Updated Guidance on the Reporting of Race and Ethnicity in Medical and Science Journals

Annette Flanagin, RN, MA; Tracy Frey, BA; Stacy L. Christiansen, MA; for the AMA Manual of Style Committee

The goal of this guidance is to provide recommendations and suggestions that encourage fairness, equity, consistency, and clarity in use and reporting of race and ethnicity in medical and science journals. As previously summarized, “terminology, usage, and word choice are critically important, especially when describing people and when discussing race and ethnicity. Inclusive language supports diversity and conveys respect. Language that imparts bias toward or against persons or groups based on characteristics or demographics must be avoided.”

With the publication of an earlier version of this guidance, comments were invited, and helpful assessments and comments were received from numerous reviewers, scholars, and researchers, who provided valuable feedback and represented diverse expertise and opinions. After thorough review of these comments (some of which did not agree with others) and additional research and discussion, the guidance was revised and updated, and additional formal review was obtained. In this Editorial, we present the updated guidance, and we sincerely thank the many reviewers for their contributions, each of whom are listed in the Acknowledgment at the end of this article.

This guidance continues to acknowledge that race and ethnicity are social constructs as well as the important sensitivities and controversies related to use of these terms and associated nomenclature in medical and health research, education, and practice. Thus, for content published in medical journals, language and terminology must be accurate, clear, and precise, and must reflect fairness, equity, and consistency in use and reporting of race and ethnicity. The guidance also acknowledges that the reporting of race and ethnicity should not be considered in isolation but should be accompanied by reporting of other sociodemographic factors and social determinants, including concerns about racism, disparities, and inequities, and the intersectionality of race and ethnicity with these other factors.

The guidance defines commonly used terms associated with race and ethnicity and acknowledges that these terms and definitions have changed, that some are out of date, and that the nomenclature will continue to evolve. Other topics addressed include relevant concerns and controversies in health care and research, including the intersectionality of ancestry and heritage, social determinants of health, and other socioeconomic, structural, institutional, cultural, and demographic factors; reporting of race and ethnicity in research articles; use of racial and ethnic collective or umbrella terms, capitalization, and abbreviations; listing racial and ethnic categories in alphabetical order vs order by majority; adjectival vs noun usage for categories of race and ethnicity; geographic origin and regionalization considerations; and brief guidance for journals and publishers that collect demographic data on editors, authors, and peer reviewers. Examples are provided to help guide authors and editors.

This updated guidance on reporting race and ethnicity is being added to the Inclusive Language section of the AMA Manual of Style: A Guide for Authors and Editors as subsection 11.2.3, Race and Ethnicity (see below). This section of the Correct and Preferred Usage chapter of the style manual also includes subsections that address reporting concerns and preferred nomenclature for sex and gender, sexual orientation, age, socioeconomic status, and persons with diseases, disorders, or disabilities. Each of these subsections also will be reviewed and updated.

With this publication, this guidance will be freely available on the JAMA website and in the AMA Manual of Style, and linked from the JAMA Network journals’ Instructions for Authors. Others are welcome to use and cite this guidance.

This guidance is not intended to be final but is presented with the understanding that monitoring will continue, and further updates will be provided as needed. Continual review of the terms and language used in the reporting of race and ethnicity is critically important as societal norms continue to evolve.

Section 11.12.3. Race and Ethnicity

Race and ethnicity are social constructs, without scientific or biological meaning. The indistinct construct of racial and ethnic categories has been increasingly acknowledged, and concerns about use of these terms in medical and health research, education, and practice have been progressively recognized. Accordingly, for content published in medical and science journals, language and terminology must be accurate, clear, and precise and must reflect fairness, equity, and consistency in use and reporting of race and ethnicity. (Note: historically, although inappropriately, race may have been considered a biological construct; thus, older content may characterize race as having biological significance.)

One of the goals of this guidance is to encourage the use of language to reduce unintentional bias in medical and science literature. The reporting of race and ethnicity should not be considered in isolation but should be accompanied by reporting of other sociodemographic factors and social determinants, including concerns about racism, disparities, and inequities, and the intersectionality of race and ethnicity with these other factors.

When reporting the results of research that includes racial and ethnic disparities and inequities, authors are encouraged to provide a balanced, evidence-based discussion of the
implications of the findings for addressing institutional racism and structural racism as these affect the study population, disease or disorder studied, and the relevant health care systems. For example, Introduction and Discussion sections of manuscripts could include implications of historical injustices when describing the differences observed by race and ethnicity. Such discussion of implications can use specific words, such as racism, structural racism, racial equity, or racial inequity, when appropriate.

**Definitions**

The definitions provided herein focus on reporting race and ethnicity. Definitions of broader terms (eg, disparity, inequity, intersectionality, and others) will be included in the overarching Inclusive Language section that contains this subsection.

“Race and ethnicity are dynamic, shaped by geographic, cultural, and sociopolitical forces.”

Race and ethnicity are social constructs and with limited utility in understanding medical research, practice, and policy. However, the terms may be useful as a lens through which to study and view racism and disparities and inequities in health, health care, and medical practice, education, and research. Terms and categories used to define and describe race and ethnicity have changed with time based on sociocultural shifts and greater awareness of the role of racism in society. This guidance is presented with that understanding, and updates have been and will continue to be provided as needed.

The terms race (first usage dating back to the 1500s) and ethnicity (first usage dating back to the late 1700s) have changed and continue to evolve semantically. The *Oxford English Dictionary* currently defines race as “a group of people connected by common descent or origin” or “any of the (putative) major groupings of mankind, usually defined in terms of distinct physical features or shared ethnicity” and ethnicity as “membership of a group regarded as ultimately of common descent, or having a common national or cultural tradition.” For example, in the US, ethnicity has referred to Hispanic or Latino, Latina, or Latinx people. Outside of the US, other terms of ethnicity may apply within specific nations or ancestry groups. As noted in a lexicographer’s post on the *Conscious Style Guide*, race and ethnicity are difficult to untangle. In general, ethnicity has historically referred to a person’s cultural identity (eg, language, customs, religion) and race to broad categories of people that are divided arbitrarily but based on ancestral origin and physical characteristics. Definitions that rely on external determinations of physical characteristics are problematic and may perpetuate racism. In addition, there is concern about whether these and other definitions are appropriate or out-of-date and whether separation of subcategories of race from subcategories of ethnicity could be discriminatory, especially when used by governmental agencies and institutions to guide policy, funding allocations, budgets, and data-driven business and research decisions. Thus, proposals have been made that these terms be unified into an aggregate, mutually exclusive set of categories as in “race and ethnicity.” (See Additional Guidance for Use of Racial and Ethnic Collective Terms.)

The term ancestry refers to a person’s country or region of origin or an individual’s lineage of descent. Another characteristic of many populations is genetic admixture, which refers to genetic exchange among people from different ancestries and may correlate with an individual’s risk for certain genetic diseases. Ancestry and genetic admixture may provide more useful information about health, population health, and genetic variants and risk for disease or disorders than do racial and ethnic categories.

Although race and ethnicity have no biological meaning, the terms have important, albeit contested, social meanings. Neglecting to report race and ethnicity in health and medical research disregards the reality of social stratification, injustices, and inequities and implications for population health, and removing race and ethnicity from research may conceal health disparities. Thus, inclusion of race and ethnicity in reports of medical research to address and further elucidate health disparities and inequities remains important at this time.

According to the “Health Equity Style Guide for the COVID-19 Response: Principles and Preferred Terms for Non-Stigmatizing, Bias-Free Language” of the Centers for Disease Control and Prevention (CDC), racism is defined as a “system of structuring opportunity and assigning value based on the social interpretation of how one looks...("race"), that unfairly disadvantages some individuals and communities, unfairly advantages other individuals and communities, and undermines realization of the full potential of our whole society through the waste of human resources.” Note that racism and prejudice can occur without phenotypic discrimination.

Jones and the CDC style guide have defined 3 levels of racism.

Systemic, institutionalized, and structural racism: “Structures, policies, practices, and norms resulting in differential access to the goods, services, and opportunities of society by ‘race’ (eg, how major systems—the economy, politics, education, criminal justice, health, etc—perpetuate unfair advantage).” The Associated Press (AP) Stylebook advises not to shorten these terms to “racism,” to avoid confusion with the other definitions.

Interpersonal and personally mediated racism: “Prejudice and discrimination, where prejudice is differential assumptions about the abilities, motives, and intents of others by ‘race,’ and discrimination is differential actions towards others by ‘race.’ These can be either intentional or unintentional.”

Internalized racism: “Acceptance by members of the stigmatized ‘races’ of negative messages about their own abilities and intrinsic worth.”

**Concerns, Sensitivities, and Controversies in Health Care and Research**

There are many examples of reported associations between race and ethnicity and health outcomes, but these outcomes may also be intertwined with ancestry and heritage, social determinants of health, as well as socioeconomic, structural, institutional, cultural, demographic, or other factors. Thus, discerning the roles of these factors is difficult. For example, a person’s ancestral heritage may convey certain health-related predispositions (eg, cystic fibrosis in persons of...
Northern European descent and sickle cell disease reported among people whose ancestors were from sub-Saharan Africa, India, Saudi Arabia, and Mediterranean countries; however, such perceptions have resulted in underdiagnosis of these conditions in other populations.\(^{15}\)

Also, certain groups may bear a disproportionate burden of disease compared with other groups, but this may reflect individual and systemic disparities and inequities in health care and social determinants of health. For example, according to the US National Cancer Institute, the rates of cervical cancer are higher among Hispanic/Latina women and Black/African American women than among women of other racial or ethnic groups, with Black/African American women having the highest rates of death from the disease, but social determinants of health and inequities are also associated with a high prevalence of cervical cancer among these women.\(^{16}\) The American Heart Association summarizes similar disparities in cardiovascular disease among Black individuals in the US compared with those from other racial and ethnic groups.\(^{17}\)

Identifying the race or ethnicity of a person or group of participants, along with other sociodemographic variables, may provide information about participants included in a study and the potential generalizability of the results of a study and may identify important disparities and inequities. Researchers should aim for inclusivity by providing comprehensive categories and subcategories where applicable. Many people may identify with more than 1 race and ethnicity; therefore, categories should not be considered absolute or viewed in isolation.

However, there is concern about the use of race in clinical algorithms and some health-based risk scores and databases because of inapplicability to some groups and the potential for discrimination and inappropriate clinical decisions. For example, the use of race to estimate glomerular filtration rates among Black adults has become controversial for several reasons.\(^{18-21}\) Oversimplification of racial dichotomies can be harmful, such as in calculating kidney function, especially with racial inequities in kidney care. In this context, health inequities among populations should be addressed rather than focusing solely on differences in racial categories (eg, Black vs White adults with kidney disease).\(^{22}\) Another example is the Framingham Risk Score, which was originally developed from a cohort of White, middle-class participants in the US included in the Framingham Heart Study and may not accurately estimate risk in other racial and ethnic populations. Similar concerns have been raised about genetic risk studies based on specific populations or that do not include participants from other groups (eg, a genome-wide association study that reports a genetic association with a specific disease or disorder based solely on a population of European descent).\(^{23}\) Use caution in interpreting or generalizing findings from studies of risk based on populations of individuals representing specific or limited racial and ethnic categories.

### Demographic Information

Aggregate, deidentified demographic information (eg, age, sex, race and ethnicity, and socioeconomic indicators) should be reported for research reports along with all prespecified outcomes. Demographic variables collected for a specific study should be indicated in the Methods section. Demographic information assessed should be reported in the Results section, either in the main article or in an online supplement or both. If any demographic characteristics that were collected are not reported, the reason should be stated. Summary demographic information (eg, baseline characteristics of study participants) should be reported in the first line of the Results section of the Abstract.

With regard to the collection and reporting of demographic data on race and ethnicity:

- The Methods section should include an explanation of who identified participant race and ethnicity and the source of the classifications used (eg, self-report or selection, investigator observed, database, electronic health record, survey instrument).
- If race and ethnicity categories were collected for a study, the reasons that these were assessed also should be described in the Methods section. If collection of data on race and ethnicity was required by the funding agency, that should be noted.
- Specific racial and ethnic categories are preferred over collective terms, when possible. Authors should report the specific categories used in their studies and recognize that these categories will differ based on the databases or surveys used, the requirements of funders, and the geographic location of data collection or study participants. Categories included in groups labeled as “other” should be defined.
- Categories should be listed in alphabetical order in text and tables.
- Race and ethnicity categories of the study population should be reported in the Results section.

### Examples

Reporting race and ethnicity in this study was mandated by the US National Institutes of Health (NIH), consistent with the Inclusion of Women, Minorities, and Children policy. Individuals participating in the poststudy survey were categorized as American Indian or Alaska Native, Asian, Black or African American, Hispanic or Latino, Native Hawaiian or Other Pacific Islander, or White based on the NIH Policy on Reporting Race and Ethnicity Data. Children’s race and ethnicity were based on the parents’ report.

Race was self-reported by study participants, and race categories (Black and White) were defined by investigators based on the US Office of Management and Budget’s Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity. Given that racial residential segregation is distinctively experienced by Black individuals in the US, the analytical sample was restricted to participants who self-identified as Black.

In this genome-wide association study, participants were from 8 African countries (ie, Kenya, Mozambique, Namibia,
Nigeria, South Africa, Sudan, Uganda, and Zambia). Any Black African group from any of the 8 African countries (mostly of Bantu descent) was included in the Black African cohort. The South African group composed primarily of multiple racial categories, comprising any admixture combination of individuals of European, Southeast Asian, South Asian, Bantu-speaking African, and/or indigenous Southern African hunter-gatherer ancestries (Khoikhoi, San, or Bushmen), was renamed admixed African individuals. The race and ethnicity of an individual was self-reported.

Data for this study included US adults who self-reported as non-Hispanic Black (hereafter, Black), Hispanic or Latino, and non-Hispanic White (hereafter, White) individuals. We excluded individuals who self-reported being Asian or of other race and ethnicity (which included those who were American Indian or Alaska Native and Native Hawaiian or Other Pacific Islander) because of small sample sizes.

Additional Guidance for Use of Racial and Ethnic Collective Terms
Specific racial and ethnic terms are preferred over collective terms, when possible. Authors should report the specific categories used in their studies and recognize that these categories will differ based on the databases or surveys used, the requirements of funders, and the geographic location of data collection or study participants.

When collective terms are used, merging of race and ethnicity with a virgule as “race/ethnicity” is no longer recommended. Instead, “race and ethnicity” is preferred, with the understanding that there are numerous subcategories within race and ethnicity. Given that a virgule often means “and/or,” which can be confusing, do not use the virgule construction in this context (see also 8.4, Forward Slash [Virgule, Solidus]).

The general term minorities should not be used when describing groups or populations because it is overly vague and implies a hierarchy among groups. Instead, include a modifier when using the word “minority” and do not use the term as a stand-alone noun, for example, racial and ethnic minority groups and racial and ethnic minority individuals.15,24 However, even this umbrella term may not be appropriate in some settings. Other terms such as underserved populations (eg, when referring to health disparities among groups) or underrepresented populations (eg, when referring to a disproportionately low number of individuals in a workforce or educational program) may be used provided the categories of individuals included are defined at first mention.29 The term minoritized may be acceptable as an adjective provided that the noun(s) that it is modifying is included (eg, “racial and ethnic minoritized group”). Groups that have been historically marginalized could be suitable in certain contexts if the rationale for this designation is provided and the categories of those included are defined or described at first mention.11

The nonspecific group label “other” for categorizing race and ethnicity is uninformative and may be considered pejorative. However, the term is sometimes used for comparison in data analysis when the numbers of those in some subgroups are too small for meaningful analyses. The term should not be used as a “convenience” grouping or label unless it was a prespecified formal category in a database or research instrument. In such cases, the categories included in “other” groups should be defined and reported. Authors are advised to be as specific as possible when reporting on racial and ethnic categories (even if these categories contain small numbers). If the numbers in some categories are so small as to potentially identify study participants, the specific numbers and percentages do not need to be reported provided this is noted. For cases in which the group “other” is used but not defined, the author should be queried for further explanation.

The terms multiracial and multiethnic are acceptable in reports of studies if the specific categories these terms comprise are defined or if the terms were predefined in a study or database to which participants self-selected. If the criteria for data quality and confidentiality are met, at a minimum, the number of individuals identifying with more than 1 race should be reported. Authors are encouraged to provide greater detail about the distribution of multiple racial and ethnic categories if known. In general, the term mixed race may carry negative connotations13 and should be avoided, unless it was specifically used in data collection; in this case, the term should be defined, if possible. To the extent possible, the specific type of multiracial and multiethnic groups should be delineated.

Example
In this study, 140 participants (25%) self-reported as multiracial, which included 100 (18%) identifying as Asian and White and 40 (7%) as Black and White.

Other terms may enter the lexicon as descriptors or modifiers for racial and ethnic categories of people. For example, the term people of color was introduced to mean all racial and ethnic groups that are not considered White or of European ancestry and as an indication of antiracist, multiracial solidarity. However, there is concern that the term may be “too inclusive,” to the point that it erases differences among specific groups.13,26-28 There are similar concerns about use of the collective and abbreviated terms for Black, Indigenous, and people of color (BIPOC) and Black, Asian, and minority ethnic (BAME) (commonly used in the UK). Criticism of these terms has noted that they disregard individuals’ identities, do not include all underrepresented groups, eliminate differences among groups, and imply a hierarchy among them.13,26-28 Although these terms may be used colloquially (eg, within an opinion article), preference is to describe or define the specific racial or ethnic categories included or intended to be addressed. These terms should not be used in reports of research, unless the terms are included in a database on which a study is based or specified in a research data collection instrument (eg, survey questionnaire).

In agreement with other guides,13,24 other terms related to colors, such as brown and yellow, should not be used to describe individuals or groups. These terms may be less inclusive than intended or considered pejorative or a racial slur.
In addition, avoid collective reference to racial and ethnic minority groups as “non-White.” If comparing racial and ethnic groups, indicate the specific groups. Researchers should avoid study designs and statistical comparisons of White groups vs “non-White” groups and should specify racial and ethnic groups included and conduct analyses comparing the specific groups. If such a comparison is justified, authors should explain the rationale and specify what categories are included in the “non-White” group.

Capitalization
The names of races, ethnicities, and tribes should be capitalized, such as African American, Alaska Native, American Indian, Asian, Black, Cherokee Nation, Hispanic, Kamba, Kikuyu, Latino, and White. There may be sociopolitical instances in which context may merit exception to this guidance, for example, in an opinion piece for which capitalization could be perceived as inflammatory or inappropriate (eg, “white supremacy”).

Adjectival Usage for Specific Categories
Racial and ethnic terms should not be used in noun form (eg, avoid Asians, Blacks, Hispanics, or Whites); the adjectival form is preferred (eg, Asian women, Black patients, Hispanic children, or White participants) because this follows AMA style regarding person-first language. The adjectival form may be used as a predicate adjective to modify the subject of a phrase (eg, “the patients self-identified as Asian, Black, Hispanic, or White”).

Most combinations of proper adjectives derived from geographic entities are not hyphenated when used as racial or ethnic descriptors. Therefore, do not hyphenate terms such as African American, African American, Mexican American, and similar combinations, and in compound modifiers (eg, African American patient).

Geographic Origin and Regionalization Considerations
Awareness of the relevance of geographic origin and regionalization associated with racial and ethnic designations is important. In addition, preferred usage may change about the most appropriate designation. For example, the term Caucasian had historically been used to indicate the term White, but it is technically specific to people from the Caucasus region in Eurasia and thus should not be used except when referring to people from this region.

The terms African American or Black may be used to describe participants in studies involving populations in the US, following how such information was recorded or collected for the study. However, the 2 terms should not be used interchangeably in reports of research unless both terms were formally used in the study, and the terms should be used consistently within a specific article. For example, among Black people residing in the US, those from the Caribbean may identify as Black but not as African American, whereas Black people whose families have been in the US for several generations may identify as Black and African American. When a study includes individuals of African ancestry in the diaspora, the term Black may be appropriate because it does not obscure cultural and linguistic nuances and national origins, such as Dominican, Haitian, and those of African sovereign states (eg, Kenyan, Nigerian, Sudanese), provided that the term was used in the study.

The term Asian is a broad category that can include numerous countries of origin (eg, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippines, Thailand, Vietnam, and others) and regions (eg, East Asia, South Asia, Southeast Asia). The term Asian American is acceptable when describing those who identify with Asian descent among the US population. However, when comparing racial and ethnic groups included and conduct analyses comparing the specific groups. If such a comparison is justified, authors should explain the rationale and specify what categories are included in the “non-White” group.

Editorial Opinion
August 17, 2021 Volume 326, Number 7
© 2021 American Medical Association. All rights reserved.
Description of people as being of a regional descent (eg, of African, Asian, European, or Middle Eastern or North African descent) is acceptable if those terms were used in formal research. However, it is preferable to identify a specific country or region of origin when known and pertinent to the study.

Example

For the genome-wide association study discovery stage, study participants of African ancestry were recruited from Ghana, Nigeria, South Africa, and the US, where the same phenotype definition was applied to diagnose primary open-angle glaucoma. The second validation meta-analysis included individuals with primary open-angle glaucoma and matched control individuals from Mali, Cameroon, Nigeria (Lagos, Kaduna, and Enugu), Brazil, Saudi Arabia, the Democratic Republic of the Congo, Morocco, and Peru.

For example, it is generally preferable to describe persons of Asian ancestry according to their country or regional area of origin (eg, Cambodian, Chinese, Indian, Japanese, Korean, Sri Lankan, East Asian, Southeast Asian). Similarly, study participants from the Middle Eastern and North African region should be described using their nation of origin (eg, Egyptian, Iranian, Iraqi, Israeli, Lebanese) when possible. Individuals of Middle Eastern and North African descent who identify with Arab ancestry and reside in the US may be referred to as Arab American. In such cases, researchers should report how categories were determined (eg, self-reported or selected by study participants or from demographic data in databases or other sources).

Note that Arab and Arab American, Asian and Asian American, Chinese and Chinese American, Mexican and Mexican American, and so on are not equivalent or interchangeable.

For studies that use national databases or include participants in a single country, a term for country of origin can be included if the term was provided at data collection (eg, Chinese American and Korean American for a study performed in the US, or Han Chinese and Zhuang Chinese for a study conducted in China). Again, how these designations were determined (eg, self-reported or selected by other means) should be reported.

Abbreviations

Generally, abbreviations of categories for race and ethnicity should be avoided unless necessary because of space constraints (eg, in tables and figures) or to avoid long, repetitive strings of descriptors. If used, any abbreviations should be clearly explained parenthetically in text or in table and figure footnotes or legends.

Guidance for Journals and Publishers That Collect Data on Editors, Authors, and Peer Reviewers

Journals and publishers that collect race and ethnicity data on editors, editorial board members, authors, and peer reviewers should follow principles of confidentiality, privacy, and inclusivity and should permit individuals to self-identify or opt out of such identification. The Joint Commitment on Action for Inclusion and Diversity in Publishing is developing an international list of terms for journals and publishers that collect information on race and ethnicity.

Journals that collect information on race and ethnicity should not permit editorial decisions to be influenced by the demographic characteristics of authors, peer reviewers, editorial board members, or editors. In addition, the collection and use of such data should respect privacy regulations and be secured to prevent disclosure of personally identifiable information. Individual personally identifiable information of authors and peer reviewers should not be accessible to anyone involved in editorial decisions. Such data may be used in aggregate to benchmark and monitor strategies to promote and improve the diversity of journals.
REFERENCES


