The *All of Us* Research Program

A Researcher's Guide

ResearchAllofUs.org

A Researcher's Guide to the *All of Us* Research Program

Table of Contents

Executive Summary

Background
The *All of Us* Research Program empowers researchers to explore the breadth of data and data types in the *All of Us* dataset and to develop their research questions within the Researcher Workbench cloud environment. With this opportunity comes a responsibility to understand the design and composition of the dataset to avoid methodological errors and data inference.

The Researcher's Guide was developed in response to requests for guidance on how to appropriately use *All of Us* data for scientific research. Researchers should use this document to understand the availability, strengths, weaknesses, and nuances of the data before developing methods to answer their scientific questions and interpret their results.

Introduction
This document provides a summary and links to detailed information on the program's design, participant eligibility, enrollment procedures, and more. The goal is to create a centralized location for information relevant to researchers and to promote responsible use and interpretation of the data.

How to use this document

- **Researchers who are new to *All of Us* data** should start at the history section, paying special attention to program design and participant recruitment.
- **Researchers who plan to use demographic details,** including self-identified race and ethnicity, should review participant characteristics prior to designing a study to understand key factors relating to data completeness, accuracy, and generalizability.
- **Researchers who are already familiar with program basics and statistical challenges of large cohort studies** outlined above can start with this article on relevant data types for their intended studies and explore the unique characteristics of each.

History of *All of Us*

The National Institutes of Health's Precision Medicine Initiative (PMI) Working Group of the Advisory Committee to the NIH Director (ACD) issued a report in 2015 that informed the framework for *All of Us*.The first participants were enrolled as beta testers in mid-2017. National enrollment in the program officially opened on May 6, 2018. In May 2020, the program launched the Researcher Workbench allowing researcher access to participant data. On March 17, 2022, *All of Us* released its initial genomic dataset, which included more than 100,000 short-read whole genome sequences. Subsequent data releases have included additional data types and participants. Enrollment, data collection, and genomic data generation are ongoing.

Currently, *All of Us* participants are invited to authorize the sharing of their electronic health records (EHRs) for research use, complete health-related surveys (known as participant-provided information or PPI in some reports) and tasks, take physical

measurements, provide biospecimens from which genomic information may be derived and other assays conducted, and share data from wearable digital health technologies, such as Fitbit devices. The depth, breadth, completeness, and accuracy of data continue to be a significant focus of the program and will continue to change as the program evolves and gathers additional longitudinal and linked participant data.

Planning Process and Study Design

Overview

Research conducted with a broad participant cohort and with meaningfully large data is key to fulfilling the promise of precision medicine to improve the lives of all. *All of Us* set out to enroll one million or more participants who reflect the U.S. population into a prospective cohort study. Research with the *All of Us* dataset is possible because of the ongoing support and engagement of the participants who generously share their health information with the program.

This section covers:

(1) The rationale for prioritizing certain sample characteristics
(2) The advantages and disadvantages of the sample and resulting dataset
(3) Best practices for responsible use and reporting of *All of Us* data

In its report, the PMI Working Group estimated prevalent and incident disease frequencies and conducted power calculations in order to determine the scale needed to detect relationships between hundreds of health outcomes and their drivers. Based on these and other considerations, the Working Group recommended that *All of Us* seek to achieve a cohort size of one million or more individuals to be followed longitudinally for at least 10 years.

All of Us prioritizes creating a cohort that will, over time, benefit the entire U.S. population. Researchers can leverage the *All of Us* participant cohort to provide new insights into factors affecting health and disease risk, as well as prevention and therapeutic strategies. Critically, the recruitment process was designed to ensure that participants from communities that have been historically less represented in biomedical research would be included in sufficient numbers to allow robust inferences within these groups.

All of Us does not focus on any particular set of diseases or health status for potential recruitment. This allows researchers to study a broad range of conditions. Inclusion of healthy individuals can assist with efforts to identify new risk factors that may predict future conditions. Inclusion of those with prevalent health conditions allows researchers to study pharmacogenomics and treatment outcomes across a wide range of conditions and medications. Finally, the broad *All of Us* participant cohort provides new opportunities to understand the factors that contribute to health and disease risk as well as response to therapies within the context of many backgrounds.

Access and Use of *All of Us* Data

The *All of Us* Research Hub has a tiered-data access model with three data tiers:

- **Public Tier:** The Public Tier contains aggregate-level data that poses negligible risks to the privacy of research participants. This data is accessible to the public through the *All of Us Research Hub*.
- **Registered Tier:** The Registered Tier contains data that poses a low risk to the privacy of research participants. This data is only accessible to authorized users within the *All of Us* Researcher Workbench; all access is logged and may be audited.
- **Controlled Tier:** Data that poses the most significant risks, although still low, to the privacy of research participants. Like the Registered Tier, this data is only accessible to authorized users within the All of Us Researcher Workbench; all access is logged and may be audited.

Researchers must complete the registration process to receive access to Registered or Controlled Tier data in the *All of Us* Researcher Workbench.

As part of the registration process, registered Workbench users must agree and comply with the program's data use policies, including the *All of Us* Data User Code of Conduct (DUCC).

For more information about the *All of Us* data use Policies and types of data available in each tier, please review the following resources:

- *All of Us* Data Use Policies
- *All of Us* Researcher Workbench Data Dictionary
- *All of Us* Data Access Framework

Making inferences from *All of Us* data

The *All of Us* cohort is a broad convenience sample, and findings are not generalizable. The inclusive and open enrollment process means that the *All of Us* participant cohort is **not** a representative sample of the U.S. population. The composition of the *All of Us* participant cohort does not match that of the country's residents.

The program recruits in communities throughout the country via partner organizations including health care provider organizations (HPOs) and engagement partners who represent and serve particular communities. Additionally, any eligible adult can enroll directly via the program website. Details of these strategies can be found in Appendix B. However, while some populations are oversampled, the *All of Us* participant cohort is still a convenience sample in that there were no weighted targets for recruitment of specific subpopulations.

Results from studies using the *All of Us* dataset cannot be generalized to the U.S. population without careful re-weighting by key demographic and socioeconomic variables.

Sample weights are necessary to even attempt to draw meaningful conclusions about the U.S. population. Because the *All of Us* participant cohort is not intended to be "representative" of the U.S. population, caution must be used in any analysis of prevalence of any disease. One approach to converting *All of Us* prevalence estimates to the broader U.S. population would be through **sample weighting**. Sample weighting is used even in more systematically sampled population studies, such as the National Health and Nutrition Examination Survey. A sample weight is assigned to each sample person. It is a measure of the number of people in the U.S. population represented by that person.

Sample weights are under development for *All of Us.* When these are developed, tutorials will be made available to assist investigators in using the weights correctly.

Findings within subsets of the *All of Us* cohort may not be generalizable to the full *All of Us* cohort. An investigator might be interested in studying only a subset of the *All of Us* cohort because not all participants contribute all data types to the program. For example, individuals may decline to provide a DNA sample or to complete surveys beyond the three initially required surveys.

Researchers must assign appropriate weights to their data to make relative frequencies more reflective of the general U.S. population. These weights must be reflective of individual study design and sampling differences if they aim to reach generalizable conclusions about the United States and its territories or the *All of Us* participant cohort as a whole.

All of Us Participants

Community engagement is foundational to *All of Us*. To that end, the Program has supported a national network of community organizations to help create an active participant and researcher community through outreach, engagement, recruitment, and retention.

Outreach includes interactions focused on fostering trust, such as providing materials and information to an audience to learn about research, precision medicine, and *All of Us*. Engagement is relationship building through bidirectional interactions, including information sharing, consultation, collaboration in decision-making, and empowered action among the program, people, awardees, and other partners. Recruitment means facilitating enrollment in the program, and retention refers to ongoing activities with participants after enrollment. To read more about these efforts, see Appendix B.

Eligibility

All eligible people living in the United States and its territories can join the program (see additional information below and in Appendix C: Consent, Authorizations, IRB Approval Process, and Security Measures).

All of Us actively encourages participation from many communities, such as self-identified racial and ethnic minorities and other groups that have been less represented in biomedical research in the past.

This broad recruitment approach is intended to enable rigorous research that may inform policy, prevention, and/or treatment approaches and potentially decrease current health disparities.

All program materials are available in English and Spanish. Additionally, the program ensures that participation is accessible to persons living with disabilities by providing site-specific accommodations.

Inclusion Criteria

Adult participants must be 18 and older with the legal authority and decisional capacity to consent, and they must currently reside in the United States and its territories.

The program piloted recruitment of children ages 0 to 4 with parental consent. Plans for further pediatric recruitment are ongoing as of March 2025.

Exclusion Criteria

All of Us aims to be to be as inclusive as possible. Although all eligible persons are considered for enrollment, it is crucial that adequate consenting procedures be in place to ensure that the rights, safety, and welfare of all participants enrolled are not compromised.

Therefore, until specific enrollment procedures are developed, the following individuals are excluded:

- Adults without decisional capacity to consent
- Children ages 5 to 18 (19 in Alabama and Nebraska, 21 in Puerto Rico)
- People who are incarcerated at the time of enrollment

In the future, the program hopes to develop policies that allow for enrollment of individuals from vulnerable groups, such as cognitively impaired and incarcerated individuals. If the program or its partners learn that a participant has become incarcerated, the program will suspend their participation, using the “deactivate” feature, until incarcerated individuals can participate or until the participant is no longer incarcerated.

The program will make researchers aware of any changes in exclusion criteria through its typical communications channels. To stay informed, researchers can subscribe to the Research Roundup, our monthly email newsletter.

Statistical considerations and limitations of the dataset

Risks to external and internal validity

Sampling biases due to the location of sites, demographics of patient populations, and reach of engagement partners may be large, varied, and are not always known. Additionally, since many recruitment sites exist within health care facilities, there is an ascertainment or selection bias towards those seeking health care. Within the *All of Us* cohort, the subset of participants with particular data types is further affected by factors such as: who voluntarily connected fitness tracker data, chose to complete follow-up surveys, or received targeted retention efforts, etc. Careful consideration of the participant sample in any given study, controlling for confounding variables where possible, and weighting, where appropriate, are needed to improve the probability of reporting generalizable findings.

For instance, most participants who share Fitbit data do not represent the U.S. population broadly or the *All of Us* participant cohort because many are young, predominantly female, white, and college-educated. The program has made efforts to address the disparities in fitness tracker data by providing some participants with Fitbit devices through the WEAR Study. Similar precautions should be taken with other data subsets, such as surveys with considerable missing data.

Confounding

Given the diverse set of data sources and collection methods, both measured and unmeasured confounding are inevitable in any study using the *All of Us* dataset. Sufficient background research must be conducted regarding the research questions of interest to understand the likelihood of confounders and whether they are included as measured variables to be controlled. In addition to the expected confounders for any given research question, the nature of *All of Us* dataset necessitates a few additional considerations outlined below.

Considerations for responsible use and reporting of *All of Us* data

The following recommendations should be considered during study design and in writing research conclusions. These recommendations may not apply to all research studies using the *All of Us* dataset. Researchers should always consult or collaborate with epidemiological and statistical experts to address study-specific questions.

Ethical use of race and ethnicity data

Researchers should avoid research that may be stigmatizing to individuals, groups, or communities. Race and ethnicity should not be used as biological variables. Research including self-reported race and ethnicity should follow recommendations from the National Academies of Science, Engineering, and Medicine.

Recruitment site

The enrollment site, site type, or process (Federally Qualified Health Center, HPO, Veterans Affairs facility, or participant self-enrollment) is a potential confounder in all analyses using *All of Us* data due to systematic differences in the sampling frames of each location. Researchers should always include this variable in multivariate analyses. This information is currently available only via SQL queries in the Researcher Workbench, though more user-friendly methods are in development. Sites can also be clustered into groups to reduce the dimensionality of the data. Unknown confounding by site may also occur if, for example, the *All of Us* recruitment room is adjacent to a particular specialty medical practice and therefore oversamples from a patient population with particular health needs. For example, proximity to a dialysis clinic would result in participants with higher rates of kidney disease than the general U.S. population.

Cross-sectional analyses

The initial (baseline) data collected by *All of Us* are *cross-sectional*, not longitudinal. The longitudinal data are integrated into the dataset over time as subsequent EHR data are collected and participant reassessments are completed.

There are existing guidelines for reporting cross-sectional data. One of the most commonly used is the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines, which include a checklist of 22 items that relate to title, abstract, introduction, methods, results, and discussion sections of observational study articles. Researchers are advised to consult the full STROBE guidelines for additional details.

Retrospective cohort or case-control analyses

All of Us incorporates historical data that preceded the recruitment into the study. Historical Fitbit and EHR data are available for some *All of Us* participants, and those data may be used to create retrospective cohort or case-control studies (See this Fitbit-based example). The STROBE guidelines cover these types of observational studies as well.

Prospective cohort analyses

Much of *All of Us'* scientific value will accrue from the use of baseline data collected at enrollment in combination with follow-up data, such as EHRs and additional surveys. Prospective studies are especially useful in epidemiology because researchers can ensure that the exposure to some type of risk factor occurred before the outcome occurred. Identification of risk exposures prior to disease occurrence can be useful for detecting new etiologies and creating prediction models that are valuable in disease prevention. A particular strength of *All of Us* is that numerous endpoints are possible, rather than a singular focus on one disease.

While there are definite strengths of the *All of Us* study design, there are certain limitations related to missing values both at baseline and at follow-up. There are also other sources of bias, even in the most carefully constructed cohort studies. Researchers should review information relating to potential sources of bias, some of which are tabulated in Appendix A, and in this *Journal of Chronic Diseases* article.

Sampling

***All of Us* was not designed to be a representative sample of the overall U.S. population nor any subpopulation.** It is a convenience sample overall and in each subpopulation. Therefore, researchers should avoid making inferences about prevalence within any given subpopulation, and researchers should avoid the temptation to make comparisons across subpopulations. For example, while it would be useful to know if prevalence of a certain disease differs between individuals of different races or ethnicities, such prevalence estimates are not valid without proper population weighting.

When reporting study design, researchers should avoid using terms like "representative." Instead, you should describe the study as a "broad convenience sample."

Population subsets

For most research questions, researchers will define a subset of the overall *All of Us* cohort. Consequently, when reporting results it is important to describe the characteristics of the participants in that subset rather than those of overall *All of Us* participants. Researchers should make very clear that the inferences from this kind of research almost certainly do not apply to the overall U.S. population. As an example, in the Fitbit study described above, researchers included the following limitation: "There are several limitations relevant to our conclusions. Our population—predominantly White women—was likely a reflection that the data provided was voluntary and biased by factors that influence accelerometer/wearable use." The authors also stated: "Future studies should aim to extend these findings to more racially diverse populations to confirm the generalizability of these findings."

Weighting

Sample weights and other statistical approaches would allow findings to be more generalizable to the U.S. population. The appropriate use of sample weighting is a complex topic and should be approached cautiously. Researchers have proposed solutions for this, including developing synthetic weights to improve generalizability of *All of Us* data.

Working with *All of Us* data types

Researchers should familiarize themselves with all of the data types that will be used in a given research analysis. This article is a good place to start and includes details about the collection,

curation, and release to researchers for each primary *All of Us* data type.

Publishing, presenting, or sharing work using *All of Us* data

Researchers should review the *All of Us* Publications, Presentation, and Poster Checklist before publishing, presenting, or otherwise sharing their research to ensure compliance with the *All of Us* Data User Code of Conduct and associated policies. These policies apply to any activities in which *All of Us* data is shared, and is not limited to publications.

The *All of Us* User Support Hub also features resources for these policies and the steps needed to ensure compliance.

Conclusion

All of Us has created this data resource for researchers to study unique combinations of data types at scale. With careful consideration for the nuances of the data and statistical methods, hugely impactful studies can be accomplished within this broad cohort using the Researcher Workbench.

Linked References

***All of Us* Research Program**

Data Sources: https://www.researchallofus.org/data-tools/data-sources/

Researcher Workbench: https://www.researchallofus.org/data-tools/workbench/

Data Access Tiers: https://www.researchallofus.org/data-tools/data-access/

Data Access Framework: https://www.researchallofus.org/faq/data-access-framework/

Registration: https://www.researchallofus.org/register/

All of Us Publication, Presentation, and Poster Checklist: http://allof-us.org/PubsChecklist

Data Use Policies:

Data User Code of Conduct: Researchallofus.org/DUCC

Publication and Presentation Policy: Researchallofus.org/PubPresPolicy

Data and Statistics Dissemination Policy: Researchallofus.org/DSDPolicy

Policy on the Ethical Conduct of Research: Researchallofus.org/PolicyEthicalConductResearch

Policy on Stigmatizing Research:
Researchallofus.org/PolicyStigmatizingResearch

Policy on Respectful Research Involving AI/AN Populations:
Researchallofus.org/PolicyRespectfulAIANResearch

Community Engagement Partner List:
https://www.joinallofus.org/community-engagement-partners

Research Roundup Newsletter Signup: https://allof-us.org/RRSignup

Article about Fitbit data and the WEAR Study:
https://allofus.nih.gov/news-events/announcements/research-roundup-all-us-participants-fitbit-data-drive-new-research

Working Group Reports and Recommendations

Precision Medicine Initiative Working Group Charge:
https://acd.od.nih.gov/working-groups/pmi.html

Precision Medicine Initiative Advisory Committee to the Director Report:
https://acd.od.nih.gov/documents/reports/PMI_WG_report_2015-09-17-Final.pdf

National Academies Report on Use of Race, Ethnicity, and Ancestry as Population Descriptors in Genomics Research:
https://www.nationalacademies.org/our-work/use-of-race-ethnicity-and-ancestry-as-population-descriptors-in-genomics-research

National Academies Report on The Use of Race and Ethnicity in Biomedical Research:
https://www.nationalacademies.org/our-work/the-use-of-race-and-ethnicity-in-biomedical-research

Sample Weighting Educational Materials

Tutorial on Constructing Weighting for NHANES:
https://wwwn.cdc.gov/nchs/nhanes/tutorials/weighting.aspx

Project Description for Designing Sample Weights for the *All of Us* Research Program:
https://reporter.nih.gov/search/5XdbuEDrVUy4tkUVc77V3g/project-details/10796237

Appendices

Appendix A: Summary of Biases from Sackett et al
Appendix B: The *All of Us* Research Program's Outreach, Engagement, Recruitment, Enrollment and Retention Strategies
Appendix C: Account Creation, Consent Process, and Compliance with state-level regulations

Appendix D: Biospecimen Processing and Storage
Appendix E: Data Access Framework
Appendix F: Data Security

Bibliography

Elm E, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies.The Lancet, Oct 20;370(9596):1453-1457 https://doi.org/10.1136/bmj.39335.541782.AD

Manolio T., Bailey-Wilson, J. & Collins, F. Genes, environment and the value of prospective cohort studies. Nat Rev Genet 7, 812–820 (2006). https://doi.org/10.1038/nrg1919

Perry AS, et al. Association of Longitudinal Activity Measures and Diabetes Risk: An Analysis From the National Institutes of Health *All of Us* Research Program. J Clin Endocrinol Metab. 2023 Apr 13;108(5):1101-1109. doi: https://doi.org/10.1210/clinem/dgac695

Sackett DL. Bias in analytic research. J Chronic Dis. 1979;32(1-2):51-63. doi: https://doi.org/10.1016/0021-9681(79)90012-2

Appendices for A Researcher's Guide to the *All of Us* Research Program

Table of Contents

Glossary and Common Acronyms

Data and Research Center (DRC)
Participant-provided information (PPI)
Health care provider organization (HPO)
Direct volunteers (DVs)
Electronic health records (EHRs)
Institutional Review Board (IRB)

Outreach: Providing materials and information to an audience (unidirectional interaction);
Engagement: Listening, responding, and supporting an audience (bidirectional interaction);
Recruitment: Facilitating enrollment in the program
Retention: Ongoing activities with participants after enrollment
Mobile engagement assets (MEAs)
Mobile clinical research units (MCRUs): Vehicles equipped to enroll participants at locations beyond the ISIA-approved institution's own physical location/property. The MCRU may be equipped with some or all of the following facilities: private interview rooms, a bathroom, a phlebotomy chair, a centrifuge, and a refrigerator/freezer.
Institution-Specific IRB Application (ISIA)
Virtual Ambassador Program (VAP)
The Support Center: The staffed response team for researcher questions about data and how to use the resource within the researcher workbench.
Genomic Return of Results (gROR)
Participant Portal: The graphical user interface supplied to participants to enroll and interact with the program.
Physical measurements and biospecimen (PM&B)
Raw Data Repository (RDR)
Curated Data Repository (CDR)
Dater User Code of Conduct (DUCC)
Data Use and Registration Agreement (DURA)
Authorization to Operate (ATO): An ATO is a formal declaration by a Designated Approving Authority (DAA) that authorizes operation of a product and explicitly accepts the risk to agency operations.
Certificates of Confidentiality: To protect participants from having their information disclosed as part of any legal demand (such as a court order or a request from federal, state, or local law enforcement) or other claims, *All of Us* Research Program awardees, including subawards, subcontracts, and vendors, are e covered by Certificates of Confidentiality.

Appendix A: Summary of Biases from Sackett et al

Major sources of bias that affect case–control and prospective cohort studies (summarized from Sackett DL, Bias in analytic research. *J Chronic Dis*. 1979)

Bias	Description
Biases that relate to subject selection	
Prevalence–incidence or survival bias.	Selection of existing cases that are currently available for study will miss fatal and short episodes, and might miss mild or silent cases.
Non-response (or respondent) bias	Differential rates of refusal or non-response to inquiries between cases and disease-free comparison subjects.
Diagnosis bias	Also known as diagnostic suspicion bias. Knowledge of a subject's exposure to a putative cause of disease can influence both the intensity and outcome of the diagnostic process.
Referral or admission-rate bias	Factors related to the probability of referral. Cases who are more likely to receive advanced care or to be hospitalized — such as those with greater access to health care or with co-existing illnesses — can distort associations with other risk factors in clinic-based studies, unless the same referral or admission biases are operative in disease-free comparison subjects
Surveillance bias	If a condition is mild or likely to escape routine medical attention, cases are more likely to be detected in people who are under frequent medical surveillance.
Biases that relate to measuring exposures and outcomes	
Recall bias	Questions about specific exposures might be asked more frequently of cases, or cases might search their memories more intensively for potential causative exposures.
Family information bias	The flow of family information about exposures or illnesses can be stimulated by, or directed to, a new case in its midst.
Exposure suspicion bias	Knowledge of a patient's disease status can influence the intensity and outcome of the search for exposure to a putative cause.

Appendix B. The *All of Us* Research Program's Outreach, Engagement, Recruitment, Enrollment and Retention Strategies

Outreach to communities and grassroots community engagement are foundational to the *All of Us* Research Program. To that end, we built and are continuing to expand a national network of community organizations that facilitate the four essential and unique components of the creation of an active participant community: outreach, engagement, recruitment, and retention. We define outreach as providing materials and information to an audience (unidirectional interaction). Engagement refers to listening, responding, and supporting that audience (bidirectional interaction). Recruitment means facilitating enrollment in the program, and retention refers to ongoing activities with participants after enrollment.

Outreach

Outreach is defined as providing materials and information about the research program in advance of creating a research program account (unidirectional interaction).

Prospective participants will learn about the *All of Us* Research Program via:

1. **Targeted advertising**, including:
    a. Print flyers, brochures, and posters
    b. Advertisements (TV, radio, online, and mobile)
    c. Billboards and bus advertisements
    d. Direct marketing (email and mail)
2. **Personal interest groups:**
    a. Social media
    b. Community events
    c. Press coverage
3. **Interactions with health care provider organizations (HPOs) or other program partners**, including:
    a. Conversations in waiting areas
    b. The regular course of clinical care at HPOs
    c. Local informational events
    d. Regional informational events organized by research program awardees, HPOs, or other program partners.
    e. Employee invitations
    f. Re-contact of consented participants in existing research programs
    g. Vsits to outpatient clinics

HPOs may use both nationally and locally developed outreach approaches to engage their patient population, members of their health plan or of an affiliate, and any interested eligible individuals in their catchment area. Advertisements direct potential participants to local program contacts or the *All of Us* website.

4. **Mobile engagement assets (MEAs)**

For additional outreach, *All of Us* has deployed MEAs to bring awareness about the research program to areas not covered by active recruitment sites. This outreach is especially valuable for engaging highly mobile populations and others from communities that may not be reached by HPOs or other partners. The MEAs offer a warm and welcoming environment where people can learn about the program. The MEA experience is carefully developed to be considerate of cultural aspects of interactions and to leverage funded community engagement partners and existing community networks.

5. **Interactions with *All of Us* partner organizations**

Partner organizations may provide education and awareness to their communities about *All of Us* and how to join. Some organizations may also teach health care providers about the program.

Engagement

Engagement is defined as listening, responding, and supporting an audience (bidirectional interaction). *All of Us* and its partners conduct dialogues with communities (e.g., by creating advisory councils, hosting educational webinars, organizing community convenings, staffing health helplines, and tabling at community events). A full list of engagement partners can be found here.

Recruitment

Recruitment is defined as facilitating enrollment in the program. Interested individuals are able to enroll in one of two ways:

1. **Through a participating HPO, a community partner, or a federally qualified healthcare organization (FQHC)**

   This approach is primarily—but not exclusively—for people who are a member of an HPO's health plan and their affiliates or have received care at any of several participating health centers across the United States and its territories. However, any eligible individual who wishes to enroll at an HPO may do so, even if they do not have a prior connection with that HPO. Participating HPOs were chosen based on their ability to reach a broad cross-section of the population, as well as for their ability to support and quickly enable the technical and scientific requirements of the study.

2. **Virtually (known as "unpaired" or self-guided enrollment)**

   This approach is for individuals without a conveniently located HPO site.

Both the unpaired and HPO-paired paths rely on program-specific digital registration tools accessible via a smartphone application and/or a program or partner website.

Enrollment

Via HPOs and community partners

Program partners may use the following enrollment techniques:

1. Placing kiosks in waiting rooms, cafeterias, corridors, or other locations at clinics.
2. Pre-screening and reaching out to potential participants using existing patient/research registries or EHR systems. Sites must obtain a waiver of consent from the Precision Medicine Initiative (PMI) IRB to access personal information in EHRs or registries for screening purposes.
3. Sending personalized invitations from a health care provider to prospective participants, using IRB-approved text.
4. Deploying mobile clinical research units (MCRUs) with trained site staff to inform hard-to-reach populations and to facilitate the enrollment and completion of study procedures as applicable. MCRUs are defined as vehicles equipped to enroll participants at locations beyond the ISIA-approved institution's own physical location/property. The MCRU may be equipped with some or all of the following facilities: private interview rooms, a bathroom, a phlebotomy chair, a centrifuge, and a refrigerator/freezer.
5. Creating "pop-up" locations to enable *All of Us* trained site staff to engage and enroll prospective participants in spaces (indoor and outdoor) equipped to enroll participants at locations/communities beyond the ISIA-approved institution's own physical location/property. These spaces may include schools, places of worship, clinics, or community events. In lieu of requiring a visit to the *All of Us* research office, the eligible mobile "pop-up" staff has the ability to enroll participants and/or conduct physical measurement and biospecimen collection from pre-consented participants at events and locations, provided all privacy and confidentiality requirements are met. Agencies or institutions that provide space or "host" an *All of Us* pop-up are generally not considered to be engaged in human subjects research.
6. Setting up a modified clinic site to create a separate space for enrolling *All of Us* participants. A modified clinic site is defined as a structure or parked vehicle that is set up on ISIA approved institution property as an alternative clinic space for participant enrollment.
7. Hosting educational or launch events geared towards enrollment, and providing enrollment materials at existing events.

8. Sending email and short message service (SMS) communications to interested individuals who have opted in to receive program messaging.

For HPOs and community partners that wish to engage participants in an inpatient or community setting, precautions will be taken to ensure the patients' safety, fitness to consent (physical, emotional, and decisional), ability and willingness to consent, and comfort (physical and emotional)

Trained *All of Us* site staff obtain the approval from the prospective participant's care team prior to approaching the individual. The member of the care team providing approval must have direct access to and knowledge of the patient and their current condition in order to assess the capacity to consent. *All of Us* site staff will work with the clinical care team to ensure that speaking with the potential participant about *All of Us* does not disrupt the individual's clinical care. If, in the opinion of the care team, a prospective participant does not have the capacity to consent, the timing is inappropriate for the person, or the approach by *All of Us* site staff would in any way be disruptive, the individual will not be approached. Care team members and *All of Us* site staff should confirm it has been at least two nights since any surgery or procedure, consider any medication the person may currently be taking, or medication previously given that may impact the person's awareness or create situational vulnerability. When engaging a person in the inpatient setting, *All of Us* site staff should first ask the person if they feel comfortable making a decision about research participation at that time and should be mindful of any indicators (e.g., drowsiness, incoherence, slurred speech, short-term memory lapses) that suggest the participant may not be fully aware and stop the interaction as appropriate. Whenever a person proceeds to enroll, *All of Us* site staff should leave information about the program with the participant as a reminder that they have joined the program.

Via self-enrollment

Targeted outreach materials are developed specifically to reach the unpaired populations, primarily in areas not served by HPO awardees.

Retention

All of Us is expected to last at least 10 years. Follow up is expected to be continuous for the life of the program. For example, data from the EHR is expected to be added to the *All of Us* dataset for participants who consented to share their EHR data and signed the Health Insurance Portability and Accountability Act (HIPAA) Authorization for Research. Participants will not receive notification each time EHR data are added. In addition, *All of Us* may periodically reassess participants via surveys or additional biospecimen collection in order to obtain longitudinal data and maximize the research value of the cohort.

All of Us has a retention strategy that uses both digital and non-digital approaches. Due to the scale and geographic range of the program, we anticipate that most long-term interactions with the program will be digital; thus, the web and mobile applications are designed to be user-friendly and engaging with a responsive and intuitive user interface. In addition, *All of Us*

works closely with its engagement partners to meet retention goals. For example, community engagement partners work within their communities to show that the program is invested in responding to participants' interests and needs and seeks to work with them for a decade or more (hosting participant convenings, regular "check-ins" via phone, ongoing social media campaigns, newsletters). See the list below for a summary of retention strategies implemented within *All of Us*:

- Periodic communication to participants from site staff (e.g., phone calls, emails, birthday cards, or mailing of *All of Us* promotional materials such as letters, brochures and invitations to events).
- Periodic printed newsletters or e-newsletters with local updates about the program (Newsletters may include profiles of participants, researchers, or research staff, as well as answers to common questions and information about community events).
- Use of social media (X, Facebook, etc.) to engage and update participants about the program and promote program-related events.
- Periodic participant appreciation events to thank *All of Us* participants and maintain relationships between the program and participants.
- Health and science educational events, such as science cafés, health fairs, or seminars, where researchers discuss their current research and how it pertains to *All of Us* and promote health literacy and where participants can meet the research staff.
- Town halls with presentations by site principal investigator(s) and/or site staff, and Q&A sessions with the audience.
- *All of Us* wallet cards to track completion of study activities and/or appointment reminders.
- *All of Us* scorecards to help participants track their physical measurements and share this information with their health care providers for further management.
- Welcome/Exit Packages to assist in informing and engaging participants on different aspects of *All of Us*. These packages are a great opportunity to show appreciation and provide participants with additional information on using the *All of Us* Participant Portal.
- A post-enrollment survey mailed or emailed to participants after they have completed the enrollment process. The survey solicits feedback on different aspects of the enrollment process and identifies opportunities for improvement. The survey results are tracked over time and provide valuable metrics on participant satisfaction.
- Postal mailings to assist with survey completion, such as a survey instruction brochure that provides instructions on portal log in and survey completion. Sites may propose efforts in addition to those noted above, subject to IRB review and approval.

Readability of Outreach and Enrollment Materials

Consistent with best practice recommendations of the National Quality Forum (NQF) and the Agency for Healthcare Research and Quality (AHRQ) for engaging participants with a broad spectrum of health literacy, outreach and enrollment materials are written at the middle school reading level. This ensures that these materials are broadly comprehensible by the greatest number of people. Reading-level experts review all public research program copy and are guided by the following target metrics:

- Flesch Reading Ease: ≥70
- Flesch–Kincaid grade level: ≤7
- Passive sentences as a % of total: ≤10%
- Sentences per paragraph: <3

Written materials are available in English and Spanish, and where possible, the enrollment materials incorporate multimodality presentation methods (aural, visual, and interactive) to aid comprehension of people with low literacy.

Appendix C. Account Creation, Consent Process, and Compliance with state-level regulations

To understand the full scope of consent and compliance, the *All of Us* Research Program Protocol is available for reference.

Participant account creation

Account creation begins after a user clicks the "Join Now" button on the website or mobile application and requires entering personal contact information, creating and confirming a password, and choosing a preferred language from a drop-down menu. This account information is stored securely in the Participant Portal host database. Once the participant completes consent, a copy of this information is also transferred to the Raw Data Repository (RDR).

Currently, all individuals must create their *All of Us* account electronically. Trained *All of Us* staff at enrollment sites or at the program's Support Center can facilitate this process and accommodate individuals who have differing levels of technological capability. Upon participant request, trained staff may provide assistance by creating login and/or password information with participants.

Consent

Following account creation, people wishing to enroll in the program answer specific questions in advance of providing informed consent:

- They are asked to confirm that they meet the program's eligibility criteria.
- They are asked their state of residence and the state where they receive most of their health care to enable compliance with state-specific requirements.

Informed Consent

Disclosure, voluntariness, and decisional capacity make up the core of valid informed consent processes. All persons wishing to participate in *All of Us* complete an informed consent process through which participants learn about the program through text and visual aids and unambiguously indicate their decision to participate. The materials presented are consistent across the research program but may be customized based on where an individual receives most of their health care, enrollment method, or site affiliation (Unpaired or paired with an enrollment HPO).

The informed consent process is initially administered and documented electronically. It is designed as a living process, with just-in-time information loops and opportunities for periodic updates. The electronic consent process is self-paced, and there is no time limit to complete it. Individuals can rapidly navigate, repeat, pause, and review according to their own information needs. The consent process can be experienced as a self-navigated, supported, or hybrid

process. Individuals are able to choose their preferred informed consent experience, either alone or with the assistance of another person; soliciting in-person support from trained site staff; or calling the Support Center. Informed consent materials are currently available in English and Spanish.

Additional Consent Modalities

The current electronic informed consent can be adapted to meet the needs of people with various learning styles and health literacy levels.

The consent process may be self-navigated or completed with support. There are circumstances where individuals intellectually capable of providing informed consent may require or prefer assistance with the consent process, due to physical, social, educational, or other limitations. *All of Us* site staff experienced in facilitating informed consent procedures are available to facilitate the *All of Us* consent procedure. They utilize approved electronic consent visual aids and text and engage the prospective participant in a discussion of informed consent to answer any additional questions or concerns a participant may have. Trained site staff who facilitate the consent process are required to co-sign the informed consent document.

These additional modalities of consent do not preclude person-to-person contact for questions and/or concerns. Trained personnel are available on site and at the Support Center to address questions or concerns about the program. Regardless of the approach to consent, participants will be given access or receive a copy of the consent form for their records.

Electronic Consent

Using an electronic informed consent process throughout the program ensures consistency of the consent information. The electronic consent process includes information on the detailed nature, purpose, procedures, benefits, and risks of and alternatives to participating in *All of Us*. Due to the longitudinal nature of the study and the patchwork of state regulations regarding research, the informed consent is modular. This also allows the program and its partners to provide a flexible participant experience. Each module requires an electronic signature from the participant, if they agree to it.

1. **Primary Consent:** The primary consent module gives an overview of participation in *All of Us*, and any potential risks and benefits of participation. Signing the primary consent indicates general understanding of *All of Us* and consent to take part in the surveys, allow the program to link in additional external health data about them (e.g., environmental data, claims data, data from disease registries), to share physical measurements and biospecimens, to have samples shared with the program's biobank partners, to have samples analyzed through assays and genomic testing, and share data from wearables (e.g., Fitbit devices).
2. **HIPAA Authorization for Research EHR/Part 2 Supplement:** This module gives details about allowing the research program access to a participant's EHR, including

health records protected by 42 CFR Part 2 (drug and alcohol abuse patient records), referred to as “Part 2” records. State-specific versions of this form are developed as needed to meet state laws and regulations regarding the release and use of health record data.

3. **Partnered Studies Consents:** Additional consent modules are frequently necessary for participation in a partnered study such as Exploring the Mind and Nutrition for Precision Health. Details about these consent modules can be found within partnered study protocols.

Compliance with state-level laws and regulations

The considerable number of distinct state laws and regulations governing the collection and use of various data types has prompted *All of Us* to conduct routine legal surveys and analysis. To augment this process, the program has leveraged expertise from the NIH Office of the General Counsel and the Office of Civil Rights. Understanding these locale-specific variations is essential in enabling enrollment of individuals from all across the United States, including the populated U.S. territories.

Appendix D: Biospecimen Processing and Storage

After collection, biospecimens are shipped to Mayo Clinic Laboratories for initial unpacking, accessioning, and sorting. They are processed and stored in the centralized biobank at Mayo Clinic. Samples are held at the centralized biobank indefinitely unless an individual participant withdraws from the study.

The centralized biobank is responsible for facilitating collection, shipment, processing, DNA isolation, sample aliquoting, storage, and future access to biospecimens. Initial sampling processing is performed at the site of collection followed by a shipping protocol that maintains the cold chain needed to prevent specimen degradation.

Processing Methodology

The collection site performs minimal sample processing. All biospecimens collected on-site or at an HPO are stored refrigerated until shipped. This does not apply to saliva kits. Specimens are generally shipped to the biobank within 24 hours of collection. The biobank tracks time of collection to the time of freezing. Samples are processed by the biobank within 40 hours of blood collection. Any samples outside of the 40-hour window are noted, and the site is notified. Specimens may continue to be used at the biobank's discretion.

All blood tubes are processed at the biobank. Information on all aliquots, including the volume for each, are recorded in the laboratory information management system and linked to a unique biobank ID.

Transport of Biospecimens

All biospecimens from HPO and non-HPO collection sites are shipped to Mayo Clinic in Styrofoam containers containing a cool pack to keep samples cool. For some HPOs, a MCL courier is responsible for the packaging of materials and containers, packing samples, and adding the cool packs in the shipping container. The logistic capability provided by MCL is used to transport the specimens from the HPO sites to the biobank. MCL utilizes a network of couriers, coupled with a direct arrangement with FedEx and other carriers, to enable daily domestic and international specimen shipment from clients to the performing laboratories in Minnesota, ensuring that shipments are made in accordance with all federal, state, and international regulations. HPO sites that do not use MCL self-ship samples.

Reliability of Sample Tracking and Identification

The collection sites utilize MayoLINK, a Mayo Clinic application that provides connectivity to MCL to order the participant's biospecimen collection. The biobank laboratory information management is built on software developed by LabVantage Solutions, Inc. Core capabilities include kit tracking, sample accessioning and annotation, sample processing and testing, storage, and shipping. All aspects of the sample lifecycle are tracked. Security within this application is robust and multilayered to keep participant and sample data secure.

The enrollment sites utilize HealthPro, a web-based application developed and managed by the *All of Us* Data and Research Center (DRC) to record information from participants' physical measurements and to complete the biospecimen ordering workflow. Within HealthPro, authorized and trained *All of Us* site staff are able to view the first name, last name, date of birth, and ZIP code of a participant to verify participant identity and biospecimen eligibility during the PM&B visit. The biobank ID is also displayed. Prior to sample collection, a sample manifest and labels for the collection tubes are printed via the HealthPro Portal. The collection tube labels contain the unique biobank ID but no other participant identifiers. The biobank ID is linked to the participant ID by the DRC. Security meets Federal Information Security Management Act (FISMA) Moderate Authorization and Accreditation standards.

Sample Receipt, Verification, and Routing

Samples are first transported to MCL. Trained staff triage incoming shipments by shipment time. Specimens are taken from their original shipping containers and stabilized at the correct temperatures. The specimens are then expedited to the internal operations area for order processing and receipt verification before being routed to the biobank. Operators manage the automation and specimens' receipt and processing. The validated transportation temperature is maintained at all times during pre-analytic processes, and specimens are promptly delivered to the biobank at the same temperature used for shipping.

Long-Term Specimen Storage

Processed blood samples are stored in robotically controlled –80°C freezers, and whole blood samples are stored in vapor phase liquid nitrogen units. Most prepared specimens are stored at the primary site in Minnesota; the Jacksonville, Florida, biobank facility serves as the off-site, secondary storage site for approximately 25% of the samples. Both biobank sites have a comprehensive disaster recovery and business continuity plan.

Appendix F. Data Security

Data Security Policy Principles and Framework

Maintaining data security and privacy within the *All of Us* Research Program is paramount to maintaining participants' trust and engagement. Extensive regulations, policies, governance, compliance, and technical safeguards are being implemented to ensure that participant data security and privacy are appropriately protected. Specifically, the Participant Portal hosts, Genome Centers, Genetic Counseling Resource, and the DRC are implementing standards at the FISMA moderate baseline, which is described in more detail below.

Security Posture

The program's security approach is a combination of regulations, policies, governance, compliance, and technical safeguards being implemented across various data flows and data types. In the case of the DRC and the Participant Portal hosts, we apply an iterative risk-based approach to implement security at all layers of the system. We leverage components from the NIST Risk Management Framework (NIST SP 800-39), the NIST Cybersecurity Framework, and the Security and Privacy Controls for Federal Information Systems and Organizations (NIST SP 800-53 rev4). Based on the risk to the system and the data contained in the system, we implement controls at the FISMA moderate baseline and select additional enhancing controls where needed, using a "pure" information security perspective to prioritize best-of-breed security methods.

FISMA and Its Significance to the *All of Us* Research Program

The Federal Information Security Management Act (FISMA) is U.S. legislation that defines a comprehensive framework to protect government information, operations, and assets against natural or manmade threats. FISMA was signed into law as part of the E-Government Act of 2002.

FISMA assigns responsibilities to various agencies to ensure the security of data in the federal government. The act requires program officials and the head of each agency to conduct annual reviews of information security programs, with the intent of keeping risks at or below specified acceptable levels in a cost-effective, timely, and efficient manner.

The National Institute of Standards and Technology (NIST) outlines nine steps toward compliance with FISMA:

1. Categorize the information to be protected
2. Select minimum baseline controls
3. Refine controls, using a risk assessment procedure
4. Document the controls in the system security plan
5. Implement security controls in appropriate information systems

6. Assess the effectiveness of the security controls once they have been implemented
7. Determine agency-level risk to the mission or business case
8. Authorize the information system for processing
9. Monitor the security controls on a continuous basis

The program is working with internal and external independent third-party security experts to define the system and its security needs, assess whether security controls are implemented, monitor and test that controls continue to be effective, and respond appropriately to incidents or anomalies to address and resolve any issues.

Relation to PMI Data Security Principles and Framework

All of Us adheres to the Data Security Policy Principles published by the White House. These principles utilize four proven design concepts:

- Authenticate: All components require authentication
- Authorize: All data, other than public data, requires explicit authorization to access
- Audit: All data access is logged (to a different system), with alerts for anomalous events
- Encrypt: All data in transit and all data at rest is encrypted

By following this principled approach, combined with meeting the FISMA compliance requirements, we implement the core data security functions of identify, protect, detect, respond, and recover at all times.

Consistent with the guidance, awardees are implementing the system to meet the PMI Security Principles and show alignment with the PMI Data Security Framework. This is achieved through the implementation of a system accreditation process following the NIST Guide for Applying the Risk Management Framework to Federal Information Systems (NIST SP 800-37). The system is authorized at the FISMA moderate classification and is assessed by a third party to meet the moderate baseline security controls in NIST-800-53, with a concentration on continuous monitoring and audit controls. Using those controls and more, it is our goal to identify likely threat sources , protect against those threats, detect incoming attacks, respond to those attacks, and recover the full integrity of all systems along with accurate event reporting.

Multiple Levels of Data Security and Privacy

We take a multilayer defense-in-depth approach to security. The DRC, Genome Centers, Genetic Counseling Resource, and Participant Portal hosts work independently and in parallel, with shared security philosophies and approaches, though some implementation details will differ. Below are specifics on how we will, over the project lifetime, implement our various layers of security.

Perimeter Security

All external-facing properties for the DRC, Genome Centers, Genetic Counseling Resource, and Participant Portal hosts have signature- and non-signature–based intrusion detection and protection systems and are scanned regularly for vulnerabilities.

Resilient Infrastructure

The DRC uses the Google Cloud Platform (GCP), which is run and maintained by Google and protected by Google's security engineering team. This platform is undergoing FedRAMP evaluation, with portions already having received an Authority to Operate (ATO) and used in several FISMA moderate projects. See https://cloud.google.com/security/whitepaper for more details.

The Participant Portal hosts use AWS East/West cloud infrastructure. The AWS cloud system is FedRAMP authorized and has been determined to have a security categorization of moderate. See https://aws.amazon.com/security/ for more details. Both of these cloud environments enable extreme redundancy and the ability to recover from lost computing assets.

Hardened Access Controls

The DRC's infrastructure and applications use Google's Access Control for both authentication and authorization, including two-factor authentication. This leverages Google's existing well-tested protections of this service, used for Google internal employees and external users (e.g., Gmail).

The Participant Portal hosts' infrastructure components utilize Amazon's Identity and Access Management (IAM) for authentication and authorization, including two-factor authentication. Application authentication uses tokens signed and validated with the latest recommended cryptographic algorithms (e.g., JSON Web Token). Across all applications and infrastructure, no user is authorized to access participants' data within the development environment without human action to approve their access, with the exception of public data. Users only have the lowest necessary access. By default, authenticated users can see nothing other than their own data. They must be explicitly authorized to access resources. All privilege escalations are logged.

Continuous Auditing and Monitoring

The Participant Portal hosts and the DRC use various auditing and monitoring tools, such as Google's StackDriver platform and CloudWatch/CloudTrail/Splunk, for handling logs. Error and anomaly detection is forwarded to both visual dashboards and real-time alerting systems to support system health remediation and security assessments.

Our systems are built on "REST APIs," so all commands are basic Web requests. All requests—external and internal—are logged.

Logs containing personal identifiers (e.g., searches for named participants from HealthPro) are treated as PII. A limited number of administrators and auditors have access to log data, and all access to logs is itself logged.

Exceptions, errors, and stack traces are sent to a specialized handler and alert response personnel, since software failures are often a precursor to an attack.

Secure Deployment and DevOps

The *All of Us* Research Program platforms—the Participant Portal hosts and the DRC—are created, destroyed, and deployed by automated code per our software development lifecycles. To reduce errors, utilizing repeatable, auditable, and remediable processes minimize direct interaction with resources.

Code Testing Before Deployment

The program uses three testing methodologies:

1. Traditional tests.
2. Static code testing utilizing automated programs (such as SoniqCube).
3. Dynamic code testing, such as running attacks against automatically instantiated fully functional environments.

In addition to dynamic code testing, *All of Us* employs both automated and human-based penetration tests across all assets on a regular basis to look for problems and to ensure our detection systems are working as expected.

Overview of Privacy and Data Confidentiality Protections

PMI Privacy and Security Principles

The PMI Privacy and Trust Principlesand the PMI Data Security Policy Principles and Framework apply to all organizations participating in the *All of Us* Research Program.

Terms and Conditions of Awards

All partners in *All of Us* are required to adhere to the PMI Privacy and Trust Principles. *All of Us* requires the awardee to work with the program and relevant stakeholders to develop a privacy plan within three (3) months of this agreement. The plan shall describe how the awardee will design and implement privacy controls and policy safeguards necessary to ensure secure data sharing, access, and use and data quality and integrity congruent with the PMI Privacy and Trust Principles. Where applicable, the plan should also describe how the awardee will comply with privacy requirements established in the Common Rule, the Public Health Service Act, the 21st Century Cures Act, HITECH, and HIPAA, including relevant supporting regulations, and

agency and program policies. The *All of Us* Research Program, through the Policy Director and in coordination with the Privacy Officer and Program Officer, may require additional privacy measures not included in the PMI Privacy Trust Principles. *All of Us* regularly monitors compliance with the privacy plan and any new privacy requirements as specified by the program and agreed upon by the awardee. The awardee reaches agreement with *All of Us* on requested terms, scope, and timing of privacy reviews.

HIPAA Privacy and Security Rules

The HPOs are required to adhere to the relevant privacy and security standards under HIPAA when aggregating and transferring data for research purposes. Some components of the unpaired participants operations in the Participant Portal hosts are also HIPAA-compliant. In accordance with the PMI privacy, trust, and security principles, participant approval for sharing of EHR data is obtained and retained in all cases.

Security Assessment and Authorization Process

The Participant Portal hosts and the DRC adhere to a security assessment and authorization process that is consistent with FISMA and NIST guidelines. The Participant Portal hosts and the DRC are developing a system security plan that are reviewed by both NIH and an independent party to ensure that controls are commensurate with the assessed risk; if the plans are satisfactory, NIH issues an Authority to Operate (ATO). The program will continuously monitor system security. The program also uses interconnection security agreements for data transferred to the DRC from the HPOs, the Participant Portal hosts, the Biobank, and other program partners.

Authorization to Operate (ATO)

An ATO is a formal declaration by a Designated Approving Authority (DAA) that authorizes operation of a product and explicitly accepts the risk to agency operations. After completing a security assessment, the head of an agency (or their designee) can authorize the system for use or grant an ATO. An agency grants an ATO according to a risk-based framework that analyzes how a vendor has implemented the security controls within their IT environment. For the *All of Us* Research Program, NIH is the DAA. Both the Participant Portal hosts and DRC infrastructures have received ATO from the NIH.

The Common Rule

The Common Rule applies to or is followed by the DRC, the HPOs, the DVs, TPC, and the Participant Portal hosts. All participants provide informed consent to participate in the program, as well as the future research use of their specimens and information that has been stripped of explicit identifiers (e.g., personal names and Social Security Numbers), as well as additional attributes that could disclose a participant's identity with minimal effort (e.g., full residential address). NIH has established a central IRB for exclusive use by the program, which approves

research only after first determining that there are adequate provisions to protect the privacy of human subjects.

Certificates of Confidentiality

To protect participants from having their information disclosed as part of any legal demand (such as a court order or a request from federal, state, or local law enforcement) or other claims, *All of Us* Research Program awardees, including subawards, subcontracts, and vendors, are e covered by Certificates of Confidentiality. NIH issues Certificates automatically to all primary awardees to cover the activities and work product of themselves and their sub-awardees. Certificates prevent the disclosure, except under specific circumstances, of any identifiable, sensitive information collected or used during the program. These protections extend to copies of *All of Us* data and prevent disclosures of such information by anyone in guardianship or possession thereof. The program expects all awardees and sub-awardees, program partners, subcontractors, and vendors to use any and all legal measures at their disposal to fight legal demands for *All of Us* data protected by a Certificate of Confidentiality. Nevertheless, should such *All of Us* information be disclosed, either legally or illegally, the Certificate makes this information immune from the legal process, without consent of the individual to whom the information pertains.

Transparency and Participant Control

Members of the program are able to set preferences about when and how they receive information or are contacted by the program. They are also able to obtain copies of information held about them. Once enrolled, participants also have the right to withdraw from further participation and to have their information and specimens withdrawn from further use by the program, with some limitations.

Account Maintenance and Review

Each site sets qualifications for job functions, hires and trains qualified people, and assesses their competence in job tasks. To ensure that only authorized personnel are able to access the system, staff access to the system requires authorization from the site's PI or point of contact. Access can be revoked or updated as needed to accommodate transfer or termination (voluntary or involuntary). Upon departure from the program or the HPO, staff credentials are revoked and the DRC system administrator is notified. In addition, as added security, account review and maintenance takes place every six months.

Technical Measures

Researcher access is limited to the curated data repository, which is electronically scrubbed of explicit personal identifiers. Researchers wishing to access data are required to agree to our Data User Code of Conduct prior to use.

Data transferred to the DRC contains links to the participants and may contain identifying information, including health care providers' clinical notes. In all cases, data is transferred with encryption and kept on secure servers. The DRC aggregates the data from all sources to create a comprehensive record for each participant.

At the DRC, the health information collected is assigned to the participant by their participant ID, personal identifiers (e.g., names) are removed from this information for creation of the curated dataset. Personal identifiers are not attached to stored biological samples. Information linking the study codes to participants' identities is stored in a secure manner and is accessible to specific individuals overseeing this program, including those involved with securing the identity of participants.

Additional Security Measures at the DRC and in the Participant Portal

A myriad of security systems, protocols, rules, and practices to safeguard participants' information are being implemented and are documented in submissions to the appropriate authorizing bodies.

Both applications are bound by FISMA, which requires procedures, techniques, and processes for protecting data.

To limit the risks of deletion or tampering—whether accidental or malicious—all Participant Portal hosts and DRC administrative accounts are required to have multi-factor authentication configured prior to accessing resources. As described above, the Participant Portal hosts and DRC architectures use a defense-in-depth approach to protect against accidental and malicious risks from a variety of actors, including the principle of least privilege (POLP), so that users must be explicitly authorized to take any action affecting participant data and maintenance of auditable access logs.

The DRC

The DRC systems are restricted to use by system operators and qualified researchers, whose access is controlled, audited, and protected, using the security mechanisms described above.

The Participant Portal

The Participant Portal host applications also protect participant information by requiring secure passwords and verification of email information for participants with emails. This approach protects against hacking of user accounts and allows for password reset verification. In addition to forcing secure user passwords to prevent improper access, policies and procedures are in place to prevent the use of social engineering to access the system. The Support Center must verify multiple components of users' data to verify their identity, such as email address and phone number.

Data Security Incidents

Data security incidents or security vulnerabilities detected during intruder testing, as defined in our FISMA compliance documentation and policies, are reported to relevant parties at NIH program leadership and a Participant Data Protections and Incident Notification Board, who take further action as needed.

The Participant Data Protections and Incident Notifications Board is constituted as an expert committee to oversee *All of Us* Research Program responses to data security incidents and risks to participant privacy resulting from such incidents. The board's responsibilities, as described in additional documentation, do not include technical oversight (provisions and conditions specified by FISMA and the ATO) but instead involve program response in the event of a data security incident, as well as communication to participants of any resultant risk to their privacy.

The primary responsibilities of the Participant Data Protections and Incident Notifications Board are to serve as the body reviewing and recording security incidents and providing notifications to the IRB, to act as the arbiter for data breach liability, and to serve as the authority for determining whether a security incident requires notification of participants.

A reportable breach is any breach where data is exposed to unauthorized parties. If the Participant Data Protections and Incident Notifications Board, the IRB, and the program determine that a breach has occurred to the extent that participants should be notified, the Participant Portal hosts and the DRC work with all program partners as necessary to notify participants according to their preferred method of contact. The Participant Data Protections and Incident Notifications Board include members from the *All of Us* Research Program awardees; participant representatives; at least one individual with ethical, legal, and social issue expertise; at least one individual with privacy and security expertise; and NIH personnel.

Appendix G: Table of Participant Data and Specimen Collection Modalities

Collection Pathway	Eligibility for Collection	Description and Duration of Activity
In-Person Visit (HPO) *Preferred pathway of PM&B collection.*	● Determined by proximity to an *All of Us* collection site and participant preference.	<u>Physical Measurement Collection</u> ● Sit for 5 minutes ○ Conduct program core physical measurements, to include: ○ Blood pressure and heart rate ○ Height and weight ○ Hip and waist circumference ● 15-20 minutes <u>Biospecimen Collection</u> ● Perform blood draw ● Collect urine specimen ● Collect saliva sample (if needed) ● 10-45 minutes
Home Visit (HPO Mobile Units and ExamOne)	● Determined by available staff resources and scope of ISIA approval ● Based on participant preference, proximity to an *All of Us* collection site, and ability to present in-person	<u>Physical Measurement Collection</u> ● Sit for 5 minutes ○ Conduct program core physical measurements, to include: ○ Blood pressure and heart rate ○ Height and weight ○ Hip and waist circumference ● 15-20 minutes <u>Biospecimen Collection</u> ● Perform blood draw ● Collect urine specimen

		● Collect saliva sample (if needed, HPO Mobile Unit only) ● 10-45 minutes
MEA	● Determined by the Mobile Engagement Asset tour schedule and access to the vehicle by participants	<u>Physical Measurement Collection</u> ● Sit for 5 minutes ○ Conduct program core physical measurements, to include: ○ Blood pressure and heart rate ○ Height and weight ○ Hip and waist circumference ● 15-20 minutes <u>Biospecimen Collection</u> ● Perform blood draw ● Collect urine specimen ● 10-45 minutes
Blood Centers/ Blood Bank (stand-alone appointments)	● Determined by enrollment efforts of the blood center.	<u>Physical Measurement Collection</u> ● Sit for 5 minutes ○ Conduct program core physical measurements, to include: ○ Blood pressure and heart rate ○ Height and weight ○ Hip and waist circumference ● 15-20 minutes <u>Biospecimen Collection</u> ● Perform blood draw ● Collect urine specimen ● 10-45 minutes *Note: PM collection at a blood bank differs based on the participant's participation in the blood diversion pouch study. See row below.*

At-Home Saliva Kit (mailed or kit distribution in person)	● Determined by proximity to or ability to get to an HPO and participant preference	<u>Physical Measurement Collection</u> ● Remote Collection of Height and Weight ○ Participants click on activity tile to enter their height and weight within their Participant Portal ○ Participants choose between Imperial or Metric measurements ○ Participants enter their height and weight ○ ~5 minutes <u>Biospecimen Collection</u> ● Collect saliva sample ● 10 minutes
Quest BioKit	● Determined by proximity to or ability to get to an HPO and participant preference	<u>Physical Measurement Collection</u> ● Remote Collection of Height and Weight ○ Participants click on activity tile to enter their height and weight within their Participant Portal ○ Participants choose between Imperial or Metric measurements ○ Participants enter their height and weight ○ ~5 minutes <u>Biospecimen Collection</u> ● Perform blood draw ● Collect urine specimen ● 10-45 minutes
Blood Diversion Pouch (BDP)	● Determined by enrollment efforts of the blood center	<u>Physical Measurement Collection</u> ● Trained site staff collect some PMs on-site: Sit for 5 minutes ○ Conduct program core physical measurements, to include: ○ Blood pressure and heart rate

		● Remote Collection of Height and Weight ○ Participants click on activity tile to enter their height and weight within their Participant Portal ○ Participants choose between Imperial or Metric measurements ○ Participants enter their height and weight ○ ~5 minutes <u>Biospecimen Collection</u> ● Perform blood draw ● Collect urine specimen ● 10-45 minutes

National Institutes of Health

AllofUs.nih.gov

ResearchAllofUs.org