All of Us Database:
Research Opportunities & Impact
April 20, 2022

Faculty: Monica J. Rodriguez, MS

Moderator: Carole Treston, RN, MPH, ACRN, FAAN

Association of Nurses in AIDS Care (ANAC) Mission

ANAC fosters the professional development of nurses and others involved in the delivery of health care for persons at risk for, living with, and/or affected by the human immunodeficiency virus (HIV) and its comorbidities. ANAC promotes the health, welfare and rights of people living with HIV around the world.

Housekeeping

- Participants lines are muted during the webinar
- Please type questions and comments in the Question or Chat area
- There will be a Q & A session at the end of the webinar, if time allows

Nursing Continuing Professional Development (NCPD)

ANAC will provide 1 contact hours of NCPD on completion of this activity.

To receive a certificate of completion, attendees must:

· Be registered to attend

· View today's webinar presentation

· Complete the online, post-activity evaluation. You will receive a link to the evaluation by email.

The deadline to claim contact hours is April 19, 2023.

ANAC is accredited as a provider of nursing continuing professional development by the American Nurses Credentialing Center's Commission on Accreditation.

NCPD questions? Email
Julie@anacnet.org

Desired Learning Outcomes

On completion of today's webinar, participants will:

- Describe the diversity of participant data that create robust *All of Us* data sets
- Discuss the potential for nurse and other clinical researcher projects thru the *All of Us* Research Hub
- List strategies that strengthen the researcher pipeline thru the *All of Us* program

Disclosures

The planners and presenters of this educational activity have no relevant financial relationships with ineligible companies to disclose.

ANAC
Association of
Nurses in AIDS Care

All of Us
of
RESEARCH PROGRAM

1 An understanding of the various data points that we are collecting.

2 An understanding of the potential for research projects through the *All of Us* Research Hub.

3 An understanding of the rich and diverse data set currently available.

4 A chance to learn about current research opportunities for growing a diverse researcher pipeline.

What is the
All of Us **Research Program?**

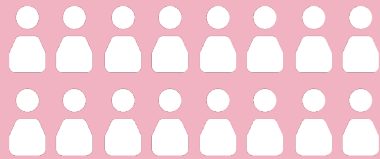# One of the largest, richest, most diverse biomedical datasets of its kind

Inviting

1 Million
or more people across the United States

Data in the Researcher Workbench

329,000+
Participants

80%
are from underrepresented communities

50% are racial and ethnic minorities

Mission & Objectives

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

Nurture partnerships for decades with at least a million participants who reflect the diversity of the United States

Deliver one of the largest, richest biomedical datasets that is broadly available and secure

Catalyze an ecosystem of communities, researchers, and funders who make *All of Us* an indispensable part of health research

Two Methods of Enrollment

HEALTH CARE PROVIDER
ORGANIZATIONS

DIRECT VOLUNTEERS

Current Protocol

Enroll, Consent and Authorize EHR

Recruiting 18+ years old initially; plan to include children in future

Online, interactive consent

Includes the authorization to share EHR data and decide if they want their genomic data returned

Answering Surveys

Current Surveys:
· The Basics
· Overall Health
· Lifestyle
· Health Care Access & Utilization
· Personal & Family Health History
· Social Determinants of Health

Additional surveys will be released on an ongoing basis

Physical Measurements*

Blood pressure

Heart rate

Height

Weight

BMI

Hip circumference

Waist circumference

*Based on diverse sampling and capacity
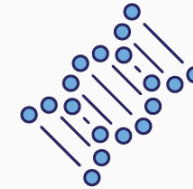
Provide Biosamples*

Blood (or saliva)

Urine specimen

Biosamples will be stored at the program's biobank

*Based on diverse sampling and capacity

Wearables and Digital Apps

Share data from wearable fitness devices, starting with Fitbit

Coming soon:
Integrated apps to track mood & cardio-respiratory fitness
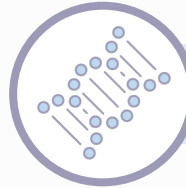
Consent and EHR Authorization

- Consent and EHR Authorization
- Participant Surveys
- Physical Measurements
- Biosamples
- Mobile/Wearable Tech

• Participants must be 18 years or older

• Online video consent

• Consent process includes authorization to share EHR data with researcher

Agree to Share Your Electronic Health Record (EHR)

As part of the consent process, we will also ask you to agree to share your electronic health record (EHR). Sharing your EHR is your choice. If you decide not to share your EHR, you can still participate by answering health surveys. But you will not be asked to take part in some other aspects of the program.

| Initial Data Types	Expanded Data Types (May Include)
• Demographics • Visits • Diagnoses • Procedures • Medications • Laboratory visits • Vital signs	• Clinical notes • Radiology, cardiology, and other reports • Mental health reports • Substance abuse, alcohol use, and tobacco use • More laboratory results, potentially including genomics

Participant Surveys

Consent and EHR Authorization
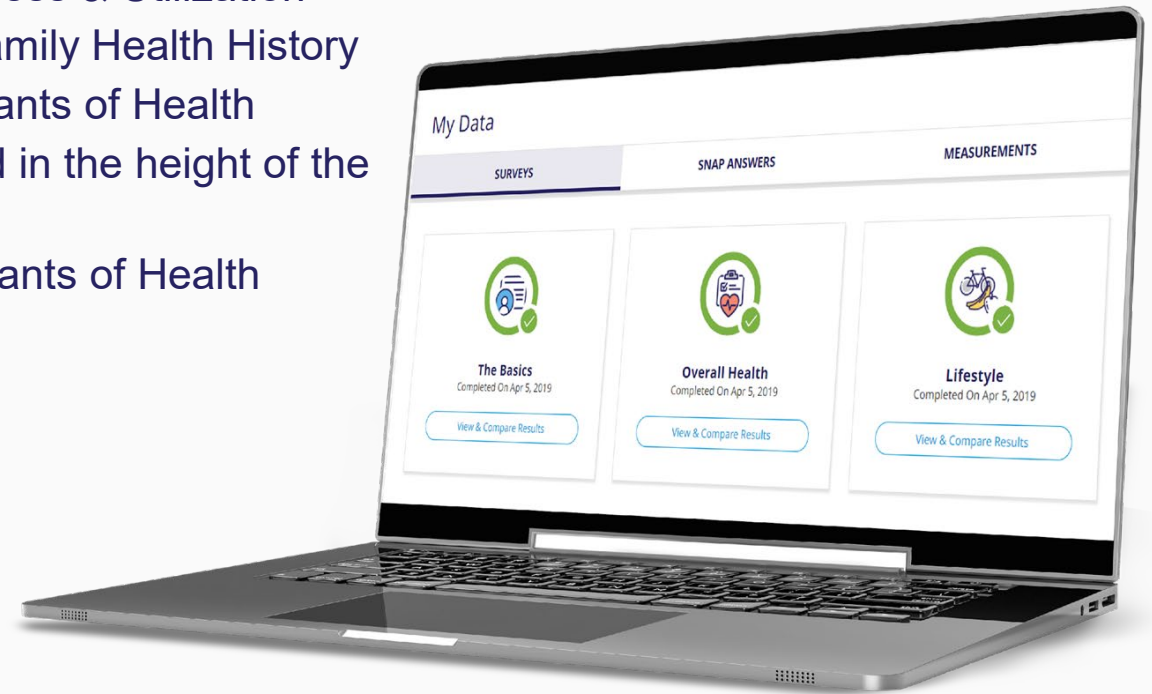
Participant Surveys

Physical Measurements
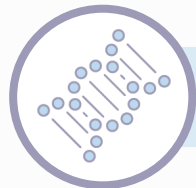
Biosamples

Mobile/Wearable Tech

Current Surveys:

- The Basics
- Overall Health
- Lifestyle
- Health Care Access & Utilization
- Personal and Family Health History
- Social Determinants of Health
- COPE (released in the height of the pandemic)
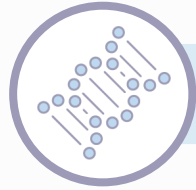- Social Determinants of Health

My Data

SURVEYS | SNAP ANSWERS | MEASUREMENTS

The Basics
Completed On Apr 5, 2019
View & Compare Results

Overall Health
Completed On Apr 5, 2019
View & Compare Results

Lifestyle
Completed On Apr 5, 2019
View & Compare Results

Physical Measurements

- Consent and EHR Authorization
- Participant Surveys
- **Physical Measurements**
- Biosamples
- Mobile/Wearable Tech

Current Measurements:

- Blood pressure
- Heart rate
- Height
- Weight
- BMI
- Hip circumference
- Waist circumference

Biosamples

Consent and EHR Authorization
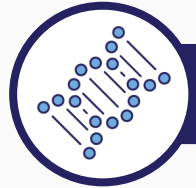
Participant Surveys

Physical Measurements

Biosamples

Mobile/Wearable Tech

Examples:

- Blood
- Saliva
- Urine

Mobile/Wearable Tech

Consent and
EHR Authorization

Participant Surveys

Physical Measurements

Biosamples

Mobile/Wearable Tech

Examples:

- Data from wearable fitness devices, including Fitbit and Apple HealthKit
- Additional integrations in the future
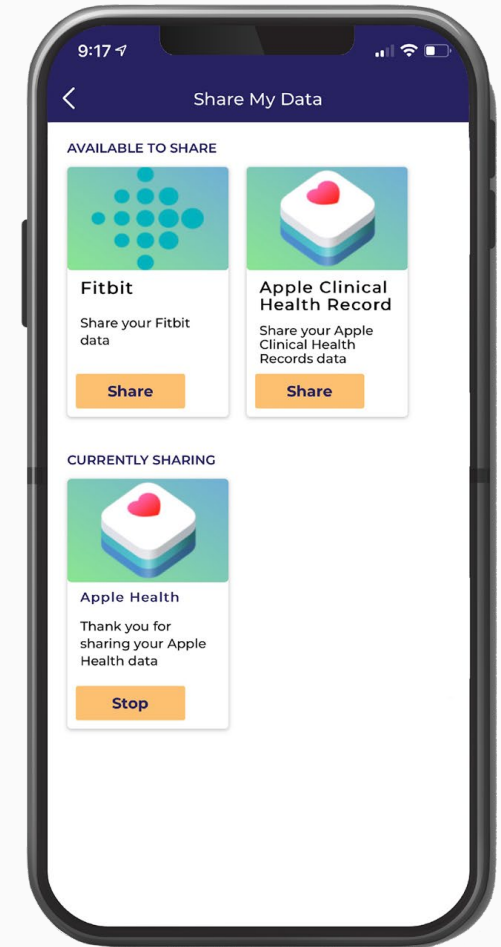
Genomics Plans

- Infrastructure
  - **Three Genome Centers**: to generate genotyping & whole genome sequencing for one million or more participants
  - **Genetic Counseling Resource**: to support the responsible return of information to interested participants

- **Return of Information**

  Over time, the program anticipates providing several kinds of information of interest to participants:

  - Traits and ancestry
  - Drug-gene interactions
  - Genetic findings connected with risk of certain diseases

Genetics Engagement Module (GEM) with Color

Ancestry Analysis
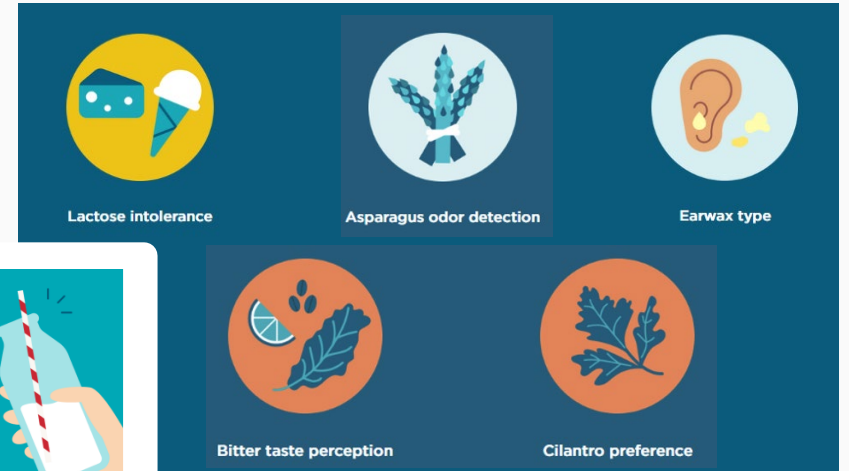
Simple Traits

Lactose intolerance

Asparagus odor detection

Earwax type

Bitter taste perception

Cilantro preference

Lactose Intolerance

Lactose intolerance means you have trouble digesting lactose contained in dairy products like milk or ice cream.

Linda, you are likely to be lactose intolerant.

What we looked at and why

A position in your DNA was analyzed that influences how much lactase your body makes. Lactase is an enzyme that breaks down lactose, a sugar found in milk and other dairy products.

- **People who do not experience lactose intolerance** produce enough lactase enzyme and are usually able to digest lactose.

DNA marker	Gene ❷	Your result ❷
rs4988235	*MCM6*	GG

*Each of your parents provides you with a nucleotide at this position, but we don't know which parent gave you

color

Scientific Framework

Scientific Framework

Enable research that will:

- Increase wellness and resilience, and promote healthy living

- Reduce health disparities and improve health equity in populations that are historically underrepresented in biomedical research (UBR)

- Develop improved risk assessment and prevention strategies to preempt disease

- Provide earlier and more accurate diagnosis to decrease illness burden

- Improve health outcomes and reduce disease impact through improved treatment and development of precision interventions

Example Use Case: Infectious Disease

Risk & Prevention
Does infection with a neurotropic virus increase risk of neurodegenerative disease later in life?

Health Equity
What are the socioeconomic factors associated with vaccination rates?

Diagnosis
How do we better predict patients at risk of nosocomial complications and prevent their occurrence?

Wellness & Resilience
What is the role of social connectivity in the ability to adhere to management of chronic infectious diseases, including HIV?
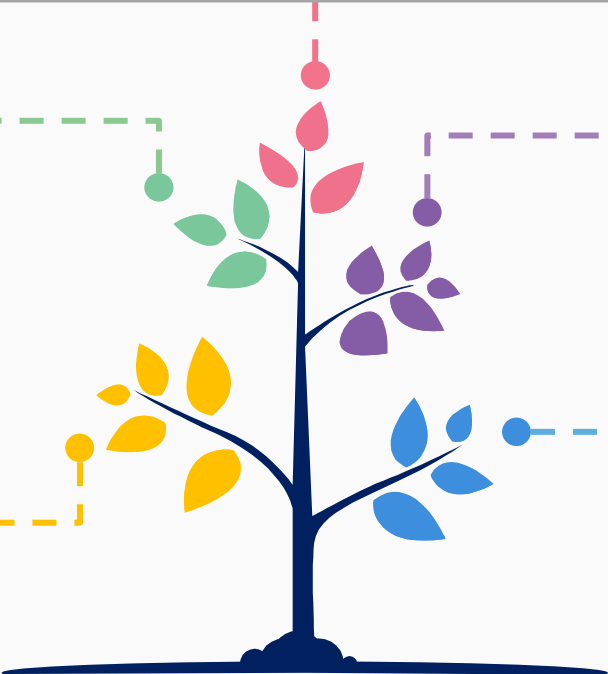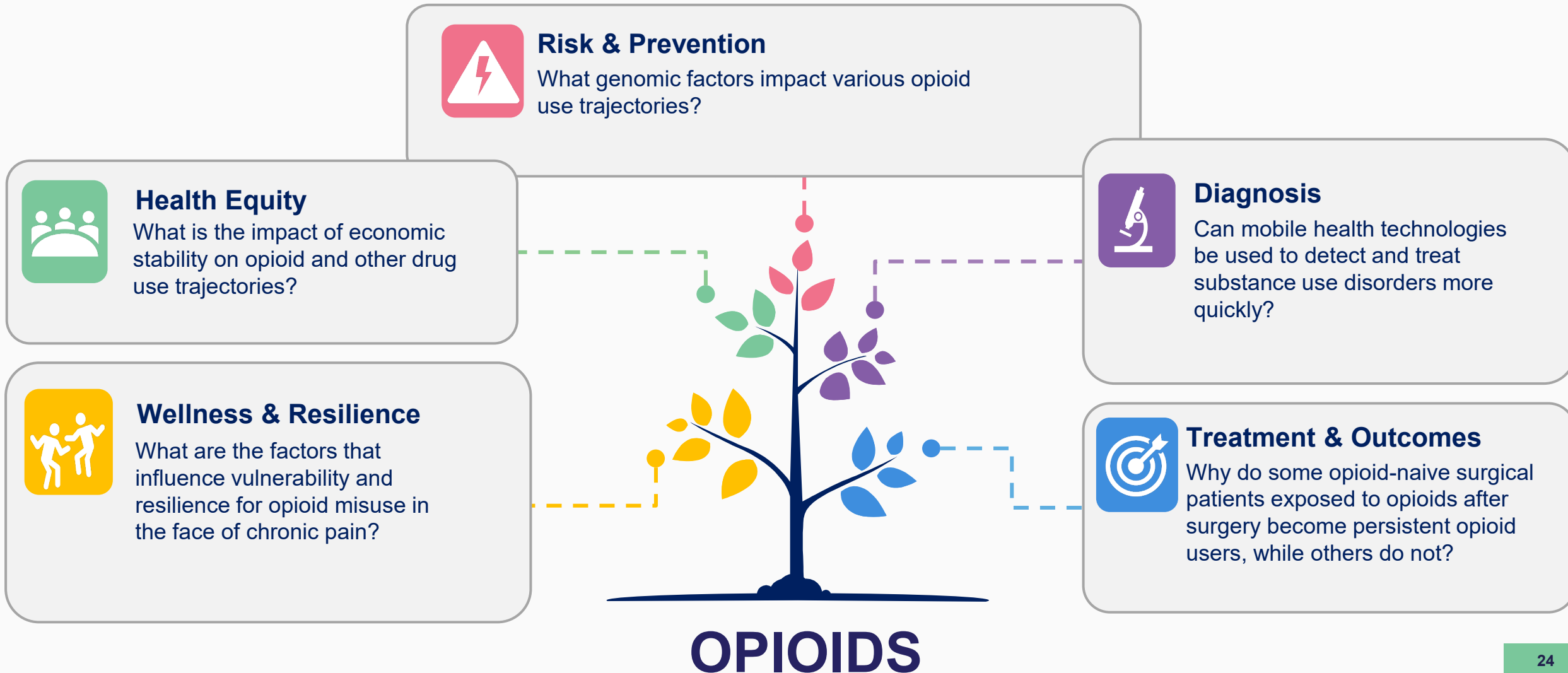
Treatment & Outcomes
What are the long-term outcomes of HIV pre-exposure prophylaxis on individuals without disease?

INFECTIOUS DISEASE

Example Use Case: Opioid Use and Dependence

Risk & Prevention
What genomic factors impact various opioid use trajectories?

Health Equity
What is the impact of economic stability on opioid and other drug use trajectories?

Diagnosis
Can mobile health technologies be used to detect and treat substance use disorders more quickly?

Wellness & Resilience
What are the factors that influence vulnerability and resilience for opioid misuse in the face of chronic pain?

Treatment & Outcomes
Why do some opioid-naive surgical patients exposed to opioids after surgery become persistent opioid users, while others do not?

OPIOIDS

24

Research Hub

All of Us Research: How it Works

All of Us RESEARCH PROGRAM

1. **Participants share their data** with the *All of Us* Research Program through multiple sources. These data are sent to a secure cloud environment, managed by the Data and Research Center.

2. **Participant data** is received and funneled through a **curation pipeline** within a secure repository that connects to the Research Hub tools.

3. **Anyone can visit the Research Hub** to learn more about the types of data *All of Us* makes available for research. The **Survey Explorer** and **Data Browser** offer more information about the unique data elements and let visitors browse aggregated participant data.

4. **Researchers register** and **apply** for access to the Researcher Workbench to analyze data.

5. **Registered researchers** in the Researcher Workbench can create research projects using collaborative workspaces, cohort-building tools, interactive notebooks, and more.

6. **Research underway** can be viewed on the **Projects Directory**. Publications related to *All of Us* Research Program data are posted on the **Publications page**.

And making them accessible to researchers across stages and settings

Our Researchers

1,500+
Registered Researchers
across a range of institutional roles and career stages

290+
Institutions

Researchers **nationwide** are using the **Researcher Workbench**

Category	Value
Academic Institutions*	204
Non-Profit Organizations	43
Health Care Organizations	44
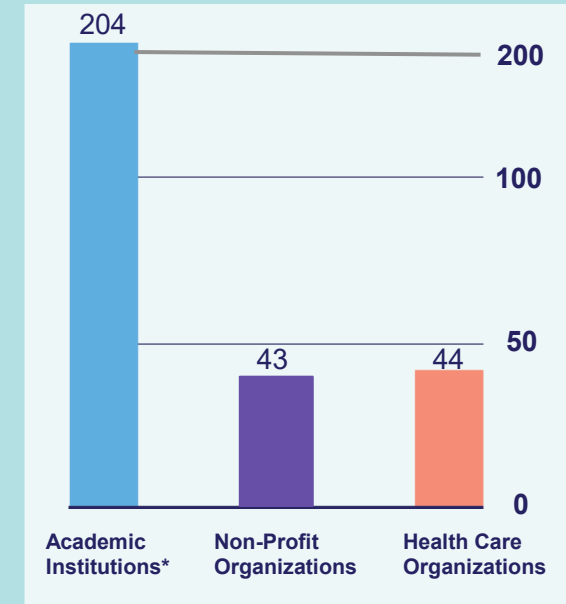
*Includes 36 Historically Black Colleges and Universities and Hispanic-Serving Institutions.

Figures accurate as of March 2022

Research Currently Underway

>1,100 Active projects

37+ Publications in peer-reviewed journals

Top conditions being studied
In the Researcher Workbench include:
- Cardiovascular disease
- Hypertension
- Mental Health
- Cancer
- Diabetes

Including the first batch of genomic data available

Genomic Data is Paired with Rich Phenotypic Data

77,000+
Have Whole Genome Sequences + Electronic Health Records + Physical Measurements + Survey Responses

95,000+
Have Whole Genome Sequences + Physical Measurements + Survey Responses

78,200+
Have Whole Genome Sequences + Electronic Health Records

3,500+
Have Whole Genome Sequences + Fitbit Records
Representing >30% of all participant Fitbit records

Within the Controlled Tier:

98,600+
Whole Genome Sequences

165,000+
Genotyping Arrays

593,500,000+
Unique Variants
Represented and searchable through public genomic data browser
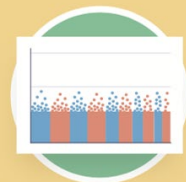
Genomics Analysis Tools
Hail and PLINK in addition to R, Python, and Jupyter Notebooks

More Individual-Level Information
- COVID-19 testing and diagnosis data
- Real dates of health events
- Residential location (first 3 digits of ZIP code)
- ICD codes

More Granular-Level Information
Non-generalized, demographic data provided by participants

- Race and ethnicity
- Sex at birth
- Gender identity
- Sexual orientation
- Education
- Employment status

Research Hub Website

ResearchAllofUs.org/explore

- **Goal: To provide more information about program data and tools.**
  - **Data Snapshots**: broad cohort metrics
  - **Data Browser**: interactive tool
  - **Survey Explorer**: source information for participant-provided information
  - **Researcher Workbench**: restricted cloud-based platform designed to execute rapid, hypothesis-driven research
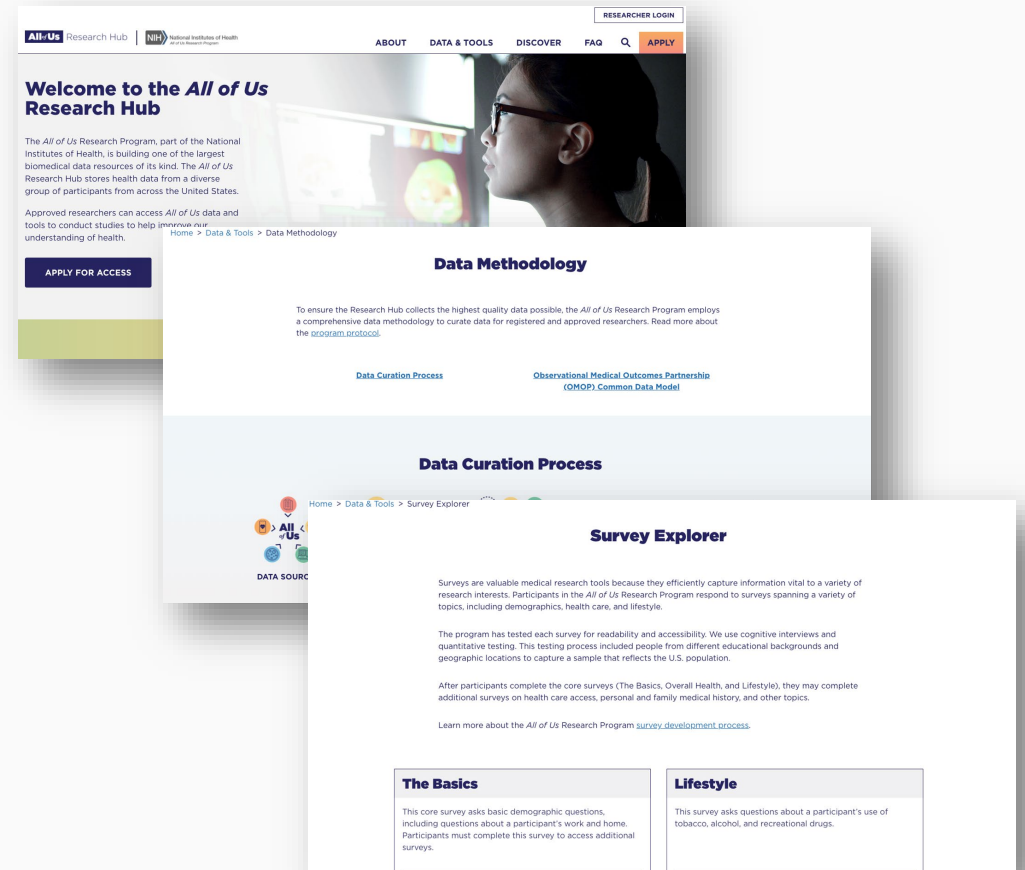
Tiered access levels enable discovery

Available to anyone

Public Tier

Anyone can visit ResearchAllofUs.org (the *All of Us* Research Hub) to learn more about the data available for research and explore aggregated participant data and summary statistics, with identifiers removed. Public resources include:

- **Data Snapshots:** Aggregated, public-facing overviews of participant characteristics and data types
- **Data Browser:** Interactive preview into the larger *All of Us* dataset through public-facing aggregate data
  - Currently includes participant-provided information such as surveys and physical measurements taken at the time of participant enrollment as well as electronic health record (EHR) data
- **Survey Explorer :** Collection of participant survey questions and source information
- **Research Projects Directory:** Information about each research project within the Researcher Workbench

RESEARCHER WORKBENCH

Available to registered researchers

Registered Tier

Registered users can access curated, individual-level data and a variety of research tools to conduct a wide range of studies.

Surveys

Electronic Health Records

Physical Measurements

Wearables
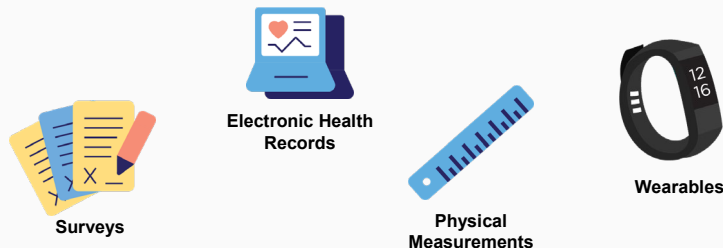
Data have been processed to protect participant privacy

Controlled Tier

Registered users with amended institutional agreements can access **all of the data in the registered tier plus expanded and NEW data** including:

- Expanded demographics
- Responses to entire COPE Survey
- COVID-19 EHR data
- Unshifted event dates
- Genomic data derived from WGS (~100k participants) and array genotyping (>165k participants)

Genomics

Selected Data Snapshots

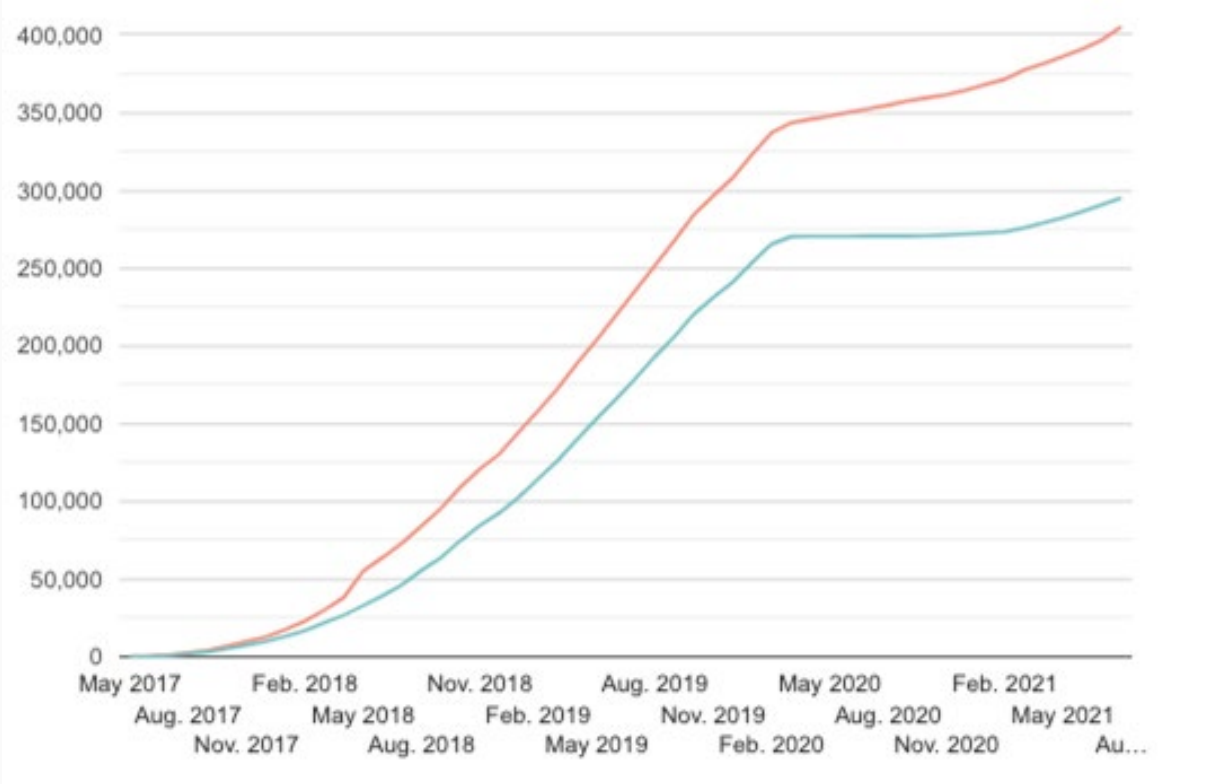(Updated 9/21/21)

411,000+
Participants

297,000+
Participants who have
completed initial steps
of the program

Enrollment Numbers

*The following numbers are approximated to protect participants' privacy.
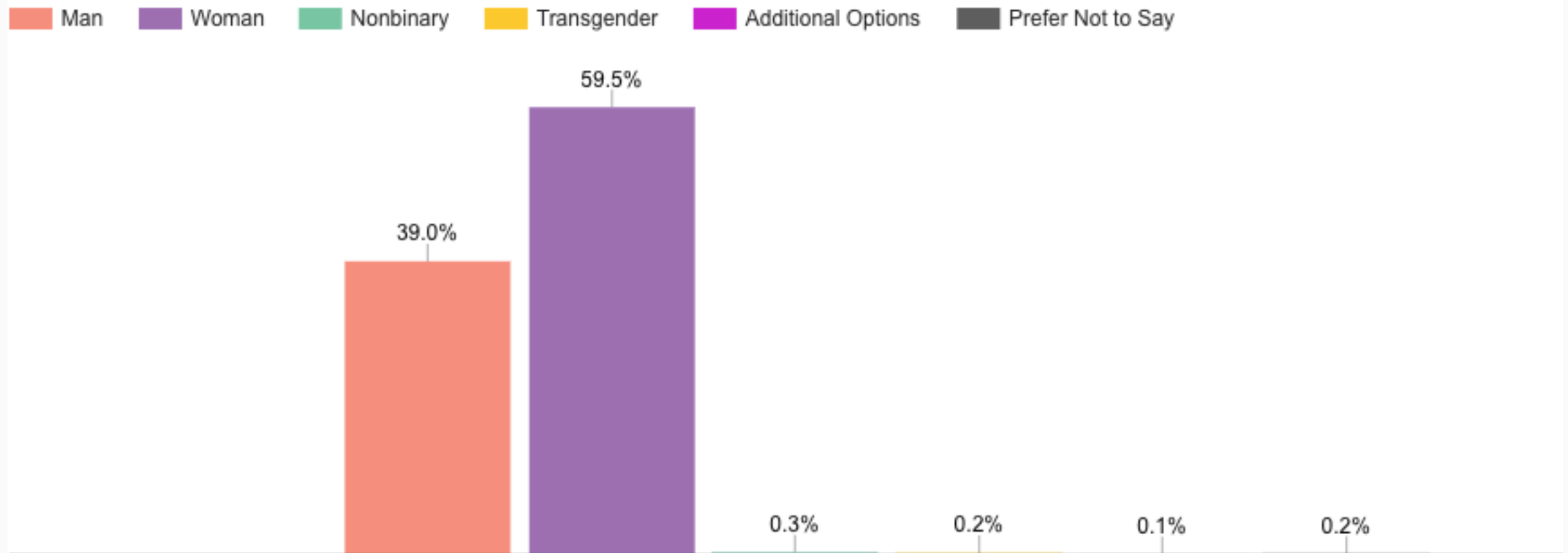Numbers are updated as of September 20, 2021.*

— Participants   — Participants who have completed the initial steps of the program

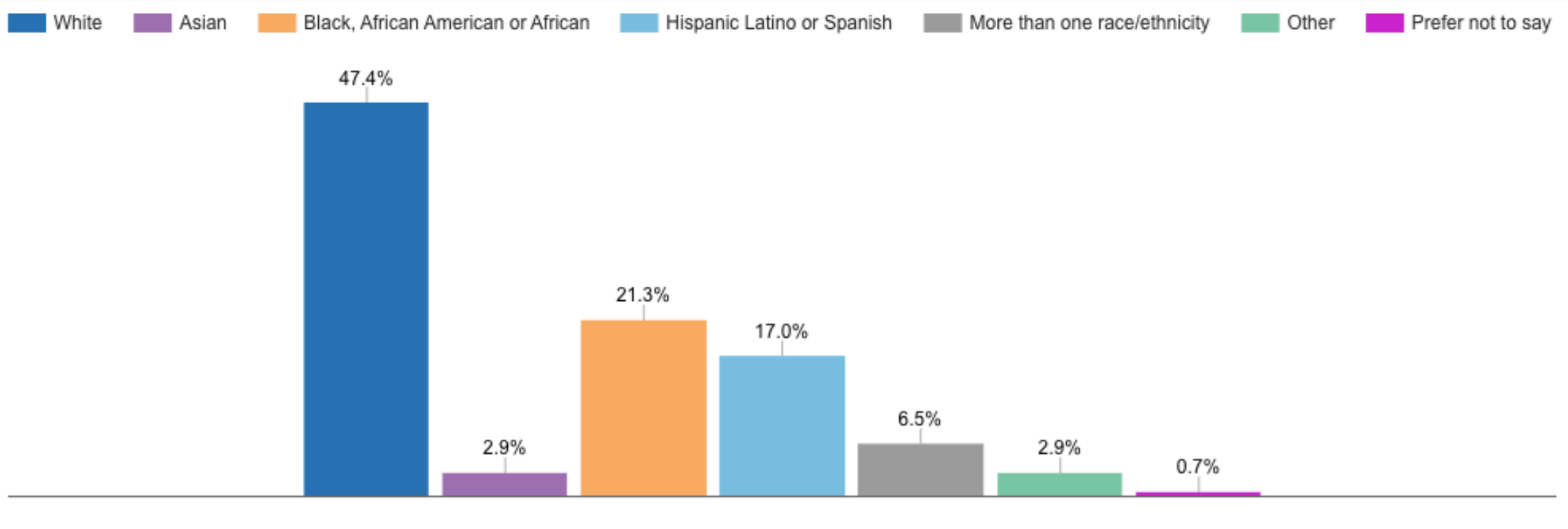# Selected Data Snapshots

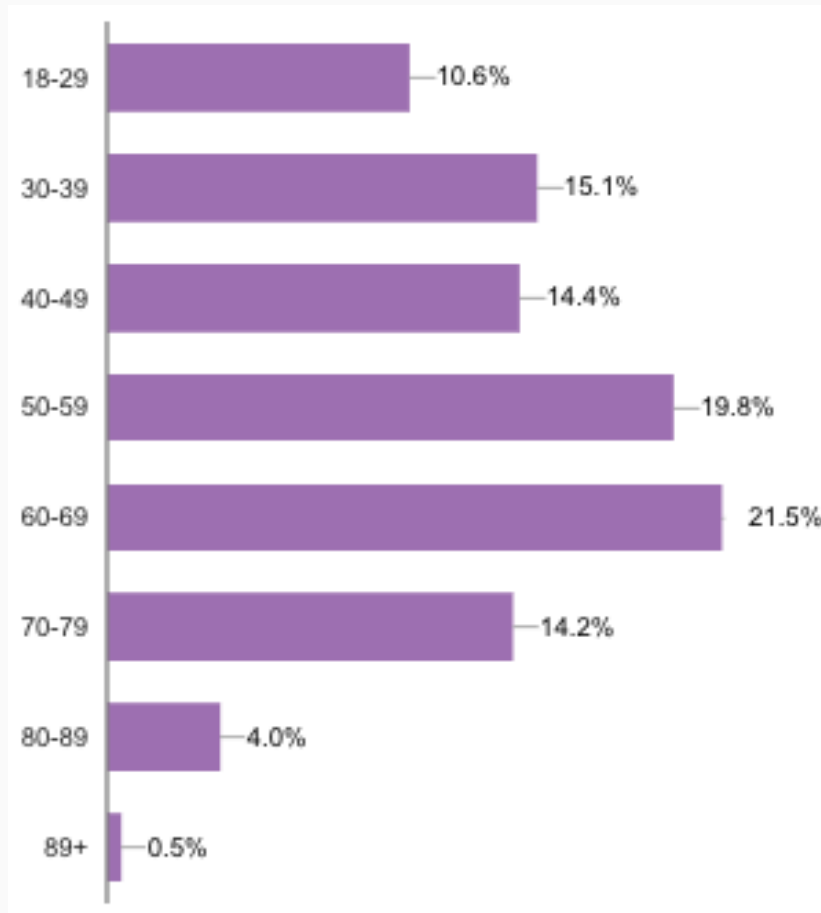(Updated 9/21/21)

Gender Identity

Selected Data Snapshots

(Updated 9/21/21)

Age

Age	Percentage
18-29	10.6%
30-39	15.1%
40-49	14.4%
50-59	19.8%
60-69	21.5%
70-79	14.2%
80-89	4.0%
89+	0.5%

Geography

20 ▬▬▬▬ 44,080

All 50 states
Bilingual enrollment
Interactive mobile exhibits
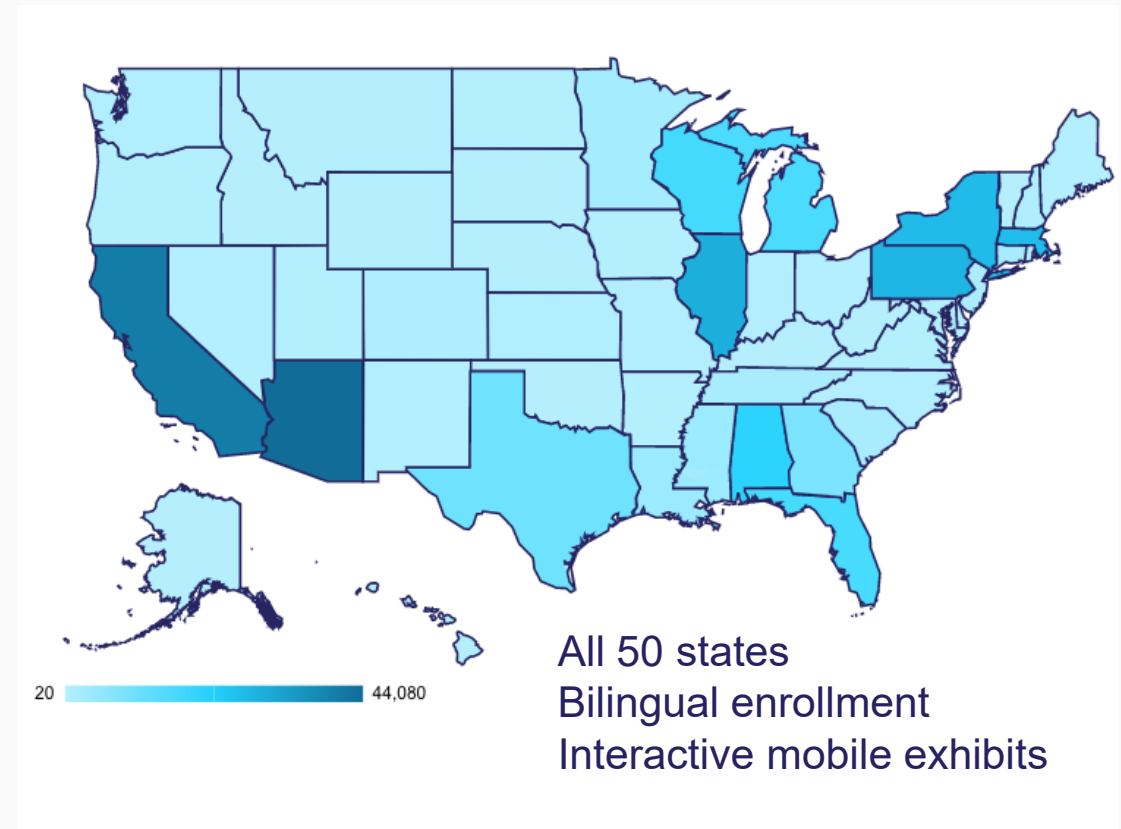
All of Us Data Browser

DataBrowser.ResearchAllofUs.org

- **Interactive tool launched in beta on May 6, 2019.**
  - Provides summary statistics from the program's growing database
  - Open to everyone – no login
  - Allows participants to understand the makeup of the cohort
  - Allows researchers to understand the characteristics of our participant population, explore the data types available, plan research questions
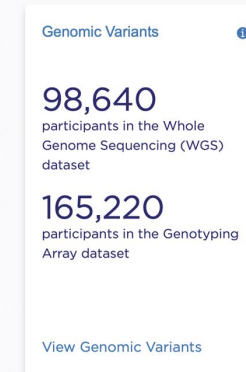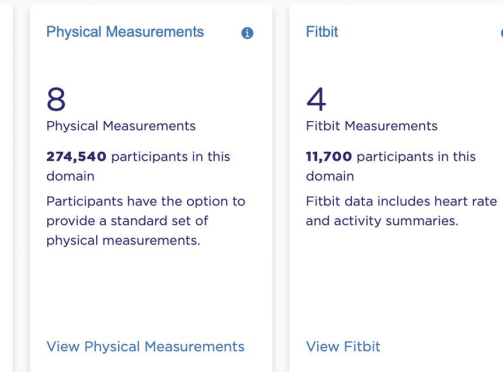
A Quick Look at the Data Browser

Search for specific keywords or browse using the different options underneath.

A Quick Look at the Data Browser

Available data gives insight into the participant cohort and research opportunities.

Top 10 by Descending Participant Counts ∨
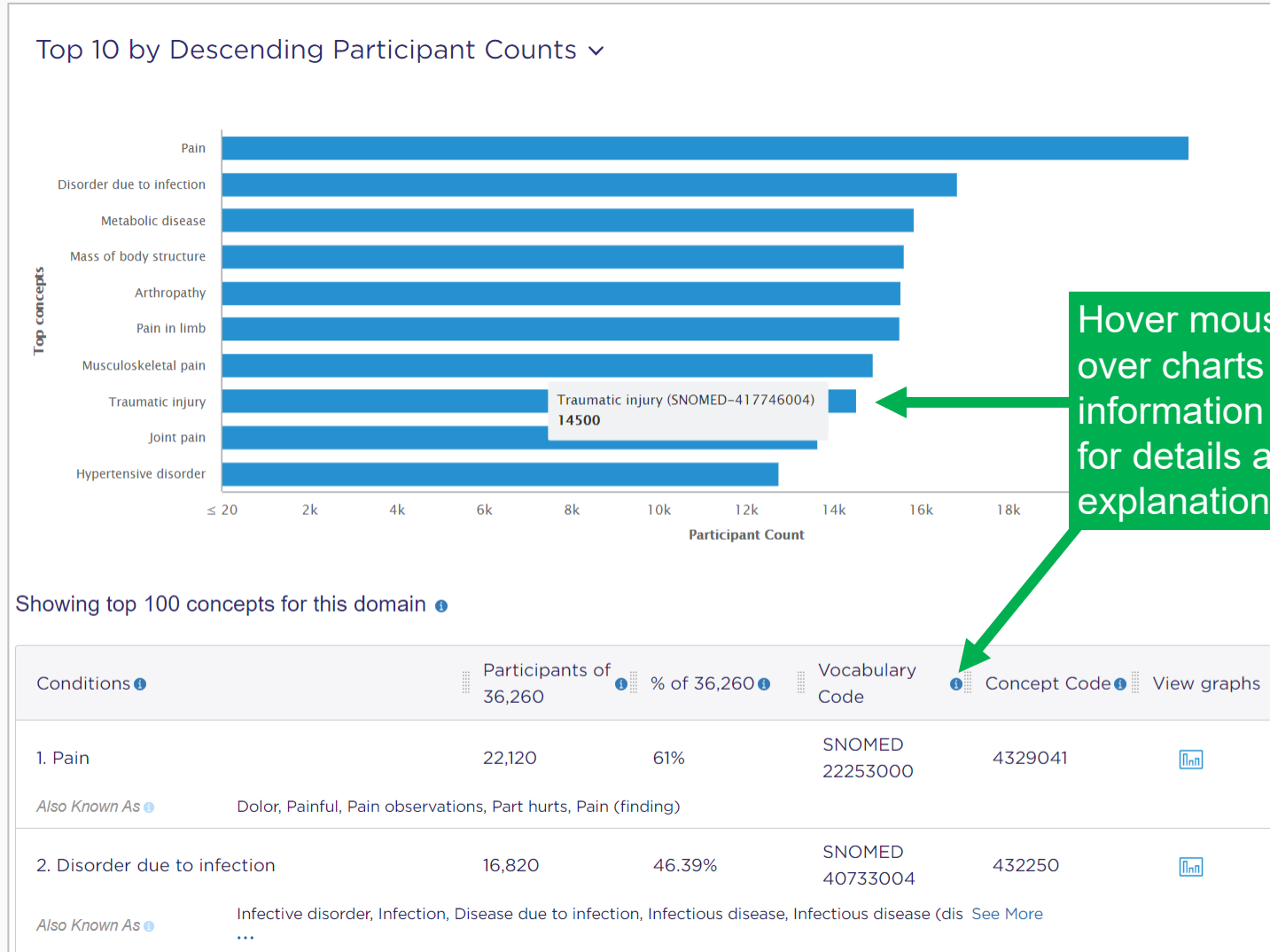
Hover mouse over charts and information icons for details and explanations.

Showing top 100 concepts for this domain ⓘ

Conditions ⓘ	Participants of 36,260 ⓘ	% of 36,260 ⓘ	Vocabulary Code	Concept Code ⓘ	View graphs
1. Pain	22,120	61%	SNOMED 22253000	4329041	▥
Also Known As ⓘ	Dolor, Painful, Pain observations, Part hurts, Pain (finding)				
2. Disorder due to infection	16,820	46.39%	SNOMED 40733004	432250	▥
Also Known As ⓘ	Infective disorder, Infection, Disease due to infection, Infectious disease, Infectious disease (dis See More ...				

A Quick Look at the Data Browser

Researchers can view the full surveys, including branching logic.

Download Survey as PDF

100,460
Participants completed this survey

7
Questions displayed

Question 1
Have you smoked at least 100 cigarettes in your entire life? (There are 20 cigarettes in a pack.)
See Answers ⌄

	Concept Code ⓘ	Participant Count	% Answered	
No	1585859	59,140	58.87%	
Yes	1585858	39,060	38.88%	⌄

↳ Do you now smoke cigarettes every day, some days, or not at all?

ANSWER	Concept Code	Participant Count	% Answered	
Not at all	1585863	22,060	21.96%	
Every day	1585861	11,420	11.37%	
Some days	1585862	5,560	5.53%	
Prefer Not To Answer	903079	1,200	1.19%	
Skip	903096	340	0.34%	

A Quick Look at the Data Browser

Explore data breakdowns by sex assigned at birth and age.
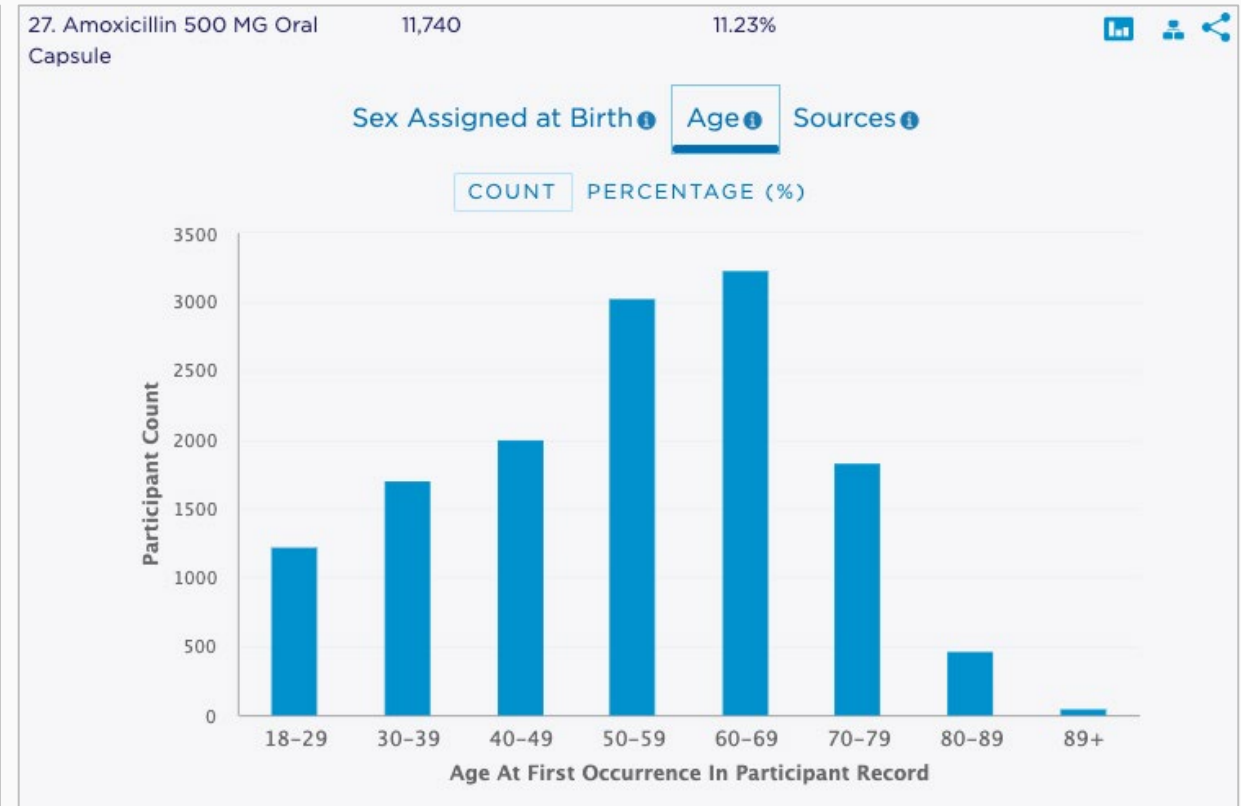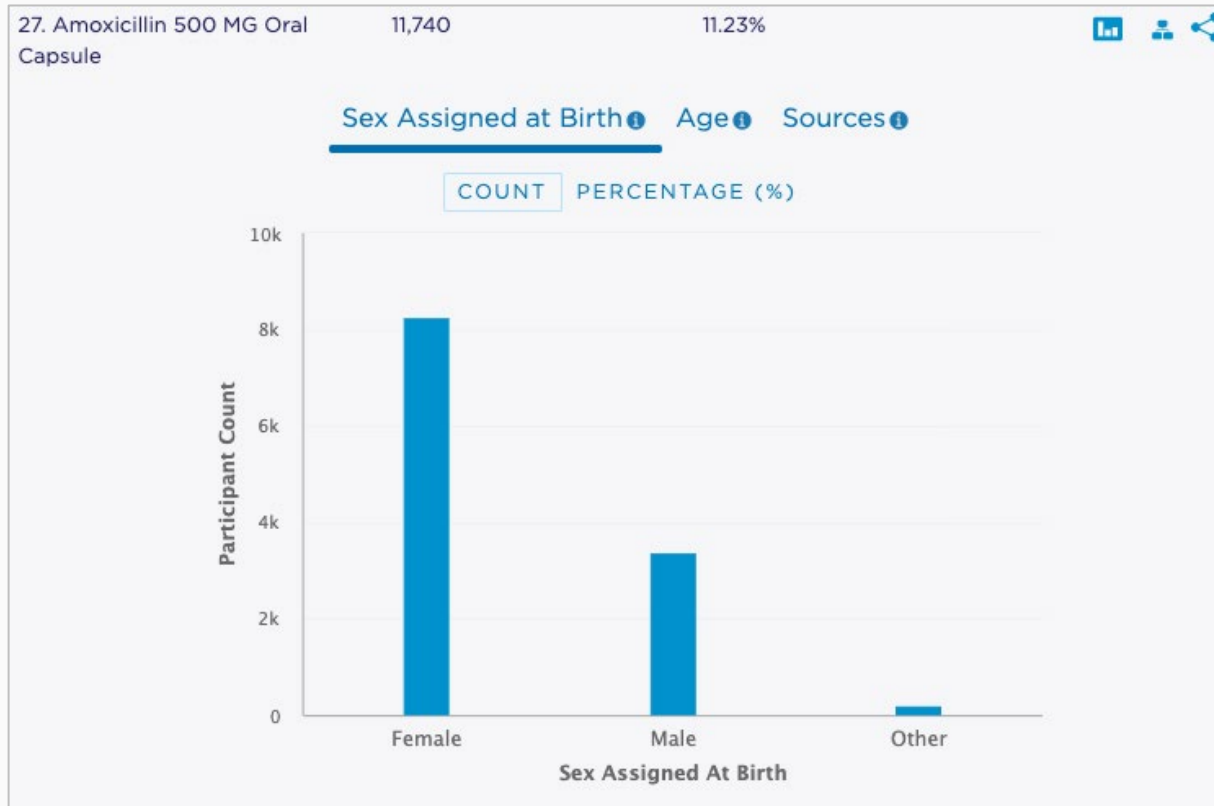
Example: COVID-19 Participant Experience (COPE) Survey

Total unique participants

62,920
Participants completed this survey

151
Questions Available

Available data gives insight into the participant cohort & research opportunities.

Survey versions

Month	Participants	Number of Questions	Download PDF
May	44080	148	Survey as PDF
June	33700	149	Survey as PDF
July/August	29920	148	Survey as PDF

Download different versions of the survey as PDFs.

In the past month, have recommendations for socially distancing caused stress for you?

See Answers ⌄

Hover mouse over charts & information icons for details & explanations.

Response of "A little" to the question, "In the past month, have recommendations for socially distancing caused stress for you?" in June:
13,200 Participants
39% of all participants who took this version of the survey
33,740 Total

Legend:
- A little
- Somewhat
- Not at all
- A lot
- Skip

databrowser.researchallofus.org/survey/covid-19-participant-experience

Powerful analytical tools & user support

WORKSPACES

USES:
- Organizing research projects
- Collaboration
- Transparency

NOTEBOOKS

USES:
- Analysis
- Queries

Uses programming languages R and Python

DATASET BUILDER

USES:
- Pre-populated analysis
- Dataset previews
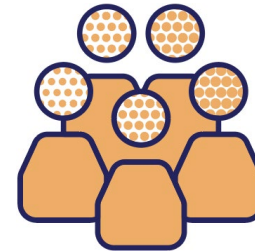
COHORT BUILDER

USES:
- Cohort creation

WORKBENCH USER SUPPORT HUB

USES:
- Learning/Training
- Personalized support
- Guides
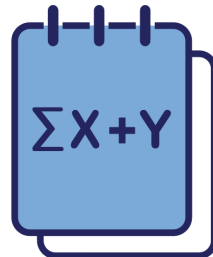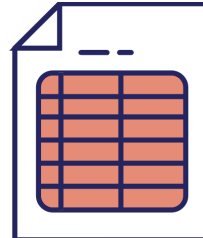
Advantages

- **POWERFUL ANALYTIC CAPABILITIES -** Advanced computing potential enables the simultaneous exploration of multiple questions on a powerful scale supported by R and Python, the most widely used open-source platforms for statistical analysis, as well as HAIL and Plink, both commonly used for genomic analysis.

- **TEAM SCIENCE APPROACH -** The Researcher Workbench provides the data, tools, and opportunity to come together with other researchers from different disciplines around project-specific data analysis (through collaborative workspaces, cohort-building tools, interactive notebooks, and more).

- **AVAILABILITY OF SAMPLE COHORTS** - For replicability and to help researchers get started
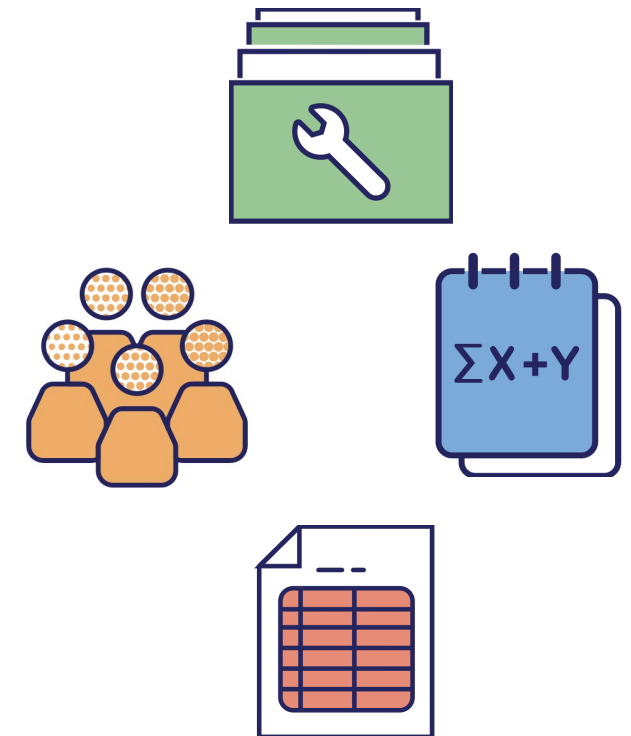
Limitations

- **LEARNING CURVE WITH DATA ANALYSIS TOOLS -** Data analysis requires proficiency in Jupyter notebooks and ability to code in R or Python as well as HAIL and Plink. Some institutions and teams may come to the platform without a background in these tools.

- **CAPACITY** - Smaller research teams may find that capacity constraints (e.g., cost, proficiency in coding tools, dependency on data scientists and/or partnerships) impact their ability to make full use of the *All of Us* platform without collaborators.

- **ADDITIONAL COSTS ASSOCIATED WITH ANALYSIS OF GENOMIC DATA** - While there is no cost to access the Researcher Workbench, there are computational costs for analyses incurred through Google Cloud Platform. The program provides $300 in initial credits to registered researchers to get them started.

Engaging Diverse Researchers

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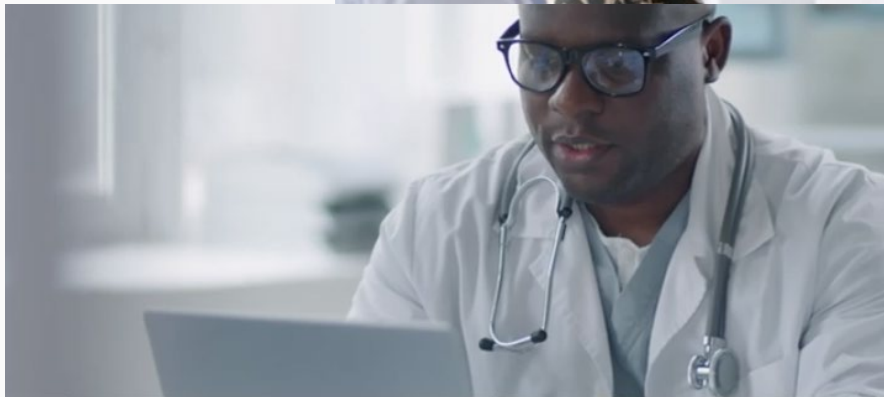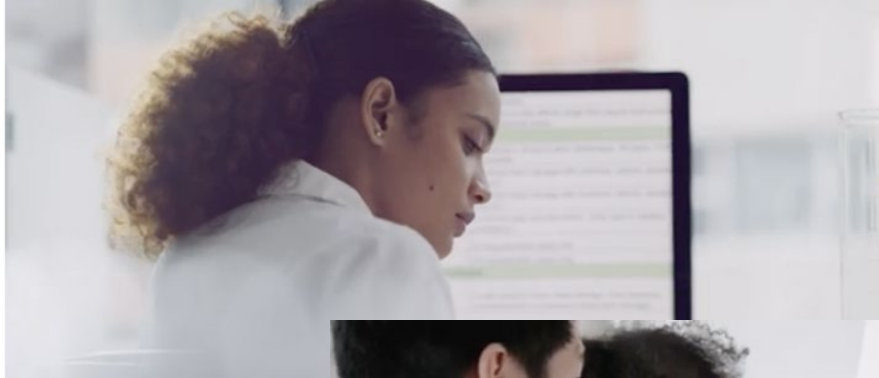
- **Creating a demographically diverse researcher cohort** that promotes responsible and ethical use of data, returns value to participant communities, and accelerates research impact.

- **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.

Explore how other researchers are using the data
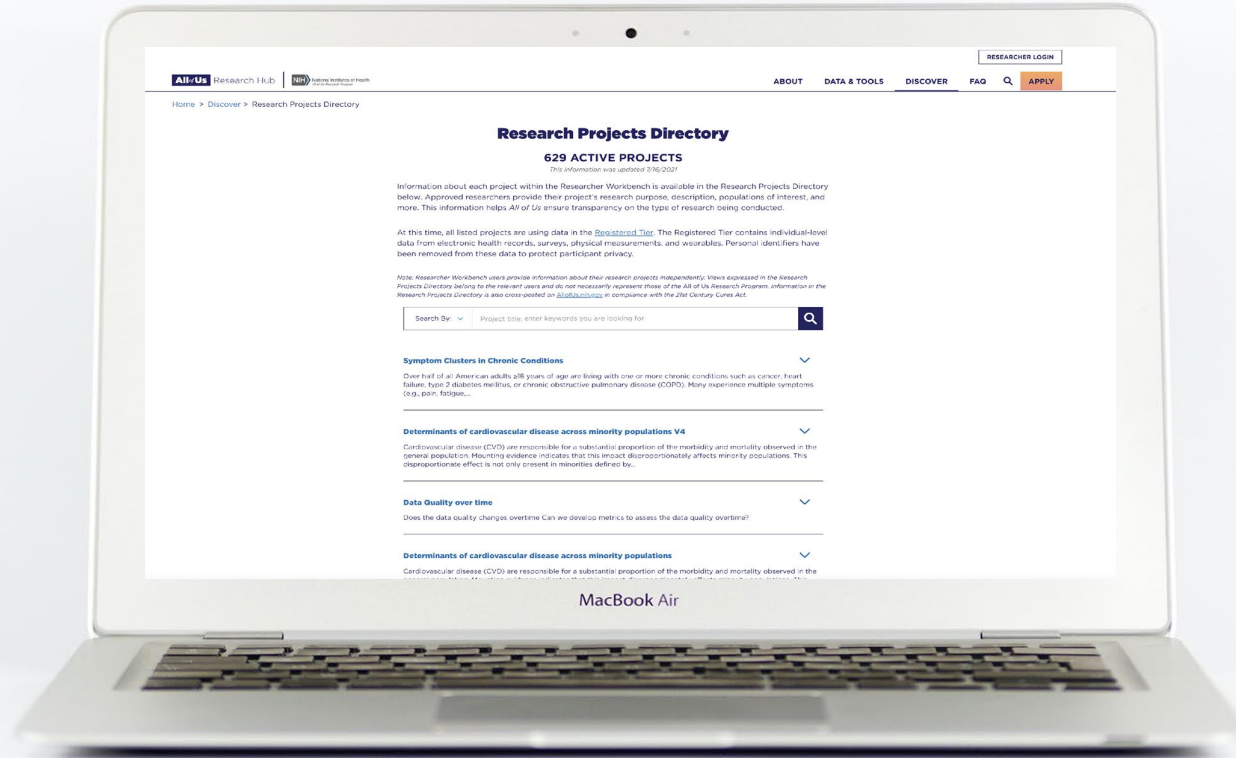
- Conducting **prospective, retrospective, and cross-sectional analyses**

- Developing **improved risk assessment** and prevention strategies to preempt disease

- Investigating **health disparities** and finding new approaches to improve health equity

- Providing **earlier and more accurate diagnoses** to reduce illness burden

- Increasing **wellness and resilience** and promoting healthy living

- Creating a line of sight to enable new **precision treatments and interventions**

researchallofus.org/explore

Building a Demographically Diverse Researcher Community

- **Inviting feedback from researchers** across all settings and career-stages

- **Applying recommendations from a diverse steering committee** including experts from various backgrounds and communities

- **Exploring opportunities to support team science** through a collaborative and accessible cloud-based platform

- **Seeking new ways to enhance accessibility and democratize access** including working to put data use agreements in place with a wide range of organizations (e.g. HBCUs, MSIs, RM1s) and encouraging students and early-stage investigators to bring fresh, creative perspectives to the dataset

- **Promoting responsible and ethical use of data,** including returning value to participant communities

The *All of Us* Researcher Workbench is an iterative platform. **Your input will help shape it.**

47

Projects

Testimonials from registered researchers

Predictive analytics for glaucoma
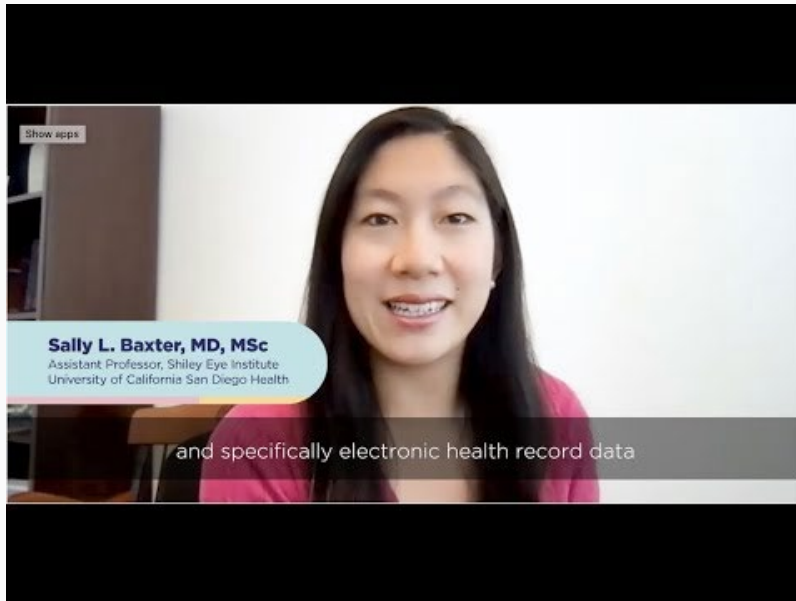
Sally Baxter, MD, MSc
Assistant Professor, Shiley Eye Institute
University of California San Diego Health

How diverse data can power more impactful research

Jason Karnes, PharmD, PhD, BCPS, FAHA
Director of Scientific Programs
University of Arizona/Banner Health

Studying health care access and utilization among adult cancer survivors

Jie Chen, PhD
Professor and Division Chief,
Biostatistics and Data Science
Medical College of Georgia, Augusta University

**Ancillary Studies
&
Researcher Opportunities**

Nutrition for Precision Health, powered by the *All of Us* Research Program

Goal: to develop predictive algorithms to inform targeted dietary approaches

FOA publication	First awards	First participant enrolled	Predictive algorithms
January 2021	Early 2022	Early 2023	FY 2027

1 To describe individual responses to "usual diets"

10,000 *All of Us* participants
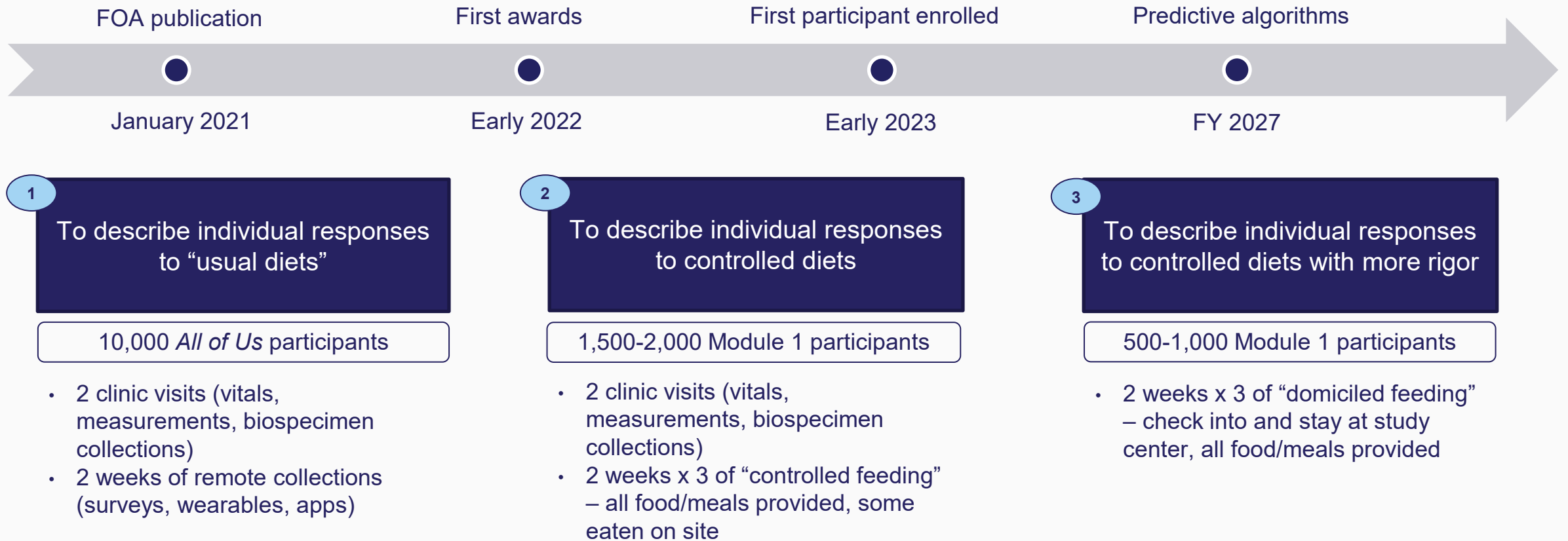
- 2 clinic visits (vitals, measurements, biospecimen collections)
- 2 weeks of remote collections (surveys, wearables, apps)

2 To describe individual responses to controlled diets

1,500-2,000 Module 1 participants

- 2 clinic visits (vitals, measurements, biospecimen collections)
- 2 weeks x 3 of "controlled feeding" – all food/meals provided, some eaten on site

3 To describe individual responses to controlled diets with more rigor

500-1,000 Module 1 participants

- 2 weeks x 3 of "domiciled feeding" – check into and stay at study center, all food/meals provided

CommonFund.NIH.gov/NutritionforPrecisionHealth

Join *All of Us*!

Apply to be an *All of Us* Researcher

STEP 1
LEARN MORE ABOUT THE DATA AVAILABLE

STEP 2
CHECK FOR YOUR INSTITUTION'S AGREEMENT

STEP 3
REGISTER AS A RESEARCHER

STEP 4
VERIFY YOUR IDENTITY

STEP 5
COMPLETE *ALL OF US* RESPONSIBLE CONDUCT OF RESEARCH TRAINING

STEP 6
SIGN DATA USER CODE OF CONDUCT

APPLY NOW

Stay in touch to learn more

SUBSCRIBE
to our quarterly email newsletter and receive the latest news and insights from the *All of Us* Research Hub.

All of Us Researchers Convention

A broad spectrum of researchers using *All of Us* data and tools offered through the Research Hub.

Two-day virtual event
- Minority Student Research Symposium (March 31st)
- Science Day (April 1st)

Summary, the *All of Us* Research Program

HOW
do I access it?

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

WHO
benefits?

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

WHY
it matters?

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

WHAT
is the *All of Us* Research Program?

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

Conclusions

o Assembling the largest, richest biomedical dataset of its kind with plans for genomic data and comprehensive clinical information

o Successfully engaging participants as partners

o New data types continuing to be rolled out

o Ancillary studies of specific populations or disease states possible

o Nurse researcher opportunities available

o Will facilitate research across multiple disease states and could serve as controls for disease-specific studies that lack healthy controls

Questions