The "All of Us" Research Program

The All of Us Research Program Investigators

The affiliations of the members of the writing committee are as follows: the Departments of Biomedical Informatics and Medicine, Vanderbilt University Medical Center, Nashville (J.C.D.); the National Center for Advancing Translational Sciences (J.L.R.) and the All of Us Research Program (G.J., E.D.), National Institutes of Health, Bethesda, MD; the Institute for Genomic Medicine and Department of Neurology, Columbia University Irving Medical Center, New York (D.B.G.); and the Broad Institute, Cambridge (A.P., J.W.S.), and the Center for Genomic Medicine and Department of Psychiatry, Massachusetts General Hospital, Boston (J.W.S.) — both in Massachusetts.

Summary

Knowledge gained from observational cohort studies has dramatically advanced the prevention and treatment of diseases. Many of these cohorts, however, are small, lack diversity, or do not provide comprehensive phenotype data. The All of Us Research Program plans to enroll a diverse group of at least 1 million persons in the United States in order to accelerate biomedical research and improve health. The program aims to make the research results accessible to participants, and it is developing new approaches to generate, access, and make data broadly available to approved researchers. All of Us opened for enrollment in May 2018 and currently enrolls participants 18 years of age or older from a network of more than 340 recruitment sites. Elements of the program protocol include health questionnaires, electronic health records (EHRs), physical measurements, the use of digital health technology, and the collection and analysis of biospecimens. As of July 2019, more than 175,000 participants had contributed biospecimens. More than 80% of these participants are from groups that have been historically underrepresented in biomedical research. EHR data on more than 112,000 participants from 34 sites have been collected. The All of Us data repository should permit researchers to take into account individual differences in lifestyle, socioeconomic factors, environment, and biologic characteristics in order to advance precision diagnosis, prevention, and treatment.

Longitudinal cohort studies have improved human health by characterizing natural histories of diseases, identifying their risk factors, and revealing new biomarkers. Advances in genomics and biosensors have set the stage for refined taxonomies of disease, which may

help to guide prognosis, improve existing treatments, and aid in the development of new therapies. Most important, advances in genomic analyses have helped to identify the underlying causes of disease in individual patients. However, many efforts have been hampered by an inadequate sample size and a lack of diversity among participants,[1] restrictive policies regarding data access, or failure to capture genotype and phenotype data comprehensively.[2] Collectively, these challenges have slowed the pace of medical discovery, decreased the generalizability of research findings, hindered reproducibility, and led to incorrect interpretations.[3,4] Population-based research, which requires large sample sizes and highly granular phenotypic data, benefits from access to populations of patients from various ancestries. The All of Us Research Program seeks to provide these data. Here, we describe the goals of the program and the extent to which they have been met.

In his State of the Union address in January 2015, President Barack Obama first announced the program (then called the Precision Medicine Initiative cohort program[5]), and the All of Us network of grant awardees received initial funding from the National Institutes of Health (NIH) in July 2016. A central goal of the program is to enroll at least 1 million persons who agree to share their electronic health record (EHR) data, donate biospecimens for genomic and other laboratory assessments, respond to surveys, and have standardized physical measurements taken. Participants will also have the opportunity to contribute data from sensors and mobile health devices and be contacted for future research opportunities. The All of Us cohort will thereby provide data for prospective, retrospective, and cross-sectional analyses. The program launched in May 2018; a year later, the program met more than one fifth of its recruitment goal of 1 million persons.

PARTICIPANT CHARACTERISTICS

The All of Us Research Program seeks to recruit persons in demographic categories that have been and continue to be underrepresented in biomedical research; such persons typically have relatively poor access to good health care.[6] Race, ethnic group, age, sex, gender identity, sexual orientation, disability status, access to care, income, educational attainment, and geographic location are therefore taken into account. Because racial and ethnic identities are more than a genetic construct, we seek to capture other social and behavioral determinants of health. Persons in underrepresented populations who are enrolled in the program will be prioritized for physical measurements and biospecimen collections. Among persons from whom biospecimens are obtained, the target percentage of persons in racial and ethnic minorities is more than 45% and that of persons in underrepresented populations is more than 75%.

The All of Us program does not focus on any particular set of diseases or health status. The inclusion of persons with a range of diseases (Table S1 in the Supplementary Appendix, available with the full text of this article at NEJM.org) will enable systematic studies of disease outcomes, medication effects, and other therapeutic approaches across various environmental, social, genomic, and economic contexts.

Participants have been involved in shaping the All of Us program since its initial conceptualization.[7] Guided by pilot studies,[8,9] the program has funded a network of 22

community partners to engage diverse populations and providers. The All of Us program is urrently recruiting adults (18 years of age) with capacity to provide informed consent. Protocols are being developed to enable enrollment of children and adolescents as well as cognitively impaired persons.

OVERVIEW OF RESEARCH PROTOCOL

Participants enroll digitally through the All of Us website (https://joinallofus.org) or a smartphone app. The electronic consent modules include explanatory videos with brief text, iconography, and formative questions. The correct answers (with explanations) are provided to participants if they answered incorrectly. After completion of the consent modules and enrollment, participants are given several baseline health surveys, each of which is designed to be completed in approximately 15 minutes (Table 1). When appropriate, the surveys use validated instruments and questions from other large cohort surveys. All questions were evaluated with cognitive and online testing in diverse populations before use.[10] The research protocol has been published previously.[11] The surveys are available in English and Spanish, and all the content for participants is targeted to a fifth-grade reading level.

Once a person consents to participate in the All of Us program, provides authorization to share EHR data, and completes the initial baseline survey of demographic information, the person is eligible to undergo an evaluation for physical measurements and biospecimen collection (Table 1) either at a health care provider organization (one that is part of the All of Us program) or through a "direct-volunteer" mechanism (Fig. 1). Health care provider organizations include regional medical centers, federally qualified health centers, and the Veterans Health Administration (Table S2 in the Supplementary Appendix). They recruit persons affiliated with their center and, in the aggregate, are expected to recruit the majority of the participants in the program. The direct-volunteer route allows persons who are not patients in a health care provider organization to enroll. These participants enroll online and visit a designated health clinic, blood bank, laboratory facility, or health care provider organization (Fig. 2).

Participants can be invited to complete additional health surveys, use smartphone and sensor devices for data collection, and join future ancillary studies that could involve additional in-person visits or biospecimen collections. Currently, participants can choose to share their Fitbit data with the All of Us program if they have a device; the program will expand to support oter devices as well. The program also uses passive means to collect longitudinal health data such as EHR data, including billing codes, laboratory and medication data, reports, and narrative content. Its data and research center (Fig. 1) will link to other data sets such as national death indexes, pharmacy data, health care claims data, and geospatially linked environmental data.

EHR data are being structured according to a common data model that has been used for other large initiatives.[13] Direct volunteers will be able to share their EHRs through Sync for Science, a new protocol developed by the Office of the National Coordinator for Health Information Technology, or through other third parties such as health information exchanges. The All of Us research protocol is expected to evolve in response to emerging

research opportunities; evaluations of the value, validity, and cost of current and potential protocol elements; and participant feedback.

RESEARCHERS' ACCESS TO DATA

Most study cohorts are required to have project-specific applications, approvals, and downloads in order to have data access. This process can take months to years, effectively excluding many researchers. Some large repositories, such as the U.K. Biobank,[14] promote broad data access by expediting project approval for researchers to access and download data within a matter of months. The All of Us program is storing data in a common cloud (online) environment. Approved researchers will be able to use Web-based tools and interactive cloud-based computing environments to explore the data and test hypotheses. Identifying information will be removed from all participant data that are available to researchers. A public browser to explore the data set (https://databrowser.researchallofus.org) was launched in May 2019, and initial tools for analysis will become available by early 2020.

Although most study cohorts that are similar to the All of Us program have project-specific criteria for researchers' access to data, the All of Us program has developed a "passport model" through which researchers will be approved for data access to study any topic that meets our criterion for allowable use.[15] To improve transparency for participants, the researchers' names and brief descriptions of their projects will be made publicly available. The researchers' identities will be verified, and they will receive training in the ethics of research involving humans and the analysis of All of Us data. The security of the software platforms has been evaluated according to the Federal Information Security Management Act and is frequently tested by both internal and external security experts. The centralized data system is easier to secure and permits better audit trails than systems that require users to download the data for analysis.

PARTICIPANTS' ACCESS TO DATA

Participants in All of Us will have access to their own data and most results of research testing. Currently, participants can see their own physical measurements and the survey information that they have supplied. We are developing protocols for the return of genetic, laboratory-assay, and EHR data to participants; this development is guided by participant feedback and internal and external experts.[9,16] While these protocols are under development, we are prioritizing the return of information regarding two classes of health-related, actionable genetic variants to participants who have elected to receive such results. Pharmacogenomic results will be returned according to the guidance of the Clinical Pharmacogenetics Implementation Consortium.[17] In addition, participants will be informed if they have highly penetrant genetic variants known to cause serious diseases for which there are established interventions; this reporting will be consistent with the recommendations of the American College of Medical Genetics and Genomics (ACMG) for reporting of incidental findings related to 59 genes.[18] We estimate that over the course of collecting and analyzing data on the entire cohort, approximately 30,000 persons will receive actionable findings for ACMG genes and more than 90% may learn of actionable

pharmacogenomic variants.[19,20] The All of Us program is in the process of establishing a genetic counseling resource for participants who receive actionable genetic results.

CURRENT STATUS OF THE PROGRAM

A national network of research program partners and community partners focusing on recruitment, data management, engagement, and technology platforms was established in July 2016 (Fig. 3). After security testing and approval of the data and Web platforms, approval by the institutional review board, initial development of methods to capture participant-provided information, establishment of a central biobank, and pilot testing across the United States, the beta phase began with the first participant enrolling on May 31, 2017. The primary goals of the beta phase were testing and improving the process and training, ensuring operational integrity, and developing additional content. More than 27,000 core participants (persons who contributed personal information, biospecimens, and physical measurements and who agreed to share EHR data) were enrolled in the beta phase.

By the national launch in May 2018, all health care provider organizations and direct-volunteer sites had started recruitment in at least one location. Currently, more than 340 recruitment sites, mostly within health care provider organizations, have been established; new sites will continue to be launched on a rolling basis (Fig. 2). As of July 2019, the All of Us program had enrolled more than 175,000 core participants and more than 230,000 total participants; an additional 40,000 participants have registered on the website. Among core participants, 51% are nonwhite and 80% meet our definition of being underrepresented in biomedical research; these percentages exceed the program goals. A total of 34 health care provider organizations have uploaded EHR data on more than 112,000 participants. The recruitment rate, which is currently approximately 3100 core participants per week, is predictably increasing as new sites begin enrollment and existing sites mature. We expect to have enrolled 1 million core participants by approximately 2024.

DISCUSSION

With a plan to follow the health and outcomes of participants over decades, the All of Us program should enable research that provides an improved understanding of health and disease, which in turn would support accurate diagnoses, rational disease-prevention strategies, treatment selection, and the development of targeted therapies (Table 2). Although we anticipate that validation studies and research projects will begin early in the program, the greater usefulness of the program will be seen as it matures, and its contributions will continue to accrue over decades, as has been noted in previous large cohort studies.[14,21–23]

To be successful, the All of Us program will need to be of value to its constituents: participants, health care providers, and researchers. Establishing authentic engagement with participants and providing value for them will be key to long-term retention and continued recruitment of persons from diverse populations. The All of Us program must resist cyberattacks and effectively communicate the legal and procedural protections of privacy such as those afforded by the 21st Century Cures Act.[24,25] The program must also minimize burdens on providers and researchers while providing a rich resource and effective tools.

National resources such as the U.K. Biobank,[14] the Million Veteran Program,[26] the China Kadoorie Biobank,[27] and other research groups[19,28] have shown the power of very large cohorts for biomedical discovery. The All of Us program has also learned from the challenges of the National Children's Study. These challenges included the need for defined scientific goals and leadership, the need for pilot studies of specific modules followed by their rapid implementation, and clear recruitment goals. Unique aspects of the program include a focus on diversity, the nationwide scale and accessibility, and the return of data to participants. "Passive" collection of data from EHRs and linkage with other data sources (Table 1) should lead to a rich collection of information and cost-efficient ascertainment of participant outcomes.

Whenever possible, All of Us software and resources will be free for use by other researchers. The research environments and access models are designed to facilitate open, collaborative, and reproducible science.[2] The All of Us program is harmonizing EHR data across sites and EHR vendors and reconciling other data, such as validated surveys and research measures, into a common data model in order to accelerate research interoperability and facilitate large-scale research such as machine-learning applications.[29,30]

The All of Us program faces substantive challenges. Reaching persons who live in rural areas or far from recruitment sites may require alternative enrollment protocols such as the substitution of the use of blood specimens for DNA collection with the use of saliva specimens from kits that can be mailed. The techniques to collect EHR data from direct volunteers are in their infancy. Harmonization of EHR data elements can be challenging, but there are examples of success.[13,31]

Since 2015, Congress has allocated $1.02 billion to the All of Us program; this includes funding for all current resources through 2019 as well as initial genome sequencing, initial bioassays, and establishment of a nationwide genetic counseling resource for participants to whom results are returned. The 21st Century Cures Act authorized an additional $1.14 billion through 2026.[24]

Some researchers have described other challenges or expressed skepticism about the effect of the program and about precision medicine more generally.[32–35] Because the All of Us program was not designed with a formal statistical sampling method, its use for some epidemiologic studies may be limited. EHR data can be fragmented, incomplete, and inaccurate[32,36]; however, in certain contexts, such data have already proved to be a powerful and efficient discovery tool.[37–39] The translation of biologic and environmental discoveries into improved health is unlikely to occur quickly, and the All of Us program faces challenges that it does not directly address, such as health care delivery and variable insurance coverage.[40] We are therefore pursuing studies to evaluate our measurements, guide development of the protocol, and examine the quality and usefulness of the data. These demonstration projects will compare information provided from different sites and types of data (e.g., EHR and survey information) and test for known genotype–phenotype and disease–exposure associations as a type of positive control.

Association does not necessarily translate into causation[32]; nonetheless, the combination of prospectively collected measures and the use of approaches such as mendelian randomization[41] and genomewide polygenic risk scores[42] can provide insight into causation and medical usefulness for future clinical trials, which could themselves leverage the All of Us program for recruitment. Our hope is that identification of risk factors and biomarkers (including environmental exposures, habits, and social determinants) will improve population health by bringing about more efficient and accurate diagnosis and screening, better understanding of diverse populations, more rational use of existing therapeutics, and the development of new treatments.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature 2016; 538: 161–4. [PubMed: 27734877]

2. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016; 3: 160018. [PubMed: 26978244]

3. Manrai AK, Funke BH, Rehm HL, et al. Genetic misdiagnoses and the potential for health disparities. N Engl J Med 2016; 375: 655–65. [PubMed: 27532831]

4. Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. Nature 2014; 505: 612–3. [PubMed: 24482835]

5. Collins FS, Varmus H. A new initiative on precision medicine. N Engl J Med 2015; 372: 793–5. [PubMed: 25635347]

6. Agency for Healthcare Research and Quality. 2016 National healthcare quality and disparities report (https://www.ahrq.gov/research/findings/nhqrdr/nhqdr16/index.html).

7. The Precision Medicine Initiative Cohort Program — building a research foundation for 21st century medicine: Precision Medicine Initiative (PMI) working group report to the advisory committee to the director, NIH. Bethesda, MD: National Institutes of Health, 2015 (http://www.nih.gov/precisionmedicine/09172015-pmi-working-group-report.pdf).

8. Kaufman DJ, Baker R, Milner LC, Devaney S, Hudson KL. A survey of U.S adults' opinions about conduct of a nationwide Precision Medicine Initiative cohort study of genes and environment. PLoS One 2016; 11(8): e0160461. [PubMed: 27532667]

9. Wilkins CH. Precision medicine for everyone. NEJM Catalyst. February 12, 2018 (https://catalyst.nejm.org/precision-medicine-initiative-everyone/).

10. Cronin RM, Jerome RN, Mapes B, et al. Development of the initial surveys for the All of Us Research Program. Epidemiology 2019; 30: 597–608. [PubMed: 31045611]

11. National Institutes of Health. The *All of Us* Research Program initial protocol. August 4, 2017 (https://allofus.nih.gov/news-events-and-media/announcements/all-us-research-program-initial-protocol).

12. Demographics Research Group. The racial dot map. Charlottesville, VA: Weldon Cooper Center for Public Service, 7 2013 (https://demographics.coopercenter.org/racial-dot-map/).

13. Hripcsak G, Ryan PB, Duke JD, et al. Characterizing treatment pathways at scale using the OHDSI network. Proc Natl Acad Sci U S A 2016; 113: 7329–36. [PubMed: 27274072]

14. Sudlow C, Gallacher J, Allen N, et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015; 12(3): e1001779. [PubMed: 25826379]

15. Workbench — All of Us research hub (https://www.researchallofus.org/data/workbench/).

16. National Institutes of Health. Genomics Working Group of the *All of Us* Research Program Advisory Panel (https://allofus.nih.gov/genomics-working-group-all-us-research-program-advisory-panel).

17. Relling MV, Klein TE. CPIC: Clinical Pharmacogenetics Implementation Consortium of the Pharmacogenomics Research Network. Clin Pharmacol Ther 2011; 89: 464–7. [PubMed: 21270786]

18. Kalia SS, Adelman K, Bale SJ, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. Genet Med 2017; 19: 249–55. [PubMed: 27854360]

19. Dewey FE, Murray MF, Overton JD, et al. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. Science 2016; 354: aaf6814-1–10.

20. Van Driest SL, Shi Y, Bowton EA, et al. Clinically actionable genotypes among 10,000 patients with preemptive pharmacogenomic testing. Clin Pharmacol Ther 2014; 95: 423–31. [PubMed: 24253661]

21. Kannel WB, Dawber TR, Kagan A, Revotskie N, Stokes J III. Factors of risk in the development of coronary heart disease — six year follow-up experience: the Framingham Study. Ann Intern Med 1961; 55: 33–50. [PubMed: 13751193]

22. Rosenquist JN, Lehrer SF, O'Malley AJ, Zaslavsky AM, Smoller JW, Christakis NA. Cohort of birth modifies the association between FTO genotype and BMI. Proc Natl Acad Sci U S A 2015; 112: 354–9. [PubMed: 25548176]

23. Chambless LE, Heiss G, Folsom AR, et al. Association of coronary heart disease incidence with carotid arterial wall thickness and major risk factors: the Atherosclerosis Risk in Communities (ARIC) Study, 1987–1993. Am J Epidemiol 1997; 146: 483–94. [PubMed: 9290509]

24. Hudson KL, Collins FS. The 21st Century Cures Act — a view from the NIH. N Engl J Med 2017; 376: 111–3. [PubMed: 27959585]

25. H.R. 34, 114th Cong.: 21st Century Cures Act (2015-2016). Congress.gov (https://www.congress.gov/bill/114th-congress/house-bill/34).

26. Gaziano JM, Concato J, Brophy M, et al. Million Veteran Program: a mega-biobank to study genetic influences on health and disease. J Clin Epidemiol 2016; 70: 214–23. [PubMed: 26441289]

27. Chen Z, Chen J, Collins R, et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. Int J Epidemiol 2011; 40: 1652–66. [PubMed: 22158673]

28. Gottesman O, Kuivaniemi H, Tromp G, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genet Med 2013; 15: 761–71. [PubMed: 23743551]

29. Rajkomar A, Oren E, Chen K, et al. Scalable and accurate deep learning with electronic health records. NPJ Digital Med 2018; 1(1): 18.

30. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. Sci Rep 2016; 6: 26094. [PubMed: 27185194]

31. Crawford DC, Crosslin DR, Tromp G, et al. eMERGEing progress in genomics-the first seven years. Front Genet 2014; 5: 184. [PubMed: 24987407]

32. Galea S, Abdalla SM. Precision medicine approaches and the health of populations: study design concerns and considerations. Perspect Biol Med 2018; 61: 527–36. [PubMed: 30613035]

33. Joyner MJ, Paneth N. Seven questions for personalized medicine. JAMA 2015; 314: 999–1000. [PubMed: 26098474]

34. Khoury MJ, Evans JP. A public health perspective on a national precision medicine cohort: balancing long-term knowledge generation with early health benefit. JAMA 2015; 313: 2117–8. [PubMed: 26034952]

35. Psaty BM, Dekkers OM, Cooper RS. Comparison of 2 treatment models: precision medicine and preventive medicine. JAMA 2018; 320: 751–2. [PubMed: 30054607]

36. Wei W-Q, Leibson CL, Ransom JE, et al. Impact of data fragmentation across healthcare centers on the accuracy of a high-throughput clinical phenotyping algorithm for specifying subjects with type 2 diabetes mellitus. J Am Med Inform Assoc 2012; 19: 219–24. [PubMed: 22249968]

37. Kho AN, Pacheco JA, Peissig PL, et al. Electronic medical records for genetic research: results of the eMERGE consortium. Sci Transl Med 2011; 3: 79re1.

38. Denny JC, Bastarache L, Ritchie MD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. Nat Biotechnol 2013; 31: 1102–10. [PubMed: 24270849]

39. Giri A, Hellwege JN, Keaton JM, et al. Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. Nat Genet 2019;51: 51–62. [PubMed: 30578418]

40. Khoury MJ, Galea S. Will precision medicine improve population health? JAMA 2016; 316: 1357–8. [PubMed: 27541310]

41. Emdin CA, Khera AV, Kathiresan S. Mendelian randomization. JAMA 2017; 318: 1925–6. [PubMed: 29164242]

42. Khera AV, Emdin CA, Drake I, et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease. N Engl J Med 2016; 375: 2349–58. [PubMed: 27959714]
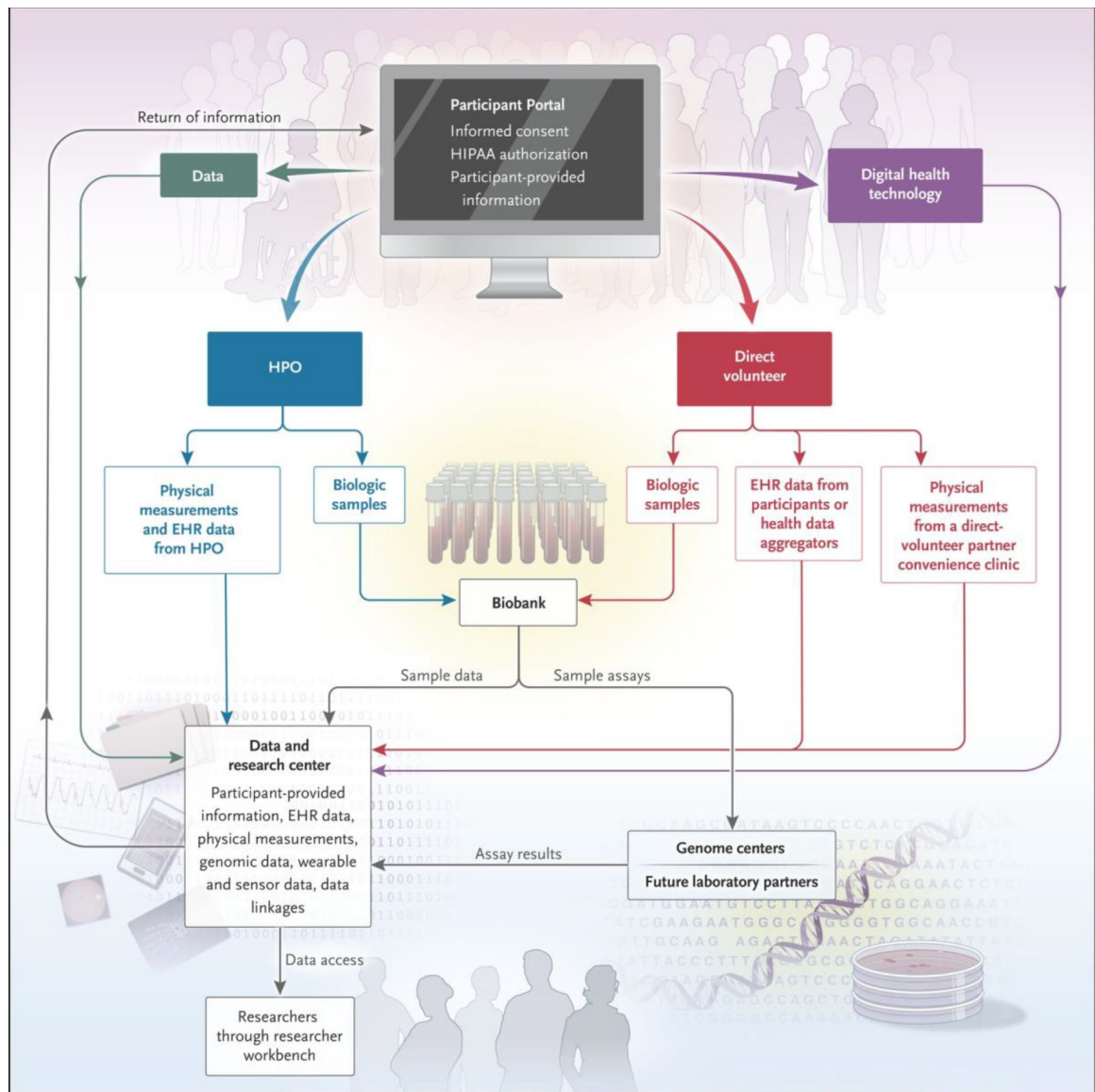
Figure 1. Participant, Data, and Biospecimen Pathways in the All of Us Program.
Genome centers in the All of Us program generate genomic data from biosamples. The
researcher workbench is the platform for data analysis in the program. EHR denotes
electronic health record, HIPAA Health Insurance Portability and Accountability Act, and
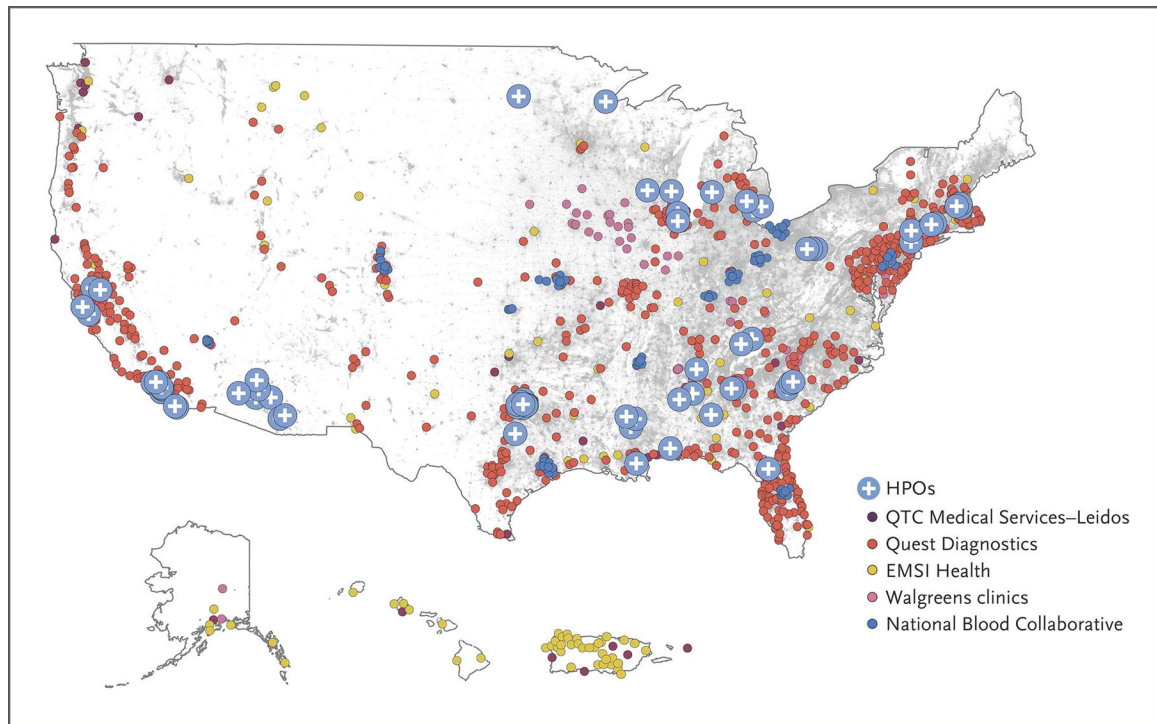HPO health care provider organization.

Figure 2. Planned Recruitment Sites and Network Partners in the All of Us Program.
Although most recruitment in the All of Us program is digital, in-person visits are required
for physical measurements and collection of biospecimens. Currently, more than 340
recruitment sites have begun enrolling participants for these in-person assessments at HPOs
and at the centers shown (QTC Medical Services–Leidos, Quest Diagnostics, EMSI Health,
Walgreens clinics, and the National Blood Collaborative. The gray background of the map
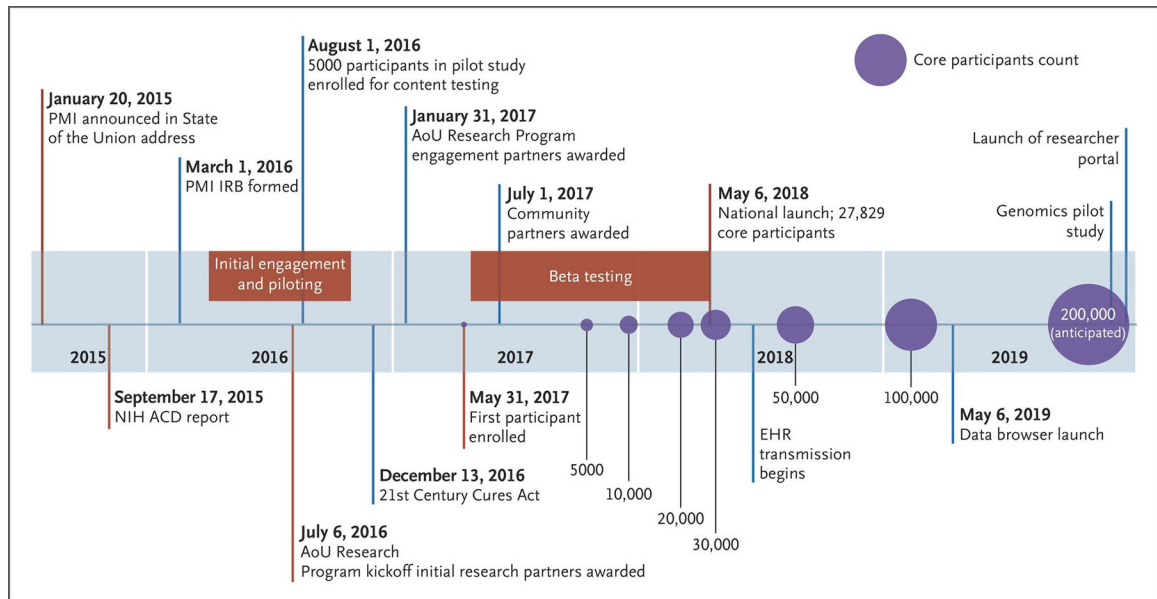indicates population density, with one dot per person in the 2010 census.[12]

Figure 3. Timeline of the All of Us Program.

The Precision Medicine Initiative (PMI) Working Group of the Advisory Committee (ACD) to the National Institutes of Health (NIH) Director generated the initial blueprint for the structure and goals for the All of Us (AoU) program.[7] The core participants count includes persons who have completed health surveys, agreed to share their EHR data, had physical measurements taken, and contributed biospecimens. Counts and dates after July 2019 are estimates. IRB denotes institutional review board.

Table 1.

Data Available to Researchers from the All of Us Cohort.[*]

Data Source	Details
Current sources	
Health surveys	Initial surveys include information on sociodemographic characteristics, overall health, lifestyle, and substance use, with subsequent modules covering personal and family medical history and access to health care.
Physical measurements	Per-protocol measurements include blood pressure, heart rate, weight, height, body-mass index, and hip and waist circumferences.
Biospecimens[†]	Blood and urine samples are tested for DNA, RNA, cell-free DNA, serum, and plasma. If blood specimens cannot be obtained, saliva specimens are obtained.
Electronic health records	Initial capture of structured data includes billing codes, medication history, laboratory results, vital signs, and encounter records from health care provider organizations. Records will be expanded to include narrative documents. Pilot studies are testing data collection through Sync for Science and other health data aggregators.
Digital health information	Data can be captured from compatible participant-owned devices such as Fitbit. Pilot studies of other devices and linkage to health apps are being explored.
Future sources	
Health surveys	Additional modules, including surveys regarding social behavioral determinants of health, are under development.
Bioassays	Pilot studies for genotyping and whole-genome sequencing are expected to begin by early 2020. Additional pilot studies of bioassays are planned.
Health care claims data	Systems for the use of claims data, including billing codes and medication data, are under development.
Geospatial and environmental data	These data include geospatial linkage to measures such as weather, air quality, pollutant levels, and census data. Assays and sensor-based measurements of exposure are under consideration.
Other sources	Voluntary contributions of data from social networks (e.g., Twitter feeds) and additional biospecimen collections are under consideration.

[*]
Additional information is available at https://www.researchallofus.org/data.

[†]
Types of biospecimens are listed in Table S3 in the Supplementary Appendix.

Table 2.

Scientific Goals of the All of Us Program and Expected Timelines.[*]

Goal	End of 2018 (N = 94,000)	End of 2019 (N =>200,000)	Years		
			2020–2022 (N = <650,000)	2023–2027 (N = >1 million)	After 2027 (N = >1 million)
Return data to participants	+	+	+++	+++	++++
Establish demonstration projects[†]		+	+++	+	+
Discover genetic and environmental correlates with disease			++	+++	++++
Improve predictions of therapeutic safety and efficacy			++	+++	+++
Discover disease biomarkers			++	+++	+++
Connect mobile health, digital health, and sensor data with clinical outcomes			++	+++	+++
Develop new disease classifications			+	+++	++++
Support clinical trials			+	+++	+++
Enable machine-learning applications			++	+++	++++
Improve understanding of health disparities			++	+++	+++
Develop and test new therapeutic agents					++

[*] The expected number of participants in the cohort is shown for each time period. The number of plus signs in each cell indicates the anticipated relative degree to which each goal may be accomplished during the estimated timeline for focused research.

[†] Demonstration projects are scientific studies implemented by the All of Us program to show the quality, usefulness, validity, and diversity of the All of Us research data set and platform. In these projects, the population and data are further characterized, and the data are evaluated with a view to determining whether known associations can be replicated.