



POLICY FORUM: GENETICS

Toward a New Vocabulary of Human Genetic Variation

Pamela Sankar and Mildred K. Cho

Recent genetic variation research has reinvigorated the dispute over the validity of race as a research variable (1–8). Proponents of using race assert that genetic differences and racial classification are strongly

Enhanced online at
www.sciencemag.org/cgi/
content/full/298/5597/1337

associated, and so support the use of race in the design of research and the application of its findings (3, 6). Critics cast race as a social construct and counter that putting subjects into racial groups fundamentally misrepresents human genetic variation and hinders research (2, 8–9). Several solutions have been offered, such as replacing race with ethnicity (10–11) or with genetic markers (12). However, although these suggestions might apply to certain kinds of research, none provides an overall solution.

This is because there are many distinct meanings to the word race, and different ways of using it as a research variable. For example, a popular definition today describes race as a social construct that incorporates beliefs about language, history, and culture (13). Here, race forms the basis on which social identity, traditions, and politics are built. This concept has been promoted as an alternative to an earlier genetic theory of race, which has been scientifically repudiated and rejected, that divided the human species into subspecies that were ranked on the basis of skill, intelligence, and morality (14). However, rejecting race as genetic hierarchy is not tantamount to rejecting the idea that human populations differ genetically.

Conceptualizing race as a social construct has helped to undermine racism by eliminating its alleged natural basis. However, it has also had the unintended consequence of eliminating a legitimate

basis for discussion of population-based genetic differences. Insistence that there is no such thing as race in a genetic or biological sense, as well as the lack of an alternative term for discussing genetic diversity leaves those who wish to discuss genetic diversity without a functional vocabulary (15–18).

Two points should help to clarify these issues. First, the debate in genetic variation research over race as a research variable is not a debate over whether human populations differ genetically. Rather the debate is over the scientific, clinical, and social significance of labeling genetic differences as race, as something else, or not at all (19).

Second, many of the disagreements about race and genetics are fostered by confusion

Strong Amerind/White Sex Bias and a Possible Sephardic Contribution Among the Founders of a Population in Northwest Colombia

I Am a Racially Profiling Doctor

Illness isn't colorblind. So why is it taboo for doctors to take note of a patient's race?



Genomics Firm Aims to Fill Asian Gene Gap

Caucasian Genes in American Blacks: New Data

Characteristics of Autoimmune Hepatitis in Patients Who Are Not of European/Caucasoid Ethnic Origin

Metric Assessment of Race from the Pelvis in South Africans

over the relationship of these two concepts. For example, recent studies have exam-

ined DNA samples from various populations and clustered them into groups based on identity of DNA sequences at several loci. In some studies (20), but not others (8), genomes examined by this method do sort out in a way that reflects race as social construct, depending on how many or which genetic loci are compared. It is not that race exists in one population and not in another. Rather, it may be that the appearance of clustering is a function of how

populations are sampled (21), of how criteria for boundaries between clusters are set, and of the level of resolution used. In the same way that the earth can be described by many different kinds of maps—from topological to economic—so, too, can the naturally occurring genetic variation among populations be divided in numerous ways and be made to highlight any chosen similarity or difference.

Proposed Alternatives to Race

This framework helps to explain why replacing race with ethnicity or genetic markers fails to solve the problem of race in genetic research. Some studies use race as a way to enroll a sample of people with genotypes more likely to be similar or diverse in a particular way. For such a study, genetic markers might effectively replace race, but only if the markers chosen happen to be distributed among the selected populations in the same way as the variable of interest in that particular study. This reasoning suggests that it may only be a matter of time before sufficient genetic markers are identified. However, results of such research still need to be socially or geographically grounded for clinical application (22).

Another popular alternative to race is ethnicity, a term with multiple, conflicting meanings (23). Anthropologists initially proposed ethnicity to direct attention away from genetics and toward social and historical factors as explanations of population variation (24). The term has gained users simply because to some researchers the word “ethnicity” seems

more acceptable. However, its actual application is often identical to race (10, 23). Nevertheless, if applied in its original sense to define a population socially or culturally, ethnicity

could replace race in research when a researcher seeks a variable that corresponds to the behavioral aspects implied by the term, such as diet, occupation, social status, or health beliefs (10, 11).

For some studies, a combination of genetic and social markers may be appropriate. However, examining more closely the range of uses to which genetic researchers put race and the problems they encounter when they do so might provide some guidance on how research should proceed.

Race as a Variable in Genetic Research

Researchers interested in genetics incorporate race into research designs in several ways. For example, pharmacogenetic studies may be inventorying the genotypes occurring in human populations and may ideally sample from a diverse set of populations to attempt to represent a broad range

P. Sankar is at the Center for Bioethics, University of Pennsylvania, Philadelphia, PA 19104–3308, USA. E-mail: sankarp@mail.med.upenn.edu. M. K. Cho is at the Stanford Center for Biomedical Ethics, Palo Alto, CA 94304, USA. E-mail: micho@stanford.edu

of genetic difference. Association studies seek genetic differences between those with and without a particular phenotype. The more genetically similar they are, the easier it is to find the specific genetic differences that account for differences in phenotype. Race is used as a proxy for genetic relatedness to control for potential confounding that occurs if the study populations differ genetically in ways not related to the phenotype in question. In contrast, epidemiological studies seeking to determine risk factors for disease may want to use race to control for population stratification, but also as a proxy for environmental exposures, including social interactions (e.g., people of certain races may be more or less likely to be referred for further treatment).

What appears as straightforward use of a variable, however, becomes complicated when researchers decide how to put it into operation. In all but the final example, the researcher is using race as a way of grouping subjects by similarity or difference of genetic sequence, which reflects population history. In the absence of known genetic markers, researchers need to access this variation through race or through ethnicity, which is the way certain aspects of genetic variation have been socially represented. Several conventions exist, such as asking subjects how they self-identify ethnically or racially, or where their grandparents were born.

Complicating Factors

Some of the complexity of race comes from its multiple, overlapping meanings that span popular and scientific use. Further confusion is generated, however, by the tendency to leave race undefined. This leads to three sorts of problems in the conduct and reporting of genetic research: (i) nonequivalent uses of race within one research report; (ii) inverting the relationship between genetics and race, or studying race as an end in itself; and (iii) an overemphasis on race.

Accessing a particular set of conditions with a variable requires choosing the right variable and using it consistently. While this may seem obvious, race-related genetic research does not always observe this rule. For example, the initial reference to race in an article is often to the racial identity of individual subjects, sometimes described as "self-assigned" by subjects. A subsequent reference to race might appear in the classification of genotypes associated with groups of the self-identified subjects. The final one might appear in the discussion section that generalizes the findings to different racial groups, i.e., massive world populations, such as Euro-Americans or Asians.

Nonequivalent use of labels is illustrated by the common juxtaposition of terms such as "white" with "African-American," where skin color and geographical location are treated as equivalent. Another example is the juxtaposition of "Asian-American" with "Mexican-American," which implies that people of Asian ancestry now living in the United States represent a level of genetic diversity that is equivalent to that of people of Mexican ancestry now living in the United States. Such examples indicate a need for more consistent attention to definition of groups and to the need to explain the rationale for their equivalence.

Another set of problems in race research results when researchers attempt to map genetics to race (or to other characteristics that are, in part, socially determined) as an end in itself. One such study set out as its goal, "to identify a set of genetic markers that would allow the confident determination of ethnicity, for use in a forensic setting" (25). A similar problem is caused by the use of language such as "white chromosome," "mutant African alleles," or "Asian gene gap" as scientific shorthand, implying that some genetic variants "belong" exclusively to some races (26–28). Even if, in rare circumstances, certain alleles have been found exclusively in one population, to call a chromosome white or Asian makes an inappropriate link between a rapidly shifting social term and a fixed biological entity.

Conclusion

Race has been retained in genetic research on the basis of the belief that it is a social or geographical unit that approximates a genetic grouping. We do not argue for the imposition of any one particular set of terms to describe genetic groupings, or for the wholesale elimination of race from genetic research. Rather, we stress that this type of boundary is not likely to be equally useful in all kinds of genetic research. Furthermore, researchers need to be clear about the choice and definition of terms, as well as to be careful about making appropriate generalizations. Funders and publishers of biomedical research should follow the suggestions of editorials in *Nature Genetics*, *Archives of Pediatrics and Adolescent Medicine*, and the *British Medical Journal* and ask researchers to define race when they use it (29–31). Editors should take care that these rules are followed. Our own preliminary analysis of articles published after these editorials reveal few if any changes in the explanations provided concerning race as a research variable.

In designing genetic studies, researchers should first consider whether they want to use race as a proxy for genet-

ic similarity or diversity, or as a proxy for nongenetic factors such as socioeconomic status, or both. Are there other, more direct, measures available that should be used instead? If not, it is important to consider the level of resolution necessary to describe populations, to use groupings that are comparable in resolution, and to describe them precisely. Sometimes, nationality might suffice, whereas other investigations might require a smaller geographical region or allow a larger one. Thus, it is important to collect data with as much precision as possible and to note always how subjects were assigned to groups, such as on the basis of records or self-assignment. It is imperative for the research community to acknowledge that the maps used in research are not the only maps used to describe the terrain they study and that careful use of language is necessary to avoid misunderstanding.

References and Notes

1. N. Angier, *New York Times*, 22 August 2000, p. F1.
2. F. S. Collins, M. K. Mansoura, *Cancer Suppl.* **91**, 221 (2001).
3. W. M. Kalow, *Clin. Pharmacol. Ther.* **70**, 1 (2001).
4. M. W. Foster, R. R. Sharp, *Genome Res.* **12**, 844 (2002).
5. D. W. Nebert, E. Bingham, *Trends Biotechnol.* **19**, 519 (2001).
6. N. Risch, E. Burchard, E. Ziv, H. Tang, *Genome Biol.* **3**, 2007.1 (2002).
7. S. G. Stolberg, *New York Times*, 13 May 2001, p. A1.
8. C. Romualdi et al., *Genome Res.* **12**, 602 (2002).
9. S. S. Anand, *Ethn. Health* **4**, 241 (1999).
10. M. A. Haynes, B. D. Smedley, Eds., *The Unequal Burden of Cancer: An Assessment of NIH Research and Programs for Ethnic Minorities and the Medically Under-served* (National Academy Press, Washington, DC, 1999).
11. H. P. Freeman, *Cancer* **82**, 219 (1998).
12. W. Gilbert, in *The Code of Codes: Scientific and Social Issues in the Human Genome Project* (Harvard Univ. Press, Cambridge, MA, 1992).
13. R. Witzig, *Ann. Intern. Med.* **125**, 675 (1996).
14. J. Marks, *What It Means to Be 98% Chimpanzee* (Univ. of California, Berkeley, 2002).
15. R. S. Schwartz, *N. Engl. J. Med.* **344**, 1392 (2001).
16. Editorial, *Nature Genet.* **29**, 239 (2001).
17. K. Owens, M.-C. King, *Science* **286**, 451 (1999).
18. www.aaanet.org/gvt/ombsumm.htm (2000).
19. L. Cavalli-Sforza, P. Menozzi, A. Piazza, *The History and Geography of Human Genes* (Princeton Univ. Press, Princeton, NJ, 1994).
20. L. Jorde et al., *Proc. Natl. Acad. Sci. U.S.A.* **94**, 3100 (1997).
21. M. Krings et al., *Am. J. Hum. Genet.* **64**, 1166 (1999).
22. M. W. Foster, R. R. Sharp, *Genome Res.* **12**, 844 (2002).
23. G. M. Oppenheimer, *Am. J. Public Health* **91**, 1049 (2000).
24. A. Montague, *Man's Most Dangerous Myth: The Fallacy of Race* (Columbia Univ. Press, New York, 1942).
25. M. Shriver et al., *Am. J. Hum. Genet.* **60**, 957 (1997).
26. D. C. Crawford, *Am. J. Hum. Genet.* **66**, 480 (2000).
27. E. Fritsche, G. S. Pittman, D. A. Bell, *Mutat. Res.* **432**, 1 (2000).
28. D. Cyranoski, *Nature* **416**, 115 (2002).
29. "Style Matters: Ethnicity, race, and culture: Guidelines for research, audit, and publication," *BMJ* **312**, 1094 (1996).
30. Editorial, *Nature Genet.* **24**, 97 (2000).
31. F. Rivara, L. Finberg, *Arch. Pediatr. Adolesc. Med.* **155**, 119 (2001).
32. Supported by NIH NHGRI grant number: R01 HGO 2189-02. P.S. acknowledges the benefit of AAAS Short Course on Research Ethics, 2000–2001.