Title

The All of Us Research Program is an opportunity to enhance the diversity of US biomedical research.

Journal

Nature Medicine, 30(2)

Authors

Bianchi, Diana
Brennan, Patricia
Chiang, Michael
et al.

Publication Date

2024-02-01

DOI

10.1038/s41591-023-02744-3

Peer reviewed

HHS Public Access

Author manuscript

Nat Med. Author manuscript; available in PMC 2025 February 18.

Published in final edited form as:

Nat Med. 2024 February ; 30(2): 330–333. doi:10.1038/s41591-023-02744-3.

The All of Us research program is an opportunity to enhance the diversity of US biomedical research

Diana W. Bianchi[1], **Patricia Flatley Brennan**[2], **Michael F. Chiang**[3], **Lindsey A. Criswell**[4], **Rena N. D'Souza**[5], **Gary H. Gibbons**[6], **James K. Gilman**[7], **Joshua A. Gordon**[8], **Eric D. Green**[9], **Susan Gregurick**[10], **Richard J. Hodes**[11], **Peter H. Kilmarx**[12], **George F. Koob**[13], **Walter J. Koroshetz**[14], **Helene M. Langevin**[15], **Jon R. Lorsch**[16], **Eliseo J. Pérez-Stable**[17], **Joni L. Rutter**[18], **Jane M. Simoni**[19], **Bruce J. Tromberg**[20], **Debara L. Tucci**[21], **Nora D. Volkow**[22], **Rick Woychik**[23], **Shannon N. Zenk**[24], **Elyse Kozlowski**[25], **Rachele S. Peterson**[25], **Geoffrey S. Ginsburg**[25], **Joshua C. Denny**[25,*]

[1]Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD

[2]National Library of Medicine, National Institutes of Health, Bethesda, MD, 2016-2023

[3]National Eye Institute, National Institutes of Health, Bethesda, MD

[4]National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, MD

[5]National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD

[6]National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD

[7]Clinical Center, National Institutes of Health, Bethesda, MD

[8]National Institute of Mental Health, National Institutes of Health, Bethesda, MD

[9]National Human Genome Research Institute, National Institutes of Health, Bethesda, MD

[10]Office of Data Science Strategy, National Institutes of Health, Bethesda, MD

[11]National Institute on Aging, National Institutes of Health, Bethesda, MD

[12]Fogarty International Center, National Institutes of Health, Bethesda, MD

[13]National Institute on Alcohol Abuse and Alcoholism, National Institutes of Health, Bethesda, MD

[14]National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD

[15]National Center for Complementary and Integrative Health, National Institutes of Health, Bethesda, MD

[16]National Institute of General Medical Sciences, National Institutes of Health, Bethesda, MD

[*]Corresponding author: Joshua C. Denny, joshua.denny@nih.gov.

Competing interests

The authors declare no conflicts of interest.

[17]National Institute on Minority Health and Health Disparities, National Institutes of Health, Bethesda, MD

[18]National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, MD

[19]Offices of the Director and Behavioral and Social Sciences Research, National Institutes of Health, Bethesda, MD

[20]National Institute of Biomedical Imaging and Bioengineering, National Institutes of Health, Bethesda, MD

[21]National Institute on Deafness and Other Communication Disorders, National Institutes of Health, Bethesda, MD

[22]National Institute of Drug Abuse, National Institutes of Health, Bethesda, MD

[23]National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, NC

[24]National Institute of Nursing Research, National Institutes of Health, Bethesda, MD

[25]All of Us Research Program, National Institutes of Health, Bethesda, MD

Abstract

The All of Us Research Program has prioritized the enrollment of people from backgrounds historically underrepresented in medical research to bring precision medicine to the full diversity of the US population and to improve health outcomes for all.

Several countries with national health systems have established large, longitudinal cohorts to advance population and precision health, each with unique features and populations, including the UK Biobank([1]) and several more. The U.S. is different from these countries: not only does it lack a national health system for its citizens, but it ranks 43rd in the world for life expectancy([2]). Important longitudinal population-specific studies of health in the United States are ongoing, such as the Million Veteran Program, but most of these studies have a dramatic underrepresentation of diverse populations([3]).

The lack of diversity in genomics research, with more than 90% of studies from populations of European ancestry([4]), has led to many challenges in equity, including non-transportability of polygenic risk scores to different populations([5]) and incorrect assignment of genomic variant pathogenicity([6]).

A platform for research

Launched in 2018, the National Institutes of Health's (NIH) All of Us Research Program has enrolled over 700,000 people and is deliberately focused on advancing health equity by inclusively engaging and enrolling a diverse population([7]). All of Us aims to include populations historically underrepresented in biomedical research. Currently, 80% of All of Us participants are underrepresented in biomedical research by race, ethnicity, age,

geography, sexual and gender identity, income, education, access to healthcare, and/or disability.

Nearly 7,000 researchers across a variety of training levels and representing more than 530 institutions (including more than 85 Historically Black Colleges and Universities and Hispanic-serving institutions) have signed up to analyze the available data repository on more than 400,000 participants. Data includes survey responses, electronic health records (EHRs), genome sequences, and FitBit records, accessed via the cloud-based Researcher Workbench. The Workbench contains three tiers of data access (Public, Registered and Controlled), with genomic data available in the Controlled Tier to registered researchers who have signed a data use agreement. More than 245,000 whole-genome sequences and data from more than 312,000 genomic variant arrays are available for analyses via the Controlled Tier. In the 245,000 genome sequences available thus far, there is genetic variation at more than a billion loci. More than 200 articles using this data have been published in peer-reviewed journals since the first data release in 2020.

More data from diverse U.S.-based populations, such as that from environmental exposures and geospatial data, are needed to drive identification of meaningful and valid risk factors, early detection tools and prevention strategies, and precision medicine applications that reach everyone. Lack of inclusion of diverse populations in research can have profound negative consequences to human health, leading to a narrow understanding of the underlying biology of diseases, impeding the development of new treatments, and limiting the effectiveness and applicability of treatment and prevention strategies for all populations. As directors of NIH institutes, centers, and offices, we encourage researchers to contribute to and leverage the All of Us Research Program as a key resource to aid in this endeavor.

Unifying the NIH

NIH has long supported cohort science beginning with the Framingham Heart Study in 1948. Many of these studies have been specific to individual disease domains or conditions, such as cardiovascular disease or aging, or tailored to certain populations. The discovery of intervenable risk factors for cardiovascular disease from the Framingham Heart Study has contributed to a >60% reduction in cardiovascular disease since the 1950s([8]).

Previous joint research initiatives across the NIH have resulted in many contributions to multi-disciplinary science, including the Brain Research through Advancing Innovative Neurotechnologies (BRAIN) Initiative([9]). There is still a critical need to ensure that research resources are applicable and valid across the entire population. All of Us aims to unite the entirety of the NIH towards creating a multimodal, comprehensive assessment of human health that is reflective of the United States.

All of Us is disease-agnostic and so is applicable to the many research interests of NIH. To better understand the landscape of health conditions reflected in data from the program, we compared top health condition prevalence in the U.S. population, as estimated in the Global Burden of Disease Study 2019([10]), to current All of Us participants who have a matching diagnosis in EHR data, and a subset of those who identified as belonging to racial or ethnic

minorities underrepresented in biomedical research. Fig 1 shows the relative prevalence of each health condition in these three populations, projected to values in 1 million participants for ease of comparison including the prevalence in underrepresented populations by race and ethnicity. All of Us generally mimics the prevalence of common diseases in the general population. There is higher relative prevalence of Substance Use Disorders, Diabetes, and Maternal and Neonatal disorders amongst underrepresented racial and ethnic groups. This provides evidence of potential health disparities for some diseases represented in All of Us.

There are several key strategic areas in which NIH institutes, centers, and offices can partner with and augment All of Us to address barriers to research for which no other adequate resources exist (Table 1).

A milestone in the symbiotic relationship between All of Us and the broader set of NIH institutes, centers, and offices is the launch of ancillary studies, which add new data and technologies in alignment with the scientific strategy of the program. This framework leverages existing relationships with NIH institutes, centers, and offices to build proactive collaborations that will inform strategies for expansion to other partners, including industry and international collaborators. Any data generated via these studies will be returned for broad researcher use, creating a robust feedback loop to enhance baseline data collection.

The largest ancillary study to date, Nutrition for Precision Health, aims to understand individual responses to different diets and develop models to provide insight into personalized nutrition. Supported by the NIH Common Fund and managed by a trans-NIH collaborative working group involving 18 institutes and centers, the study is also a key component of the NIH Nutrition Research strategic plan. Future ancillary studies will allow for the addition of new data types to expand the depth of data available to researchers, including environmental, geospatial, exposomics, imaging, and others.

Synergies to address deficiencies

We have identified nine key areas of synergy between NIH institutes, centers, and offices and All of Us that address common and unique deficiencies in the current clinical, epidemiological, and genomics research landscape (Box 1).

The nine themes in Box 1 are already present among the more than 7,700 studies ongoing in the All of Us Researcher Workbench. Several studies are investigating ways to optimize existing algorithms and develop new models to predict clinical outcomes by leveraging the population diversity represented in All of Us data. This includes a new model trained with multisite, multi-EHR All of Us data predicting the need for surgery among people with glaucoma, which outperformed a previous model from a less diverse, single site dataset([11]). This work highlights the need to include robust diversity in artificial intelligence and machine learning models for disease prediction to ensure equitable utility across populations.

Other studies have investigated disparities in underrepresented groups across All of Us data for cardiovascular disease([12]) and cancer([13]), among others. Researchers are beginning to use All of Us data to test validation of polygenic risk scores in diverse populations([14]) as

well as the pathogenicity of rare genomic variants([15]). All of Us will empower translation of research findings to clinical outcomes.

Data missingness

Achieving the promise and potential of All of Us is not without challenges. Data missingness, harmonizing EHRs across vendors and environments, and continuing to engage and enroll a diverse population are just some of the areas in which the program is actively seeking solutions. We are working to make the Researcher Workbench easier and more accessible to diverse researchers, including development of standardized phenotype definitions and introduction of new tools. More research is needed into understanding enrollment biases and evaluate the generalizability of results, including analytical approaches to address missing data from surveys and EHRs. The program also continuously monitors survey completion rates across populations and has introduced new methods to reduce gaps in completion, such as computer-assisted telephone interviews and other specialized interventions. The program promotes open science through sharable workspaces in the Researcher Workbench, and methods and tools created to overcome these limitations are freely available for use by other researchers.

Ensuring equitable participant access to All of Us is an ongoing process, and All of Us is continuing to identify and address operational constraints identified as barriers to enrollment. Inherent in its mission and core values is the ability to acquire informed consent from the full complement of participants. The consent process was designed to enable fully self-navigated consent by the maximal number of eligible participants and ensure an inclusive experience for all regardless of race, ethnicity, gender identity, sexual orientation, disability status, income, education, or access to care. However, many participants may still desire or require additional support to complete the consent process. Assisted and facilitated consent are presently permitted, and consent-by-proxy where an enrollment relies on a legally authorized representative is an initiative in planning stages in partnership with American Association on Health and Disability to provide input on the operationalization of participation by proxy, including subject matter expert consultation, a landscape analysis, and report anticipated in 2024.

All of Us was intentionally designed as a resource that addresses diversity and inclusion at all levels: diversity in participant demographics; diversity in data types; and diversity in the researchers. It aims to be a world-accessible scientific resource that provides the data, tools, and cloud-based analysis infrastructure needed to enable biomedical research to support the missions of all NIH institutes, centers, and offices, extending across the biomedical research community and enhancing the work of researchers at all career levels.

We represent the NIH institutes, centers, offices, and All of Us, and invite researchers to join us in expanding the All of Us platform to all disease domains, diverse populations, common and rare genomic variants, social and commercial determinants of health, and other modalities to ensure the advancement of precision health, medicine, and equity for everyone.

Acknowledgements

The authors thank the participants of the All of Us Research Program for their trust, time, and contributions.

References

1. Sudlow C. et al. PLoS Med 12, e1001779 (2015).
2. Foreman KJ et al. Lancet 392, 2052–2090 (2018). [PubMed: 30340847]
3. Popejoy AB, Fullerton SM. Nature 538, 161–164 (2016). [PubMed: 27734877]
4. Mills MC, Rahal C. Nat Genet 52, 242–243 (2020). [PubMed: 32139905]
5. Martin AR et al. Nat Genet 51, 584–591 (2019). [PubMed: 30926966]
6. Manrai AK et al. N Engl J Med 375, 655–665 (2016). [PubMed: 27532831]
7. Denny JC et al., N Engl J Med 381, 668–676 (2019). [PubMed: 31412182]
8. Mensah GA et al. Circ Res 120, 366–380 (2017). [PubMed: 28104770]
9. Bargmann CI, Newsome WT. JAMA Neurol 71, 675–676 (2014). [PubMed: 24711071]
10. GBD 2019 Diseases and Injuries Collaborators. Lancet 396, 1204–1222 (2020). [PubMed: 33069326]
11. Baxter SL et al. Am J Ophthalmol 227, 74–86 (2021). [PubMed: 33497675]
12. Acosta JN et al. J Am Heart Assoc 10, e021724 (2021).
13. Aschebrook-Kilfoy B. et al. PLOS ONE 17, e0272522 (2022).
14. Lennon NJ et al., medRxiv, (2023). doi: 10.1101/2023.05.25.23290535.
15. Geck RC, Powell NR, Dunham MJ, Am J Hum Genet 110, 228–239 (2023). [PubMed: 36681081]

Box 1.

Nine areas of synergy with All of Us.

1. Enhancing the diversity of genomic data across all conditions.
2. Building an open foundation for artificial intelligence and machine learning across human health and diverse populations, EHR vendors, and healthcare systems.
3. Establishing models and technologies for disease prediction, surveillance, screening approaches across the lifespan.
4. Leveraging linkage between biospecimens, EHR, digital health platforms, imaging, and other participant-provided data streams to prospectively identify disease biomarkers prior to diagnosis.
5. Linking domains of behavioral, emotional, and cognitive traits and their impact on various aspects of medical illness, treatment adherence, and cognitive changes, including the extension to children and adolescents.
6. Equitably defining risk factors (such as actionable genetic variants and genetic predisposition across race and ethnicity) to improve care guidelines and outcomes for the leading health conditions in the USA across diverse populations, including health disparities.
7. Deciphering complex traits associated with differential responses to the exposome.
8. Investigating diseases of childhood that have largely been underexplored in biomedical research, but often represent a foundation for future health.
9. Creating a longitudinal sampling of environmental, geospatial, social determinants, and other data to capture the effect of all experiences of an individual across the lifespan.

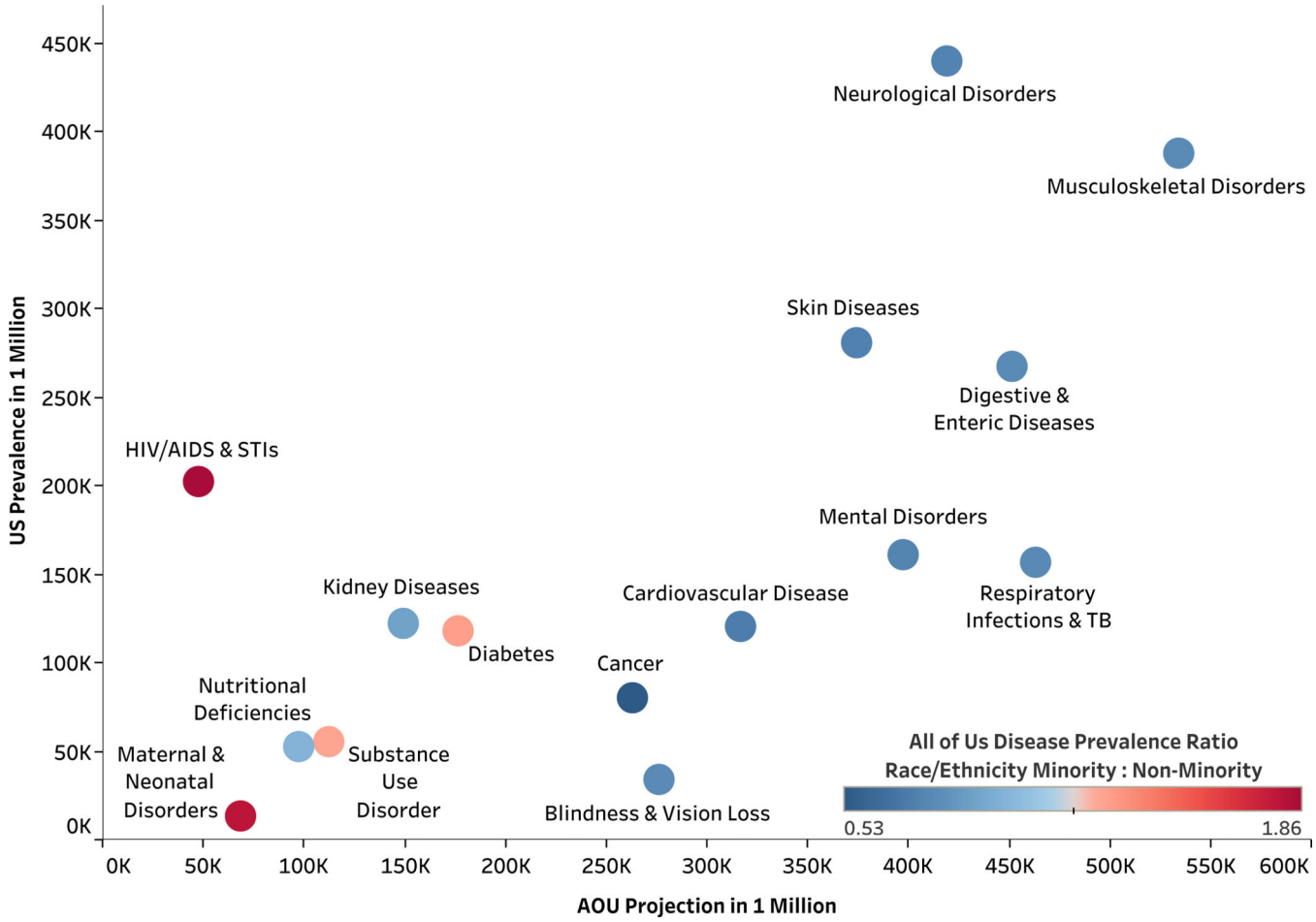

Figure 1. Prevalence of common health conditions in the USA and among All of Us Research Program Participants.

National prevalence of each disease is estimated by the Global Burden of Disease Study 2019. Utilizing the International Classification of Disease 9 (ICD-9) and the International Classification of Disease 10 (ICD-10) codes from the GBD2019 Study, mapped to OHDSI Common Data Model condition concepts, prevalence was estimated from participants who have provided access to their EHR records in All of Us, including the subset cohort identified as being Underrepresented Racial and Ethnic minorities in Biomedical Research. All estimates were adjusted to project predicted prevalence in 1 million people in the United States.

Table 1.

Strategic areas for partnership with All of Us.

Strategic area	Goal
Diversity, equity, and inclusion	Intentionally develop diversity and engagement strategies that address historical barriers to inclusion in research, to ensure that All of Us participants are reflective of the U.S. population and inclusive of populations under-represented in biomedical research.
Data linkage	Integrate and harmonize disparate sources of EHR data throughout healthcare systems, including federally qualified health centers, and link environmental exposures, healthcare claims and mortality data.
Data completeness	Improve data completeness across diverse populations by developing special considerations for survey completion (such as computer-assisted telephone interviews), with an intentional focus on providing participants with a variety of ways to contribute to the longitudinally of the program (such as through opportunities to complete new surveys and link other data sources such as wearables and imaging), and assessment of retention rates across populations.
Dynamic infrastructure	Create a dynamic infrastructure that allows for shifting priorities in response to changes in the national health landscape and accommodates a variety of health topics and approaches, such as the COVID-19 pandemic and emerging opportunities such as artificial intelligence.