

World view



By Alice B. Popejoy

Too many scientists still say Caucasian

Racist ideas of categories for human identity continue to warp research and medicine.

Of the ten clinical genetics labs in the United States that share the most data with the research community, seven include ‘Caucasian’ as a multiple-choice category for patients’ racial or ethnic identity, despite the term having no scientific basis. Nearly 5,000 biomedical papers since 2010 have used ‘Caucasian’ to describe European populations. This suggests that too many scientists apply the term, either unbothered by or unaware of its roots in racist taxonomies used to justify slavery – or worse, adding to pseudoscientific claims of white biological superiority.

I work at the intersection of statistics, evolutionary genomics and bioethics. Since 2017, I have co-led a diverse, multidisciplinary working group funded by the US National Institutes of Health to investigate diversity measures in clinical genetics and genomics (go.nature.com/3su2t8n).

Many working in genomics do have a nuanced understanding of the issues and want to get things right. Still, I have been dismayed by how often the academics and clinicians I’ve encountered shy away from examining, or even acknowledging, how racism warps science. Decades of analyses have shown that ‘racial groups’ are defined by societies, not by genetics. Only the privileged have the luxury of opining that this is not a problem. As a white woman, I too have blind spots that need constant examination.

Pioneering works in social science such as Dorothy Roberts’ *Fatal Invention* (2012), Kim Tallbear’s *Native American DNA* (2013) and *The Social Life of DNA* (2016) by Alondra Nelson, have eloquently pointed out many of the flawed assumptions and approaches that plague human genomics.

A common theme of this scholarship is that groupings depend more on dominant culture than on ancestry. In Singapore, the government mandates that individuals are identified explicitly as Chinese, Malay, Indian or Other, which affects where they can live and study. In the United States, people with ancestry from the world’s two most populous countries, India and China, along with every other country on the continent, are collapsed into a single racial category called ‘Asian’. Similarly, the term ‘Hispanic’ erases a multitude of cultural and ancestral identities, especially among Indigenous peoples of the Americas.

Erroneous ideas about genetic ‘races’ live on in the broad, ambiguous ‘continental ancestry’ groups such as ‘Black, African’ or ‘African American’, that are used in the US Census and are ubiquitous in biomedical research. These collapse incredible amounts of diversity and erase cultural and ancestral identities. Study participants deemed not to fit within such crude buckets are often excluded from

‘Racial groups’ are defined by societies, not by genetics.”

Alice B. Popejoy is a computational public-health geneticist in the Department of Biomedical Data Science and Center for Biomedical Ethics at Stanford University, California, USA.
e-mail: apopejoy@stanford.edu

analyses, despite the fact that fewer and fewer individuals identify with a single population of origin.

One practical way forwards is to move away from having people identify themselves using only checkboxes. I am not calling for an end to the study of genetic ancestry or socio-cultural categories such as self-identified race and ethnicity. These are useful for tracking and studying equity in justice, health care, education and more. The goal is to stop conflating the two, which leads scientists and clinicians to attribute differences in health to innate biology rather than to poverty and social inequality.

We need to acknowledge that systemic racism, not genetics, is dominant in creating health disparities. It shouldn’t have taken the inequitable ravages of a pandemic to highlight that. Furthermore, every researcher and physician should be aware of the racial bias that abounds in medical practice: some pulse oximeters give more accurate readings for light-skinned people than for those with dark skin; Black Americans are undertreated for pain; and historical biases in data used to train algorithms to make medical decisions can lead to worse outcomes for vulnerable groups. Hence the ongoing revisions to the subsection on race and ethnicity in the American Medical Association’s *Manual of Style*, and why medical schools are examining how their curricula reinforce harmful misconceptions about race.

Thankfully, more researchers are collecting self-reported data on geographical family origins, languages spoken at home and cultural affiliations. I’d like to see data-collection forms with open-ended questions, rather than those that force fixed choices or reduce identity to a box labelled ‘other’. These self-reported indicators could be combined with genetic data to improve on current approaches to mapping the dimensions of diversity in our populations.

Approaches to genetic ancestry based on known reference populations are inadequate, in part because so much global diversity is missing from our data. I am working with the Human Pangenome Reference Consortium, which aims to generate a more accurate and inclusive resource for global genomic diversity. It will include communities, especially Indigenous peoples, in developing protocols for data collection, storage and use. This respects Indigenous data sovereignty, and makes for more accurate and inclusive studies.

The more precisely we can measure genetic and non-genetic contributors to health and disease, the less researchers will rely on biologically meaningless designations that reinforce faulty assumptions and cause harm. The use of sequence data in clinical care could, for instance, facilitate recommendations for drug dosage that are genotype-based, rather than race-based.

Simply picking another word to replace ‘Caucasian’ won’t be enough to root out racism in research and medicine. But all should be aware of the harms the word represents.