

## Meeting Summary

# Workshop on the Use of Race and Ethnicity in Genomics and Biomedical Research

October 24-25, 2016 | Rockville, Maryland

## Workshop Objective

On October 24 and 25, 2016, the National Human Genome Research Institute (NHGRI) and the National Institute on Minority Health and Health Disparities (NIMHD) convened a workshop to discuss the use of race and ethnicity data in genomics, biomedical, and clinical research, and their application to minority health and health disparities. Participants included genomic, clinical, epidemiologic, and social science researchers, NIH staff, government stakeholders, and other experts to discuss what efforts are needed to explore how human diversity including race and ethnicity is studied and reported now and how it should be in the future.

The usefulness of social and political population labels, including those based on race and ethnicity in genomics and biomedical research is a topic of debate. This discussion is significant not only for defining rigorous scientific study designs, but for influencing how scientists and the public conceptualize, discuss, and react to human differences. Misuse of population descriptors in biomedical research has the potential to perpetuate misinformation, stigmatize certain groups, and simplify the complex relationships between individual identity, genetics, and health.

The workshop's objective was to explore how genomics and biomedical research can describe research participants' diverse backgrounds and experiences in ways that are scientifically and socially meaningful. Meeting participants were asked to identify research questions to advance understanding of self-identified race and ethnicity (SIRE) and ancestry informative markers (AIMs) data in biomedical research, and generate recommendations for the appropriate use of race, ethnicity, and other population descriptors in scientific research.

The two-day workshop consisted of presentations and discussions during consecutive sessions, each focused on a particular aspect of race and ethnicity in research. Presenters were asked to discuss their own scholarship around the sessions' theme, and give examples that help illustrate their perspective. The main topics of these sessions included the appropriate use of SIRE and AIMs in various research contexts, the impact to minority health research and health disparities research, the reporting of genetic variation including in non-research contexts such as clinical and direct-to-consumer testing, collaboration between genomics and social sciences, utilizing electronic health records (EHRs) in personalized medicine and public health research, and potential roles for the National Academy of Sciences (NAS) and scientific journals in exploring and implementing new research and reporting frameworks.

## Workshop Discussion

Although the workshop did not aim to create unanimous recommendations or formal consensus positions, the statements here came out of the moderated discussion with broad agreement among participants. Broadly, discussants agreed that there are compelling reasons to study race and ethnicity in biomedical research from both a biological and social perspective. Race and ethnicity have historically been used as basic demographic factors. Today, they are often used as a proxy for socioeconomic status, and to track inequity in health care. The definition of race and ethnicity by self-identification (SIRE) is seen as the gold standard in research. An individual's SIRE is an important component of his or her identity and experience in society. Perceived race, or how an individual's race is defined by society, is also important in determining the effects of race on health.

However, while race and ethnicity are often framed as predictors of health outcomes, they are social and political constructs - not biological realities - that constitute self-identification with groups defined by categories such as ancestry, geography, and culture. The lack of a clear definition of race in biomedical research results in conflation of self-identified race categories with older, socio-political definitions of race, and can lead to negative perceptions of research with race and ethnicity measures. Presenters showed evidence that current SIRE largely correlates with shared genetic ancestry in broadly classified groups (African, Asian, European), but generally agreed that finer and more standardized understanding of genetic differences and environmental contributors are important factor in advancing knowledge of disease mechanisms and predicting variance in clinical outcomes. The growing number of "mixed" or "multiracial" individuals emphasizes the need for better ways to study and report biological and social influencers of health that complement SIRE, and illustrates the limitations of using the current race categories on health disparities research.

The broader scientific community also has a role to play in reevaluating how research utilizing race and ethnicity measures is reviewed, and encouraging participation of underrepresented groups. There is a long history of minority underrepresentation in biomedical research that has resulted in health care disparities, including disparities in the development of effective diagnostic tools and therapies. The difficulty of recruiting broad racial and ethnic groups was acknowledged; the number of participants from minority population are usually small and are often excluded from results for not having enough statistical power for analysis. The lack of minority participation stems from a variety of sources including mistrust after prior and historical research abuses, lack of funding and time for community involvement, poor communication and outreach, and unclear benefits of participating in research. The scientific community needs to improve outreach to minority groups and to increase the transparency and accessibility of their research in order to rectify the inequality of representation in the scientific data.

Participants also discussed the need to exercise caution with the language used in gathering and reporting race-associated biomedical data so that the lay public does not draw erroneous conclusions, such as that an entire racial group has increased risk for a disease. When reporting on race- and ethnicity-associated data to the general public, scientists should avoid technical genetic information and should instead use actionable language that is accurate about disease risk.

With large cohort studies occurring around the world, international dimensions to the issues about race and ethnicity are also important to the future of this discussion. The scientific community is still learning how best to use genomic tools and leverage other resources to improve medicine, and it needs to define the types of research studies that will best inform the future of biomedical science. More discussion and study is needed into how to make the data planned and currently gathered from these international cohort studies the most comparable and useful it can be. The categorization of population groups should be guided by research questions. For both recruiting and reporting purposes, researchers need to be explicit about the basis of group classifications and the outcomes for which the data are being used.

The workshop concluded with two discussion sessions during which recommendations to share with the broader scientific community were generated. The themes are recommendations presented here reflect both individual and collective opinions of the workshop participants but are not necessarily consensus views or those of the NHGRI, NIMHD, or other U.S. government bodies. The workshop agenda and participant list are included as appendices to this summary.

## Discussion and Highlighted Themes

**Without abandoning current OMB categories in US-based research, researchers should expand the depth of information collected regarding race and ethnicity in a comprehensive and standardized way.**

- Examine multiple dimensions of racial/ethnic identity, which can include SIRE, self-identified ancestry, and genomically-identified ancestry, assessment by others (other identified), how individuals believe others perceive them (reflected appraisal).
- Make resulting data comparable and searchable across clinical, research, and public health data.
- Establish standardized data collection for utilization of electronic health care records for genomic and biomedical research.
- Use OMB categories for US-based research in the context of recruitment and enrollment reporting, but do not require their use in reporting scientific data.

**Encourage the expansion of data categories that need to be collected in addition to race and ethnicity that will help explain population differences (e.g. SES, region, zip code, immigration status).**

- When limited data can be collected, prioritize available measures.
- Discourage the use of race and ethnicity as a proxy for expanded data categories.

**Examine and, if necessary, revise the instructions to researchers regarding collecting and reporting race and ethnicity to ensure they do not limit the ways in which population data can be reported.**

- Encourage or require researchers be able to explain how race and ethnicity is collected and used in their research and why.
- Develop methods for meaningfully analyzing small sample populations, including those with unknown, “other,” or “mixed” racial or ethnic categorization.

**Examine and develop a global paradigm of common descriptors for population groups.**

- Explore the use of race and ethnicity in international research and the challenges and opportunities for making data comparable.
- Recommend terminology, collection, and analysis guidelines that are equally applicable to global research as to US-based research.

**Research longitudinal frameworks of racial and ethnic identity.**

**Incorporate research into the use of race and ethnicity into existing program announcements.**

