

**Comment on OMB Proposal to Revise Race and Ethnicity Guidelines
from
The Data Equity Center at the UCLA Center for Health Policy Research**

April 12, 2023

Bob Sivinski
Chair, Interagency Technical Working Group on
Race and Ethnicity Standards
Office of Management and Budget
1650 17th St. NW
Washington, DC 20500

Re: Feedback on Initial Proposals for Updating OMB's Race and Ethnicity Statistical Standards

Dear Mr. Sivinski,

In response to the Federal Register notice 88 FR 5375, and representing the UCLA Data Equity Center (DEC) at the UCLA Center for Health Policy Research, we would like to provide feedback on the Initial Proposals for Updating OMB's Race and Ethnicity Statistical Standards. The DEC, funded by the Robert Wood Johnson Foundation, works to provide resources and assistance to organizations seeking to improve the equity of their data through improvements to the collection, processing and dissemination of this data that enhance representativeness of often-overlooked populations, and enabling those populations access to that data. Equitable data delivers a product that truly reflects the population it purports to represent but also remains interoperable and in line with standards such that data can be seamlessly shared and used for policy making.

The race and ethnicity guidelines issued by OMB under Directive 15 have wide-ranging effects on data collection efforts not only within the federal government but also throughout the country. We believe that it is important that the revisions to these guidelines reflect the growing racial and ethnic diversity of the population of the United States. This will improve the overall quality of the data and ensure that public policy can better respond to the specific needs and experiences of all. To this end, it will be important for the proposed revisions to both align with the ways in which the general public identify race and ethnicity and be consistent with the overall goal(s) of collecting this data. We recognize that there are different ways to define race and ethnicity and many purposes for which this data may be collected. For this reason, we begin this memo by providing some background material discussing the concepts of race and ethnicity and how these relate to their measurement. We then provide specific comments on and recommendations for the proposed revisions to Directive 15.

BACKGROUND

Defining Race and Ethnicity

In the United States, individuals are racialized (assigned a race or ethnicity) under the current OMB construct of ethnicity as "Hispanic or Latino or Not Hispanic or Latino" and race as "American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or other

Pacific Islander, and White” (Ponce, Shimkhada, Adkins-Jackson, 2023). Race and ethnicity are conceptually distinct systems of classification that can be used to place individuals into socially meaningful categories. For the purposes of this document, we define race and ethnicity as follows:

- *Race* is a classification system that is based on a set of phenotypic characteristics that are culturally-identified as visual markers of membership in socially-meaningful groups.
- *Ethnicity* is a classification system based on cultural characteristics, such as language, ancestry, heritage, culture, and religion.

Purpose of Race and Ethnicity Data Collection

From its beginnings, the United States has legally afforded different rights and privileges to individuals based on their race and/or ethnicity. Race and ethnicity have been foundational organizing principles of social hierarchy at the federal, state, and local levels, which means that classifications of race and ethnicity have been and currently remain integral to the establishment and perpetuation of social, political, cultural, and economic hierarchies and inequality in the U.S. The historical legacies of both legal and extralegal discrimination remain apparent today, despite the passage of the Civil Rights Act of 1964, which for the first time prohibited discrimination on the basis of race or national origin throughout the United States.

Collection of race and ethnicity information by federal agencies is necessary for identifying the presence and magnitude of the effects of both current and historical legacies of disparate and discriminatory treatment on the basis of race and ethnicity. In addition to the enforcement needs of the Civil Rights Act, a number of laws, regulations, and treaties require the collection of race, ethnicity, and other related data. For example, Public Law 94-311¹ calls for the improvement and expansion of the collection, analysis, and publication of data on Americans of Spanish origins or descent. Similarly, the Patient Protection and Affordable Care Act Section 4302 requires the development and implementation of data collection standards for race and ethnicity in all national population health surveys (to the extent practicable). Further, the U.S. federal government has treaty obligations with American Indian and Alaska Native tribes to collect data on their behalf.

Race and ethnicity as a Multidimensional Concept

Because race operates on interpersonal, organizational, and structural levels, the mechanisms through which racism is generated are multilevel and complex, which means that multiple types of measures may be necessary to fully identify and understand these processes (see e.g., Roth, 2016). These could include:

- Self-reported race and ethnicity
- Racial and ethnic identity
- Direct measurement of specific phenotypic characteristics, such as skin color, or of clusters of phenotypic characteristics
- Perceived race and ethnicity (how others perceive the respondent’s race and ethnicity)
- Parent race and ethnicity

¹ <https://uscode.house.gov/statutes/pl/94/311.pdf#:~:text=PUBLIC%20LAW%2094-311%E2%80%94JUNE%2016%2C%201976%20Public%20Law%2094-311,statistics%20for%20Americans%20of%20Spanish%20origin%20or%20descent>.

- Contextual measures of racial and ethnic composition (e.g., neighborhood, school, workplace)

The most common measure of race and ethnicity asks respondents to self-report their own race and/or ethnicity—or proxy report this information for another person. This self-reported measure may take into account not only how an individual identifies their own race, but also other factors, such as the individual's perceived race, their ancestry, the context in which the data is being collected, and the individual's social status and experiences. Because it incorporates all of these dimensions, it can be considered the broadest measure of race and for this reason, our comments and recommendations in this memo focus on this measure. It is important to note that for some respondents, this form of race and ethnicity reporting can vary over time and by context (Saperstein and Penner, 2014; Freedman et al., 2012; Lopez, 2003).

The way in which others classify an individual's race (perceived race) is important for understanding the ways in which discrimination and racism occur within interpersonal interactions. This is most often measured by asking respondents to report what race other people see them as (López et al., 2018). This measure asks respondents to report on other people's views without necessarily specifying which other people or under what circumstances—which may reduce accuracy or interpretability. It also may be less relevant for studying the effects of structural racism or of more intimate or persistent relationships (such as an employment context) within which an individual's racial identity is more likely to be known to others.

Another method is to measure a specific phenotypic characteristic—such as skin color—either subjectively or using objective measurement tools and examine how outcomes change based on differences or gradations in this measure (Ma et al., 2018). This provides valuable information regarding the variability within racial ethnic categories and how perceived distance between different racial categories affects disparities. However, because racial categories usually depend on a cluster of phenotypic characteristics and there is significant overlap in phenotypic characteristics across racial categories, this type of measure is more often used to assess variation within racial categories or identify boundaries between them rather than evaluate disparities between them.

It's important to note that the level of detail or disaggregation about race and ethnicity that respondents can reasonably be expected to be able to report will differ across measures. Individuals who are self-reporting their race and ethnicity have greater knowledge of their own racial and ethnic background, how they are perceived by others, and the impact of revealing their own race or ethnicity in a specific context. When individuals are reporting on the race or ethnicity of others, particularly of others of a different race or ethnicity, they may not be able to provide the same level of detail or report with the same degree of accuracy. Thus, our recommendations for detailed racial and ethnic categories may not be as appropriate for these types of measures.

Distinguishing Race and Ethnicity in Theory and Practice: Implications for Self-Reported Race

OMB's proposal to modify Directive 15 includes the suggestion to combine the racial and ethnic components of the existing two question format into a single combined race and ethnicity question. We support this change because this is more consistent with the way in which individuals process and assign race and ethnicity information. Although definitionally we make a conceptual distinction between racial and ethnic classification systems, in practice this

distinction is not always clear or practicable. The focus on phenotypic characteristics suggests that racial classification systems are rooted in biology, while ethnic classifications are rooted in cultural ancestry; however, this biological vs cultural distinction is both inaccurate and harmful. It is inaccurate because in daily life, classification into racial categories is often based on a combination of phenotypic characteristics *and* cultural signals and expectations. It is harmful because it ascribes a biological justification for socially-based hierarchies.

The relevance of individual or clusters of phenotypic characteristics for defining racial group membership is culturally specific and dependent on sociopolitical context and existing social status hierarchies to give them meaning. Often, these prototypical phenotypes and cultural characteristics—such as language or clothing—that mark group membership are laden with beliefs or assumptions drawn from otherwise unrelated stereotypes—usually about biologically ingrained ability and/or culturally-based characteristics and behavior—that are used to justify continuing race- and ethnicity-based social status hierarchies. Similarly, phenotypic characteristics may be used in conjunction with cultural characteristics to evaluate ethnic typicality (Mitchell et al., 2018) Thus, in practice, often there is no clear distinction between the ways in which racial and ethnic classification and discrimination practices operate.

Many phenotypic characteristics are not discrete characteristics but variations on a continuum without clear lines that demarcate between racial categories. In fact, there is considerable overlap across racial categories and where individuals mark the boundaries between categories may shift based on the presence or absence of other phenotypic or ethnic characteristics (Ma et al., 2018), social context (Loveman, 2014; Telles and Paschel, 2015), or individual characteristics of the rater (Krosch, Jost, and Bavel, 2021) or the person being classified (Carpinella et al., 2018; Freedman, et al., 2011). For example, there is an extensive literature on colorism within racial categories that demonstrates that there is considerable variation in skin tone within each of the currently recognized OMB racial categories, as well as overlap across them (Fuentes et al., 2021; Gonlin, 2022; Ryabov, 2015).

This phenotypic variation means that although racial categories may be conceptually defined based on clusters of phenotypic characteristics, individuals and ethnic groups classified within a racial category vary in their prototypicality—the degree to which they embody these characteristics (Monk, 2022; Ma et al., 2018). Similarly, although ethnicity is defined as being based on shared cultural characteristics, within any ethnic group there is no one set of cultural characteristics that all members of this group share and individuals also vary in their ethnic typicality (Mitchell et al., 2018).

It is important to note that the characteristics associated with a prototypic phenotype for a racial category are culturally specific (Telles and Paschel, 2015; Goh and McCue, 2021). For example, although “Asian” is a racial category that encompasses individuals with ancestry from most of the Asian continent in both the United States and Britain, due to differences in their histories of colonialization and migration, the phenotypic characteristics considered most prototypically Asian differ between the two countries: Americans are more likely to consider East Asian characteristics more prototypically Asian, while Britons are more likely to consider South Asian characteristics more prototypically Asian (Goh and McCue, 2021).

Cultural understandings of the characteristics that define prototypicality for racial groups have implications for how these racial categories are viewed; the assumptions, prejudices, and stereotypes associated with racial categories; and how individuals racially classify themselves and are classified (and treated) in daily life. For example, because Americans view East Asians as more prototypically “Asian,” when asked to consider the classification of specific ethnic

groups within racial categories, they often exclude South Asians from the Asian category (Lee and Ramakrishnan, 2020).

The variation in racial and ethnic typicality within racial and ethnic categories combined with the fuzziness of group boundaries and overlap in phenotypic characteristics suggest two things. First, that the boundaries between race and ethnicity are porous and, because they are socio-politically determined, they are continually shifting with demographic and historical change. And second, these demographic and historical changes mean that official definitions may not reflect the ways in which race and ethnicity are understood by the larger public, particularly when segments of that public have immigrated to the U.S. and are more familiar with—and identify with—racial and ethnic classification systems used in other countries.

Classification of Race and Ethnicity under OMB Directive 15

The current OMB race and ethnicity classification system attempts to skirt the issue of racial phenotypic complexity by defining racial categories primarily based on geographic location, most commonly country or nation of origin. This is consistent with the statutory language of the Civil Rights Act, which bans discrimination based on nation of origin. In addition, by defining race based on country (or, more broadly, continent) of origin, the OMB definitions (seem to) provide objective criteria for classification, which could potentially improve the consistency and quality of responses, particularly for those who are less familiar with the U.S. racial classification system. However, the implementation of this system of classification also includes a number of inconsistencies that are tied to underlying essentialist assumptions about race and ethnicity and the connection between them.

In 1977, OMB issued Directive No. 15 (42 FR 1926 May 12, 1977²) the U.S. federal government's first effort to standardize the collection of race and ethnicity data. These guidelines introduced the two-question format for race and ethnicity in which respondents were asked to report Spanish origin separately from their race. The inclusion of the Spanish origin question was prompted by the passage of Public Law 94-311³, which called for the improvement and expansion of the collection, analysis, and publication of data on Americans of Spanish origins or descent. Moreover, the standards outlined in OMB Directive No. 15 (1977) "reflected legislative and agency needs, and not efforts by population groups to be specifically identified" (60 FR 166 August 28, 1995: 44675⁴).

The introduction of the two-question format made it possible to collect the mandated information on Spanish origin, while recognizing that, because Spanish colonialization was not limited to the Americas and that the Spanish colonialists in many of these nations were engaged in the African slave trade and intermixed (often by force) with the nations' indigenous populations, this newly created "Hispanic" category could draw individuals from multiple racial groups. By collecting Hispanic origin using a separate question, the guidelines preserved the original racial classification system that restricted respondents to reporting a single race.

After the implementation of the 1997 revisions to Directive 15, which recommended that respondents be allowed to select more than one racial and ethnic category, a separate Hispanic

² <https://tile.loc.gov/storage-services/service/ll/fedreg/fr042/fr042092/fr042092.pdf>

³ <https://uscode.house.gov/statutes/pl/94/311.pdf#:~:text=PUBLIC%20LAW%2094-311%E2%80%94JUNE%2016%2C%201976%20Public%20Law%2094-311,statistics%20for%20Americans%20of%20Spanish%20origin%20or%20descent.>

⁴ <https://www.govinfo.gov/content/pkg/FR-1995-08-28/pdf/95-20787.pdf>

origin question was no longer strictly necessary for fully identifying the Hispanic/Latino population. At the same time, the limitations of the two-question approach became more difficult to ignore as a substantial percentage of Hispanic/Latino respondents struggled to find themselves in the existing race categories. For example, as discussed in greater detail in the comment submitted by the California Health Interview Survey (CHIS) in the 2021 CHIS 38.5% of adult Latino respondents either selected “Other (Specify)” or skipped the race question altogether, and only one-fifth of those who specified a response could be recoded into an existing racial category based on their response. Similarly, after data processing and recoding of open-ended “other race” responses, 36.6% of single-race adult Latinos in the 2021 American Communities Survey (ACS)⁵ remained classified in the “Other race” category. In contrast, less than 2% of non-Latino respondents in either survey reported their race as “Other race.”

The 1997 OMB Directive 15 guidelines do not define race or ethnicity or explain what these questions are developed to measure, and this creates internal inconsistencies and contradictions that affect the quality of measurement. The definitions it does provide are for the five racial groups and these are predominantly based on country (or continent) of origin rather than racial phenotype or cultural characteristics. This further blurs the boundaries between race and ethnicity and freezes the complex political and migration histories of many ethnic populations to specific locations. Additionally, it produces a number of internal contradictions and inconsistencies between how individuals understand and report their own race and ethnicity and how this data is coded and understood by the federal government.

In particular, the notion of “origin” suggests there is an “original home” of each racial group and elides the ways in which historical experiences of slavery, colonialization, and migration problematize the geography-based definitions of these categories. It fails to interrogate the ways in which the concept of “origin” has built into it implicit essentialist assumptions about when and where the differences between racial (and ethnic groups) arose and implies a static persistence over time. It ignores changes in geopolitical contexts that shape whether and how these racial and ethnic identities are practiced, embodied, and experienced in the present day, including whether and how they are retained and modified across generations and within people with multiracial and multiethnic backgrounds.

Further, by defining race as a set of super categories within which most ethnic origins are nested, this method implies that there is an underlying assumption of racial homogeneity within ethnic groups (outside of the Americas) and poses difficulties for classifying ethnic groups with cross-national—and potentially, cross-racial—ties.

This is evident in the definition of the African American or Black category. In order to exclude White African descendants of European colonizers and North Africa populations that have historically been considered White and maintain the social boundaries between these categories, the definition of the category African American or Black includes self-referential language on skin color: “A person having origins in any of the black racial groups of Africa.” The African American or Black category is the only racial category that specifically references a phenotypic characteristic: an individual is considered Black when an individual is descended from Black people.

The problematic nature of these definitions can also be seen in the definition of the American Indian and Alaska Native (AIAN) population that appears in the 1997 guidelines. This definition

⁵ Calculated from the U.S. Census Bureau American Community Survey 1-year public use data file using the Beta version of the U.S. Census Bureau’s online API query tool.

treats this community not as a race, but as a political and cultural group. Although it begins by defining AIANs as “all individuals having origins in any of the original peoples of North, Central, and South America,” it then restricts this community to those who “maintain tribal affiliation or community attachment,” effectively imposing a cultural component on a racial category. This further demonstrates a lack of clarity regarding the boundaries between race and ethnicity. At the same time, it leaves those with AIAN ancestry or origins and no cultural attachment to an AIAN community outside of this category, reflecting the long history of forced relocation and racial/cultural integration of this population within the U.S. and Americas more broadly.

Limitations of the current approach also become apparent when we consider populations at the boundaries between racial categories. For example, the category “Hispanic” was popularized by the federal government to refer to individuals of Spanish origins. Because of the global reach of Spanish colonialization, this encompasses nearly all of Central and South America—excluding Brazil. Because Brazil was colonized by the Portuguese rather than the Spanish, Brazilians are not considered Hispanic or Latino. However, this distinction is not always clear to respondents and Brazilians sometimes report themselves to be Hispanic or Latino. Moreover, because Brazil is a diverse, multiracial society, Brazilians cannot all be nested within a single race and may appear across multiple racial categories. This means that they are not often identifiable within most data.

People with ancestry from countries in the Middle East and North Africa (MENA) are considered White under OMB Directive 15; however, within these countries there is considerable variation in skin color and other phenotypic and cultural characteristics that differ from other ethnic groups who are classified as White (Maghbouleh et al., 2022). This leads a significant percentage of this population to report their race as “other race”. However, when respondents report as “other race” and write in a MENA country as their nation of origin, these individuals are also reclassified as White during data processing. These respondents do not perceive themselves as White, and generally are not perceived (or treated) by others as White in the United States (Maghbouleh et al., 2022; Aziz, forthcoming), and thus this reclassification based on country of origin misrepresents their experience as a minoritized population and renders this experience invisible.

Filipinos occupy a contested classification space at the boundary between two race categories: Asian and Native Hawaiian and Pacific Islanders (NHPI). The United Nations and U.S. federal government (appropriately) classify the Philippines as a Southeast Asian nation; however, based on geography, the Philippines are both an Asian country and located on the Pacific Ocean. As such, a significant proportion of Filipinos identify as NHPI rather than as Asian. Additionally, because the Philippines were at one time colonized by the Spanish, some Filipinos identify as Hispanic/Latinx, although their claim to Spanish origin is sometimes contested because their official language is not Spanish (Lopez, Krogstad, and Passel, 2022).

COMMENTS ON OMB’S PROPOSED REVISIONS AND RECOMMENDATIONS

Guiding Principles for Improving Data Equity through Data Collection and Reporting

The underlying purpose of collecting race and ethnicity information is to measure race- and ethnicity-based inequities and understand the ways in which these are produced and maintained over time. Although the 1997 OMB Directive 15 indicates that these guidelines lay out the minimum race and ethnicity categories that should be collected by the federal government, often these guidelines are treated as the most detailed option. However, an

abundance of research has demonstrated that inequities between subgroups within these broad categories can often be as large or larger than those between the minimum categories. This has sparked a movement calling for the disaggregation of racial and ethnic categories as one step toward improving data equity (Ponce, Shimkhada, and Adkins-Jackson, 2023).

In response, the Biden Administration has made data equity a central goal⁶. As a result of this push toward racial and ethnic data disaggregation and data equity, a number of agencies and federal data collection efforts have expanded the number of ethnic subcategories included as response options, either by listing additional response options—usually as subcategories within the Directive 15 race or Hispanic origin categories—or by providing respondents the opportunity to provide a write-in response. For example, the minimum guidelines issued by the Department of Health and Human Services under the Affordable Care Act, Section 4302 recommend that several Asian subcategories be listed as response categories in order to allow for disaggregation of the Asian race category. But to-date many of these efforts have been voluntary.

In a 2016 study on the status of racial and ethnic measurement practices among population health data managers, we found that the OMB standards and guidance for collection and processing are key to guiding data disaggregation (Ponce, Scheitler, and Shimkhada, 2016). However, even with the guidance, we found several barriers in increasing the availability of disaggregated data including sample size, question wording, and difference in reporting and tabulation. The revisions to OMB Directive 15 can play an important role in addressing these challenges.

When more detailed race and ethnicity information is collected, this serves multiple purposes. First, detailed information on ethnic subgroups makes it possible to disaggregate the data by these ethnic subgroups. Each of the broad race and ethnicity categories includes individuals from a wide variety of ethnic backgrounds, each of which have their own demographic, cultural, and socioeconomic patterns of migration, segregation, and integration that affect outcomes across generations in the United States.

Data disaggregation allows users to both 1) evaluate the specific experiences of these subgroups, which may differ substantially from those of the broader racial category; and 2) assess the ways in which changes in the ethnic composition of these broader categories affect racial disparities over time. The ability to assess disparities and outcomes within more targeted subgroups is important for improving equity by ensuring that the specific needs of these communities are visible and measurable, which is the first step towards meeting these needs. Understanding compositional changes within each of the broad racial categories is essential for accurately assessing progress and monitoring changes in disparities over time. The importance of this has been highlighted by research that convincingly demonstrates that changes in the ethnic composition within marginalized race groups (see e.g., Hamilton and Hummer, 2011) and/or of the White comparison category (Read, Lynch, and West, 2021) are responsible for a significant portion of the narrowing of racial disparities over time, suggesting that there has been less progress in reducing race-based disparities than was previously believed.

⁶ See Executive Order on Advancing Racial Equity and Support for Underserved Communities through The Federal Government, issued January 21, 2021. Available at: <https://www.whitehouse.gov/briefing-room/presidential-actions/2021/01/20/executive-order-advancing-racial-equity-and-support-for-underserved-communities-through-the-federal-government/>

Data disaggregation also plays an important role in improving data quality. In addition to allowing for the disaggregation of data by ethnic subgroups within racial categories, detailed ethnic subgroup data is often used to ensure that respondents of the same ethnic background are consistently classified within their expected racial category. This minimizes the impact of differential reporting of race by some ethnic subgroups who have difficulty finding themselves within the OMB racial categories (some of whom are discussed above in the Background section).

When these data processing procedures are applied inconsistently across data collection efforts (for example, between the U.S. Census or American Communities Survey and population health surveys), this can create inconsistencies in the classification and distribution of race and ethnicity between data sources that are not easy for users to measure or identify. Specifically, when only the minimum racial and ethnic information is collected, it is not possible to use ethnic subgroup information to identify those who have reported their race in ways that are inconsistent with the federal government's classification of ethnic subgroups. This means that these individuals will remain misclassified in the data. When this misclassification involves racial communities with small populations, such as Native Hawaiians and Pacific Islanders or American Indians and Alaska Natives, it can have a large impact on the measurement of these communities. Thus, collecting detailed race and ethnicity information allows individuals to be classified consistently and ensures that the data more accurately represent the experiences of not just ethnic subgroups, but also the minimum race categories.

Thus, the collection of detailed racial and ethnic data is important because data equity is a necessary first step toward a more just and equitable society, and also because it improves data quality and consistency over time. This in turn will ensure that this data can be used for its intended purpose: to identify and measure racial and ethnic inequities that pervade U.S. society.

Protecting respondents from data disclosure and identification is too often cited as a reason to avoid collecting detailed racial and ethnic data. We recognize that data collection efforts often occur within populations that have little representation from some racial and ethnic populations. Within these populations, the current minimum standards may pose a challenge, because data is likely to include an insufficient number of respondents from these small groups to make that data reportable. However, too often, these evaluations consider only the immediate need of the data collector and fail to consider the ways in which this information may be used to improve the overall quality of the data or could be combined with other data to produce usable estimates for these populations. This has wider ramifications for data quality and representation.

For example, when collecting mortality data, a number of states collect race and ethnicity using categories that do not meet the minimum guidelines in the 1997 Directive 15, because they combine the Native Hawaiian and Pacific Islander (NHPI) and Asian categories. Although the number of deaths within the NHPI population may not be independently reportable within the state, when combined with data from other states, this information is reportable at the national level. However, such national reporting is not possible because these states report Asian and NHPI data in aggregate. The importance of this information was highlighted during the COVID-19 pandemic where data from the subset of states that disaggregated the Asian and NHPI categories demonstrated that NHPIs consistently experienced the highest COVID-19 case and death rates. Because the Asian population is much larger than the NHPI population, in states that combined these categories, the elevated case and mortality rates of NHPIs were rendered invisible.

For these reasons, we recommend that revisions to OMB Directive 15 should consider each of the proposed changes by evaluating their impact on data equity. We propose the following data equity principles to guide this evaluation:

1. Ensuring that wherever possible, small racial and ethnic groups that are often underserved and rendered invisible in data collection efforts are able to be identified.
2. Decisions regarding which racial and ethnic (sub)groups are explicitly listed/included should not solely be based on relative population size, but also consider factors such as:
 - a. The likelihood that the unique experiences and history of the subgroup will mean that outcomes for that subgroup will differ from those of the larger group;
 - b. Whether members of a subgroup often misclassify themselves or are misclassified by others into broader racial or ethnic groups;
 - c. Whether data collected will or can be combined with other data, for example, state data combined within national data sets or the pooling across waves of data.
3. When data is reported, it should be reported at the most granular level possible; reporting practices should be reviewed to assess the possibility of relaxing statistical standards to enable reporting—with clear caveats regarding the quality of estimates.
4. The classification of racial and ethnic subgroups into intermediate reporting categories should reflect the interests and experiences of the communities being classified.

To the first point, we suggest that due to the geographic concentration of some racial and ethnic groups, the representation and measurability of different communities will differ across the country; however, data collection efforts are often stymied by lack of knowledge about how to expand data collection beyond the minimum standards without sacrificing comparability to national gold standards, such as the Census and the American Communities Survey. Therefore, OMB should not only provide minimum guidelines for data collection, but also guidance for how to broaden data collection efforts beyond these standards to allow a wider array of subgroups to be measured and seen.

To the second point, we acknowledge that the proposal to determine which racial and ethnic subgroups should be explicitly listed based on population size makes intuitive sense. These groups will be more likely to have a sufficient number of respondents to allow reporting of subgroup differences. However, the collection of ethnic subgroups serves multiple purposes, and not all of them are served by focusing on relative population size. First, to the extent that the experiences of the most common ethnic subgroup dominate our estimates for a broader category due to population size, identifying this group provides less additional information on the diversity of experiences within this category. Second, the most dominant groups within a category are the least likely to be misclassified by respondents, because they are more likely to be associated as the prototypical member of that group. Thus, their inclusion does not (necessarily) improve data quality.

Often, decisions regarding the collection of detailed race and ethnicity categories are based on whether an individual sample will yield sufficient numbers of respondents to allow reporting. However, with the growing push toward linking and pooling data, this concern may be short-sighted. Even when sufficient samples of ethnic subgroups are not available within a particular state or data collection period, more detailed data collection allows for this data to be linked and/or combined across states or years to augment sample sizes and report at a more disaggregated level. None of this is to argue that the inclusion of some of the more populous ethnic subgroups should not be considered, but to point out that there are other—equally valid—

criteria that can be used to determine which groups should be represented within a specific data set.

If detailed racial and ethnic data is collected, many data collection efforts will contain too few responses to allow data to be publicly reported or released for some racial and ethnic subgroups. To ensure that data can be reported as granularly as possible, it will be necessary for the federal government to develop a set of intermediate categories into which respondents can be “rolled-up” to allow consistent reporting. The classification system used in the 2020 Census Detailed Redistricting File may be a starting point for such a system; however, a number of communities have raised concerns about this classification system (for example, its classification of Hmong under East Asian instead of South Asian).

We also strongly recommend that the communities being classified be consulted in a way that allows the diverse perspectives within those communities to be heard rather than allowing a single individual or group to speak for all members of that group and that these efforts should be part of routine and sustained community engagement and outreach. As noted in the comment submitted by the NHPI Data Policy Lab at the UCLA Center for Health Policy Research, this should include “institutionalizing community advisory boards and appointing community leaders and subject matter experts with relevant knowledge, expertise, and experience working with and utilizing demographic data.” Such outreach will be crucial to ensuring accountability and transparency so that the data collected truly represents these populations (Ponce, Shimkhada, and Adkins-Jackson, 2023).

We understand that collection of detailed race and ethnicity information increases the burden of data collection and processing, but the benefits of improvements to data quality and data equity outweigh these costs, which are otherwise disproportionately borne by communities who remain invisible and underserved. It is easy for data managers to find reasons not to collect such detailed information, thus OMB should not follow a voluntary approach to the collection of detailed race and ethnicity information. We agree with the approach recommended by the NHPI Data Policy Lab that states: “If OMB were to allow federal agencies the ability to avoid utilizing the proposed minimum categories, we recommend that stringent requirements be put in place that require review of such decisions by community advisory boards, publication of the determination and supporting analysis, that the agency seek public comments, and require the agency to respond to the comments.”

All of these decisions affect and are affected by the ability to use and report data for these smaller groups. OMB should recommend that wherever possible, efforts should be made to report estimates at the most granular level possible, while still maintaining necessary privacy protections for respondents. Too often, statistical or privacy standards are used to excuse data collection efforts that do not meet minimum standards. It may be necessary to relax reporting or statistical standards for groups whose unique experiences would otherwise remain invisible. Clear guidelines on when and where this could occur would be welcome.

Specific Recommendations for Data Collection

The Combined Race and Ethnicity Question

We support the move to combine the race and Hispanic origin questions. Our review of results from the 2015 National Content Test, other testing conducted by the Pew Research Center (Patten, 2015), and the extant literature on Latinx racial identity and data collection (Pew Research Center, 2021; Pew Research Center, 2015; Garbarski et al., 2022), as well as our

long-standing experience in collecting race and ethnicity data for the California Health Interview Survey lead us to believe that a combined question will provide higher quality data by reducing both nonresponse and reporting of “another race” among the segment of the Latino population that find the current two-question format confusing and difficult to answer, while continuing to allow the disaggregation of the Hispanic/Latino category by race.

We recognize that the Afro-Latino community has legitimate concerns about this change; specifically, that:

1. the combined question could lead to an underreporting of the Afro-Latino population if these respondents are more likely to select a single response, and
2. the combined question could lead Afro-Latinos to be considered “multiracial” for the first time, which will result in this population being combined with other multiracial populations into a single heterogeneous category that will be rendered invisible through data tabulation decisions.

With respect to the first point, the results from 2015 National Content Test demonstrate that the single question format does not reduce reporting of Hispanic/Latino origins either overall or within the Afro-Latino population. However, one commonly noted limitation of the National Content Test is that it failed to oversample racial and ethnic minority populations in areas with a high-concentration of Afro-Latino residents, such as New York City. Thus, the test may not fully reflect the impact of this change on this population. We are sympathetic to this argument; however, we note that research conducted by Pew (Lopez, 2023) suggests that the current two-question format *already* significantly undercounts the Afro-Latino population. Additionally, their concerns about changes in reporting have been largely speculative; these critics have not provided data to suggest that the move to a single question will lead to a more substantial undercount of this population. In contrast, the evidence we do have suggests it will not.

One important caveat, however, is that racial and ethnic reporting practices are not only shaped by current understandings of race and ethnicity, but also shape the understandings of respondents in return. Current race and ethnicity reporting practices may be influenced by the long-standing practice of collecting Hispanic/Latino origin separately from race. This encourages (many) Latino respondents to consider their Hispanic/Latino origin as an ethnicity, separate from their race, and this could influence how they report on the testing of a combined question. Broader implementation of the single-question format could change the way in which the Latino population thinks about this distinction between their race and ethnicity, and subsequently change their reporting. For example, it could create a consolidated Hispanic/Latino racial identity within the Latino population, leading to fewer Latino/Hispanic respondents who also select another race category. This would reduce the ability of data users to identify the very real differences in outcomes by skin tone within the Hispanic/Latinx population. However, we note that colorism is not exclusive to the Hispanic/Latinx population; it is prevalent within every racial group, but rarely measured in practice.

This does not mean, however, that the ability to identify the Afro-Latinx population should be sacrificed to improve data collection among other Latinx subgroups. Thus, to reduce the likelihood that Afro-Latinos will change their racial and ethnic reporting categories, we recommend that the OMB guidelines should explicitly include an Afro-Latino subcategory under either (or both of) the Hispanic/Latino or Black/African American categories to ensure that these populations can continue to be identified and consistently coded in the data.

With respect to the second concern raised regarding the inclusion of Afro-Latinos within a heterogeneous multiracial category: although it will likely improve data quality overall, the shift to a single question format is also likely to lead to an increase in the proportion of the U.S. population that identifies with two or more racial and ethnic groups (Patten, 2015). If the current practice of combining all respondents who fall in this category into a single, heterogeneous group continues, this will minimize the political and social impact of the Latino population—as well as of smaller subgroups within the multiracial population. This will have a disparate impact on smaller racial groups, such as Native Hawaiians and Pacific Islanders or American Indians and Alaska Natives.

For this reason, we strongly recommend that if OMB implements a single question format for race and ethnicity, OMB must also issue race and ethnicity tabulation and reporting guidelines that minimize the impact of this change for Latinos who also report a race, as well as for other populations that report more than one race. We provide additional recommendations for these guidelines below in the section on tabulation of race and ethnicity.

Minimum Categories for Data Collection: Middle East and North Africa

Although historically individuals from the Middle East and North Africa (MENA) have been classified as White in the United States, this is largely a result of historical necessity. In the eras before the passage of the Civil Rights Act of 1964, MENA people could evade race-based laws that placed restrictions on non-Whites by declaring themselves and being legally recognized as White. While this status offered advantages at the time, today the inclusion of MENAs within the White racial category has rendered their experiences of exclusion invisible.

As discussed above, research clearly demonstrates that Middle Eastern and North African individuals do not view themselves as White and are not viewed as White by other Americans. In the 2010 Census, there was an organized effort within MENA communities to encourage MENA respondents to report their race as “other” to improve their ability to be seen and their minoritized status to be recognized. They routinely experience discrimination and racial- and ethnic-based mistreatment. The 1997 Directive 15 classification system prevents their experiences of discrimination from being identified and measured and encourages a significant percentage to report as “other” race.

For these reasons, we support the inclusion of a new MENA race category. We note, however, that there remains ambiguity regarding which nationalities and ethnicities should be classified within this category. To resolve this question, OMB should place considerable weight on the comments from members of this community. Further, because this category will be newly added, small in number, and the boundaries less familiar to respondents, it will be particularly important for the OMB guidelines to recommend that respondents be able to report ethnic subgroups to ensure that respondents are being accurately included in this category.

Minimum Categories for Data Collection: Asian

We recommend that the detailed ethnic subgroups listed for Asians be drawn from an expanded version of the minimum guidelines developed by the Department of Health and Human Services in response to Affordable Care Act Section 4302 to be adopted by OMB as the minimum standard for data collection. These categories disaggregate the Asian category into seven ethnic subgroups: Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, and Other Asian (Specify).

This set of categories has four clear benefits:

1. it encompasses the largest Asian subgroups in the United States,
2. it includes categories from multiple regions of Asia (South Asia, Southeast Asia, and East Asia);
3. it includes categories that are most often misclassified as Pacific Islander, such as Filipino; and
4. it includes an open-ended write-in option to identify those from ethnic subgroups not already listed.

As noted above, the Asian racial category includes people from a diverse array of nations, many of whom share little in common beyond their geographic region. If the goal of collecting race and ethnicity data is to identify differential treatment based on race or nation of origin, the diversity of nations represented within the Asian category makes it ill-equipped to do so. The inclusion of the largest Asian subgroups in the United States as separate reporting categories is a first step toward recognizing this diversity and assessing the disparate outcomes within this broadly defined racial group.

We recognize that the Asian population makes up a small proportion of the U.S. population and in 2021, over half of the nation's Asian population resided in five states: California (7.0 million⁷), New York (2.0 million), Texas (1.8 million), New Jersey (1.0 million), and Illinois (0.9 million). This means that it will not be possible to report disaggregated data within some individual states or national surveys with smaller sample sizes of Asian respondents. However, in addition to the benefits of disaggregating racial and ethnic categories, the collection of more detailed data can improve the consistency and quality of the collected data and allow for this data to be aggregated across states or pooled across years of data. This will not be possible if ethnic subgroup information is not collected.

The boundary between Asian and Pacific Islander nations is not always well-understood by respondents. For example, as noted above, although the U.S. government considers them to be Asian, it is common for Filipino respondents to classify themselves as Pacific Islanders. Although we agree that Filipino respondents are appropriately classified as Asian, there are a number of reasons—geographic, political, and cultural—that some Filipinos consider themselves to be Pacific Islanders rather than Asian; however, to the extent that racial and ethnic classifications are meant to be used to measure disparities between groups, it is important that members of a specific racial and ethnic group be classified consistently within the same category rather than based on their reported identity.

Filipinos are the third largest category of Asians, while Pacific Islanders comprise the smallest racial group within the United States. This means that misclassification of Filipinos in the Pacific Islander category can lead to biased estimates of disparities experienced by the Pacific Islander population. When data is collected using the 1997 Directive 15 minimum categories, it is not possible to identify Filipino respondents who report as Pacific Islander. Disaggregating the Asian category to identify the Filipino and other subpopulations makes it possible to ensure that respondents are classified consistently within racial categories. This improves data quality and

⁷ State-specific population estimates drawn from 2021 American Communities Survey data based on the respondents who report their race as Asian alone or in combination with another race. They are calculated using the Beta version of the U.S. Census Bureau's online API query tool.

comparability across data collection contexts—even when small sample sizes prevent disaggregation of these categories in public data and reports.

Finally, the diverse array of nations and ethnicities that fall within the Asian category mean that many ethnic subgroups will have too few respondents to allow for reporting. This makes it crucial to be able to create meaningful intermediate reporting categories for this group. The inclusion of an open-ended write-in category is necessary to ensure that such reporting categories can be generated. A closed-ended “other Asian” category would force data users to collapse all of these smaller ethnic subgroups together into a single heterogeneous group for whom the results would be difficult to interpret and prevent data users from grouping these respondents into the more meaningful categories.

Minimum Categories for Data Collection: Native Hawaiian and Pacific Islander

Although the 1997 OMB Directive 15 guidelines recommended treating Native Hawaiians and Pacific Islanders (NHPI) as a separate category, in practice, many states and other data collection efforts continue to combine the Asian and Pacific Islander categories. We strongly recommend that public data systems report disaggregated NHPI statistics as a separate race group and make disaggregated data available to the public for analysis. We also recommend the further disaggregation of NHPI subpopulations when possible, because there is great heterogeneity in sociodemographic factors across subgroups within the NHPI category.

The NHPI population encompasses a diverse group of subpopulations with considerable variation in sociodemographic and economic characteristics and in their political relationship to the U.S. government. For this reason, data collection efforts should focus on allowing respondents to report within detailed NHPI categories that can be classified as NHPI when sample sizes do not allow data disaggregation. At the minimum, we believe that OMB should follow guidance outlined in the comments provided by the NHPI Data Policy Lab housed at the UCLA Center for Health Policy Research. Respondents should be allowed to select from the following NHPI categories: Native Hawaiian, Samoan, CHamorro, Tongan, Marshallese, Fijian, and Other Pacific Islander (Specify).

Although each of these populations is small in number, they are fast-growing communities. As with the Asian racial category, the NHPI category includes a diverse group of subpopulations and collecting data using detailed ethnic subgroups will be crucial for ensuring that these subgroups can be reclassified using meaningful intermediate categories. The ability to “roll-up” these subgroups into meaningful intermediate categories depends on the inclusion of an open-text “Other Pacific Islander (Specify)” category.

It is even more imperative that these NHPI sub-groups have their data collected more accurately given that the two states with the largest NHPI populations in the U.S., California and Hawai'i, have reported that within the NHPI community, in 2020, COVID-19 cases and deaths impact non-Hawaiian Pacific Islanders to a much more severe extent than Native Hawaiians.

Recommendations on Changes to Terminology

The proposal to revise Directive 15 includes several changes to terminology within the definitions of the racial and ethnic categories that we support. These include:

1. The replacement of language referring to “people with origins in...” with references to “people who identify as...”

2. The renaming Native Hawaiian and Other Pacific Islander category to Native Hawaiian and Pacific Islander
3. The removal of “who maintain tribal affiliation or community attachment” from the definition of American Indian and Alaska Native
4. The replacing of “Far East” with “East Asian”
5. The discontinuation of the terms “majority” and “minority”

In addition, we support the recommendation from the NPI Data Policy Lab to update the spelling of “Chamorro” to “CHamorro.”

Recommendations for Data Cleaning and Processing

Although the proposed changes to the race and ethnicity data collection guidelines should improve the consistency and quality of race and ethnicity reporting, some respondents will continue to report in ways that will require data cleaning and processing. These cleaning and processing decisions are usually made internally and remain opaque to the public, but they can be important for understanding the composition of racial and ethnic groups, as well as the assumptions underlying the methods used to classify race and ethnicity.

Therefore, **we recommend that OMB require producers of public-facing data to develop documentation describing the ways in which their data cleaning and processing practices modify respondent reports of race and ethnicity, and we further recommend that they make this documentation readily available to the public.**

On the Use of Ethnicity Write-In Responses to (Re)classify Race and Ethnicity

As we have discussed in detail above, the use of detailed ethnic subgroup reports can be used to improve data consistency and ensure that the experiences of these subgroups are being accurately measured. We have touched on a few examples of where this could improve data quality. In general, when there is evidence that respondents have a different understanding of how a specific race or ethnicity should be classified, this type of data processing and cleaning can be useful and appropriate.

However, there are also circumstances when this type of reclassification may not be appropriate and may undermine the overall purpose of collecting race and ethnicity data. These occur when an individual reports having ancestry or origin from a racial and ethnic group that is not in alignment with their reported race. In some of these cases, the respondent is reclassified in ways that imply they are multiracial. The United States is a pluralistic society and many of those who have family who have lived in the U.S. for multiple generations can (accurately) claim multi-ethnic and/or multi-racial ancestry or origins, but many of these individuals do not (Mitchell et al., 2018). Increasing use of genetic testing has changed the way in which some respondents view their racial and ethnic identities. This in turn has changed how they report their race; those who have undergone genetic testing are more likely to report as multiracial and are more likely to report three or more races when provided the opportunity to do so (Johfre, Saperstein, and Hollenbach, 2021).

Some respondents take an alternate tack; they primarily identify with and report as a single race, but when asked to report their more diverse ancestry they provide this information. For the first time, the 2020 Census asked White and Black respondents to provide follow-up write-in responses describing their ancestry or origins. When these respondents reported ancestries or origins consistent with another racial category, their race reports were recoded to reflect this

multiracial background.⁸ In an effort to be consistent with the practices of the U.S. Census Bureau, whose population data is used to standardize racial and ethnic population distributions, the California Health Interview Survey introduced a similar detailed origin question for White and Black respondents in 2021. As they describe in greater detail in the comment letter submitted by CHIS, the results of this change suggest that the vast majority of respondents whose race was changed (92%) shifted from single-race White to multiracial White-AIAN.

The claiming of AIAN ancestry among White Americans has a long and fraught history and such reclassification should be taken only with an abundance of caution. It is not clear that these respondents were failing to report as AIAN because they did not understand the racial classification system. These respondents could have been reporting ancestral backgrounds that go back several generations and have little effect on how the individual currently views their own race, or on how others view their race. This ancestral background may be correct but could also be based on false or falsified family histories or on DNA tests that may not accurately identify genetic heritage. Their failure to report AIAN as part of their racial identity could reflect their distance from this ancestral background. In general, White Americans see their ancestry or ethnic origins as less central to who they are than other Americans (Cohen, Brown, and Lopez, 2021), so they may be less discriminating in which origins they report.

In part, the Census Bureau's reclassification of these respondents could be due to their interpretations of "origin" or "ancestry". On its face, such reclassifications appear to apply the "one-drop rule" broadly to race, which is problematic in a pluralistic society in which forced intercourse and rape were routinely employed as methods of asserting control and dominance over minoritized populations. It is also problematic because the purpose of measuring race and ethnicity is to be able to measure and assess the effects of race- and/or ethnic-based inequities in individual experiences. The reclassification of these respondents as multiracial AIAN when they do not report themselves to be AIAN can undermine that purpose. If these respondents identify as and are identified by others as White, their distant AIAN ancestry may have little effect on their current experiences. Moreover, because the size of the AIAN population is small, the inclusion of these respondents in this category may have an outsized impact on the measurement of disparities between AIANs and other populations, leading to an underestimate of the challenges they face.

For this reason, we recommend that data processing practices that result in the reclassification of respondents:

1. be transparent in order to allow affected communities to identify and comment on potential harms;
2. consider the impact of reclassification on small and minoritized groups;
3. distinguish between misreporting based on common misunderstandings of racial and ethnic categories and reporting of ancestry or origins that are not incorporated into self-reported race.

We further recommend that additional research be conducted to develop clear rules for the use of origin and ancestry data to reclassify race that are consistent with data equity goals of accurately capturing these populations.

Recommendations for Tabulation of Race and Ethnicity Responses

⁸ Please note that this reflects our understanding of conversations between CHIS staff and employees of the U.S. Census Bureau and is not drawn from citable documentation.

Asian Subgroups

Asian subgroups should be disaggregated to the greatest degree possible, but at the minimum, public-facing data should disaggregate the Asian category into four categories: Central Asian, East Asian, Southeast Asian, and South Asian. These broad categories demonstrate different patterns of health and mortality, as well as different histories and patterns of migration to the United States.

Intermediate region-based categories can be used to roll-up smaller disaggregated groups. The U.S. Census Bureau has previously used intermediate categories for smaller groups, including Central Asian, East Asian, Southeast Asian, and South Asian categories. However, the U.S. Census Bureau has provided little justification for inclusion in these categories. For example, the Hmong population has been classified as East Asian; however, given differing immigration patterns and sociodemographics of Hmongs in America compared to East Asians, this group is more appropriately classified among Southeast Asians.

We recommend the following grouping, which follows the 2020 Census Detailed Race tabulation with a modification of classifying Hmongs under Southeast Asians rather than under East Asians:

Central Asian: *Afghan, Kazakh, Kyrgyz, Turkmen, Uzbek*

East Asian: Chinese, Japanese, Korean, Taiwanese, Mongolian, Okinawan

Southeast Asian: Burmese, Cambodian, Hmong, Indonesian, Laotian, Malaysian, Thai, Filipino, Vietnamese

South Asian: Asian Indian, Bangladeshi, Bhutanese, Nepalese, Pakistani, Sri Lankan

These suggested lists are non-exhaustive but provide a starting point for guidance for classification of detailed Asian subgroups into intermediate categories. As noted above, we strongly recommend that OMB consult with community leaders to determine the appropriate classification of Asian subgroups.

Native Hawaiian and Pacific Islanders

NHPI race data should not be aggregated with Asian race data. NHPI race data should also be reported separately from the “other race” when possible. Although epidemiologists and statisticians often hesitate to report the small numbers for NHPs due to unstable rates or lack of reaching a certain statistical threshold, we contend that the data should be reported anyway, with caveats outlining the limitations of the data.

Health disparities were central to the creation of the revised OMB 15 disaggregated NHPI category as the categories federal agencies must use when collecting racial and ethnic data. However, for example, the CDC’s COVID-19 surveillance system currently depends on reporting from local and state health agencies that represent an inconsistent patchwork of data collection and reporting practices. The lack of uniformity and representation of NHPs in the data presents serious challenges to researchers attempting to accurately measure the depth of COVID-19’s impact. In the case of the COVID-19 pandemic, these numbers were essential in the early days of the crisis to mobilize grassroots community responses to the spread of the virus, even when the initial case numbers were low.

NHPIs and other relatively smaller populations have a higher transaction cost to access their own community's data. They may be required to pay data usage fees, to complete an application process, travel to a secure location where data are stored, and have advanced statistical knowledge in order to access their restricted data. Therefore, agencies should make efforts to lower these costs for communities like NHPIs who are underrepresented in data reporting, so that community researchers have equitable and ethical access to data.

Where possible, NHPI data should be disaggregated into subcategories. If data from detailed sub-categories must be restricted due to privacy requirements, intermediate subcategories that are already defined and utilized by the U.S. Census can be utilized with one modification. We recommend that OMB follow the recommendation of the UCLA NHPI Data Policy Lab and include a separate category for those from the Compacts of Free Association (COFA) countries due to their unique political relationship to the United States. The NHPI intermediate subcategories that data reporting systems could use would include:

COFA Countries: Federated States of Micronesia, the Republic of the Marshall Islands, and the Republic of Palau

Melanesian: Fijian, Melanesian, Papua New Guinean, Solomon Islander, Ni-Vanuatu

Micronesian: CHamorro, Mariana Islander, Carolinian, Kosraean, Pohnpeian, Saipanese, I-Kiribati, Chuukese, Yapese

Polynesian: Native Hawaiian, Samoan, Tongan, Tahitian, Tokelauan

These suggested lists are non-exhaustive but provide OMB a starting point for guidance for roll-up into intermediate categories. We strongly recommend that OMB consult with community leaders when developing this classification system.

Respondents Who Report More than One Race

The proposed changes to Directive 15, which include both the shift toward a single question format and the introduction of a new Middle East and North Africa (MENA) race category, are likely to increase the proportion of respondents who report as more than one race (Patten, 2015). Although a common practice is to classify all respondents who report more than one race into a single, heterogeneous category, this **cannot** be standard practice as the size and diversity of this category increases.

The creation of a combined multiracial category has an outsized impact on the smallest race groups: Native Hawaiians and Pacific Islands (NHPI) and American Indians and Alaska Natives (AIAN). As a result of colonialization practices, including forced relocation and integration, these populations are more likely than other race groups to identify as more than one race and/or Hispanic/Latino. Common race and ethnicity tabulation practices restrict racial groups to those who are non-Hispanic/Latino and single race. However, more than half of those who identify as NHPI or AIAN also identify as Hispanic/Latino and/or with another race, so the restriction of tabulations to non-Hispanic/Latino single race members of these groups artificially reduces the sizes of these two populations—usually by more than half.

Because NHPI and AIAN are the smallest racial groups, these restrictions often mean that data do not contain sufficient sample sizes to produce reliable estimates for these race groups, rendering them invisible in public-facing data. Unlike those in the largest racial and ethnic groups based on population, these communities must then expend additional resources to access restricted data that allows them to be identified. Thus, the use of a combined multiracial category is not benign. Nor does it only affect the NHPI and AIAN groups. This classification

system has also been used to limit the political power of Black Americans. In 2021, when defending newly drawn Congressional districts that reduced the influence of the state's Black population, the state of Alabama argued that multiracial Black residents should not be considered as part of the Black racial group but as part of a separate (smaller) multiracial category.

The combined question format is likely to increase the proportion of respondents who identify as more than one race, as well as increase the diversity of responses within this group. This means that if Directive 15 recommends a single question format, it is crucial that this Directive also include new instructions for tabulating data to ensure that this change in format is not used to further dilute the political, social, and economic power of people of color within the United States.

We recommend that OMB include guidelines that recommend that the distribution of race and ethnicity be tabulated in the following ways:

1. Using separate, overlapping racial and ethnic categories that count all individuals who report as each racial and ethnic group, either alone or in combination with another race and ethnicity, and when possible, divide each racial and ethnic group into two groups: "alone" or "in combination with another race and ethnicity";
2. If overlapping categories are not possible, priority should be given to ensuring that data from small racial and ethnic groups that are likely to have high proportions of respondents reporting more than one race—such as NHPI, AIAN, or MENA—can be reported for a combined "alone or in combination with another race and ethnicity" with other racial and ethnic groups reported as residual categories.

While computationally more complex, the use of overlapping categories will more accurately represent respondent reports and not artificially dilute the power of smaller racial and ethnic groups through tabulation decisions.

If you have any questions or would like to discuss this further, please contact Professor Ninez Ponce at nponce@ucla.edu.

REFERENCES

- Aziz, Sahar. Forthcoming. Legally White, Socially Brown: Racialization of Middle Eastern Americans. In *Routledge Handbook on Islam and Race* (ed. Zain Abdullah).
- Bonilla-Silva, Eduardo. 1997. "Rethinking Racism: Toward a Structural Interpretation." *American Sociological Review* 62(3):465–80.
- Carpinella, Colleen M., Chen, Jacqueline M., Hamilton, David L., Johnson, Kerri L. 2015. "Gendered Facial Cues Influence Race Categorizations." *Personality and Social Psychology Bulletin*, 41(3), 405-419.
- Cohn, D'Vera, Brown, Anna, Lopez, Mark Hugo. 2021. "Black and Hispanic Americans See Their Origins as Central to Who They Are, Less So for White Adults." Pew Research Center Report (May 10, 2021).
- Freedman, Jon, Penner, Andrew M., Saperstein, Aliyah, Scheutz, Mattias, Ambady, Nalini. 2012. "Looking the part: Social status cues shape racial perception." *PLoS ONE*, 6, e25107.
- Fuentes, Milton A., Reyes-Portillo, Jazmin A., Tineo, Petty, Gonzalez, Kenny, Butt, Mamona. 2021. Skin Color Matters in the Latinx Community: A Call for Action in Research Training, and Practice. *Hispanic Journal of Behavioral Sciences*, 43(1-2): 32-58.

- Garbarski, Dana, Dykema, Jennifer, Neman, Tiffany, Jone, Cameron, Schaeffer, Nora Cate, Farrar Edwards, Dorothy. 2022. Questioning Identity: How a Diverse Group of Respondents Answer Standard Questions about Ethnicity and Race. Presented at 2022 Annual Meeting of the American Association of Public Opinion Researchers in Chicago, IL.
- Gonlin, Vanessa. 2022. Mixed-Race Ancestry ≠ Multiracial Identification: The Role Racial Discrimination, Linked Fate, and Skin Tone Have on the Racial Identification of People with Mixed-Race Ancestry. *Social Sciences*, 11: 160.
- Hamilton, Tod G., Hummer, Robert A. 2011. Immigration and the Health of U.S. Black Adults: Does Country of Origin Matter?" *Social Science & Medicine*, 73(10): 1551-1560.
- Johfre, Sasha Shen, Saperstein, Aliyah, Hollenbach, Jill A. 2021. Measuring Race and Ancestry in the Age of Genetic Testing." *Demography*, 58(3): 785-810.
- Krosch, Amy R., Jost, John T., Van Bavel, Jay J. 2021. "The Neural Basis of Ideological Differences in Race Categorization," *Philosophical Transactions, The Royal Society, B*, 376, 20200139.
- Lee, J., & Ramakrishnan, K. (2020). Who counts as Asian. *Ethnic and Racial Studies*, 43(10), 1733–1756.
- Lopez, Alejandra. 2003. Collecting and Tabulating Race and ethnicity Data with Diverse and Mixed Heritage Populations: A Case-Study with U.S. High School Students. *Ethnic and Racial Studies*, 26(5): 931-961.
- Lopez, Mark Hugo. 2023. More than Two Centuries of Measuring Race and Ethnicity in the United States. Presentation the National Academy of Medicine for the Committee on Promoting a Culture of Health (March 2023).
- Lopez, Mark Hugo, Krogstad, Jens Manuel, and Passel, Jeffrey S. 2022. Who Is Hispanic? Pew Research Center Report.
- López, Nancy, Vargas, Edward, Juarez, Melina, Cacari-Stone, Lisa, Bettez, Sonia. 2018. "What's Your 'Street Race'? Leveraging Multidimensional Measures of Race and Intersectionality for Examining Physical and Mental Health among Latinxs." *Sociology of Race and Ethnicity*, 4(1): 49-66.
- Loveman, Mara. 2014. *National Colors: Racial Classification and the State in Latin America*. Oxford, UK: Oxford University Press.
- Ma, Debbie S. 2018. "Race Signaling Features: Identifying Markers of Racial Prototypicality among Asians, Blacks, Latinos, and Whites." *Social Cognition*, 36(6), 603-625.
- Mitchell, Laren L. Kathawalla, Ummul-Kiram, Ajayi, Alex A., Fish, Jillian, Nelson, Sarah C., Peissig, Lovey H.M., Syed, Moin. 2018. Ethnic-Racial Typicality and Its Relation to Ethnic Identity and Psychological Functioning. *Cultural Diversity and Ethnic Minority Psychology*, 24(3): 400-413.
- Naghbouleh, Neda, Schacter, Ariela, Flores, Renee D. 2022. Middle Eastern and North African Americans May Not Be Perceived, or Perceive Themselves, To Be White," *Proceedings of the National Academy of Sciences*, 119(7): e2117940119.
- Monk, Ellis P. 2022. "Inequality without Groups: Contemporary Theories of Categories, Intersectional Typicality, and the Disaggregation of Difference," *Sociological Theory*, epub ahead of print, 1-25. Available: https://scholar.harvard.edu/files/monk/files/monk_-_inequality_without_groups_-_stx.pdf
- Patten, Eileen. 2015. Who Is Multiracial? Depends on How You Ask: A Comparison of Six Survey Methods to Capture Racial Identity. Pew Research Center Report.
- Pew Research Center. 2015. Multiracial in America: Proud, Diverse and Growing in Numbers. Pew Research Center Reports.
- Pew Research Center. 2021. Majority of Latinos Say Skin Color Impacts Opportunity in America and Shapes Daily Life. Pew Research Center Reports.
- Ponce, Ninez, Scheitler, A.J., Shimkhada, Riti. 2016. Understanding the Culture of Health for Asian American, Native Hawaiian and Pacific Islanders (AANHPIs): What Do Population-

Based Health Surveys Across the Nation Tell Us About the State of Data Disaggregation for AANHPIs? Report for the Robert Wood Johnson Foundation. Available at:

<https://www.policylink.org/sites/default/files/AANHPI-draft-Report-9-262016.pdf>

- Ponce, Ninez, Shimkhada, Riti, Adkins-Jackson, Paris B. 2023. Making Communities More Visible: Equity-Centered Data to Achieve Health Equity. *Millbank Quarterly*, (forthcoming).
- Read, Jen'nan, Lynch, Scott Michael, West, Jessica. 2021. Disaggregating Heterogeneity among Non-Hispanic Whites: Evidence and Implications for U.S. Racial/Ethnic Health Disparities. *Population Research and Policy Review*, 40(1): 9-31.
- Roth, Wendy D. 2016. "The Multiple Dimensions of Race," *Ethnic and Racial Studies*, 39(8), 1310-1338.
- Ryabov, Igor. 2015. Educational Outcomes of Asian and Hispanic Americans: The Significance of Skin Color. *Research in Social Stratification and Mobility*, 44: 1-9.
- Saperstein, Aliyah, Penner, Andrew M. 2014. "Beyond the looking glass: Exploring fluidity in racial self-identification and interviewer classification." *Sociological Perspectives*, 57, 186-207.
- Telles, Edward, Paschel, Tianna. 2015. Who Is Black, White or Mixed Race? How Skin Color, Status, and Nation Shape Racial Classification in Latin America. *American Journal of Sociology*, 120(3): 864-907.