

ERRATUM

How Fitbit data are being made available to registered researchers in All of Us Research Program

Hiral Master, Aymone Kouame, Kayla Marginean, Melissa Basford, Paul Harris
Vanderbilt University Medical Center Nashville TN, USA
*Email: hiral.master@vumc.org, aymone.kouame@vumc.org, kayla.marginean@vumc.org,
melissa.basford@vumc.org, paul.a.harris@vumc.org*

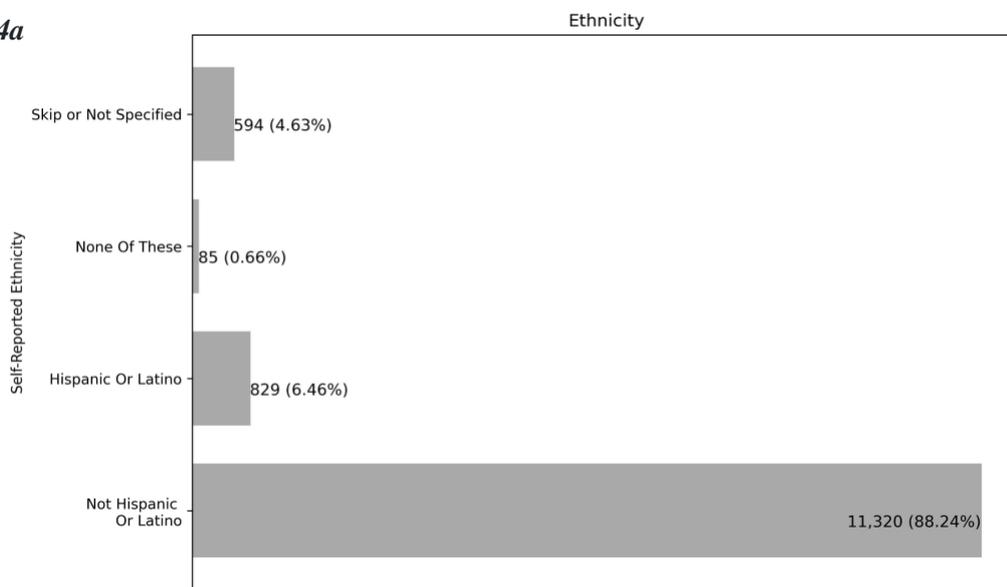
Michelle Holko
Google Public Sector Washington DC, USA
Email: michelle.holko@gmail.com

*In the above PSB article published in Biocomputing 2023: Proceedings of the Pacific Symposium, pp. 19-30; PMID: PMC9811842; PMID: [36540961](https://pubmed.ncbi.nlm.nih.gov/36540961/)
The following correction has been made.*

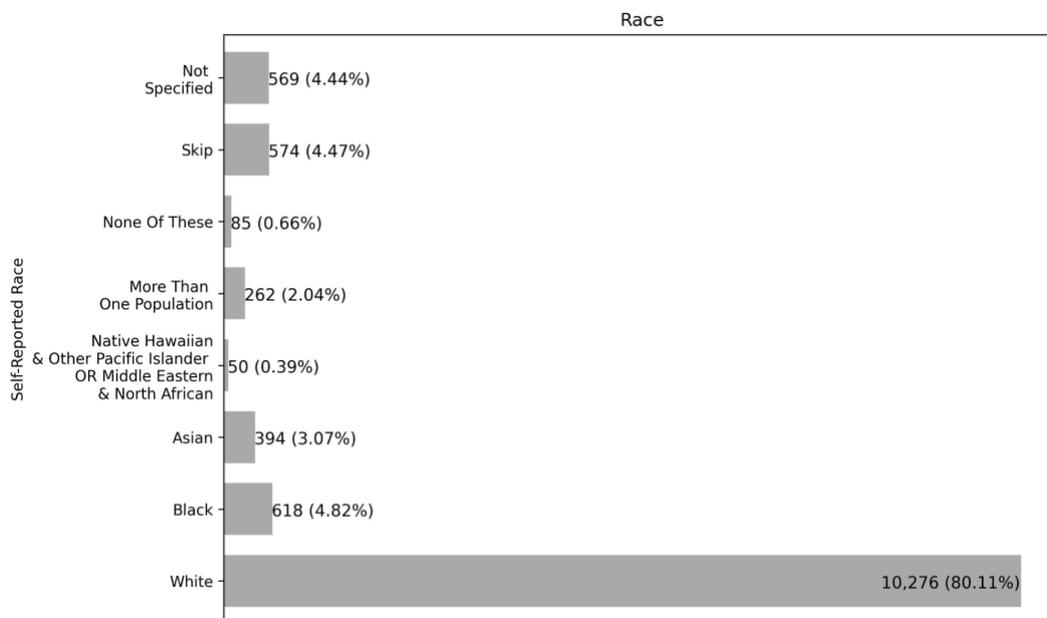
Author Correction

In the version of [this](#) PSB 2023 conference full-length paper, the authors had mistakenly shown counts in the figures such that counts <20 for some categories could be derived using mathematical formula. The authors apologize for this unintentional error. Therefore, authors have updated figures 4a, 4b and 4d to ensure exact participants counts for categories <20 cannot be derived using mathematical formula. This update was done to ensure participants' privacy and follow *All of Us* Data and Statistics Dissemination Policy.

4a



4b



4d

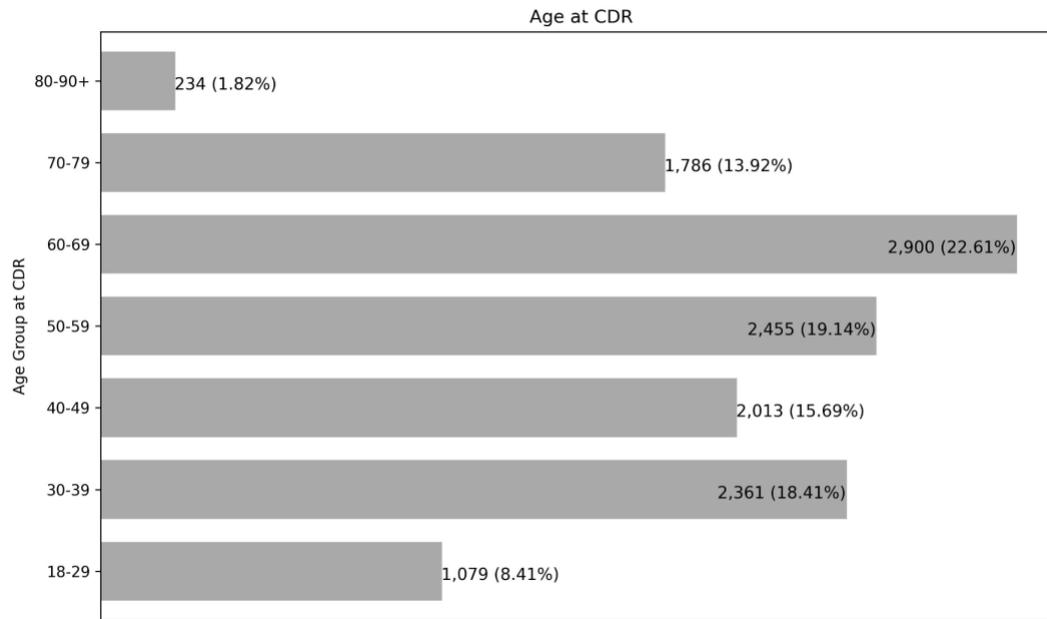


Fig. 4. Self-reported a) ethnicity, b) race, d) age of participants with Fitbit data in June 2022 curated data repository, which can be accessed by registered users on Researcher Workbench.

Figures have been updated to ensure exact participants counts for categories <20 cannot be derived using mathematical formula as per the *All of Us* Data and Statistics Dissemination Policy.

How Fitbit data are being made available to registered researchers in *All of Us* Research Program

Hiral Master, Aymone Kouame, Kayla Marginean, Melissa Basford, Paul Harris

Vanderbilt University Medical Center

Nashville TN, USA

Email: hiral.master@vumc.org, aymone.kouame@vumc.org, kayla.marginean@vumc.org, melissa.basford@vumc.org, paul.a.harris@vumc.org

Michelle Holko

Google Public Sector

Washington DC, USA

Email: michelle.holko@gmail.com

The National Institutes of Health's (NIH) *All of Us* Research Program aims to enroll at least one million US participants from diverse backgrounds; collect electronic health record (EHR) data, survey data, physical measurements, biospecimens for genomics and other assays, and digital health data; and create a researcher database and tools to enable precision medicine research [1]. Since inception, digital health technologies (DHT) have been envisioned as essential to achieving the goals of the program [2]. A "bring your own device" (BYOD) study for collecting Fitbit data from participants' devices was developed with integration of additional DHTs planned in the future [3]. Here we describe how participants can consent to share their digital health technology data, how the data are collected, how the data set is parsed, and how researchers can access the data.

Keywords: Wearables, Digital health technologies, Precision medicine

1. Introduction

In 2016, the U.S. Congress, via the 21st Century Cures Act, authorized a total of \$1.5 billion over ten years to fund the *All of Us* Research Program at the National Institutes of Health (NIH). This program is publicly funded, with resources appropriated each year by the U.S. Congress. The program was borne out of the Precision Medicine Initiative, and strives to nurture relationships with participants, build a robust ecosystem of communities and researchers, and strives to deliver the largest and most diverse biomedical dataset. The program is accumulating multiple streams of health-related information such as electronic health records (EHRs), genomics, physical measures, participant surveys and wearables (such as Fitbit) from 1,000,000 or more Americans, with a focus

on populations usually under-represented in biomedical research to date [1, 2].

In addition to EHR, genomics, physical measures and surveys, the program has an interest in incorporating digital health data, e.g., data from wearable devices like fitness trackers, to promote research in this space by diverse academic researchers on a diverse data set. The program currently invites participants to donate Fitbit and Apple HealthKit data in a “bring your own device” (BYOD) model [3, 4]. As of June 2022, Fitbit data for 12,844 *All of Us* Research Program participants were provided to registered researchers on the secure, cloud-based Researcher Workbench platform. This report is focused on the back-end process by which participants can link their own Fitbit device, and what happens to this Fitbit data once they are shared with the program. We will discuss the current processes that are being employed to provide these data to the research community, and how researchers can access these data via *All of Us* Researcher Workbench platform.

In this report, we provide a high-level overview of the Fitbit data process from data ingestion to delivery. This report also provides a high-level overview on demographic characteristics, such as ethnicity, race, sex at birth, age, income, and employment of participants who contributed any Fitbit data in the *All of Us* Research Program. Additional digital health technology data streams are planned for the longer term of this study. Lastly, the report also highlights some unique opportunities on how digital health data from *All of Us* Research Program can be leveraged by registered researchers to advance healthcare for all.

2. Methods

2.1. How are participants consented to be part of AoU and share Fitbit data?

Participants may log on to the *All of Us* participant portal at <https://participant.joinallofus.org> to participate in the program. Participants need to provide primary consent to be part of the *All of Us* program, which aims to collect at least 10 years of data from participants. Given it is a long-term research program, participants remain in touch with the program via phone, email, and/or app. They may also connect their family members, in case participants cannot be reached. They might also use social media or public databases to help keep participant’s contact information up to date. If participants have a fitness tracker, they may be asked to share data from it. **Figure 1** shows the *All of Us* Research Program participant facing portal where participants can elect to share Fitbit data with the program. Participants can withdraw from *All of Us* any time. Consent to share electronic health record (EHR) data is mandatory before participants can start sharing digital data. Once the consent to EHR is completed, participants can share their digital health data.

Data sharing process on participant portal (<https://participant.joinallofus.org>) lists the steps for deciding whether or not to share or not share data from their own Fitbit devices:

1. First, participants are provided with the program’s working definitions for wearables. “Mobile apps and wearable devices can collect data outside of a hospital or clinic.”
2. Participants are then shown the steps for securely sharing digital health data.
3. After confirming that they would like to share their data, participants are prompted to log into their Fitbit account to pair their device with their *All of Us* account.
4. Once a participant selects “approved”, they are then redirected to the participant portal and are shown a success message.

Donation of digital health data is optional for participants. Participants may withdraw from participation or stop contributing data via the Connector at any time by revoking access for each individual data record type, or all data record types related to *All of Us* Research Program as a data sharing endpoint via the appropriate application. Participants may choose to re-enable their data sharing at any time, for each individual data record type, or all data record types. Data previously contributed by participants will remain with the *All of Us* Research Program after a participant’s program withdrawal and will not be retroactively scrubbed.

2.2. *What happens to participants' data?*

The Participant Technology Systems Center (PTSC) securely stores all the Fitbit data on the cloud platform. Files are delivered by the PTSC to the Data and Research Center (DRC) at Vanderbilt University Medical Center. Specifically, files are uploaded in the Raw Data Repository (RDR) daily. **Figure 2** shows the flow of participant digital health data from the PTSC to the DRC. These data are structured as json files. These data then undergo curation in BigQuery, and are made available to researchers on the *All of Us* Researcher Workbench, a cloud-based platform.

3. *Results*

3.1. *How are Fitbit data parsed (Schema development)?*

The DRC uses a hands-off approach to data processing and delivery to support a wide range of scientific research investigations. Specifically, Fitbit data are available in json format, which is considered raw data. The contents of the filename are mapped to a single field and contents within each file are mapped into another field. These file contents are then parsed into a series of tables for data types, including:

- Heart Rate (By Zone Summary)
- Heart Rate (Minute-Level)
- Activity (Daily Summary)
- Activity Intraday Steps (Minute-Level)

Files are then mapped from the bucket to a Postgres database in a secure FISMA VM. The final output is mapped on the BigQuery database, which undergoes curation pipeline. During this process, the data are de-identified and are then being made available on Researcher Workbench as supplemental (non-OMOP) tables.

3.2. *How can Fitbit data be accessed by researchers?*

The program's goal is to share data widely but wisely to ensure rigorous measures are taken to protect participants' privacy. Therefore, the Fitbit data is delivered to researchers in a tiered approach. Specifically, the summary level information regarding the data can be accessed publicly via the website (<https://www.researchallofus.org/>). Researchers can access row-level de-identified data via the *All of Us Researcher workbench*, which is a secured cloud-based platform. On Researcher Workbench, researchers can access de-identified Fitbit data in Registered and Controlled tiers. In the registered tier, Fitbit data are date-shifted by random number between 1 to 365 to ensure participant's privacy. No date-shifting is performed in the controlled tier.

In order to access the de-identified data on the secured, cloud based platform, researchers need to create a Researcher Workbench account. The researcher must be a part of an institution that has a data use agreement. Currently, the list of institutions that have agreements in place can be viewed publicly on the website (<https://www.researchallofus.org/institutional-agreements/>). If the researcher's organization does not currently have a data use agreement in place, they can initiate this process by submitting a form online. Upon submission of request, the contracting officer from Vanderbilt University Medical Center reaches out to contacting contact from the requestor's institution within a couple of business days. Timeline to complete this process and obtain agreement varies based on workflows around the contracting processes at the requestor's institution. Once the institutional agreement is in place, the individual researcher can create an account and go through the relevant questionnaires and ethics training to validate their account. At present, any US-based academic, nonprofit, or health care institution can obtain data use agreement and there is no process for researchers outside the United States, or for researchers in the private sector to access the Workbench. However, expanding access to these groups is a priority for the program and a goal for future development.

3.3. *What Fitbit data are currently being made available on the Researcher Workbench?*

Currently, 12,844 *All of Us* Research Program participants provide any Fitbit data, which can be accessed via Researcher Workbench (**Table 1**). Nearly 13% of participants who provided Fitbit data resided in California state at the time of enrollment in the program (**Figure 3**). Of the participants who provide any Fitbit data, 80% are white, 88% are Non-Hispanic or Latino, 67% are Female at birth and 52% report being employed for wages (**Figure 4**). The detailed cohort characterization report is now publicly available on User Support Hub article [5].

4. Discussion

4.1. Research Utility

Digital health data on Researcher Workbench represent the data that are parsed from json files to structured tables. Specifically, this Fitbit data is longitudinal in nature. Thus, these device-generated summary and high-resolution intraday data are robust in nature and allow a wide-range of research, including method development and longitudinal study design. Currently, registered users can subset their analytical sample by presence of any Fitbit data by using graphical interface tools (e.g. cohort and dataset builder). However, there is an opportunity to develop various tools that would further wearable research. For instance, researchers on the platform can work on innovative projects and share their work with other registered users on Researcher Workbench. A couple examples of tools and methods that would be helpful to incorporate into the platform are for feature detection, e.g. periods of exercise and user behavior for wearing a Fitbit device). Time-series based tools, and methods to deal with data missingness over time (e.g. when charging or generally when the device is not worn or not functioning) will also be useful. Thus, these data support the program's overarching mission of accelerating health research and medical breakthroughs by enabling researchers to conduct various types of studies, including cross-sectional and longitudinal research designs.

4.2. Lessons Learned

Our initial work has provided insight and lessons that may be generalizable and applicable for other programs aiming to collect and share BYOD digital health data. These include establishing the system to integrate digital health data in cloud platforms and making decisions on how to deliver this large digital health data in sustainable and accessible fashion. Currently, we provide the digital health data as separate structured data tables on the cloud platform. Since the digital health data is collected from participant's own devices, the data is collected right from the time their Fitbit account was created, which gives opportunity for researchers to conduct longitudinal study design research projects.

4.3. Limitations of dataset

The characterization for digital health data is limited to specific data types such as activity and heart rate. Today, the standardized fashion of managing digital health data is in its infancy state, therefore, these data are being made available as separate datatables on Researcher Workbench. We acknowledge that the majority of participants whose Fitbit data is being made available on Researcher Workbench is biased, i.e., majority of participants who provided Fitbit data reported being White and employed for wages. However, these data represent participants who had their own Fitbit devices and consented to share EHR data. The program is currently expanding the

efforts by providing Fitbit devices to *All of Us* Research Program participants who do not own Fitbit devices so they can participate and share their data [6]. Lastly, we acknowledge that access to row-level deidentified data is currently available to researchers who are part of an institution in the United States that has an institutional data use agreement in place. However, the program has initiated efforts to expand access globally and foster public-private relationships, ensuring programs' goals and mission are met.

4.4. *Future plans*

We plan to expand the digital technology data offerings not only in terms of providing more participant's data but also adding more data types and includes data from devices (e.g., Apple HealthKit, Garmin, etc.) in addition to Fitbit. For instance, in future, we plan to provide sleep and device information from Fitbit, which will expand the research use cases.

5. *Conclusion*

Digital Health Technologies are increasingly being used for health-related applications. The *All of Us* Research Program has a unique opportunity to continue to drive research using these devices, to understand how these data can be used to support individuals in their health journeys. Integrating additional devices, and collecting and making additional data available to researchers, will help contribute to a robust ecosystem for researchers. In addition, tools to help researchers analyze these data are needed. These can be developed both by the program and by the researcher community. Finally, promoting diversity, not only in the data set but also in the researchers analyzing the data, is important for reducing bias and inequity of results.

5.1. *Table*

Fitbit data type	Count of participant ids
Any Fitbit data type	12,844
Activity summary	12,794
Heart summary	11,575
Step intraday	12,790
Heart rate intraday	11,575

Table 1. Counts of participants who provide Fitbit by data type as of data, which is available on Researcher Workbench, starting June 22, 2022 (N Fitbit Pid = 12,844)

5.2. *Figures/Illustrations*

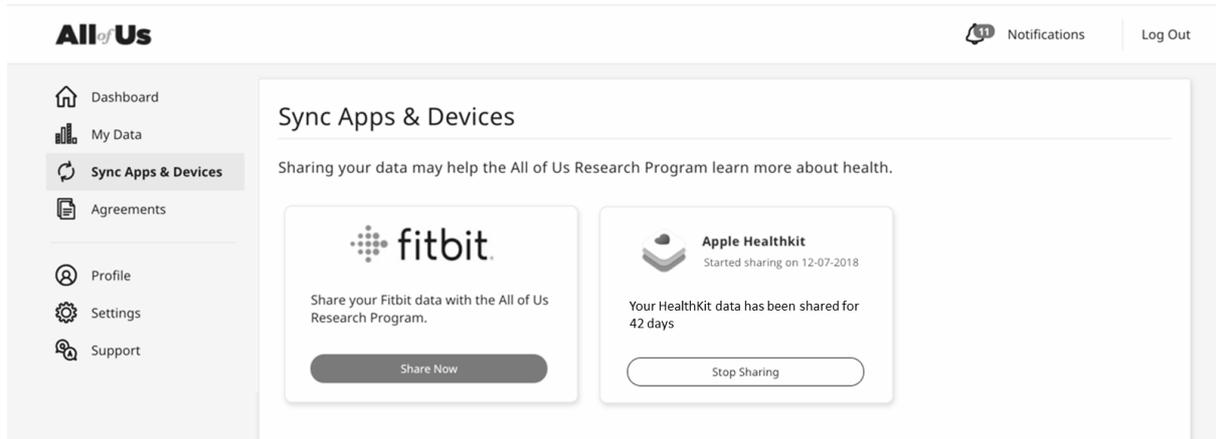


Fig. 1. *All of Us* Research Program participant facing portal where participants can share Fitbit data

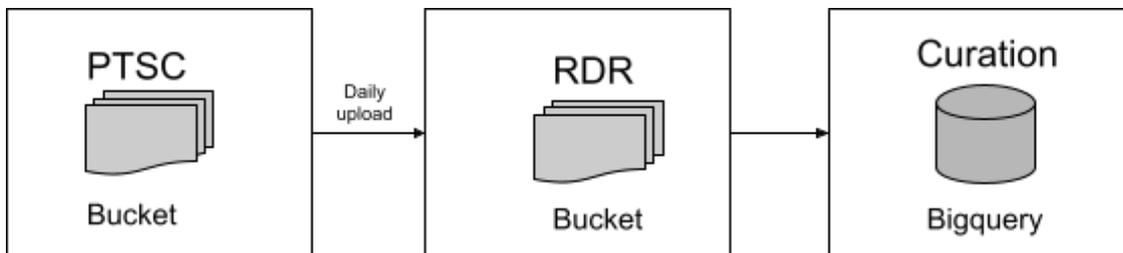


Fig. 2. *Flow of Fitbit data from the participant portal to the Data and Resource Center's raw data repository and curated data repository.*

US Map of All of Us participants with Fitbit data, N= 12,844

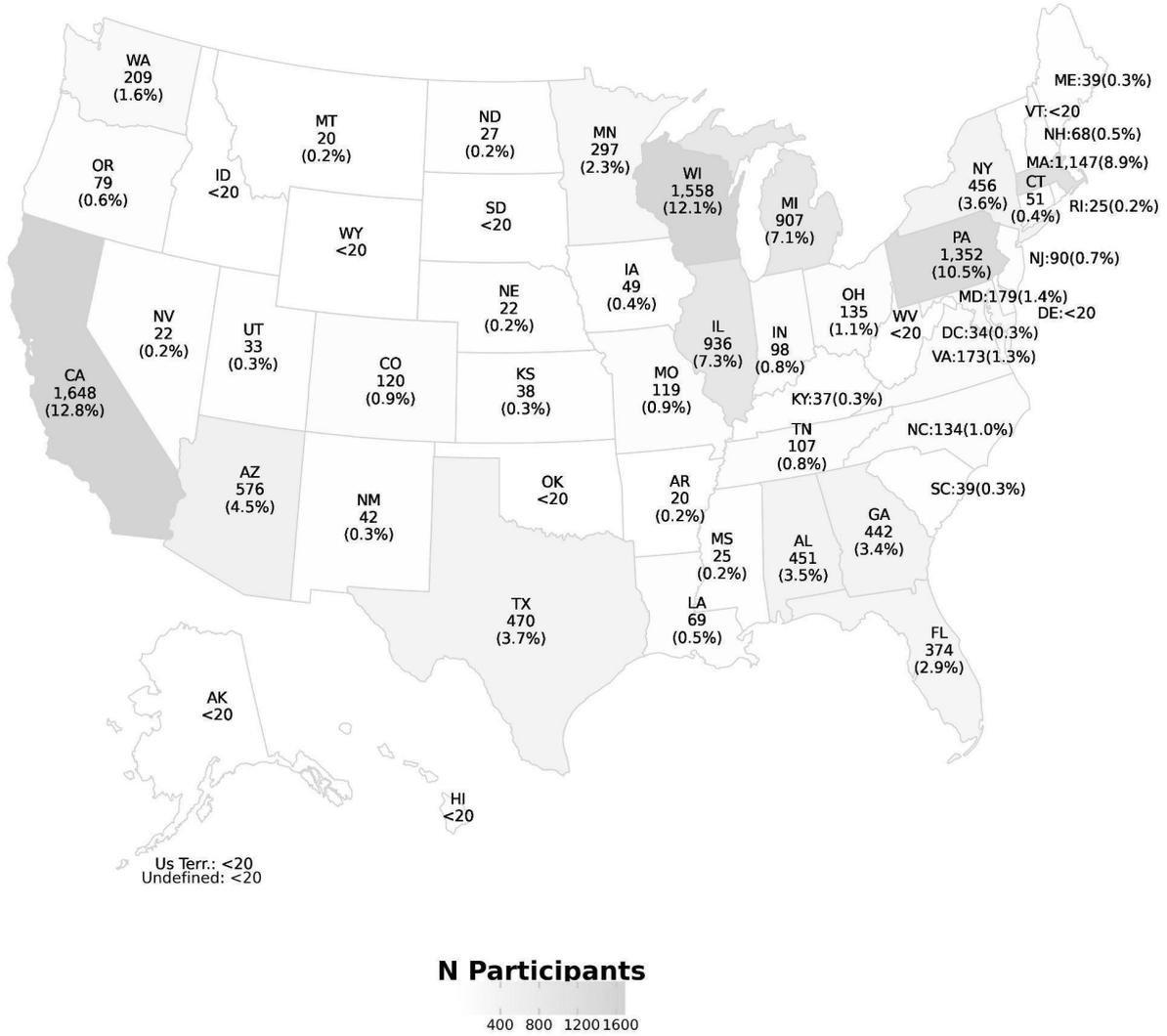
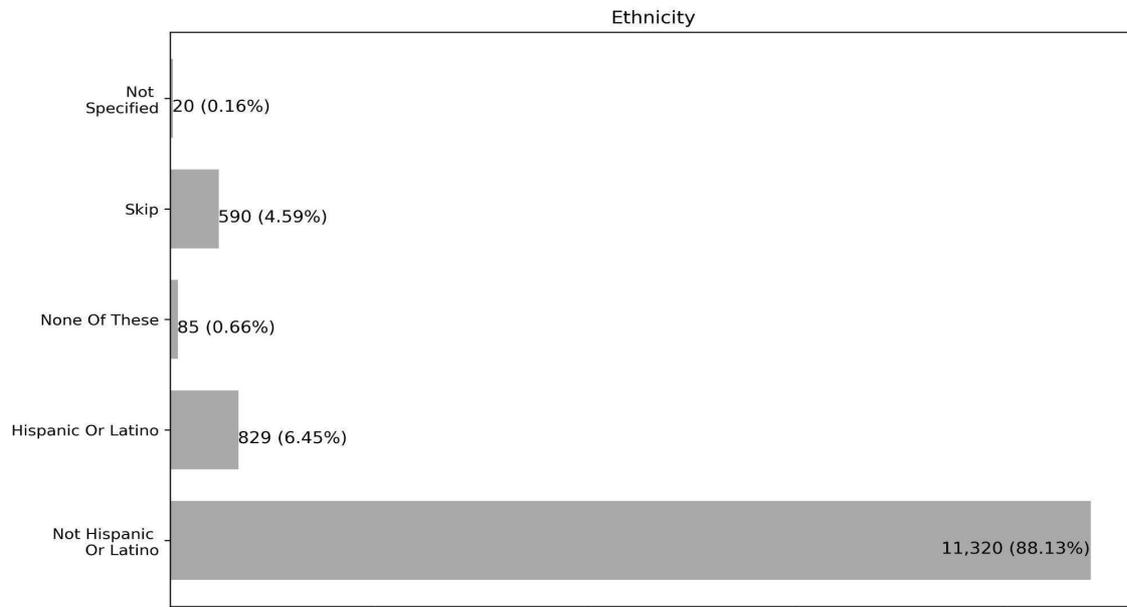
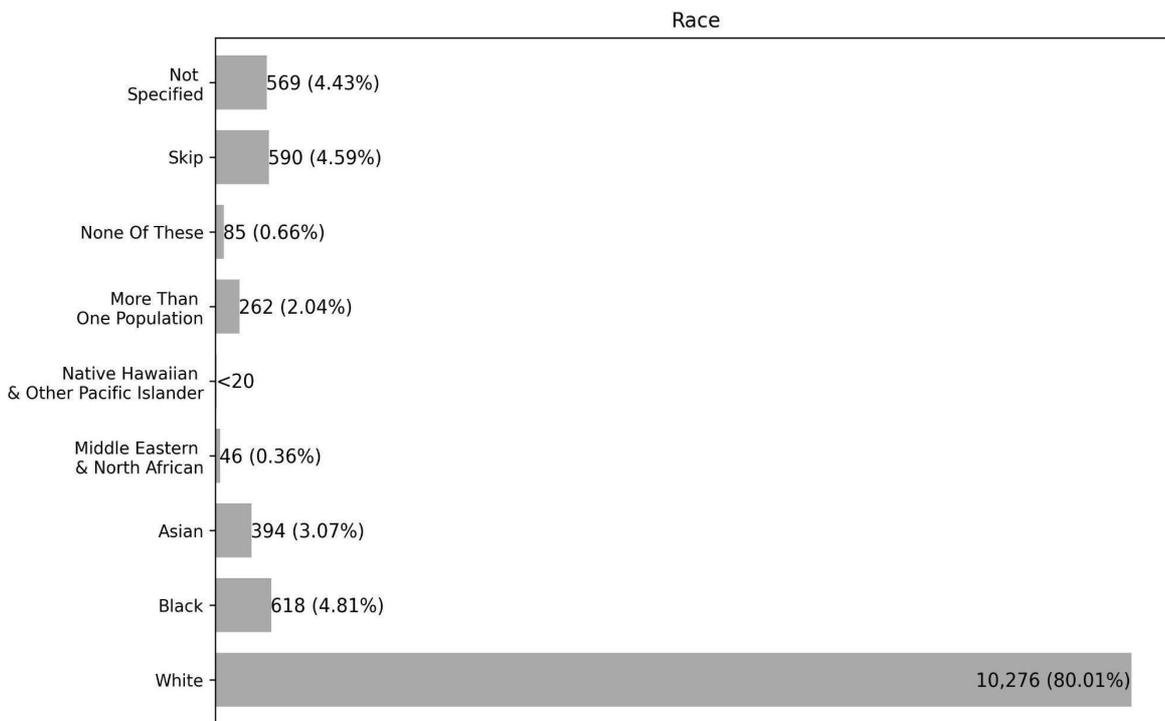


Fig. 3. State-wise distribution of participants who provided Fitbit data in the *All of Us* Research Program (N Fitbit Pid = 12,844).

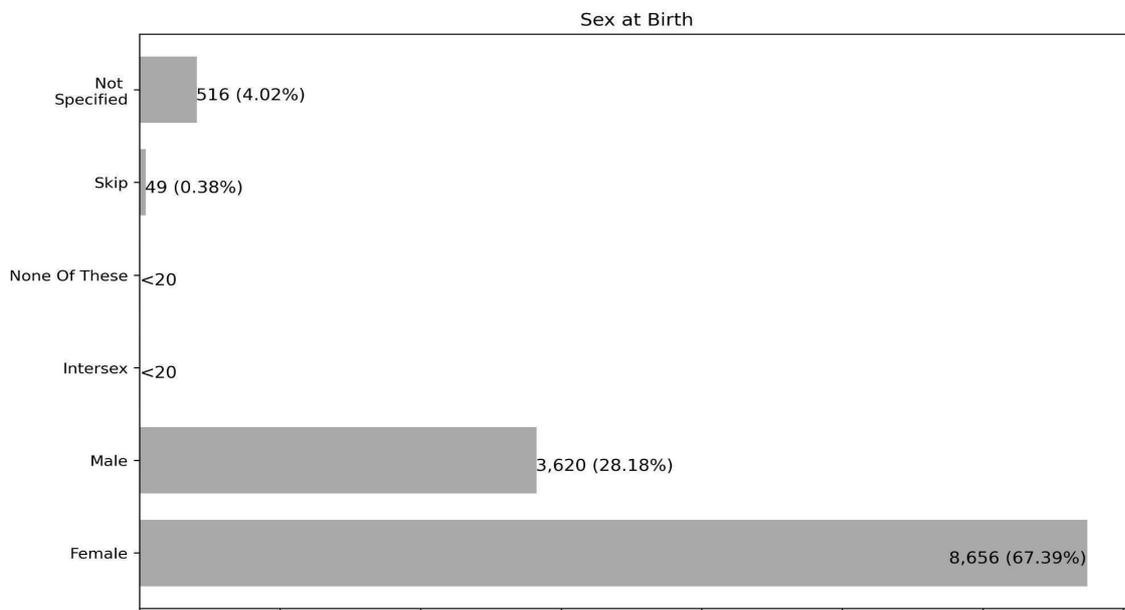
a



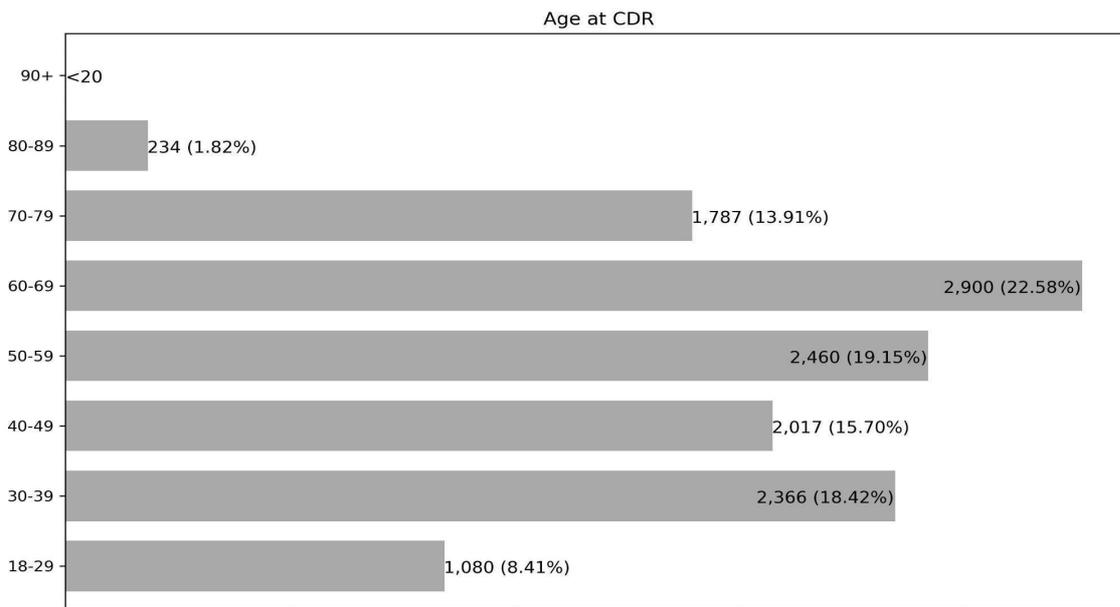
b



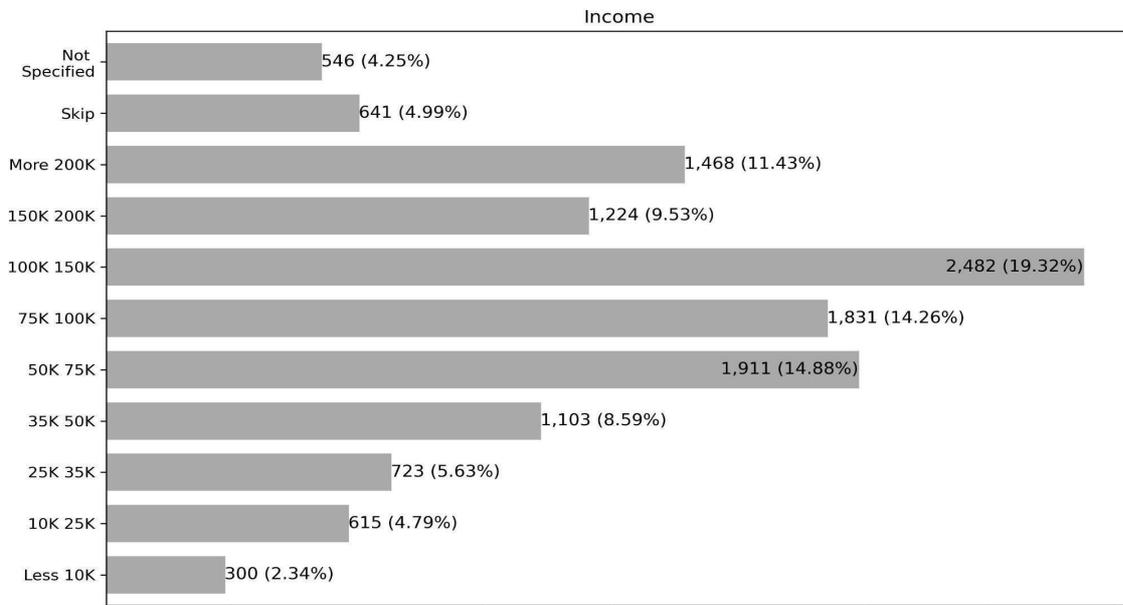
c



d



e



f

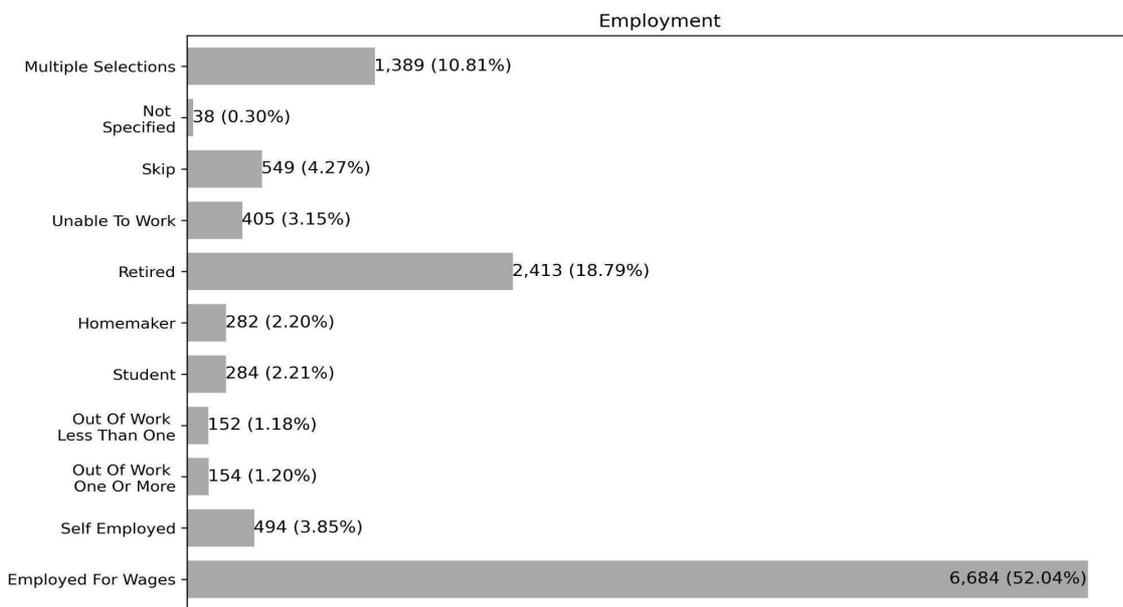


Fig. 4. Self-reported a) ethnicity, b) race, c) sex at birth, d) age e) income, and f) employment of participants with Fitbit data in June 2022 curated data repository, which can be accessed by registered users on Researcher Workbench (N Fitbit Pid = 12,844).

6. Acknowledgments

The *All of Us* Research Program would not be possible without the partnership of its participants. The *All of Us* Research Program is supported by the National Institutes of Health, Office of the Director: Regional Medical Centers: 1 OT2 OD026549; 1 OT2 OD026554; 1 OT2 OD026557; 1 OT2 OD026556; 1 OT2 OD026550; 1 OT2 OD 026552; 1 OT2 OD026553; 1 OT2 OD026548; 1 OT2 OD026551; 1 OT2 OD026555; IAA: AOD21037, AOD22003, AOD16037, AOD21041; Federally Qualified Health Centers: HHSN 263201600085U; Data and Research Center: 5 U2C OD023196; Biobank: 1 U24 OD023121; The Participant Center: U24 OD023176; Participant Technology Systems Center: 1 U24 OD023163; Communications and Engagement: 3 OT2 OD023205; 3 OT2 OD023206; and Community Partners: 1 OT2 OD025277; 3 OT2 OD025315; 1 OT2 OD025337; 1 OT2 OD025276.

References

1. *All of Us* Research Program Investigators, Denny JC, Rutter JL, et al. The "All of Us" Research Program. *N Engl J Med.* 2019;381(7):668-676.
2. Hudson K, Lifton R, Patrick-Lake B, and the Precision Medicine Initiative Working Group. The Precision Medicine Initiative Cohort Program: Building a Research Foundation for 21st Century Medicine. Working Group Report to the Advisory Committee to the Director, NIH. 2015. Available from: <https://www.nih.gov/sites/default/files/research-training/initiatives/pmi/pmi-working-group-report-20150917-2.pdf>
3. Holko, M., Litwin, T.R., Munoz, F. et al. Wearable fitness tracker use in federally qualified health center patients: strategies to improve the health of all of us using digital health devices. *npj Digit. Med.* 2022;5, 53.
4. *All of Us* Research Program News and Events [Internet]. *All of Us* Research Program Expands Data Collection Efforts with Fitbit. 2019. Available from: <https://allofus.nih.gov/news-events-and-media/announcements/all-us-research-program-expands-data-collection-efforts-fitbit>
5. Master, H., Labrecque, S., Kouame, A., et al [Internet]. 2022. 2022Q2R2 v6 Data Characterization Report: Overall *All of Us* Cohort Demographics. All of Us Research Program. Available from <https://aousupporthelp.zendesk.com/hc/en-us/articles/7906441956244-2022Q2R2-v6-Data-Characterization-Report-Overall-All-of-Us-Cohort-Demographics>
6. Scripps Research [Internet]. Through '*All of Us*' program, Scripps Research launches wearable technology study to accelerate precision medicine. 2021. Available from: <https://www.scripps.edu/news-and-events/press-room/2021/20210224-aou-fitbit-study.html>