



“Race” and “ethnicity” in biomedical research: How do scientists construct and explain differences in health?☆

Catherine Lee*

Rutgers University, Sociology, 54 Joyce Kilmer Ave., Lucy Stone Hall A268, Piscataway, NJ 08854, USA

ARTICLE INFO

Article history:

Available online 29 January 2009

Keywords:

USA
Race
Ethnicity
Biomedical research
Genetics
Health disparities

ABSTRACT

Social and biomedical scientists, journal editors, and public health officials continue to debate the merits of the use of race and ethnicity in health-related research. As biomedical research focuses on issues of racial or ethnic health disparities, it remains unclear how biomedical scientists investigate race or ethnicity and health. This paper examines how biomedical researchers construct and analyze race or ethnicity in their studies and what conclusions they make about difference and health. Using content analysis of 204 biomedical research journal publications, which were supported by grants won from the National Cancer Institute of the National Institutes of Health in the USA, I demonstrate that although authors tended to see race or ethnicity as important and significant in their research, they rarely defined or operationalized the concepts adequately. Moreover, when presenting findings of racial or ethnic difference, authors generally did not provide explanations of the difference. I argue that this under-theorized and unspecified use of race or ethnicity and the biological conclusions drawn about health and difference have the potential to reify “race” and to limit our thinking about what these biomedical differences suggest about health disparities and inequalities in general.

© 2008 Elsevier Ltd. All rights reserved.

Introduction

Despite the repudiation of scientific racism, scientists remain divided over what race means or ought to mean in scientific investigations as well as in lay discussions and policymaking. The concept may represent social, biological, and even genetic differences. Such variance is antithetical to the tenets of scientific research, which, in its ideal form, demands that analytical variables be consistent and their categories mutually exclusive (Bowker & Star, 1999; Timmermans & Berg, 2003). Due in part to this concern over research methods and perhaps apprehension over the biological reduction of race, biomedical researchers, social scientists, journal editors, and public health officials have all weighed in on the matter of what, if anything, race ought to mean in scientific investigations, the data collections that enable such research, the reporting of findings, and resulting policies that are informed by these discoveries (British Medical Journal, 1996; Cooper, Kaufman,

& Ward, 2003; Fullilove, 1998; Kaplan & Bennett, 2003; Phimister, 2003; Schwartz, 2001).

It is unclear what researchers *do* and *mean* when they use race or ethnicity in their investigations. The studies that have examined the use of these constructs in research have shed light on how scientists define and use them, but their scope has been limited to health services research (Drevdahl, Taylor, & Phillips, 2001; Williams, 1994) and genetics (Race, Ethnicity, and Genetics Working Group, 2005; Sankar, Cho, & Mountain, 2007). This study expands this inquiry by exploring how biomedical researchers conceptualize and incorporate race or ethnicity in their investigations. The issue of if and how race or ethnicity are used in biomedical research is particularly important now given the rise in attention to health disparities as a political issue and emphasis on biomedical solutions for such matters (Halfmann, Rude, & Ebert, 2005). In this paper, I examine how biomedical researchers, funded by the National Cancer Institute (NCI) of the National Institutes of Health (NIH), construct and use the concepts of race or ethnicity as analytical variables and what kinds of conclusions they make about health and difference. The study investigated 204 biomedical research journal articles, which were supported by grants won from the NCI in the years between 1990 and 1999—a watershed decade that ushered in numerous policy changes tied to health and biomedical research. Using content analysis of the publications, I demonstrate a number of findings. Authors tended to see race or

☆ The author thanks Rod Hayward, Joanna Kempner, Scott Murry, Bhaven Sampat, Kristen Springer, and the anonymous reviewers for their comments; Meghan Cameron and Feng Pan for their assistance; and the Robert Wood Johnson Scholars in Health Policy Research Program for its support.

* Tel.: +1 732 445 4240.

E-mail address: clee@sociology.rutgers.edu

ethnicity as important and significant in their research. However, despite seeing the importance of race or ethnicity in their research, authors rarely defined or operationalized the concepts adequately. Moreover, when presenting findings of racial or ethnic difference, authors generally did not provide explanations of the difference. I argue that this under-theorized and unspecified use of race or ethnicity and the biological conclusions drawn about health and difference have the potential to reify “race” and to limit our thinking about what these biomedical differences suggest about health disparities and inequalities in general.

Race and ethnicity in biomedical research

Before going further, I first provide an overview of the debates and issues surrounding the use of race and ethnicity in biomedical research and historical background of related government regulations. In this section and throughout the paper, I adopt a social constructionist approach to understanding the contestations and uses of race or ethnicity in biomedical research. I recognize that as a social construct its meanings are neither fixed nor essential. Despite this constructivist approach, I do not use quotes around the word “race” for stylistic simplicity (Jacobson, 1998, p. ix). The meaning, significance, and use of the concept of race are historically variable and contingent upon a host of economic, political, social, and cultural practices (Marx, 1998; Omi & Winant, 1994). The social constructionist approach to race dominates the social sciences, and social scientists refute the existence of any “natural” ordering of people discernable by physical characteristics or ancestral origins. Instead, this approach to race recognizes that modern classification systems such as the one dominant in contemporary U.S. are borne out of socio-political processes. This, however, does not mean that race is somehow inconsequential. The lived experience of a racialized identity is not only real but also potentially devastating. The effect of racism on health, for example, has been well documented (Karlsen & Nazroo, 2002; Krieger, 2000; Williams, Neighbors, & Jackson, 2003).

Ethnicity is defined as a collective with putative common ancestry that shares cultural symbols and practices, including language, diet, religion, values, and norms (Cornell & Hartmann, 1998; Schermerhorn, 1978). Many social scientists claim that ethnicity is group-defined and voluntaristic. They argue that an ethnic group is self-consciously ethnic, and one’s ethnicity is an *achieved* status while race is imposed from without and is an ascribed status (Cornell & Hartmann, 1998, p. 19). Citing this difference, both social and biomedical researchers have advocated the use of ethnicity over race as I explain below (Crews & Bindon, 1991; Loveman, 1999). However, a strictly achieved versus ascribed distinction may suggest falsely that there is greater analytical difference between the two than there really is. Furthermore, research on ethnic identity suggests that non-members may categorically proscribe ethnicity (Brubaker, 2004). While racial construction may have a different history than ethnic construction, both concepts are dynamic with fluid boundaries (Wade, 1997).

In biomedical research, scientists fall into two main camps: one that refutes a biological basis for race and another that sees race as being *potentially* biologically meaningful. For researchers who see no biological underpinning to race, race stands as a proxy for socio-cultural, economic, and particular historical processes and experiences. It is used to capture behavioral and structural differences between racialized groups. Advocates of this position recognize that there is genetic variance amongst humans, but they insist that these variations do not overlap with contemporary notions of race nor do they necessarily overlap with modern racial categories used in places like the United States. These scientists insist that while the experience of a racialized life may affect health outcomes, the

concept of race itself has no biological or genetic basis (Bhopal & Donaldson, 1998; Braun, 2002; Cooper et al., 2003; Graves, 2001; Krieger, 1996, 2000; Leslie, 1990; Schwartz, 2001; Williams, 2002; Witzig, 1996).

Scientists who argue that there is a biological meaning to race disagree over the significance and precision of this claim. These scientists’ evidence and arguments are increasingly rooted in our expanded understanding of and, for many, hope in a genomic future that can generate “personalized medicine” (Phimister, 2003, p. 1082; Shriver & Kittles, 2004). Not surprisingly, many of these scientists are geneticists. In a number of influential papers, population geneticists have argued that there are genetic variations that overlap with ancestral or continental origin (Bamshad & Olson, 2003; Bamshad et al., 2003; Risch, Burchard, Ziv, & Tang, 2002; Rosenberg et al., 2002; Shriver et al., 2004). Thus, some geneticists and biomedical researchers argue that modern understandings of race and self-identification of racial categories by subjects are good approximations of ancestral origin. Furthermore, since genetic variations identified in different ancestral origin groups probably yield dissimilar responses and outcomes, geneticists reason that data on race should be collected and that it ought to be used as a research variable (Burchard et al., 2003; Ioannidis, Ntzani, & Trikalinos, 2004).

Scientists who accept some biological meaning of race but question its use in research argue that race as it is identified, categorized, and used in the contemporary United States is a bad proxy for continental or ancestral origin. While still suggesting that there may be populations that can be genetically identified as being separate and unique from one another to pursue pharmacogenomics, they state that race should not be used as a proxy for population genetic variation, especially if the goal is to determine the specifics of differences in disease outcomes or pharmacokinetics (Jorde & Wooding, 2004; Rotimi, 2004; Tishkoff & Kidd, 2004).

Responding to this debate, as noted above, some scientists and journal editors have advocated the use of ethnicity over race, believing the former to be free of the problematic history of scientific racism and biological reductionism (British Medical Journal Editorial, 1996; Cooper, 1994; Crews & Bindon, 1991; JAMA Editorial, 2005; Schwartz, 2001). They argue that ethnicity as a concept more fully or accurately captures the environmental, cultural, behavioral, or socio-political experiences that patients or clinical test subjects face, which may affect their disease etiology or responses to therapeutic interventions. However, research can be conducted and conclusions reached that reify and essentialize the concept of ethnicity as equally as race. Consider research related to BRCA1 and 2 gene mutations and breast cancer amongst Ashkenazi Jewish women. Should its prevalence amongst Ashkenazi Jewish women be seen as an ethnic variation? How should carriers of the mutations who are Hispanic or Asian American be identified? It is unclear whether ethnicity or race is the appropriate categorical grouping for analysis (see discussion below and John et al., 2007), and scientists and journal editors cannot agree which is more appropriate for research.

In part, the confusion and uncertainty about terms and definitions arise from historical debates regarding racial classification in health data. Since the early 20th century, modern government public health agencies and organizations have collected data on health indicators by gender and race and ethnicity (Krieger & Fee, 1994). Health officials have seen and treated these categories as static, self-evident, and easily recordable. The continued collection of such data has helped to make the categories themselves relevant for health. Furthermore, the history and development of the women’s movement and the Civil Rights Movement have helped to politicize these categories, making them more significant for ordering political action and understandings of difference. As Steve Epstein argues, “everyday political relevance of gender and racial

identification in the US only increases the likelihood that these categories will be emphasized in biomedical classification" (2004, p. 192). The veracity of this statement has been more fully realized as government policies to address health disparities have grown, leading to increased data collection, public health initiatives, and greater support for biomedical research.

The decade of the 1990s introduced many of these policy changes, beginning auspiciously with the start of the Human Genome Project. Other critical developments affecting biomedical research, especially on health disparities, were implemented throughout the decade. The NIH created the Office of Research on Minority Health in 1990. The year 1990 also marked the end of the decline in federally financed research and development (R&D) spending across all science and engineering fields that began in 1975. From 1990 through 2004, NIH's funding of academic R&D grew at an estimated annual rate of 6.4% per year (in 2000 dollars). The NIH's R&D budget between 1998 and 2004 doubled, and its share of federal funding of academic R&D rose from 52% in 1990 to 63% in 2004 (National Science Board, 2008, pp. 5–14).

Perhaps the most significant policy change affecting biomedical research has been the overhaul at the NIH regarding the inclusion of women and minorities in clinical research. Following Congress's passage of the NIH Revitalization Act of 1993, the NIH issued an "inclusion mandate" in 1994, requiring all grant awardees of the NIH to include women and minorities in clinical research beginning with fiscal year 1995 (National Institutes of Health, 1994). Furthermore, the NIH instructs its grantees to use the racial and ethnic categories outlined in the Office of Management and Budget (OMB) Directive 15, which is the official schema used in most government record-keeping and data collection related to race and ethnicity (National Institutes of Health, 2001). The inclusion of women and minorities and the implementation of the official racial categories were and continue to be contentious. As Epstein (2007) chronicles, scientists in particular objected to the intrusion of "affirmative action," "quotas," or "political correctness" into science.

Regardless of these criticisms, we do not have a clear picture of how race or ethnicity is investigated. In research related to health services or to genetics, a small number of studies have been conducted. Williams (1994) found that the concepts race and ethnicity were widely used in the journal *Health Services Research*. He also found that the terms were seldom defined, although they were employed to stratify or adjust results. In a study of race and ethnicity as variables in the journal *Nursing Research*, Drevdahl et al. (2001) found similar results. While researchers often incorporated race and ethnicity in their investigations, they rarely defined what they meant in using such constructs.

More recently, there have been examinations of genetics research and the use of race, ethnicity, ancestry, or other population terms. The Race, Ethnicity, and Genetics Working Group of the National Human Genome Research Institute (2005) identified the diverse ways that genetics researchers convey group differences, including the use of race or ethnicity. This may reflect partly the myriad of recent calls by genetics and biomedical journal editors to carefully define population terms (Nature Genetics Editorial, 2000, 2001). To evaluate the extent to which authors have responded to such requests, Sankar et al. (2007) analyzed reports of genetics research. They found that just 9.1% of articles from 2001 and 2004 explained the basis for using race or ethnicity terms, and no article defined these concepts. This paper builds on these previous studies but examines biomedicine more generally.

Data and methods

The study examined how race and/or ethnicity are defined, operationalized, and utilized in publications that have been

supported by research grants from the National Cancer Institute in the USA. It did not look at the impact of any given institutional shift on the use of race or ethnicity in research and cannot provide a causal explanation for why the use of race or ethnicity takes the forms that they do. Nevertheless, I provide tentative results that suggest some possible influence that changes in research and journal editorial policies may be exerting on publications.

I focused on biomedical research funded by the NCI, which is the largest and oldest institute of the 27 institutes and centers that make up the NIH. For fiscal year 2007, the NCI awarded 6376 research grants totaling over \$3 billion (NIH Office of Extramural Research, 2008). Each institute and center has a unique character and varying history of supporting research on race or ethnicity. They also have varying levels of adherence to the NIH's inclusion mandate (Epstein, 2007, p. 159). By focusing on one institute, I was able to "control," in some sense, the heterogeneity found at the NIH. An in-depth study of one institute can generate a more nuanced content analysis of the resulting journal publications. NCI is not an institute whose main mission is to address racial health disparities, and researchers have identified both racial and non-racial etiology for different types of cancer, the institute's disease focus. I concluded that the NCI would provide an important, albeit perhaps not representative, snapshot of the NIH.

Using the search engine CRISP (Computer Retrieval of Information on Scientific Projects) I selected all new research projects (identified by the NIH as "investigator initiated, basic scientific support to principal investigator in his/her area of competency") awarded from 1990 through 1999 that supported biomedical research and had the following key terms in the grants: race (including racial and racism), ethnicity (including ethnic), minority, black, African American, white, Caucasian, Native American, Mexican American, Hispanic, Latino, or Asian. Because I investigate the production of knowledge, I excluded grants that did not result in a published article. A unique dataset that links NIH grant numbers to publications identification numbers (PMIDs) allowed me to determine which grants were appropriate. Using the above criteria, I generated over 200 grant abstracts. I excluded grants that did not deal with biomedical research, broadly defined to include any biological studies of pathways to disease, disease outcomes, and treatment. This resulted in 89 grants. I eliminated grants that were U01 (cooperative agreements between the government and investigators) and P01 (research program projects or center grants). Sixty-six grants were R01s (research project), 20 were R03s (small research grants), two were R29s (first independent research support and transition—FIRST—award), and one was an R35 (outstanding investigator grant). Grants provided support for an average of 3.6 years and covered various cancers. Breast and prostate cancer studies received the most number of grants; 20 grants were awarded for the study of breast cancer, and 18 were awarded for prostate cancer research. Thirty-six grants identified some component of genetic research as one of its focus. Forty-six proposed to study some aspect of racial or ethnic difference in disease.

For each grant award, I selected the first three published articles (determined by publication date) tied to the grant. Not all grants had three articles while some grants had dozens. All 89 grants had at least one article related to the grant. Sixty-six had two or more published article related to the grant, and 53 grants had three related publications. There were four articles tied to more than one grant. Eliminating the duplicate articles reduced the total number of unique publications to 204. These articles were published in a range of biomedical and scientific journals. The most popular journal was *Cancer Epidemiology, Biomarkers & Prevention*, which published 34 of the articles in the sample. *The American Journal of Epidemiology* was the site of 18 publications, and *Cancer Research*

included 15. Ten articles were found in *Journal of the National Cancer Institute* while nine were published in the *International Journal of Cancer*. The remaining articles were published in 75 different journals.

All articles were read by at least two coders and tested for coder reliability. Articles were coded for information on use of prior knowledge of racial or ethnic difference, data and methods, definition and use of race or ethnicity, presentation of racial or ethnic demographic and results data, and discussion of racial or ethnic differences. I conducted quantitative and qualitative content analyses of the 204 publications.

Racial or ethnic differences are important

Given the initial search criteria of the grants, it is not surprising that most of the articles mentioned race or ethnicity in some form. Authors referred to “race,” “ethnicity,” or other euphemistic terms in 166 or 81% of the 204 articles, which included studies of both multi-ethnic or multi-racial and single-group samples. In 47 or 23% of the articles, authors used the term “race” only. In 55 or 27% of the articles, authors used the term “ethnicity” but not “race.” Researchers used both “race” and “ethnicity” in 38 or 19% of the articles. In 26 or 13% of the articles, authors chose other terms such as “minority,” “birthplace,” or reference to a specific ethnic group (Fig. 1).

Of the 204 publications, 87 articles were supported by an NCI grant from the pre-inclusion mandate era. Sixty-six or 76% of these articles mentioned race or ethnicity in some form. Similarly, of the 117 articles that arose from NCI grants awarded between 1995 and 1999, 100 or 85% of the articles mentioned race or ethnicity. Again, given the inclusion criteria of the grants for the study, the high mention of these terms in resulting publications is not unexpected. These numbers cannot demonstrate the impact of the inclusion mandate or other NIH policy changes; the data are unable to state what a statistically significant trend is. Nevertheless, it does show that the reference to race or ethnicity is not new to the post-inclusion mandate period.

Authors expressed the sentiment that examining racial or ethnic differences is important in a number of ways. Some researchers pointed to a large or multi-ethnic study population, which could enable statistical analyses with race as a variable as demonstration of its robustness. For example, in the grant abstract for a study of cancer and diet, Laurence Kolonel (1993) stated, “Strengths of this project include its prospective design, large size, ethnic diversity, [and] minority component.” Another investigator mentioned

Kolonel’s work and suggested the validity of her own by stating, “The multiethnic, multicultural nature of the present population provided a wide range of iodine exposure, similar to the equally diverse group previously studied by Kolonel et al. in Hawaii” (Horn-Ross et al., 2001, p. 983). A diverse population provided opportunities for analyses and comparisons across racialized groups. Also noting an advantage of their study in relation to diversity were researchers Wu, Wan, and Bernstein, who wrote, “Our study represents one of the only three large population-based epidemiologic studies that has been designed specifically to investigate further the etiologies of these [stomach and esophageal] tumor types. Our study also allows comparison of risk estimates in Whites and non-Whites” (2001, p. 731). Researchers clearly indicated that race or ethnicity mattered and having the ability to evaluate related differences contributed to a study’s increased merit.

Even in single-group sample studies, authors mentioned the importance of race or ethnicity. While a diverse sample perhaps may address between-group questions, single-group sample studies can enable in-depth analyses, which may offer clues to a disease’s ethnic or racial variability. For example, researchers noted the higher rates of pre-invasive cervical lesion amongst Alaskan women despite their steady decline in rates for cervical cancer (MacLehose et al., 1999). Very broadly conceived, there were 15 articles that focused on a single ethnic or racial group, ranging from studies of Ashkenazi Jewish women to non-Hispanic whites. In some cases, the studies were conducted to identify a group-specific concern, and in others, the single-group samples were the result of convenience sampling or limited data access. Despite the lack of racial or ethnic heterogeneity in their studies, authors still invoked race or ethnicity, made both implicit and explicit between-group comparisons, and underscored the importance of these constructs in their work.

Most commonly, authors denoted the significance of race or ethnicity by employing four procedures. One, some authors cited a prior understanding of racial or ethnic difference. In 52 or 31% of the 166 articles that mentioned race or ethnicity, investigators cited some prior findings of racial or ethnic difference—in risk or rate of morbidity or mortality or in genetic variation. Second, without necessarily mentioning this prior knowledge, many authors also used race or ethnicity in their analyses.¹ Of the 166 articles, 100 or 60% of them used race or ethnicity as an analytical variable. Researchers reported that their studies “adjusted for race” in their statistical analyses even in studies in which racial or ethnic variation was not a main focus of the investigation. In her study of epidemiologists, Shim writes that scientists have a “habitual mention of including and controlling for race” in their research. While there may have been rationalizations previously, the practice is now so common that “controlling” or “adjusting for race” is “standard operating procedure” (Shim, 2005, p. 413).

Some authors presented racial or ethnic information visually to emphasize the significance of race or ethnicity in their studies. They employed a third practice, offering demographic data in tabular form in 61 or 37% of the 166 articles that mentioned race or

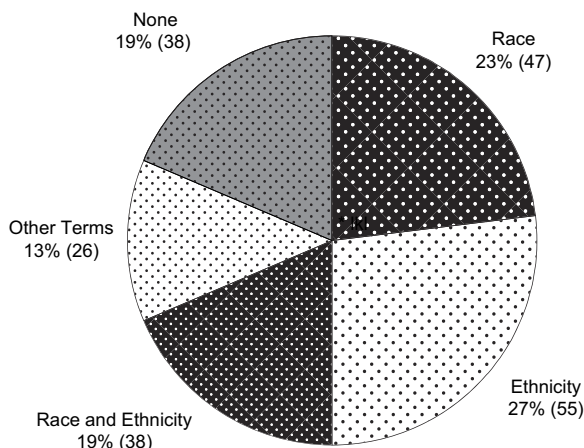


Fig. 1. Use of “race” and “ethnicity” in articles ($N = 204$). Categories total more than 100% due to rounding.

¹ As explained above, authors of studies with single-group samples invoked race or ethnicity and used many of the methods employed by researchers in multi-ethnic or multi-racial samples to convey the significance of race. For example, 10 out of 15 articles provided discussions of prior racial or ethnic knowledge, and four out of 15 articles provided racial or ethnic demographic data in tabular form. Nevertheless, authors of single-group sample studies may be less likely to use race or ethnicity as an analytic variable or to present such results in tabular form since their data are not conducive to between-group analyses. Thus, I recalculated the percentages with these 15 articles removed. Of the remaining 151 articles, 99 or 66% used race or ethnicity as an analytic variable, and 58 or 38% of the articles presented racial or ethnic results in tabular form. These 15 articles are included in the percentages presented in Fig. 2 for ease of comparison.

ethnicity. Fourth, researchers also offered racial or ethnic differences in outcomes in tabular form in 58 or 35% of the 166 articles. While 44 or 27% of the 166 articles did not employ any of these procedures, most authors incorporated one or more. Forty-three or 26% used one; 32 or 19% used two; 32 or 19% used three; and 17 or 10% of the 166 articles that mentioned race or ethnicity included all four practices (Fig. 2).

At times authors seemed to over-emphasize the significance of race or ethnicity in their research. For example, in studies in which racial or ethnic variation was not found, some authors continued to suggest that there were differences. In a study of gene polymorphism and predictability of clinical response in colorectal cancer patients, Park et al. (2002) explained, “We could not show significant association between clinical response and ethnicity.” The authors had expected to find ethnic variation in the clinical response and were therefore surprised with their null finding. They eventually concluded, “We hypothesize that the small sample size precluded us from detecting this potential association, and that larger studies are needed to answer this question” (Park et al., 2002, p. 48). The authors suggested that the null finding should not be accepted.

In a study of prostate cancer, the authors argued that variations that they had expected did not materialize in their data, because it lacked statistical power. They wrote, “Our study did not have the power to detect small but meaningful differences between subgroups of our population. [Biomarkers] did not differ significantly by disease status or by race ... Our study had 90% power to find a relatively large (0.35) difference between correlations, but smaller differences in correlation between racial groups or between cases and controls were much less likely to be detected” (Godley et al. 1996, p. 117). These authors indicated that even small differences by race, had they been detected, would have been meaningfully significant. Both the *expectation* and actual findings of difference were important.

Defining and using “race” and “ethnicity”

Despite the frequent invocation of the terms “race” and/or “ethnicity,” authors rarely offered a definition or operationalization of the concepts. There was an *a priori* assumption that readers would simply recognize or understand what the terms meant. In just 39 or 23% of the 166 articles that mentioned race or ethnicity,

authors provided some sort of explanation of race or ethnicity. No author explicitly articulated a definition of race—that is, the investigators did not explain if they conceived and utilized race as a biological construct, a socio-political identity, or social proxy for cultural and behavioral practices. In 35 of the 39 cases in which authors provided some sort of definition, they explained that race or ethnicity was “self-reported.” Authors defined ethnicity as birthplace (of father or grandfather) in three of the articles, and the remaining article that provided a definition referred to “ancestry”.

Recognizing the problem associated with defining race or ethnicity as well as the growing contentious debates surrounding the use of race or ethnicity in research, editors of biomedical journals implemented new instructions for authors. Epstein considers this establishment a spillover effect from the inclusion mandate implementation and related changes at the NIH, which have increasingly focused on women and minority health (2007, p. 176). In 1997, the International Committee of Medical Journal Editors (ICMJE) issued the fifth edition of their “Uniform Requirements for Manuscripts Submitted to Biomedical Journals” (ICMJE, 1997). Most American journals adopted the policy and began instructing authors to “identify the age, sex, and other important characteristics” of subjects. The guideline also stated “the definition and relevance of race and ethnicity are ambiguous,” warning authors to be “particularly careful about using these categories” (ICMJE, 1997, p. 311). Thirty-nine articles were published between 1990 and 1996. Authors did not define race or ethnicity in 32 or 82% of these articles. In seven or 18% of the articles, these terms were defined. For publications that followed the fifth edition of the Uniform Requirements from 1997 to 2003, 127 articles mentioned race or ethnicity. In 95 or 75% of the articles, authors did not define race or ethnicity. They defined the terms in 32 or 25% of the articles. As with grant year and the NIH inclusion mandate, the data cannot provide statistically significant tests of the impact of the Uniform Requirements on authors’ use of race or ethnicity. As stated earlier, the paper does not seek to provide a causal explanation for how and why race or ethnicity is used in biomedical research. Nevertheless, this study has identified a pattern of continued, and perhaps increased, use over the 1990s when these crucial policy changes affecting biomedical research occurred.

Despite the calls for authors to clarify their use of race or ethnicity in their research, scientists generally did not offer very

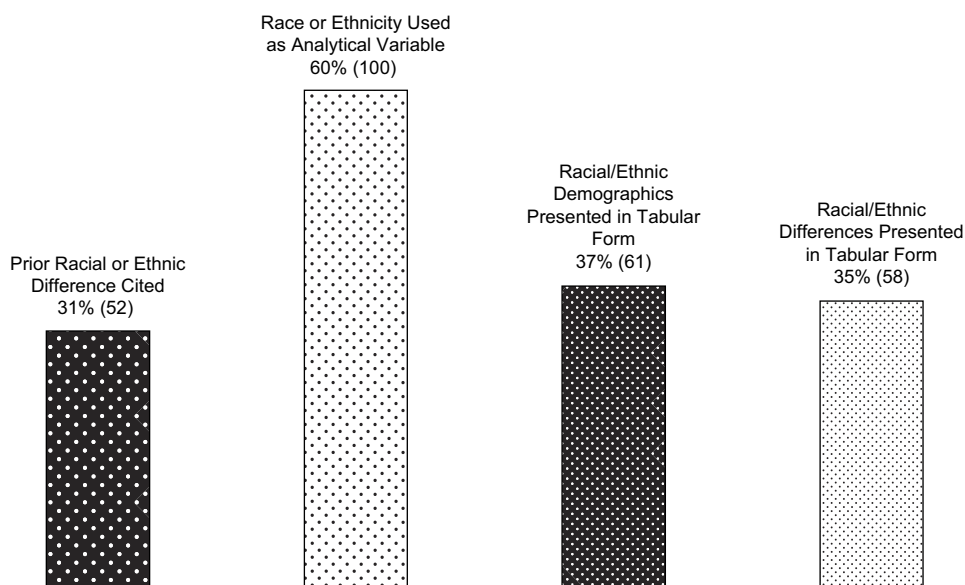


Fig. 2. Use and presentation of racial or ethnic data in articles mentioning race or ethnicity ($N = 166$). Categories are not mutually exclusive and do not total 100%.

detailed explications. The most detailed discussion of what the authors meant by race or ethnicity was provided in a study of smoking and pancreatic adenocarcinoma risk. The authors stated “Race was self-reported according to three broadly defined categories: Caucasian, African American, and Asian. Hispanic participants were classified as Caucasian, Asian, or African American depending on which of these racial categories was selected by the respondent” (Duell et al., 2002, p. 299). This is less of a definition and more of a reporting of the official racial and ethnic categories outlined in the government’s Office of Management and Budget (OMB) Directive 15. This effort to define race leaves a basic question unanswered. Though the reader can determine that race is something that respondents self-identified, whether this variable should be interpreted as a biological, social, or other category was not explained. It also raises other questions, such as how race (determined by “three broadly defined categories”) differs from ethnicity (assessed by the Hispanic question) in understanding pancreatic cancer. This example illustrates the important point that using ethnicity over race will not necessarily remove ambiguities, especially if researchers do not clarify what the social and biological mechanisms the concepts are supposed to capture.

Findings of difference are not explained

Whether or not definitions were offered, when authors used race or ethnicity as an analytical variable or presented racial or ethnic data, they rarely provided explanations of how or *why* race or ethnicity was important. Authors reported findings of racial or ethnic differences in 58 articles and similarities across racial or ethnic groups in 12 articles. Scientists did not always explain what could account for the racial or ethnic difference. In 40 or 69% of the 58 articles in which authors reported racial or ethnic differences, authors simply stated their findings but did not offer an explanation. In 19 or 33% of the 58 articles, authors tried to offer an explanation, referring to socio-economic, behavioral, environmental, genetic, biological, and/or other reasons.

These explanations were not elaborate, and authors offered little detail. Wu et al. (2001) provided one of the more elaborate discussions of racial or ethnic differences found in a study of prostate cancer. They wrote, “Reasons for the racial differences, in particular, the lower levels of 3 α -diol G and AG levels in Asian American men, are not known. Environmental influences, genetic control, and an interplay of genetic and environmental factors are likely explanations for these differences” (2001, p. 537). In an article on mortality after second primary breast carcinoma, Bernstein, Lapinski, Lynch, Holford, and Thompson (2002) also attempted to offer a more thorough explanation for the finding of racial difference. They cited other studies that found socio-economic reasoning for increased mortality for African American women after primary breast carcinoma and suggested this as a possible explanation for higher mortality in African American women after a second primary breast carcinoma. In both of these examples, the investigators did not research the social or environmental factors that may explain some, or even all, of the racial variation. While these investigators recognized that there are important feedback mechanisms between genetics and environment in relation to race (Duster, 2003; Williams, 1997), they did not test these complex interactions.

In nine or nearly half of the 19 articles that offered an explanation, authors relied on biological or genetic explanations (with no reference to social or environmental factors). A reliance on biological or genetic variation for explanation can help reduce racial health disparities to immutable facts and reify the concept of race. In a study of prostate cancer, researchers studied Afro-Caribbean men in Tobago and hypothesized that as descendents of West Africans, the Tobago men would have high rates of prostate cancer,

like African Americans. In discovering their hypothesis to be true, they then tried to theorize what could explain the high rates of prostate cancer amongst both groups. They wrote, “One of the known risk factors for prostate cancer is ethnicity, i.e. African descent, although we do not know how this risk is mediated. One hypothesis is that genetic factors contribute to the high risk for prostate cancer among populations of African origin. If the Caucasian admixture rate in the Tobago population is indeed low, then this population may carry a higher burden of high-risk genes of African descent than the more admixed populations in the United States ... [P]opulations of African descent ... share ancestral genetic factors that increase susceptibility to prostate cancer” (Bunker et al., 2002, p. 729). Though they referred to genetics as the main explanation, the researchers did not conduct genetic screenings of the study subjects and had no proof regarding levels of admixture.

A publication in which authors define what they mean by race or ethnicity, find differences across racial or ethnic groups, and offer an explanation for the variation can still run into a muddled area when its authors try to make conclusive arguments about health and difference. Despite some claims by geneticists and advocates of ancestral population research, the very act of doing biomedical research relies on folk taxonomies of race and ethnicity that do not neatly and clearly dissect groups into distinct, mutually exclusive categories for analysis. I illustrate this by presenting in detail a study by Modan et al. (1996) on the BRCA1 gene and ovarian cancer in Israel.

In the study, the investigators used place of birth as a proxy for Jewish subgroup or ethnicity. Ashkenazi Jews were defined as those born in Europe or America. Researchers did not identify the other groups by name, but they categorized the subjects according to birth in Israel or Asia and Africa. In assessing the relationship between a BRCA1 185delAG mutation and ovarian cancer, Modan et al. found that “With 1 exception all the mutation-positive cases were of European extraction (‘Ashkenazi’)” (1996, p. 1824). The authors tried to make an argument for the ethnic basis of ovarian cancer, at least in some women. Nevertheless, Modan et al. seemed to recognize how complicated such a finding would be for understanding health and difference and clinical practice even if the observed outcomes were true. They wrote, “The impact of the ability to detect mutations in ovarian cancer susceptibility genes on clinical decision making is still unclear” (1996, p. 1825). They insisted that limiting screening to Jewish Ashkenazi women would be premature and even questioned the term “Ashkenazi,” conceding it was “imprecise and immeasurable for historic and scientific reasons.” As a matter of practice, the investigators acknowledged that it may be problematic to use birthplace as a proxy for determining the at-risk population, especially given the mixed origin of a number of their own subjects (Modan et al., 1996, p. 1825).

This example illustrates an important theoretical point about the use of racial or ethnic classification in research. In biomedical research, racial or ethnic classification is based on prototype theory, one that is “fuzzier” than we realize or that scientists want to admit. Such classification is not rooted in binary, discrete categories (Bowker & Star, 1999, p. 62). Instead, when we use prototype theory, we have a “broad picture” in our minds about what we think we are classifying. Different social groups have different prototypes in mind. Thus, for Modan et al., ethnicity defined as birthplace made sense in a study of subjects in Israel, while this may not be appropriate elsewhere especially since the meaning of race or ethnicity is not only fluid but also spatially variable (Davis, 2001; Marx, 1998; Nobles, 2000).

Conclusion

The potentially fuzzy and imprecise nature of the use of race or ethnicity in biomedical research is not discipline or field specific.

Social scientists who investigate race or ethnicity can also be faulted for not defining or ill-defining these terms. They regularly make conclusions about racial or ethnic variations then fail to explicate the social mechanisms by which race or ethnicity is meaningful for determining social outcomes such as educational achievement or income attainment (Loveman, 1999). Nevertheless, the potentials for misuse of racial or ethnic constructs and harm lie more significantly in the realm of biomedical research given the increasing importance of biomedicine and genomics as well as the history of racism and science. This study examined research supported by the NCI, and as such, the findings may not be generalizable to all biomedical research. However, investigations sponsored by the largest institute of the biggest provider of funds for biomedical research in the world represent a significant portion of the field. Researchers in other fields may produce more nuanced analyses and interpretations of race or ethnicity and health. This is particularly true in public health where scholars have long advocated a more critical examination of the relations between race or ethnicity and health (Krieger, 1996; Williams, 2002). In biomedical research funded by the NCI, this paper has shown that scientists sensed the importance of race or ethnicity in biomedical research and thus used these concepts. However, they rarely defined them or articulated how race or ethnicity operated in their models. When racial or ethnic variation was found, most researchers did not provide an explanation for how and why such findings resulted or their medical significance.

Despite the controversies and at times limited data, biomedical scientists continue to study racial or ethnic differences and increasingly the genetic roots of such variations. Certainly, the completion of the Human Genome Project, critical changes at the NIH and other governmental organizations that support scientific research, and scientific journals' publication policies all play a role in encouraging this development. Leading scientists, both biomedical and social scientific, have offered remedies for addressing possible pitfalls (Bonham, Warshauer-Baker, & Collins, 2005). Some of these efforts, such as genetics journals' editorial guidelines requiring authors to define population terms, have yielded limited success (Nature Genetics Editorial, 2000, 2001; Sankar et al., 2007). Government regulations and funding arrangements may make the use of race or ethnicity in biomedical research easier without necessarily demanding scientists to be more critical. At heart, the findings of this study simply may illustrate the continuing debates and conflicts over the use of race and ethnicity in biomedical research.

Scientists' construction, utilization, and discussion of race or ethnicity highlight a number of obstacles we face in addressing health disparities and other inequalities. The finding that biomedical researchers do not provide conclusive results about racial or ethnic health outcomes presents crucial public policy challenges since Halfmann et al. (2005) have shown that biomedical initiatives are increasingly offered as the solution for addressing racial or ethnic disparities in health. This suggests that it is even more imperative that policymakers fully understand what biomedical researchers are investigating and discovering about racial or ethnic health differences. Biomedical research that purports to find racial or ethnic differences may get offered as evidence to support greater biomedical research, perhaps at the cost of important public health research. This biomedical emphasis may encourage drug companies to focus their research and development resources increasingly on pharmacogenomics and the development of ethnic-specific drugs (Lee, 2005). At the very least, public attention to supposed differences could generate interest and demand for "ethnic drugs." In turn, this may affect clinical practice and prompt more medical "racial profiling," which may not yield the best outcomes for patients (Satel, 2002, 2004). Health care providers may come to

accept poor health measures that exist for racialized minorities as they get framed in static biological or even genetic terms.

Most significantly, a growing faith in biomedicine and genomics and an uncritical acceptance of scientific studies of race or ethnicity may foster essentialized and biologically reductionist approaches to not only addressing health disparities but also other racial or ethnic inequalities. This biomedical and genetic focus may lead to biomedical solutions and the withdrawal of social, political, or economic approaches to easing social and economic inequalities. Furthermore, we may inadvertently accept the validity and legitimacy of a biological understanding of race (Duster, 2003, 2005). As we move toward an increasingly more medicalized and genetic understanding of our bodies and conditions, we may concede to medical authority and biomedical research the right to externally validate a biological construction of race that is actually rooted in socio-historical processes.

References

- Bamshad, M. J., & Olson, S. E. (2003). Does race exist? *Scientific American*, 289(6), 78–85.
- Bamshad, M. J., Wooding, S., Scott Watkins, W., Ostler, C. T., Batzer, M. A., & Jorde, L. B. (2003). Human population genetic structure and inference of group membership. *American Journal of Human Genetics*, 72(3), 578–589.
- Bernstein, J. L., Lapinski, R., Lynch, C., Holford, T., & Thompson, W. D. (2002). Factors influencing mortality among young women with second primary breast carcinoma. *Cancer*, 95(10), 2051–2058.
- Bhopal, R., & Donaldson, L. (1998). White, European, Western, Caucasian, or what? Inappropriate labeling in research on race, ethnicity, and health. *American Journal of Public Health*, 88(9), 1303–1312.
- Bonham, V. L., Warshauer-Baker, E., & Collins, F. S. (2005). Race and ethnicity in the genome era: the complexity of the constructs. *American Psychologist*, 60(1), 9–15.
- Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. Cambridge: MIT Press.
- Braun, L. (2002). Race, ethnicity, and health: can genetics explain disparities? *Perspectives in Biology and Medicine*, 45(2), 159–174.
- British Medical Journal Editorial. (1996). Style matters: describing race, ethnicity, and culture in medical research. *British Medical Journal*, 312(7038), 1054–1055.
- Brubaker, R. (2004). *Ethnicity without groups*. Cambridge, MA: Harvard University Press.
- Bunker, C. H., Patrick, A. L., Konety, B. R., Dhir, R., Brufsky, A. M., Vivas, C. A., et al. (2002). High prevalence of screening-detected prostate cancer among Afro-Caribbeans: the Tobago Prostate Cancer Survey. *Cancer Epidemiology, Biomarkers & Prevention*, 11(8), 726–729.
- Burchard, E. G., Ziv, E., Coyle, N., Gomez, S. L., Tang, H., Karter, A. J., et al. (2003). The importance of race and ethnic background in biomedical research and clinical practice. *New England Journal of Medicine*, 348(12), 1170–1175.
- Cooper, R. S. (1994). A case study in the use of race and ethnicity in public health surveillance. *Public Health Reports*, 109(1), 46–52.
- Cooper, R. S., Kaufman, J. S., & Ward, R. (2003). Race and genomics. *New England Journal of Medicine*, 348(12), 1166–1170.
- Cornell, S., & Hartmann, D. (1998). *Ethnicity and race: Making identities in a changing world*. Thousand Oaks, CA: Pine Forge Press.
- Crews, D. E., & Bindon, J. R. (1991). Ethnicity as a taxonomic tool in biomedical and biosocial research. *Ethnicity & Disease*, 1, 42–49.
- Davis, F. J. (2001 [1991]). *Who is black? One nation's definition*. Tenth Anniversary Edition. University Park: The Pennsylvania State University Press.
- Drevdahl, D., Taylor, J. Y., & Phillips, D. A. (2001). Race and ethnicity as variables in *Nursing Research*, 1952–2000. *Nursing Research*, 50(5), 305–313.
- Duell, E. J., Holly, E. A., Bracci, P. M., Liu, M., Wiencke, J. K., & Kelsey, K. T. (2002). A population-based, case-control study of polymorphisms in carcinogen-metabolizing genes, smoking, and pancreatic adenocarcinoma risk. *Journal of the National Cancer Institute*, 94(4), 297–306.
- Duster, T. (2003). Buried alive: the concept of race in science. In A. H. Goodman, D. Heath, & M. Susan Lindee (Eds.), *Genetic nature/culture: Anthropology and science beyond the two-culture divide*. Berkeley: University of California Press.
- Duster, T. (2005). Enhanced: race and reification in science. *Science*, 307(5712), 1050–1051.
- Epstein, S. (2004). Bodily differences and collective identities: the politics of gender and race in biomedical research in the United States. *Body and Society*, 10(2–3), 183–203.
- Epstein, S. (2007). *Inclusion: The politics of difference in medical research*. Chicago: University of Chicago Press.
- Fullilove, M. T. (1998). Comment: abandoning "race" as a variable in public health research—an idea whose time has come. *American Journal of Public Health*, 88(9), 1297–1298.

- Godley, P. A., Campbell, M. K., Miller, C., Gallagher, P., Martinson, F. E., Mohler, J. L., et al. (1996). Correlation between biomarkers of omega-3 fatty acid consumption and questionnaire data in African American and Caucasian United States males with and without prostatic carcinoma. *Cancer Epidemiology, Biomarkers & Prevention*, 5(2), 115–119.
- Graves, J. (2001). *The Emperor's new clothes: Biological theories of race at the millennium*. New Brunswick, NJ: Rutgers University Press.
- Halfmann, D., Rude, J., & Ebert, K. (2005). The biomedical legacy in minority health policy-making, 1975–2002. *Research in the Sociology of Health Care*, 23, 245–275.
- Horn-Ross, P. L., Morris, J. S., Lee, M., West, D. W., Whittemore, A. S., McDougall, I. R., et al. (2001). Iodine and thyroid cancer risk among women in a multiethnic population: the Bay Area Thyroid Cancer Study. *Cancer Epidemiology, Biomarkers & Prevention*, 10(9), 979–985.
- ICMJE. (1997). Uniform requirements for manuscripts submitted to biomedical journals. *New England Journal of Medicine*, 336(4), 309–315.
- Ioannidis, J. P. A., Ntzani, E. E., & Trikalinos, T. A. (2004). 'Racial' differences in genetic effects for complex diseases. *Nature Genetics*, 36(12), 1312–1318.
- Jacobson, M. F. (1998). *Whiteness of a different color: European immigrants and the alchemy of race*. Cambridge, MA: Harvard University Press.
- JAMA Editorial. (2005). Instructions for authors. *JAMA*, 294(1), 119–126.
- John, E. M., Miron, A., Gong, G., Phipps, A. I., Felberg, A., Li, F. P., et al. (2007). Prevalence of pathogenic BRCA1 mutation carriers in 5 US Racial/ethnic groups. *JAMA*, 298(24), 2869–2876.
- Jorde, L. B., & Wooding, S. P. (2004). Genetic variation, classification and 'race'. *Nature Genetics*, 36(11), S28–S33.
- Kaplan, J. B., & Bennett, T. (2003). Use of race and ethnicity in biomedical publication. *JAMA*, 289(20), 2709–2716.
- Karlsen, S., & Nazroo, J. Y. (2002). Relation between racial discrimination, social class, and health among ethnic minority groups. *American Journal of Public Health*, 92(4), 624–631.
- Kolonel, L. (1993). *Grant application: Multiethnic minority cohort study of diet and cancer*. National Cancer Institute. Grant Number 1R01CA054281-01A2. Acquired by author through Freedom of Information Act request to NCI.
- Krieger, N. (1996). Inequality, diversity, and health: thoughts on 'race/ethnicity' and 'gender'. *Journal of American Medical Women's Association*, 51(4), 133–134.
- Krieger, N. (2000). Discrimination and health. In L. Berkman, & I. Kawachi (Eds.), *Social epidemiology* (pp. 36–75). Oxford: Oxford University Press.
- Krieger, N., & Fee, E. (1994). Social class: the missing link in US health data. *International Journal of Health Services*, 24(1), 25–44.
- Lee, S.-J. (2005). Racializing drug design: implications of pharmacogenomics for health disparities. *American Journal of Public Health*, 95(12), 2133–2138.
- Leslie, C. (1990). Scientific racism: reflections on peer review, science, and ideology. *Social Science & Medicine*, 31(8), 891–905.
- Loveman, M. (1999). Is race essential? *American Sociological Review*, 64(6), 891–898.
- MacLehose, R. F., Harpster, A., Lanier, A. P., Schiff, M. A., Koehler, K., Murphy, N., et al. (1999). Risk factors for cervical dysplasia in Alaska native women: a pilot study. *Alaska Medicine*, 41(4), 76–85.
- Marx, A. (1998). *Making race and nation: A comparison of the United States, South Africa, and Brazil*. Cambridge: Cambridge University Press.
- Modan, B., Gak, E., Sade-Bruchim, R. B., Hirsh-Yechezkel, G., Theodor, L., Lubin, F., et al. (1996). High frequency of BRCA1 185delAG mutation in ovarian cancer in Israel. *JAMA*, 276(22), 1823–1825.
- National Science Board. (2008). *Science and engineering indicators 2008, Vol. 1*. Arlington, VA: National Science Foundation (NSB 08-01).
- Nature Genetics Editorial. (2000). Census, race and science. *Nature Genetics*, 24(2), 97–98.
- Nature Genetics Editorial. (2001). Genes, drug and race. *Nature Genetics*, 29(3), 239–240.
- National Institutes of Health. (1994). NIH guidelines on the inclusion of women and minorities as subjects in clinical research. Available from: <http://grants.nih.gov/grants/guide/notice-files/not94-100.html>.
- National Institutes of Health. (2001). NIH policy on reporting race and ethnicity data: subjects in clinical research. Available from: <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-01-053.html>.
- NIH Office of Extramural Research. (2008). Number and total award amount of all NIH competing and noncompeting research grants by NIH institute/center FY 2007–2004. Available from: <http://report.nih.gov/award/resgr.cfm>.
- Nobles, M. (2000). *Shades of citizenship: Race and the census in modern politics*. Palo Alto: Stanford University Press.
- Omi, M., & Winant, H. (1994). *Racial formation in the United States: From the 1960s to the 1990s*. New York: Routledge/Kegan Paul.
- Park, D. J., Stoehmacher, J., Zhang, W., Tsao-Wei, D., Groshen, S., & Lenz, H.-J. (2002). Thymidylate synthase gene polymorphism predicts response to capecitabine in advanced colorectal cancer. *International Journal of Colorectal Disease*, 17(1), 46–49.
- Phimister, E. G. (2003). Medicine and the racial divide. *New England Journal of Medicine*, 348(12), 1081–1082.
- Race, Ethnicity, and Genetics Working Group. (2005). The use of racial, ethnic, and ancestral categories in human genetics research. *American Journal of Human Genetics*, 77(4), 519–532.
- Risch, N., Burchard, E., Ziv, E., & Tang, H. (2002). Categorization of humans in biomedical research: genes, race and disease. *Genome Biology*, 3(7), 1–12.
- Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A., et al. (2002). Genetic structure of human populations. *Science*, 298(5602), 2381–2385.
- Rotimi, C. N. (2004). Are medical and nonmedical uses of large-scale genomic markers conflating genetics and 'race'? *Nature Genetics*, 36(11), S43–S47.
- Sankar, P., Cho, M. K., & Mountain, J. (2007). Race and ethnicity in genetic research. *American Journal of Medical Genetics Part A*, 143A(9), 961–970.
- Satel, S. (5 May, 2002). I'm a racially profiling doctor. *New York Times*.
- Satel, S. (10 December, 2004). Race and medicine can mix without prejudice. *Medical Progress Today*.
- Schermerhorn, R. A. (1978). *Comparative ethnic relations: A framework for theory and research*. Chicago: University of Chicago Press.
- Schwartz, R. S. (2001). Racial profiling in medical research. *New England Journal of Medicine*, 344(18), 1392–1393.
- Shim, J. K. (2005). Constructing 'race' across the science-lay divide: racial formation in the epidemiology and experience of cardiovascular disease. *Social Studies of Science*, 35(3), 405–436.
- Shriver, M. D. G. C., Kennedy, E. J., Parra, H. A., Lawson, V., Sonpar, J., Huang, J. M., et al. (2004). The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. *Human Genomics*, 1(4), 274–286.
- Shriver, M. D., & Kittles, R. A. (2004). Genetic ancestry and the search of personalized genetic histories. *Nature Review Genetics*, 5(8), 611–618.
- Timmermans, S., & Berg, M. (2003). *The gold standard: The challenge of evidence-based medicine and standardization in health care*. Philadelphia: Temple University Press.
- Tishkoff, S. A., & Kidd, K. K. (2004). Biogeography of human populations: implications for race and medicine. *Nature Genetics*(Suppl. 1), S21–S27.
- Wade, P. (1997). *Race and ethnicity in Latin America*. Chicago: Pluto Press.
- Williams, D. (1994). The concept of race in *Health Services Research: 1966–1990*. *Health Services Research*, 29(3), 261–273.
- Williams, D. (1997). Race and health: basic questions, emerging directions. *Annals of Epidemiology*, 7(5), 322–333.
- Williams, D. (2002). Ethnicity, race, and health. In N. F. Smelser, & P. B. Baltes (Eds.), *International encyclopedia of the social and behavioral sciences*. Oxford, UK: Elsevier Sciences Ltd.
- Williams, D. R., Neighbors, H. W., & Jackson, J. S. (2003). Racial/ethnic discrimination and health: findings from community studies. *American Journal of Public Health*, 93(2), 200–208.
- Witzig, R. (1996). The medicalization of race: scientific legitimization of a flawed social construct. *Annals of Internal Medicine*, 125(8), 675–679.
- Wu, A. H., Whittemore, A. S., Kolonel, L. N., Stanczyk, F. Z., John, E. M., Gallagher, R. P., et al. (2001). Lifestyle determinants of 5 α -reductase metabolites in older African American, White, and Asian-American Men. *Cancer Epidemiology, Biomarkers & Prevention*, 10(5), 533–538.