CHARACTERIZATION OF CLINICAL STUDY POPULATIONS BY RACE AND ETHNICITY IN BIOMEDICAL LITERATURE

Priyanka Kanakamedala, AB; Susanne B. Haga, PhD

Objective: The importance of race and ethnicity in biomedical research has long been a subject of debate, recently heightened by data revealed by the completion of the sequencing of the human genome and the mapping of human genetic variation. We aimed to determine whether and how the reporting of race has changed over the last three decades and how the practice may differ given study location, where the journal of publication is based, and decade of publication.

Design: We analyzed a sample of studies published in the *Journal of the American Medical Association, The Lancet*, and the *Canadian Medical Association Journal* from 1980 to 2009.

Main Outcome Measures: The number of articles that reported race by journal and decade and the descriptors used.

Results: Of 1,867 articles analyzed, 17.30% reported race. The reporting of race and number of populations reported increased over time for all three journals. In addition, the diversity of race/ethnicity descriptors increased, with increased use of race/ethnicity combinations and nationality of research subjects.

Conclusion: Though it has increased over the past few decades, the reporting of race/ethnicity of study populations is relatively low, ambiguous and inconsistent, likely influenced by the uncertain relevance of these variables to the study's outcomes, study location, researcher views, and the policies of journals and funding agencies. Thus, due to the inconsistent and ambiguous practice of reporting race/ethnicity, comparison of study outcomes can result in misleading conclusions. Improvements in standardization of terms and new approaches to characterize research participants related to race/ethnicity are imperative. (*Ethn Dis*. 2012;22(1):96–101)

Key Words: Race, Ethnicity, Clinical Trials, Publications

From Sanford School of Public Policy, Duke University, Durham, North Carolina.

Address correspondence to Susanne B. Haga, PhD; Institute for Genome Sciences & Policy and Sanford School of Public Policy; Duke University; 304 Research Drive; Box 90141; Durham, NC 27708; 919.684.0325; 919.613.6448 (fax); susanne.haga@duke.edu

INTRODUCTION

There has been much debate about the biological validity of race, particularly spurred by recent findings from genomic analyses of human populations.[1,2] Overall, it has been found that there are greater genetic differences between individuals of the same racial group than between individuals of different groups.[3] Although some rare genetic variants are only found in one population, the majority of genetic variations are found in all populations.[4,5] Despite the uncertainty of the biological validity of race, the construct of race is still used in biomedical research as a primary descriptor of study participants. However, its continued use may contribute to confounding outcomes instead of narrowing or identifying factors associated with study outcomes.

One of the primary reasons for the continued use of race in the literature is enforcement of requirements by funding agencies, regulatory agencies and biomedical journals. For example, in 2001, the US National Institutes of Health issued a policy specifying characterization of human subjects by race and ethnicity for research proposals and annual progress reports.[6] The racial and ethnicity classifications recommended are based on the 1997 Office of Management and Budget (OMB) Directive.[7] The US Food and Drug Administration (FDA) has issued similar guidelines recommending that racial and ethnicity data should be collected for participants in US clinical trials.[8] Furthermore, the International Committee of Medical Journal Editors developed a set of guidelines for reporting such descriptors and categorizations, which are followed by more than 500 journals. The guidelines state, "when authors use variables such as race or ethnicity, they should define how they measured the variables and justify their relevance."[9] Some journals have developed additional guidelines on how race and ethnicity should be reported.[10]

Descriptors for race and ethnicity have changed over time as evidenced from the continually evolving categories in the US Census since its inception.[11,12] Changes in terminology and differences between countries may lead to confusion and misinterpretation of conclusions involving race/ethnicity when making comparisons between studies.[13] This study aimed to document the characterization of the race and ethnicity of clinical study populations in three top medical journals based in three different countries over the last three decades. In particular, we were interested in investigating the use of race/ethnicity descriptors in journals based in different

This study aimed to document the characterization of the race and ethnicity of clinical study populations in three top medical journals based in three different countries over the last three decades.

English-speaking countries given the different descriptors used in census-taking and the national make-up of each country. In addition to surveying the prevalence and types of descriptors used to describe the race and ethnicity of study populations, we compared the terms used in publications to those used in national census.

METHODS

Generation of Dataset

To assess the evolution of the use of race/ethnicity categories in biomedical research over the past three decades, we selected a sample of publications for analysis in three top medical journals based on impact factor, country of publication, and online access from our library: the *Journal of the American Medical Association* (*JAMA;* based in the United States), *The Lancet* (*Lancet;* based in the United Kingdom) and the *Canadian Medical Association Journal* (*CMAJ*; based in Canada). A PubMed search of each journal was performed in April 2009 to identify a sample of articles published between 1981 and 2008. The PubMed search criteria included original research articles and clinical trials of human participants and specifically excluded editorials, letters, reviews, commentaries and guidances. To achieve a feasible sample size, articles from every third year of publication from 1981 onwards were selected for our final dataset for analysis. This yielded a total of 1,867 journal articles: 464 from *JAMA*, 1,318 from *Lancet*, and 85 from *CMAJ*. Articles were then manually curated to ensure that the aforementioned criteria were met.

A numerical code was created for each data category that was captured to facilitate data entry and analysis. Each article was identified by its PubMed identification number. Several data points were extracted from each article for analysis: journal, year/decade of publication, term(s) used to report race, ethnicity or heritage of the study subjects, the study trial location, and the country of the corresponding author. Articles with a single study population with respect to race were included in the dataset and coded accordingly to terminology used to describe population. Each article was reviewed and coded by one coder (PK). A random sample of articles (10%) was reviewed (SBH) and agreement was greater than 95%. All data were recorded and stored in an Excel file for analysis.

Data Analysis

Frequency distributions, *t* tests, and two proportion Z-tests (using $\alpha=.05$) were conducted to determine the statistical significance of the bivariate distributions between the independent variables (eg, journal and decade) and the dependent variable (race/ethnicity descriptors). More specifically, the usage of specific terms describing race/ethnicity was analyzed and compared within and between journals and each decade studied. In addition, we compared Census categories and use of race/ethnicity descriptors to explore the influence of the evolving Census categories on race/ethnicity descriptors used in the biomedical literature for each decade and country.

RESULTS

Prevalence of Articles Reporting Race/Ethnicity

Of the 1,867 journal publications analyzed overall, 17.2% of articles reported race/ethnicity; 39.22% of *JAMA* articles, 10.32% of *Lancet* articles and 5.88% of *CMAJ* articles (Table 1). The reporting of race increased through the decades in *JAMA* from 14.08% of articles reporting in the 1980s to 40.66% in the 1990s ($P<.0001$) and 46.45% in the 2000s. Similarly, the reporting of race/ethnicity increased in *Lancet* from 3.68% of articles in the 1980s to 7.69% in 1990s ($P=.004$) and 24.15% in 2000s ($P<.0001$). In *CMAJ*, no articles reported race in the 1980s or 1990s, but 13.89% of articles reported race in the 2000s.

Table 1. Prevalence of articles reporting race by journal and decade

Journal/Decade	Percentage Reporting Race (actual)
JAMA	
1980s ($n=71$)	14.08% (10)
1990s ($n=182$)	40.66% (74)[a]
2000s ($n=211$)	46.45% (98)
Overall ($n=464$)	39.22% (182)
Lancet	
1980s ($n=462$)	3.68% (17)
1990s ($n=533$)	7.77%[b] (41)
2000s ($n=323$)	24.15%[a] (78)
Overall ($n=1,318$)	10.32% (136)
CMAJ	
1980s ($n=22$)	0.00% (0)
1990s ($n=27$)	0.00% (0)
2000s ($n=36$)	13.89% (5)
Overall ($n=85$)	5.88% (5)
Total (1,867)	17.30%

[a] Significantly greater than previous decade ($P<.0001$).
[b] Significantly greater than previous decade ($P<.01$).

Prevalence of Populations Reported

The number of study populations (based on the number of race/ethnicity descriptors reported) was calculated for all articles that described the race/ethnicity of study participants. The average number of populations reported per article in *JAMA* that reported race increased significantly from 2.40 to 2.70 between the 1980s and 1990s ($P=.004$), and to 3.49 populations reported per article in the 2000s ($P=.04$) (Figure 1). We also found significant differences in the number of populations described in *Lancet*, with an average of 1.00 population per article reported in the 1980s, significantly increasing to 1.88 in the 1990s ($P=.03$) and 2.14 in the 2000s ($P=.01$). The difference in the number of populations described between *JAMA* and *Lancet* was statistically significant only in the 1990s ($P=.02$). A significant difference in the number of populations described in *JAMA* and *CMAJ* ($P=.005$) and *Lancet* and *CMAJ* ($P=.01$) in the 2000s was observed.

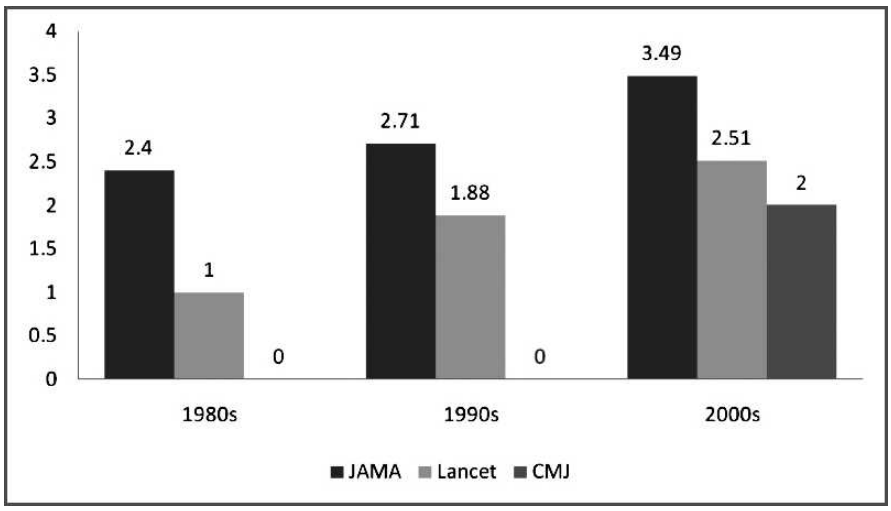

Figure 1. Average number of populations reported by articles by journal and decade

Variability of Race/ Ethnicity Descriptors

Overall, there were a total of 104 descriptors used to describe race/ethnicity in the articles analyzed (60 of these were in *JAMA*; 61 in *Lancet*; 6 in *CMAJ*). However, after grouping of synonymous terms (based on presumed ancestry), there were 9, 12, and 5 unique groups in *JAMA*, *Lancet*, and *CMAJ*, respectively. The descriptor Black was used most often to describe populations of African descent and White was used to describe populations of European descent. In *JAMA* articles, the most common descriptor used overall was White (100% of articles in the 1980s; 75.68% in the 1990s; 81.63% in the 2000s), followed by Black, Hispanic, and other, throughout all three decades studied (Table 2). In *Lancet*, the top descriptors used in the 1980s were Gambian, Chinese, White, and Caucasian; while in the 1990s were White (65.85%), Black, other, African American and Asian; and in the 2000s were White (78.21%), Black, Hispanic and other. In *CMAJ* in the 2000s, the most common descriptors were White (100%), Black, Asian, other, nonwhite and South Asian. The descriptors used in *Lancet* became increasingly descriptive, combining race and ethnicity, and describing populations by nationality or region, such as Arab, Chinese, Bangladeshi, Indian and North African. The use of these descriptors also aligns with the UK Census categories, which includes many nationalities in contrast to the US Census categories.

Relationship between Reporting of Race and Study Location

Studies published in *JAMA* and conducted in the United States wholly ($N=123$) or in part (international multi-site studies; $N=14$) were more likely to report race/ethnicity than non-US studies ($N=7$) ($P<.0001$.). However, studies published in *Lancet* and conducted in the United States in whole ($N=21$) or in part ($N=11$) were less likely to report race than studies that did not include a US site ($N=82$) ($P<.0001$). No studies conducted in the United States published in CMAJ were included in our dataset.

DISCUSSION

Our findings suggest that while the use of race/ethnicity descriptors in the biomedical literature has increased, the high but ambiguous number of descriptors used to describe study populations may lead to misleading conclusions across studies. The role of race in biomedical research continues to be debated, yet the practice of reporting race/ethnicity is both accepted and expected, if not required. Some have expressed concerns that the excessive focus on differences in race and ethnicity may reify the racial divide and inadvertently ascribe to the biological variations between races.[14,15] Others argue that the social, economic, environmental and perhaps genetic factors that surround race and ethnicity can be useful for testing hypotheses regarding medical outcomes.[16] Despite the FDA and OMB statements that race and ethnicity are not scientific, the organizations have upheld the belief that such descriptors are important for public health surveillance, research and collection of health-related data.[8] Our analysis of a sample of biomedical articles published in three leading medical journals over the last three decades shows that a significant amount of disparity exists with respect to the reporting of race/ethnicity and that the addition of ethnicity may reflect greater prominence of cultural heritage as well as recognition of multiracial heritage.

In summary, this study demonstrated three major findings: 1) the reporting of race/ethnicity of clinical trial participants has increased over the past three decades, though remains quite low; 2) the number of populations described per article is increasing, but continues to be dominated by White populations; and 3) when race/ethnicity are reported, study populations are described inconsistently between and within journals. Our finding regarding the overall low reporting of race/ethnicity (17.30%) is supported by several previous analyses,[17–19] though higher rates (>50%) have been reported in pediatric,[20–22] nursing,[23] and public health[24,25] research. Similarly, the increase in the number of articles reporting race over the past decades is supported by earlier work.[21,26] Despite the large number of terms used to describe study

Table 2. Top race/ethnicity descriptors reported per journal and decade

	JAMA			*Lancet*			*CMAJ*		
	Population	**Total**	**%[a]**	**Population**	**Total**	**%[a]**	**Population**	**Total**	**%[a]**
	(Total number of articles reporting race/ethnicity=10)			(Total number of articles reporting race/ethnicity=17)					
1980s	White	10	100.00	Gambian	4	23.53			
	Black	7	70.00	Chinese	4	23.53			
	Hispanic	4	40.00	White	2	11.76			
	Other	2	20.00	Caucasian	2	11.76			
	(Total number of articles reporting race/ethnicity=74)			(Total number of articles reporting race/ethnicity=41)					
1990s	White	56	75.68	White	27	65.85			
	Black	32	43.24	Black	10	24.39			
	Hispanic	25	33.78	Other	10	24.39			
	Other	21	28.38	African American	4	9.76			
				Asian	4	9.76			
	(Total number of articles reporting race/ethnicity=98)			(Total number of articles reporting race/ethnicity=78)			(Total number of articles reporting race/ethnicity=5)		
2000s	White	80	81.63	White	61	78.21	White	5	100.00
	Black	45	45.92	Black	25	32.05	Black	1	20.00
	Hispanic	42	42.86	Hispanic	14	17.95	Asian	1	20.00
	Other	37	37.76	Other	18	23.08	Other	1	20.00
							Nonwhite	1	20.00
							South Asian	1	20.00

[a] Percentages may be greater than 100% due to reporting of more than one population per article. Percentages were calculated based on total number of articles reporting race/ethnicity per journal and decade.

populations observed here and other studies,[17,25,27] the most common group described and represented in clinical studies continues to be White.[21,27] The apparent homogeneity of clinical trial populations has been a continuing challenge over the past several decades,[28,29] though evidence suggests that the representation in US cancer trials is consistent with disease burden in minority populations.[30] Corbie-Smith et al[19] found that studies of diseases associated with racial or ethnic disparities often did not report the race/ethnicity of participants or analyses based on such descriptors. Despite the reporting of race/ethnicity of study participants in some studies, it has been reported that few publications actually discuss race/ethnicity.[20]

The inconsistent reporting of race/ethnicity poses a challenge for comparing clinical trials globally and, in some instances, even those conducted in the same country. Many journals have policies in place regarding the use of race/ethnicity in human studies, however, there is no clear standard for race/ethnicity descriptors or defining study populations.[10,31] The wide range of researchers' views on the use and definition of race/ethnicity also likely contributes to the variation in practice observed.[10,32] The large number of terms used to describe seemingly similar groups in articles in *JAMA*, *Lancet* and *CMAJ* may lead to erroneous comparisons of outcomes of presumably similar study populations. For example, in the United Kingdom, Asian could refer to persons of Central, South, or East Asian descent whereas in the United States, the term refers to persons primarily of East or South Asia.[33,34] If race/ethnicity descriptors serve as surrogates for other variables potentially affecting health outcomes such as environment, the lack of a clear description of the population studied may lead to inaccurate comparisons. In addition, the difference in prevalence of reporting race between countries (higher in US-based studies)[26,35] limits comparisons with global studies. Describing population groups based on nationality (eg, Gambian) appears more prevalent for studies conducted outside of the United States and published in *Lancet*, perhaps due to greater diversity and cultural propensity to describe nationality than use broad race descriptors.

The inconsistent reporting of race/ethnicity poses a challenge for comparing clinical trials globally and, in some instances, even those conducted in the same country.

Based on the sample of biomedical literature analyzed in this study, racial/ethnicity descriptors appear to have become more specific over time, incorporating ethnicity in the 1990s. Earlier articles from *JAMA* predominantly used terms that were used as the race categories in the US census. In contrast,

articles in *Lancet*, while also using terms such as White and Black, included other terms not listed in their census such as Hispanic and African American. Compounding the problem of inconsistent reporting, many publications do not explain their criteria or methods for assigning race or ethnicity as an independent variable, despite the fact that many articles report conclusions about relationships between race/ethnicity, genetics and health outcomes.[17,25,36,37]

This ongoing ambiguity supports the development of a standard vocabulary as suggested by many others,[13,17,24,38–47] however, the strong national social and political values influencing race/ethnicity terminology would seriously impede global consensus. Alternatively, a new practice to accurately measure factors for which race serves as a surrogate could be developed and implemented such as collecting data on ancestry through grandparental birthplace or using genetic ancestry markers.[48–50] The use of genetic markers would allow for consideration of admixture,[51–53] however, it too is not an option without its own challenges.[54] First, genetic analysis of all clinical study populations would not be economically feasible until sequencing or genotyping technologies costs decline further and technical expertise becomes widespread and routine. Second, no standard set of markers has been developed, which could result in similar ambiguity in defining study participants' ancestry. Third, by replacing race with genetic markers, we would potentially be overlooking environmental factors associated with or causative of health outcomes. Thus, genetic characterization alone would not be sufficient to describe the biological and/or social factors that may be important to a study's outcomes and some combination of factors enabling more detailed characterization of research subjects' ancestry and environment are needed.

In conclusion, even if reporting of study participants' race/ethnicity continues to become more common, the lack of uniformity of the terms may pose problems for understanding the factors that affect health outcomes and lead to erroneous and potentially harmful conclusions. Thus, it becomes unclear how to interpret differences in study outcomes across populations in different countries, cultures, health systems and the like, beyond the United States, the United Kingdom, and Canada. Greater enforcement of existing journal policies regarding how populations are defined and clarification of the relevance of race/ethnicity to the study's outcomes as well as development of international consensus of race/ethnicity descriptors or consideration of alternative methods to characterize study populations are needed.

ACKNOWLEDGMENTS

This work was supported in part by the Duke Clinical Translational Science Award (National Institutes of Health: 5UL-1-RR-204128). We thank Ms. Genevieve Tindall for her kind assistance in preparation of the manuscript.

REFERENCES

1. National Institutes of Health, National Human Genome Research Institute. International Human Genome Sequencing Consortium describes finished human genome sequence; 2004. http://www.genome.gov/12513430. Accessed September 8, 2011.
2. Caulfield T, Fullerton SM, Ali-Khan SE, et al. Race and ancestry in biomedical research: exploring the challenges. *Genome Med*. 2009;1(1):8.
3. The use of racial, ethnic, and ancestral categories in human genetics research. *Am J Hum Genet*. 2005;77(4):519–532.
4. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. *Science*. 2002;296(5576): 2225–2229.
5. Rosenberg NA, Pritchard JK, Weber JL, et al. Genetic Structure of Human Populations. *Science.* 2002. 2002;298(5602):2381–2385.
6. National Institutes of Health. NIH policy on reporting race and ethnicity data: subjects in clinical research (NOT-OD-01-053); 2001. http://grants.nih.gov/grants/guide/notice-files/not-od-01-053.html. Accessed September 8, 2011.
7. Office of Management and Budget. Revisions to the standards for classification of federal data on race and ethnicity. 62 *Federal Register*, 58781–58790 (1997).
8. US Food and Drug Administration. Guidance for industry: collection of race and ethnicity data in clinical trials; September 2005. http://www.fda.gov/RegulatoryInformation/Guidances/ucm126340.htm. Accessed September 8, 2011.
9. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals; 2009. http://www.icmje.org/manuscript_1prepare.html. Accessed September 8, 2011.
10. Smart A, Tutton R, Martin P, Ellison GT, Ashcroft R. The standardization of race and ethnicity in biomedical science editorials and UK biobanks. *Soc Stud Sci*. 2008;38(3):407–423.
11. Nobles M. *Shades of Citizenship*. Stanford: Stanford University Press; 2000.
12. Nobles M. Racial Categorization and Censuses. In: Kertzer D, Artel D, eds. *Census and Identity*. Cambridge: Cambridge University Press; 2002.
13. Bamshad M. Lost in translation: meaningful policies for writing about genetics and race. *Am J Med Genet A*. 2007;143A(9):971–972.
14. Duster T. Medicine. Race and reification in science. *Science*. 2005;307(5712):1050–1051.
15. Goldstein DB, Willard HF. A medicine for one race only; 2005. http://news.duke.edu/2005/01/raceonly.html. Accessed September 8, 2011.
16. Burchard EG, Ziv E, Coyle N, et al. The importance of race and ethnic background in biomedical research and clinical practice. *N Engl J Med*. 2003;348(12):1170–1175.
17. Ma IW, Khan NA, Kang A, Zalunardo N, Palepu A. Systematic review identified suboptimal reporting and use of race/ethnicity in general medical journals. *J Clin Epidemiol*. Jun 2007;60(6):572–578.
18. Swanson GM, Bailar JC, 3rd. Selection and description of cancer clinical trials participants–science or happenstance? *Cancer*. 2002;95(5):950–959.
19. Corbie-Smith G, St George DM, Moody-Ayers S, Ransohoff DF. Adequacy of reporting race/ethnicity in clinical trials in areas of health disparities. *J Clin Epidemiol*. 2003;56(5):416–420.
20. Ackerman PD, Kelly ML, Walsh CA, Ross LF. Do peer guidelines or editorial policies affect the reporting and discussion of race and ethnicity in pediatric research? *Account Res*. 2005;12(1):17–31.
21. Brahan D, Bauchner H. Changes in reporting of race/ethnicity, socioeconomic status, gender, and age over 10 years. *Pediatrics*. 2005;115(2): e163–166.
22. Anderson MR, Moscou S. Race and ethnicity in research on infant mortality. *Fam Med*. 1998;30(3):224–227.

23. Drevdahl DJ, Philips DA, Taylor JY. Uncontested categories: the use of race and ethnicity variables in nursing research. *Nurs Inq*. 2006; 13(1):52–63.
24. Ahdieh L, Hahn RA. Use of the terms 'race', 'ethnicity', and 'national origins': a review of articles in the American Journal of Public Health, 1980–1989. *Ethn Health*. 1996;1(1): 95–98.
25. Comstock RD, Castillo EM, Lindsay SP. Four-year review of the use of race and ethnicity in epidemiologic and public health research. *Am J Epidemiol*. 2004;159(6):611–619.
26. Berger JS, Melloni C, Wang TY, et al. Reporting and representation of race/ethnicity in published randomized trials. *Am Heart J*. 2009;158(5):742–747.
27. Drevdahl D, Taylor JY, Phillips DA. Race and ethnicity as variables in Nursing Research, 1952–2000. *Nurs Res*. 2001;50(5):305–313.
28. Murthy VH, Krumholz HM, Gross CP. Participation in cancer clinical trials: race-, sex-, and age-based disparities. *JAMA*. 2004; 291(22):2720–2726.
29. Stewart JH, Bertoni AG, Staten JL, Levine EA, Gross CP. Participation in surgical oncology clinical trials: gender-, race/ethnicity-, and age-based disparities. *Ann Surg Oncol*. 2007;14(12): 3328–3334.
30. Tejeda HA, Green SB, Trimble EL, et al. Representation of African-Americans, Hispanics, and Whites in National Cancer Institute cancer treatment trials. *J Natl Cancer Inst*. 1996;88(12):812–816.
31. Winker MA. Race and ethnicity in medical research: requirements meet reality. *J Law Med Ethics*. 2006;34(3):520–525, 480.
32. Corbie-Smith G, Henderson G, Blumenthal C, Dorrance J, Estroff S. Conceptualizing race in research. *J Natl Med Assoc*. 2008;100(10): 1235–1243.
33. Bhopal RS, Phillimore P, Kohli HS. Inappropriate use of the term 'Asian': an obstacle to ethnicity and health research. *J Public Health Med*. 1991;13(4):244–246.
34. Rankin J, Bhopal R. Current census categories are not a good match for identity. *BMJ*. 1999; 318(7199):1696.
35. Frampton GK, Shepherd J, Dorne JL. Demographic data in asthma clinical trials: a systematic review with implications for generalizing trial findings and tackling health disparities. *Soc Sci Med*. 2009;69(8):1147–1154.
36. Shanawani H, Dame L, Schwartz DA, Cook-Deegan R. Non-reporting and inconsistent reporting of race and ethnicity in articles that claim associations among genotype, outcome, and race or ethnicity. *J Med Ethics*. 2006;32(12): 724–728.
37. Sankar P, Cho MK, Mountain J. Race and ethnicity in genetic research. *Am J Med Genet A*. 2007;143A(9):961–970.
38. Hahn RA. The state of federal health statistics on racial and ethnic groups. *JAMA*. 1992;267(2):268–271.
39. Kaplan JB, Bennett T. Use of race and ethnicity in biomedical publication. *JAMA*. 2003;289(20):2709–2716.
40. Trevino FM. Standardized terminology for hispanic populations. *Am J Public Health*. 1987;77(1):69–72.
41. Haga SB, Venter JC. Genetics. FDA races in wrong direction. *Science*. 2003;301(5632):466.
42. The unexamined 'Caucasian'. *Nat Genet*. 2004;36(6):541.
43. Osborne NG, Feit MD. The use of race in medical research. *JAMA*. 1992;267(2):275–279.
44. Bhopal R, Donaldson L. White, European, Western, Caucasian, or what? Inappropriate labeling in research on race, ethnicity, and health. *Am J Public Health*. 1998;88(9): 1303–1307.
45. Bhopal R, Kohli H, Rankin J. Editors' practice and views on terminology in ethnicity and health research. *Ethn Health*. 1997;2(3): 223–227.
46. Sankar P, Cho MK. Genetics. Toward a new vocabulary of human genetic variation. *Science*. 2002;298(5597):1337–1338.
47. Williams DR. The concept of race in Health Services Research: 1966 to 1990. *Health Serv Res*. 1994;29(3):261–274.
48. Schwartz AG, Wenzlaff AS, Bock CH, et al. Admixture Mapping of Lung Cancer in 1,812 African Americans. *Carcinogenesis*.2010.
49. Shields AE, Blumenthal D, Weiss KB, Comstock CB, Currivan D, Lerman C. Barriers to translating emerging genetic research on smoking into clinical practice. Perspectives of primary care physicians. *J Gen Intern Med*. 2005;20(2):131–138.
50. Kumar R, Seibold MA, Aldrich MC, et al. Genetic ancestry in lung-function predictions. *N Engl J Med*. 2010;363(4):321–330.
51. Halder I, Yang BZ, Kranzler HR, Stein MB, Shriver MD, Gelernter J. Measurement of admixture proportions and description of admixture structure in different U.S. populations. *Hum Mutat*. 2009;30(9):1299–1309.
52. Klimentidis YC, Miller GF, Shriver MD. Genetic admixture, self-reported ethnicity, self-estimated admixture, and skin pigmentation among Hispanics and Native Americans. *Am J Phys Anthropol*. 2009;138(4): 375–383.
53. Scanlon PD, Shriver MD. ''Race correction'' in pulmonary-function testing. *N Engl J Med*. 2010;363(4):385–386.
54. Bamshad M, Wooding S, Salisbury BA, Stephens JC. Deconstructing the relationship between genetics and race. *Nat Rev Genet*. 2004;5(8):598–609.

AUTHOR CONTRIBUTIONS

Study design: Kanakamedala, Haga
Acquisition of data: Kanakamedala
Data analysis and interpretation: Kanakamedala, Haga
Manuscript draft: Kanakamedala, Haga
Statistical expertise: Kanakamedala
Administrative: Kanakamedala
Supervision: Haga