PLOS ONE

RESEARCH ARTICLE

Diversity and inclusion for the *All of Us* research program: A scoping review

Brandy M. Mapes [ID][1]*, **Christopher S. Foster**[2], **Sheila V. Kusnoor**[3], **Marcia I. Epelbaum**[3], **Mona AuYoung**[4], **Gwynne Jenkins**[2], **Maria Lopez-Class**[2], **Dara Richardson-Heron**[2], **Ahmed Elmi**[2], **Karl Surkan**[5], **Robert M. Cronin** [ID][6], **Consuelo H. Wilkins**[7], **Eliseo J. Pérez-Stable**[8], **Eric Dishman**[2], **Joshua C. Denny**[9], **Joni L. Rutter** [ID][10]*, **the *All of Us* Research Program**[¶]

1 Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America, 2 Office of the Director, National Institutes of Health, Bethesda, Maryland, United States of America, 3 Center for Knowledge Management, Vanderbilt University Medical Center, Nashville, TN, United States of American, 4 Scripps Whittier Diabetes Institute, Scripps Health, San Diego, California, United States of American, 5 Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America and *All of Us* Research Program Participant Representative, 6 Department of Biomedical Informatics, Medicine, and Pediatrics, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America, 7 Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America, 8 National Institute on Minority Health and Health Disparities, Bethesda, Maryland, United States of America, 9 Departments of Biomedical Informatics and Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, United States of America, 10 National Center for Advancing Translational Sciences, National Institutes of Health, Bethesda, Maryland, United States of America

¶ Membership of the the All of Us Research Program is provided in the Acknowledgments.
* brandy.mapes@vumc.org (BMM); joni.rutter@nih.gov (JLR)

OPEN ACCESS

Abstract

The *All of Us* Research Program (*All of Us*) is a national effort to accelerate health research by exploring the relationship between lifestyle, environment, and genetics. It is set to become one of the largest research efforts in U.S. history, aiming to build a national resource of data from at least one million participants. *All of Us* aims to address the need for more diversity in research and set the stage for that diversity to be leveraged in precision medicine research to come. This paper describes how the program assessed demographic characteristics of participants who have enrolled in other U.S. biomedical research cohorts to better understand which groups are traditionally represented or underrepresented in biomedical research. We 1) reviewed the enrollment characteristics of national cohort studies like *All of Us*, and 2) surveyed the literature, focusing on key diversity categories essential to the program's enrollment aims. Based on these efforts, *All of Us* emphasizes enrollment of racial and ethnic minorities, and has formally designated the following additional groups as historically underrepresented: individuals—with inadequate access to medical care; under the age of 18 or over 65; with an annual household income at or below 200% of the federal poverty level; who have a cognitive or physical disability; have less than a high school education or equivalent; are intersex; identify as a sexual or gender minority; or live in rural or non-metropolitan areas. Research accounting for wider demographic variability is critical. Only by ensuring diversity and by addressing the very barriers that limit it, can we position *All of Us* to better understand and tackle health disparities.

Introduction

The *All of Us* Research Program (*All of Us*), funded by the National Institutes of Health (NIH), is a national effort to accelerate health research by exploring the relationship between lifestyle, environment, and genetics [1]. The program is set to become one of the largest health care research efforts in U.S. history, aiming to build a longitudinal resource of multiple data types and biosamples from at least one million participants.
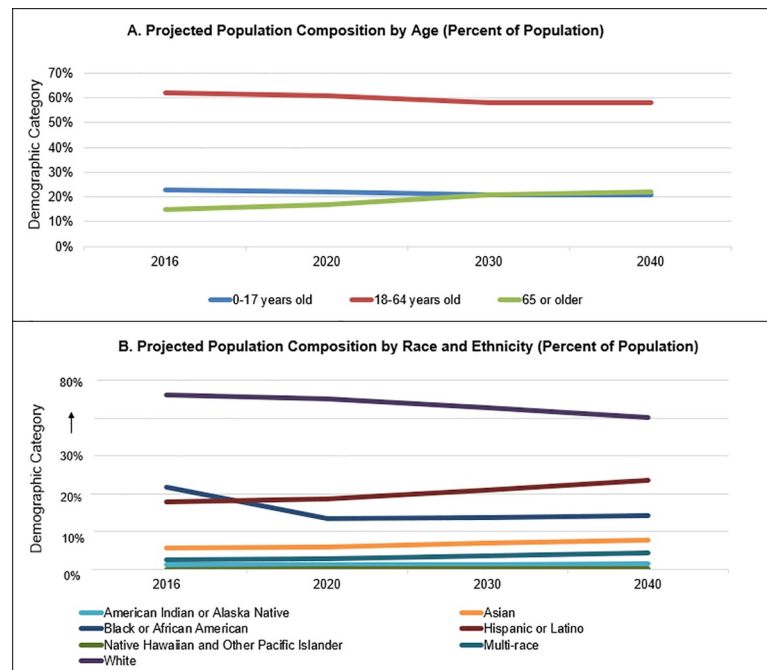
A main goal of the NIH Precision Medicine Initiative (PMI) Working Group [2] was that *All of Us* "should broadly reflect the diversity of the U.S." As stated in the 21st Century Cures Act, *All of Us* will "ensure inclusion of a broad range of participants, including consideration of biological, social, and other determinants of health that contribute to health disparities." [3] The recognition that not all groups have benefited equally from research is a major step towards health equity. A key contribution that *All of Us* can make is setting the stage for more diverse participation in precision medicine research to come. We describe the efforts made to date to guarantee that *All of Us* reflects the broad diversity of the U.S. and enables precision medicine research for all.

To ensure that *All of Us* is rigorous in its efforts to enroll a diverse group of one million participants, we assessed the demographic characteristics of other U.S. biomedical research cohorts. To better understand which groups are represented or underrepresented in biomedical research, we 1) conducted a broad survey of the literature focused on ten demographic categories essential to the program's diversity aims, and 2) reviewed the enrollment characteristics of national cohort studies similar to *All of Us*. By identifying these populations, we can understand the barriers and gaps in enrollment of key underrepresented groups. Based on previous evaluations of diversity and inclusion by race and ethnicity, non-White groups have already been established as underrepresented. [4–6] It is important to note here a distinction between minority health and underrepresentation of other groups, especially with regards to health disparities research and biomedical research representation. Minority health refers to "health characteristics and attributes of racial and/or ethnic minority groups (defined by Office of Management and Budget), who are socially disadvantaged due in part by being subject to potential discriminatory acts." [7] Historically, women have been underrepresented as well, though their proportion now exceeds 50% in NIH-funded research. [8] In addition to racial and ethnic diversity in biomedical research, this work proposes that researchers examine other demographic and social determinants when considering inclusion of diverse participant populations. This review established a landscape of demographic characteristics and social determinants meant to guide *All of Us* towards measuring diversity among its participants.

A Guide for diversity inclusion

Defining underrepresentation for *All of Us* and stakeholder input

All of Us has relied on designated definitions of diversity guided by the leading authorities on health disparities. In addition, we have considered the inclusion of groups that may be underrepresented in biomedical research, specifically, through consultation with a variety of resources and stakeholders, as follows. The National Institute on Minority Health and Health Disparities (NIMHD), congressionally mandated to designate disparity groups, [9] has designated the following groups as most affected by health disparities in the U.S.: Blacks/African Americans, Hispanics/Latinos, American Indians/Alaska Natives, Asian Americans, Native Hawaiians and other Pacific Islanders, socioeconomically disadvantaged populations, underserved rural populations, and sexual and gender minorities. We also used the guidelines for including women and minorities in clinical research set forth by the NIH [10], as well as a

A. Projected Population Composition by Age (Percent of Population)

B. Projected Population Composition by Race and Ethnicity (Percent of Population)

Fig 1. (A-B). Projected Age, Race, & Ethnicity, and Sex Composition of the United States through 2040.

review of the U.S. Census data (Fig 1) [11]. We sought input from NIH leadership and associated stakeholders, including the NIMHD; the Office of Research on Women's Health; the Sexual and Gender Minority Research Office; the National Human Genome Research Institute; the National Institute of Child Health and Human Development; the National Institute on Aging; the Office of Extramural Research; and the *All of Us* Research Program. Lastly, input was obtained from the *All of Us* Participant Provided Information Committee (PPIC), an *All of Us* governance body who developed the program's baseline demographic survey that uses existing measures from other national surveys. [1] A diverse team of PPIC experts, including *All of Us* participant representatives, and *All of Us* participant engagement teams vetted these measures to ensure appropriate terminology and phrasing and ensured consistency in terminology for guiding the diversity categories developed here, some of which are being implemented as of this writing. We sought to examine if these major groups have historically been underrepresented in the biomedical research literature, including clinical trial studies, and whether these categories would help measure the impact of the All of Us Research Program for these groups.

Based on these major inputs, *All of Us* prioritized nine essential diversity categories to complement racial and ethnic minority inclusion. An important point to note is that, aside from racial/ethnic minorities and sex assigned at birth, the other categories have not been broken out demographically in most other studies. Therefore, it is not explicitly known whether they are truly underrepresented in biomedical research. Explicitly defining and evaluating all these demographic characteristics will be important for understanding their impact on social determinants of health. These include (listed alphabetically): access to care, age, annual household income, disability, educational attainment, gender identity, geography (i.e. rurality), sex assigned at birth, and sexual orientation. We anticipate that eventually 75% of the *All of Us* cohort will be from historically underrepresented groups in biomedical research and expect nearly half of this diversity to be from racial or ethnic minorities. [1] Of the nine categories, six

are included in the NIMHD definitions of "disparity populations," three are related to socio-economic status, two to sexual and gender minorities, and one to rurality. For each category, we verified a designation of groups as either historically represented or underrepresented through a review of the literature.

Source selection

We established a team with pertinent expertise to plan and execute the review. The team included representatives from the NIH and participating *All of Us* institutions with expertise in library and information science, enrollment monitoring and reporting, underserved populations, health disparities research, and strategic planning for large-scale national consortia.

Searches of English language articles were conducted using PubMed and Google Scholar along with a mixture of Medical Subject Headings (MeSH) terms and keywords relevant to the nine diversity categories (S1 Table for search strategies and updated retrieval counts). Excluded from the results were duplicate articles, those not relevant to human subjects research, and those that did not contain data to either support or refute justification for considering designation of representation. Meta-analyses and systematic review articles published since 2012, specific to biomedical cohort studies conducted in the United States, and emphasizing results or conclusions specific to the categories of interest were prioritized for review by a subgroup of the authors. We also used handsearching and consulted the subject matter experts on our team to identify additional potentially relevant white papers and unpublished reports. Examples of ways evidence about representation in research was assessed included: evaluation of data from existing U.S. biomedical research studies and trials, evidence depicting how study eligibility criteria contributed to underrepresentation of these groups, and evidence outlining barriers to participation unique to these groups. Results were refined to include a small set of distinctive articles that emphasized findings specific to underrepresentation of the key groups (i.e., why they were considered underrepresented, whether the groups were included in study recruitment, data analysis, or conclusions). The literature was reviewed over the course of several months in 2017 through 2018 (S2 Table for final counts of screened and selected articles, S2 File for selected bibliography).

Complementary assessment of core diversity characteristics for similar cohort studies

Additionally, a selection of other national cohort studies was evaluated for demographic characteristics benchmarking a diversity baseline for *All of Us*. We included cohort studies with an enrollment of >75,000 and a study design or scope like *All of Us*. Those selected were U.S. based except for one international study, the U.K. Biobank. [12] Information was collected using publicly available sources including websites, datasets, and publications. When demographic information was not readily searchable, we made attempts to gather the information from study authors.

Of the national cohort studies evaluated, we identified fifteen as most relevant to informing *All of Us* based on the aims of the research and mixture of data types collected from participants. Five (33%) were still enrolling participants at the time of our review, and seven (47%) had begun enrollment in 2006 or later. We confirmed that thirteen (87%) collected questionnaire data, collected biospecimens, and linked to external data sources. Eight (53%) collected electronic health records. We also prioritized review of diversity categories including age, educational attainment, household income, U.S. geographic region, disability, race and ethnicity, sex assigned at birth, and sexual and gender minority status. See Fig 2 for a brief demographic snapshot of the cohorts we evaluated.

	BioVU	California Teachers Study	Cancer Prevention Studies Baseline II	Cancer Prevention Study III	Kaiser Permanente Research Program on Genes, Environment and Health	Kaiser Permanente Southern California Children's Health Study	The Millennium Cohort Study	Million Veteran Program Active Duty Cohort	Multiethnic Cohort Study of Diet and Cancer	NIH-AARP Diet and Health Study	Nurses' Health Study I	Nurses' Health Study II	Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial	UK BioBank	Women's Health Initiative
Enrollment Dates	ongoing	1995-1996	ongoing	2006-2013	2008- ongoing	2007-2009	ongoing	ongoing	1993-1997	1995-1996	1976	1989	1993-2001	2006-2010	1994-2015
Total	249,208	133,479	1,185,296	303,682	270,570	920,034	77,047	397,104	215,251	545,784	121,700	116,430	154,898	503,944	93,676
Demographics Characteristics															
Age															
0-17	8%	*	*	*	*	*	*	*	*	0%	*	*	*	0%	*
18-64	58%	<66%	*	100.0%	*	*	*	*	*	63%	100%	100%	64%	100%	*
65+	34%	<17%	*	0.0%	*	*	*	<32%	*	37%	*	*	36%	0%	*
Education Attainment															
< High School/GED	*	*	15%	1%	*	*	6%	*	*	6%	*	*	*	30%	5%
> High School/GED	*	*	85%	97%	*	*	94%	*	*	91%	100%	100%	*	70%	95%
Race/Ethnicity															
White	80%	87%	93%	83%	*	21%	69%	77%	23%	93%	97%	92%	86%	94%	83%
African American	10%	3%	4%	5%	*	8%	14%	14%	16%	4%	*	*	5%	1%	8%
Asian or Pacific Islander	2%	3%	>1%	*	*	7%	8%	*	33%	1%	*	*	*	1%	3%
Hispanic/Latino	3%	4%	>1%	8%	*	51%	6%	*	22%	2%	*	*	*	0%	4%
Sex															
Male	43%	*	43%	23%	*	50%	73%	92%	30%	59%	*	*	49.5%	45%	*
Female	57%	100%	57%	77%	*	50%	27%	8%	70%	41%	100%	100%	50.5%	54%	100%

Data presented were collected September - November 2017; data may not reflect current demographics for those studies still actively enrolling. * indicates instances in which information is unknown, not applicable, not collected, not reported, or not part of the study eligibility criteria. Some totals do not equal 100% because of instances in which demographic data is missing and/or not reported.

Fig 2. Core demographic characteristics of similar cohorts evaluated to support development of the *All of Us* guide for diversity and inclusion.

Results

Roughly 645 articles were screened throughout the literature review and 170 were selected as significant for informing the *All of Us* Guide for Diversity and Inclusion (S2 Table for exact counts by search category). Based on the body of literature and consultation with stakeholders, *All of Us* emphasizes inclusion of racial and ethnic minorities and has prioritized the following additional groups as being possibly underrepresented in biomedical research: individuals with inadequate access to medical care; <18 or >65 years; with an annual household income at or below 200% of the federal poverty level (FPL); who have a cognitive or physical disability; or have less than a high school education or equivalent. Also included are those individuals who: are intersex; identify as a sexual or gender minority; or live in rural or non-metropolitan areas. Including the racial and ethnic groups, Table 1 shows historically represented and underrepresented characteristics of each category and how they are measured by the survey responses. [13] Researchers interested in studying specific subsets of these groups (for example, individuals < 3 or >80), are able to do so based on the way in which these demographic data have been collected.

These designations are supported by assessing the other national cohort studies. Historically underrepresented groups, where we were able to obtain data, were underrepresented across all fifteen studies evaluated. Annual household income was unknown, not collected, or not reported for thirteen (87%). Of those reporting income, one (7%) is U.S. based and only a small portion of the participants had an annual household income under $25,000. [15] Educational attainment data were obtained for seven (47%) cohort studies. Percentage of participants with less than a high school degree or GED ranged from 1% to 15% for those U.S. based studies. [15–18] We confirmed inclusion of older adults in only five (33%) studies and their representation ranged from <17% to 37%. [16,19–22] Participants under 18 were enrolled in two (13%) studies. [22,23] Only one (7%) study reported on representation of participants with a disability; these participants made up 18% of that total sample. [17]

Fourteen (93%) studies reported demographic representation of participants by sex. Males were not eligible for four of the studies. Ten (67%) were eligible to both males and females and reported on sex. Of those ten, females were overrepresented in five (50%). [12,18,21,22,24] None of the fifteen studies reported intersex. Five (33%) reported participants' geographic region. Those mostly recruited from only one U.S. geographic region though two (13%) reported geographic dispersity stretching across at least four regions. [15,25] Data to

Table 1. The *All of Us* Guide for diversity and inclusion.

Diversity Category	Represented in Biomedical Research	Underrepresented in Biomedical Research
Race and Ethnicity[a]	Individuals who identify as White and non-Hispanic	Individuals who identify as other than White and non-Hispanic (i.e. Asian; Black, African or African American; Hispanic, Spanish, or Latino; Native Hawaiian or Pacific Islander; Middle Eastern or North African)
Access to Care	Individuals who have had a needed medical visit in the past 12 months or can obtain and pay for medical care as needed	Individuals who have not had a needed medical visit in the past 12 months or cannot easily obtain or pay for medical care as needed
Age[b]	Adults ages 18–64	Children 17 or younger and adults 65 or older
Annual Household Income	Individuals with household incomes above 200% of the Federal Poverty Level	Individuals with household incomes equal to or below 200% of the Federal Poverty Level
Disability	Individuals without a physical or cognitive disability	Individuals with either a physical or cognitive disability
Educational Attainment	Individuals with a high school degree or equivalent	Individuals with less than a high school degree or equivalent
Gender Identity[a]	Individuals who identify as either a man or a woman	Individuals who identify as gender variant, non-binary, transgender, or something else
Geography	Individuals who reside in urban metropolitan areas	Individuals who reside in rural and non-metropolitan areas
Sex Assigned at Birth[a]	Male or female individuals	Individuals who are neither male nor female (i.e. intersex)
Sexual Orientation	Individuals who identify as straight	Individuals who identify as asexual, bisexual, gay or lesbian or something else

[a] Racial and ethnic minorities and women: The NIH Revitalization Act of 1993 set forth policy that mandates the inclusion of women and minorities in NIH-funded clinical research unless their exclusion is justifiable and approved by the relevant NIH institute. The resulting policy guided the enrollment criteria developed by *All of Us*. For the purpose of this work, however, women are not considered UBR. Representation of women in NIH-funded research has steadily risen to over 50%. [4]

[b] Adults aged 65 or older: NIH Policy and Guidelines on the Inclusion of Individuals Across the Lifespan as Participants in Research Involving Human Subjects defines older adults as individuals 65 years of age or older and underrepresentation of this entire group has been confirmed by the literature. Adults 75 or older, however, may be especially underrepresented; there is additional federal guidance on ensuring their inclusion. [14]

specifically assess rurality were not readily available. Only two (13%) studies reported overrepresentation of racial and ethnic minorities compared to White participants. [23,24]

Lastly, most did not draw an intentional distinction between sex assigned at birth and gender. Data to permit evaluation of gender minority representation, however, were not collected from participants or not publicly reported for any of the fifteen studies. This was the case for sexual orientation data as well. See Fig 2 for aggregate findings from the cohort assessment and Table 1 for *All of Us* Guide for Diversity and Inclusion.

Discussion

Most of the articles sought to illustrate major health care and research barriers impacting racial and ethnic populations, along with other underrepresented groups. The literature demonstrated that racial/ethnic minorities and other groups are typically impacted by at least one of several barriers including financial (e.g., insurance coverage, income), cognitive (e.g., health literacy, technical proficiency), language (e.g., lack of translation resources), communication and cultural factors (e.g., cultural humility, lack of trust), structural (e.g., transportation, geographic accessibility), racial and ethnic discrimination, or study design (e.g., eligibility criteria limiting participation or inclusion, failure to reach diverse groups due to recruitment limitations). [7,26–34] These barriers have significant effects on multiple groups. Access to research opportunities is critically impacted by structural barriers like transportation, but also by language, communication, and unnecessary restrictions in study design. For example, willingness to participate in research does not significantly differ between individuals who reside in rural

areas and those who live in urban locations. Rather, the limiting factor is accessibility impacted by either structural barriers or recruitment limitations. [31,34,35] Older adults are also impacted by these barriers. The literature demonstrated that older adults have been marginalized in research due to a number of reasons including not meeting eligibility criteria (e.g., clinical trial restrictions on comorbid conditions due to risk of side effects), perceived burden on study staff (e.g., longer screening process or more "high touch" retention tactics needed), and cognitive or physical impairments that affect the ability to consent or fully participate. [36–38] Similar barriers are faced by racial or ethnic minorities. [39–41] Cultural barriers, specifically lack of trust, and not being trustworthy, are defining barriers across racial or ethnic minorities, particularly Black and African American individuals, [34,42] and also across many other groups including rural populations and sexual or gender minorities. Specifically in cancer-related clinical trials, studies have shown no racial or ethnic differences in participation rates [43], but this does not seem to be generally applicable to other types of clinical research settings. [44] Importantly, overcoming a lack of trust is through engagement and carefully designed and strategies [45] by which participants are invited to participate, in a respectful way, in research programs.

Representation of underrepresented groups is significantly affected by the lack of consistent demographic data collection and gaps in research domains specific to the health disparities that impact them. For example, unlike sex assigned at birth, which is a biological construct, gender is a construct that includes many dimensions. Domain experts recommend researchers include measures to capture both an individual's assigned sex at birth and gender identity to fully represent one's physical and internal sense of self. [46] The literature confirmed that gender identity as a social determinant of health is not only understudied but also unrecognized. [47,48] This body of literature is comprised of articles on evaluation of instruments that gather this information (e.g., questionnaires, electronic medical records, etc.), their limited use within healthcare and research [49,50] and the lack of knowledge and sensitivity of medical personnel about different gender identities. [51] Similarly, sexual orientation measures are not often collected from patients and research participants, especially noted for older adults. [52,53] These omissions were observed in our own review of national cohorts. The literature also noted that biomedical research has generally neglected exploration of sexual orientation, apart from research focused on understanding sexually transmitted disease. This implies a substantial gap in research that could lead to improvements in broader health outcomes for sexual and gender minorities. [54,55] The lack of data collection measures and demographic reporting does not allow us to confirm whether individuals that identify as sexual or gender minorities are truly underrepresented. Data collection is a mandatory first step, however, in addressing the obvious gaps in research results that directly impact health disparities faced by this population. For this reason, sexual and gender minorities are included as an underrepresented group.

We identified 12 articles that examined sex as a biological factor in biomedical research. Several of the articles call for more comparison of research findings by sex, noting that, like race and ethnicity, sex differences in many domains are still widely understudied. [56–60] Regardless of actual representation of females in biomedical research, many health outcomes specific to them will remain unknown unless researchers prioritize exploring differences across sex assigned at birth, race, and ethnicity. In addition, sex is predominantly perceived as a binary biological construct and individuals born as intersex, or those with disorders of sexual development, have difficulties "fitting" into a male/female binary category. Including intersex as a biological construct is important for health care since individuals' needs may not apply to either the male or female classification. [59] The literature indicates that this is also true for transgender individuals who have undergone gender affirming surgeries or who receive hormonal treatments as well. [61]

We identified literature that explained underrepresentation of select minority populations due to issues in how race and ethnicity data are traditionally collected and reported. [62,63] For example, there has been confusion among survey respondents over the categorization of Hispanic or Latino as an ethnicity rather than a race. [63] A lack of sufficient categories for race and ethnicity data, often in combination with underrepresentation within the research samples, means results are often reported in aggregated categories that can mask underlying disparities. For example, Middle Eastern and North African populations have often been classified as White in research questionnaires that do not specifically call out a category for Middle Eastern and North African, similarly Asian American, Native Hawaiian, and other Pacific Islanders have historically been grouped as Asian or Other. [62,63]

Limitations of this approach

The broad scope and depth of this review was limited since complex strategies were needed to canvass the literature, returning exceptionally large retrieval sizes. This work is meant as a survey of trends, relying mainly on reviews obtained from the literature to best inform the state of the science on groups historically underrepresented in biomedical research. Our evaluation of other large national cohort studies was also limited because we conducted most of this work using publicly available sources. In some instances, we were unable to confirm the accuracy or currency of the information. In addition, *All of Us* is a hypothesis and disease-agnostic research program with minimal inclusion/exclusion criteria. Most of the studies we identified in our evaluation differ from *All of Us* in both areas. For example, clinical trials-enrolled participants with specific medical conditions or those that may benefit from a therapeutic treatment. They may, by their very nature, exclude certain demographic groups. Finally, many studies do not collect comprehensive socio-demographic data, therefore, we do not know whether their populations include groups we assert to be underrepresented.

Commentary

There is a rich diversity within the U.S. that must be leveraged for understanding factors contributing to health and disease for the entire population. Historically, biomedical research has neglected this diversity to our detriment. A recent report [64] showed that, of the drugs removed from the U.S. market between 1997 and 2000, 80% were removed due to side effects or fatalities that occurred in women. Another report [65] found that racial factors, although important, do not solely account for the disparities in health care. Societal and environmental factors are equally important to account for context. [66] A widely-used antiplatelet drug, clopidogrel, needs to be converted to an active form to be effective. [67] However, due to a genetic variation that affects the body's ability to metabolize it, only about 50% of individuals of Asian ancestry are able to benefit, along with other individuals who carry the variant. This is one example of a modern scientific advancement, in this case a life-saving medication, offering a high benefit for some individuals within populations, but significant risks for others. [68]

An early research opportunity for *All of Us* will be examining health disparities among these racial and ethnic populations intersected across other demographic groups and their impact on social determinants of health. For a summary of significant conclusions, see Table 2. The *All of Us* Research Program Guide for Diversity and Inclusion (Table 1) is an essential product for *All of Us*, serving as the program foundation for 1. future demographic enrollment monitoring, 2. gauging the success of new recruitment strategies as they are implemented and providing opportunity to revise when key enrollment gaps are identified, and 3. gauging the variability and strength of the resulting research data repository, and ensuring that it will support health disparity research for all historically underrepresented groups as it grows

Table 2. Key points learned in developing the *All of Us* guide for diversity and inclusion.

• The rich, cultural, geographic, and demographic diversity of the U.S. is underrepresented in U.S. biomedical research.

• Significant barriers impact representation of key populations. The biomedical research community must acknowledge and address them to ensure inclusivity.

• A lack of data collection for many key demographic characteristics and social determinants is a critical gap across the entirety of biomedical research. For example, educational attainment and many other socio-economic status factors have a robust relationship with overall mortality. Yet, these measures are almost completely absent from the research landscape. *All of Us*, and the biomedical research community at large, must address this.

• Sexual and gender minority measures, except for research exploring sexually transmitted disease, have mostly been ignored. Although a lack of measures does not necessarily suggested underrepresentation, disparities impacting these groups are understudied. To understand representation of these groups, *All of Us* and the biomedical research community must start asking about sexual orientation and gender identity.

• A guide for inclusion is necessary to ensure equity in research outcomes. The guide, however, must be viewed as a starting point, not a final product in understanding and addressing barriers that impact participation, can we position *All of Us* to better understand and tackle health disparities for all.

https://doi.org/10.1371/journal.pone.0234962.t002

over time. This guide is a novel attempt to emphasize and define diversity and inclusion for *All of Us* based on a series of biologically and socially significant constructs. There are considerable nuances that impact diversity and inclusion monitoring and we respect that they should not be treated independently of one another or considered universally equal. For example, there is an enormous body of literature confirming the importance of racial and ethnic diversity in biomedical research. The *All of Us* guide emphasizes minority representation in biomedical research, and the additional demographic characteristics and social determinants highlighted are meant to complement, not dilute, that inclusion.

Conclusion

Research that accounts for wider demographic variability is critical. It is time that we move beyond relying on research generated from largely homogeneous populations of usually White, male, urban, and higher socio-economic status. [69] Only by incorporating more diverse factors as key variables, ensuring inclusion, and identifying and addressing barriers that limit research participation, can we position *All of Us* to better understand and adequately tackle health disparities.

Research is used to provide the basis by which we discover new pathways and identify new treatment approaches. As such, only by narrowing the gaps of health disparities and increasing our understanding of how social determinants of health impact health outcomes can we realize a future of research that is more precise. We note that discrimination is a standing issue in workforce diversity, [70] which impacts minority participation in research studies like ours. Based on the work presented, and in conjunction with community partners, *All of Us* has developed resources for engaging diverse communities (https://www.joinallofus.org/en/community/community-resources). These provide a springboard for researchers to more readily incorporate diversity and inclusion strategies to local settings [45,71] or disease specific contexts, or to use the https://www.researchallofus.org/ platform to build diverse cohorts. Within *All of Us*, the attention paid to these important factors can help surface the intersectionality of these barriers to research participation and help the research community to be better equipped to engage diverse populations.

The approach taken by *All of Us* was broad, methodical, and intentional in conceptualizing an inclusive guide. Despite the limitations, it encompasses a depth of inclusion that is unprecedented in other national cohorts. The emphasis on historically underrepresented minorities and other populations provides a reference for any future study that aims to ensure diversity

and equity. The *All of Us* Research Program is well on its way to creating a national resource that reflects and supports the broad diversity of the U.S. necessary for advancing precision medicine for all.

Supporting information

S1 Table. Search strategies by diversity category and January 2020 retrieval counts.
(DOCX)

S2 Table. Number of articles screened and selected in original literature search.
(DOCX)

S1 File. Bibliography of selected references by search category.
(DOCX)

S2 File. Bibliography of selected references by search category.
(DOCX)

Acknowledgments

Author Contributions

Conceptualization: Brandy M. Mapes, Christopher S. Foster, Mona AuYoung, Karl Surkan, Robert M. Cronin, Consuelo H. Wilkins, Eliseo J. Pérez-Stable, Joshua C. Denny, Joni L. Rutter.

Formal analysis: Christopher S. Foster, Sheila V. Kusnoor, Marcia I. Epelbaum, Mona AuYoung.

Methodology: Brandy M. Mapes, Christopher S. Foster, Sheila V. Kusnoor, Marcia I. Epelbaum, Mona AuYoung, Robert M. Cronin, Joshua C. Denny, Joni L. Rutter.

Project administration: Christopher S. Foster, Gwynne Jenkins, Maria Lopez-Class, Dara Richardson-Heron, Ahmed Elmi, Eric Dishman, Joni L. Rutter.

Visualization: Brandy M. Mapes.

Writing – original draft: Brandy M. Mapes, Sheila V. Kusnoor, Marcia I. Epelbaum, Mona AuYoung, Maria Lopez-Class, Joshua C. Denny, Joni L. Rutter.

Writing – review & editing: Brandy M. Mapes, Christopher S. Foster, Sheila V. Kusnoor, Marcia I. Epelbaum, Mona AuYoung, Gwynne Jenkins, Karl Surkan, Robert M. Cronin, Consuelo H. Wilkins, Eliseo J. Pérez-Stable, Joshua C. Denny, Joni L. Rutter.

References

1. The "All of Us" Research Program. New England Journal of Medicine. 2019 Aug 15; 381(7):668–76. https://doi.org/10.1056/NEJMsr1809937 PMID: 31412182

2. The Precision Medicine Initiative Cohort Program–Building a Research Foundation for 21st Century Medicine.: 108.

3. Bonamici S. H.R.34 - 114th Congress (2015–2016): 21st Century Cures Act [Internet]. 2016 [cited 2019 Feb 1]. Available from: https://www.congress.gov/bill/114th-congress/house-bill/34

4. Popejoy AB, Fullerton SM. Genomics is failing on diversity. Nature. 2016 Oct 12; 538(7624):161–4. https://doi.org/10.1038/538161a PMID: 27734877

5. Knepper TC, McLeod HL. When will clinical trials finally reflect diversity? Nature. 2018 May; 557 (7704):157–9. https://doi.org/10.1038/d41586-018-05049-5 PMID: 29743700

6. Fisher ER, Pratt R, Esch R, Kocher M, Wilson K, Lee W, et al. The role of race and ethnicity in views toward and participation in genetic studies and precision medicine research in the United States: A systematic review of qualitative and quantitative studies. Mol Genet Genomic Med. 2019 Dec 23;e1099. https://doi.org/10.1002/mgg3.1099 PMID: 31867882

7. Duran DG, Pérez-Stable EJ. Novel Approaches to Advance Minority Health and Health Disparities Research. Am J Public Health. 2019 Jan 1; 109(S1):S8–10. https://doi.org/10.2105/AJPH.2018.304931 PMID: 30699017

8. Office of Research on Women's Health and NIH Support for Research on Women's Health. Report of the Advisory Committee on Research on Women's Health Fiscal Years 2015–2016 [Internet]. National Institutes of Health; 2017. Available from: https://orwh.od.nih.gov/sites/orwh/files/docs/ORWH_Biennial_Report_WEB_508_FY-15-16.pdf

9. Legislative History [Internet]. NIMHD. [cited 2019 May 8]. Available from: https://www.nimhd.nih.gov/about/legislative-info/legislative-history.html

10. Inclusion of Women and Minorities as Participants in Research Involving Human Subjects—Policy Implementation Page | grants.nih.gov [Internet]. [cited 2019 Jan 25]. Available from: https://grants.nih.gov/grants/funding/women_min/women_min.htm

11. Bureau UC. 2017 National Population Projections Tables [Internet]. [cited 2019 May 8]. Available from: https://www.census.gov/data/tables/2017/demo/popproj/2017-summary-tables.html

12. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018; 562(7726):203–9. https://doi.org/10.1038/s41586-018-0579-z PMID: 30305743

13. Cronin RM, Jerome RN, Mapes B, Andrade R, Johnston R, Ayala J, et al. Development of the Initial Surveys for the All of Us Research Program. Epidemiology. 2019 Apr 29;

14. Research C for DE and. E7 Studies in Support of Special Populations: Geriatrics [Internet]. U.S. Food and Drug Administration. 2019 [cited 2019 Jun 26]. Available from: http://www.fda.gov/regulatory-information/search-fda-guidance-documents/e7-studies-support-special-populations-geriatrics

15. Design of the Women's Health Initiative clinical trial and observational study. The Women's Health Initiative Study Group. Control Clin Trials. 1998 Feb; 19(1):61–109. https://doi.org/10.1016/s0197-2456(97)00078-0 PMID: 9492970

16. Participants' Profile—NIH_AARP Diet and Health Study [Internet]. [cited 2018 Feb 13]. Available from: https://dietandhealth.cancer.gov/profile.html

17. Smith TC, Millennium Cohort Study Team. The US Department of Defense Millennium Cohort Study: career span and beyond longitudinal follow-up. J Occup Environ Med. 2009 Oct; 51(10):1193–201. https://doi.org/10.1097/JOM.0b013e3181b73146 PMID: 19786902

18. History of the Cancer Prevention Studies | American Cancer Society [Internet]. [cited 2018 Feb 13]. Available from: https://www.cancer.org/research/we-conduct-cancer-research/epidemiology/history-cancer-prevention-study.html

19. California Teachers Study [Internet]. California Teachers Study. [cited 2018 Feb 13]. Available from: https://www.calteachersstudy.org

20. Gaziano JM, Concato J, Brophy M, Fiore L, Pyarajan S, Breeling J, et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. Journal of Clinical Epidemiology. 2016 Feb 1; 70:214–23. https://doi.org/10.1016/j.jclinepi.2015.09.016 PMID: 26441289

21. Zhu CS, Pinsky PF, Kramer BS, Prorok PC, Purdue MP, Berg CD, et al. The prostate, lung, colorectal, and ovarian cancer screening trial and its associated research resource. J Natl Cancer Inst. 2013 Nov 20; 105(22):1684–93. https://doi.org/10.1093/jnci/djt281 PMID: 24115361

22. Roden DM, Pulley JM, Basford MA, Bernard GR, Clayton EW, Balser JR, et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. Clin Pharmacol Ther. 2008 Sep; 84 (3):362–9. https://doi.org/10.1038/clpt.2008.89 PMID: 18500243

23. Koebnick C, Coleman KJ, Black MH, Smith N, Der-Sarkissian JK, Jacobsen SJ, et al. Cohort Profile: The KPSC Children's Health Study, a population-based study of 920 000 children and adolescents in southern California. Int J Epidemiol. 2012 Jun 1; 41(3):627–33. https://doi.org/10.1093/ije/dyq252 PMID: 21257603

24. Kolonel LN, Altshuler D, Henderson BE. The multiethnic cohort study: exploring genes, lifestyle and cancer risk. Nat Rev Cancer. 2004 Jul; 4(7):519–27. https://doi.org/10.1038/nrc1389 PMID: 15229477

25. Bao Y, Bertoia ML, Lenart EB, Stampfer MJ, Willett WC, Speizer FE, et al. Origin, Methods, and Evolution of the Three Nurses' Health Studies. Am J Public Health. 2016 Sep; 106(9):1573–81. https://doi.org/10.2105/AJPH.2016.303338 PMID: 27459450

26. Carrillo JE, Carrillo VA, Perez HR, Salas-Lopez D, Natale-Pereira A, Byron AT. Defining and targeting health care access barriers. J Health Care Poor Underserved. 2011 May; 22(2):562–75. https://doi.org/10.1353/hpu.2011.0037 PMID: 21551934

27. Neutens T. Accessibility, equity and health care: review and research directions for transport geographers. Journal of Transport Geography. 2015 Feb; 43:14–27.

28. Campbell DJ, Manns BJ, Hemmelgarn BR, Sanmartin C, King-Shier KM. Development of a conceptual framework for understanding financial barriers to care among patients with cardiovascular-related chronic disease: a protocol for a qualitative (grounded theory) study. CMAJ Open. 2016 Jun; 4(2): E304–8. https://doi.org/10.9778/cmajo.20160030 PMID: 27398378

29. Bonevski B, Randell M, Paul C, Chapman K, Twyman L, Bryant J, et al. Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. BMC Med Res Methodol. 2014 Mar 25; 14:42. https://doi.org/10.1186/1471-2288-14-42 PMID: 24669751

30. George S, Duran N, Norris K. A systematic review of barriers and facilitators to minority research participation among African Americans, Latinos, Asian Americans, and Pacific Islanders. Am J Public Health. 2014 Feb; 104(2):e16–31. https://doi.org/10.2105/AJPH.2013.301706 PMID: 24328648

31. Virani S, Burke L, Remick SC, Abraham J. Barriers to recruitment of rural patients in cancer clinical trials. J Oncol Pract. 2011 May; 7(3):172–7. https://doi.org/10.1200/JOP.2010.000158 PMID: 21886499

32. Woodall A, Morgan C, Sloan C, Howard L. Barriers to participation in mental health research: are there specific gender, ethnicity and age related barriers? BMC Psychiatry. 2010 Dec 2; 10:103. https://doi.org/10.1186/1471-244X-10-103 PMID: 21126334

33. Townsley CA, Selby R, Siu LL. Systematic review of barriers to the recruitment of older patients with cancer onto clinical trials. J Clin Oncol. 2005 May 1; 23(13):3112–24. https://doi.org/10.1200/JCO.2005.00.141 PMID: 15860871

34. Kim SH, Tanner A, Friedman DB, Foster C, Bergeron CD. Barriers to clinical trial participation: a comparison of rural and urban communities in South Carolina. J Community Health. 2014 Jun; 39(3):562–71. https://doi.org/10.1007/s10900-013-9798-2 PMID: 24310703

35. Friedman DB, Foster C, Bergeron CD, Tanner A, Kim SH. A qualitative study of recruitment barriers, motivators, and community-based strategies for increasing clinical trials participation among rural and urban populations. Am J Health Promot. 2015 Jun; 29(5):332–8. https://doi.org/10.4278/ajhp.130514-QUAL-247 PMID: 24670073

36. Cherubini A, Oristrell J, Pla X, Ruggiero C, Ferretti R, Diestre G, et al. The persistent exclusion of older patients from ongoing clinical trials regarding heart failure. Arch Intern Med. 2011 Mar 28; 171(6):550–6. https://doi.org/10.1001/archinternmed.2011.31 PMID: 21444844

37. Zulman DM, Sussman JB, Chen X, Cigolle CT, Blaum CS, Hayward RA. Examining the evidence: a systematic review of the inclusion and analysis of older adults in randomized controlled trials. J Gen Intern Med. 2011 Jul; 26(7):783–90. https://doi.org/10.1007/s11606-010-1629-x PMID: 21286840

38. Tahhan AS, Vaduganathan M, Greene SJ, Fonarow GC, Fiuzat M, Jessup M, et al. Enrollment of Older Patients, Women, and Racial and Ethnic Minorities in Contemporary Heart Failure Clinical Trials: A Systematic Review. JAMA Cardiol. 2018 Oct 1; 3(10):1011–9. https://doi.org/10.1001/jamacardio.2018.2559 PMID: 30140928

39. Yancey AK, Ortega AN, Kumanyika SK. Effective recruitment and retention of minority research participants. Annu Rev Public Health. 2006; 27:1–28. https://doi.org/10.1146/annurev.publhealth.27.021405.102113 PMID: 16533107

40. Wendler D, Kington R, Madans J, Van Wye G, Christ-Schmidt H, Pratt LA, et al. Are racial and ethnic minorities less willing to participate in health research? PLoS Med. 2006 Feb; 3(2):e19. https://doi.org/10.1371/journal.pmed.0030019 PMID: 16318411

41. Bolen S, Tilburt J, Baffi C, Gary TL, Powe N, Howerton M, et al. Defining "success" in recruitment of underrepresented populations to cancer clinical trials: moving toward a more consistent approach. Cancer. 2006 Mar 15; 106(6):1197–204. https://doi.org/10.1002/cncr.21745 PMID: 16453333

42. Hamel LM, Penner LA, Albrecht TL, Heath E, Gwede CK, Eggly S. Barriers to Clinical Trial Enrollment in Racial and Ethnic Minority Patients With Cancer. Cancer Control. 2016 Oct; 23(4):327–37. https://doi.org/10.1177/107327481602300404 PMID: 27842322

43. Langford AT, Resnicow K, Dimond EP, Denicoff AM, Germain DS, McCaskill-Stevens W, et al. Racial/ethnic differences in clinical trial enrollment, refusal rates, ineligibility, and reasons for decline among patients at sites in the National Cancer Institute's Community Cancer Centers Program. Cancer. 2014; 120(6):877–84. https://doi.org/10.1002/cncr.28483 PMID: 24327389

44. Cunningham-Erves J, Villalta-Gil V, Wallston KA, Boyer AP, Wilkins CH. Racial differences in two measures of trust in biomedical research. J Clin Transl Sci. 2019 Jun 17; 3(2–3):113–9. https://doi.org/10.1017/cts.2019.378 PMID: 31660234

45. Wilkins CH. Effective Engagement Requires Trust and Being Trustworthy. Med Care. 2018 Oct; 56(10 Suppl 1):S6–8.

46. Best Practices for Asking Questions to Identify Transgender and Other Gender Minority Respondents on Population-Based Surveys—Williams InstituteWilliams Institute [Internet]. [cited 2018 Jun 20]. Available from: https://williamsinstitute.law.ucla.edu/research/census-lgbt-demographics-studies/geniuss-report-sept-2014/

47. Reisner SL, Conron KJ, Tardiff LA, Jarvi S, Gordon AR, Austin SB. Monitoring the health of transgender and other gender minority populations: validity of natal sex and gender identity survey items in a U.S. national cohort of young adults. BMC Public Health. 2014 Nov 26; 14:1224. https://doi.org/10.1186/1471-2458-14-1224 PMID: 25427573

48. Pega F, Veale JF. The case for the World Health Organization's Commission on Social Determinants of Health to address gender identity. Am J Public Health. 2015 Mar; 105(3):e58–62. https://doi.org/10.2105/AJPH.2014.302373 PMID: 25602894

49. Deutsch MB, Keatley J, Sevelius J, Shade SB. Collection of gender identity data using electronic medical records: survey of current end-user practices. J Assoc Nurses AIDS Care. 2014 Dec; 25(6):657–63. https://doi.org/10.1016/j.jana.2014.04.001 PMID: 24880490

50. Schneider C, Cerwenka S, Nieder TO, Briken P, Cohen-Kettenis PT, De Cuypere G, et al. Measuring Gender Dysphoria: A Multicenter Examination and Comparison of the Utrecht Gender Dysphoria Scale and the Gender Identity/Gender Dysphoria Questionnaire for Adolescents and Adults. Arch Sex Behav. 2016 Apr; 45(3):551–8. https://doi.org/10.1007/s10508-016-0702-x PMID: 26883025

51. Carabez R, Pellegrini M, Mankovitz A, Eliason M, Scott M. Does your organization use gender inclusive forms? Nurses' confusion about trans* terminology. J Clin Nurs. 2015 Nov; 24(21–22):3306–17. https://doi.org/10.1111/jocn.12942 PMID: 26263919

52. Fredriksen-Goldsen KI, Kim H-J. Count me in: response to sexual orientation measures among older adults. Res Aging. 2015 Jul; 37(5):464–80. https://doi.org/10.1177/0164027514542109 PMID: 25651579

53. Sell RL, Becker JB. Sexual orientation data collection and progress toward Healthy People 2010. Am J Public Health. 2001 Jun; 91(6):876–82. https://doi.org/10.2105/ajph.91.6.876 PMID: 11392926

54. Boehmer U. Twenty years of public health research: inclusion of lesbian, gay, bisexual, and transgender populations. Am J Public Health. 2002 Jul; 92(7):1125–30. https://doi.org/10.2105/ajph.92.7.1125 PMID: 12084696

55. Mitchell M, Howarth C, Kotecha M, Kreegan C. Sexual orientation research review 2008. Equality and Human Rights Commission; 2008.

56. Segarra I, Modamio P, Fernandez C, Marino EL. Sex-Divergent Clinical Outcomes and Precision Medicine: An Important New Role for Institutional Review Boards and Research Ethics Committees. Front Pharmacol. 2017; 8:488. https://doi.org/10.3389/fphar.2017.00488 PMID: 28785221

57. Powers MS, Smith PH, McKee SA, Ehringer MA. From sexless to sexy: Why it is time for human genetics to consider and report analyses of sex. Biol Sex Differ. 2017; 8:15. https://doi.org/10.1186/s13293-017-0136-8 PMID: 28473910

58. Mansukhani NA, Yoon DY, Teter KA, Stubbs VC, Helenowski IB, Woodruff TK, et al. Determining If Sex Bias Exists in Human Surgical Clinical Research. JAMA Surg. 2016 Nov 1; 151(11):1022–30. https://doi.org/10.1001/jamasurg.2016.2032 PMID: 27551816

59. Nolan MR, Nguyen TL. Analysis and reporting of sex differences in phase III medical device clinical trials-how are we doing? J Womens Health (Larchmt). 2013 May; 22(5):399–401.

60. Weinberger AH, Smith PH, Kaufman M, McKee SA. Consideration of sex in clinical trials of transdermal nicotine patch: a systematic review. Exp Clin Psychopharmacol. 2014 Oct; 22(5):373–83. https://doi.org/10.1037/a0037692 PMID: 25133506

61. Deutsch MB, Buchholz D. Electronic health records and transgender patients—practical recommendations for the collection of gender identity data. J Gen Intern Med. 2015 Jun; 30(6):843–7. https://doi.org/10.1007/s11606-014-3148-7 PMID: 25560316

62. NW 1615 L. St, Suite 800Washington, Inquiries D 20036USA202-419-4300 | M-857-8562 | F-419-4372 | M. Census may change questions on race, Hispanic origin for 2020 [Internet]. Pew Research Center. [cited 2019 Mar 26]. Available from: https://www.pewresearch.org/fact-tank/2017/04/20/seeking-better-data-on-hispanics-census-bureau-may-change-how-it-asks-about-race/

63. 2015 National Content Test: Race and Ethnicity Analysis Report [Internet]. [cited 2019 Mar 26]. Available from: https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/final-analysis/2015nct-race-ethnicity-analysis.html

64. Nowogrodzki A. Clinical research: Inequality in medicine. Nature. 2017 Oct 4; 550:S18–9. https://doi.org/10.1038/550S18a PMID: 28976953

65. LaVeist T, Pollack K, Thorpe R, Fesahazion R, Gaskin D. Place, Not Race: Disparities Dissipate In Southwest Baltimore When Blacks And Whites Live Under Similar Conditions. Health Affairs. 2011 Oct 1; 30(10):1880–7. https://doi.org/10.1377/hlthaff.2011.0640 PMID: 21976330

66. Leong AB, Ramsey CD, Celedón JC. The Challenge of Asthma in Minority Populations. Clinic Rev Allerg Immunol. 2012 Aug 1; 43(1):156–83.

67. Mega JL, Simon T, Collet J-P, Anderson JL, Antman EM, Bliden K, et al. Reduced-function CYP2C19 genotype and risk of adverse clinical outcomes among patients treated with clopidogrel predominantly for PCI: a meta-analysis. JAMA. 2010 Oct 27; 304(16):1821–30. https://doi.org/10.1001/jama.2010.1543 PMID: 20978260

68. Reference GH. Clopidogrel resistance [Internet]. Genetics Home Reference. [cited 2019 Mar 25]. Available from: https://ghr.nlm.nih.gov/condition/clopidogrel-resistance

69. Clinical Trials Still Don't Reflect The Diversity Of America [Internet]. NPR.org. 2015 [cited 2019 Jan 15]. Available from: https://www.npr.org/sections/health-shots/2015/12/16/459666750/clinical-trials-still-dont-reflect-the-diversity-of-america

70. Laurencin CT. The context of diversity. Science. 2019 Nov 22; 366(6468):929–929. https://doi.org/10.1126/science.aba2319 PMID: 31753971

71. Friedman EA, Gostin LO. From local adaptation to activism and global solidarity: framing a research and innovation agenda towards true health equity. International Journal for Equity in Health. 2017 Feb 21; 16(1):18. https://doi.org/10.1186/s12939-016-0492-8 PMID: 28219436