NATIONAL ACADEMIES
*Sciences
Engineering
Medicine*

**Consensus Study Report**
Highlights

Using Population Descriptors in Genetics and Genomics Research

A New Framework for an Evolving Field

Genetic and genomic information has become far more accessible, and research using human genetic data has grown exponentially over the past decade. The use of genetic information is now widespread across biomedical research. Genetics and genomics research is conducted by a wide range of investigators across disciplines, who often use population descriptors inconsistently and/or inappropriately to capture the complex patterns of continuous human genetic variation. The committee emphasized that a population descriptor is a concept or classification scheme that categorizes people into groups (or "populations") according to a perceived characteristic or dimension of interest. A few examples are race, ethnicity, and geographic location, although this is a non-exhaustive list. Because there is a misconception that humans can be grouped into discrete, innate categories, researchers can inappropriately rely on descent-associated population descriptors such as race to describe population genetic differences. Therefore, there is a need to create clear guidance about the use of population descriptors as well as an opportunity to implement substantive changes to the ways they are used.

In response to a request from the National Institutes of Health, the National Academies assembled an interdisciplinary committee of expert volunteers to conduct a study to review and assess existing methodologies, benefits, and challenges in using race, ethnicity, ancestry, and other population descriptors in genomics research. The resulting report focuses on understanding the current use of population descriptors in genomics research; examining best practices for researchers in the use of race, ethnicity, and ancestry as population descriptors; and identifying how best practices could
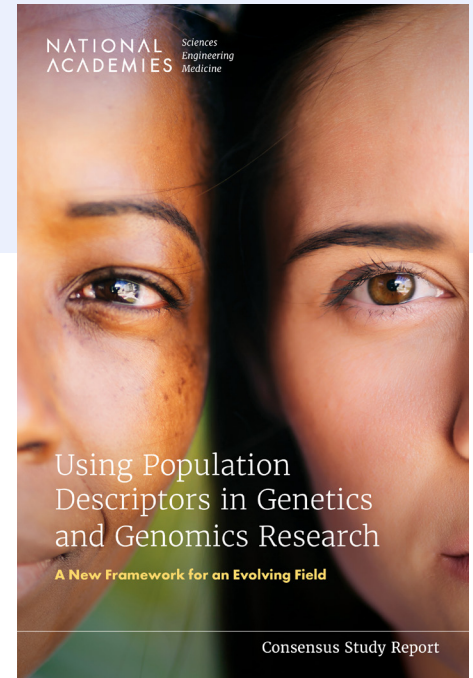
be widely adopted within the biomedical and scientific communities to strengthen genetics and genomics research.

GUIDING PRINCIPLES

Learning from the mistakes of the past and current use of population descriptors is critical to transforming not only the use of these descriptors but also the field of genomics research as a whole. For the recommendations that follow to successfully advance the appropriate use of population descriptors in genetics and genomics research, they must be grounded in ethical and empirical principles that engender trust and enhance the credibility of research. Accordingly, the study committee developed a set of guiding principles to serve as a foundation for its recommendations (see Figure 1): respect, equity and justice, beneficence, validity and reproducibility, and transparency and replicability.

COMMITTEE RECOMMENDATIONS

Requisites for Sustained Change

The committee identified three overarching approaches that are paramount to the long-term success of any effort to address challenges surrounding the use of population descriptors in genetics and genomics research: avoiding typological thinking, including environmental factors in study design, and engaging communities. Though not new, these topics warrant

FIGURE 1 A framework for change. Guiding principles undergird the subsequent recommendations, which fall into three categories: requisites for transforming the use of population descriptors in human genomics research, guidance for researchers conducting different types of genomics studies, and implementation that includes relevant parties supporting researchers and promoting change throughout the genomics research ecosystem.

increased attention and discussion to spark needed progress.

Typological categories, such as the racial and ethnic categories established by the U.S. Office of Management and Budget, are a poor fit for capturing human biological diversity. The committee cautioned against their use for many analytical purposes in genetics research. Specifically, researchers should not use race as a proxy for human genetic variation (**Recommendation 1**). In general, researchers should avoid typological thinking, including the assumption and implication of hierarchy, homogeneity, distinct categories, or stability over time of groups of people in studies of human genetic variation (**Recommendation 2**). Researchers should also be attentive to the connotations and impacts of the terminology they use to label groups (**Recommendation 3**).

Genetic effects cannot be adequately explained without nongenetic, or environmental, contexts. Sometimes population descriptors are used as proxies for environmental factors or exposures. Rather than relying on inadequate proxies, researchers conducting human genetics studies should directly evaluate the environmental factors or exposures that are of potential relevance to their studies (**Recommendation 4**). If it is necessary to use population descriptors as proxies, researchers should explain how they are used and why they are relevant. Collaborating with experts in the social sciences, epidemiology, environmental sciences, or other relevant disciplines can help advance this effort.

Community engagement improves communication, study coordination, and trustworthiness of research. Researchers should work in ongoing partnerships with study participants and community experts to integrate the perspectives of communities and to inform the use of population descriptors (**Recommendation 5**).

Guidance for the Selection and Use of Population Descriptors in Genetics and Genomics Research

It is important to have a clear understanding of what population descriptors represent—and a rigorous rationale for using them—before selecting the appropriate group categories and labels with which to work.
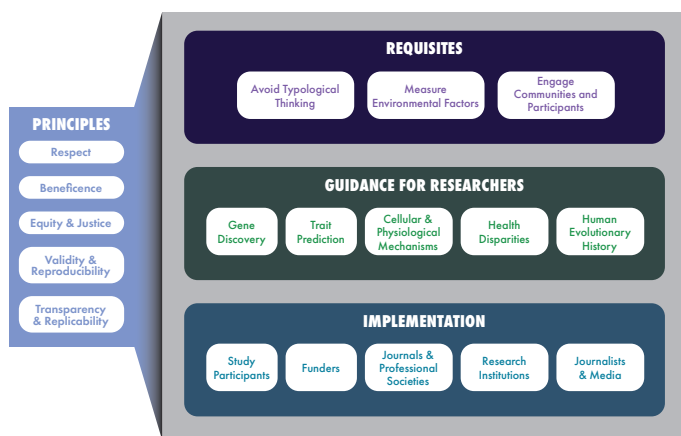
There is no one-size-fits-all solution to the challenge of using population descriptors; rather, the appropriate population descriptor depends on the scientific question that is being addressed. Therefore, researchers should tailor their use of population descriptors to the type and purpose of the study and explain how and why they used those descriptors (**Recommendation 6**). Researchers can also consider using multiple descriptors for each study participant to improve clarity. See Table S-1 in the report for recommended approaches organized by genomics study type.

When communicating their research methods, findings, and conclusions, researchers should be as transparent as possible about the specific procedures used to name groups within their data sets. Transparent communication includes stating how and why particular decisions were made. For every descriptor selected, labels should be applied consistently to all participants (**Recommendation 7**). For example, if ethnicity is the descriptor, all participants should be assigned an ethnicity label, rather than labeling some by race, others by geography, and yet others by ethnicity or nationality. Researchers should also disclose the process by which they selected and assigned group labels and the rationale for any grouping of samples, including for legacy samples (**Recommendation 8**).

Implementation and Accountability

Many aspects of the current systems that fund, support, evaluate, and reward genomics research must adapt to better facilitate implementation of the committee's recommendations. The genomics research ecosystem has many players, including funders of genetics and genomics research, professional societies, research journals, individual researchers, and research institutions, who all share responsibility for making these changes. These key partners should ensure that policies

and procedures are aligned with these recommendations and invest in developing new strategies to support implementation when needed (**Recommendation 12**).

To facilitate the inclusion of environmental measures and the engagement of diverse communities in genomics research, funding agencies and research institutions should develop strategies to encourage and reward interdisciplinary collaboration among researchers with different areas of expertise, including genetics and genomics, social sciences, epidemiology, and community-based research (**Recommendation 10**). In addition, funders and research institutions should create new initiatives to advance the study and development of best practices for population descriptor usage in genetics and genomics research (**Recommendation 11**). Funding agencies, research institutions, research journals, and professional societies should also offer tools widely to their communities to facilitate the implementation of these recommendations (**Recommendation 9**).

Importantly, because the field of genomics research is dynamic and evolving, responsibility for periodic reevaluation of these recommendations should be overseen by effective, multidisciplinary advisory groups (**Recommendation 13**). These advisory groups could periodically reevaluate established best practices on the use of descent-associated population descriptors to ensure that they reflect the current state of the science and an ongoing commitment to ethical and empirical principles.

LOOKING FORWARD

It will take a concerted effort by all relevant parties, patience, and time to achieve the desired goal of properly using and reporting population descriptors routinely and consistently. The recommendations in this report provide a pathway to generate lasting change in an evolving field.

COMMITTEE ON THE USE OF RACE, ETHNICITY, AND ANCESTRY AS POPULATION DESCRIPTORS IN GENOMICS RESEARCH

ARAVINDA CHAKRAVARTI (*Co-Chair*), New York University Grossman School of Medicine; **CHARMAINE ROYAL** (*Co-Chair*), Duke University; **KATRINA ARMSTRONG,** Columbia University; **MICHAEL BAMSHAD,** University of Washington and Seattle Children's Hospital; **LUISA N. BORRELL,** City University of New York; **KATRINA CLAW,** University of Colorado Anschutz Medical Campus; **CLARENCE C. GRAVLEE,** University of Florida; **MARK D. HAYWARD,** The University of Texas at Austin; **RICK KITTLES,** Morehouse School of Medicine; **SANDRA SOO-JIN LEE,** Columbia University; **ANDRÉS MORENO-ESTRADA,** Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Mexico; **ANN MORNING,** New York University; **JOHN P. NOVEMBRE,** University of Chicago; **MOLLY PRZEWORSKI,** Columbia University; **DOROTHY E. ROBERTS,** University of Pennsylvania; **SARAH A. TISHKOFF,** University of Pennsylvania; **GENEVIEVE L. WOJCIK,** Johns Hopkins Bloomberg School of Public Health

STAFF SARAH H. BEACHY, Study Director; **SAMANTHA SCHUMM,** Associate Program Officer; **LEAH CAIRNS,** Study Co-Director (*until October 2022*); **KATHRYN ASALONE,** Associate Program Officer; **MEREDITH HACKMANN,** Associate Program Officer; **LYDIA TEFERRA,** Research Assistant; **APARNA CHERAN,** Senior Program Assistant (*from June 2022*); **MICHAEL K. ZIERLER,** Science Writer; **ANDREW M. POPE,** Senior Director, Board on Health and Sciences Policy (*until July 2022*); **CLARE STROUD,** Senior Director, Board on Health and Sciences Policy (*from July 2022*); **MALAY K. MAJMUNDAR,** Director, Committee on Population

Health and Medicine Division
Division of Behavioral and Social Sciences and Education

NATIONAL ACADEMIES

Sciences
Engineering
Medicine