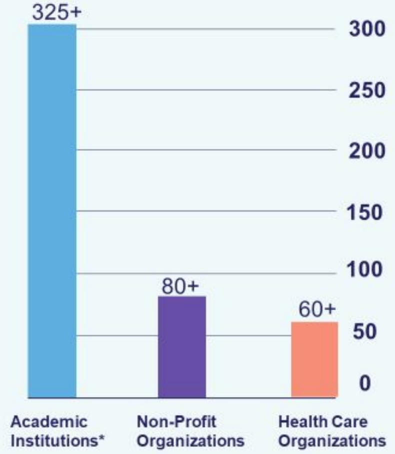
Registered Researchers

across a range of institutional roles and career stages

490+

Institutions

Researchers nationwide are using the Researcher Workbench



***Includes 75+ Historically Black Colleges and Universities and Hispanic-Serving Institutions.**

Figures accurate as of September 2022

Figures accurate as of March 2023

Research Currently Underway

4,000+
Active projects

130
Publications in peer-reviewed journals

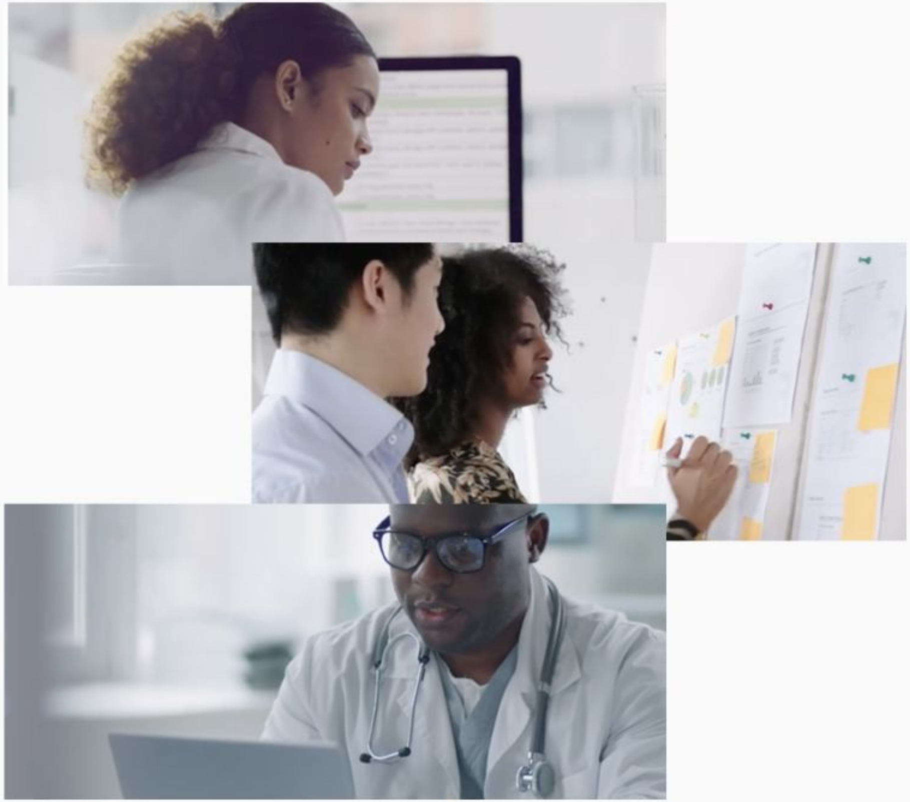
Top conditions being studied

In the Researcher Workbench include:

- Cardiovascular disease
- Hypertension
- Mental Health
- Cancer
- Diabetes

Figures accurate as of March 2023

Building a diverse researcher cohort



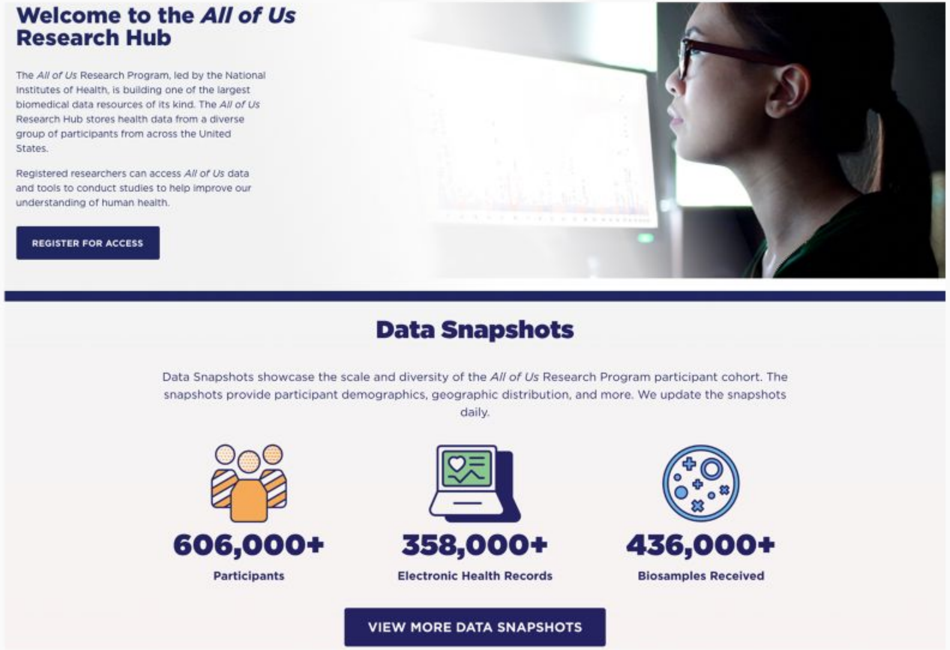
- **Creating a demographically diverse researcher cohort** that promotes responsible and ethical use of data, returns value to participant communities, and accelerates research impact.

- **Encouraging student assemblies and early-stage investigators** to bring fresh, creative perspectives & innovative research outcomes.

- **Ensuring access for researchers from various institutions/organizations** to establish a truly equitable resource for all.

With aggregated overviews and interactive previews available to everyone

ResearchAllofUs.org



Welcome to the *All of Us* Research Hub

The *All of Us* Research Program, led by the National Institutes of Health, is building one of the largest biomedical data resources of its kind. The *All of Us* Research Hub stores health data from a diverse group of participants from across the United States.

Registered researchers can access *All of Us* data and tools to conduct studies to help improve our understanding of human health.

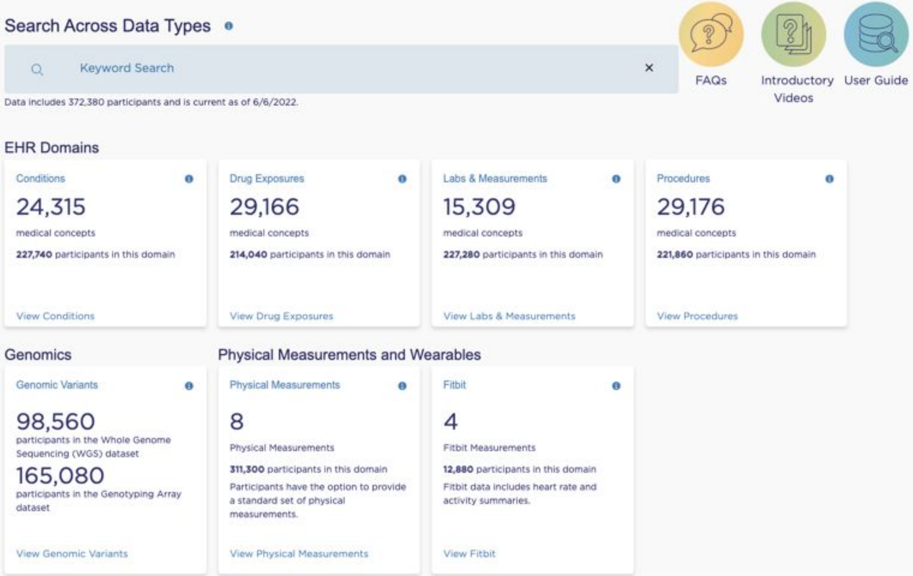
Data Snapshots

Data Snapshots showcase the scale and diversity of the *All of Us* Research Program participant cohort. The snapshots provide participant demographics, geographic distribution, and more. We update the snapshots daily.

606,000+ Participants

358,000+ Electronic Health Records

436,000+ Biosamples Received



Search Across Data Types

Data includes 372,380 participants and is current as of 6/6/2022.

EHR Domains

Conditions
24,315
medical concepts
227,740 participants in this domain
View Conditions

Drug Exposures
29,166
medical concepts
214,040 participants in this domain
View Drug Exposures

Labs & Measurements
15,309
medical concepts
227,280 participants in this domain
View Labs & Measurements

Procedures
29,176
medical concepts
221,860 participants in this domain
View Procedures

Genomics

Genomic Variants
98,560
participants in the Whole Genome Sequencing (WGS) dataset
165,080
participants in the Genotyping Array dataset
View Genomic Variants

Physical Measurements and Wearables

Physical Measurements
8
Physical Measurements
311,300 participants in this domain
Participants have the option to provide a standard set of physical measurements.
View Physical Measurements

Fitbit
4
Fitbit Measurements
12,880 participants in this domain
Fitbit data includes heart rate and activity summaries.
View Fitbit

Opportunities and Challenges in Implementing a Large-Scale Participant Digital Research Data Collection

April 11-12, 2023

Izzy Seo, MPH (NIH)
Mark Begale (Vibrent Health)
James McClain, Ph.D MPH (NIH)

All of Us RESEARCH PROGRAM

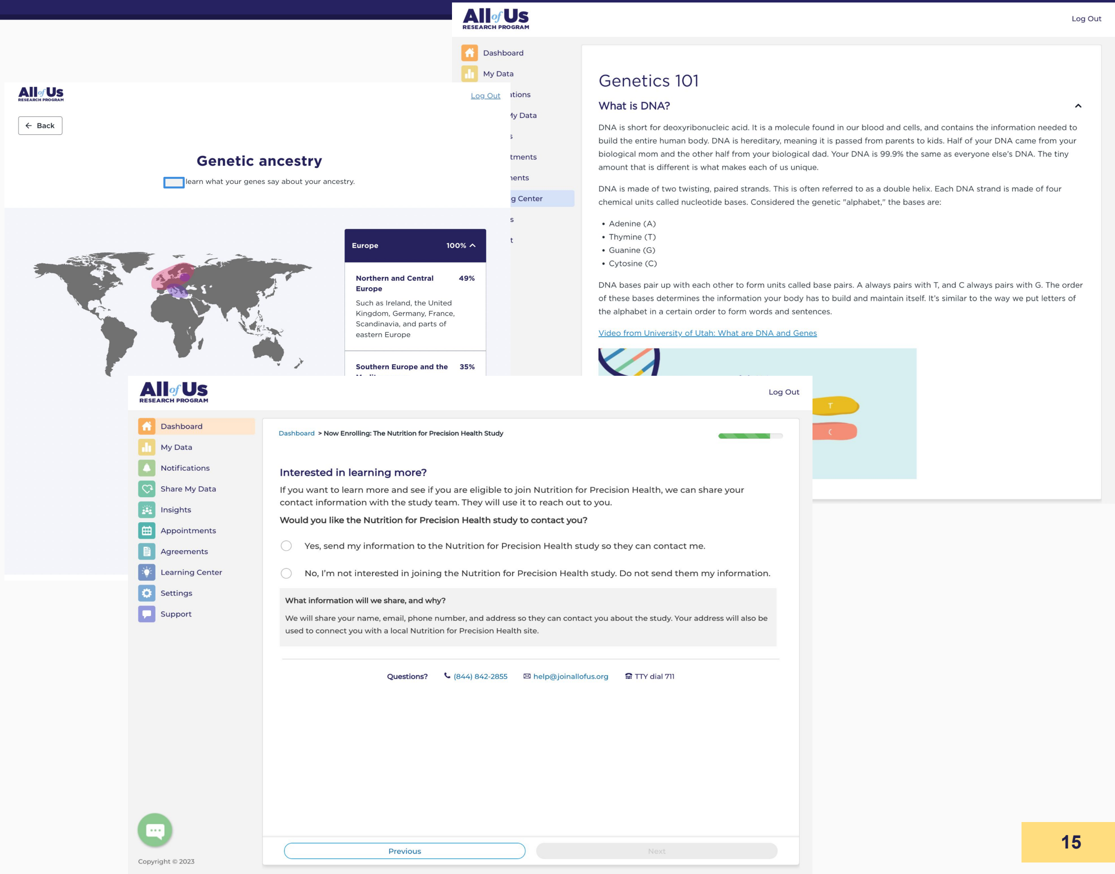
NIH National Institutes of Health

A Quick Primer:
The All of Us Participant Portal

Participants In All Of Us Engage with the All of Us Participant Portal (iOS, Android, Web)

The Portal enables participants to:

- Enroll
- Consent
- Complete surveys and other data collection modules
- Share EHR data
- Share wearable data (e.g. Apple Health / Fitbit)
- Schedule visits with partner biosample collection institutions
- Receive return of information (genomics, survey data, etc.)
- Receive tailored engagement messaging
- Ask for help
- Receive educational content
- More…



Genetics 101

What is DNA?

DNA is short for deoxyribonucleic acid. It is a molecule found in our blood and cells, and contains the information needed to build the entire human body. DNA is hereditary, meaning it is passed from parents to kids. Half of your DNA came from your biological mom and the other half from your biological dad. Your DNA is 99.9% the same as everyone else's DNA. The tiny amount that is different is what makes each of us unique.

DNA is made of two twisting, paired strands. This is often referred to as a double helix. Each DNA strand is made of four chemical units called nucleotide bases. Considered the genetic "alphabet," the bases are:

- Adenine (A)
- Thymine (T)
- Guanine (G)
- Cytosine (C)

DNA bases pair up with each other to form units called base pairs. A always pairs with T, and C always pairs with G. The order of these bases determines the information your body has to build and maintain itself. It's similar to the way we put letters of the alphabet in a certain order to form words and sentences.

Video from University of Utah: What are DNA and Genes

Genetic ancestry

learn what your genes say about your ancestry.

Europe	100%
Northern and Central Europe Such as Ireland, the United Kingdom, Germany, France, Scandinavia, and parts of eastern Europe	49%
Southern Europe and the	35%

Dashboard > Now Enrolling: The Nutrition for Precision Health Study

Interested in learning more?

If you want to learn more and see if you are eligible to join Nutrition for Precision Health, we can share your contact information with the study team. They will use it to reach out to you.

Would you like the Nutrition for Precision Health study to contact you?

○ Yes, send my information to the Nutrition for Precision Health study so they can contact me.

○ No, I'm not interested in joining the Nutrition for Precision Health study. Do not send them my information.

What information will we share, and why?

We will share your name, email, phone number, and address so they can contact you about the study. Your address will also be used to connect you with a local Nutrition for Precision Health site.

Questions? (844) 842-2855 help@joinallofus.org TTY dial 711

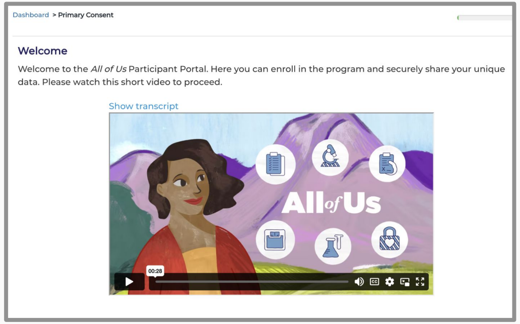
Enabling National Enrollment and Engagement

Opportunity: Enabling Broad Access for Participation

AoU supports nationwide digital consent processes accessible via web & mobile apps

- **Program platforms enable participant consent processes and workflows**
    - Available in English and Spanish language
    - Form variations account for federal and state requirements
- **Electronic consent delivery and signature**
    - Multimedia experience
        - Short text summaries
        - Video content
        - Digital long-form consent documents

Example: *All of Us* Participant Electronic Consent Process Samples

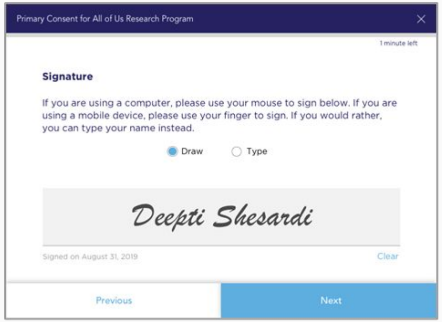


Dashboard > Primary Consent

Welcome

Welcome to the *All of Us* Participant Portal. Here you can enroll in the program and securely share your unique data. Please watch this short video to proceed.

Show transcript



Primary Consent for All of Us Research Program

1 minute left

Signature

If you are using a computer, please use your mouse to sign below. If you are using a mobile device, please use your finger to sign. If you would rather, you can type your name instead.

● Draw ○ Type

Deepti Shesardi

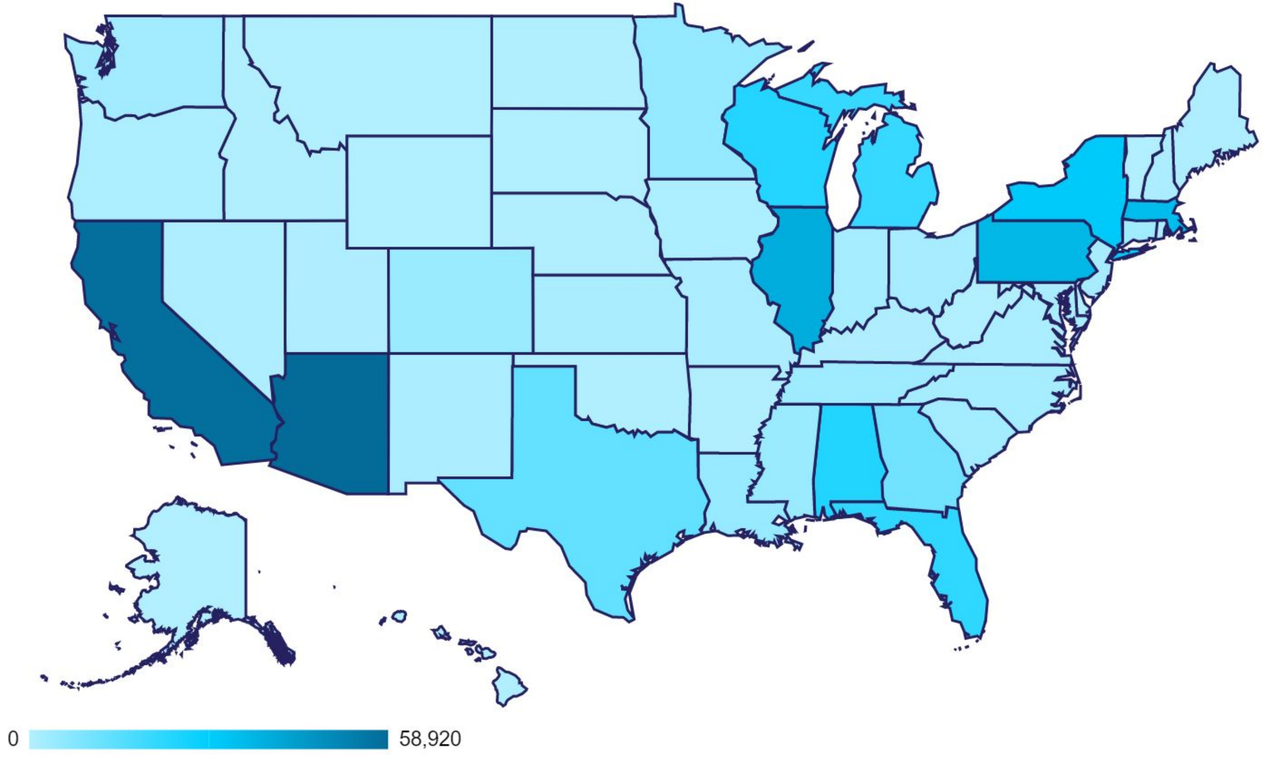
Signed on August 31, 2019

Enrollment Method	Primary Consent	Authorization to Share Electronic Health Records	Genetic Return of Results Consent	Sub-Study Consents	Ancillary Studies Consents	Special Purpose Consents
Standard	• Program Consent • Primary Consent - Cohort 1 Update	• Four versions (variation in content authorization process & expiry period)	• Single consent to perform genetic testing	Current Ex.: • WEAR Study • Exploring the Mind	Anticipated support for future ancillary studies conducted on *All of Us* Platforms	California Experimental Subject's Bill of Rights
U.S. Veterans Affairs (VA)	• VA Program Consent • VA Primary Consent - Cohort 1 Update	• VA specific version				

Challenge: Establishing Nationwide Program Awareness & Participant Support

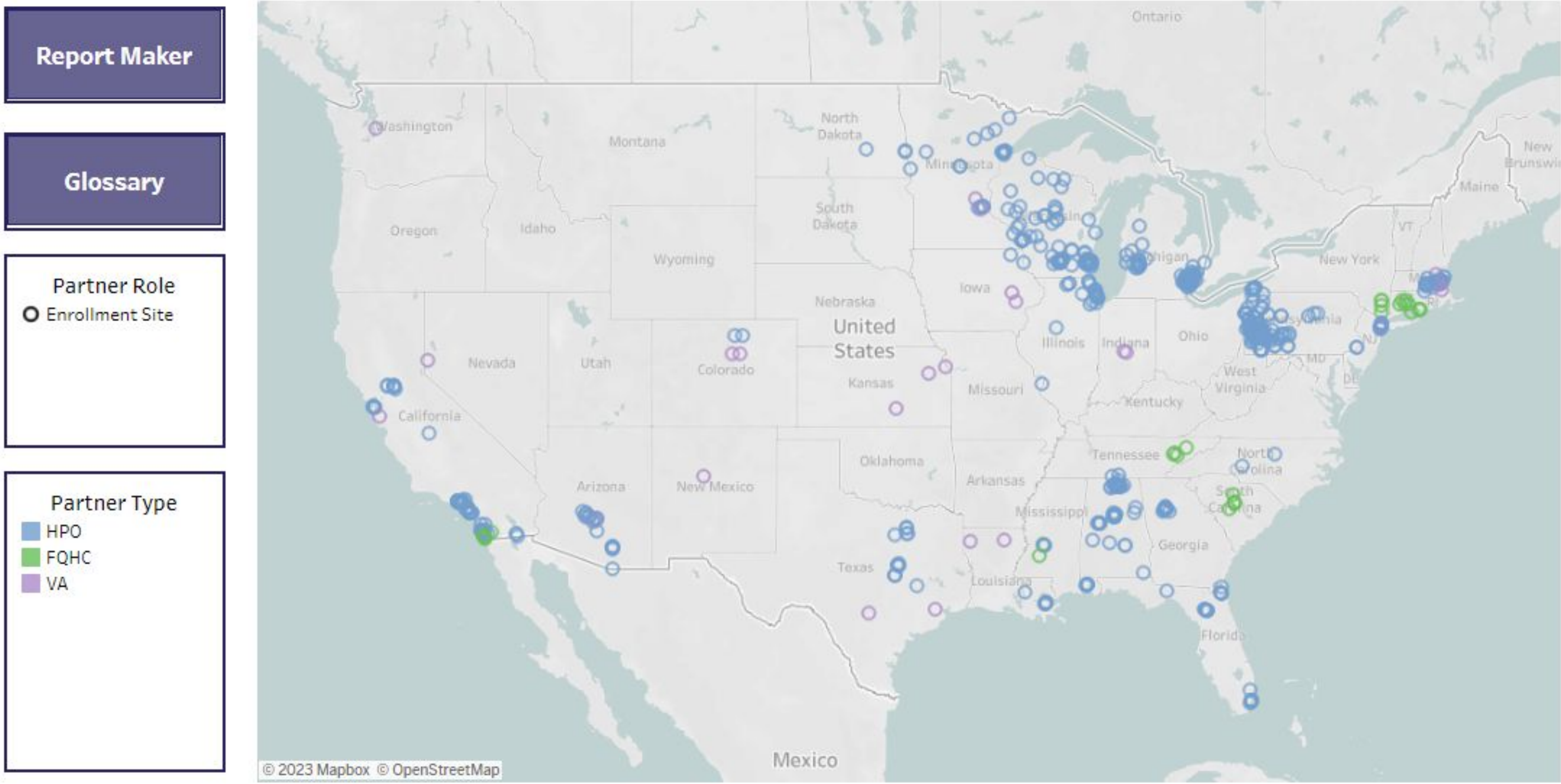
AoU regional partner networks & direct participant engagement are key to enrollment

All of Us supports nationwide enrollment in all 50 states through a combination of virtual recruitment, regional enrollment centers, and interactive mobile exhibits. Participants can currently enroll and participate in English or Spanish.



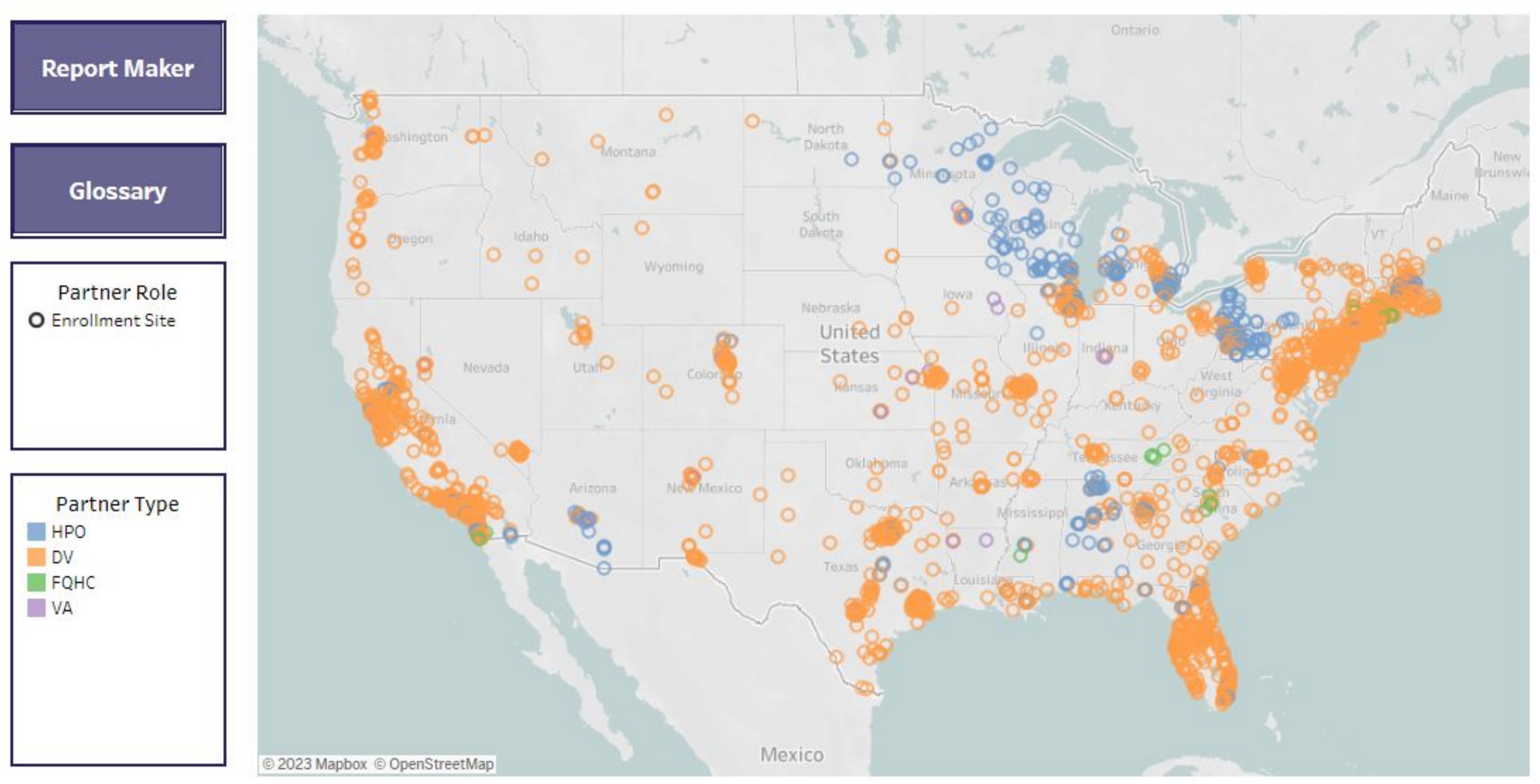
Challenge: Geographic Limits to Current Full Service Enrollment Locations

AoU enrollment locations are geographically concentrated around partner networks



Opportunity: Expanding Biosample Collection Options Through Clinical Labs

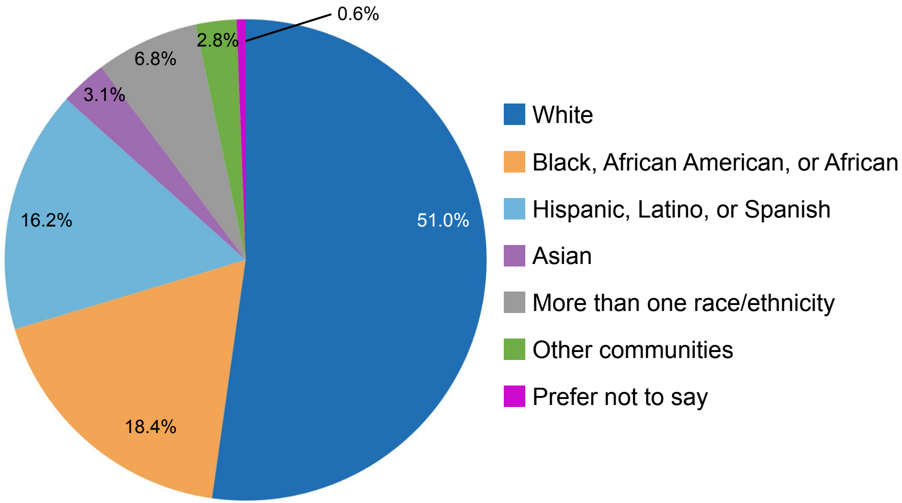
AoU piloted and is preparing to expand the use of Quest Labs for sample collections



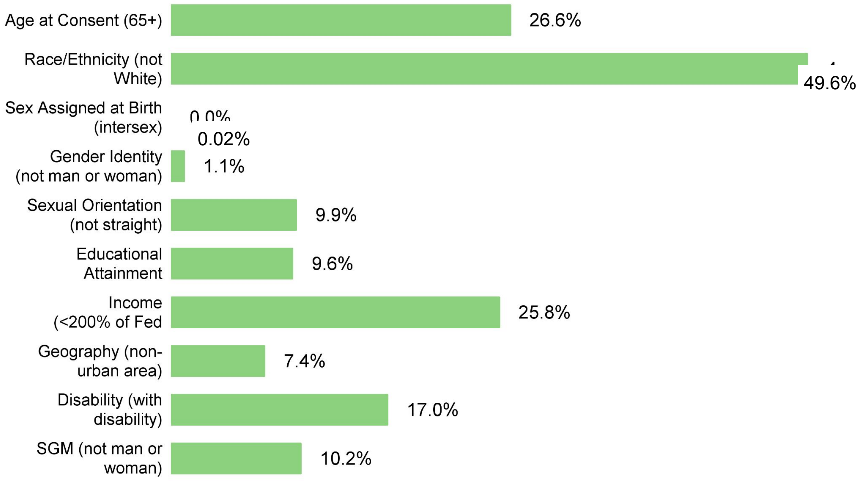
Opportunity: Diversity and Representation of Participant Enrollment

Investment in community engagement & trust-based relationships critical to success

Race & Ethnicity of Participants



UBR Category



About 80% of *All of Us* participants identify with a group that is underrepresented in biomedical research (UBR)

Numbers current as of March 21, 2023

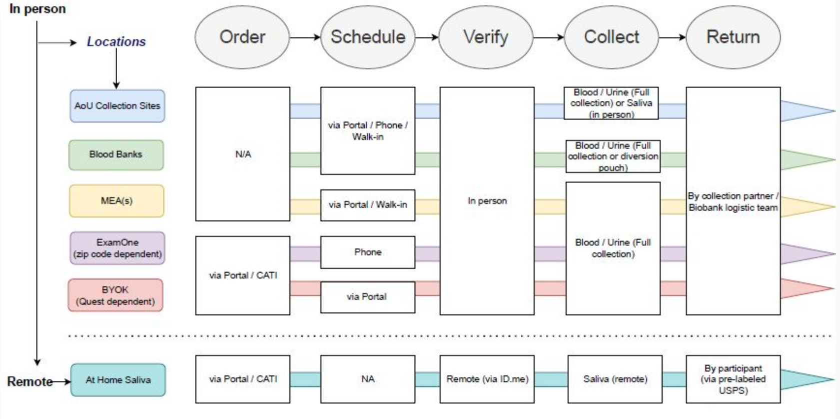
Enabling Participant Choice & Protocol Access

Opportunity: Move Beyond the Clinic & Meet Participants Where They Are

AoU enables biosample collection through mailed saliva kits, events, mobile teams & home visits

Home Saliva Collection	Total	%
Invited / Eligible	107,161	
Kit Ordered	53,780	50.2% (of eligible)
Received @ Biobank	39,368	73.2% (of ordered)

- **Over 39k biosamples remotely collected in home & returned by participants.**
  - 73.2% of kits ordered were returned.
- **Kits ordered and returned from 50 states + Puerto Rico and Guam**



	Locations	Order	Schedule	Verify	Collect	Return
In person	AoU Collection Sites	N/A	via Portal / Phone / Walk-in	In person	Blood / Urine (Full collection) or Saliva (in person)	By collection partner / Biobank logistic team
	Blood Banks				Blood / Urine (Full collection or diversion pouch)	
	MEA(s)		via Portal / Walk-in		Blood / Urine (Full collection)	
	ExamOne (zip code dependent)	via Portal / CATI	Phone			
	BYOK (Quest dependent)		via Portal			
Remote	At Home Saliva	via Portal / CATI	NA	Remote (via ID.me)	Saliva (remote)	By participant (via pre-labeled USPS)

All of Us Journey and Events

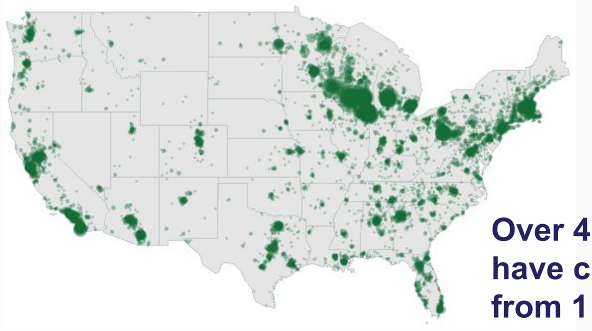
Our traveling exhibits and partner events bring the program to local communities.

Data last updated 3/28/2023

Opportunity: Expanding EHR Sharing Through Participant Mediated Exchange

Direct EHR sharing expands both participant opportunities and program reach

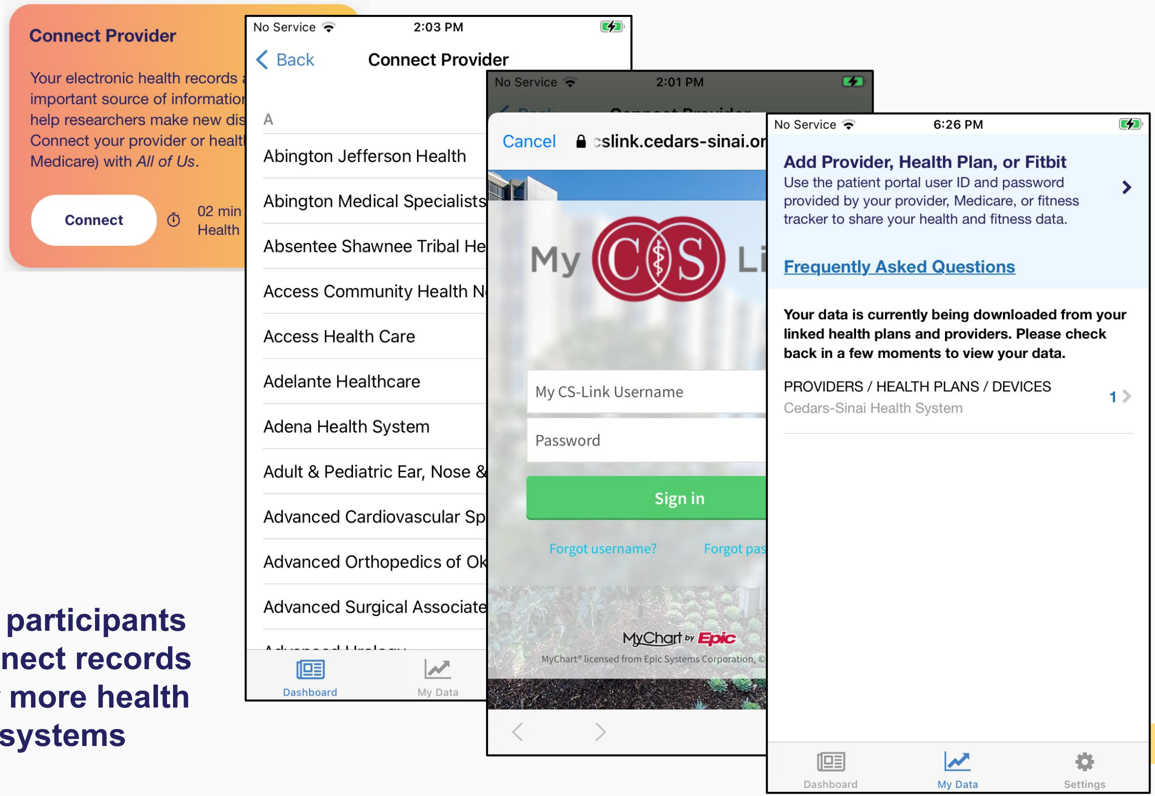
- ***All of Us*** **protocol supports data authorization and sharing from select external data sources**
- **Participants can authorize access and share data from**
    - **EHR / Patient Portals**
        - Connect to health provider organization nationally



*HI, AK, PR, GU connections not displayed

Over 40k participants have connect records from 1 or more health provider systems

Example: *All of Us* Participant EHR Data Authorization and Sharing



Connect Provider

Your electronic health records a important source of informatio help researchers make new dis Connect your provider or healt Medicare) with *All of Us*.

Add Provider, Health Plan, or Fitbit

Use the patient portal user ID and password provided by your provider, Medicare, or fitness tracker to share your health and fitness data.

Your data is currently being downloaded from your linked health plans and providers. Please check back in a few moments to view your data.

PROVIDERS / HEALTH PLANS / DEVICES 1

Cedars-Sinai Health System

Opportunity: Exposure Assessment Through Wearable Devices & Mobile Apps

Participant data sharing & device distribution enable collection of objective longitudinal data

- *All of Us* continues to accrue DHT data through Bring Your Own Device (BYOD) and WEAR (Fitbit device distribution) protocol connections.
- 20k unit participant WEAR protocol expansion will improve representation and drive participant Fitbit account connection above 50k in 2023.

Data Source	PTSC Portal Sharing	CE Portal Sharing	Total Participants
Fitbit	20,372	12,144	32,516
Apple Health	12,700	8,141	20,841
Google Fit	0	3,456	3,456
Total AoU DHT Connections	33,072	23,741	56,813

Data last updated 3/28/2023

Driving Recruitment and Engagement with Systematic Tooling

“No Login” Activity Completion

Problem Statements

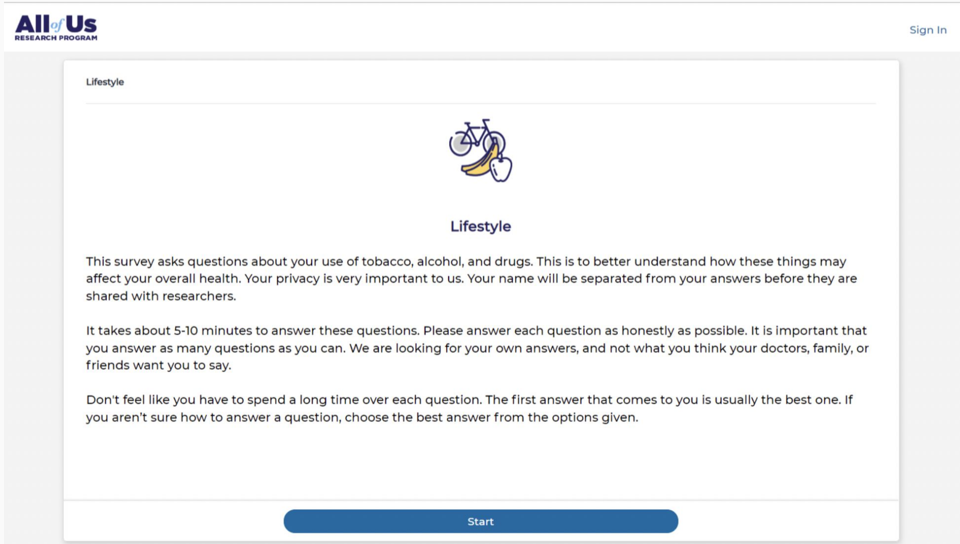
- Participants within All of Us and larger digital health research studies are often required to log-in to the program experience to complete tasks (like surveys).
- This places a burden on participants to login which increases the risk of non-completion
- This is further exacerbated in situations where participants cannot easily recall their account credentials

Givens

- All participants are required to have an email address or phone number to register
- Participant PII and result data must be secured behind the login process to support program privacy and security

Solution

- Participants are sent unique secure links to complete surveys and tasks outside of a logged in portal experience
- No tasks that require PII utilize this method



Lifestyle

This survey asks questions about your use of tobacco, alcohol, and drugs. This is to better understand how these things may affect your overall health. Your privacy is very important to us. Your name will be separated from your answers before they are shared with researchers.

It takes about 5-10 minutes to answer these questions. Please answer each question as honestly as possible. It is important that you answer as many questions as you can. We are looking for your own answers, and not what you think your doctors, family, or friends want you to say.

Don't feel like you have to spend a long time over each question. The first answer that comes to you is usually the best one. If you aren't sure how to answer a question, choose the best answer from the options given.

“No login” activity completion resulted in a ~20% improvement in task completion where implemented

Standardizing Participant Referral Processes (Refer a Friend)

Problem Statements

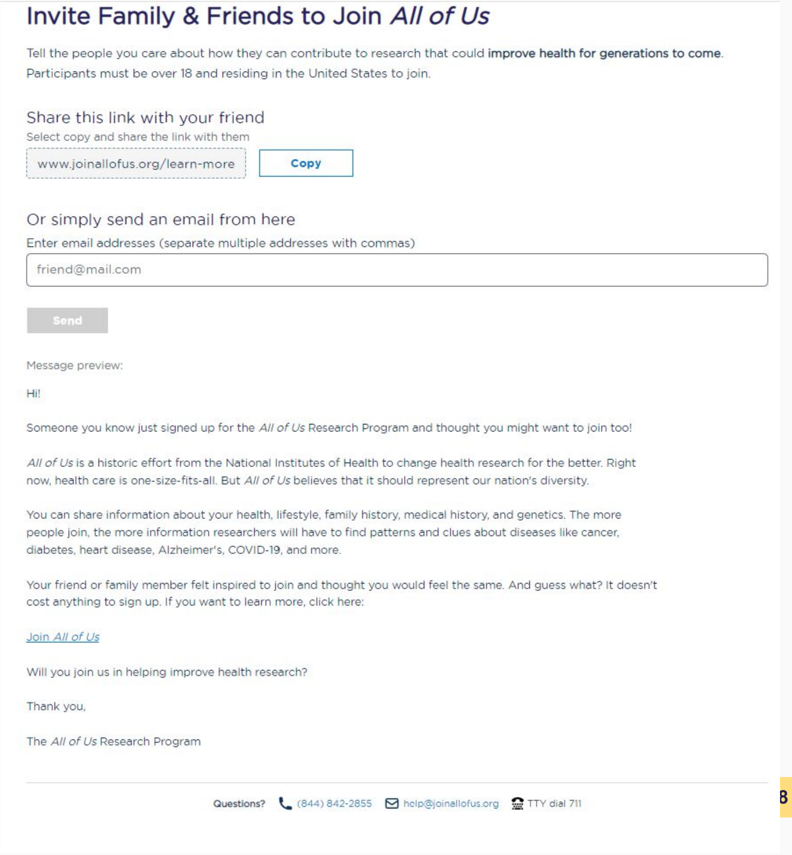
- Additional participant enrollment opportunities are needed to achieve the overall program objective of recruiting 1M+ participants
- Registered and unregistered participants should be able to inform prospective about the program
- There was no standardized, trackable mechanism for doing this

Givens

- Participant PII and result data must be secured behind the login process to support program privacy and security

Solution

- Provide a generalizable solution that is usable within and outside the participant experience for enabling participant referral



Invite Family & Friends to Join *All of Us*

Tell the people you care about how they can contribute to research that could **improve health for generations to come**. Participants must be over 18 and residing in the United States to join.

Share this link with your friend

Select copy and share the link with them

Or simply send an email from here

Enter email addresses (separate multiple addresses with commas)

Message preview:

Hi!

Someone you know just signed up for the *All of Us* Research Program and thought you might want to join too!

All of Us is a historic effort from the National Institutes of Health to change health research for the better. Right now, health care is one-size-fits-all. But *All of Us* believes that it should represent our nation's diversity.

You can share information about your health, lifestyle, family history, medical history, and genetics. The more people join, the more information researchers will have to find patterns and clues about diseases like cancer, diabetes, heart disease, Alzheimer's, COVID-19, and more.

Your friend or family member felt inspired to join and thought you would feel the same. And guess what? It doesn't cost anything to sign up. If you want to learn more, click here:

Join *All of Us*

Will you join us in helping improve health research?

Thank you,

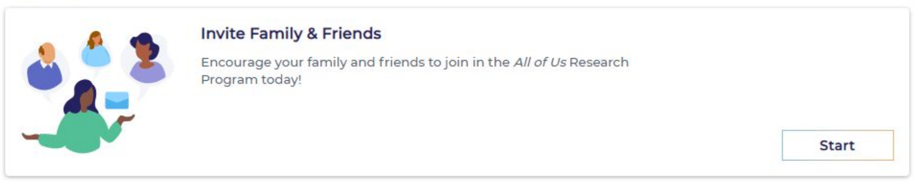
The *All of Us* Research Program

Questions? ☎ (844) 842-2855 ✉ help@joinallofus.org ☎ TTY dial 711

Standardizing Participant Referral Processes (Refer a Friend) Stats

Launched on September 23, 2022, preliminary data from Refer a Friend demonstrated:

- 6,627 Participant Portal-based referrals transmitted via the program email platform
- 17.8% of email referral recipients “click” through the the joinallofus.org “Learn More” page, of these:
  - 22.7% complete registration
  - 26.0% complete consent



Invite Family & Friends

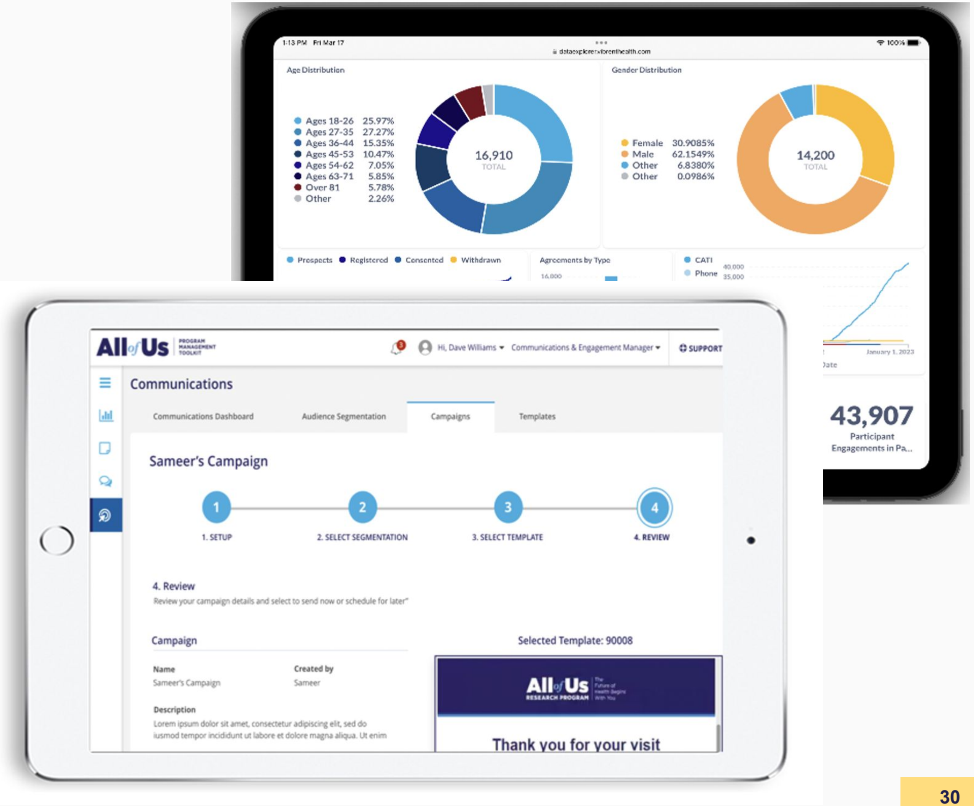
Encourage your family and friends to join in the *All of Us* Research Program today!

Data last updated 3/28/2023

Tools for Staff to Engage with Participants (Program Management Toolkit)

Program Management Toolkit enables more than 800+ study staff to:

- Create and manage study sites
- Engage and communicate with participants using comprehensive rules via email and SMS (1:1 and campaigns)
- Complete data collection activities on behalf of participants
- Track task completion for participants
- Manage staff goals
- Build recruitment microsites
- More..



Sameer's Campaign

4. Review

Review your campaign details and select to send now or schedule for later"

Campaign

Name	Created by
Sameer's Campaign	Sameer

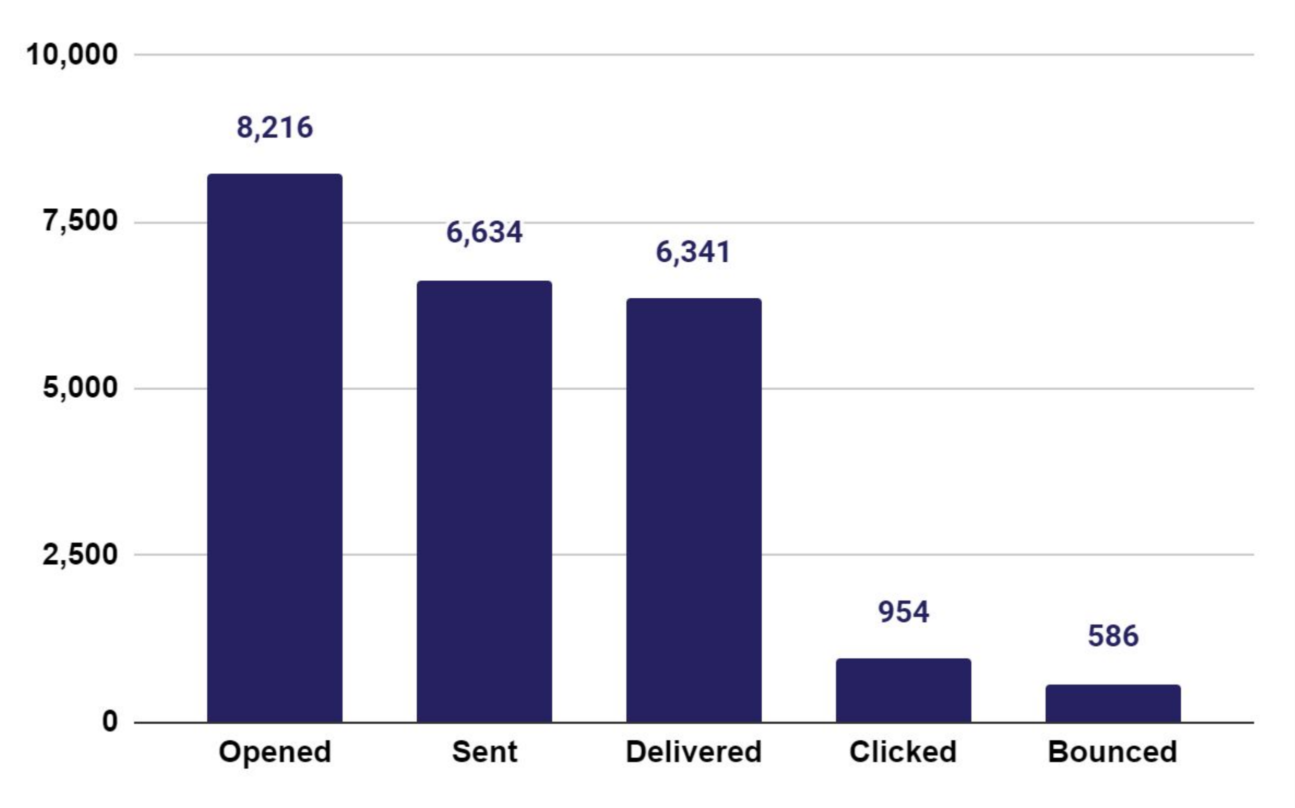
Description

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do iusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim

Selected Template: 90008

Thank you for your visit

Participant Referral and Iterable Email Campaign Analytics



Data last updated 3/28/23

Developing Standardized, Open-Access Reusable Tools for Expand

Standardizing Open Source Data Collection and Return of Information Development

Integration partners, app developers, and survey authors…

- **Data Integration Partners** (e.g. FitBit, Blue Button)
- **Experience Developers** (e.g. mini-apps, RDoC)
- **Survey Authorship Committees** (e.g. AoURP surveys)

…are provided with the standardized systems and support tools necessary to…

REDCap HL7 FHIR

apigee All of Us Contribute

- **Authorization / Authentication**
- **Data Integration SDKs**
- **APIs**
- **Experience SDKs**

…provide engaging experiences, data types, and surveys to participants…

- **New Data Types / Insights**
- **Portal / App Experiences**
- **Portal Surveys**

…whose data empowers researchers.

All of Us Research Hub

Standardizing Open Source <u>Data Collection</u> Modules

Value Propositions

- Individual cost to develop engaging return of information can be difficult to standardize using off the shelf tools
- Enabling consortium partners to build return of information modules in a standardized way

Approach

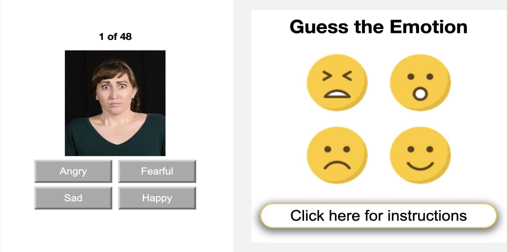
- Provide a generally available standard and toolkit for surfacing modular participant experiences that:
  - Use standard APIs / SDKs
  - Can be shared to other research projects using similar measure or data sets

Example

Working with the Many Brains team (an open science cognitive task development non-profit), All of Us delivered completely independently developed cognitive testing tools to a pilot cohort in December 2022, with expansion to all participants planned for Q2, 2023

All of Us Application Wrapper / SDKs

Many Brains TestMyBrain Exploring the Mind Module



↓

HL7® FHIR®

↓

DRC Research Workbench

Standardizing Open Source Return of Information Modules

Value Propositions

- Providing participants with high quality, meaningful/useful return of information is critical to ensure ongoing engagement
- Individual cost to develop engaging return of information can be difficult to standardize using off the shelf tools
- Enable consortium partners to build return of information modules in a standardized way

Approach

Provide a generally available standard and toolkit for surfacing modular participant experiences that:

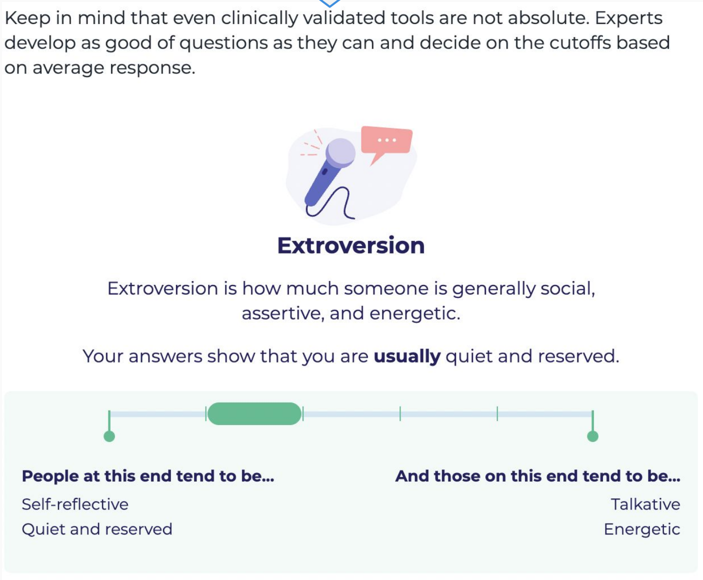
- Use standard APIs / SDKs
- Use FHIR as a data exchange format
- Can be shared to other research teams building return of information for participants

Example

Provide reusable open-source modules for standardized personality assessments (planned release in June!)

Personality Assessment Module

HL7® FHIR®



Keep in mind that even clinically validated tools are not absolute. Experts develop as good of questions as they can and decide on the cutoffs based on average response.

Extroversion

Extroversion is how much someone is generally social, assertive, and energetic.

Your answers show that you are **usually** quiet and reserved.

People at this end tend to be...
Self-reflective
Quiet and reserved

And those on this end tend to be...
Talkative
Energetic

Working with NIH *AoU* Researcher Workbench
A Guided Tour

April 11-12, 2023

Gage Rion, ME (Vanderbilt University Medical Center)
Danielle Wilfong, M.S., Ed.M. (Vanderbilt University Medical Center)
Samantha Lee Stewart, M.S.(Vanderbilt University Medical Center)

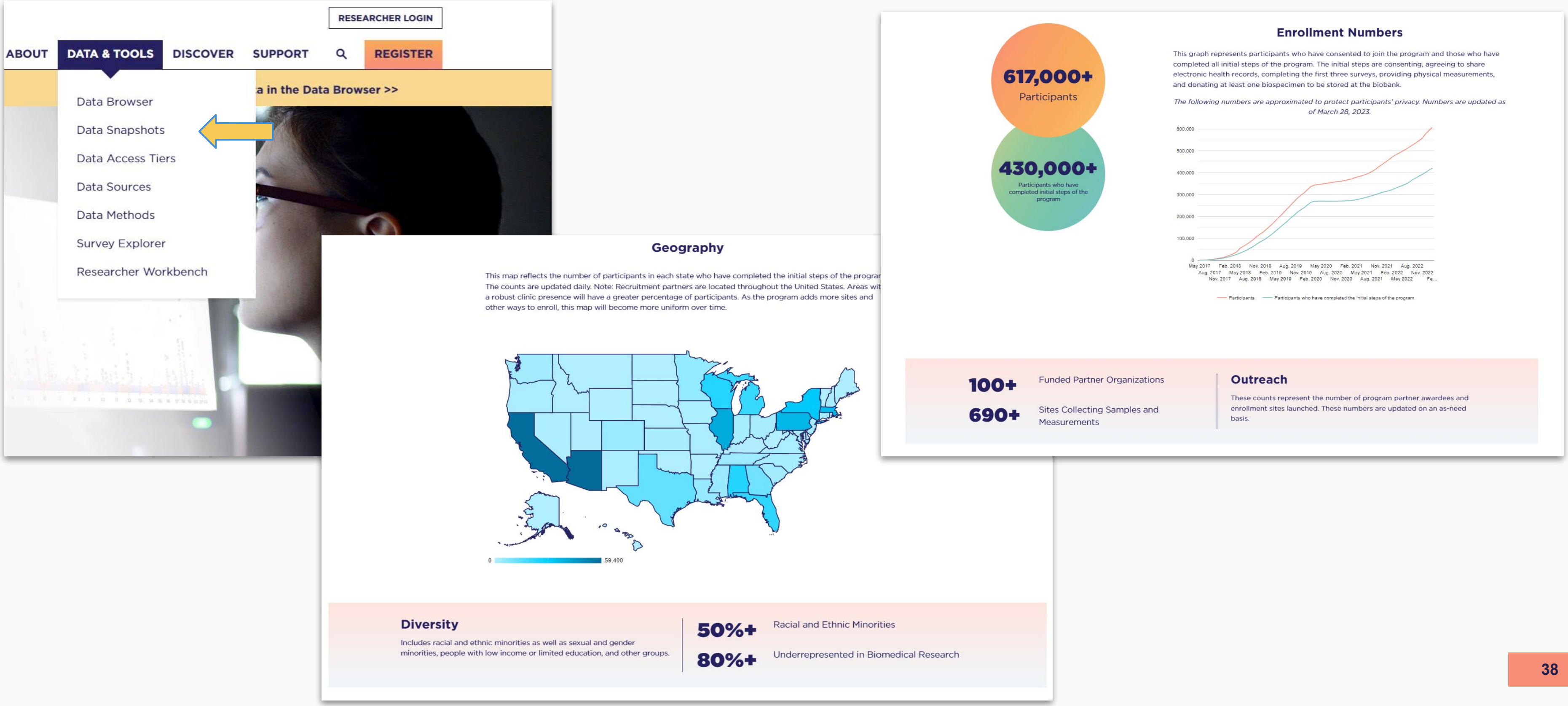
NIH National Institutes of Health

Researchallofus.org is home to *All of Us* data, research tools, & projects.

“The Public Website”

“The Workbench”

Data Snapshot provide an overview of participant data.



617,000+ Participants

430,000+ Participants who have completed initial steps of the program

Enrollment Numbers

This graph represents participants who have consented to join the program and those who have completed all initial steps of the program. The initial steps are consenting, agreeing to share electronic health records, completing the first three surveys, providing physical measurements, and donating at least one biospecimen to be stored at the biobank.

The following numbers are approximated to protect participants' privacy. Numbers are updated as of March 28, 2023.

100+ Funded Partner Organizations

690+ Sites Collecting Samples and Measurements

Outreach

These counts represent the number of program partner awardees and enrollment sites launched. These numbers are updated on an as-need basis.

Geography

This map reflects the number of participants in each state who have completed the initial steps of the progra
The counts are updated daily. Note: Recruitment partners are located throughout the United States. Areas wit
a robust clinic presence will have a greater percentage of participants. As the program adds more sites and other ways to enroll, this map will become more uniform over time.

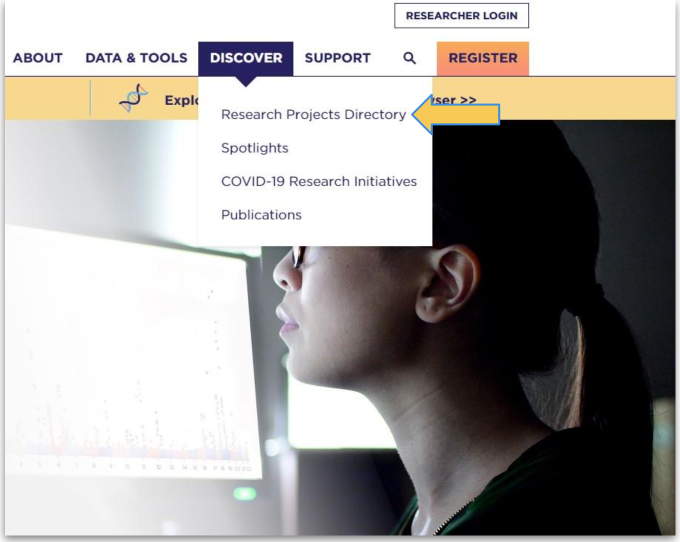
Diversity

Includes racial and ethnic minorities as well as sexual and gender minorities, people with low income or limited education, and other groups.

50%+ Racial and Ethnic Minorities

80%+ Underrepresented in Biomedical Research

Use the Project Directory to view all active projects.



Research Projects Directory

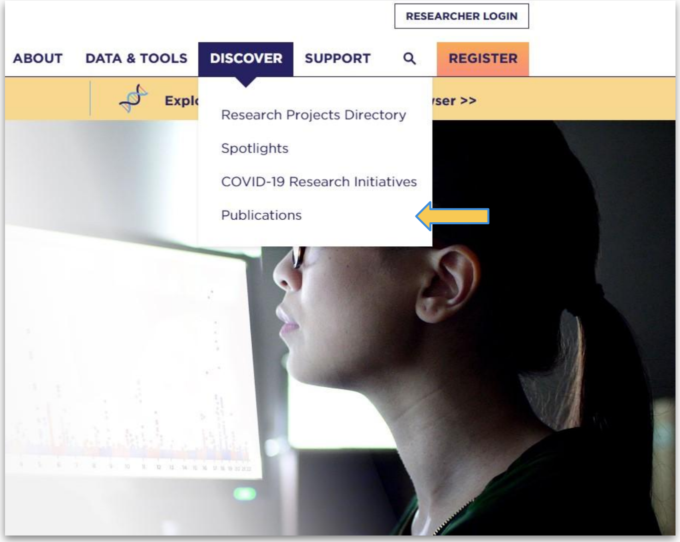
4,297 ACTIVE PROJECTS

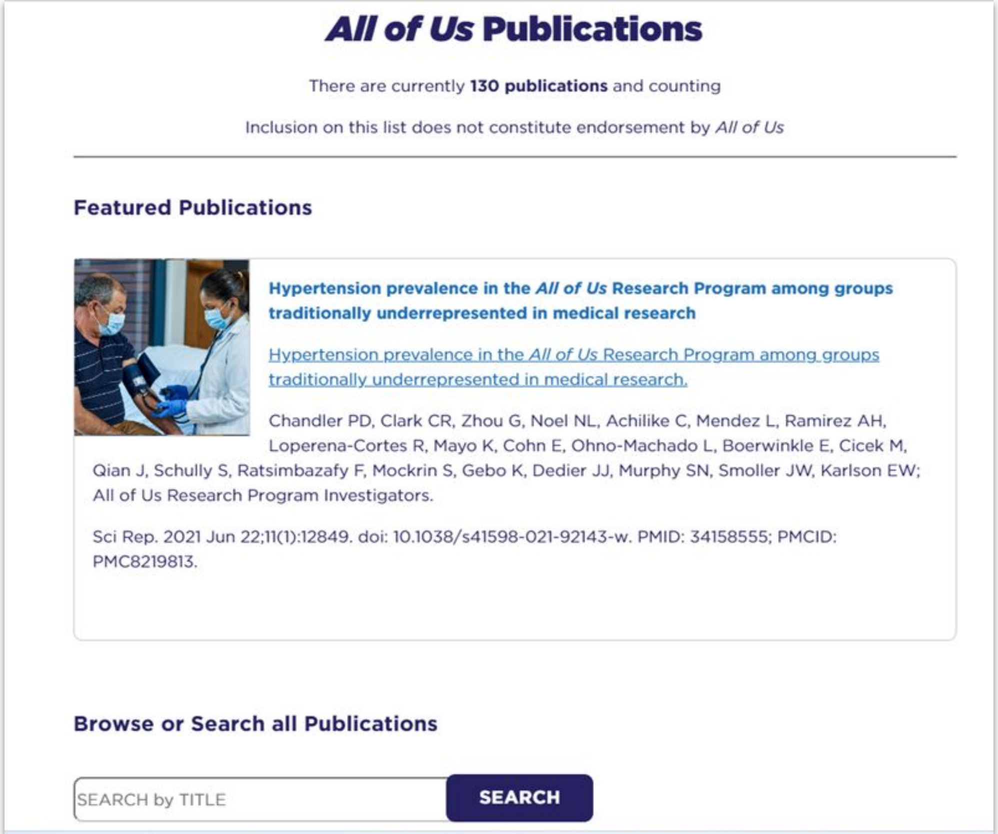
This information was updated 3/29/2023

The Research Projects Directory includes information about all projects that currently exist in the Researcher Workbench to help provide transparency about how the Workbench is being used. Each project specifies whether Registered Tier or Controlled Tier data are used.

Note: Researcher Workbench users provide information about their research projects independently. Views expressed in the Research Projects Directory belong to the relevant users and do not necessarily represent those of the All of Us Research Program. Information in the Research Projects Directory is also cross-posted on AllofUs.nih.gov in compliance with the 21st Century Cures Act.

Read researchers' peer-reviewed *All of Us* publications





All of Us Publications

There are currently **130 publications** and counting

Inclusion on this list does not constitute endorsement by *All of Us*

Featured Publications

Hypertension prevalence in the *All of Us* Research Program among groups traditionally underrepresented in medical research

Hypertension prevalence in the *All of Us* Research Program among groups traditionally underrepresented in medical research.

Chandler PD, Clark CR, Zhou G, Noel NL, Achilike C, Mendez L, Ramirez AH, Loperena-Cortes R, Mayo K, Cohn E, Ohno-Machado L, Boerwinkle E, Cicek M, Qian J, Schully S, Ratsimbazafy F, Mockrin S, Gebo K, Dedier JJ, Murphy SN, Smoller JW, Karlson EW; All of Us Research Program Investigators.

Sci Rep. 2021 Jun 22;11(1):12849. doi: 10.1038/s41598-021-92143-w. PMID: 34158555; PMCID: PMC8219813.

Browse or Search all Publications

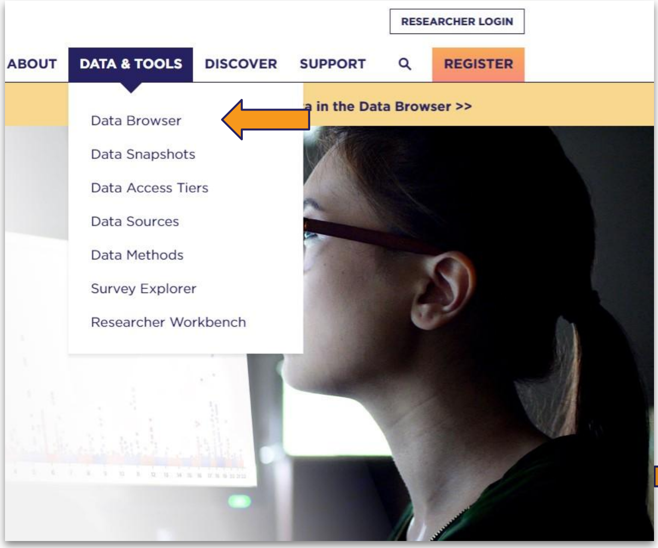
SEARCH by TITLE

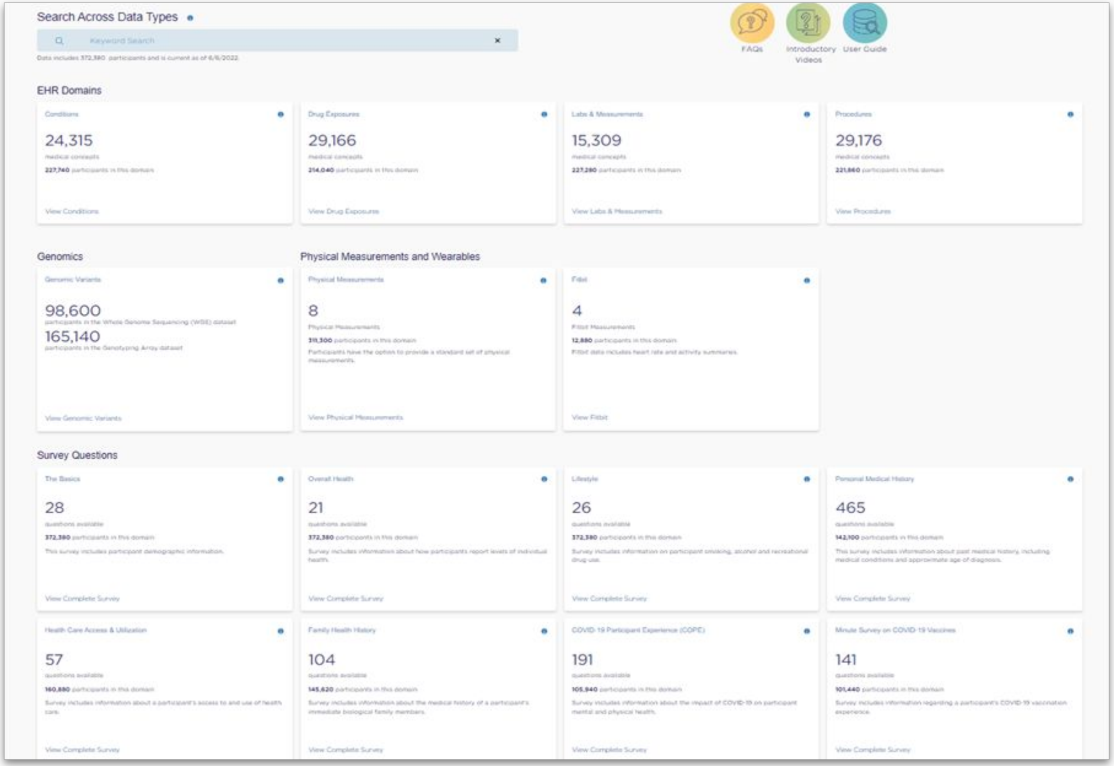
SEARCH

All of Us RESEARCH PROGRAM

Exploring the Data - The Data Browser & Survey Explorer

The Data Browser is an interactive way to view the data.





Data includes 372,380 participants and is current as of 6/6/2022

EHR Domains

Conditions
24,315
medical concepts
227,740 participants in this domain
View Conditions

Drug Exposures
29,166
medical concepts
214,040 participants in this domain
View Drug Exposures

Labs & Measurements
15,309
medical concepts
227,280 participants in this domain
View Labs & Measurements

Procedures
29,176
medical concepts
221,860 participants in this domain
View Procedures

Genomics

Genomic Variants
98,600
participants in the Whole Genome Sequencing (WGS) dataset
165,140
participants in the Genotyping Array dataset
View Genomic Variants

Physical Measurements and Wearables

Physical Measurements
8
Physical Measurements
311,300 participants in this domain
Participants have the option to provide a standard set of physical measurements.
View Physical Measurements

Fitbit
4
Fitbit Measurements
12,880 participants in this domain
Fitbit data includes heart rate and activity summaries.
View Fitbit

Survey Questions

The Basics
28
questions available
372,380 participants in this domain
This survey includes participant demographic information.
View Complete Survey

Overall Health
21
questions available
372,380 participants in this domain
Survey includes information about how participants report levels of individual health.
View Complete Survey

Lifestyle
26
questions available
372,380 participants in this domain
Survey includes information on participant smoking, alcohol and recreational drug use.
View Complete Survey

Personal Medical History
465
questions available
142,100 participants in this domain
This survey includes information about past medical history, including medical conditions and approximate age of diagnosis.
View Complete Survey

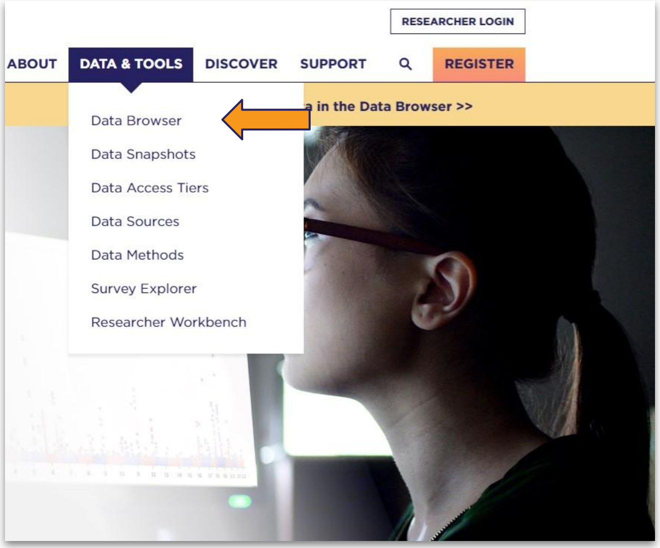
Health Care Access & Utilization
57
questions available
160,880 participants in this domain
Survey includes information about a participant's access to and use of health care.
View Complete Survey

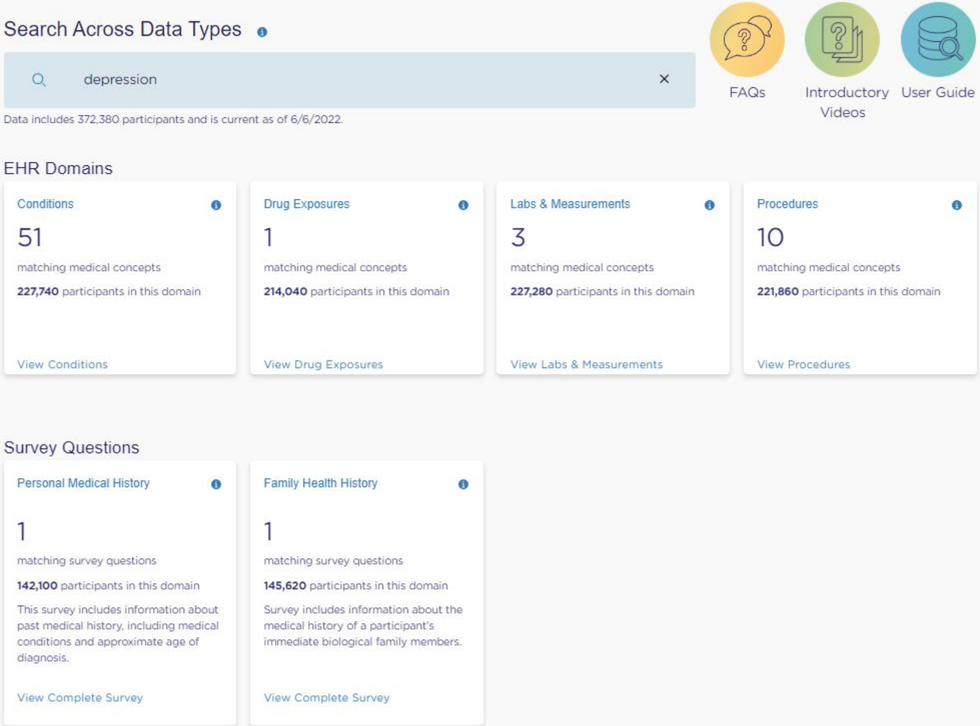
Family Health History
104
questions available
145,620 participants in this domain
Survey includes information about the medical history of a participant's immediate biological family members.
View Complete Survey

COVID-19 Participant Experience (COPE)
191
questions available
105,940 participants in this domain
Survey includes information about the impact of COVID-19 on participant mental and physical health.
View Complete Survey

Minute Survey on COVID-19 Vaccines
141
questions available
101,440 participants in this domain
Survey includes information regarding a participant's COVID-19 vaccination experience.
View Complete Survey

The Data Browser is an interactive way to view the data.





Search Across Data Types

depression

Data includes 372,380 participants and is current as of 6/6/2022.

EHR Domains

Conditions
51
matching medical concepts
227,740 participants in this domain
View Conditions

Drug Exposures
1
matching medical concepts
214,040 participants in this domain
View Drug Exposures

Labs & Measurements
3
matching medical concepts
227,280 participants in this domain
View Labs & Measurements

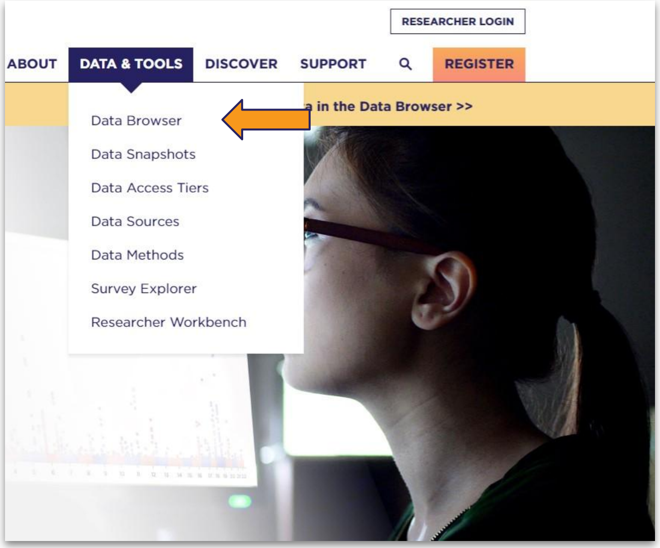
Procedures
10
matching medical concepts
221,860 participants in this domain
View Procedures

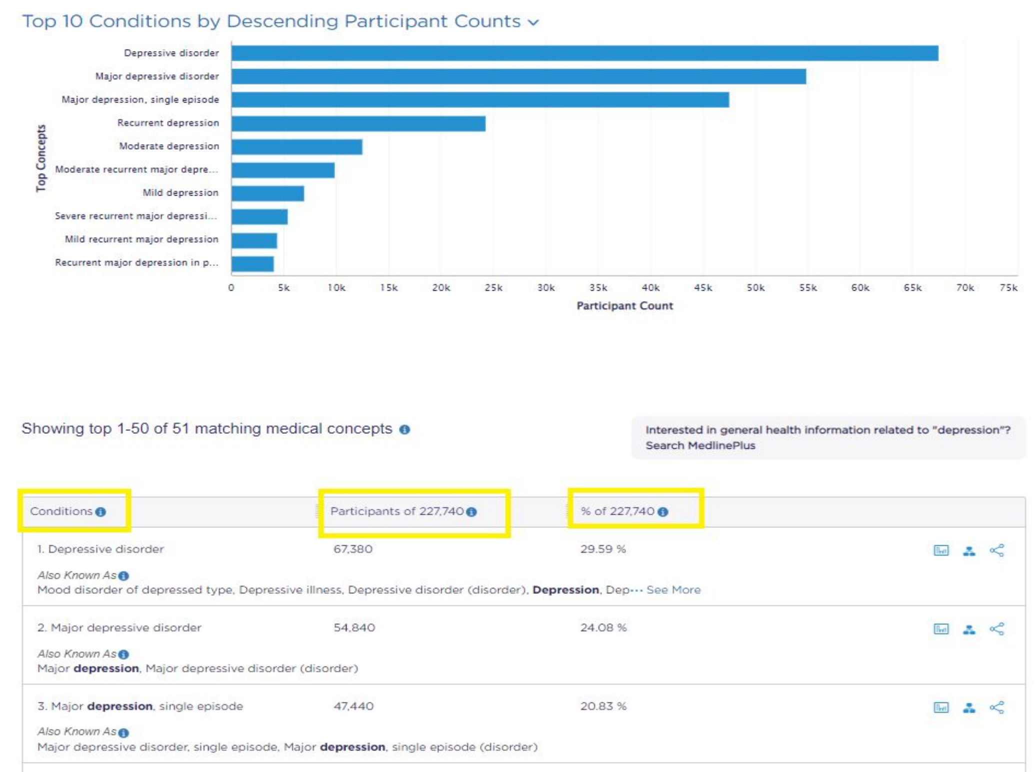
Survey Questions

Personal Medical History
1
matching survey questions
142,100 participants in this domain
This survey includes information about past medical history, including medical conditions and approximate age of diagnosis.
View Complete Survey

Family Health History
1
matching survey questions
145,620 participants in this domain
Survey includes information about the medical history of a participant's immediate biological family members.
View Complete Survey

The Data Browser is an interactive way to view the data.



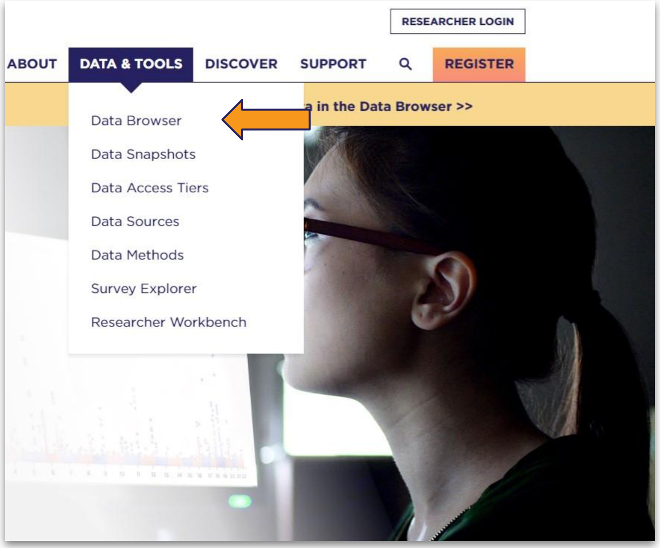


Showing top 1-50 of 51 matching medical concepts

Interested in general health information related to "depression"? Search MedlinePlus

Conditions	Participants of 227,740	% of 227,740
1. Depressive disorder *Also Known As* Mood disorder of depressed type, Depressive illness, Depressive disorder (disorder), **Depression**, Dep··· See More	67,380	29.59 %
2. Major depressive disorder *Also Known As* Major **depression**, Major depressive disorder (disorder)	54,840	24.08 %
3. Major **depression**, single episode *Also Known As* Major depressive disorder, single episode, Major **depression**, single episode (disorder)	47,440	20.83 %

The Data Browser is an interactive way to view the data.



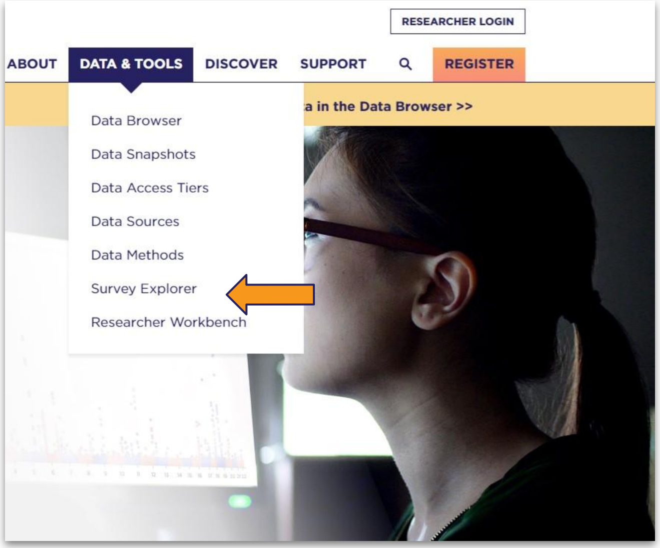
Showing top 1-50 of 51 matching medical concepts

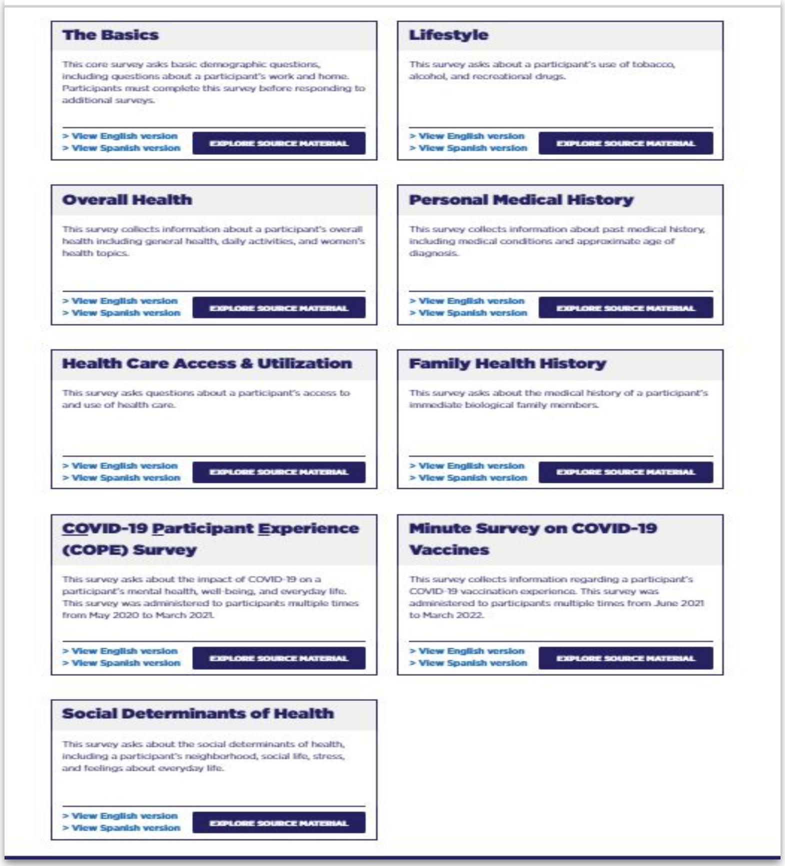
Interested in general health information related to "depression"? Search MedlinePlus

Conditions	Participants of 227,740	% of 227,740
1. Depressive disorder	67,380	29.59 %

Also Known As
Mood disorder of depressed type, Depressive illness, Depressive disorder (disorder), **Depression**, Dep··· See More

The Survey Explorer hosts all surveys available to participants.





The Basics

This core survey asks basic demographic questions, including questions about a participant's work and home. Participants must complete this survey before responding to additional surveys.

> View English version
> View Spanish version
EXPLORE SOURCE MATERIAL

Lifestyle

This survey asks about a participant's use of tobacco, alcohol, and recreational drugs.

> View English version
> View Spanish version
EXPLORE SOURCE MATERIAL

Overall Health

This survey collects information about a participant's overall health including general health, daily activities, and women's health topics.

> View English version
> View Spanish version
EXPLORE SOURCE MATERIAL

Personal Medical History

This survey collects information about past medical history, including medical conditions and approximate age of diagnosis.

> View English version
> View Spanish version
EXPLORE SOURCE MATERIAL

Health Care Access & Utilization

This survey asks questions about a participant's access to and use of health care.

> View English version
> View Spanish version
EXPLORE SOURCE MATERIAL

Family Health History

This survey asks about the medical history of a participant's immediate biological family members.

> View English version
> View Spanish version
EXPLORE SOURCE MATERIAL

COVID-19 Participant Experience (COPE) Survey

This survey asks about the impact of COVID-19 on a participant's mental health, well-being, and everyday life. This survey was administered to participants multiple times from May 2020 to March 2021.

> View English version
> View Spanish version
EXPLORE SOURCE MATERIAL

Minute Survey on COVID-19 Vaccines

This survey collects information regarding a participant's COVID-19 vaccination experience. This survey was administered to participants multiple times from June 2021 to March 2022.

> View English version
> View Spanish version
EXPLORE SOURCE MATERIAL

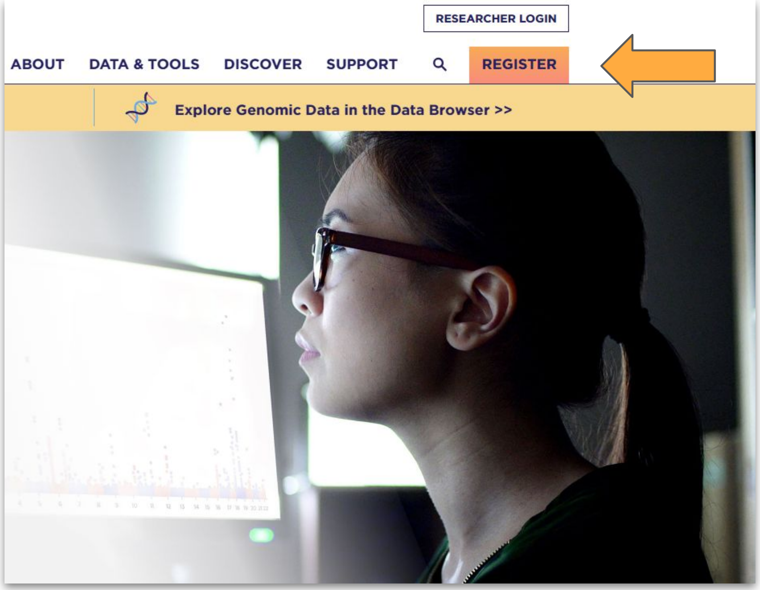
Social Determinants of Health

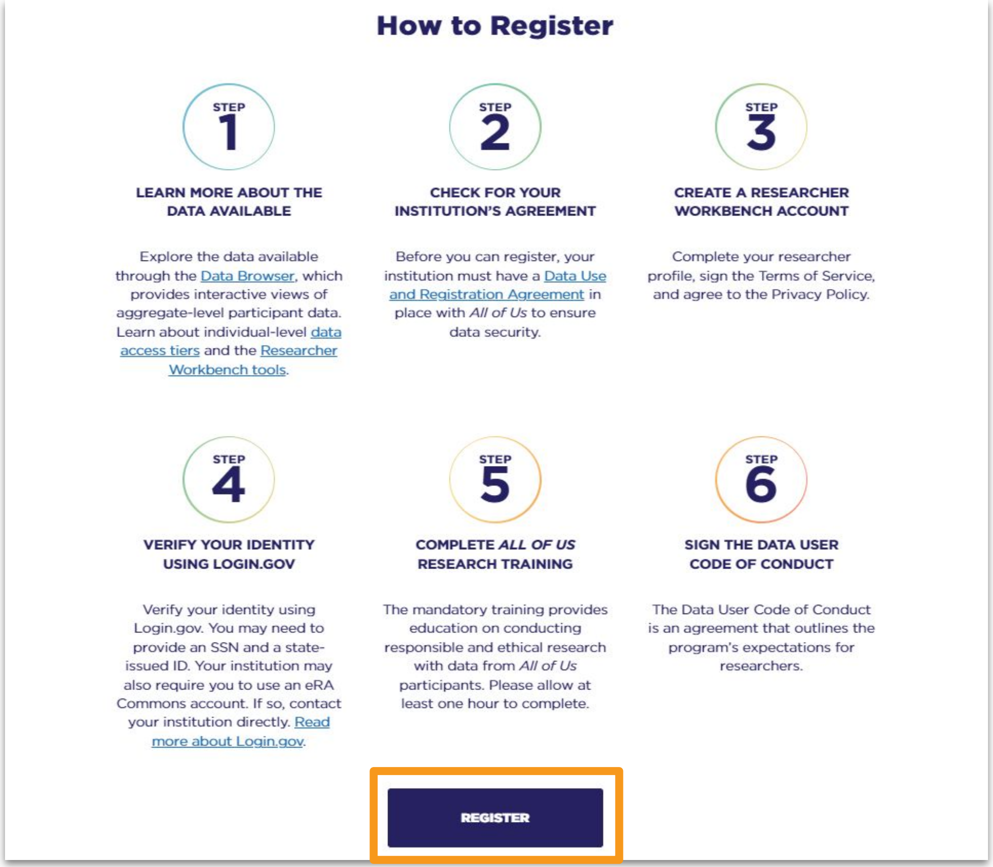
This survey asks about the social determinants of health, including a participant's neighborhood, social life, stress, and feelings about everyday life.

> View English version
> View Spanish version
EXPLORE SOURCE MATERIAL

How to Register

Registration is easy!





How to Register

STEP 1

LEARN MORE ABOUT THE DATA AVAILABLE

Explore the data available through the Data Browser, which provides interactive views of aggregate-level participant data. Learn about individual-level data access tiers and the Researcher Workbench tools.

STEP 2

CHECK FOR YOUR INSTITUTION'S AGREEMENT

Before you can register, your institution must have a Data Use and Registration Agreement in place with *All of Us* to ensure data security.

STEP 3

CREATE A RESEARCHER WORKBENCH ACCOUNT

Complete your researcher profile, sign the Terms of Service, and agree to the Privacy Policy.

STEP 4

VERIFY YOUR IDENTITY USING LOGIN.GOV

Verify your identity using Login.gov. You may need to provide an SSN and a state-issued ID. Your institution may also require you to use an eRA Commons account. If so, contact your institution directly. Read more about Login.gov.

STEP 5

COMPLETE *ALL OF US* RESEARCH TRAINING

The mandatory training provides education on conducting responsible and ethical research with data from *All of Us* participants. Please allow at least one hour to complete.

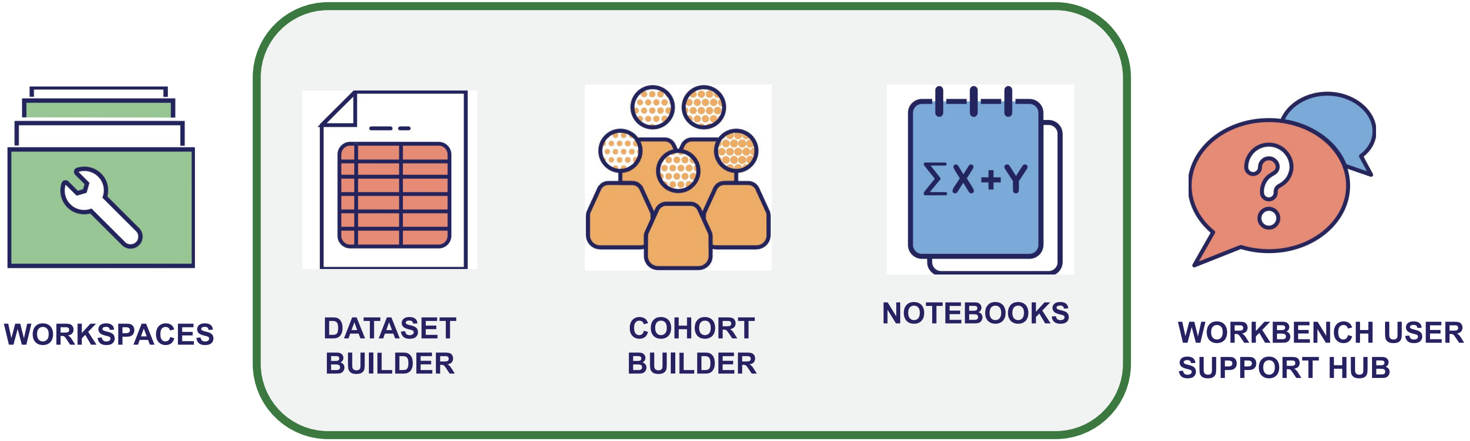
STEP 6

SIGN THE DATA USER CODE OF CONDUCT

The Data User Code of Conduct is an agreement that outlines the program's expectations for researchers.

The Researcher Workbench and how to start a research project

The Researcher Workbench = secure cloud analysis platform



WORKSPACES

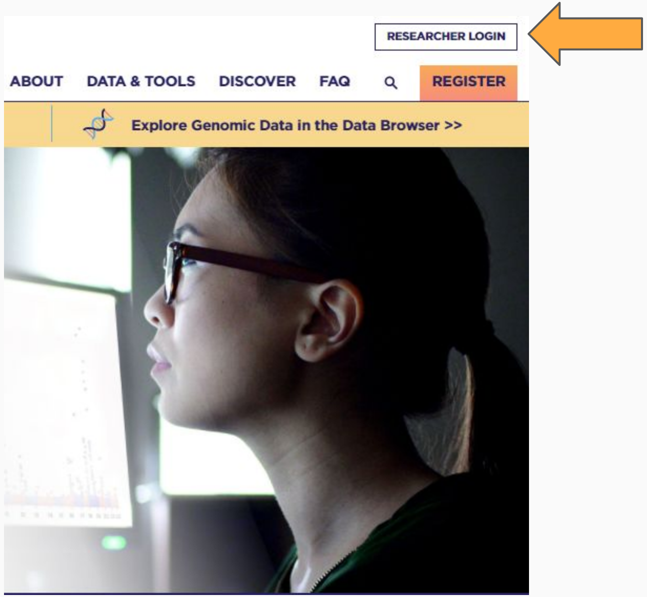
DATASET
BUILDER

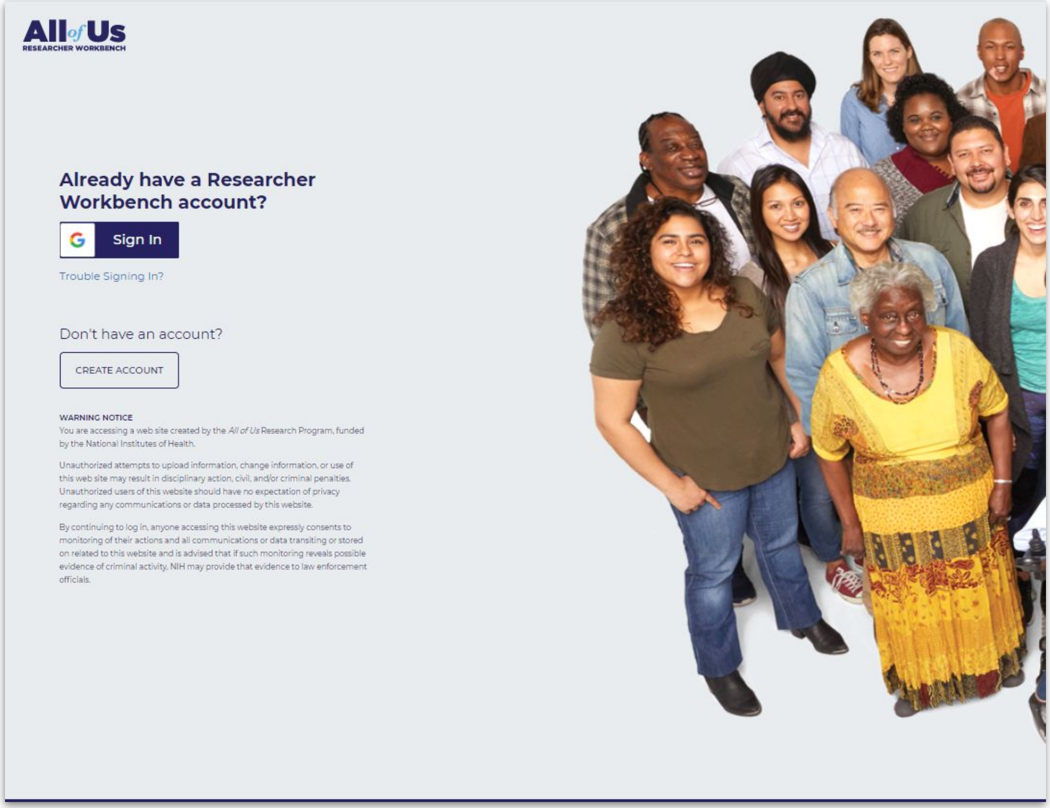
COHORT
BUILDER

NOTEBOOKS

WORKBENCH USER
SUPPORT HUB

Login from the Hub to access your Workspaces.





Already have a Researcher Workbench account?

Don't have an account?

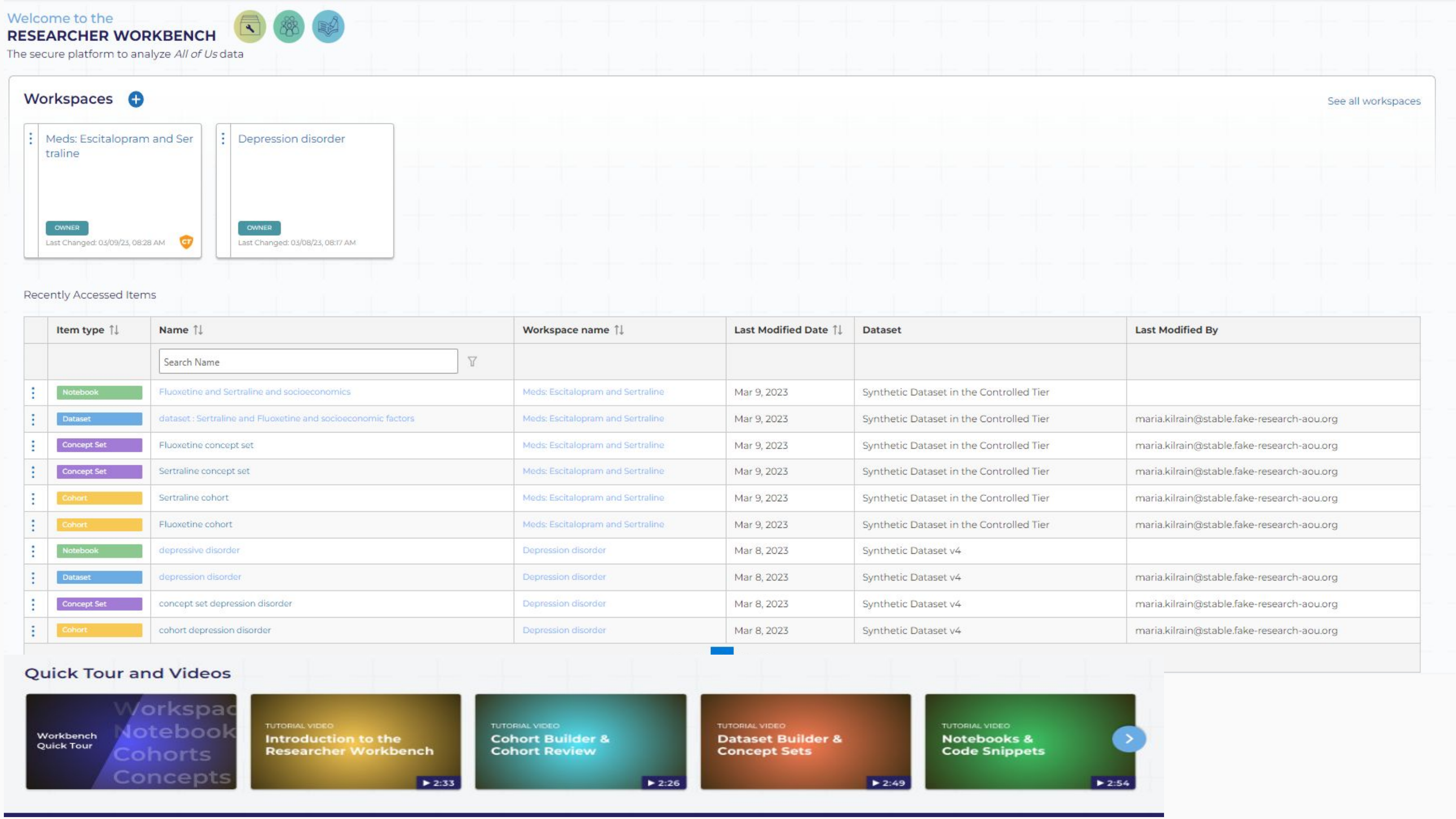
WARNING NOTICE

You are accessing a web site created by the *All of Us* Research Program, funded by the National Institutes of Health.

Unauthorized attempts to upload information, change information, or use of this web site may result in disciplinary action, civil, and/or criminal penalties. Unauthorized users of this website should have no expectation of privacy regarding any communications or data processed by this website.

By continuing to log in, anyone accessing this website expressly consents to monitoring of their actions and all communications or data transiting or stored on related to this website and is advised that if such monitoring reveals possible evidence of criminal activity, NIH may provide that evidence to law enforcement officials.

The Workbench landing page



Welcome to the
RESEARCHER WORKBENCH
The secure platform to analyze *All of Us* data

Workspaces

See all workspaces

Meds: Escitalopram and Sertraline
OWNER
Last Changed: 03/09/23, 08:28 AM

Depression disorder
OWNER
Last Changed: 03/08/23, 08:17 AM

Recently Accessed Items

Item type	Name	Workspace name	Last Modified Date	Dataset	Last Modified By
	Search Name				
Notebook	Fluoxetine and Sertraline and socioeconomics	Meds: Escitalopram and Sertraline	Mar 9, 2023	Synthetic Dataset in the Controlled Tier	
Dataset	dataset : Sertraline and Fluoxetine and socioeconomic factors	Meds: Escitalopram and Sertraline	Mar 9, 2023	Synthetic Dataset in the Controlled Tier	maria.kilrain@stable.fake-research-aou.org
Concept Set	Fluoxetine concept set	Meds: Escitalopram and Sertraline	Mar 9, 2023	Synthetic Dataset in the Controlled Tier	maria.kilrain@stable.fake-research-aou.org
Concept Set	Sertraline concept set	Meds: Escitalopram and Sertraline	Mar 9, 2023	Synthetic Dataset in the Controlled Tier	maria.kilrain@stable.fake-research-aou.org
Cohort	Sertraline cohort	Meds: Escitalopram and Sertraline	Mar 9, 2023	Synthetic Dataset in the Controlled Tier	maria.kilrain@stable.fake-research-aou.org
Cohort	Fluoxetine cohort	Meds: Escitalopram and Sertraline	Mar 9, 2023	Synthetic Dataset in the Controlled Tier	maria.kilrain@stable.fake-research-aou.org
Notebook	depressive disorder	Depression disorder	Mar 8, 2023	Synthetic Dataset v4	
Dataset	depression disorder	Depression disorder	Mar 8, 2023	Synthetic Dataset v4	maria.kilrain@stable.fake-research-aou.org
Concept Set	concept set depression disorder	Depression disorder	Mar 8, 2023	Synthetic Dataset v4	maria.kilrain@stable.fake-research-aou.org
Cohort	cohort depression disorder	Depression disorder	Mar 8, 2023	Synthetic Dataset v4	maria.kilrain@stable.fake-research-aou.org

Quick Tour and Videos

Workbench Quick Tour
Workspaces Notebooks Cohorts Concepts

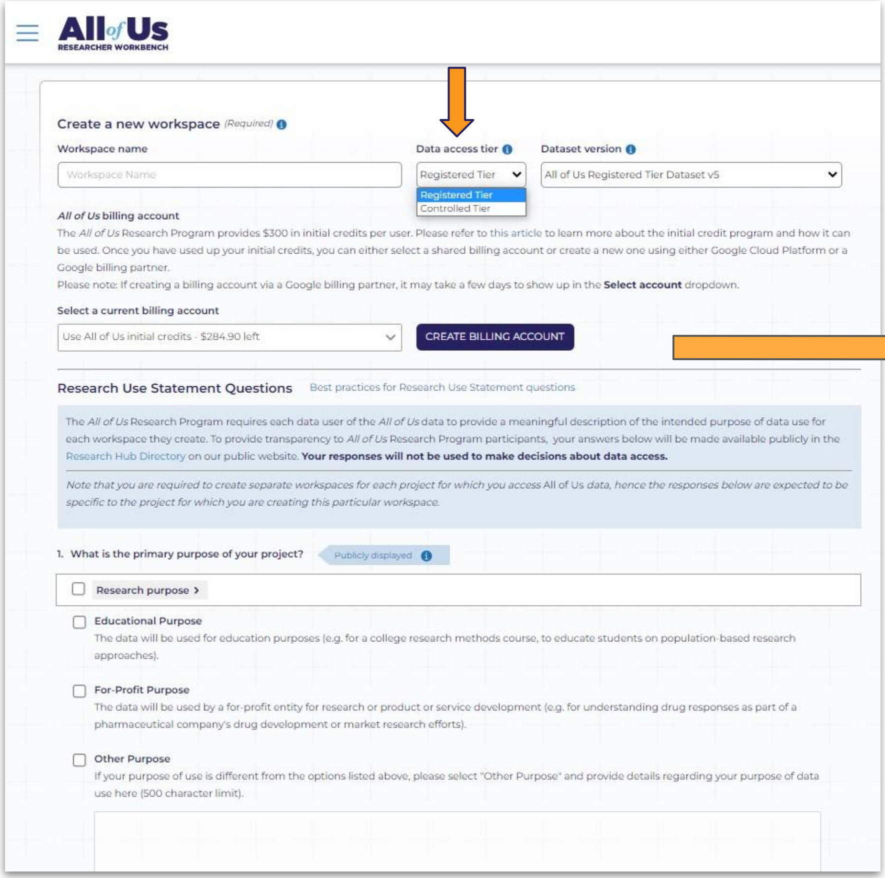
TUTORIAL VIDEO
Introduction to the Researcher Workbench
▶ 2:33

TUTORIAL VIDEO
Cohort Builder & Cohort Review
▶ 2:26

TUTORIAL VIDEO
Dataset Builder & Concept Sets
▶ 2:49

TUTORIAL VIDEO
Notebooks & Code Snippets
▶ 2:54

Answer questions about your Workspace.



Create a new workspace *(Required)*

Workspace name: Workspace Name

Data access tier: Registered Tier
- Registered Tier
- Controlled Tier

Dataset version: All of Us Registered Tier Dataset v5

All of Us billing account

The *All of Us* Research Program provides $300 in initial credits per user. Please refer to this article to learn more about the initial credit program and how it can be used. Once you have used up your initial credits, you can either select a shared billing account or create a new one using either Google Cloud Platform or a Google billing partner.

Please note: If creating a billing account via a Google billing partner, it may take a few days to show up in the **Select account** dropdown.

Select a current billing account: Use All of Us initial credits - $284.90 left

Research Use Statement Questions   Best practices for Research Use Statement questions

The *All of Us* Research Program requires each data user of the *All of Us* data to provide a meaningful description of the intended purpose of data use for each workspace they create. To provide transparency to *All of Us* Research Program participants, your answers below will be made available publicly in the Research Hub Directory on our public website. **Your responses will not be used to make decisions about data access.**

Note that you are required to create separate workspaces for each project for which you access All of Us *data, hence the responses below are expected to be specific to the project for which you are creating this particular workspace.*

1. What is the primary purpose of your project? Publicly displayed

☐ Research purpose ›

☐ Educational Purpose
The data will be used for education purposes (e.g. for a college research methods course, to educate students on population-based research approaches).

☐ For-Profit Purpose
The data will be used by a for-profit entity for research or product or service development (e.g. for understanding drug responses as part of a pharmaceutical company's drug development or market research efforts).

☐ Other Purpose
If your purpose of use is different from the options listed above, please select "Other Purpose" and provide details regarding your purpose of data use here (500 character limit).

Depression and eye diseases

SCIENTIFIC QUESTIONS BEING STUDIED

Questions: What is the risk of developing depression among patients with glaucoma, macular degeneration, and/or diabetic retinopathy? What are the associated risk factors?

Relevance: Visual impairment due to ocular diseases is known to affect our quality of life and mental status. It is important to assess the risk of depression in patients suffering from prevalent ocular diseases, particularly the less easily diagnosed and treated ones, in order to understand their possible impact on the patients' mental outcome and whether they cause similar or discrepant mental burden of visual impairment. Additionally, knowledge about risk factors associated with worse mental outcome in individual ocular disease can inform on the differing follow-up care and outreach measures needed for these patients.

PROJECT PURPOSE(S)

- Disease Focused Research (Glaucoma, macular degeneration, diabetic retinopathy)
- Population Health
- Social / Behavioral

SCIENTIFIC APPROACHES

Datasets: Regular All of Us demographic and survey dataset and the COVID-19 Participant Experience (COPE) survey
Research methods: We will use survival analysis to analyze the risk and time of developing depression for patients with one or multiple of the aforementioned ocular conditions. We also hope to do a multivariable odds ratio analysis that identify risk factors associated with higher risk of developing depression in each individual disease. Supplemental analysis may be done among patients with and without self-reported blindness to examine the effect of self-reported visual impairment on the incidence of depression.

ANTICIPATED FINDINGS

We anticipated a higher risk and shorter time to development of depression for:
1. Patients with ocular conditions vs healthy controls
2. Patients with central vision loss (e.g. macular degeneration) vs peripheral vision loss (e.g. glaucoma)
3. Having multiple ocular conditions vs only one condition
4. Having self-reported blindness vs not having self-reported blindness
Also, the risk factors associated with risk of depression may slightly differ among the 3 conditions

DEMOGRAPHIC CATEGORIES OF INTEREST

This study will not center on underrepresented populations.

DATA SET USED

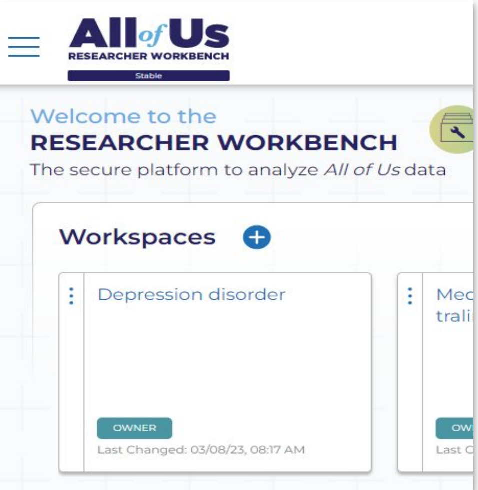
Registered Tier

RESEARCH TEAM

Owner:

- Jo-Hsuan Wu - Research Fellow, University of California, San Diego

The About tab provides information about your Workspace.

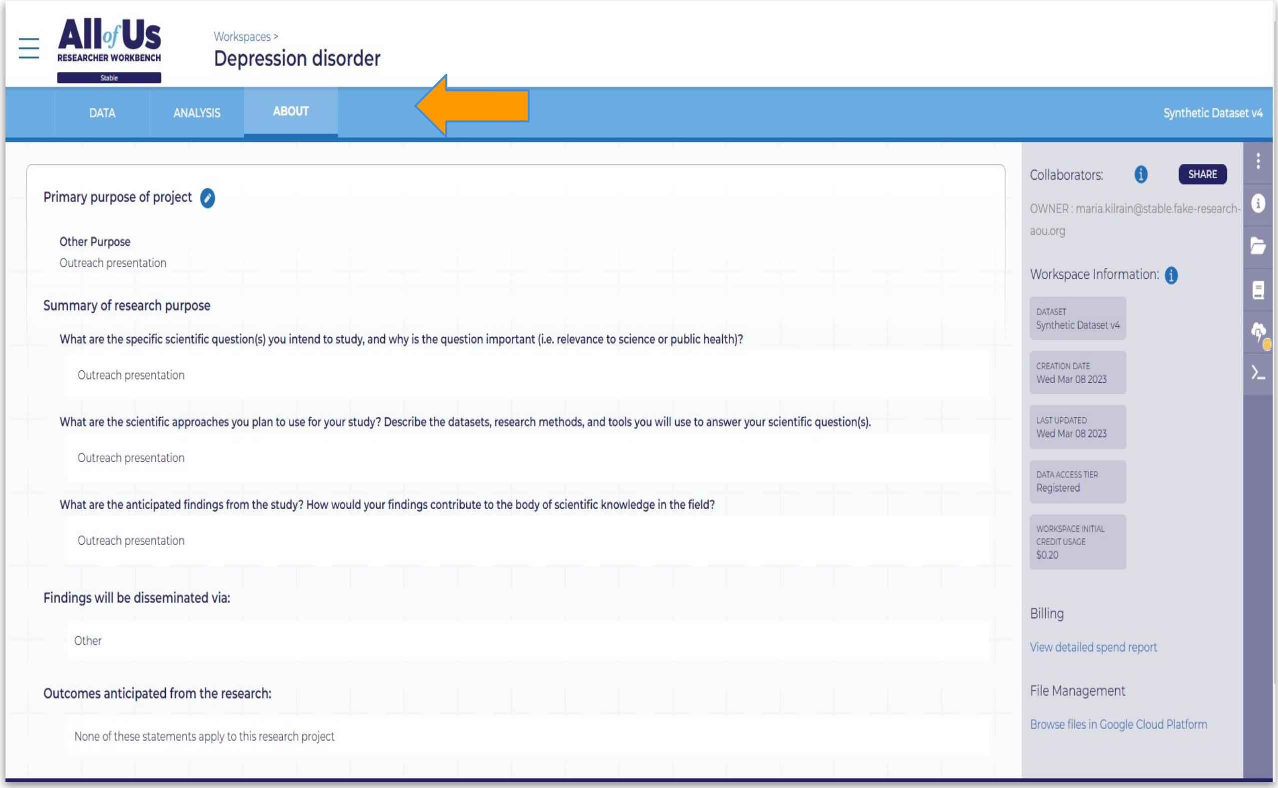


Welcome to the
RESEARCHER WORKBENCH
The secure platform to analyze *All of Us* data

Workspaces

Depression disorder

Last Changed: 03/08/23, 08:17 AM



Depression disorder

Synthetic Dataset v4

Primary purpose of project

Other Purpose
Outreach presentation

Summary of research purpose

What are the specific scientific question(s) you intend to study, and why is the question important (i.e. relevance to science or public health)?

Outreach presentation

What are the scientific approaches you plan to use for your study? Describe the datasets, research methods, and tools you will use to answer your scientific question(s).

Outreach presentation

What are the anticipated findings from the study? How would your findings contribute to the body of scientific knowledge in the field?

Outreach presentation

Findings will be disseminated via:

Other

Outcomes anticipated from the research:

None of these statements apply to this research project

Collaborators:

OWNER : maria.kilrain@stable.fake-research-aou.org

Workspace Information:

DATASET
Synthetic Dataset v4

CREATION DATE
Wed Mar 08 2023

LAST UPDATED
Wed Mar 08 2023

DATA ACCESS TIER
Registered

WORKSPACE INITIAL CREDIT USAGE
$0.20

Billing

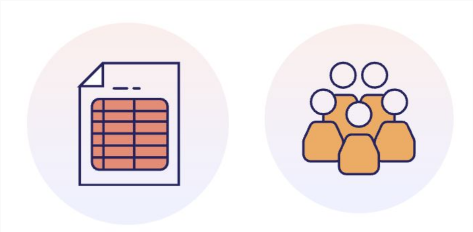
View detailed spend report

File Management

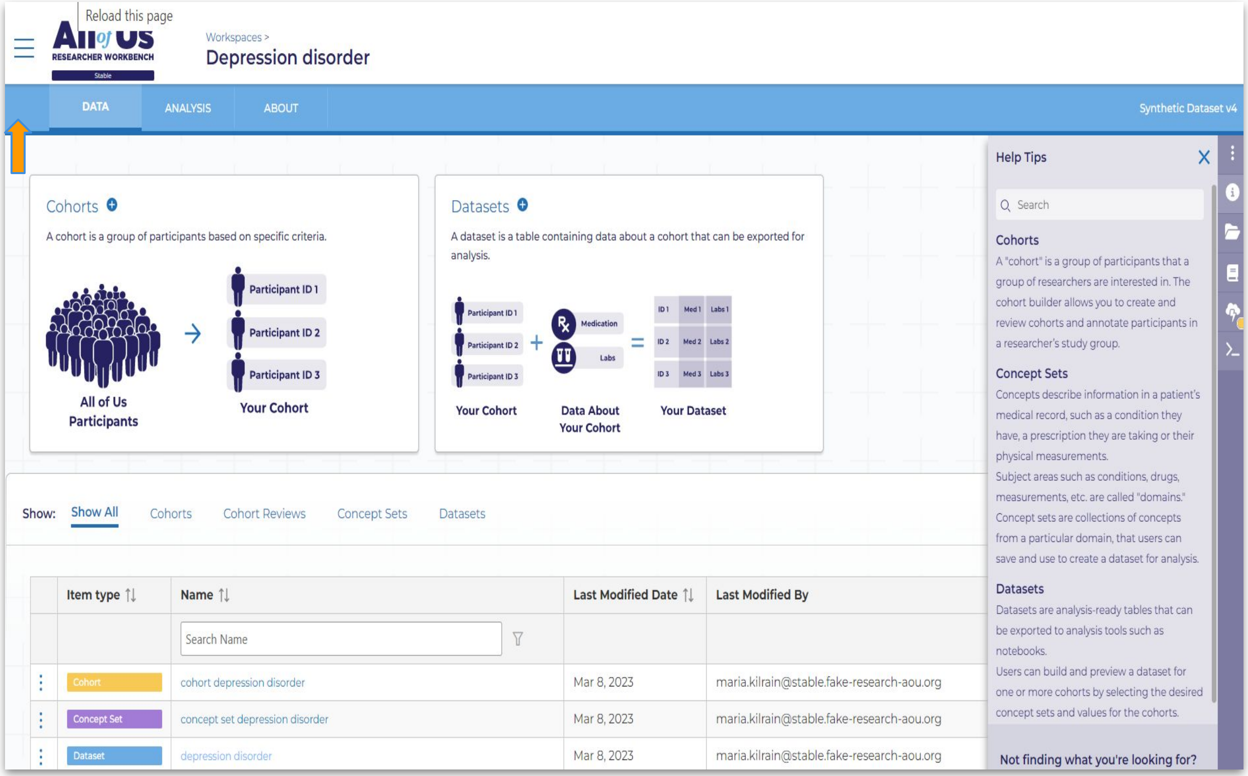
Browse files in Google Cloud Platform

WORKSPACES

The Data tab is where you begin building your research.



Cohort + Dataset Builders



A cohort is a group of participants based on specific criteria.

A dataset is a table containing data about a cohort that can be exported for analysis.

Item type	Name	Last Modified Date	Last Modified By
Cohort	cohort depression disorder	Mar 8, 2023	maria.kilrain@stable.fake-research-aou.org
Concept Set	concept set depression disorder	Mar 8, 2023	maria.kilrain@stable.fake-research-aou.org
Dataset	depression disorder	Mar 8, 2023	maria.kilrain@stable.fake-research-aou.org

Cohorts

A "cohort" is a group of participants that a group of researchers are interested in. The cohort builder allows you to create and review cohorts and annotate participants in a researcher's study group.

Concept Sets

Concepts describe information in a patient's medical record, such as a condition they have, a prescription they are taking or their physical measurements.

Subject areas such as conditions, drugs, measurements, etc. are called "domains."

Concept sets are collections of concepts from a particular domain, that users can save and use to create a dataset for analysis.

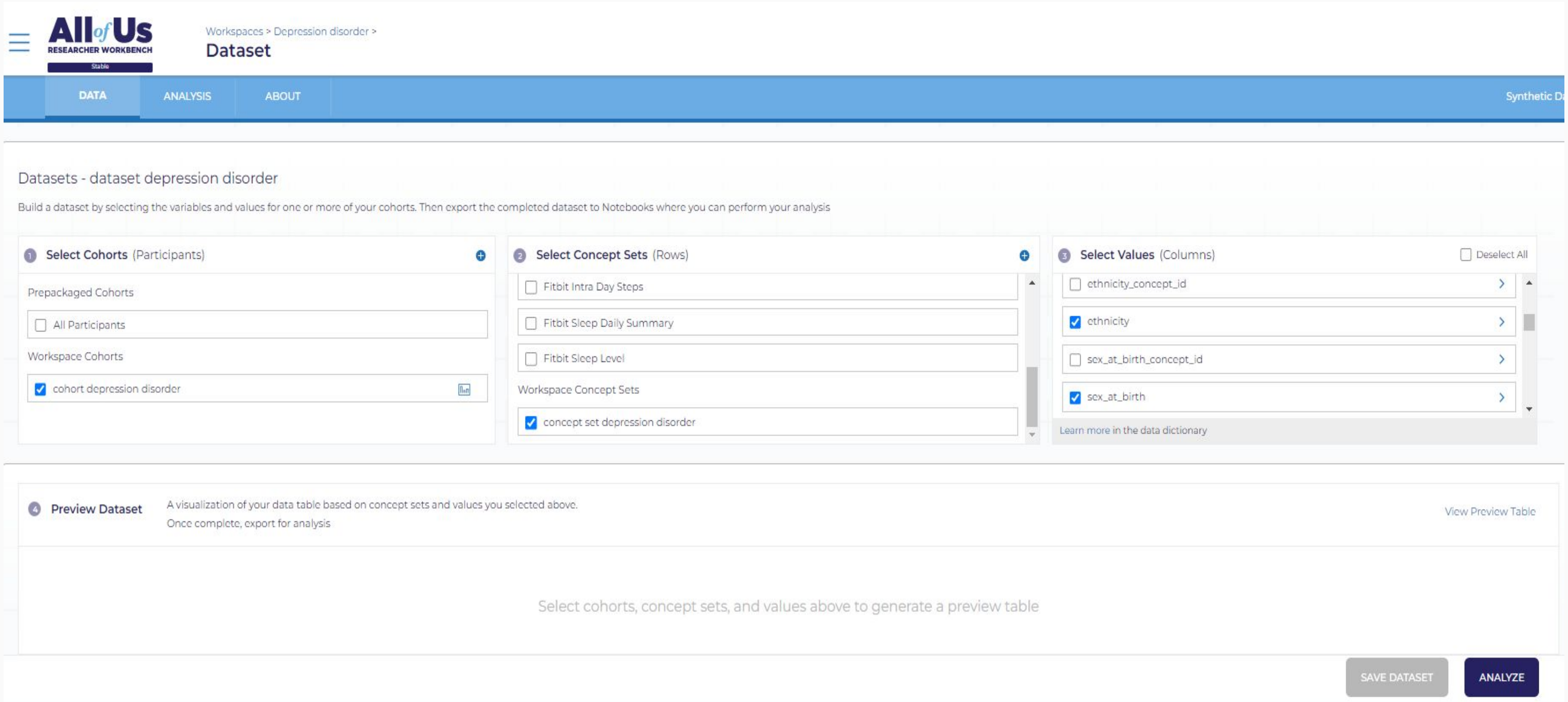
Datasets

Datasets are analysis-ready tables that can be exported to analysis tools such as notebooks.

Users can build and preview a dataset for one or more cohorts by selecting the desired concept sets and values for the cohorts.

Not finding what you're looking for?

Dataset builder



Datasets - dataset depression disorder

Build a dataset by selecting the variables and values for one or more of your cohorts. Then export the completed dataset to Notebooks where you can perform your analysis

1 **Select Cohorts** (Participants)

Prepackaged Cohorts

☐ All Participants

Workspace Cohorts

☑ cohort depression disorder

2 **Select Concept Sets** (Rows)

☐ Fitbit Intra Day Steps
☐ Fitbit Sleep Daily Summary
☐ Fitbit Sleep Level

Workspace Concept Sets

☑ concept set depression disorder

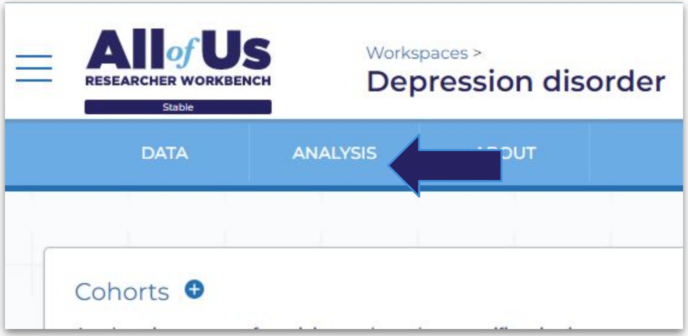
3 **Select Values** (Columns)

☐ ethnicity_concept_id
☑ ethnicity
☐ sex_at_birth_concept_id
☑ sex_at_birth

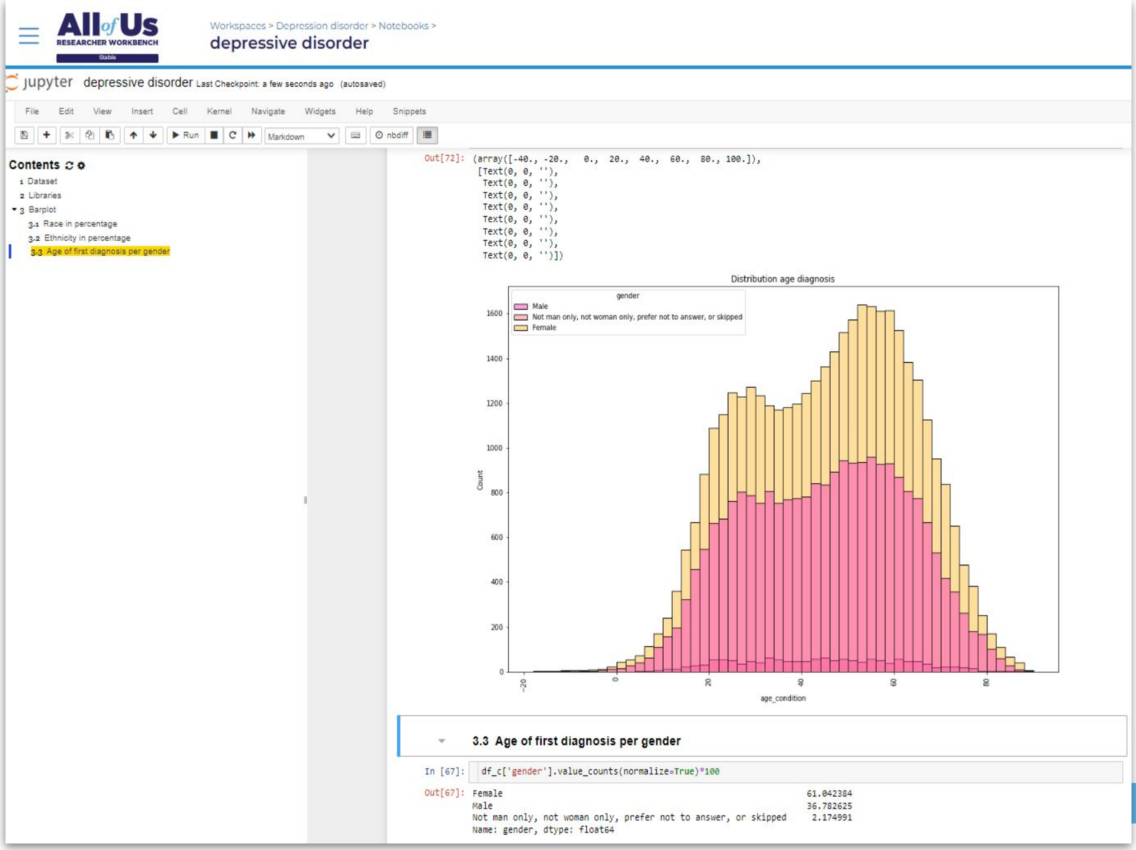
4 **Preview Dataset** A visualization of your data table based on concept sets and values you selected above.
Once complete, export for analysis

Select cohorts, concept sets, and values above to generate a preview table

The Jupyter Notebook is where you will run your analysis.



Notebooks for analysis



```
Out[72]: (array([-40., -20.,   0.,  20.,  40.,  60.,  80., 100.]),
          [Text(0, 0, ''),
           Text(0, 0, ''),
           Text(0, 0, ''),
           Text(0, 0, ''),
           Text(0, 0, ''),
           Text(0, 0, ''),
           Text(0, 0, ''),
           Text(0, 0, '')])
```

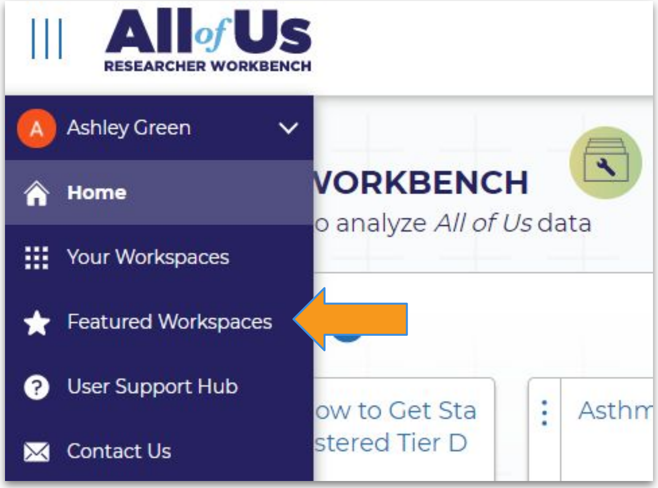
3.3 Age of first diagnosis per gender

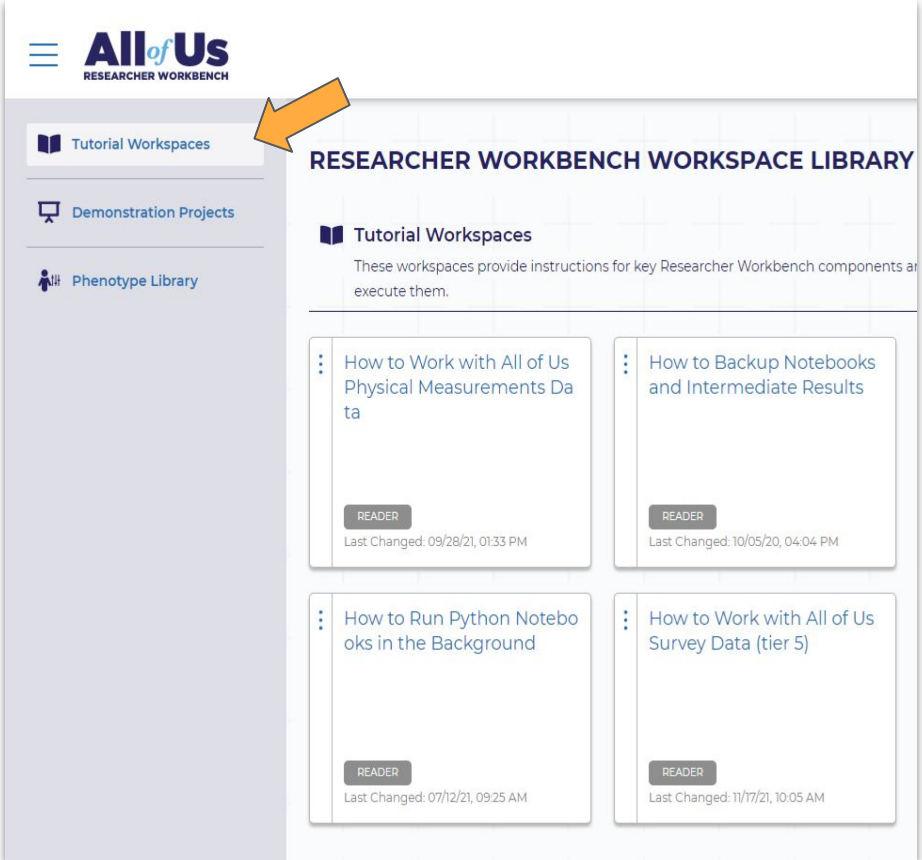
```
In [67]: df_c['gender'].value_counts(normalize=True)*100

Out[67]: Female                                                            61.042384
         Male                                                              36.782625
         Not man only, not woman only, prefer not to answer, or skipped     2.174991
         Name: gender, dtype: float64
```

Support Resources

The Featured Workspaces can help you during your research.





RESEARCHER WORKBENCH WORKSPACE LIBRARY

Tutorial Workspaces

These workspaces provide instructions for key Researcher Workbench components ar
execute them.

How to Work with All of Us Physical Measurements Da ta
READER
Last Changed: 09/28/21, 01:33 PM

How to Backup Notebooks and Intermediate Results
READER
Last Changed: 10/05/20, 04:04 PM

How to Run Python Notebo oks in the Background
READER
Last Changed: 07/12/21, 09:25 AM

How to Work with All of Us Survey Data (tier 5)
READER
Last Changed: 11/17/21, 10:05 AM

We also provide regular office hours & webinars.

Support Desk

Send us your questions.

Email us or use our “Help” icon to:
- Resolve technical issues
- Ask questions about our data or the research program
- Give us feedback or make suggestions
- and more…

Response in 1-2 business days.
support@researchallofus.org

New User Orientation

Zoom in for a live orientation.

Webinar objectives:
- Understand data & data organization
- Learn how to get started on Researcher Workbench
- Learn how to utilize the User Support Hub

Monthly
link emailed to all registered users

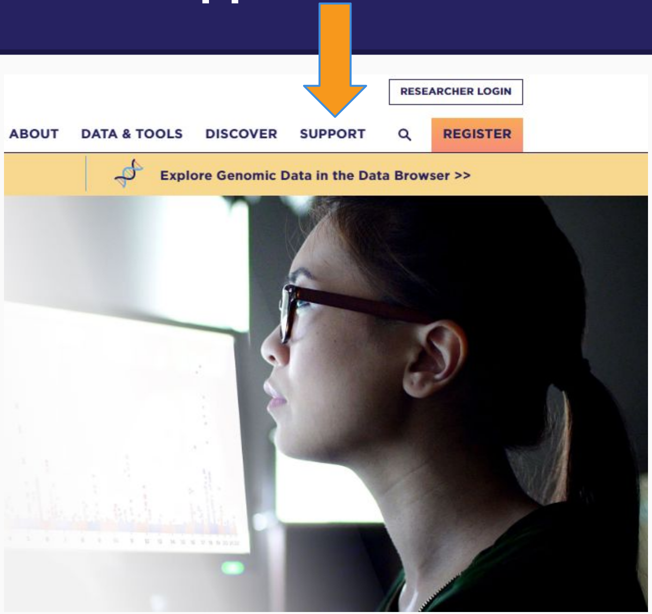
Office Hours

Get live, virtual help with questions.

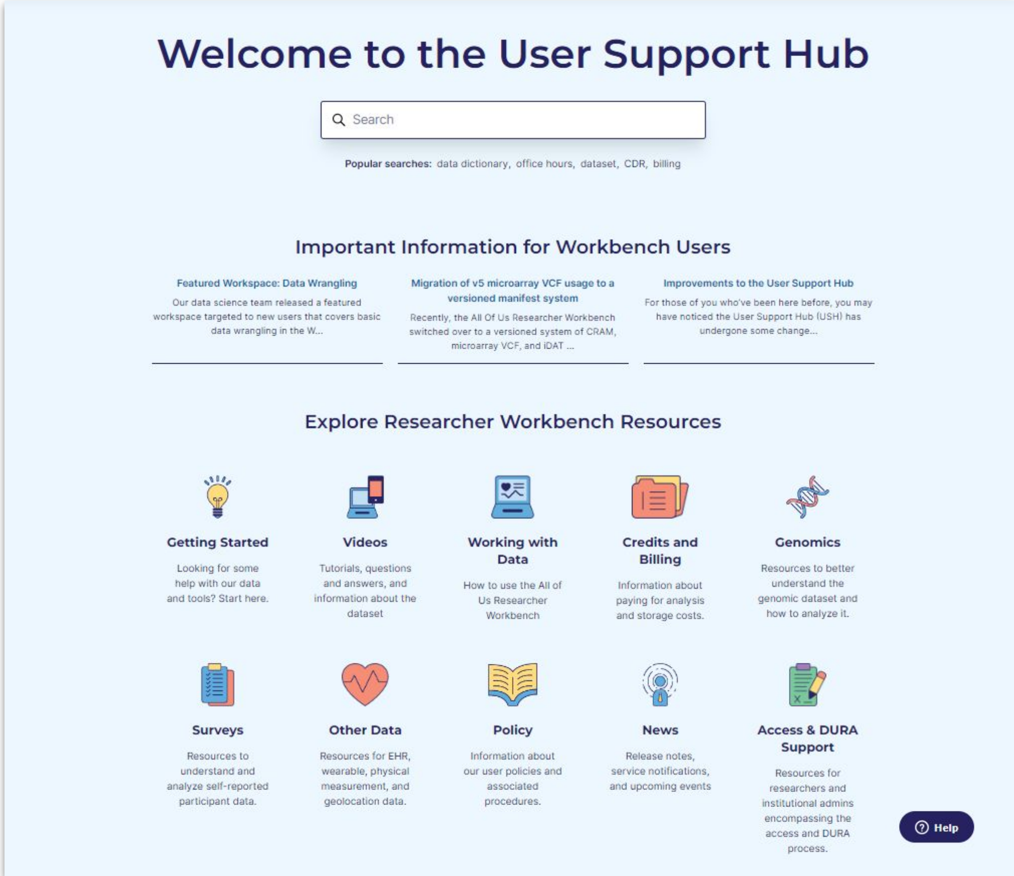
Ask your questions live:
- Workflow implementation questions
- Data refreshers and updates
- Billing questions
- and more…

Weekly (Tues) & Bi-Weekly (Fri)
*link emailed to all **registered users***

The Support Hub and Team are here to help you!



https://aousupporthelp.zendesk.com/hc/en-us



Welcome to the User Support Hub

Popular searches: data dictionary, office hours, dataset, CDR, billing

Important Information for Workbench Users

Featured Workspace: Data Wrangling

Our data science team released a featured workspace targeted to new users that covers basic data wrangling in the W...

Migration of v5 microarray VCF usage to a versioned manifest system

Recently, the All Of Us Researcher Workbench switched over to a versioned system of CRAM, microarray VCF, and iDAT ...

Improvements to the User Support Hub

For those of you who've been here before, you may have noticed the User Support Hub (USH) has undergone some change...

Explore Researcher Workbench Resources

Getting Started

Looking for some help with our data and tools? Start here.

Videos

Tutorials, questions and answers, and information about the dataset

Working with Data

How to use the All of Us Researcher Workbench

Credits and Billing

Information about paying for analysis and storage costs.

Genomics

Resources to better understand the genomic dataset and how to analyze it.

Surveys

Resources to understand and analyze self-reported participant data.

Other Data

Resources for EHR, wearable, physical measurement, and geolocation data.

Policy

Information about our user policies and associated procedures.

News

Release notes, service notifications, and upcoming events

Access & DURA Support

Resources for researchers and institutional admins encompassing the access and DURA process.