

ARTICLES

Proposed Changes in Federally Mandated Race and Ethnic Classifications May Undermine Their Original Purpose

Andrew A. Beveridge¹, Margo Anderson²

¹ Queens College and Graduate Center CUNY and Social Explorer, ² University of Wisconsin--Milwaukee

Keywords: Race, Ethnicity, US Census, Self-identification

<https://doi.org/10.29115/SP-2023-0027>

Survey Practice

Vol. 16, Issue 1, 2023

The Office of Management and Budget (OMB) is proposing an overhaul of "Statistical Policy Directive 15," commonly referred to informally as OMB15, the mandated method for collecting data on and classifying the race and ethnic status of the American population. Initiated in 1977, the standards for collecting data on race and ethnicity are mandated to be used by the Census Bureau and all federal agencies, and eventually all state, local, and private entities that receive federal funds, and/or come under civil rights compliance provisions designed to monitor disparities and correct discrimination. The classification was developed to monitor compliance with several civil rights laws affecting many sectors of American life that were passed in the 1960s and in later years. The widespread use and enormous implications of changes in how the United States collects race and ethnicity data mean that we should be very cautious about changes like the ones that have been proposed, and we think it would be wise to make these mandated data classifications both as simple as possible for the U.S. population to understand their meaning and to apply them in the many areas where they are to be used.

Introduction

The Office of Management and Budget (OMB) is proposing an overhaul of "Statistical Policy Directive 15," commonly referred to informally as OMB15,¹ the mandated method for collecting data on and classifying the race and ethnic status of the American population. Initiated in 1977, the standards for collecting data on race and ethnicity are mandated to be used by the Census Bureau and all federal agencies, and eventually all state,

¹ To make our discussion easier to follow, we have provided three web-based reference documents. The first includes a list of 11 civil rights acts that require monitoring of various sectors of the society, the standards for OMB15 adopted in 1977, those adopted in 1997, and those proposed in 2017 and in 2023, the relevant Census questions from 1980, 1990, 2000, 2010, and 2020, as well as forms and instructions for birth and death certificates, and example forms for employer monitoring, public school monitoring, police arrest monitoring, and traffic stop monitoring. The link for this document is https://www.dropbox.com/s/popvswtshzyk59t/Revised_Appended_Doc.pdf?dl=0. We also include Appendix F from the Technical Documentation for PL94-171, which includes detailed race and ethnic classifications. Its link is https://www.dropbox.com/s/o1io08jcsk79in/PL94-171_Appendix_F.pdf?dl=0. We also provide a table of race and Hispanic classifications, including those of race in combination for the 2010 and 2020 Census, the 2019 and 2021 American Community Survey, and the 2010 and 2020 Census Estimates. The estimates eliminate the "Some Other Race" category in a "modified race file." The link for this table is https://www.dropbox.com/s/icjyqoenlmt5rw/Ref_Table_Final_Census_ACS_Est.pdf?dl=0. Our discussion refers to each of these documents.

local, and private entities that receive federal funds, and/or come under civil rights compliance provisions designed to monitor disparities and correct discrimination.

The classification was developed to monitor compliance with several civil rights laws affecting many sectors of American life that were passed in the 1960s and in later years. The widespread use and enormous implications of changes in how the United States collects race and ethnicity data mean that we should be very cautious about changes like the ones that have been proposed, and we think it would be wise to make these mandated data classifications both as simple as possible for the U.S. population to understand their meaning and to apply them in the many areas where they are to be used.

How Does Monitoring Disparities and Discrimination Currently Function?

The Census Bureau collects basic data on the distribution of the American population by racial and ethnic group. An employer with 100 or more employees, or working on a federal contract, for example, reports the distribution of its labor force by race and ethnic group. The Census tabulations for the local labor market area provide the denominators for comparison with the counts reported by the employer. Thus, it is possible to compare the count of individuals from a particular racial or ethnic group residing in a specific region, possessing a defined set of educational qualifications and other credentials, with those holding the same job title but belonging to different racial and ethnic groups employed by a particular employer. Similarly, comparisons can be made for those served or employed by a specific business, those enrolled or accepted in a specific school, those who suffered from a specific medical condition, and how they were treated based on their race and ethnic status. Thus, one could easily see if there were disparities and if possible that discrimination existed. So long as the data continued to be collected in a comparable manner, it is possible to monitor and track any changes in such disparities or disproportions. But this only works to the extent that the Census Bureau data for the denominators is consistent with the tabulations produced by the agencies collecting the “numerator” data, or valid calculations of disparities are impossible.

The Proposed Changes and a History of Changes

The original categories were set in 1977 and modified slightly in 1997, but no new major groups were added, with the 1997 Statistical Policy Directive (SPD) providing the ability to check one or more races. The 1977 and 1997 SPDs called for two questions: one on a person’s race and the other on their ethnicity, defined as Hispanic or Latino, or not Hispanic or Latino. The current OMB effort continues much of the work started during the Obama administration to consolidate the two-question format into a single

question on race and ethnicity and add an ethnic category for Middle East and North Africa (MENA). These proposed changes were considered by the OMB; however, they proved to be controversial and were strongly opposed by several entities, including the Heritage Foundation. When Trump assumed the presidency, the proposed changes were not moved forward. The two-question format in the 1997 SPD remained in force. Combining the current race and ethnicity questions into a single question, as currently proposed, and continuing to allow for checking more than one box is expected to reduce the number of Hispanic respondents who in the past have checked Hispanic and then checked Some Other Race in the two-question format. The current approach to monitoring civil rights issues, counts all Hispanics of whatever race together, and generally assesses only the non-Hispanic race category. The proposed change to a single question format should not cause much difference in reporting but should be tested carefully, and a crosswalk should be provided.

The addition of the MENA category, however, represents a major change. In short, the proposal is to add a checkbox for MENA. Adding a checkbox will mean that some proportion of the population is expected to check it. Based on the answers to the ancestry question on the American Community Survey (ACS), about 1% of the population are expected to choose that answer. However, unless the MENA category is developed and explained in a manner that is broadly understandable, it may not yield useful data. The OMB's proposal does not state plainly which groups are to be included in the MENA category.² Though the current PL94-171 documentation (see note 1 above, Appendix F of the PL 94-171 technical documentation) classifies those from Turkey as White, the ACS ancestry question considers them Mideastern. In the initial content test from 2015 analyzing the MENA question (note 2), the bureau did not list a Turkish response as Mideastern, but enough respondents did so that they added it as a possibility for the research study.

Before a decision is made on adding the MENA category, more research is needed so that respondents, and question administrators beyond the Census Bureau, can administer the question (or assign individuals to the proper group) in a simple and unambiguous way. The current array of race groups in

² The Census Bureau's 2015 Race and Ethnicity Content Test defined MENA in a variety of diverse ways. As noted, "The Census Bureau's working classification of Middle Eastern and North African groups was geographically based and includes both Arab groups, such as Egyptian and Jordanians, and non-Arab groups, such as Iranian and Israeli. It also included ethnic groups from the region such as Assyrian and Kurdish. "The working classification of MENA included the following 19 nationalities: Algerian, Bahraini, Egyptian, Emirati, Iraqi, Iranian, Israeli, Jordanian, Kuwaiti, Lebanese, Libyan, Moroccan, Omani, Palestinian, Qatari, Saudi Arabian, Syrian, Tunisian, and Yemeni." These were supplemented to include transnational groups, as well as some countries suggested by stakeholders, as well as pan ethnic terms such as Arab, Middle Eastern, and North African." What exact groups would eventually be used was not decided; however, the number is expected to be approximately 1% of the U.S. population based on the ancestry groups in the 2021 ACS. See 2015 National Content Test: Race and Ethnicity Analysis Report (pp 21–22). Available at: <https://www2.census.gov/programs-surveys/decennial/2020/program-management/final-analysis-reports/2015nct-race-ethnicity-analysis.pdf>

the standard have stood the test of time and do include quite easily definable groups based upon historical patterns of discrimination. It is not obvious that this is currently true for MENA.

The OMB review must be concerned with the categories into which individuals are to be sorted, but beyond that, to be successful, it must also be constructed and administered in such a way that the mode of collection itself, including the way the responses are processed, does not get in the way of such consistency. Furthermore, since some data are based upon self-response (including the Census), some involve a third party assigning the information to individuals, while still others allow opting out of a response altogether, it is important that each method results in very similar distributions. In short, how the data are collected (the mode effects), the ways in which the data are elicited (demand and other effects), and the effects of recoding answers should all be tested and minimized.

Effects of Small Changes: The Case of the 2020 Census

Even slight changes in how these questions are presented can have enormous impacts, as seen in the results of the 2020 Census. Despite no official change in the OMB's classification, it appears that the Census Bureau reported race and Hispanic status of individuals based upon examining new open-ended responses and thus reported dramatic shifts in the proportions of Americans in various racial and ethnic categories. The new open-ended fill-in space for Whites and Blacks asked for additional information from some 300 million individuals. The Census Bureau scanned up to 200 characters of each fill-in using over 1,600 categories reproduced in the PL94-171 Technical Documentation, Appendix 7. If they found an open-ended answer that named a second race or a Hispanic/Latino response according to their coding guide, they then added that information to the respondent's first race, often creating a "multiple race" response. Even with over 1,600 responses, this approach lead to ambiguous results. Although a significant proportion of the coding guide highlighted national origin as the more detailed category, ethnic groups do not always necessarily follow such boundaries. For instance, many African and other countries have different groupings that are well known internally. Also, it is fully possible for one to be born or migrate from a given society and be of a variety of backgrounds.

The Federal Register Notice from the OMB asks for comments on a form using six checkboxes for all race and ethnic categories except American Indian or Alaskan Native. (See page 27 of the linked document.) Such an approach means there will be mode effects and demand effects that will affect responses. As with the MENA category, providing a checkbox or not providing a checkbox means that those groups without a checkbox on the form will have a lower number of reported respondents than if they had a checkbox. A natural experiment regarding this occurred in New York City with the 2000 Census, where the number of respondents who identified as Dominican

was undercounted by 150,000 out of the 650,000 found in other Census surveys.³ Providing no checkboxes, but only fill-ins for detailed classifications may also change the results. Some respondents will fill out the open-ended classifications, others will not or will not understand what such fill-ins are supposed to indicate. Further, those who fill out the Census questionnaire online may have an easier time using the fill-ins than those receiving a questionnaire by mail or filling it out with the help of an enumerator. At minimum, research needs to be conducted and training needs to be produced to assess the impact of the fill-in option. The extensive use of such options to modify the checked answers in the 2020 Census highlights the possible effects of having a novel approach to coding such answers. Further, the capacities of the Census Bureau to process complex responses may not extend to other agencies, though the goal of the statistical policy directive is consistent collection and reporting across many different users.

The results of these question and coding changes mean that the 2020 Census race and Hispanic categories are not consistent with earlier data collections or, indeed, any of the time-series analyses used for tracking the data. Changing the questions again in the 2030 Census will only exacerbate the issue.

The effect of this change in collecting and coding race and ethnicity data in 2020 can be seen by comparing the results from the 2019 and the 2021 ACS. The 2019 used the old method of race and Hispanic coding identical to the Census method from 2010 while the 2021 used the new method of coding. When one compares the number of multi-race responses in both administrations, one cannot disentangle what was caused by the new methods and what was caused by demographic change. Comparing the two ACS administrations indicates that most of the differences between the two can be attributed to the methodological changes.⁴

Overall, the ACS showed a 1.1% population growth rate from 2019 to 2021. However, the bureau reported 33.5 million fewer people in the single race White population. The reported “two or more” race population went up by over 30.5 million, the “some other race” population by over 7.5 million. When split by Hispanic and non-Hispanic, the non-Hispanic single race population declined by over 4 million. Those reported as “two or more” race had increased by almost 6 million and those reported as “some other race” by just over one million. For the Hispanic population, those reported as single race White had declined by just under 29.5 million, while those reported as two or more races had increased over 24.5 million, and those reported as “some other race” alone had increased by over 6.5 million. These dramatic changes were due to subtle changes in the question format and

³ See *New York Times*, June 27, 2001, Janny Scott, “A Census Query is Said to Skew Data on Latinos.”

⁴ The reference table provides information comparing the 2010 to 2020 Census, the 2019 to the 2021 ACS and the 2010 to the 2020 Census Estimates. See note 1.

changes in how the results were coded. Adding a MENA category will not only move answers that may have been placed in White, “some other race” or some combination of races to the MENA category but will also affect the distribution of the other groups. See accompanying table for more details, and for similar comparisons for the 2010 and 2020 Census and 2010 and 2020 Census estimates.

The bureau has acknowledged the issue in its reports on the racial and ethnic distributions in the 2020 Census, noting that the changes “could be attributed to several factors, including demographic change since 2010. But we expect they were due to the improvements to the design of the two separate questions for race and ethnicity, data processing and coding, which enabled a more thorough and accurate depiction of how people prefer to self-identify.” To date, however, the bureau has not produced a crosswalk disentangling the impact of demographic change and methodological format change on the reported numbers.⁵ Nor has the bureau produced information on exactly how they recoded responses and/or added new races to those who had checked off only one race.

Recommendations

Based upon this review and the information we have to date from the 2020 Census, we make the following recommendations:

If the new single-question format or the inclusion of the MENA category is adopted, research studies should be done to see what the impact of such a change would have on the Census distribution of answers and on the ability of non-Census question administrators (or those doing the classification) to administer the same standard questions in a consistent manner. Such studies should include the development of clear instructions for all agencies and entities that will be charged with using the new classification, and specific examples of how that question should be tabulated and presented. Both the new single-question format and inclusion of the category MENA would require documentation for those who do not professionally collect data but will be mandated to do so. Training, not just for those collecting data at federal level, but for all entities involved should be provided.

In addition, the construction of a crosswalk will be crucial, since many uses of this classification (e.g., for birth certificates, initial school enrollment, and initial employment) are not amenable to change without a complete reclassification of millions of individuals. As such, it is important to be able to compare responses given earlier to those being collected. Such a crosswalk should be created before implementation of the new criteria. This

⁵ See U.S. Census Bureau, “Improved Race and Ethnicity Measures Reveal U.S. Population Is Much More Multiracial,” <https://www.census.gov/library/stories/2021/08/improved-race-ethnicity-measures-reveal-united-states-population-much-more-multiracial.html>, accessed June 7, 2023.

would make it possible to assess the efficacy of any changes and ensure that consistent time series analysis of this classification was possible. If it was not plainly consistent, other approaches to the modification of SPD15 should be assessed.

Finally, it is vital that the Census and the OMB working group do further research on the best way to solicit the more complex racial and ethnic identifications from residents of the United States. This includes asking whether more complex and nuanced responses on the Census or any other data collection might affect the distribution of responses to the basic categories. If such an effect was likely, they should seriously consider less intrusive changes. The more complex pattern of ethnicity and race could be studied and monitored either using the ACS or a special supplement to the ACS for such a purpose. Using an open-ended question for this purpose for a census of over 330 million residents seems an especially heavy and expensive burden, especially for a project that is not mandated, and may cause severe difficulties in continuing to implement mandated monitoring of disparities and civil rights among ethnic and racial groups who have been historically disadvantaged.

.....

Corresponding author

Andrew A. Beveridge, Queens College and Graduate Center CUNY and Social Explorer, andy@socialexplorer.com

Submitted: June 20, 2023 EDT, Accepted: October 12, 2023 EDT