Digital data collection:

The *All of Us* Research Project perspective

National Institutes of Health

Chris Lunt

Chief Technology Officer

March 24, 2020

The *All of Us* Research Program: An Innovative Research Effort

Diversity at the scale of **1 million people** or more (250K+ joined, 80% underrepresented)

Focus on participants as **partners**

Longitudinal design; ability to recontact participants

Multiple data types: EHR, surveys, baseline physical measurements, biospecimens, genomics, and more

National, open resource for all: broadly accessible to all researchers with open source software & tools

Hypothesis-neutral

All of Us
RESEARCH PROGRAM

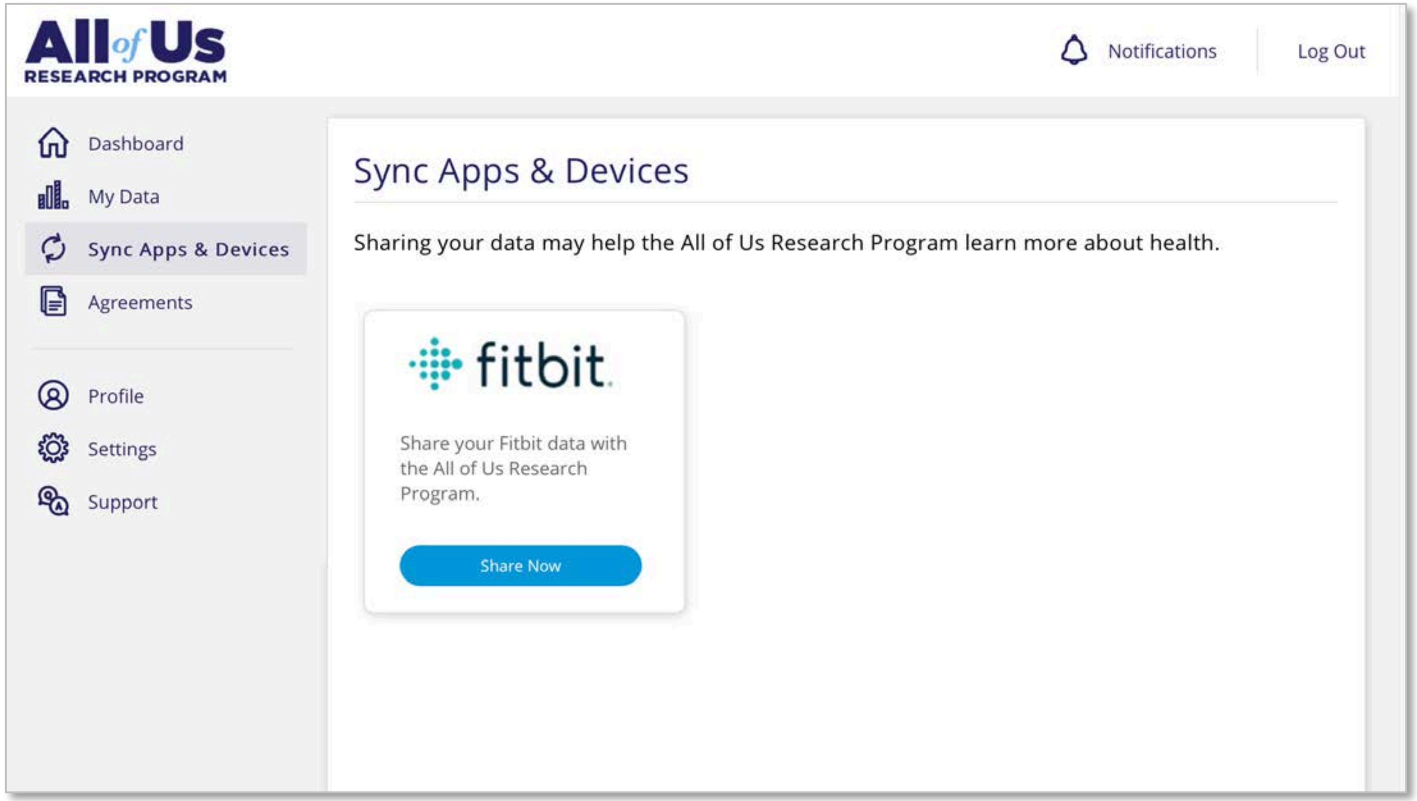
Definition

Digital Health Technology: devices that can be used to collect health measures outside of the clinic, both actively and passively, including phones, laptops, wearables, portable, and in-home devices.

Wearables and Digital Apps

- “Bring your own device” (BYOD) program:
  - Fitbit (now)
  - Apple HealthKit (in March)

- Pilots of specific smartphone-based apps
  - Mood app
  - Others in development



Sync Apps & Devices

Sharing your data may help the All of Us Research Program learn more about health.

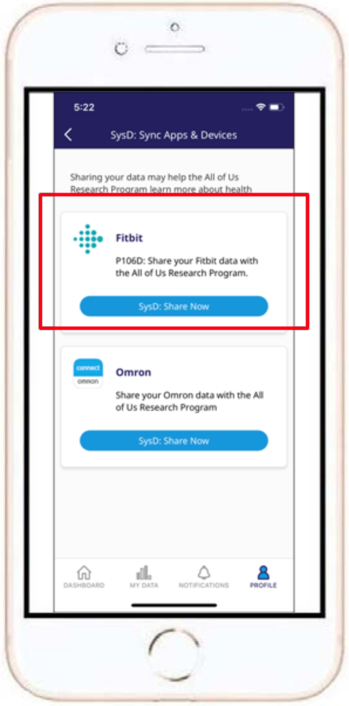
Share your Fitbit data with the All of Us Research Program.

Value of DHT data

- Pro: information from outside the clinic
- Pro: longitudinal, intensive, and repeated
- Pro: sensor data not subject to self-reporting biases
- Pro: can be passive (low participant burden)
- Pro: builds on existing and growing infrastructure we don’t pay for
- Con: narrow, and generally deductive: building on existing hypotheses (“steps matter”)
- Con: relying on external infrastructure (mobile phones) skews the data collected
- Con: security risk, some require engagement with external partners
- Con: some control goes to the technology provider

Fitbit Flow

Fitbit Screens - *Not Modifiable by AoURP*

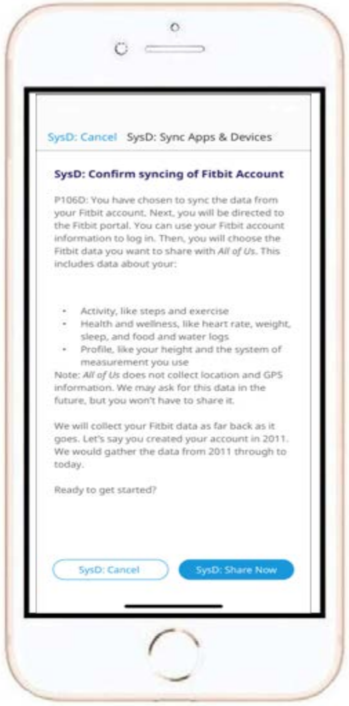


Sharing your data may help the All of Us Research Program learn more about health

P106D: Share your Fitbit data with the All of Us Research Program.

Share your Omron data with the All of Us Research Program

Participant requests to share their data



SysD: Confirm syncing of Fitbit Account

P106D: You have chosen to sync the data from your Fitbit account. Next, you will be directed to the Fitbit portal. You can use your Fitbit account information to log in. Then, you will choose the Fitbit data you want to share with *All of Us*. This includes data about your:

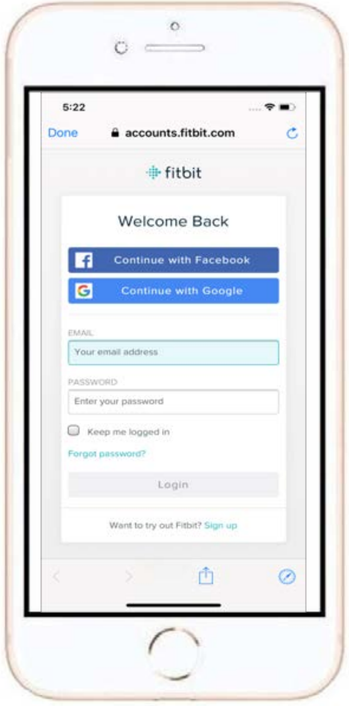
- Activity, like steps and exercise
- Health and wellness, like heart rate, weight, sleep, and food and water logs
- Profile, like your height and the system of measurement you use

Note: *All of Us* does not collect location and GPS information. We may ask for this data in the future, but you won't have to share it.

We will collect your Fitbit data as far back as it goes. Let's say you created your account in 2011. We would gather the data from 2011 through to today.

Ready to get started?

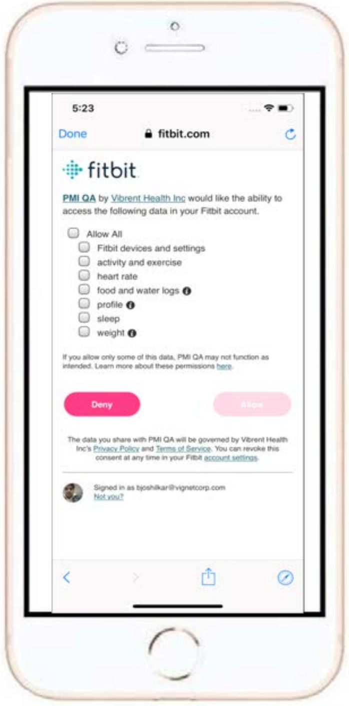
Participant reads informative screen on system screen they are about to see, which they consent to



☐ Keep me logged in

Want to try out Fitbit? Sign up

Browser screen redirects to Fitbit login page



PMI QA by Vibrent Health Inc would like the ability to access the following data in your Fitbit account.

☐ Allow All
  ☐ Fitbit devices and settings
  ☐ activity and exercise
  ☐ heart rate
  ☐ food and water logs
  ☐ profile
  ☐ sleep
  ☐ weight

If you allow only some of this data, PMI QA may not function as intended. Learn more about these permissions here.

The data you share with PMI QA will be governed by Vibrent Health Inc's Privacy Policy and Terms of Service. You can revoke this consent at any time in your Fitbit account settings.

Signed in as bjoshilkar@vignetcorp.com
Not you?

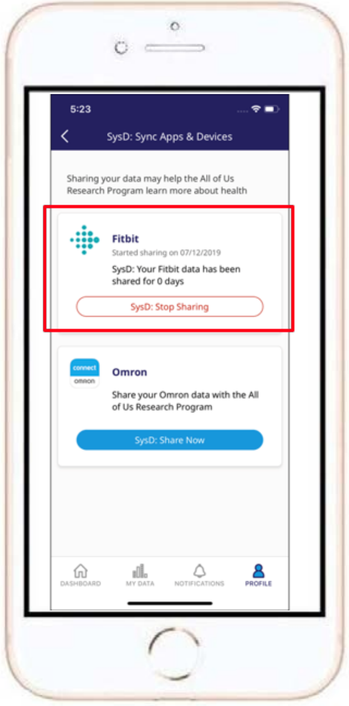
Browser screen from Fitbit indicates App is requesting permission to access (read only)

syncDevices.devices.fitbit.successMessage.alt.text

P106D: Thank you! You are now sharing your Fitbit data with the All of Us Research Program

SysD: Please close this screen to continue.

Browser screen showing success and close option



Sharing your data may help the All of Us Research Program learn more about health

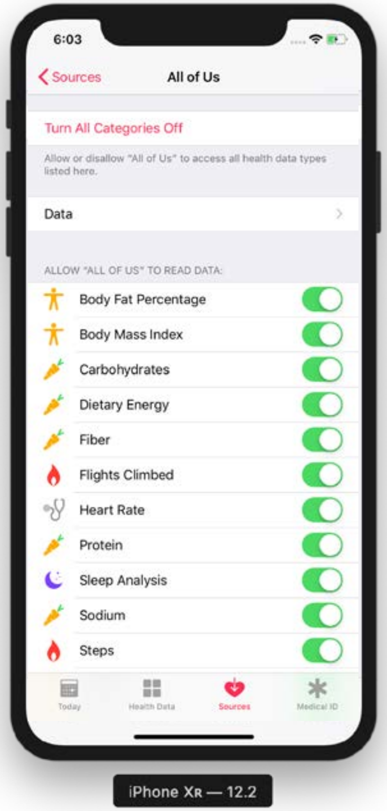
Started sharing on 07/12/2019

SysD: Your Fitbit data has been shared for 0 days

Share your Omron data with the All of Us Research Program

Information is being shared and participant can stop sharing at any time

The participant controls what they’d like to share



All of Us

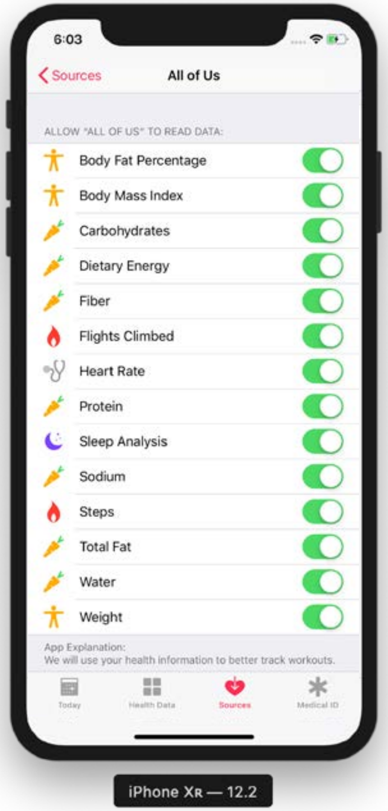
Turn All Categories Off

Allow or disallow "All of Us" to access all health data types listed here.

Data

ALLOW "ALL OF US" TO READ DATA:

Body Fat Percentage
Body Mass Index
Carbohydrates
Dietary Energy
Fiber
Flights Climbed
Heart Rate
Protein
Sleep Analysis
Sodium
Steps



All of Us

ALLOW "ALL OF US" TO READ DATA:

Body Fat Percentage
Body Mass Index
Carbohydrates
Dietary Energy
Fiber
Flights Climbed
Heart Rate
Protein
Sleep Analysis
Sodium
Steps
Total Fat
Water
Weight

App Explanation:
We will use your health information to better track workouts.

All of Us

No data found

Participant can modify what data can be read by All of Us Research Program in Apple Health/HealthKit

Apple Health showing that data is not being written to by *AoU*RP

Prioritization Criteria

Why choose a particular assessment or device strategy?

- **Science**: helps advance the scientific agenda of the program
- **Recruitment**: lets us tap into the existing audience (e.g. Apple)
- **Engagement**: helps keep participants interested and feeling valued
- **Partnership**: helps keep program partners interested and feeling valued
- **Cost**: monetary, program attention, participant burden (time and know-how)
- **Logistics:** accuracy, availability by location, and more

*details at the end of this

Criteria: Science (for example)

What are the considerations for scientific value of a datatype?

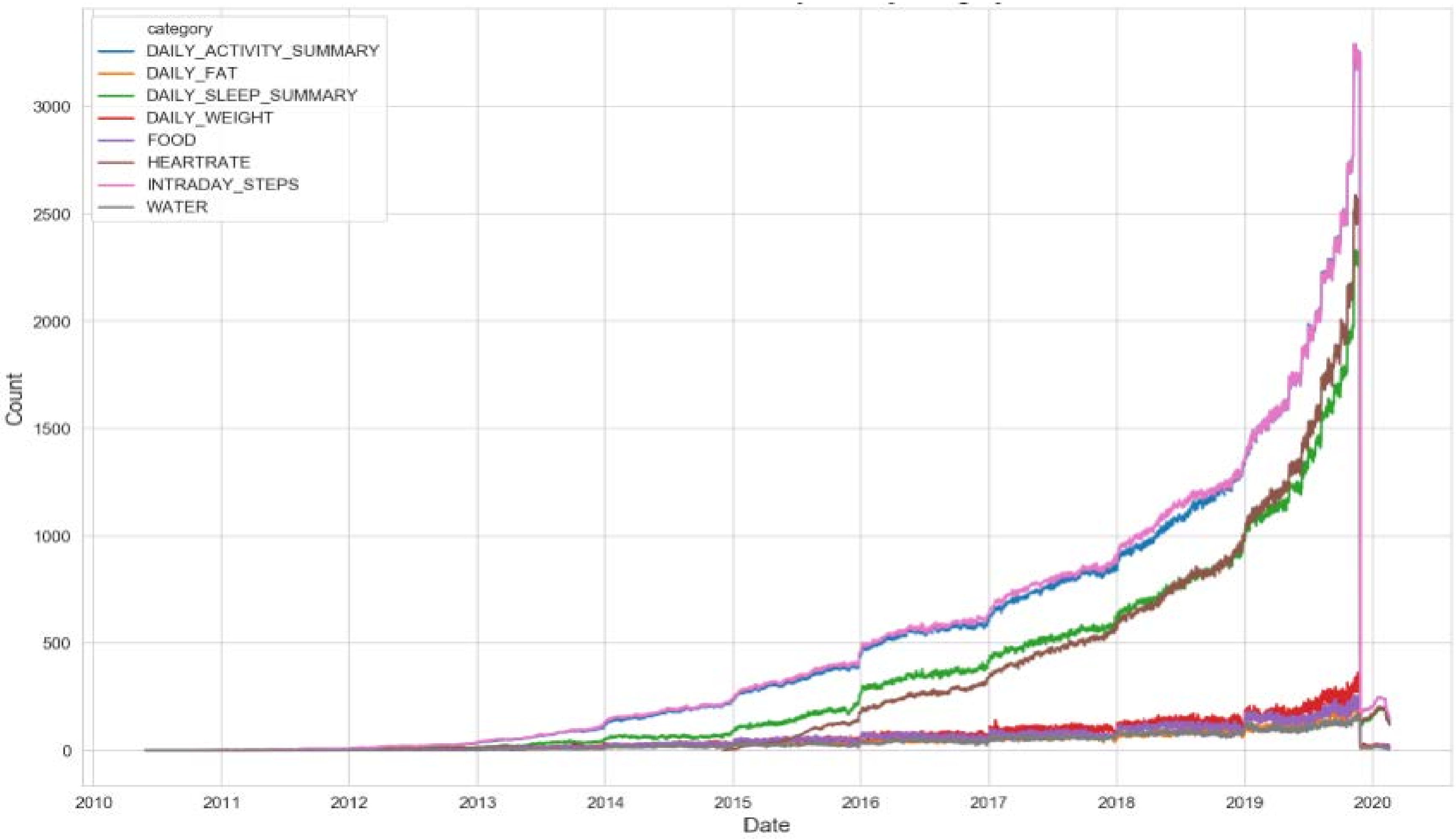
- **Novel**: is this a new type of data? (hopefully with some prior evidence)
- **Context**: does little of this data exists from outside the clinic? (e.g. blood pressure)
- **Audience**: is there little of this data for UBR populations?
- **Volume**: has this data been captured at high frequency or longitudinally?
- **Association**: has this data been captured in conjunction with genetics et al?
- **External Validation**:

Bring Your Own Device (BYOD)

- Pro: immediate engagement
- Pro: lower cost
- Pro: people show up with existing data
- Con: self-selection skews data
- Con: differing devices skews data
- Con: limited audience
- Con: full curation cost for limited data

- ~12% people have wearables
- 30% in some segments
- 77% have smartphones
- Smartphones primary device for many SES families

AoU Fitbit data over time by data element



Amount of data for each Fitbit data element over time – for the first set of data received from DRC in Feb 2020 (~30% of total Fitbit data the program has access to) – represents data from 4656 participants – most of the data is from steps and HR.

Core strategy

1. **Drive** DHT strategy from overall Science strategy **and** from engagement needs
2. Develop a long-term, cross-component **pipeline**
3. **Balance** opportunity-taking and thesis-driven selection
4. Use **off-the-shelf** consumer technology (no study watch)
5. Start with **BYOD** for earlier movement and lower learning cost

For more information…

ResearchAllofUs.org

@AllofUsResearch
#JoinAllofUs

AllofUs.nih.gov

databrowser.researchallofus.org

Appendix

Criteria: recruitment

Can we tap into an existing audience?

- **Size**: how big is the audience we may be able to convert?
  The Stanford Apple Watch Heart study has over 400,000 participants in less than 6 months.
- **Composition**: will this audience dilute our UBR population?
  The Stanford Apple Watch Heart study has an audience that matches the Apple audience: high literacy and high income. But we can use selection criteria to prioritize the core participants.
- **Partner willingness**: here be lawyers

Criteria: Engagement (value to participants)

- **Learning about one's self**: Both clinically and from an identity perspective
- **Altruism**: contributing to the greater good
- **Belonging**: Feeling of being part of an effort larger than themselves by seeing others doing the same thing
- **Discipline**: Doing what one says one will do, especially if it's difficult and seen as virtuous (see jogging)
- **Learning about science**: Health, medicine, and research
- **Attention and Stuff**: You gave me your time and a thing
- **Fun**: Like "brain games"

Criteria: Partnership

How do we find mutual value with program partners?

- **Seize windows of opportunity**: arbitrage, like getting free prior-generation wearables, or co-launching efforts (Apple Research Kit)
- **Value their contribution**: value the time they've given us, and their expertise (reaching and appealing to particular audiences)

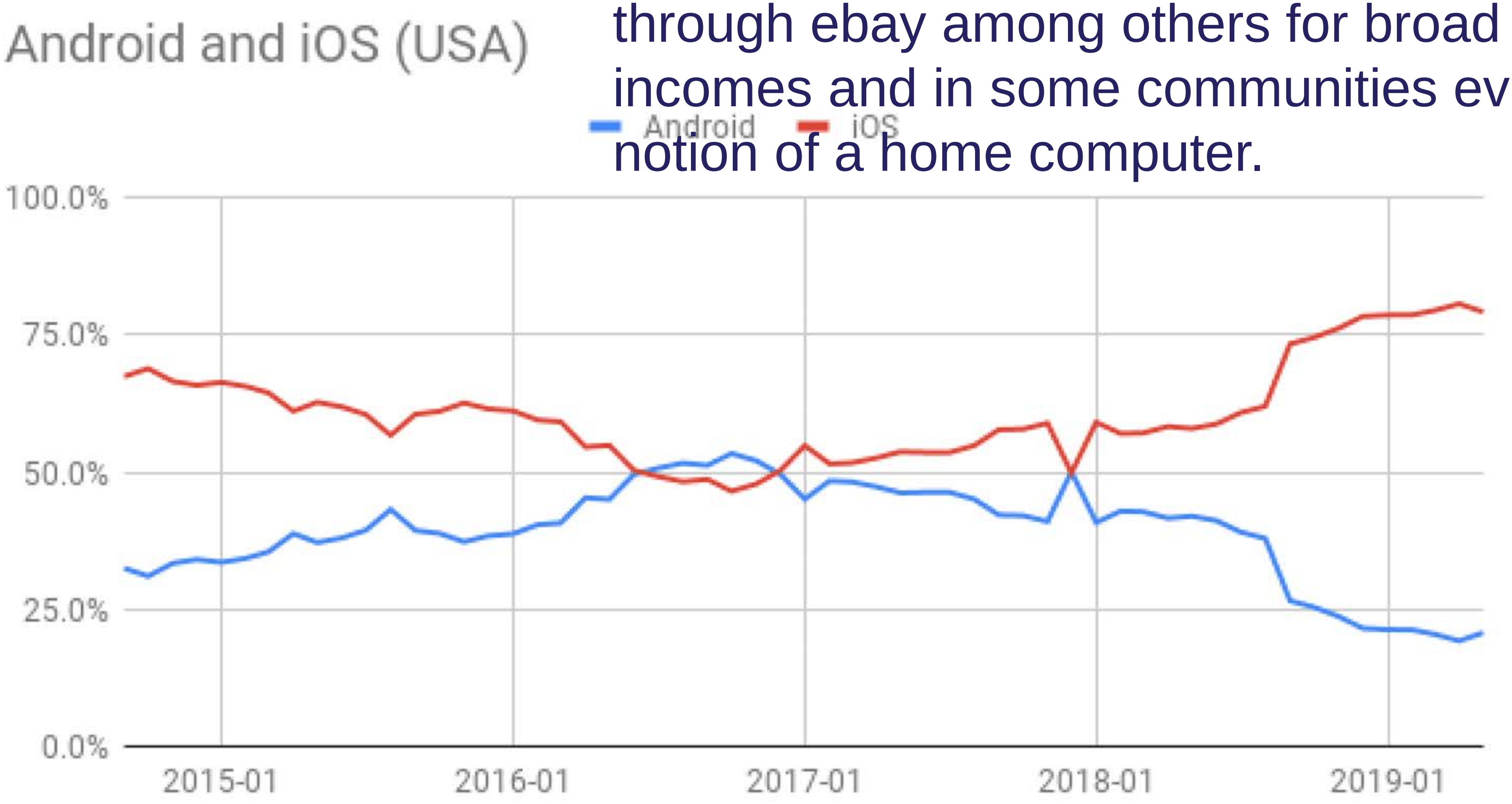
Criteria: Cost

- **Monetary**: What’s the cost to the program per participant
- **Program attention**: How much work to get it launched (privacy, security, IRB approval, FDA waiver, Protocol modification, etc)
- **Participant time**: How much time to provide the data?
- **Participant cognitive load**: How hard is it to do?
- **Bad press risk**: Are there health, security, or privacy risks?

Criteria: Logistics (practical stuff)

- **Accuracy**: is the data gathered accurate enough to be scientifically useful?
- **Availability**: where can this be done?
- **BYOD vs distributed**: Are we providing the means to give the data, or relying on people to already have it? (Keep in mind things like “donate my search history”)
- **Passive vs Active**: what is required of a participant to do this?
- **Supported vs Unsupported**: does this require the involvement of staff? For example, retina scans.
- **Selected vs Volunteered**: and how do we select them
- **Formats and Standards**

iOS in the US is becoming more dominant that Android due to its stability, security and ease of use in its technical design. This also supports the cost model of devices available through cellular service as well as re-sell through ebay among others for broad adoption across incomes and in some communities even replaces the notion of a home computer.



All of Us 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: Blood Pressure

Risk & Prevention

How do age-related changes in blood pressure in children and young adults impact the development of hypertension and hypertension-related conditions in adulthood?

Health Equity

What is the impact of economic stability on rates of screening, likelihood of receiving treatment, and blood pressure levels?

Diagnosis

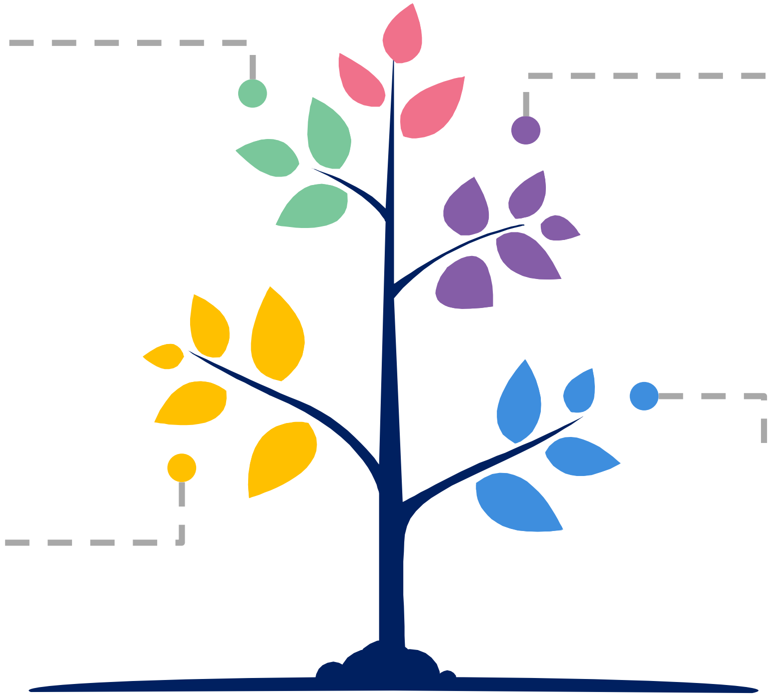
Does ambulatory blood pressure monitoring (ABPM) or home blood pressure monitoring (HBPM) provide a more accurate estimate of cardiovascular risk?

Wellness & Resilience

What genomic, environmental, and lifestyle factors underlie the different patterns in age-related trajectories of blood pressure, thereby increasing or reducing the risk of high blood pressure?

Treatment & Outcomes

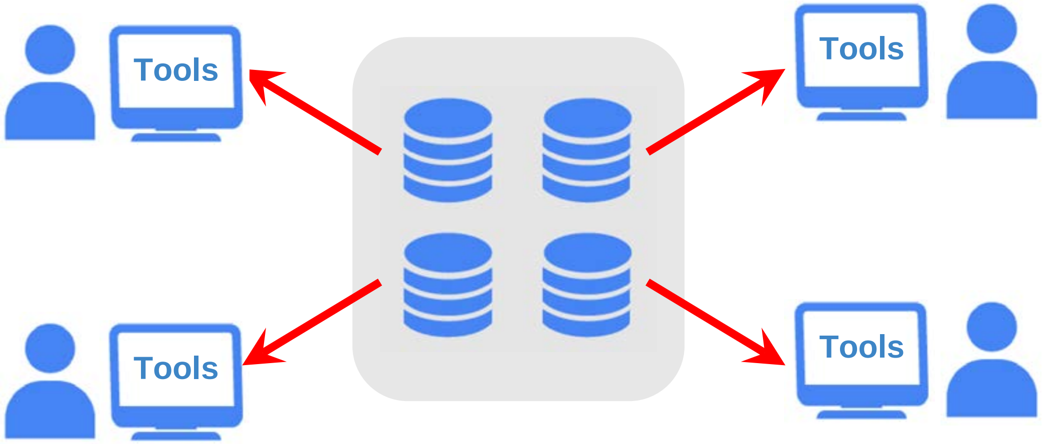
What are effective and scalable community-based interventions to improve blood pressure levels, medication prescription and medication fill rates?



BLOOD PRESSURE

Traditional approach

Bring data to researchers



Discourages shared research

“Weakest link” security

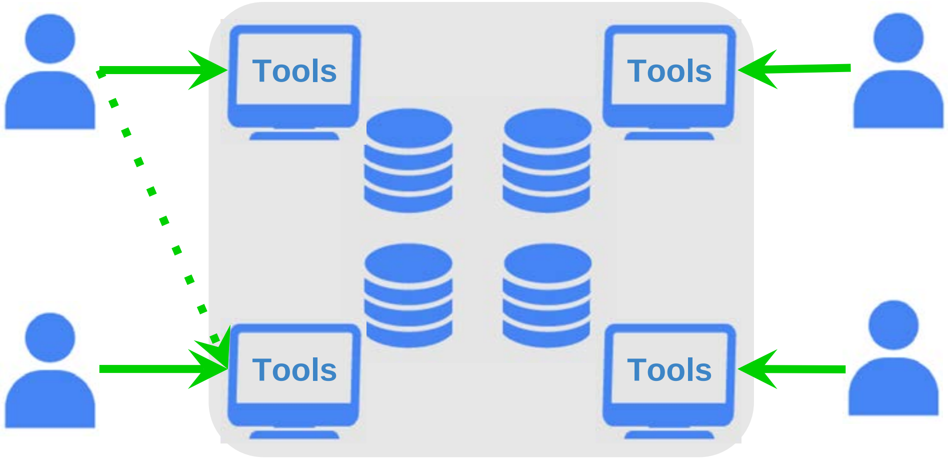
Huge infrastructure needed

Pay for multiple copies

Bespoke & unsupported tools

Cloud-centric approach

Bring researchers to data



Facilitates collaboration

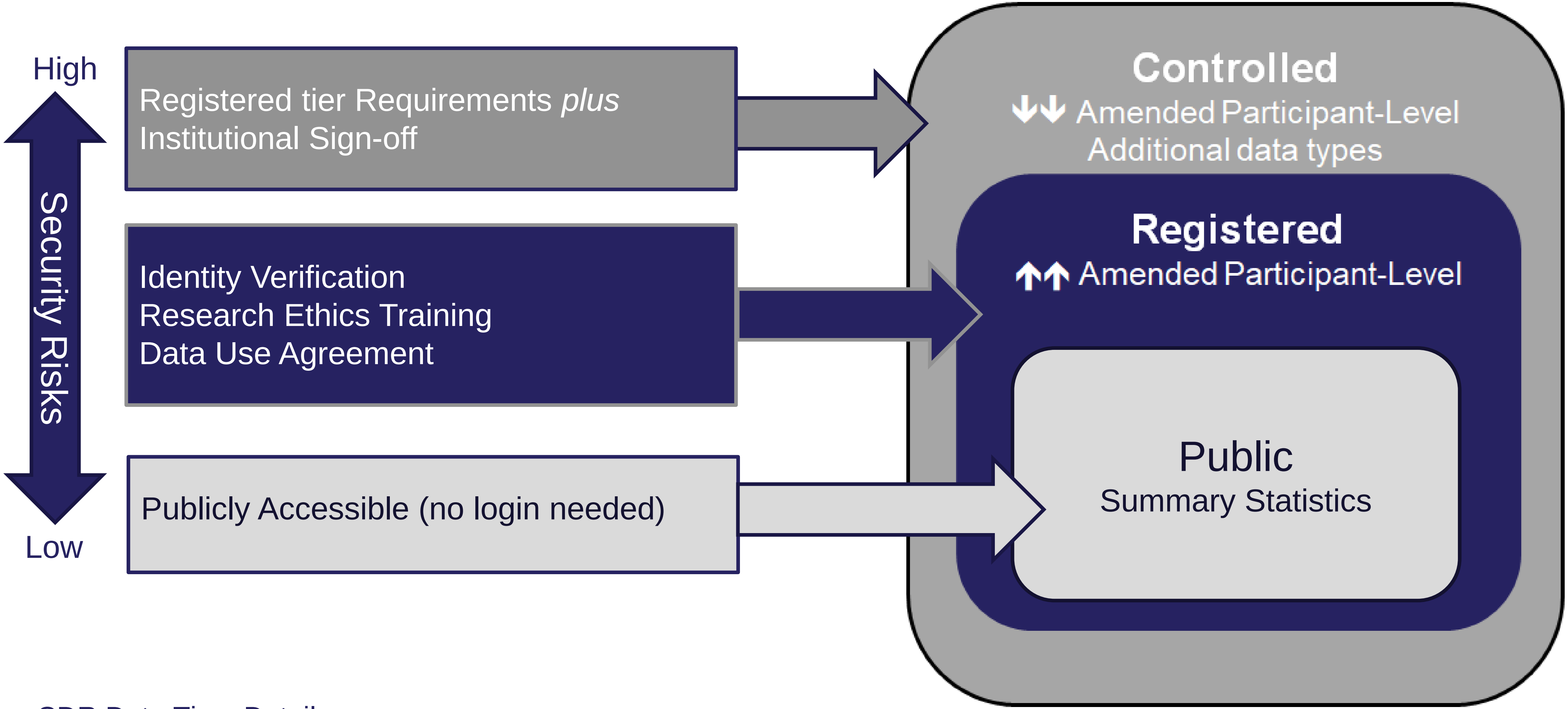
Centralized security controls

Accessible to all researchers

Decreased cost of storage

Shared tool ecosystem

All of Us Data Tiers



High

Security Risks

Low

Registered tier Requirements *plus*
Institutional Sign-off

Identity Verification
Research Ethics Training
Data Use Agreement

Publicly Accessible (no login needed)

Controlled
↓↓ Amended Participant-Level
Additional data types

Registered
↑↑ Amended Participant-Level

Public
Summary Statistics

CDR Data Tiers Details
Summary of Identity Protection Rules

Current Progress

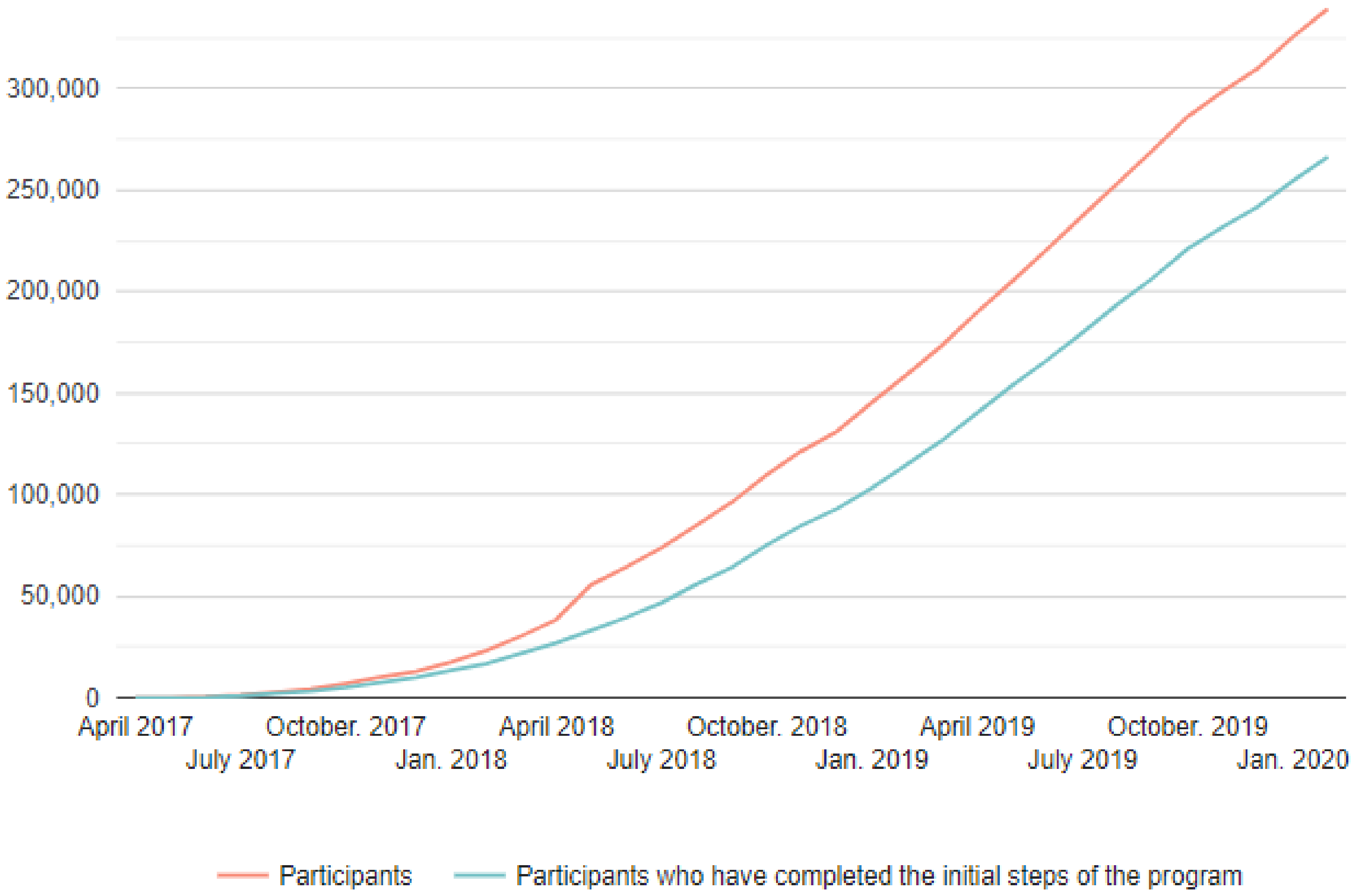
340,000+

Participants

266,000+

Participants who have completed initial steps of the program

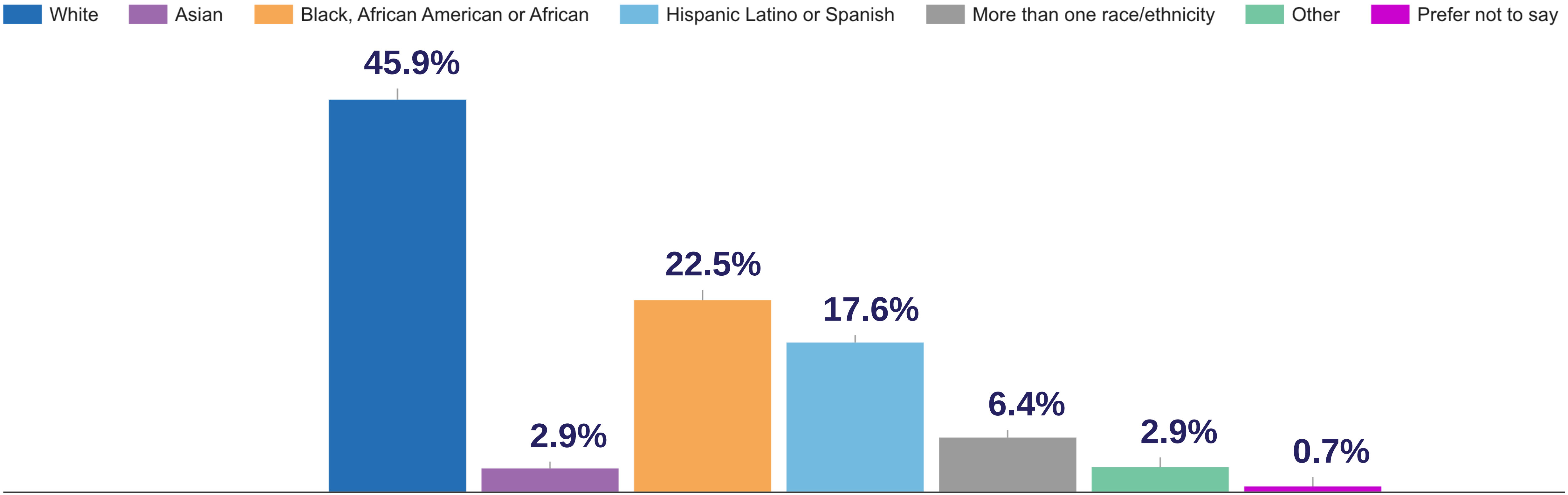
The following numbers are approximated to protect participants' privacy. Numbers are updated as of March 2, 2020.

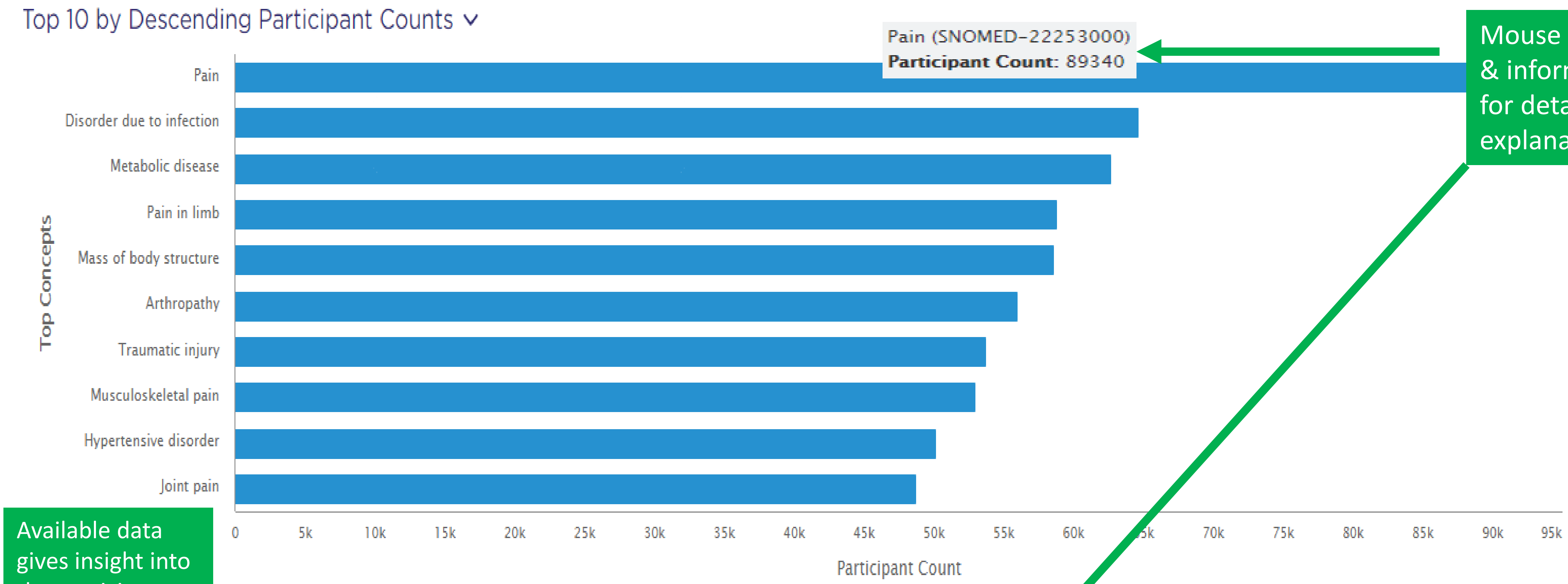


Selected Data Snapshots

(Updated 3/3/20)

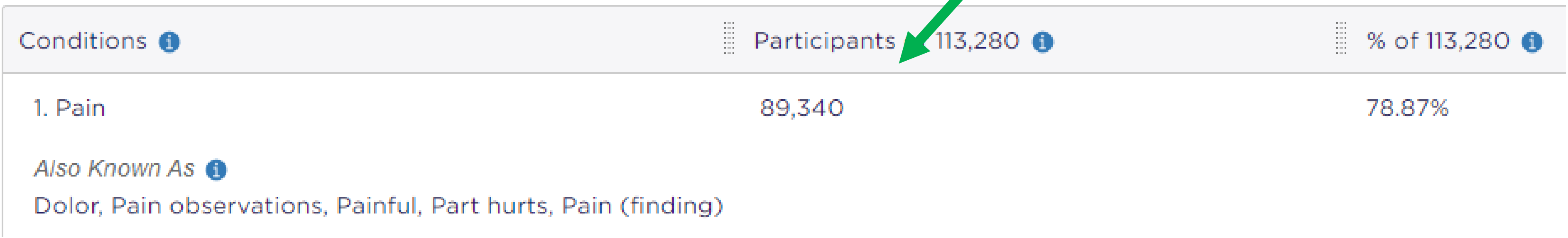
Race & Ethnicity





Mouse over charts & information icons for details & explanations.

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



Conditions ℹ	Participants 113,280 ℹ	% of 113,280 ℹ
1. Pain	89,340	78.87%

Also Known As ℹ

Dolor, Pain observations, Painful, Part hurts, Pain (finding)

Shaping the future of *All of Us*

- What research is best suited to this large cohort?
  - Long term outcomes (rare will still be hard)
  - Gene-Environment interaction
  - Longitudinal research
  - Diverse populations
  - Comparative effectiveness research

- What data are crucial for your field of research?
  - *Must-have* data points: lab results, diagnostic codes, questionnaires
  - Population characteristics
  - Sample types and storage
  - Frequency of data

All of Us RESEARCH PROGRAM