





Pioneering Precision: Strategic Insights for the Growth and Sustainability of the *All of Us* Research Program

September 17, 2024

All of Us Reflection and Advancement Working Group Report to the *All of Us* Research Program Advisory Panel

Working Group Members

Russ Altman, M.D., Ph.D., Working Group Chair
Stanford University
Kenneth Fong Professor of Bioengineering, Genetics, Medicine, Biomedical Data Science and (by courtesy) Computer Science

Marylyn DeRiggi Ritchie, Ph.D., M.S., Working Group Vice Chair
University of Pennsylvania
Edward Rose, M.D. and Elizabeth Kirk Rose, M.D. Professor of Genetics
Vice President for Research Informatics
Director of the Institute for Biomedical Informatics
Director for the Division of Bioinformatics, Department of Biostatistics
Vice Dean of Artificial Intelligence and Computing

Rhonda Robinson Beale, M.D., Working Group Vice Chair
UnitedHealth Group
Senior Vice President, Mental Health Services
Deputy Chief Medical Officer, Mental Health Services

Aaron Abend, M.B.A
Autoimmune Registry, Inc.
Executive Director

Naomi Allen, D.Phil., M.Sc., B.Sc.
UK Biobank
Chief Scientist
University of Oxford
Professor of Epidemiology, Nuffield Department of Population Health

Wendy Chung, M.D., Ph.D.
Boston Children's Hospital
Chief of the Department of Pediatrics
Harvard Medical School
Professor of Pediatrics

Gary Miller, Ph.D., M.S.
Columbia University Mailman School of Public Health
Vice Dean, Research Strategy and Innovation

Christopher O'Donnell, M.D., M.P.H., FAHA, FACC
Cardiovascular & Metabolism at Novartis Institutes for Biomedical Research
Global Head of Translational Medicine

Elizabeth O. Ofili, M.D., M.P.H., FACC
Morehouse School of Medicine
Professor of Medicine
Morehouse Healthcare
Practicing Cardiologist

Evelyn Ortiz
New York City Housing Authority, Coordinator (Retired)
Health Equity Advocate
Participant Ambassador, *All of Us* Research Program

Jesus Ramirez-Valles, Ph.D., M.P.H.
University of California San Francisco
Chief of the University of California San Francisco Division of Prevention Science

Scout, Ph.D., M.A.
National LGBT Cancer Network
Executive Director

Prashant Shah, M.S.
Intel Corporation
Director of Engineering of Health and Life Sciences
Artificial Intelligence Products Group

Hannah Valantine, M.D., M.R.C.P.
Stanford University
Professor of Medicine
National Institutes of Health
Former Chief Officer for Scientific Workforce Diversity

Roberto Vargas, M.D., M.P.H.
Charles R. Drew University of Medicine and Science
Assistant Dean, Health Policy and Inter-Professional Education, College of Medicine
Charles R. Drew University of Medicine and Science
Director, Health Policy Pillar of the Urban Health Institute

Xiaobin Wang, M.D., Sc.D., M.P.H.
Johns Hopkins Bloomberg School of Public Health and School of Medicine
Zanvyl Krieger Professor in Children’s Health
Professor of Pediatrics
Director, Center on the Early Life Origins of Disease

Ex Officio Members

Sadiqa Mahmood, D.D.S., M.P.H.
Health Catalyst
General Manager and Senior Vice President

Josh Mandel, M.D.
Harvard Medical School
Lecturer on Biomedical Informatics
Chief Architect, SMART Health IT
Chief Architect, Microsoft Healthcare

Elizabeth Rubinstein
Patient Advisor, Henry Ford Health System Board of Directors Quality and Safety Committee
Advisor, *All of Us* TACH Organ Procurement Organization
Former Participant Ambassador, *All of Us* Research Program

Karen Wall, Ed.D., M.A.
U.S. Army Nurse Corps
Major, Retired (Operation Desert Storm Combat Veteran)
Western Institute for Social Research
Graduate Professor, Marriage and Family Therapy
Former Participant Ambassador, *All of Us* Research Program

All of Us Research Program Support Members

Rachele Peterson, M.S., M.B.A.
National Institutes of Health
Chief of Staff

Hannah Coleman, M.S.W.
Axle Informatics
Special Assistant to the Chief of Staff

Victoria Palacios, M.P.H.
Leidos
Biomedical Life Scientist

Table of Contents

Executive Summary

The *All of Us* Research Program (*All of Us*) was conceived in 2015, issued its first awards in 2016, and launched nationally in 2018. It has made tremendous strides in creating a cohort of people across the United States who have contributed information about their health, environment, and lifestyle to spur scientific discovery, with the goal of translating new knowledge into more precise and more equitable medicine. It has already catalyzed discoveries (and associated publications) across a wide range of disciplines according to an *All of Us* year in review published in the *American Journal of Human Genetics*.

A critical and unprecedented strategy that contributed to the creation and success of the cohort was funding key national community engagement partners and making engagement a priority for every awardee—even technology, data, and enrollment awards. The *All of Us* Reflection and Advancement Working Group was charged with focusing on eight questions that span four areas in the report. The key findings can be summarized here:

1. The diversity of the *All of Us* cohort—unmatched by any other large cohort in existence—is its primary strength. It is critical to nurture the relationships that helped build and maintain this engaged and diverse cohort, so communication, transparency, and effective engagement with participants in asking and answering scientific questions—as well as demonstrating trustworthiness—are absolutely critical for future success.

2. As of August 2024, the cohort includes 834,000 consented participants—more than 566,000 of whom have completed all initial steps—and is strong. But enrollment must continue in order to achieve the initial goal of at least one million participants, with attention to inclusion across all U.S. populations—especially those historically underrepresented in biomedical research (UBR). Such diversity is essential to ensure the scientific and statistical power necessary to generate meaningful insights into both common and rare diseases, ultimately advancing health equity and reducing disparities across health-related outcomes. It is important that the participants engage and contribute actively, taking pride in their participation and the biomedical discoveries that emerge from their data.

3. The inclusion of a pediatric cohort is a critical and unique opportunity. Ideally, recruitment of children will be linked to family members in the cohort (to observe generational effects, for example). The opportunity to observe the evolution of health throughout the full life of an individual will provide insight of profound value in creating a health care system that optimizes health across the lifespan.

4. The platform for intake and dissemination of diverse data is functional and robust—and will be maintained, secured, and extended for decades of use. The *All of Us* leadership will need to make "buy versus build" decisions for all aspects of the information

technology infrastructure. Critical to this will be the ability to have competitive awards, with terms that allow the resource to evolve its partners when necessary.

5. The diversity of the cohort should continue to be matched by a diversity of scientific users—from academia, private industry, government, and foundations, and across many scientific disciplines—in order to guarantee that the questions answered with the resource reflect the needs and priorities of many.

6. The long-term and sustained trust of the participants and community leaders of the program is critical for the long-term viability of the resource. This trust will stem from high-quality and relevant communications and transparency that maintain the program's trustworthiness. It is equally important to engage participants in new research projects, so that they have a continued sense of pride in and ownership of the fruits of the research and the impact of *All of Us*.

7. *All of Us* is becoming a primary national resource for translational research—the process of moving discoveries from the lab to clinical benefits to patients—thus complementing other initiatives across NIH. *All of Us* has already started to provide sequencing data to researchers and, through DNA results, individual information to improve the health of participants; it should be the central resource of expertise and resources for learning to effectively and efficiently translate scientific knowledge into practical biomedical and health practices and interventions. In addition, it provides a setting that enables foundational discoveries in genomics, lifestyle, and environment that are immediately relevant and translatable to clinical care through its large and national population. The goal of building an equitable health care system that delivers biomedical innovation to all populations will benefit from the research conducted with *All of Us* data and participants.

8. *All of Us* has an opportunity to engage partners and collaborators who can use the resources of the program to conduct ancillary, supplemental, and/or collaborative studies that enrich *All of Us* while reducing the direct cost to the taxpayer. For example, many datasets of interest to the private sector may be obtained in partnership with companies that pay for the data, thus making the resource more potent at a low cost (or no cost). Furthermore, NIH Institutes and Centers will be able to conduct trials more rapidly and at a fraction of current costs, leveraging the cohort of previously recruited and studied participants who are anxious to help. Other partners will bring questions, approaches, and data that enrich *All of Us* in unexpected ways and thus ensure its relevance for decades to come.

9. *All of Us* has the opportunity to develop sustainable and cost-effective practices that are compatible with its anticipated lifetime of several decades. Decision processes can be streamlined and unencumbered by bureaucracy; decisions should be consistent with the long-term maintenance of the resource. *All of Us* will evolve partnerships as the country's

needs and the fields of science and health change over time. The *All of Us* leadership will pursue collaborations and partnerships that are cost-effective. Staff size can be managed to ensure sustainability and the ability to weather periods of fiscal tightening. If stable funding is not ensured regardless of source, the program may not achieve its long-term intended outcomes for the country.

10. *All of Us* can become a research resource for responding to national health emergencies. It has already demonstrated this capability in the early research efforts related to COVID-19. With appropriate resources, the program can extend this capability so that it becomes an essential research component of national emergency response efforts. Even before emergencies, *All of Us* can engage in multi-institutional and multi-agency research projects that test its ability to participate in and help coordinate complex efforts to address fundamental research questions raised by critical health challenges, such as the opioid epidemic, mental health, cancer, and dementia.

Overview of the *All of Us* Research Program

Precision medicine—the use of individual variability in genes, environment, and lifestyle to inform disease treatment and prevention—offers a powerful model for continuously improving human health. It has the potential to influence all levels of intervention (i.e., screening, diagnosis, prognosis, and treatment) and reduce disease burden. To enable precision medicine, large numbers of participants and researchers, representing a wide range of experiences and perspectives, are needed to advance our fundamental understanding of health and disease. Novel therapeutics and diagnostics will follow from this understanding.

The *All of Us* Research Program was established following President Barack Obama's 2015 announcement of the Precision Medicine Initiative® (PMI). Guided by scientific opportunities and the imperative of addressing health disparities, the program deliberately focused on developing relationships with partner organizations and engaging communities historically underrepresented in biomedical research (UBR). Since its launch, the program has achieved important milestones (Figure 1) that have stemmed from a commitment to a future of precision health accessible to all populations. *All of Us* is committed to advancing health equity and ensuring that broad and diverse populations have the opportunity to contribute to and benefit from research, either by donating data as participants and/or by using the data as researchers.

Figure 1: Progress Toward *All of Us'* Five-Year Goals (2021–2026)
Description: This infographic outlines the program's key achievements from its launch through August 2024. The accomplishments are organized according to the program's five strategic goals, which guide its evolution, chart its future through 2026, and position *All of Us* as a critical resource for advancing health equity and health research.

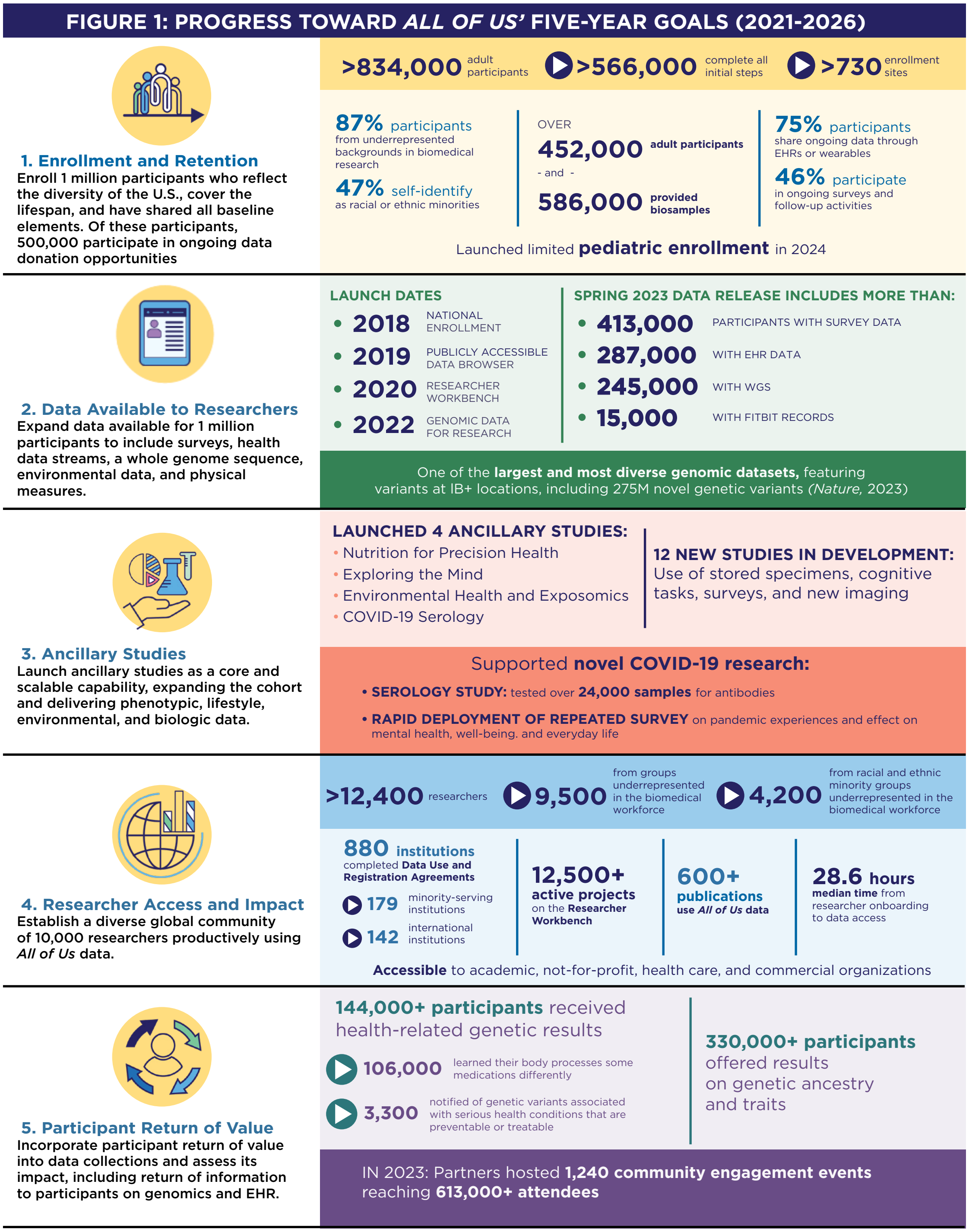


FIGURE 1: PROGRESS TOWARD *ALL OF US'* FIVE-YEAR GOALS (2021-2026)

1. Enrollment and Retention

Enroll 1 million participants who reflect the diversity of the U.S., cover the lifespan, and have shared all baseline elements. Of these participants, 500,000 participate in ongoing data donation opportunities

>834,000 adult participants ▶ **>566,000** complete all initial steps ▶ **>730** enrollment sites

87% participants from underrepresented backgrounds in biomedical research

47% self-identify as racial or ethnic minorities

OVER **452,000** **adult participants** - and - **586,000** **provided biosamples**

75% participants share ongoing data through EHRs or wearables

46% participate in ongoing surveys and follow-up activities

Launched limited **pediatric enrollment** in 2024

2. Data Available to Researchers

Expand data available for 1 million participants to include surveys, health data streams, a whole genome sequence, environmental data, and physical measures.

LAUNCH DATES

2018	NATIONAL ENROLLMENT
2019	PUBLICLY ACCESSIBLE DATA BROWSER
2020	RESEARCHER WORKBENCH
2022	GENOMIC DATA FOR RESEARCH

SPRING 2023 DATA RELEASE INCLUDES MORE THAN:

413,000	PARTICIPANTS WITH SURVEY DATA
287,000	WITH EHR DATA
245,000	WITH WGS
15,000	WITH FITBIT RECORDS

One of the **largest and most diverse genomic datasets,** featuring variants at IB+ locations, including 275M novel genetic variants (*Nature,* 2023)

3. Ancillary Studies

Launch ancillary studies as a core and scalable capability, expanding the cohort and delivering phenotypic, lifestyle, environmental, and biologic data.

LAUNCHED 4 ANCILLARY STUDIES:
- Nutrition for Precision Health
- Exploring the Mind
- Environmental Health and Exposomics
- COVID-19 Serology

12 NEW STUDIES IN DEVELOPMENT: Use of stored specimens, cognitive tasks, surveys, and new imaging

Supported **novel COVID-19 research:**
- **SEROLOGY STUDY:** tested over **24,000 samples** for antibodies
- **RAPID DEPLOYMENT OF REPEATED SURVEY** on pandemic experiences and effect on mental health, well-being. and everyday life

4. Researcher Access and Impact

Establish a diverse global community of 10,000 researchers productively using *All of Us* data.

>12,400 researchers ▶ **9,500** from groups underrepresented in the biomedical workforce ▶ **4,200** from racial and ethnic minority groups underrepresented in the biomedical workforce

880 institutions completed **Data Use and Registration Agreements**
▶ **179** minority-serving institutions
▶ **142** international institutions

12,500+ active projects on the **Researcher Workbench**

600+ publications use ***All of Us*** **data**

28.6 hours **median time** from researcher onboarding to data access

Accessible to academic, not-for-profit, health care, and commercial organizations

5. Participant Return of Value

Incorporate participant return of value into data collections and assess its impact, including return of information to participants on genomics and EHR.

144,000+ participants received health-related genetic results
▶ **106,000** learned their body processes some medications differently
▶ **3,300** notified of genetic variants associated with serious health conditions that are preventable or treatable

330,000+ participants offered results on genetic ancestry and traits

IN 2023: Partners hosted **1,240 community engagement events** reaching **613,000+ attendees**

All of Us has established a centralized, cloud-based open platform that houses data from diverse participants and supports the study of health across the lifespan and many different types of disease. The platform, called the Research Hub, contains the *All of Us* dataset and provides tools that allow anyone to begin exploring the types and quantities of data being collected by the program. Researchers can further use the diverse data to conduct studies by registering to use the cloud-based Researcher Workbench. Once they complete registration, researchers can create collaborative research projects, build research cohorts, and analyze data.

The Research Hub benefits from the experiences of other cohort studies such as UK Biobank, the Million Veteran Program, and FinnGen. These cohorts use ongoing data collection, including repeated measurements and longitudinal outcome assessments, to supply data to researchers for longitudinal study of disease. The Researcher Workbench innovated by creating a cloud-based straightforward access model (the passport model) that has dramatically accelerated the creation of a diverse researcher community.

Purpose and Activities of the *All of Us* Reflection and Advancement Working Group

Charge

The *All of Us* Research Program Reflection and Advancement Working Group, established in April 2024, comprises select individuals from the *All of Us* Research Program Advisory Panel, as well as external voices and experts in biomedical research. *All of Us* Chief Executive Officer Dr. Josh Denny charged the working group with assessing the program's evolution, evaluating achievements and challenges since the national launch in 2018, and identifying opportunities for improvement and advancement. The overall aim of this working group is to shape the program's strategic plan toward maximizing its scientific impact and trajectory leading up to 2031.

Objective

This report reflects on how the program progressed from its initial vision—outlined in the 2015 Precision Medicine Initiative Working Group Report to the Advisory Committee to the NIH Director (ACD)—to subsequent milestones: the issuance of initial awards, in 2016; the national launch of participant enrollment, in 2018; the first data release, in 2020; and the establishment of the program's strategic goals, in 2021. This effort assessed achievements and challenges, areas for continued improvement, and opportunities for maximizing the program's scientific impact and trajectory. Furthermore, the working group explored the broader impact of *All of Us* on the research ecosystem, diversity of research, community engagement, biomedical workforce development, and advances in scientific knowledge. We provided an authentic assessment that ensured the achievements were noted, challenges identified constructively, and opportunities called out with equal consideration.

Since the working group's kickoff meeting, on April 17, 2024, and through a series of four subsequent input sessions, the report writers have gathered insights from *All of Us* Advisory Panel members, program leadership, subject matter experts, and external stakeholders. This report lays the foundation for shaping the future of the *All of Us* Research Program, with a forward-looking vision focused on maximizing the program's impact over the next decade.

Charting Progress: Key Metrics and Program Differentiators

1. What are the most important differentiators of the program? What is *All of Us* uniquely positioned to achieve, or should be?

2. Given the vision outlined in the 2015 ACD Working Group Report, which key indicators or metrics best reflect the program's progress and success thus far?

In the 2015 ACD Working Group Report, the committee suggested that the program (now known as *All of Us*) "will build a large research cohort of one million or more Americans that will provide the platform for expanding our knowledge of precision medicine approaches and that will benefit the nation for many years to come." The recommendations in the ACD report included (1) recruitment of a diverse cohort of one million Americans, such that the cohort reflects the diversity of the United States; (2) long-term engagement with participants in terms of data collection, as well as soliciting feedback and input during the planning and implementation of the research program; and (3) collection of diverse streams of data and development of accessible, secure platforms for data access. Now is a good time to reflect on the success and progress of the program thus far.

First, evaluation of enrollment metrics is a critically important indicator in reviewing the progress of *All of Us*. The program's enrollment framework has efficiently (meaning both more rapidly and less expensively) reached more than 834,000 diverse participants through a nationwide network of more than 180 consortium partners. Through these partnerships, 730 sites have supported the collection of samples and measurements. In terms of the demographic characteristics of the cohort, participants are incredibly diverse across multiple domains. *All of Us* has a special focus on engaging participants from populations that have been historically underrepresented in biomedical research, as characterized by race and ethnicity (47%), health care access and utilization (28.1%), age at date of consent (26.4%), income (35.9%), disability status (20.8%), gender identity (1.3%), sex assigned at birth (0.06%), sexual orientation (10.1%), educational attainment (8.6%), and geography (7.8%). Many other research programs focus on one or two global populations; *All of Us* is unique in having such breadth and depth of diversity in its participant community.

From this diverse cohort, *All of Us* collects longitudinal health and wellness data (including valuable biospecimens that can be studied and restudied as capabilities evolve), generates whole genome sequence (WGS) data, and obtains survey data, including information on social determinants of health (SDoH). More than 413,000 participants have contributed data in the program's spring 2023 release. At this point, *All of Us* is well positioned to meet its recruitment goals with a cohort of participants that reflects the diversity of the United States. This will enable researchers to understand the basis of health and wellness, holistic risk factors, and the origin and trajectories of disease, and to develop novel biomarkers and establish polygenic risk scores. For example, the fact that *All of Us* is collecting SDoH data is a huge benefit of the program, given that very few large cohorts collect these data. Thus *All of Us* data can help drive novel research that will affect health policies and clinical guidelines by combining data that were previously difficult to combine at a large scale. In one case, guidelines on the use of genetic information (on the *DPYD* gene) to guide chemotherapy were updated based on data in *All of Us*.

If we view this progress in diversity of enrollment through one lens, such as women's health in the United States, it is notable that *All of Us* has far surpassed the enrollment in the Women's Health Initiative—one of the largest women's health projects ever launched in the country,—in which 161,000 women participated. By comparison, *All of Us* has data from nearly 250,000 women currently available to researchers from the spring 2023 data release. In addition to electronic health record data on conditions unique to women, the program also has information on women's health from participant surveys. The magnitude of these data makes it possible for researchers to quickly organize subsets of female participants based on traits, diagnoses and conditions, genetics, menopausal status, and more; study correlations and quantify the prevalence of diseases among women; and inform screening, prevention, and personalized care. Researchers can use these data to conduct a rapid genome-wide association study (GWAS) and a phenome-wide association study (PheWAS) to study genetic causes of conditions that disproportionately affect women; identify non-genetic risk factors; and study drug response, efficacy, and adverse effects among women. It would be feasible to establish an intramural team of NIH researchers to use *All of Us* data and study priority questions in women's health research, and for NIH to organize extramural research teams to study or validate findings using *All of Us* data. In addition, the program could administer a comprehensive women's health survey regarding sexual and reproductive health history, menstruation, the age of menarche, and more to further build the research value of this dataset for advancing women's health.

Second, articulating metrics of success in participant engagement is another important indicator. The program has developed a unique infrastructure for participant engagement. It involves governance models that include participant voices at all levels of operation, and communication and outreach strategies that are enabling longitudinal retention of diverse participants. Their success can be evaluated both quantitatively and qualitatively. Quantitatively, participants have completed retention activities, providing researchers with varied information, including surveys from more than 413,000 participants, electronic health records (EHRs) from more than 287,000 participants, physical measurements from more than 337,000 participants, and 15,000 Fitbit records, as of the spring 2023 data release. Many participants are using *All of Us* support services for assistance in understanding genetics, answering surveys, and providing biospecimens. Qualitatively, we can look at the types of engagement activities at partner sites, including health care provider organizations and community engagement partners. Many partner sites have hosted events to engage with diverse participants. Additionally, *All of Us* has created targeted strategies to build trust in science and health communications with participants. Staff and participants can describe the quality of their interactions and observe other positive impacts of the program.

All of Us can continue to monitor and develop metrics of evaluation regarding the return of value to participants. The original ACD report envisioned the program eventually returning results to participants. The program has begun to do this, and continues to expand the set of clinically relevant pharmacogenomic and genomic results that will be offered to participants. For example, 1,000 to 5,000 participants per week have been invited to decide whether they want their health-related DNA results, according to a study on *All of Us* participant perspectives on the return of value in research. As a result, many participants have chosen to receive individual reports based on the sequencing of their genome that have direct relevance to their health and wellness. Indeed, approximately 3% of participants (more than 3,300) who received their hereditary disease risk report have a potentially life-changing genetic variant, and about 90% of participants (more than 106,000) had a result in their pharmacogenetics report that could inform how their body processes one or more medications. Continuing to capture qualitative measures of participant satisfaction with this return-of-results process would be beneficial.

Third, evaluation of the collection of diverse streams of data is an essential component of measuring the impact of the program. The variety of data in *All of Us* is unprecedented. Core data types collected include surveys, EHRs, biosamples (blood, saliva, and urine), physical measurements, digital health data, whole genome sequences, and genotype arrays. Not only is the breadth of data types well beyond what most studies include; *All of Us* is also one of the most diverse cohorts in existence, and thus the only one with this rich blend of accessible data across diverse participants. As such, the coupling of SDoH, EHR, and genomic data across diverse participant populations would be difficult, if not impossible, at the scale of *All of Us* in any other dataset. There is no broadly accessible cohort in existence with the depth of diversity and sequence data that *All of Us* has. Furthermore, the program has the potential to continue to diversify and improve genomic studies worldwide. Finally, *All of Us* has developed processes for obtaining new data types through novel awards such as the Center for Linkage and Acquisition of Data as well as partnered ancillary studies.

Another axis of diversity in the data that can be assessed is the clinical diversity of diseases and conditions documented in the cohort (Table 1). The *All of Us* enrollment strategy and broad inclusion criteria means that participants from across the full spectrum of health and diseases are recruited into the program. While many other cohorts are disease-specific, *All of Us* is open to all, and thus will have an unprecedented variety of diseases—particularly across UBR groups.

Table 1: Common U.S. Causes of Death and Prevalence Among *All of Us* Participants
Description: This table presents the number of *All of Us* participants affected by the common causes of death in the United States, as of the April 2023 curated data release into the Researcher Workbench. It also highlights the percentage of participants affected by each condition who are from communities that are UBR. The top five diseases—ischemic heart disease, lung cancer, chronic obstructive pulmonary disease (COPD), Alzheimer's disease, and colorectal cancer—were identified according to the Global Burden of Disease Study 2019.

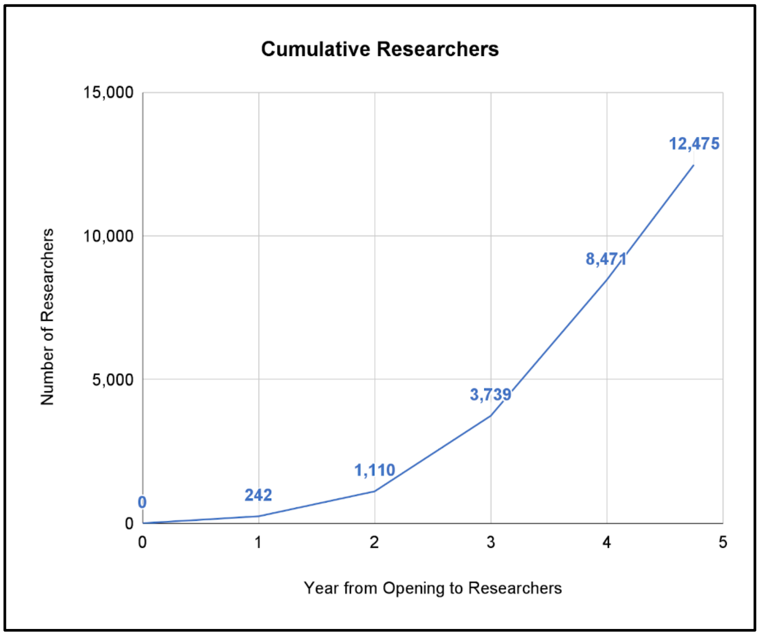
Condition	# of *All of Us* Participants	% of UBR Participants	% of UBR by Race/Ethnicity
Ischemic heart disease (IHD)	47,991	91.9%	36.7%
Chronic obstructive pulmonary disease (COPD)	25,021	94.7%	40.1%
Colon and rectum cancer	6,550	87.8%	30.0%
Alzheimer's disease and other dementias	5,713	95.3%	38.1%
Tracheal, bronchus, and lung cancer	2,761	92.0%	29.7%

Biospecimen data (blood, urine, and other samples) are particularly valuable because they can be saved and evaluated in the future when new tests are created that can detect molecules and pathogens that are today unknown. During the COVID-19 pandemic, *All of Us* tested samples provided by participants in early 2020 and found evidence that there were SARS-CoV-2 infections in several states earlier than had initially been reported.

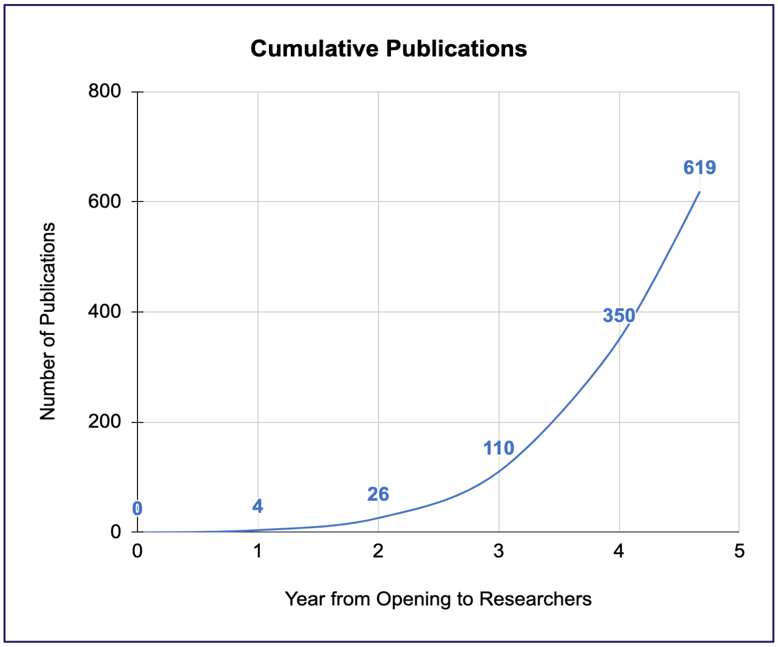
The diversity of the cohort will also help improve the application of artificial intelligence (AI) to population data. AI is a critical application area for health, and *All of Us* is positioned to ensure that AI applications in health care are of high quality. Health care AI models that are used for clinical decision support often suffer from bias because of a lack of diverse real-world data; some populations may not be present in the data used to train the AI system. This can lead to an inequitable allocation of resources. *All of Us* is uniquely positioned to help minimize bias in health care AI models and improve real-world model performance (generalizability).

Fourth, evaluating the development of accessible, secure platforms for data access must be a key indicator of success; if the *All of Us* Research Program generates an exceptional precision health cohort in the world, but researchers are unable to access it for research in a safe, secure manner, then the program's impact, the donations of participants, and our country's investment will never be fully realized. To this end, the program has prioritized making data readily accessible to researchers. *All of Us* decided on an early data access model—minimizing the time from data collection to research availability. This deliberate choice greatly facilitated researcher access to the data and accelerated the pace of discoveries. Since 2019, the program has released seven versions of data. Quantitative metrics of the program's success in data access by researchers include the following: 12,400 researchers worldwide, from 880 institutions, are active in the Researcher Workbench. Of these, 9,500 researchers are from groups that are underrepresented in the biomedical workforce. As of August 2024, more than 12,500 projects are in progress in the Researcher Workbench. In terms of scientific impact, 619 publications are now using *All of U*s data, often combining multiple data types for an integrated analysis (Figure 2).

Figure 2: Growth of *All of Us* Researchers and Publications Using *All of Us* Data
Description: This figure includes two charts: a line graph showing the growth of registered researchers on the *All of Us* Researcher Workbench from January 2020 to August 26, 2024, and another illustrating the number of publications using *All of Us* data from January 2020 to August 26, 2024. Together, these charts highlight the expanded use of the program's data and the increasing scientific discovery over time.



Data from DRC Analytics Hub, January 2020–August 26, 2024



Data from [CDC All of Us Publications Database](): January 2020–August 26, 2024

Monitoring these metrics as well as new ones that become important will be vital to continue as the program continues to grow in its participant membership, data assets, and researcher community. For example, continuing to monitor the number of trainees and early-stage investigator (ESI) researchers using *All of Us* data is critical in the long term. As more researchers become familiar with the platform, they will incorporate the *All of Us* cohort data into their research efforts, and all of their trainees will learn to use the program data. This will create a new generation of researchers for whom *All of Us* is a basic research infrastructure upon which they can always rely. Thus, the trainee and ESI research base should grow over time.

Similarly, as more researchers use the data from *All of Us*, we should see more research grants using *All of Us* data and, ultimately, the volume of discoveries (and associated publications and patents) grow at a comparable rate—an excellent ongoing indication of data quality and impact. Finally, as *All of Us* is beginning to enable data access to commercial entities, we will be able to count the number of startup companies using its data, the Small Business Innovation Research (SBIR) and Small Business Technology Transfer (STTR) program applications including its data, and the commercial researchers using its data for drug discovery programs.

All of Us has been intentional and creative about engaging a diverse researcher community to access and use the data for research purposes. By enlisting a diversified biomedical research workforce, *All of Us* is uniquely positioned to have a high impact on advancing health equity and increasing the likelihood that research outcomes benefit everyone.

Cultivating Established Trust: Future Growth Opportunities for *All of Us*

3. Where are there opportunities for growth or areas for improvement within the program?

The initial eight years of *All of Us* have established the core infrastructure for the program, so it is critical now to consider the prospective priorities and future growth areas. This is particularly important during a time of fiscal uncertainties, in order to maximize the value gained from investments. The working group identified several areas for future growth.

First, although the cohort is sizable—it has 834,000 participants, more than 566,000 of whom have completed all initial steps of the protocol—there is still the original goal of one million participants. These should not be simply one million participants who have signed up but should include one million or more "active" participants who are engaged in the resource, contributing data used for discoveries. There are many ways to define "active," and *All of Us* should consider defining useful levels of activity so that it can monitor the numbers of participants at each level. *All of Us* can seek to (1) continue enrollment to reach one million active participants and (2) ensure the full implementation of the plans for communication and retention so that the cohort maintains its size and activity over the coming decades. There are good baseline estimates from UK Biobank and other programs about the level of participation that can be expected, though these longitudinal cohorts are composed of different populations and from different health systems. Given the diversity of the *All of Us* cohort and the unlikeliness that one size will fit all, it is important to have a variety of mechanisms and channels for communication and retention. Thus basic growth and maintenance of the cohort are critical. The participants signed up for *All of Us* with the expectation of continued participation and, as outlined in the consent, the expectation that the program would last for at least 10 years. So there is a risk in not actively engaging them in research programs and not allowing them to clearly perceive that their participation is generating knowledge and making a difference.

Second, also in relation to the size of the *All of Us* cohort, there is the possibility of adding robust pediatric representation. The importance of children's health for the long-term health of the United States—and the world—cannot be overstated. The program appropriately and thoughtfully began enrollment with adult participants, with plans to expand to pediatric age groups (discussed in more detail in response to Question 6). Given the experience of the program and the acceleration of pediatric activities in 2023, including development and implementation of an institutional review board–approved protocol to enroll the youngest age group, there is now a chance to complete planning for longitudinal pediatric participation and ensure sustainable support for pediatric engagement, enrollment, and retention goals (for example, by focusing on prospective pediatric participants related to existing cohort members). Importantly, the pediatric subset of the *All of Us* cohort needs to be of sufficient size to support precision medicine research. This may lead to a total cohort size that is greater than one million.

Third, it is critical for *All of Us* to create an infrastructure for rich and varied ancillary studies that can serve as a platform for supporting many research activities. As the *All of Us* cohort matures and has established baseline data, it becomes an extremely attractive target for use in a wide variety of studies, across NIH and for the external research community. These studies may include collaborations where additional measures are gathered on part or all of the cohort in the development of medical knowledge, products, and services. It would also include creating opportunities for participants to take part in clinical trials that span intramural and extramural research, which could provide a huge cost savings and, at the same time, reduce the need to recruit entirely new cohorts that are studied once and then disbanded. The financial pressures on *All of Us* and on NIH (which sponsors many of these studies) may make this approach financially attractive. Ongoing *All of Us* partnered ancillary studies include Nutrition for Precision Health, Exploring the Mind, and Environmental Health and Exposomics, with the COVID-19 Serology Study already completed.

Presumably, the additional data would be paid for by collaborators (industry; NIH Institutes, Centers, and Offices; and so on), but would ultimately end up stored and available to *All of Us* researchers. However, to support these studies, *All of Us* can create a nimble process to evaluate opportunities—costs in terms of participant time/effort, the use and consumption of limited biosamples, staff time, and the value of data to the cohort over the long run. One advantage of these studies will be a more active engagement of the *All of Us* cohort, which will allow participants to feel more involved and critical to the enterprise. It will also allow research to begin on the best way to translate findings into meaningful clinical interventions. Building out the ability to collaborate with external entities in such ancillary studies can be a high-priority activity for *All of Us* that would enable growth of its scientific value and reduce costs to both industry and NIH research studies.

Fourth, *All of Us* has an opportunity to help bridge research and clinical implementation. The "learning health care system" has been a vision for medical research for more than a decade; it is based on the idea that every patient can and should generate data and knowledge that are used to improve the care of future patients. That has proven difficult to implement in practice, but *All of Us* has the scale and data to model how such a system can work. A critical first step has been the genetic return of results to participants, in which the genomic information from DNA sequencing has been processed and the clinical implications summarized and provided to participants with genetic counseling. They are then able to follow up with their physicians about any worrisome findings. This effort has become a laboratory for learning how to analyze the genome for critical findings, deliver this information in an understandable way to participants and clinical providers, and measure the impact of this information on health. The answers to these critical research questions will inform the entire health care system in the most impactful and cost-effective method for linking research findings to patient health.

Finally, core to the *All of Us* Research Program are the quality of data and productivity of researchers. The quality of data fuels everything: It allows subsets of the cohort to be identified for ancillary studies and researchers to creatively combine information to ask and answer new scientific questions. It also establishes standards that can be promulgated outside of *All of Us* to promote more uniform and consistent data collection across all sectors.

As the complexity of data increases with the addition of new types of data from both internal efforts and as part of collaborations, *All of Us* is developing methods for incorporating, documenting, and linking the data to enable research. One of the most important activities for *All of Us* is ensuring that the effective size of the cohort—the number of participants with data sufficiently complete to allow for routine use in research studies—remains high, maintained around one million or greater with children included. There should not be a rapidly growing "ghost" cohort of participants whose level of data does not support routine use in research. Stable, predictable funding for engaging these participants regularly through ancillary studies and ongoing data collection is critical to avoid squandering this investment and infrastructure.

At the same time, it is also critical that *All of Us* continues to support researcher productivity by ensuring that the scientific workforce is prepared to use its data. This includes well-organized and up-to-date training materials and documentation for the entire resource, adequate computing power (at reasonable prices) to allow analysis of the data using the latest algorithms, and continued outreach to a wide range of scientists in disciplines relevant to scientific discovery and clinical translation.

4. How can the program amplify the trust established among diverse participants, expanding their engagement and leveraging their involvement and voices?

Participants in *All of Us* have volunteered to provide their personal health information over time, as their health evolves, to support biomedical discovery. These generous participants are motivated by many goals but share an altruistic interest in improving health care through research. At the same time, they have expectations that (1) their privacy will be protected to the greatest extent possible, (2) they will be invited to learn about personal findings that may have an impact on their health, (3) they will receive regular communications about the general scientific progress made by *All of Us*, and (4) they will be notified about additional opportunities to share data or take part in other research studies associated with *All of Us*. Fundamental to all of these activities is the requirement that *All of Us* maintain the trust of participants established through engagement and the comprehensive consent, permission, and assent processes.

Failure to maintain trust could lead to an erosion of the size and quality of the cohort. This could ultimately threaten the utility of the resource. It could also undermine efforts to address historical health disparities and efforts to restore trust in biomedical research. Community partners—who serve as liaisons, ambassadors, and translators between *All of Us* and community members who participate in *All of Us*—are a critical resource for recruitment and retention. Participant partners are individual participants who support the program's design, implementation, and governance—providing critical voices to monitor and improve activities that build trust—including support of the scientific goals of the resource. All of these partners are critical because they help ensure that *All of Us* serves all communities, populations, and voices across the United States and considers their questions, concerns, and priorities. The working group identified several ways to ensure that the existing methods for establishing and maintaining trust are amplified.

First, when it comes to the *All of Us* cohort, it is important for participants to be offered continuing opportunities and information about ongoing studies—including options to participate in new studies. This will occur, for example, with the expansion of ancillary studies. It is critical that participants self-identify as "*All of Us* participants" and carry this identity with pride. Thus communication programs with participants can be robust, active, and purposeful. Ideally, *All of Us* will continue to meet with participants in their communities and listen to their feedback, and the program could develop metrics to track not only individual participants but also communities. Communication should be through a variety of media, to serve the preferences of participants, including face-to-face events whenever feasible, as well as social media, chats, emails, webinars, radio, television, and so on. During these communications, there must continue to be transparency about what data are being collected, how data are being used, what scientific progress has been enabled, and how clinical translation will work to bring the fruits of the research to the health of the population. This may also include frank discussions about *All of Us* financial resources, how priorities are set, and why choices are made. The program is piloting such virtual participant convenings now. The return of results—as for the genomic information—to participants should be continued so that it becomes an ongoing source of renewal for participant motivation, understanding, and support for the program.

Second, with the development of collaborative and ancillary studies, *All of Us* will establish methods for communicating the scientific value of these efforts and the ways in which participant trust can be extended to the collaborating organizations. Thus there will be transparent disclosure of (1) the scientific goals of the collaborations and the program values and principles that drive the selection of such offerings; (2) how private companies, nongovernmental organizations, or even other NIH Institutes and Centers will be collecting and using the data; (3) where the funding comes from and what benefits might accrue to funders (e.g., financial arrangements, patents, and copyrights); (4) what the terms of engagement (e.g., embargo of data, return of individual participant results) between *All of Us* and the collaborators are; and (5) how the relationship will be governed and how potential conflicts will be resolved. These disclosures will likely require that *All of Us* explain (particularly to skeptics) the importance and value of these collaborations and the ways that they extend and add to the resource and to biomedical discovery and translation. The working group notes a special opportunity within NIH for *All of Us* to provide cohorts for nascent studies that currently require an expensive process of patient identification and recruitment. In a cohort whose size is effectively one million participants, all common diseases as well as many rare diseases may be present in numbers that allow for novel research projects at a relatively low cost compared with the creation of novel one-off cohorts.

Third, *All of Us* can maintain the trust of participants and researchers, monitor the satisfaction of individual participants and their communities, and understand the factors associated with participant loss so that the program can continue to develop efforts that retain participants at a central level and locally, where feasible. This is particularly critical for vulnerable populations and those that are underrepresented in biomedical research. The working group recommends the continued development of a trust model, focusing on engaging participant partners and ambassadors to share their diverse lived experiences that offer valuable insights and help to intimately inform and influence the program.

This is consistent with the "participant-first" approach that *All of U*s has taken, where participants receive communications before others, and the priority is understanding and integrating the participant voice, translating participant needs and wants, amplifying partnership connections, and recognizing the emotional value of participation. *All of Us* can use its communication channels to assess shifts in participant perceptions or priorities and then respond.

With respect to the research community, *All of Us* can continue to monitor the publications that emanate from the program, as well as incoming grant proposals and presentations at scientific conferences, to understand both the current and emerging uses of the resource. As for participants, regular opportunities to elicit research views of the data quality, abundance, and analytic capabilities would ensure enduring relevance of the resource. For example, the Research Highlights feature showcases recent research powered by *All of Us* data and tools by selecting publications and summarizing them for participants in English and in Spanish with an infographic. These highlights are then shared with participants in their digital portal, as well as in monthly participant newsletters and online research highlights. The transparency of the Researcher Workbench, with public workspace descriptions and flaggable options as well as ongoing audits, offers additional examples of similar positive efforts by the program.

There is an opportunity to integrate siloed community groups with internal participant ambassadors through regular meetings and workshops, fostering a more cohesive and collaborative environment. Additionally, transparent partnerships should be amplified by promoting affiliations between health care provider organizations, community groups, and participant advisory groups. This approach aims to enhance trust and transparency through clear lines of communication and collaboration.

A look ahead reveals several opportunities for future participant engagement. The comprehensive and user-friendly *All of Us* Participant Portal consolidates news, events, research highlights, and educational content, enhancing engagement and retention while simplifying communication and fostering a sense of community and excitement. Virtual symposia involving Participant Ambassadors and community partners (currently in the pilot phase for regional convenings) can be organized to facilitate transparent engagement and problem-solving. Establishing a diverse Participant Researcher Advisory Board, composed of trained Participant Partners with a research orientation, could foster culturally relevant research questions and enhance the program's inclusivity. Lastly, revisiting and updating the Ethical, Legal, and Social Implications (ELSI) white paper can revitalize participant integration through progress mapping, ensuring that ethical considerations remain central to the program's development.

Ensuring Health Equity: The Current and Future Impact of *All of Us* as a National Resource

5. What is the impact of the *All of Us* Research Program, and what are the best opportunities for it to change the equitable practice of medicine now and in the future?

The core discoveries catalyzed by *All of Us* have already had a substantial impact in biomedical research according to an *All of Us* year in review published in the *American Journal of Human Genetics*. A key opportunity is for *All of Us* to help foster equitable research practices that advance precision medicine for all—research that informs the design of clinical systems to ensure that all patients receive high-quality health care, regardless of personal characteristics (e.g., gender, ethnicity, geographical location, socioeconomics, cultural lifestyle). Research enabled by the *All of Us* data and platform will generate knowledge that drives the practice of medical decisions that the public can trust.

Capabilities Needed to Support Research on the Equitable Practice of Medicine

Equitable clinical practice requires robust research capabilities for collecting and disseminating data on diverse populations—particularly UBR populations. Research should identify mechanisms to expand and translate data into relevant, usable information, facilitated by trusted researchers using methodologies that yield reliable and accurate scientific and practical knowledge. To maintain continuous relevance, the system should enable novel future research, focusing on identifying and addressing health disparity knowledge gaps and creating a virtuous cycle of discovery, implementation, and refinement. An equitable research system must also investigate the best methods to facilitate the efficient translation of knowledge into action—particularly for frontline clinicians in health care systems. Part of promoting data equity is providing feedback and information to public participants and communities to foster continuous trust and engagement. Finally, researchers must assess the appropriate roles of partners, such as industry investors, policy makers, philanthropic funders, and the public, who are crucial to the system's sustainability.

All of Us uniquely fills key roles and functions within this described system by supporting research relevant to the equitable practice of medicine. *All of Us* currently has operationalized crowdsourcing and a system for enrolling diverse populations. It has also created intake platforms for collecting relevant basic and clinical data and an output platform that makes data easily and securely accessible to researchers. As part of NIH, *All of Us* has built partnerships with diverse communities, making it a credible and scalable program to ensure the ongoing production of knowledge backed by scientific rigor.

The program can continue to support research relevant to the equitable practice of medicine by building a culturally aligned, systematic process that broadens the recruitment of both public and organizational-based participant populations—thus providing the core data required for research. It can also increase the numbers of active short- and long-term community participants while prioritizing the creation of pathways that allow a broader range of constituents in the clinical care delivery system to participate. Additionally, expanding the scope and connections with entities (e.g., health insurance companies, large employers, and disease and population advocacy groups) that can actively contribute and effectively utilize data will produce meaningful and credible information, particularly relevant to underrepresented populations. Finally, the program should seek to prioritize partnerships with entities within the health care system that have the necessary capabilities to advance, disseminate, and sustain the *All of Us* Research Program's objectives and processes in a timely, impactful manner.

With its strong foundation, *All of Us* has opportunities to build sustainability by focusing on its core capabilities and/or building affiliated capabilities to drive growth; strategically developing or strengthening focused alignments internally within NIH and with related governmental entities; and, most importantly, exploring new external partnerships.

NIH Internal Opportunities

Internally, *All of Us* leaders may explore partnerships and strategies to consolidate and reduce duplicative budget allocations across Institutes and Centers designated for activities that are similar or aligned. Through strategic partnerships and cross-division agreements, *All of Us* can proactively encourage the use of its data collection processes to support the creation of new cohorts and registries. *All of Us* could allow (and incentivize, where possible) external grantees to use the *All of Us* cohort to more efficiently and inexpensively create their research cohorts. Additionally, *All of Us* data parameters could be integrated into study datasets, allowing transparent analysis and reporting of cohort diversity.

External Opportunities

Nontraditional Researcher Opportunities

Diverse populations are important not only to academic researchers who seek to generate new knowledge, but also for population-based provider researchers within health systems, directly accountable for delivering equitable, efficient, and effective care to their designated populations. These direct frontline clinical providers are making care decisions based on observational data that may come from a relatively closed circle of patients. Expanding exposure and access to the diverse real-world data from *All of Us* can deepen these researchers' understanding of the broad population's needs, risk factors, and differential responses to interventions. It also provides an opportunity to gain knowledge from comparisons to larger but similar populations. To catalyze research that will impact frontline clinical care, *All of Us* can strategically increase the number of health system researchers with access to the data platform. With a joint exchange of data and research opportunities between *All of Us* and these health-system researchers, there is an opportunity to increase and accelerate the impact of the resource.

There are also an increasing number of nontraditional researchers who are community scientists and citizen scientists, and these groups offer important opportunities to expand the scope and impact of *All of Us*. In other disciplines, organized nonprofessional scientists have made remarkable contributions.

Opportunities With the Private Sector

The *All of Us* dataset can be valuable for commercial companies engaged in biomedical research to enable new breakthroughs. *All of Us* should continue creating a sustainable funding model to enable access to the data while respecting participant consent and ensuring that the privacy of the participants is protected. This may involve a fee-based approach to defray costs.

Highly valued *All of Us* capabilities to academic researchers are also highly valued capabilities to other groups in this space, including private-sector pharmaceutical companies, health systems providers, device manufacturers, and product investors. *All of Us* is a circular value system: The more participants are engaged, the more data are collected, and the more valuable the data and the information generated become. *All of Us* can consider placing a value on data access; it can discount access costs to those who successfully promote participant engagement. Some researchers may bring different values such as access to new datasets that accelerate analyses, contribute expertise and infrastructure for efficient IT modifications, or have robust communication capabilities. *All of Us* has been created to catalyze discoveries and analyses by diverse researchers (as measured by publications and patents) who advance the health of diverse communities. An additional value proposition can be the inclusion of those who contribute to the effort by paying for access and computation while also bringing additional value through their complementary resources and capabilities. *All of Us* can engage in and evaluate these promising partnerships.

Opportunities With External Organizations With Population Registries

Population-based health systems and provider specialty organizations with patient registries can be valuable partners that support *All of Us* datasets by enhancing population recruitment and exploring potential bidirectional data exchanges with these external data registries and other operational data exchanges.

Opportunities With Funders, Foundations, and Philanthropic Organizations

All of Us has the opportunity to be an important influencer and partner to other governmental funders through an expansion of its work with the Foundation for the NIH and, most importantly, to external foundations and partners. These major funders could fuel and facilitate health system research and care delivery demonstration pilots. As a partner, *All of Us* can help provide strong and credible data on diverse population characteristics and allow more strategic and specific funding objectives. Funding resources that partner with *All of Us* can encourage grantees to work with their populations to register with the program, build volume, and increase diversity. There may be additional advantages in routinely annotating the use of *All of Us* in reporting external grant outcomes.

Participant Opportunities

The future of the *All of Us* Research Program is actually in the hands of the participants. The program can empower them to be change agents and positive disruptors to the traditional research process. This could ensure that the program does not become one of many programs vying for the same audience. The foundation is built, the potential is there, and it is time for big steps to grow into the future.

Participants are valued not only as data contributors but also as active members of the research team, thus eliminating the silos that isolate researchers from participants. Participants can be publicly visible. Positive and sincere participant testimonies give credibility and trustworthiness to the program and help build the trust of volunteers—particularly those from underrepresented populations. Similarly, this has also been key to the success of Patient-Centered Outcomes Research Institute.

6. As the program exceeds 800,000 participants and 10,000 researchers, how should it focus future investments to sustain and grow as a vital national research resource?

Since 2016, *All of Us* as an NIH program has been funded through congressional appropriations, including the NIH Office of the Director base appropriation and 21st Century Cures Act authorized appropriation. With its initial performance goals achieved and the establishment of clear functionality driven by its operational capabilities, *All of Us* is at a pivotal juncture: positioned as a vital national research resource uniquely focused on public engagement, data collection, and the facilitation of research on diverse populations.

Given that *All of Us* is continually enrolling participants, it is too soon to see the full impact of investment on scientific progress and downstream changes for many diseases and conditions (both treatment and/or prevention). The program is on an upward trajectory of scientific discovery, as evidenced by all figures in this report. Already, it is possible to undertake research into the genetic determinants of disease; these research opportunities will become even more robust as more participants have their genome sequenced. In particular, the availability of whole genome sequencing data on diverse participants makes this dataset a globally unique resource to better understand how genetic factors influence disease across different populations. This resource will ultimately lead to more precise risk prediction—which could, for example, inform targeted screening—and treatment strategies that are more equitable across different ethnic and racial population subgroups.

To maximize the utility of this resource for the research community, two key areas require focus. First, continuing to expand the amount of data collected on all participants (and evaluating its utility) is crucial. This includes prioritizing the linkage to EHRs for all participants, as the data utility is significantly enhanced with this connection. Additionally, linking environmental data and incorporating biomarkers—particularly proteomics, but also metabolomics and RNA sequencing, if feasible—will substantially enhance the resource's overall value. The program should examine methods to facilitate combining its data with other national and international biobank data. *All of Us* should establish a robust pediatric cohort and establish familial linkages across participants.

Second, continuing to increase the usage of the resource is vital for accelerating the number and rate of scientific discoveries. This can be achieved by raising awareness of access for international researchers, as well as enabling researchers from commercial companies to utilize the resource. By addressing these areas, *All of Us* will become more valuable and impactful for the broader scientific community. The power of the program's data lies in the diversity of data and populations, and the intersections between them. *All of Us* can help drive precision medicine by fostering interdisciplinary research that capitalizes on these intersections. Generative AI should be responsibly integrated into the Researcher Workbench, because it can help integrate and synthesize data from diverse sources that bridge gaps between disciplines, generate novel hypotheses, integrate vast amounts of literature, and help researchers with limited programming expertise write code to conduct their analysis. This may help lower barriers to entry and attract more researchers.

The 34% budget cut from fiscal year (FY) 2023 to FY 2024 and the potential for an additional 56% cut from FY 2024 to FY 2025 is a significant concern for the long-term future of the program. Predictable and stable funding is necessary for *All of Us* to achieve its long-term intended outcomes. The program should consider additional changes to its model for sustaining and/or expanding the roles it plays in the practice system. This includes rethinking its strategy and methodologies, and diversifying its revenue stream. Significantly reduced budgets call for a razor-sharp focus on strategy, consolidation of like functions and resources, and recruitment of different partnerships with aligned goals and value propositions or—in the most stringent financial scenarios—a pause on new enrollment to focus only on existing participants and researcher access without additional funding sources.

To move forward with its goal, *All of Us* will need to rethink its strategy and operational model to manage sustainability, be less dependent on shifting budgetary allocations, and grow with the demands of the market. Following principles for business sustainability, *All of Us* will need to reevaluate its internal budgetary allocations while considering a strategy to sustain and facilitate growth. This involves rethinking operations and associated costs and migrating to a different revenue strategy to effectively replace losses in budgetary allocations.

Enrolling volunteer participants is a key and powerful capability of *All of Us*. It is a major component of the plan for sustainability and provides a unique position in contributing to an equitable system of clinical practice. The volume, diversity, and engendered trust of volunteer participation is the key asset of *All of Us*, driving the research and demonstrating the effectiveness of its processes and trusted positioning. Therefore, increasing the number of participants needs to be a major priority going forward in near- and long-term planning. Internal processes to enhance the process need to be targeted for the investment of time, resources, and focus. To ensure that the cohort fully reflects the diversity of the country, including multiple UBR groups, the program should continue to consider the rotation of major enrollment centers across the country. Care should be taken that this is not an abandoning of certain areas or populations, but rather a necessary rotation to ensure inclusion.

All of Us will need to innovate ways to financially support the value delivered to participants (e.g., building crowdsourcing support systems and personal data driven profiles). Engagement of participant ambassadors in the *All of Us* governance structure drives and ensures commitment to the importance of their participation and engagement to the growth and strength of the program. *All of Us* will need to continue to actively engage Participants Ambassadors and use their feedback to protect the trust that initially fueled participant engagement. *All of Us* should continue to prioritize enrolling underrepresented participants and maintain global leadership by having one of the largest and most diverse biomedical datasets in the world.

All of Us should continue expanding to currently underrepresented populations—especially pediatrics—and consider remodeling its approach to data collection and maintenance to support long-lifespan data. The program will have to structure pediatric-related datasets within the data platform to support research on developmental changes over time. Over the years, *All of Us* and NIH have made progress in cultivating relationships with participants and community organizations. As part of its growth strategy, *All of Us* will need to further strengthen these relationships by building on the pillars that support them. This includes developing and operationalizing a participant trust model, incorporating siloed community groups, amplifying transparent partnerships, and providing participant education, as suggested by Participant Partners.

A Special Opportunity for Pediatric Health

As *All of Us* matures, the program needs to truly reflect the U.S. population, including the youngest members of society: infants, children, and adolescents. There are many conditions that affect only children (e.g., neonatal jaundice, bronchiolitis, congenital hearing loss). Other conditions (e.g., severe congenital anomalies and many rare, lethal genetic conditions) affect mostly children because the manifestations are so severe that individuals may not live long enough to become adults. Some lifelong conditions (e.g., asthma, type 1 diabetes, autism) are diagnosed in childhood. Studying individuals before and around the time of symptom onset therefore necessitates studying children.

Additionally, for most common conditions across the lifespan—such as obesity, hypertension, and atherosclerosis—physiology, pathogenesis, and modification of risk via health behaviors begin during childhood and adolescence. Studying factors starting with prenatal and perinatal exposures and continuing over the life course will inform when and how we can improve the health of future generations with early intervention. Including children and strategies to engage and retain them once they become adults will provide a critical investment to understand health over the life course in a way that other cohorts have not.

Children, while comprising about 25% of the U.S. population, represent 100% of our future. Including participants from birth to 18 years old will make *All of Us* the first and largest longitudinal cohort to encompass participants across the entire lifespan. By building a robust pediatric cohort and dataset, the *All of Us* Research Program would set itself apart from previous major adult and pediatric cohorts in several key ways.

First, it will create the world's largest longitudinal cohort across the entire life course. This expansion will significantly advance early prediction, primordial prevention, and primary prevention of chronic diseases, many of which—such as cardiovascular diseases, asthma, diabetes, cancer, and Alzheimer's disease—may have origins in early life, even during the fetal period. These diseases are potentially preventable or reversible if detected and addressed at an early stage.

Second, this addition will attract and expand the pool of researchers interested in promoting health across the life course, including pediatricians, obstetricians, gynecologists, family medicine practitioners, and specialists in maternal and child health and women's health, which are national priority areas.

Third, enrolling children of current adult participants in *All of Us* will enable linkages across families and create a two-generation cohort, laying the groundwork for a sustainable, multigenerational study that could eventually encompass three or even four generations within one to two decades. This approach provides a unique opportunity to understand why major health disparities persist in the United States. Given its significant proportion of marginalized populations, *All of Us* has the opportunity to offer unparalleled opportunities to investigate the origins and persistence of these disparities across generations, potentially elucidating epigenetic transmission from one generation to the next.

Envisioning the Future: The Sustainability and Longitudinality of *All of Us*

7. What prospective actions can *All of Us* take to minimize long-term infrastructure costs to taxpayers while improving overall efficiency and effectiveness?

The *All of Us* Research Program represents a significant investment, and it is imperative that it produce scientific knowledge and initial demonstrations of effectiveness in order to deliver value to taxpayers. The forms that this value can take are myriad. *All of Us* should provide the data and research infrastructure to accelerate basic biomedical discovery that yields insight into the mechanisms of health and disease. These insights will be informed by a diverse cohort of participants reflective of the diversity of individuals living in the United States. It should provide a cohort of patients upon which many clinical trials—funded by NIH and by private companies or consortia—can be run. It can provide these cohorts in a manner that is significantly less expensive than current clinical trial approaches. It should serve as a proving ground for clinical translation of basic discoveries into clinical practice across diverse settings (e.g., community settings). The infrastructure for recruitment and retention, as well as the research infrastructure for data storage and analysis, should provide a model that can be inexpensively adopted by other research programs, thus ensuring the compatibility of *All of Us* with these programs. *All of Us* has already made investments in establishing the cohort, the communication programs, the advisory structures, the data formats, and the analytic platforms to establish a world-class research infrastructure. Going forward, the program will have to assess its strengths and weaknesses and make choices about how to allocate resources to do what it can uniquely do best. The working group identified several areas that require focus.

First, *All of Us* can develop approaches for seeking new sources of funding. The 21st Century Cures Act encourages *All of U*s to develop and utilize public-private partnerships. There is no doubt that basic "core" funding must be provided from the federal budget, so *All of Us* should also develop approaches for working with Congress and the Administration. Core support funds are critical for "keeping the lights on"—participant communication and retention, researcher support, physical facilities for storing biospecimens, and computational facilities for storing the data securely and supporting data analysis. However, these core capabilities can be supplemented with additional endeavors to fully benefit from the value contained in the participant data and biospecimens.

The most obvious source would be ancillary and collaborative studies where the primary funding comes from outside sources: private companies, foundations (including disease-specific ones), other government agencies, and nongovernmental organizations. This would add to the rich data in *All of Us* and might be essentially free. UK Biobank provides an outstanding example of managing the added value of ancillary/collaborative studies to enrich the resource at relatively low cost.

In some cases, *All of Us* might want to seek co-investment and create attractive cost-matching programs to acquire mission-critical data at a discount. These arrangements may require the data to be embargoed for a short time (e.g., six months) in order to allow the collaborators to analyze it before releasing it for public analysis.

Another source of income could be from researchers via charges for using the data via its compute infrastructure. The working group recognizes this as a potential funding stream, but worries that it would reduce enthusiasm for and use of the *All of Us* platform by researchers—particularly those with limited resources. Therefore, a careful payment scale consistent with current funding models of science, and likely including "scholarships" for low-resource institutions, would be required. Consultation with academic and industrial user groups about their willingness and ability to pay would be a critical step.

Second, in order for *All of Us* to become a longitudinal resource over the next several decades, it should consider management efficiencies to reduce core costs. *All of Us* can evaluate the complexity and time required to enter new partnerships, collaborations, and consortia. It can continue to identify services that are critical for it to build and maintain in house, versus those that can be purchased and recompeted regularly to keep costs down. *All of Us* can investigate ways to streamline the processes of enrollment and retention, including the methods by which participants are informed of and recruited to ancillary/collaborative studies. The participants want to participate, so engagement should be easy for them. The speed and effectiveness in initiating and executing exciting scientific research projects can be tracked in order to minimize lost opportunities. *All of Us* can also create an academic ecosystem of studies of its cost-effectiveness, impact, and outcomes. Health economists, outcomes researchers, and others have the skills to evaluate the impact of the program and stimulate these evaluations. In addition, consultation with health care organizations, health insurance companies, and others in the business of health care will also provide insight into the opportunities and impacts of *All of Us*.

Third, *All of Us* can be bold in leading new consortia to attack the most pressing health problems facing society today. The COVID-19 pandemic struck early in the life of *All of Us*; nonetheless, the program contributed to the literature on the pandemic by mobilizing researchers and participants to ask and answer key questions. The working group expects that a more mature *All of Us* will not just participate but also lead research initiatives in areas of critical need. This will involve creating coalitions of researchers, clinicians, health care systems, and payers to form multidisciplinary teams and formulate rapid and effective research programs to understand and intervene in disease. These coalitions can be created immediately—not while waiting for the next emergency. Datasets can be examined and harmonized in advance (at a minimum, a plan for harmonization should be created). Governance rules can be drafted for such research collaborations so that they can be stood up quickly. Initial coalitions should be created soon to test these ideas, identify the challenges, and make research progress. When the next pandemic or other large-scale health emergency happens, *All of Us* should be a central player in organizing the research response to characterizing and intervening. And it should do so using tested methods that have been prototyped and implemented in advance. Even a single effective discovery and intervention program in the setting of a major health threat could essentially pay for the entire *All of Us* investment.

8. How can *All of Us* strategically contribute to future cost savings for researchers within the scientific ecosystem, leveraging initiatives like the passport and cloud model?

The creation of a cohort of one million or more Americans with data from surveys, EHRs, biosamples, physical measurements, digital health technologies, whole genome sequences, and genotype arrays is the goal of the *All of Us* Research Program. Part of the motivation for creating such a cohort is to be a disruptor and enabler for precision medicine research across the United States, and to serve as a living laboratory for researchers to ask important questions regarding health and disease to improve the health of all Americans. Now that *All of Us* is well underway and achieving its goals of recruiting a diverse group of people across the country with rich data, an important consideration is how the program can strategically contribute to future cost savings for researchers within the scientific ecosystem. Precision medicine research using large-scale, multimodal data from human populations can be costly and time-intensive. The *All of Us* Research Program has an opportunity to reduce both factors for researchers across the spectrum.

First, *All of Us* data can transform biomedical and population health research and health care discovery. Having a readily accessible cohort of one million diverse people nationwide with a spectrum of health and disease states allows for countless research studies without the need for new recruitment of research participants. The recruitment cost savings that researchers gain by removing this need is huge. Additionally, for junior researchers, including trainees (students, fellows) and junior faculty, having access to the data assets and computational infrastructure provided by *All of Us* is unprecedented. In many cases, there is no longer a need for researchers to develop a plan to recruit participants, develop a consent protocol, and spend the time and resources building a cohort. Through the Researcher Workbench, researchers can create their study cohorts seamlessly and for very low cost. The investment in developing the *All of Us* cohort, data, and computing platform is a strategic contribution to future cost savings for research in the scientific ecosystem.

Second, the Researcher Workbench levels the playing field in terms of access to data and computation. It also enables a diverse research community to participate in state-of-the-art research in precision medicine. The *All of Us* team has been working diligently to create the Researcher Workbench and make it accessible to the national research community. It uses a "data passport" model to give registered researchers broad access rather than granting data access on a project-by-project or question-by-question basis. Once they complete the requisite steps, researchers can create workspaces in the tier or tiers to which they have access. When they set up their workspaces, they are required to provide public-facing project descriptions to support transparency. Researchers must ensure that their research complies with the program's data use policies, and regular monitoring and auditing occurs by the *All of Us* Resource Access Board.

The team has made great progress in the past few years and should continue to push toward making the Researcher Workbench both easier to use and more flexible for experienced bioinformatics researchers. In addition, developing a strategy for users to easily upload and share other data across projects in the Researcher Workbench will be important for future research programs. Perhaps even incentivizing data sharing—S-index to calculate how many times data has been shared—would be beneficial. Additionally, developing a mechanism to connect researchers who are doing similar work on the Researcher Workbench would accelerate research outcomes. Perhaps developing incentives for these collaborations will increase productivity and decrease overlap on projects on the Researcher Workbench.

Third, developing a plan for making the *All of Us* data more easily interoperable with other national and local datasets would be paradigm changing and significantly reduce costs of research. For example, the model of making *All of Us* data sharable and easily accessible for other large datasets in other NIH consortia would be very beneficial. Many large datasets exist, but integrating these datasets is challenging. The *All of Us* team may be able to make this process simpler for researchers. As more researchers start to use *All of Us* and other large datasets, such as UK Biobank, for joint analyses, it may be useful to invest effort into harmonizing the data types and formats (e.g., the Observational Medical Outcomes Partnership [OMOP], Phecodes). If this is done centrally and then shared with researchers who have access to both datasets, it will reduce researcher costs for research and expedite the research as well.

Finally, the program should consider an approach for leveraging federated data analysis frameworks to accommodate using *All of Us* with other large datasets that live exclusively in other cloud platforms. These include UK Biobank (in DNAnexus), TOPMed (in BioData Catalyst), and National Human Genome Research Institute datasets (in Analysis, Visual, and Informatics Lab-space). Perhaps through something like a cloud passport model, researchers can go through the data access approval process for all requested datasets, then be given access to a portal that provides federated access to all datasets. This would completely change the way that these large datasets are accessed and result in huge cost savings.

There is a rich emerging ecosystem of biobanks, and *All of Us* should participate in and lead collaborative efforts across it. By aligning data standards, such as OMOP or Phecodes, researchers can more effectively conduct joint analyses, ensuring that findings are consistent and comparable across cohorts. This harmonization will not only streamline research processes but also maximize the potential for meaningful discoveries that can affect global health. Additionally, fostering awareness and collaboration on new data collections—such as web-based surveys, wearables, and environmental modeling—will further strengthen the robustness of research findings. By ensuring that these new data sources are as comparable as possible across different cohorts, researchers can better replicate studies and validate results, ultimately leading to more reliable and impactful outcomes in biomedical research.

The *All of Us* Research Program and its leadership team have done a tremendous amount already to reduce the cost of research for the broader research community. Though there is still more that can be done, they should be commended for all they have done so far. This program and data asset has enabled a significant increase in the research community conducting precision medicine research and will continue to do so for years to come.