

---

# Racial and Ethnic Categories in Biomedical Research: There is no Baby in the Bathwater

*Mildred K. Cho*

There are deep divides over the use of racial and ethnic categories in biomedical research and its application in both medical and non-medical contexts. On one side of a roughly described dividing line are practitioners who need to use every piece of information at their disposal to solve pressing, real-world problems in real time, such as making clinical diagnoses or identifying perpetrators of crime. On the other side are scientists and policy makers committed to meeting a scientific and social need for accuracy and thus trying to avoid miscategorization.

As Jay Cohn describes in this issue, medical practitioners in particular have used racial and ethnic categories to “enhance diagnostic and therapeutic precision.”<sup>1</sup> He argues for retaining this practice. The plea, motivated by genuine concern for patients, is to avoid “throwing the baby out with the bathwater.” However, Cohn and others mischaracterize the nature of the debate. The argument is not about whether differences among populations exist, or even whether differences among “races” exist. There are clearly phenotypic and physiological differences within the human population, and some of these roughly track socially-defined groupings that we call “race” or “ethnicity.” However, there is no single accepted set of racial and ethnic categories; we cannot even clearly define what “race” or “ethnicity” means. Second, while some clinically significant differences map onto racial or ethnic boundaries (however defined), many do not.<sup>2</sup> This means that basing diagnostic and therapeutic decisions in part on perceived race or ethnicity will be imprecise. Further, this mapping will ultimately rely largely on visual identification, which is notoriously unreliable.<sup>3</sup> Finally, reliance on racial and ethnic categories distracts from information that might actually be more relevant to research, diagnosis, or therapy, such as environmental factors or finer-grained differences in ancestral origins than the crude grouping of “race.”

The point of current research to characterize human genetic, environmental, and phenotypic variation is to bring precision to genomic analysis, and to help us understand when environmental, rather than genomic variation, is the major determinant of disease. The problem with using race or ethnicity as a measure is that it is really used as a proxy for an as-yet undetermined mix of genetic, biological, and environmental factors. While this may be perceived as “good enough” for use in daily clinical practice, it reinforces inaccurate perceptions about “racial” and “ethnic” groups.

---

**Mildred K. Cho, Ph.D.**, is Associate Professor of Pediatrics at Stanford University and Associate Director of the Stanford Center for Biomedical Ethics. Her background is in biological sciences and health policy.

Race is real, but as noted by Troy Duster and others,<sup>4</sup> it is often not a measure of an individual, but an interactive measure of a perception of an individual by another. As such, it can be a useful measure in certain studies such as research on health disparities, where the effects of perceived race on social interactions and health are specifically of interest. In this situation, the “race” variable is appropriately treated as a combination of biological and social factors.

**The point of current research to characterize human genetic, environmental, and phenotypic variation is to bring precision to genomic analysis, and to help us understand when environmental, rather than genomic variation, is the major determinant of disease.**

There is a long tradition in medicine and other professions of using racial groups to categorize individuals, and these categories are deeply embedded in the legal and social systems of the United States as well as the national psyche. However, the history of our classification scheme argues against, rather than for, continuing this tradition.<sup>5</sup> Scientists and clinicians do not intend to imply hierarchy when they use racial classifications, but it is naïve to think that hierarchy can be surgically removed from the concept of race. Hierarchy was an integral part of the concept as originally defined.

When introduced into the scientific taxonomy by Linnaeus in 1758, there were four racial groups.<sup>6</sup> Since then, the scientific literature has documented as many as thirty-four different races.<sup>7</sup> More recent literature describing groupings of human populations based on genetic analysis has suggested the existence of anywhere from two to six groups.<sup>8</sup> However, even Darwin questioned the distinctiveness of racial groupings in the human species: “It may be doubted whether any character can be named which is distinctive of a race and is constant.”<sup>9</sup> This doubt is echoed in modern times by geneticists: “Thus, populations are never pure in a genetic sense, and definite boundaries between individuals or populations (e.g., “races”) will necessarily be somewhat inaccurate and arbitrary.”<sup>10</sup>

As described by Keita, Rotimi, and many others,<sup>11</sup> “race” is used in human populations in a way that it is not used in other species. Definitions of race in other species include: “An interbreeding, usually geographically isolated population of organisms differing from other populations of the same species in the frequency of hereditary traits. A race that has been given formal taxonomic recognition is known as a subspecies.”<sup>12</sup> Another definition is a “group of organisms (all of the

same species) that is genetically self-sustaining and isolated geographically or temporally during reproduction.”<sup>13</sup> Neither of these definitions applies to our use of the term “race” to describe human populations today. As Rotimi has noted, isolated groups such as Old Order Amish might be considered genetically isolated enough to be considered a “race” by this type of definition,<sup>14</sup> but this clearly does not square with typically-used racial categories in the United States.<sup>15</sup>

This conceptual mismatch illustrates the most serious problem with the use of racial and ethnic categories in biomedical research. Even if “race” and “ethnicity” could be described consistently by genetic or biological measures (and they cannot), categorization of individuals in the clinic relies on

classification by visually-identified characteristics, self-definition, or a combination of the two.<sup>16</sup> Even though we may feel confident of our visual perceptions and racial or ethnic conclusions, we know that this kind of classification is dismally inaccurate. Further, self-definition may capture an individual’s identification with a particular social group, rather than biological ancestry or anything genomic.

For example, in a study comparing the racial classification on birth and death certificates of infants in the United States who died within a year of birth, “inconsistency in the coding of race is low for whites (1.2%), greater for blacks (4.3%), and greatest for races other than white or black (43.2%).”<sup>17</sup> In another example, a study comparing the “true” ethnic classification (by visual identification by police officers) with the predicted classification (by analysis of short tandem repeat, or STR loci that are used for identification of individuals because they vary highly from person to person) of individuals in the United Kingdom, no group was classified the same way by both methods more than sixty-seven percent of the time, and those genetically assessed to be in the “Middle-Eastern” category were perceived that way visually only thirty percent of the time.<sup>18</sup>

Relying on self-reporting or replacing race with ethnicity does not increase precision. “[I]n one American study where people had to assign themselves to an ethnic group in two consecutive years, one third of the population chose a different ethnic group on the second occasion.”<sup>19</sup> As noted above, self-identification may have more to do with social identification and affiliation than genomics.

One of the most insidious consequences of clinging to racial and ethnic classifications is the assumption

that if a difference between “racial groups” is found, (a) race is the best way to categorize this difference, and (b) race is the most relevant factor contributing to this difference. For example, pharmacogenetic differences among groups that differ by “race” are often cited as a reason to retain racial categorization. The authors of one study state that “5-10% of Europeans, but only ~1% of Japanese, have loss-of-function variants at [the CYP2D6 locus] that affect the metabolism of more than 40 drugs.”<sup>20</sup> However, the authors also note that “[t]he CYP2D6 ultra-rapid metabolizer alleles also vary in frequency, even within Europe, from ~10% in Northern Spain to 1-2% in Sweden,” suggesting that differences within a “racial group” can be just as great and just as relevant as differences between such groups.

Even if genetic differences among populations might be associated with differences in, for example, drug reactions, the genes implicated in these difference might not be the most important factors driving the phenotype. Racial differences in thiopurine methyl transferase genotype<sup>21</sup> and adverse reactions to chemotherapy are an example of a clinical use for racial classification.<sup>22</sup> However,

[a]n analysis of six clinical studies correlating adverse thiopurine effects and TPMT genotype revealed that an average of seventy-eight percent of adverse drug reactions were not associated with TPMT polymorphisms. Pharmacogenetic testing will thus not eliminate the need for careful clinical monitoring of adverse drug reactions.<sup>23</sup>

All of this means that there is no “baby in the bathwater,” no clinical or scientific utility to racial and ethnic categories unless one is studying perceived race or ethnicity or self-perception. There are certainly clinically significant differences among individuals and among groups. However, what defines these individuals and groups is not what we call “race” or “ethnicity” because there is no consistent definition of racial or ethnic categories. Because social perceptions of the meaning of race and ethnicity are extremely fluid, basing research findings on these categories or applying scientific findings based on perceived race or ethnicity is fraught with problems. Thus, attempts to “better define [the racial and ethnic] structure [of drug response]” will be futile.<sup>24</sup>

## References

1. J. Cohn, “The Use of Race and Ethnicity in Medicine: Letters from the African-American Heart Failure Trial,” *Journal of Law, Medicine & Ethics* 34, no. 3 (2006): 552-554.
2. P. Sankar and M. Cho, “Toward a New Vocabulary of Human Genetic Variation,” *Science* 298 (2002): 1337-38.
3. M. Cho and P. Sankar, “Forensic Genetics: ELSI Beyond the Clinic,” *Nature Genetics* 36 (2004): S8-S12.
4. P. Ossorio and T. Duster, “Controversies in Biomedical, Behavioral, and Forensic Sciences,” *American Psychologist* 60 (2005): 115-128; T. Duster, “Lessons From History: Why Race and Ethnicity Have Played a Major Role in Biomedical Research,” *Journal of Law, Medicine & Ethics* 34, no. 3 (2006): 487-496.
5. S. Lee, J. Mountain, and B. Koenig, “The Meanings of ‘Race’ in the New Genomics: Implications for Health Disparities Research,” *Yale Journal of Health Policy, Law and Ethics* 1 (2001): 33-75.
6. C. Linnaeus, *Systema Naturae* (Stockholm: Salvius, 1758): at 20-33.
7. J. Deniker, *The Races of Man: An Outline of Anthropology and Ethnography* (Paris: Charles Scribner, 1900): at 280.
8. L. Jorde and S. Wooding, “Genetic Variation, Classification and ‘Race,’” *Nature Genetics* 36 (2004): S28-S33; N. Risch, “Categorization of Humans in Biomedical Research: Genes, Race, and Disease,” *Genome Biology* 3 (2003): 2007.1-12; N. Rosenberg, et al., “Genetic Structure of Human Populations,” *Science* 298 (2002): 2381-85.
9. C. Darwin, *The Descent of Man, and Selection in Relation to Sex* (London: John Murray, 1871): Chapter 7.
10. L. Jorde and S. Wooding, “Genetic Variation, Classification and ‘Race,’” *Nature Genetics* 36 (2004): S28-S33.
11. S. Keita and A. Boyce, “‘Race’: Confusion about Zoological and Social Taxonomies, and Their Places in Science,” *American Journal of Human Biology* 13 (2001): 569-75.; S. Keita, et al., “Conceptualizing Human Variation,” *Nature Genetics* 36 (2004): S17-S20.
12. *American Heritage Dictionary of the English Language*, 4th ed. (Boston: Houghton Mifflin Co., 2003).
13. See <<http://dictionary.reference.com>> (2005).
14. C. Rotimi, article of file with author.
15. Executive Office of the President, Office of Management and Budget, *Standards for the Classification of Federal Data on Race and Ethnicity* (Washington, D.C.: Government Printing Office, 1995).
16. American Anthropological Association, *Response to OMB Directive 15: Race and Ethnic Standards for Federal Statistics and Administrative Reporting* (Arlington, VA: 1997).
17. R. Hahn, J. Mulinare, and S. Teutsch, “Inconsistencies in Coding of Race and Ethnicity Between Birth and Death in US Infants: A New Look at Infant Mortality, 1983 through 1985,” *JAMA* 267 (1992): 3151-52.
18. A. Lowe, A. Urquhart, L. Foreman, and I. Evett, “Inferring Ethnic Origin by Means of an STR Profile,” *Forensic Science International* 119 (2001): 17-22.
19. K. Leech, “A Question in Dispute: The Debate About an ‘Ethnic’ Question in the Census,” in *Runnymede Research Report* (London: Runnymede Trust, 1989).
20. J. Wilson, et al., “Population Genetic Structure of Variable Drug Response,” *Nature Genetics* 29 (2001): 265-69.
21. D. Otterness, et al., “Human Thiopurine Methyltransferase Pharmacogenetics: Gene Sequence Polymorphisms,” *Clinical Pharmacology & Therapeutics* 62 (1997): 60-73.
22. V. Burroughs, R. Maxey, and R. Levy, “Individualized Pharmaceutical Treatment,” *Journal of the National Medical Association* 94 (2002): 1-26.
23. J. van Aken, M. Schmedders, G. Feuerstein, and R. Kollek, “Prospects and Limits of Pharmacogenetics: The Thiopurine Methyl Transferase (TPMT) Experience,” *American Journal of Pharmacogenomics* 3 (2003): 149-55.
24. Cohn, *supra* note 1.