



April 12, 2023

Karin A. Orvis, PhD
Chief Statistician of the United States
Office of Information and Regulatory Affairs
Office of Management and Budget
725 17th Street NW
Washington DC 20503

RE: Comments on the initial proposals from the Federal Interagency Technical Working Group on Race and Ethnicity Standards (Working Group) for revising OMB's 1997 Statistical Policy Directive No. 15: Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity (SPD 15); OMB-2023-0001

Comments submitted electronically via www.regulations.gov

Dear Dr. Orvis:

The American Medical Informatics Association (AMIA) appreciates the opportunity to comment on *initial proposals from the Federal Interagency Technical Working Group on Race and Ethnicity Standards (Working Group) for revising OMB's 1997 Statistical Policy Directive No. 15: Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity (SPD 15); OMB-2023-0001*. AMIA is the professional home for more than 5,500 informatics professionals, representing frontline clinicians, researchers, and public health experts who bring meaning to data, manage information, and generate new knowledge across the health and healthcare enterprise.

AMIA appreciates the efforts of the Working Group and the Office of the Chief Statistician to improve the quality and usefulness of federal race and ethnicity data. The current (1997) race and ethnicity standards should be revised to capture data that reflect the diversity of the U.S. population more accurately. We commend the effort to ensure that any changes to existing categories are the product of interagency collaboration.

AMIA offers the following comments for consideration.

Dignity and Respect for All People

First, we steadfastly support the following Working Group guiding principle:

Respect individuals. Respect for individual dignity should guide the processes and methods for collecting data on race and ethnicity; respondent self-identification should be facilitated to the greatest extent possible.

AMIA suggests that not only is self-identification (not by observer) the gold standard, self-reporting (not by proxy) should be facilitated to the greatest extent possible. Additionally, emphasis must be placed on confidentiality of data in all settings, especially among data providers who collect race and ethnicity data that is then sent to a federal agency.

Purpose/Definition

The stated purpose of the standards is to provide a common language for uniformity and comparability in the collection and use of data on race and ethnicity by federal agencies. Reference is made to the defined scope of data collection:

Race and ethnicity are socio-political constructs. For purposes of these standards, the race and ethnicity categories set forth are sociopolitical constructs and are not an attempt to define race and ethnicity biologically or genetically.

We believe it is important for OMB to address the definition of “sociopolitical constructs” and to place this in context for the public. We also note the text reference to the public not distinguishing the difference between race and ethnicity, often considering the terms to be the same. It would be helpful for the Working Group to define “race” and “ethnicity” to provide clarity, with subsequent inclusion of definitions on data collection instruments.

Further, the text creates confusion with reference to ‘categories represent a social-political construct,’ followed by repeated inquiries about largely geographically focused existing or proposed “categories.” This is a core inconsistency that needs to be addressed. Most of the examples provided reflect nationalities and/or geographies. The OMB should partner with data scientists, scholars, and organizations representing the diversity of ethnic groups, including groups that represent oppressed and marginalized ethnic groups, to co-develop categories and examples that are reflective of the diversity of ethnic groups across any “category.”

Use Cases for Race and Ethnicity Data

It would benefit the public to better understand the comprehensive use cases for the collection of race and ethnicity data across the federal government, beyond the collection of demographic characteristics of the population. The text references ‘evolving needs and uses for data’ and

'societal needs.' The inclusion of more concrete examples would better enable all stakeholders to understand the scope of this review of race and ethnicity standards and why it is important.

As an example, archival explanatory text from the Federal Register (*Federal Register / Vol. 62, No. 210 / Thursday, October 30, 1997*) provides the following language to address the need for race and ethnicity standards. An update would be appropriate to illustrate more contemporary use cases.

They were developed in cooperation with Federal agencies to provide consistent data on race and ethnicity throughout the Federal Government. Development of the data standards stemmed in large measure from new responsibilities to enforce civil rights laws. Data were needed to monitor equal access in housing, education, employment, and other areas, for populations that historically had experienced discrimination and differential treatment because of their race or ethnicity. The standards are used not only in the decennial census (which provides the data for the "denominator" for many measures), but also in household surveys, on administrative forms (e.g., school registration and mortgage lending applications), and in medical and other research. The categories represent a social-political construct designed for collecting data on the race and ethnicity of broad population groups in this country and are not anthropologically or scientifically based.

Testing of Definitions of Minimum Categories

This proposal references "Previously Tested Definitions of Minimum Categories." Testing methods are addressed in *Federal Register / Vol. 62, No. 210 / Thursday, October 30, 1997*). It would be beneficial to all stakeholders to understand how the current proposed category modifications will be tested. We concur with the following Working Group guiding principle:

Category changes are based on sound research. Any changes in the categories should be based on sound methodological research and should include evaluations of the impact of any changes not only on the usefulness of the resulting data but also on the comparability of any new categories with the existing ones.

Here again, it would be beneficial for OMB and/or the Working Group to partner with data scientists, scholars, and organizations representing the diversity of ethnic groups across the continent, to ensure meaningful testing of any definitions.

Data Collection / Pediatric Population

Guidance/standards are needed on how to document language/race/ethnicity in pediatric populations, to not obfuscate caregiver/parent data with patient data.

Data Collection / Unknown Category

An “Asked but Unknown” category or “Could Not Collect” category is necessary for inclusion in revised standards. Such categories are preferable to race and ethnicity assigned based upon the perception of an observer or leaving the response blank (which makes it difficult for those using the data and may cause data quality issues especially when researchers use imputation to assign values to missing data). As noted in this proposal, special consideration will be needed to address data collection in adults with cognitive impairment. In cases where no family member is present, “Asked but Unknown” or “Could Not Collect” would represent the most appropriate choices and support data completeness and quality. Importantly, the reason “Why” the “Asked but Unknown”/“Could Not Collect” option was selected must be documented by the person collecting the data, again to facilitate data completeness and quality.

Source of Data Collection

Standards are needed to differentiate whether data were abstracted by health systems (asked at check-in or new patient registration by staff) or directly entered by patients/caregivers to improve data accuracy and empower patients/families to enter their own demographic data (gold standard).

Recommendation 3

- Require the collection of detailed race and ethnicity categories by default unless an agency determines that the potential benefit of the detailed data would not justify the additional burden to the agency and the public or the additional risk to privacy or confidentiality. See figures above.

OMB requests comments on:

3c. Some Federal information collections are able to use open-ended write-in fields to collect detailed racial and ethnic responses, while some collections must use a residual closed-ended category (e.g., “Another Asian Group”). What are the impacts of using a closed-ended category without collecting further detail through open-ended written responses?

AMIA Response: The answer depends on the intent of the revision of the standards. If the intent is to support accurate data gathering in the United States, there is likely to be more value in a set of broad categories to choose from with additional, open ended write-in supplementary fields. The categories people use in practice can eventually be mined for a more granular taxonomy.

Recommendation 4

4a. What term (such as “transnational”) should be used to describe people who identify with groups that cross national borders (e.g., “Bantu,” “Hmong,” or “Roma”)?

AMIA Response: This question brings forth an important distinction that is confounded in the current way that ethnicity appears to be conceptualized in this proposed recommendation. This may be due to the lack of a clear definition of ethnicity. Ethnicity is not synonymous with or equal to nationality – countries across the globe have within them various ethnic groups that are not the same as their nationality. Therefore, 4a is irrelevant because the ethnicity data collected should be only about ethnicity, not nationality. If nationality data are something of interest to collect, those data need to be collected in a question separate from ethnicity.

Since the specific groups listed would each fall entirely within one of categories defined below, is there a reason it would not be deemed sufficient to have them as examples within the categories? In addition, this would reduce the concern raised that the categories below are too rooted in nationalities as opposed to ethnicities.

1. If a combined race and ethnicity question is implemented, what term should be used for respondents who select more than one category? For example, is the preferred term “multiracial,” “multiethnic,” or something else?

AMIA Response: Having one selector for multiple identities reduces data quality, as all data are lost about which races or ethnicities the person is identifying as, therefore making it impossible for researchers to know how to use and interpret these data when aggregated in this way. Instead of having one option for multiracial/ethnic, allow respondents to select multiple races and identities. This will allow for better data quality and support appropriate use and interpretation of findings.

2. *Section D, Previously Tested Definitions of Minimum Categories.*

In all definitions referenced below, the Working Group and/or OMB should partner with data scientists, scholars, and organizations to co-develop categories and examples that reflect the diversity of ethnic groups across the continent, including groups that represent oppressed and marginalized ethnic groups.

- *American Indian or Alaska Native:* The category “American Indian or Alaska Native” includes all individuals who identify with any of the original peoples of North, Central, and South America. It includes people who identify as “American Indian” or “Alaska Native” and includes groups such as Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village of Barrow Inupiat Traditional Government, Tlingit, etc.

AMIA Response: There is confounding of tribal affiliation (US and Canadian designations) with ethnic groups, which are not always interchangeable.

- *Asian:* The category “Asian” includes all individuals who identify with one or more nationalities or ethnic groups originating in East Asia, Southeast Asia, or the Indian subcontinent. Examples of these groups include, but are not limited to, Chinese, Filipino, Asian Indian, Vietnamese,

Korean, and Japanese. The category also includes groups such as Pakistani, Cambodian, Hmong, Thai, Bengali, Mien, etc.

AMIA Response: Most of these examples are nationalities.

- *Black or African American:* The category “Black or African American” includes all individuals who identify with one or more nationalities or ethnic groups originating in any of the Black racial groups of Africa. Examples of these groups include, but are not limited to, African American, Jamaican, Haitian, Nigerian, Ethiopian, and Somali. The category also includes groups such as Ghanaian, South African, Barbadian, Kenyan, Liberian, Bahamian, etc.

AMIA Response: These are all nationalities, many of which do not represent ethnicities of indigenous groups due to the history of European colonialism and genocide in Africa and the Caribbean. Especially given that the examples for the western hemisphere only include the Caribbean, these options exclude Black communities across Central and South America.

- *Hispanic or Latino:* The category “Hispanic or Latino” includes all individuals who identify with one or more nationalities or ethnic groups originating in Mexico, Puerto Rico, Cuba, Central and South American, and other Spanish cultures. Examples of these groups include, but are not limited to, Mexican or Mexican American, Puerto Rican, Cuban, Salvadoran, Dominican, and Colombian. The category also includes groups such as Guatemalan, Honduran, Spaniard, Ecuadorian, Peruvian, Venezuelan, etc.

AMIA Response: The OMB needs to partner with Hispanic and Latino data scientists, scholars, and organizations representing the diversity of ethnic groups across the Americas to develop a clear definition of Hispanic and Latino ethnicities as well as co-develop categories and examples that are reflective of the diversity of Hispanic and Latino ethnic groups.

- *Middle Eastern or North African:* The category “Middle Eastern or North African” includes all individuals who identify with one or more nationalities or ethnic groups originating in the Middle East or North Africa. Examples of these groups include, but are not limited to, Lebanese, Iranian, Egyptian, Syrian, Moroccan, and Israeli. The category also includes groups such as Algerian, Iraqi, Kurdish, Tunisian, Chaldean, Assyrian, etc.

AMIA Response: It is unclear what is the factor that is being used to distinguish Middle Eastern and North African ethnicities—if it is cultural and religious similarities, then some Eastern European countries are excluded, such as Albania and Kosovo (which, along with other Balkan countries, were colonized by the Turkish Empire).

- *Native Hawaiian or Pacific Islander:* The category “Native Hawaiian or Pacific Islander” includes all individuals who identify with one or more nationalities or ethnic groups originating in Hawaii, Guam, Samoa, or other Pacific Islands. Examples of these groups include, but are not

limited to, Native Hawaiian, Samoan, Chamorro, Tongan, Fijian, and Marshallese. The category also includes groups such as Palauan, Tahitian, Chuukese, Pohnpeian, Saipanese, Yapese, etc.

AMIA Response: OMB should seek out partnerships with Native Hawaiian and Pacific Islander data scientists, scholars, and organizations representing the diversity of ethnic groups across the Pacific.

• *White*: The category “White” includes all individuals who identify with one or more nationalities or ethnic groups originating in Europe. Examples of these groups include, but are not limited to, German, Irish, English, Italian, Polish, and French. The category also includes groups such as Scottish, Norwegian, Dutch, Slavic, Cajun, Roma, etc.

AMIA Response: Most of these are nationalities, not ethnicities. Also, the term “Slavic” is too broad – there are many Slavic ethnic groups that are distinct from each other that the broad “Slavic” term does not provide specificity to be meaningful. The equivalent would be to have “Scandinavian” option rather than listing the various ethnic groups living in Scandinavia.

Are these draft definitions:

- i. Comprehensive in coverage of all racial and ethnic identities within the U.S.?
- ii. Using equivalent criteria?
- iii. Reflective of meaningful distinctions?
- iv. Easy to understand?
- v. Respectful of how people refer to themselves?

Please suggest any alternative language that you feel would improve the definitions.

AMIA RESPONSE: Definitions

These categories contain a mixture of nationality, race, and ethnicity options. Nationalities do not equal ethnicities.

4b. As seen in *Figure 2*, based on the Working Group's initial proposal, the question stem asks “What is your race or ethnicity?” Do you prefer a different question stem such as: “What is your race and/or ethnicity?”, “What is your race/ethnicity?”, “How do you identify?”, etc.? If so, please explain.

AMIA Response: A more inclusive phrasing might be “How do you identify your race and/or ethnicity?” This asks for the person’s perspective, but is still specific to the topic, while a question like “How do you identify?” could elicit a variety of responses. There are challenges in using the word “identify” when it appears OMB is seeking to “classify.”

The OMB should provide more clarity about how the process, and the categories, reflect classification (used for these administrative purposes) vs. identification, which has a completely different purpose and confounds what OMB appears to aim for in revision of the race and ethnicity standards. There should be discussion of the opportunity to shift the focus from socially assigned race and ethnicity arbitrary groups as the marker for “identity” to continental ancestry. Then from within these continental (not national) ancestral boundaries, self-identifiers can be designated. The continental ancestral options: Europe, Asia, Africa, North America by definition includes the Caribbean), South America and Australia. Simplifying the choices to 7 widely accepted locations as how one self-identifies, potentially makes selections of then sub-categories potentially more specific.

Recommendation 6

- Solicitation of comments on additional topics or future research

OMB requests comments on:

6a. SPD 15 does not dictate the order in which the minimum categories should be displayed on Federal information collections. Agencies generally order alphabetically or by population size; however, both approaches have received criticism. What order, alphabetical or by population size, do you prefer and why? Or what alternative approach would you recommend?

AMIA Response: Presentation of Race and Ethnicity Categories

AMIA suggests that categories be presented in alphabetical order versus by population size. With current data limitations, population size cannot be accurately captured and can reinforce marginalization, hierarchies, and oppression of smaller populations and ethnic groups. In addition, anything resembling a ranking should be eschewed.

6c. How can Federal surveys or forms collect data related to descent from enslaved peoples originally from the African continent? For example, when collecting and coding responses, what term best describes this population group (e.g., is the preferred term “American Descendants of Slavery,” “American Freedmen,” or something else)? How should this group be defined? Should it be collected as a detailed group within the “Black or African American” minimum category, or through a separate question or other approach?

AMIA Response: In order to answer this question, the OMB needs to partner with data historians, ethicists, scientists, scholars, and organizations representing Black Americans whose ancestors were enslaved and brought to the Western Hemisphere.

6d. The proposals in this FRN represent the Working Group's initial suggestions for revisions to SPD 15 to improve the accuracy and usefulness of Federal race and ethnicity data. The Working Group and OMB welcome comments and suggestions on any other ways that SPD 15 could be revised to produce more accurate and useful race and ethnicity data.

AMIA Response: Addressing Inequities

The collection of race and ethnicity are important. However, asking a follow up question to determine if the person has experienced discrimination based on any of their characteristics is equally important to better understand the basis of discrimination and inequities in healthcare, employment, housing, or other sectors. This must, however, be accompanied with a clear explanation of what discrimination means. Any definition must include clarity about systemic and/or systematic structures that intentionally denies one access to resources granted to a reference group. Again, OMB should partner with scholars to make the designation of discrimination apolitical and verifiable.

Thank you or your consideration of these comments. If you have any questions, please contact Tayler Williams, AMIA Public Policy Manager, at twilliams@amia.org.

Sincerely,



Gretchen Purcell Jackson, MD, PhD, FACS, FACMI, FAMIA
President and Board Chair, AMIA
Vice President & Scientific Medical Officer, Intuitive Surgical
Associate Professor of Surgery, Pediatrics, and Biomedical Informatics, Vanderbilt University
Medical Center