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.
Utilizing the *All of Us* database: Research Opportunities & Impact

April 20, 2022

Monica Rodriguez, MS
Pyxis Partners
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 \textit{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

- 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.

<table>
<thead>
<tr>
<th>Initial Data Types</th>
<th>Expanded Data Types (May Include)</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Demographics</td>
<td>• Clinical notes</td>
</tr>
<tr>
<td>• Visits</td>
<td>• Radiology, cardiology, and other reports</td>
</tr>
<tr>
<td>• Diagnoses</td>
<td>• Mental health reports</td>
</tr>
<tr>
<td>• Procedures</td>
<td>• Substance abuse, alcohol use, and tobacco use</td>
</tr>
<tr>
<td>• Medications</td>
<td>• More laboratory results, potentially including genomics</td>
</tr>
<tr>
<td>• Laboratory visits</td>
<td></td>
</tr>
<tr>
<td>• Vital signs</td>
<td></td>
</tr>
</tbody>
</table>

Participant Surveys
Physical Measurements
Biosamples
Mobile/Wearable Tech
Participant Surveys

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
Current Measurements:
- Blood pressure
- Heart rate
- Height
- Weight
- BMI
- Hip circumference
- Waist circumference
Biosamples

Examples:
- Blood
- Saliva
- Urine

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

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 lactose 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.

<table>
<thead>
<tr>
<th>DNA marker</th>
<th>Gene</th>
<th>Your result</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4988235</td>
<td>MCM6</td>
<td>0/0</td>
</tr>
</tbody>
</table>

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

Simple Traits

- Lactose Intolerance
- Asparagus odor detection
- Earwax type
- Bitter taste perception
- Cilantro preference
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?

**Wellness & Resilience**
What is the role of social connectivity in the ability to adhere to management of chronic infectious diseases, including HIV?

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

**Treatment & Outcomes**
What are the long-term outcomes of HIV pre-exposure prophylaxis on individuals without 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?

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

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

**Treatment & Outcomes**
Why do some opioid-naive surgical patients exposed to opioids after surgery become persistent opioid users, while others do not?
Research Hub
All of Us Research: How it Works

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

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

*Includes 36 Historically Black Colleges and Universities and Hispanic-Serving Institutions.

Figures accurate as of March 2022
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

**Within the Controlled Tier:**

- **98,600+** Whole Genome Sequences
- **165,000+** Genotyping Arrays
- **593,500,000+** Unique Variants

**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

**Public Tier**

Anyone can visit [ResearchAllofUs.org](http://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**

**Registered Tier**

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

**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)

*Data have been processed to protect participant privacy*
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.
Selected Data Snapshots

(Updated 9/21/21)

**Gender Identity**

![Chart showing gender identity distribution]

- **Man**: 39.0%
- **Woman**: 59.5%
- **Nonbinary**: 0.3%
- **Transgender**: 0.2%
- **Additional Options**: 0.2%
- **Prefer Not to Say**: 0.1%
Selected Data Snapshots

(Updated 9/21/21)

Race & Ethnicity

- White: 47.4%
- Black, African American or African: 21.3%
- Hispanic Latino or Spanish: 17.0%
- More than one race/ethnicity: 6.5%
- Other: 2.9%
- Prefer not to say: 0.7%
Selected Data Snapshots

(Updated 9/21/21)

**Age**

<table>
<thead>
<tr>
<th>Age</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>18-29</td>
<td>10.6%</td>
</tr>
<tr>
<td>30-39</td>
<td>15.1%</td>
</tr>
<tr>
<td>40-49</td>
<td>14.4%</td>
</tr>
<tr>
<td>50-59</td>
<td>19.8%</td>
</tr>
<tr>
<td>60-69</td>
<td>21.5%</td>
</tr>
<tr>
<td>70-79</td>
<td>14.2%</td>
</tr>
<tr>
<td>80-89</td>
<td>4.0%</td>
</tr>
<tr>
<td>89+</td>
<td>0.5%</td>
</tr>
</tbody>
</table>

**Geography**

All 50 states

- Bilingual enrollment
- Interactive mobile exhibits
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.

Hover mouse over charts and information icons for details and explanations.
A Quick Look at the Data Browser

Researchers can view the full surveys, including branching logic.

<table>
<thead>
<tr>
<th>Question 1</th>
<th>Participant Count</th>
<th>% Answered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have you smoked at least 100 cigarettes in your entire life? (There are 20 cigarettes in a pack.)</td>
<td>1585859</td>
<td>59,140</td>
</tr>
<tr>
<td>Yes</td>
<td>1585856</td>
<td>39,060</td>
</tr>
<tr>
<td>L3. Do you now smoke cigarettes every day, some days, or not at all?</td>
<td>1585863</td>
<td>22,060</td>
</tr>
<tr>
<td>No</td>
<td>1585861</td>
<td>11,420</td>
</tr>
<tr>
<td>Every day</td>
<td>1585862</td>
<td>5,560</td>
</tr>
<tr>
<td>Some days</td>
<td>903079</td>
<td>1,200</td>
</tr>
<tr>
<td>Prefer Not To Answer</td>
<td>903096</td>
<td>340</td>
</tr>
<tr>
<td>Skip</td>
<td>903096</td>
<td>340</td>
</tr>
</tbody>
</table>
A Quick Look at the Data Browser

Explore data breakdowns by sex assigned at birth and age.

Sex Assigned at Birth  Age  Sources

COUNT  PERCENTAGE (%)

Participant Count

Sex Assigned At Birth

Female  Male  Other

27. Amoxicillin 500 MG Oral Capsule 11,740 11.23%

COUNT  PERCENTAGE (%)

Participant Count

Age At First Occurrence In Participant Record

18-29  30-39  40-49  50-59  60-69  70-79  80-89  89+

11,740  11.23%
Example: COVID-19 Participant Experience (COPE) Survey

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

Total unique participants

62,920
Participants completed this survey

151
Questions Available

Survey versions

<table>
<thead>
<tr>
<th>Month</th>
<th>Participants</th>
<th>Number of Questions</th>
<th>Download PDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>May</td>
<td>44080</td>
<td>148</td>
<td>Download as PDF</td>
</tr>
<tr>
<td>June</td>
<td>33700</td>
<td>149</td>
<td>Download as PDF</td>
</tr>
<tr>
<td>July/August</td>
<td>29920</td>
<td>148</td>
<td>Download as PDF</td>
</tr>
</tbody>
</table>

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

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

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.
Projects
Testimonials from registered researchers

Predictive analytics for glaucoma

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

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

Studying health care access and utilization among adult cancer survivors

Jason Karnes, PharmD, PhD, BCPS, FAHA
Director of Scientific Programs
University of Arizona/Banner Health
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: January 2021
- First awards: Early 2022
- First participant enrolled: Early 2023
- Predictive algorithms: 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

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

Stay in touch to learn more.
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

**WHAT**

is the *All of Us* Research Program?

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

**WHY**

it matters?

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

**WHO**

benefits?

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

**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.
Conclusions

- Assembling the largest, richest biomedical dataset of its kind with plans for genomic data and comprehensive clinical information
- Successfully engaging participants as partners
- New data types continuing to be rolled out
- Ancillary studies of specific populations or disease states possible
- Nurse researcher opportunities available
- Will facilitate research across multiple disease states and could serve as controls for disease-specific studies that lack healthy controls
Thank You!

researchallofus.org

AllofUs.nih.gov

@AllofUsResearch
#JoinAllofUs
Questions