

“Race-Ethnicity”: A Dubious Scientific Concept

Buehler and coauthors (1) are entrapped by the persistent errors made when epidemiologists designate race and ethnicity (2). The fault is not that of the authors, but of the penetration into science of our culture's muddled thinking about race. For example, Buehler and coauthors uncritically note that the categories recommended for specifying “race-ethnicity” in the Epidemiology Surveillance Project are “white non-Hispanic, black non-Hispanic, Hispanic, Native American (includes American Indians, Eskimos, Aleuts), Asian and Pacific Islander, and unspecified.” This classification scheme obviously derives from Directive No. 15: Race and Ethnic Standards for Federal Statistics and Administrative Reporting (3), which is still in effect. What has escaped Buehler and coauthors, and many other authors, is that Directive No. 15 acknowledges that its own classifications “should not be interpreted as being scientific or anthropological in nature . . . They have been developed in response to needs expressed by both the executive branch and the Congress . . .”

In fact, there can be no logic in a scheme that classifies a sample of humans into selected races or selected ethnic groups—all on the same list. Even if we assume that the concepts of race and ethnicity are meaningful, they are entirely different from one another. Why should an Hispanic be disenfranchised from having his race count? After all, an Hispanic is defined in Directive No. 15 as “a person of Mexican, Puerto Rican, Cuban, Central or South American or other Spanish culture or origin, *regardless of race*” (italics added). And if ethnicity is important, why should a so-called black non-Hispanic or white non-Hispanic be disenfranchised from having his *ethnicity* count?

Many researchers think of race as a useful marker for genetic variation. It is time for public health workers and epidemiologists, at least when wearing their scientific hats, to start turning away from this flawed notion. The reality of racial classification in the United States is that many people are classified as “blacks” even though a large majority of their ancestors would be classified as “whites.” Many people are classified racially as “Hispanics,” even though Hispanic is properly an ethnic characterization—and a very imprecise one. Many people of Swedish or Iranian ancestry are classified together as “whites,” and many people of Japanese or Filipino ancestry are classified together as “Orientals” despite genotypic and phenotypic differences between such pairs at least as great as those between many “whites” and many “blacks.” And many people classified as “Native Americans” have an ancestry that is indistinguishable from that of many people classified as “Hispanics.” So much

for race as a marker for genetic variation.

Our cultural heritage often makes us perceive race, however defined, to be more focal than it really is. Even the gene responsible for the sickle cell trait, almost always considered in the United States to be a “black gene,” is as widely distributed in the Middle East and India as in Africa, and is absent from many parts of tropical Africa (4). Likewise, we often unduly emphasize race as a relevant concept in epidemiology, even while we ignore demographic concepts and variables likely to be of considerable importance, such as salt intake, pet ownership, residence on a reservation, or handedness.

Regardless of the means whereby public health workers classify populations, the bases for classifications ought to be purposeful, logical, and explicit. Workers concerned with genetics might be expected to actually use techniques of modern genetics to define and subdivide populations, and to state what their techniques are. Perhaps some investigators wish to resort to old-fashioned calipers or color charts, or to devise new rating scales of demonstrated reliability. If so, they should identify their methods. Workers concerned with the concept of ethnicity, on the other hand, should describe whether place of birth, language spoken at home, favorite food, surname, duration (in years or generations) of residence at a stated location, or some other characteristic provides the objective criterion for the designation, and they, too, should present in conventional detail how their data were collected. And if workers study variables conceptualized otherwise, such as economic status, birth order, disease history, drug use, diet, and so forth, then their criteria should be described.

As things stand, rationales for grouping people into so-called race-ethnicity groups, as in the paper by Buehler and coauthors, are obscure. Methods for making groupings too often consist of self-classifications by untutored lay subjects, of casual classifications by admissions clerks, or of nonscientific definitions suiting political or social needs. Most often, details of the methods or criteria defining race or ethnicity are simply unstated. Valid scientific conclusions can not emanate from such information.

Albert Weissman, PhD, Department of Clinical Research, Pfizer, Inc., Groton, CT

References

1. Buehler, J. W., Stroup, D. F., Klaucke, D. N., and Berkelman, R. L.: The reporting of race and ethnicity in the National Notifiable Diseases Surveillance System. *Public Health Rep* 104: 457-465, September-October 1989.
2. Weissman, A.: On the designation of race in clinical pharmacology reports. *Drug Information J.* In press.

3. Directive No. 15: Race and ethnic standards for federal statistics and administrative reporting. *Federal Register* 43: 19269-19270, May 4, 1978.
4. Livingstone, F. B.: Anthropological aspects of the distribution of the human hemoglobin variants. *In Hemoglobin variants in human populations*, edited by W. P. Winter. CRC Press, Inc., Boca Raton, FL, 1986, pp. 17-28.

Authors' Response

By agreement between State epidemiologists and the Centers for Disease Control (CDC), States use five categories for race and ethnicity in weekly reports to the National Notifiable Diseases Surveillance System. These categories oversimplify the diversity of our population, but they offer data on race and ethnicity not available through this system before new reporting methods were implemented in 1985. The answer to one question often raises others, and there is a natural tendency to want more data. However, in concluding that these data are not useful, Dr. Weissman does not adequately consider the purpose and process of public health surveillance.

Surveillance is used to monitor disease trends and identify groups at risk (1). Public health professionals use these data to identify situations that require investigation or intervention. As an ongoing process, surveillance must focus on the information that is most important for disease prevention and control and must balance conflicting needs, such as detail versus timeliness. Clearly, the weekly system could not collect all of the information Dr. Weissman recommends. Physicians may not routinely collect such data; health departments may not have sufficient staff to conduct such extensive investigations; and the entire process would delay reports unacceptably.

Some of the questions raised by Dr. Weissman are more appropriately addressed in disease-specific surveillance systems or studies. For example, the incidence of the acquired immunodeficiency syndrome (AIDS) is higher in Hispanics than in non-Hispanic whites. As documented by AIDS surveillance, this is largely due to the high rates of AIDS associated with intravenous drug use in persons of Puerto Rican ancestry—information that can lead to targeted interventions (2). Similarly, AIDS surveillance in San Francisco collects more detailed information on race and ethnicity for persons of Asian ancestry, permitting identification of those at highest risk and development of culturally relevant information campaigns (3). Other disease-specific surveillance systems complement the weekly reporting system. Typically, they collect more detailed information but do not provide a weekly "snap-shot" of disease trends.

Dr. Weissman dismisses "race as a marker for genetic variation." Regardless of the merits of this argument, we did not suggest that it is. Instead, we used race and ethnicity only as a proxy for the "social, environmental, and economic disadvantages associated . . . with an increased risk of infectious diseases." In the future, we may be able to gather additional information that more directly measures these underlying problems, but in the

meantime, we should use the information that we have. The five broad categories for race and ethnicity are associated with substantial differences in disease risk—a strong argument in itself for their usefulness in public health surveillance. Differences in disease occurrence identified by these categories for race and ethnicity support further efforts to characterize and reduce disparities.

James W. Buehler, MD, and Ruth L. Berkelman, MD, Division of HIV/AIDS, Center for Infectious Diseases, and Donna F. Stroup, PhD, and Douglas N. Klaucke, MD, MPH, Division of Surveillance and Epidemiologic Studies, Epidemiology Program Office—all four are at the Centers for Disease Control, Public Health Service, Atlanta, GA.

References.....

1. Thacker, S. B., Berkelman, R. L.: Public health surveillance in the United States. *Epidemiol Rev* 10: 164-190 1988.
2. Selik, R. M., Castro, K. G., Pappaioanou, M., Buehler, J. W.: Birthplace and the risk of AIDS among Hispanics in the United States. *Am J Public Health* 79: 836-839, July 1989.
3. Woo, J. M., et al.: The epidemiology of AIDS in Asian and Pacific Islander populations in San Francisco. *AIDS* 2: 473-475 (1988).