

**INTERIM REPORT TO THE OFFICE OF MANAGEMENT AND BUDGET:
Review of Standards for Maintaining, Collecting, and Presenting
Federal Data on Race and Ethnicity**

*Prepared by the Federal Interagency Working Group for
Research on Race and Ethnicity*

<u>I. MANDATE AND SCOPE OF REVIEW</u>	1
<u>A. Background</u>	2
<u>B. Summary of initial proposal</u>	2
<u>II. REVIEW PROCESS</u>	6
<u>A. Guiding Principles</u>	6
<u>B. Areas Identified for Review</u>	7
<u>C. Data Sources Consulted</u>	9
<u>D. Statutory Need, Operational Feasibility, and Public Cost</u>	9
<u>III. INTERIM FINDINGS</u>	13
<u>A. Question Format and Nonresponse</u>	13
<u>B. Classification of Middle Eastern or North African (MENA) Race/Ethnicity</u>	21
<u>C. Additional Minimum Reporting Categories</u>	26
<u>D. Relevance of Terminology</u>	30
<u>IV. CONCLUSION</u>	36
<u>Appendix 1: Implementation of Current Standards on the Separate Questions</u>	37
<u>Appendix 2: Combined vs Separate Question: Existing Research on Non-response in Surveys and Administrative Data Collections</u>	40
<u>Appendix 3: Detailed Analysis of Subgroup 1 Research Questions on Combined Question</u>	43
<u>Appendix 4: Detailed Analyses on a New MENA Category</u>	52
<u>Appendix 5: Detailed Results of Analyses on Relevance of Terminology</u>	56

I. MANDATE AND SCOPE OF REVIEW

Under the *Paperwork Reduction Act (PRA)*, OMB is required to ensure the efficiency and effectiveness of the Federal Statistical System as well as the integrity, objectivity, impartiality, utility, and confidentiality of information collected for statistical purposes. OMB also develops and oversees the implementation of Federal-wide principles, policies, standards, and guidelines concerning the development, presentation, and dissemination of statistical information.

OMB’s *Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity* (<https://www.gpo.gov/fdsys/pkg/FR-1997-10-30/pdf/97-28653.pdf>) were developed to provide consistent and comparable data on race and ethnicity throughout the Federal government for statistical and administrative programs. The standards were developed using

OMB's established, objective process to produce statistical policy directives, which includes Federal technical evaluation, public comment, and expert statistical analysis and were last updated in 1997.

A. Background

Since the standards were last revised, the ways in which people in the U.S. self-identify their race and ethnicity may have evolved. To ensure that measures of race/ethnicity remain relevant for policy making purposes, in 2014 OMB formed an Interagency Working Group for Research on Race and Ethnicity (Working Group) to exchange research findings, identify implementation issues, and collaborate on a shared research agenda to improve Federal data on race and ethnicity. The Working Group comprises representatives from ten cabinet departments and three other agencies that collect or use race and ethnicity data.

Through its systematic review of the implementation of the 1997 revisions across Federal and state governments, and examining stakeholder feedback, the Working Group identified particular areas where revisions to the 1997 standards could potentially improve the quality of race and ethnicity information collected and presented by Federal agencies. On September 30, 2016, OMB requested public comment on the principles that should govern its work and the areas initially identified for review. (See <https://www.regulations.gov/document?D=OMB-2016-0002-0001>).

Specifically, the Working Group proposed further exploration in four areas. Subgroups were formed to prepare initial proposals for the Working Group's consideration.

1. The use of separate questions versus one combined question to measure race and ethnicity and question phrasing as a solution to race/ethnicity question nonresponse;
2. The classification of a Middle Eastern and North African group and distinct reporting category;
3. The description of the intended use of minimum reporting categories; and
4. The salience of terminology used for race and ethnicity classifications and other language in the standards.

Over the 30 day period, 3,750 comments were received from the public. After reviewing these public comments and continuing its analysis, the Working Group developed this Interim Report, which describes the Working Group's progress to date and requests further public comment. In this Interim Report, first, the principles guiding the Working Group's review (Section IIA) and the results of the September 30's Federal Register Notice (Section IIB) are summarized. Second, the review process, taking into account public comment, empirical analysis (Section IIC), statutory requirements and public burden (Section IID), is described. This is followed by an in-depth discussion of each area identified for review (Section III). The report concludes with initial proposals and requests for further public comment. Additional detail on the analyses is included in the attachments.

B. Summary of initial proposal

1. Separate Questions or Combined Question for Race and Ethnicity

The current standard of Separate Questions calls for **collection** of information with a question first asking ethnicity (Hispanic or non-Hispanic) and then in a separate question asking race. Self-reported information collected in Federal surveys more often follows this standard than non-Federal administrative data collections. However, Federal statistics often

are not **reported** using the separate categories. The two major barriers to fully implementing the standards for collection seem to be: (1) how agencies calculate race when there is no response to race or “other” is selected, and (2) the limitation of non-Federal administrative data systems in accommodating a “mark all that apply” approach that allows respondents to select multiple races. Barriers on the reporting side seem to be the size of the sampled population and the associated risk to respondent confidentiality and statistical reliability when the size of the race group is small in the sample. Federal agencies have tried to achieve the spirit of the current standards despite these barriers, but challenges remain. However, the use of a Combined Question format in certain circumstances, where race and ethnicity are not asked separately, may address nonresponse challenges if sufficient data quality and comparability can be assured across the Federal government. The subgroup will explore these approaches further before preparing a proposal for the Working Group’s consideration.

2. Classification of Middle Eastern or North African (MENA) Race/Ethnicity

Since the Federal standards on race and ethnicity first were issued in 1977, the Federal government has classified White as people having origins in any of the original peoples of Europe, the Middle East, or North Africa. As part of periodic review of the standards during the mid-1990s, OMB considered several suggestions to improve the accuracy and reliability of Federal race and ethnicity statistics. Among the suggestions proposed was the inclusion of an additional, distinct minimum reporting category for respondents identifying as “Arabs or Middle Easterners.” At the conclusion of that review, agreement could not be reached among public stakeholders on the intended measurement concept (i.e., whether the category should be based on language, geography, etc.) nor on a definition for this category. As a result, an additional, minimum reporting category for this group was not created. Instead, the *1997 Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity* advised that further research be done to determine the best way to improve data for “Arabs/Middle Easterners.”

Based on public comment and analyses to date (see the Census Bureau’s 2015 Forum on Ethnic Groups from the Middle East and North Africa and a review public comments on *Proposed Information Collection; Comment Request; 2015 National Content Test (12/2/2014)*), the Subgroup proposes that a Middle Eastern or North African classification be added to the standards and be geographically based. The MENA classification should be defined as: A person having origins in any of the original peoples of the Middle East and North Africa. This includes, for example, Lebanese, Iranian, Egyptian, Syrian, Moroccan, Israeli, Iraqi, Algerian, and Kurdish¹. However, many questions remain that require additional input from the public before the Working Group as a whole would adopt this proposal and recommend a change in the standard to OMB.

¹ The rationale for using these examples is to include the two largest Middle Eastern Arab nationalities (Lebanese and Syrian), the two largest North African Arab nationalities (Egyptian and Moroccan), and the two largest non-Arab nationalities within the Middle Eastern / North African region (Iranian and Israeli) as the first six examples. This is followed by the next largest Middle Eastern Arab nationality (Iraqi), the next largest North African Arab nationality (Algerian); as well as an example of a transnational, non-Arab group (Kurdish).

The current standards already allow for the breakout of *racial* subcategories from the required minimum standards when the quality of the collected data can be assured. However, many public commenters signaled that an *ethnic* classification for the MENA population should be added because those with roots, origins, and heritage from the Middle East and North Africa are racially diverse. Although results from the Census 2015 National Content Test show that those in the Census Bureau’s working classification of MENA largely identify with the MENA response category when it is available, the 2015 test did not break out MENA as an *ethnicity*. Rather, it was presented as a checkbox where respondents could check multiple boxes, such as MENA and White or MENA and Black or just MENA, all with the ability to also write in a country of origin or ethnicity.

In addition, some of the groups proposed for inclusion under a MENA classification were also ethnoreligious groups. A challenge to ethnicity measurement can be the intersection of ethnicity with religious affiliation. The race/ethnicity standards are not intended to measure religion (see P.L. 94-521), and it is unclear how to address inclusion of ethnoreligious groups while clearly maintaining the intent and use of the resulting measure as not indicating religion. Further, the cost and burden of requiring this additional reporting category when race/ethnicity is measured across the Federal government is unclear. Also, the size of the overall group is small as a percentage of the total population, and reporting out the data could encounter many of the problems encountered with other small population groups such as Native Hawaiians and Pacific Islanders in surveys or data collections where there are small samples. These cost and quality issues need to be explored further.

3. Additional Minimum Reporting Categories

During OMB’s review of Federal standards on collecting and reporting data on race/ethnicity in the 1990s, approximately 10,000 letters and postcards were received on the issue of classifying data on Native Hawaiians. The Interagency Committee recommended to OMB that data on Native Hawaiians continue to be classified in the Asian or Pacific Islander category. This recommendation was opposed by the Hawaiian congressional delegation, the 7,000 individuals who signed and sent preprinted yellow postcards, the State of Hawaii departments and legislature, Hawaiian organizations, and other individuals who commented on this recommendation. Based largely on evidence presented at public hearings in Hawaii, OMB did not accept the Interagency Committee recommendation and decided to break apart the Asian or Pacific Islander category into two categories—one called “Asian” and the other called “Native Hawaiian or Other Pacific Islander.”

The initial review of the 1997 standards did not identify additional, minimum reporting categories for detailed race/ethnicity groups as an element for evaluation. However, during the public comment period for September 30, 2016’s Federal Register Notice, the Working Group received more than 1,200 comments expressing the need for further disaggregated data for Asian communities and Native Hawaiian or Other Pacific Islander communities. Other comments express a similar need for disaggregated data, including 10 comments advocating for the disaggregation of the Black or African American category.

There are numerous examples of Federal agencies collecting detailed race and ethnicity data in their statistical reporting; these are not limited to decennial censuses or the American Community Survey (ACS). Nonetheless, OMB has learned that the minimum reporting categories as described in the current standards are often misinterpreted as the only

permissible reporting categories rather than allowing more detailed categories, provided that detail group data can be aggregated in a systematic way to allow for comparison across data sources.

The Subgroup proposes that rather than changing the standards, OMB should issue guidelines for the collection of detailed data for American Indian or Alaska Native, Asian, Black or African American, Hispanic or Latino, Native Hawaiian or Other Pacific Islander, and White groups for self-reported race and ethnicity collections. By providing these guidelines, comparable detailed race and ethnicity data will be supported across Federal agencies. However, many questions remain, such as under what conditions detailed data should *not* be collected. In addition, the Subgroup will continue to discuss whether OMB should require or, alternatively, strongly support but not require Federal agencies to collect detailed data.

4. Relevance of Terminology

Although many respondents report within the race and ethnicity categories specified by the 1997 standards, the standards themselves may not be well understood by the public. Over time, some terms may be preferred over others. For example, from comments received, the Subgroup noted some confusion about the opportunity to select more than one category, and about the use of the terms of race, origin, and ethnicity. Some respondents do not identify themselves within current Federal definitions of the race and ethnicity minimum categories.

Additionally, references to geographic location in descriptions of race and ethnicity categories may be incomplete, unclear, inconsistent, or confusing. Given both the low prevalence of these geographic locations appearing as write-in responses, as well as public comments to the September 30, 2016 Federal Register Notice advocating against use of geographic regions in race/ethnicity group definitions, the Subgroup recommends no changes be made to the current standards to specifically incorporate the following geographic locations into any existing race or ethnicity category: Australian (including the original people of Australia/the Aborigines), Brazilian, Cape Verdean, New Zealander, Papua New Guinean.

The Subgroup observed a lack of clarity in the classification of several groups in the current standards. Further, there is an error in the 1997 OMB classification standards with “Cuban” being listed twice in the reporting category of “Hispanic or Latino.” The Subgroup noted that population size currently does not govern the listing of detailed groups across all minimum reporting categories but recommends that population size should be used to determine which duplicate initial category mention of “Cuban” should remain. The Subgroup also considered whether the current ordering of the classification list should be updated to reflect current population size. As a next step, the Subgroup plans to apply this rationale to the classification listing and determine the magnitude and benefit of any resulting changes.

The Subgroup noted that some terms used in the current standards to describe race/ethnicity may not (or no longer) resonate with the public, such as “Negro” and “Far East.” Because respondents may find certain terms to be outdated and/or offensive, the Subgroup recommends that these terms be removed from the standards.

The Subgroup examined reporting patterns of South and Central American Indian

respondents to determine if self-identification and reporting behavior is consistent with the current standard, which includes them in the American Indian or Alaska Native reporting category. The Subgroup noted more research and public input would be necessary before adding specific South or Central American subgroups to the current description of the AIAN category

The Subgroup considered the use of “Principal Minority Race” in the current standard. Given the changes in the US population and the language in some legislation and Executive Orders, the Subgroup examined whether the current standards still refer to and provide guidance on reporting “principal minority race,” and if the term is still meaningful, whether the designation should be guided by minority group population size, historical disadvantage, or some other principle. The Subgroup considered whether “principal minority race” be expanded to include ethnicity. In particular, the Subgroup considered whether referring to Black or African American as the “principal minority race” is still relevant, meaningful, accurate, and acceptable. The Subgroup also plans to consider if “Hispanic” should be among the groups considered, and the salience of alternative terms (e.g., “principal minority race/ethnicity”).

The remainder of this interim report provides detail supporting the initial findings and proposals of each Subgroup.

II. REVIEW PROCESS

Within the Working Group, Subgroups were formed to identify areas for possible revision; review public comments; conduct empirical analyses of potential improvements; and consider statutory requirements and potential public burden and cost. The Subgroups prepared initial proposals for consideration by the Working Group as a whole, and, subsequently, by OMB. Each Subgroup comprised Federal statisticians and/or Federal policy analysts. Several agencies were represented in each Subgroup, and Subgroup co-chairs facilitated work processes. Each Subgroup prepared its own analysis plan, which was shared and discussed across the Working Group.

A. Guiding Principles

In the deliberations leading to the 1977 standards, principles were established to guide interagency consideration. These principles were further elaborated and framed the 1997 revisions.

On September 30, 2016, OMB requested input from the public regarding principles that should govern the Working Group’s review. With one exception (noted below), no substantive responses regarding these principles were received. Accordingly, the Working Group adopted the principles listed below for the current review.

1. The racial and ethnic categories set forth in the standards should not be interpreted as being genetic, scientific, or anthropological in nature. For the purposes of these standards, race is a socio-political construct.
2. Respect for individual dignity should guide the processes and methods for collecting data on race and ethnicity; respondent self-identification should be facilitated to the greatest extent possible.
3. To the extent practicable, the concepts and terminology should reflect clear and generally understood definitions that can achieve broad public acceptance.

4. The racial and ethnic categories should be comprehensive in coverage and produce compatible, non-duplicated, exchangeable data across Federal agencies.
5. Foremost consideration should be given to data aggregations by race and ethnicity that are useful for statistical analysis, program administration and assessment, and enforcement of existing laws and judicial decisions, bearing in mind that the standards are not intended to be used to establish eligibility for participation in any Federal program.
6. While Federal data needs for racial and ethnic data are of primary importance, consideration should also be given to needs at the State and local government levels, including American Indian tribal and Alaska Native village governments, as well as to general societal needs for these data.
7. The categories should set forth a minimum standard; additional categories should be encouraged provided they can be aggregated to the standard categories. The number of standard categories should be kept to a manageable size, as determined by statistical concerns and data needs.
8. A revised set of categories should be operationally feasible in terms of burden placed upon respondents and the cost to agencies and respondents to implement the revisions.
9. Any changes in the categories should be based on sound methodological research and should include evaluations of the impact of any changes not only on the usefulness of the resulting data but also on the comparability of any new categories with the existing ones.
10. Any revision to the categories should provide for a crosswalk at the time of adoption between the old and the new categories so that historical data series can be statistically adjusted and comparisons can be made.
11. Because of the many and varied needs and strong interdependence of Federal agencies for racial and ethnic data, any changes to the existing categories should be the product of an interagency collaborative effort.
12. ADDITIONAL PRINCIPLE: Any Middle Eastern or North African classification or category, if proposed, should be constructed in a manner that adheres to public law.²

The Working Group recognizes that these principles may in some cases represent competing goals for the standards. Through the review process, it is necessary to balance public sentiment, empirical analysis, statutory data needs, social concerns, cost, and public burden.

B. Areas Identified for Review

On September 30, 2016, OMB also requested public comment on the areas identified by the Working Group where revision to the current standards might improve the quality of Federal data on race and ethnicity. (See [LINK](#)) Specifically, these four areas are:

1. The use of separate questions versus a combined question to measure race and ethnicity and question phrasing;
2. The classification of a Middle Eastern and North African group and distinct reporting category;
3. The description of the intended use of minimum reporting categories; and
4. The salience of terminology used for race and ethnicity classifications and other language in the standards.

Over the 30 day period, 3,750 comments were received from the public through

² This principle has been expressed in certain other circumstances. For example, PL 94-521 enacted in October 17, 1976 states, in part, "(c) Notwithstanding any other provision of this title, no person shall be compelled to disclose information relative to his religious beliefs or to membership in a religious body."

regulations.gov and/or Race-Ethnicity@omb.eop.gov. Comments sometimes pertained to more than one topic. By areas identified for review:

- Question format and nonresponse issues: 98 comments
- Measurement of Middle Eastern or North African (MENA) race/ethnicity: 2,289 comments
- Disaggregation of data during collection or presentation: 1,297 comments of which 1,247 related to requiring additional minimum reporting categories for the Asian race/ethnicity and the Native Hawaiian or Other Pacific Islander race/ethnicity
- Relevance of terminology used in the current standards: 46 comments
- Out of scope: 102 comments
- Unrelated to current standards: 61 comments

Stakeholder engagement was supported prior to and concurrent with the standards' review. For example:

- On July 10, 2014, OMB attended the National Association of Latino Elected and Appointed Officials³ (NALEO) Summit "Ensuring an Accurate Count of the Nation's Population."
- OMB participated in U.S. Census Bureau's 2015 Forum on Ethnic Groups from the Middle East and North Africa, held on May 29, 2015.⁴ At this forum of community members, academics, organization leaders, and other experts, feedback was received on testing a Middle Eastern or North African category in the 2015 National Content Test.
- The White House Initiative on Asian Americans and Pacific Islanders (WHIAAPI) held a conference entitled Elevate: AAPI Data Challenge and Convening on October 7, 2016. At this convening, OMB described and answered questions about September 30, 2016's Federal Register Notice.⁵
- WHIAAPI held a conference call for OMB to answer any questions from the larger WHIAAPI stakeholder community. This conference call was held on October 14, 2016.
- OMB and the Working Group conducted a webinar describing the intent and established process for the 2016-17 review of Federal race and ethnicity data standards. A link to the webinar was posted on <https://www.youtube.com/watch?v=s-II36wS9xE>.
- A Frequently Asked Questions document was posted on the September 30, 2016 Federal Register Notice page at www.regulations.gov on October 27, 2016.

Although the initial examination of the 1997 standards did not identify additional, minimum reporting categories for Asian detailed groups and Native Hawaiian or Other Pacific Islander detailed groups as an element for evaluation, in recognition of the public comments received, the Working Group is now reviewing this proposal.

³ <http://www.naleo.org/>.

⁴ U.S. Census Bureau. (2016). 2015 Forum on Ethnic Groups from the Middle East and North Africa. Washington, DC: Buchanan, A., Marks, R., Alvarez, M. Retrieved from: <http://www.census.gov/library/working-papers/2015/demo/2015-MENA-Experts.html>

⁵ Retrieved from: <http://sites.ed.gov/aapi/elevate-aapi-data-challenge/>.

C. Data Sources Consulted

In addition to consulting research literature on measuring race/ethnicity, and drawing upon their own professional expertise, the Working Group members also contributed to the design of ⁶ and/or analyzed results from a number of information collections by the individual agencies. The data sources listed below were used as the basis of much of the Working Group's empirical analyses.

The **American Community Survey (ACS)** is a detailed annual update of the decennial census that provides data required by Federal agencies to administer programs and distribute Federal funds to states and localities. Because of the ACS, we know more about jobs and occupations, educational attainment, veterans, commuting patterns, whether people own or rent their homes, and other topics. Federal, State, and local governments, as well as businesses, use this information to assess the past and plan the future. Among the demographic information collected is race/ethnicity and national origin. See <https://www.census.gov/programs-surveys/acs/>.

The **2010 Census Race and Hispanic Origin Alternative Questionnaire Experiment (AQE)** focused on improving the race and ethnicity questions by testing a number of different questionnaire design strategies. The primary research objectives were to design and test questionnaire strategies that would increase reporting in the 1997 standards' racial and ethnic categories, lower item nonresponse, increase accuracy and reliability of the results, and elicit reporting of detailed racial and ethnic groups. See https://www.census.gov/2010census/pdf/2010_Census_Race_HO_AQE.pdf.

The **2015 National Content Test (NCT)** was the Census Bureau's primary mid-decade opportunity to compare different race/ethnicity question design strategies prior to making final decisions about the 2020 Census content. The 2015 NCT included a nationally representative sample of 1.2 million housing units in all fifty states and Puerto Rico. Housing units in areas with relatively high populations of people who identify as Middle Eastern or North African, American Indian or Alaska Native, Asian, Native Hawaiian or Pacific Islander, Black, or Hispanic were oversampled to ensure a diverse sample with representation from a variety of race and ethnic groups. The survey was available in English and Spanish. The 2015 NCT also included a re-interview component where a sample of respondents were asked a series of questions about their racial/ethnic background with the goal of ascertaining an accurate picture of the respondent's racial/ethnic identity. See <http://census.gov/programs-surveys/decennial-census/2020-census/planning-management/final-analysis/2015NCT-race-ethnicity-study-plan.html>.

Additionally, based on Working Group analysis, a follow up study has been initiated by the National Center for Health Statistics to provide **further cognitive testing of Middle Eastern and North African classification**. The Subgroup plans to make the results of that study publicly available prior to OMB's decision regarding possible revisions to the standards.

D. Statutory Need, Operational Feasibility, and Public Cost

In addition to its review of public feedback and empirical research, the Working Group also

⁶ For example, contributions were made to the design of upcoming cognitive testing of Middle Eastern or North African categories and further question format testing within the 2017 Census Test.

considered operational feasibility, cost, and public burden in its deliberations. These concerns merit general consideration because they must be considered to responsibly provide proposals that, on balance, best reflect the public’s interest.

1. Statutory and Program Need

The data are used for congressional reapportionment, legislative redistricting, and enforcement of the Voting Rights Act. Following each decennial census, congressional reapportionment—the redistribution of the 435 seats in the U.S. House of Representatives among the 50 States—is calculated using the population totals for each State and the formula of “equal proportions” adopted by the Congress in 1941 (United States Code, Title 2, Section 2a). Redistricting is the process of redrawing the boundaries of congressional, State, and local legislative districts in accordance with applicable law. Changes to the standards could affect congressional reapportionment and compliance in one-person/one-vote in redistricting to the extent that such changes affect the overall response to the decennial census or other requirements regarding the enforcement of the Voting Rights Act of 1965.⁷

Agencies that use these data for monitoring and enforcing civil rights laws include the Equal Employment Opportunity Commission, the U.S. Commission on Civil Rights, the Civil Rights Division of the Department of Justice, the Office of Federal Contract Compliance Programs in the Department of Labor, the Office for Civil Rights in the Department of Education, and the Office for Civil Rights in the Department of Health and Human Services. State and local governments, educational institutions, and private sector employers use the categories when providing data on race and ethnicity to meet Federal reporting requirements. Changes to the standards could affect the evaluation of the provision of services to the full public.

2. Operational Feasibility

New ways of reporting race and ethnicity could require implementation guidance and bridging studies so that the public can assess the magnitude of changes resulting from any revisions to the standard to existing statistical time series--some that may go back 20 years.

Regarding data collection, revisions to survey forms may prove minimal for agencies that conduct surveys of individuals or households, or other data collections that allow for respondents to identify multiple answers to the combined race/ethnicity question. However, changes to administrative data collections would tend to be more complex, with both general implementation and systems challenges. Federal administrative data collections often are based on record keeping systems maintained by State, local, or other governmental organizations, or by institutions, businesses, or other entities. This is an efficient and cost-effective means for leveraging collected data to meet Federal needs.

However, local and State administrative records are not required to conform to the Federal standards. A change in the Federal data collection standards will require complementary revisions throughout the system in order for the data to be internally comparable. The 1997 OMB standards on race/ethnicity took more than 10 years to implement in the National Center for Education Statistics’ Common Core of Data school survey. The revisions required changes to the internal data collection systems maintained by the Nation’s nearly 100,000

⁷ See 52 U.S.C. § 10301 (2015).

public schools, as well as by their school districts and State education agencies. Additionally, in many systems, race or ethnicity may be entered into the system in one period when an individual enters the program or system, and maintained within the system, and reported out by the system for a period of years. As an example, a student may be enrolled in an education system, open a record with one race or ethnicity group, and the record may not be modified to take into account new race or ethnicity reporting standards. Although the ease with which information technology systems may be altered may have improved over the past two decades, the scope of the systems affected and the associated public cost must be considered.

Another example of the complexities of administrative reporting of race/ethnicity data is vital records. The recording of a vital event is governed by State law, regulation, and policy. The National Center for Health Statistics (NCHS), manages the National Vital Statistics System (NVSS), a national demographic and event data set compiled from data recorded on birth and death certificates, including 4 million births and 2.5 million deaths annually. The authority for the methods and procedures for the registration of deaths lies with the States, and the Federal government depends upon the States to enact laws and regulations that provide methods of registration and data collection comparable from State to State. As a result, the production of national vital statistics involves long-standing collaboration and cooperation across many levels of government and other entities. Any proposed change to the collection of race and ethnicity data would require substantial involvement and collaboration with the States over a considerable period of years, because changes to the Federal collection of race and ethnicity cannot be mandated for State activities that are under the jurisdiction of the State. Inconsistency between self-reported Federal survey information and reporting on administrative records results in discrepancies between major sources of information for the Nation, which could be made worse by changing the standards

For example, the Census Bureau is dependent upon the State vital records to produce population estimates that serve as the foundation for Federal allocations and programs. NCHS provides the Census Bureau with birth and death counts for use in their Post-Censal Estimates Program, which forms the basis for a majority of national official statistics. The Census Bureau is constrained by the State-provided information on the birth and death certificates when combining these data with other data to estimate the population each year.

In another example, eligible community health centers receive partial funding from the Department of Health and Human Services (HHS) and are required to report to the Federal government on services provided to patients by race and ethnicity using the current Federal standards. However, they are not required to conform to the Federal standards when collecting this information from patients at the point of origin. Reconciling differences in measurement at point of origin with Federal reporting standards will remain a challenge as the use of administrative records to meet Federal data needs continues to expand.

If OMB were to make revisions to the 1997 standards, it would, according to customary practice, indicate a date of effect. It would also, according to customary practice, provide implementation guidance to Federal agencies. In doing so, the Working Group would, as its similar group did in the 1990s, work to ensure the Federal statistical system had adequate

implementation guidelines on how, for example, time series and longitudinal data would continue, despite the potential for a break in series.

3. Public Cost

Revisions to the standards for data on race and ethnicity would require Federal agencies to make changes to data collection forms, computer programs, interviewer training and coding manuals, and other related materials for their data systems. These costs may be considerable, depending on the magnitude of revision.

Actual published cost estimates are few. The National Center for Education Statistics (NCES) estimated that it would take 25 million person-hours to complete implementation of the 1997 standards in all 50 states.⁸ The statement on fiscal effect⁹ in a proposed bill to collect additional granular race and ethnicity data across the State of California for Asian detailed groups cited a cost of approximately \$2.7 million, require four to five additional staff across State agencies, and the updating of at least 60 paper forms to be compliant.

Federal standards would apply and costs would accrue to the Federal agency to adjust State-provided administrative data to conform to the Federal standards. This situation exists today but could be made more complex and costly by changing the standards.

In cases where local and State agencies and private sector entities follow the Federal standards for collection, record keeping, and presentation of data on race and ethnicity, costs to localities and States for revising these systems should also be considered. As an example, the National Association of Public Health Statistics and Information Systems (NAPHSIS) estimated a onetime cost of \$65 million for all States to reengineer systems as necessary to implement the 2003 revision of the birth certificate, including changes necessary to conform with the 1997 revision to the Federal race/ethnicity standards.¹⁰

Although financial costs and public burden would be incurred if changes were made to the standards, other types of costs may be associated with not making changes leading to improvement in data quality. Problems that exist with use of the current standards would not be resolved. These continuing problems include the diminished utility of race/ethnicity information where nonresponse occurs and when individuals are misclassified; substantial processing costs for addressing nonresponse or misclassification of race/ethnicity questions; suboptimal interpretation of evidence used to inform policy making; and possibly misallocation of funds and failed protection under the law from discrimination.

⁸ National Forum on Education Statistics, Race/Ethnicity Data Implementation Task Force. (2008). *Managing an Identity Crisis: Forum Guide to Implementing New Federal Race and Ethnicity Categories (NFES 2008802)*. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

⁹ State of California. Assembly Committee on Appropriations, Chair Lorena Gonzalez. *Assembly Bill 1726, Chapter 607*. Date of Hearing April 27, 2016. Retrieved from: http://www.leginfo.ca.gov/pub/15-16/bill/asm/ab_1701-1750/ab_1726_cfa_20160425_162753_asm_comm.html

¹⁰ National Center for Health Statistics. (2007). *Report to the NCHS BSC on the Review of the NCHS Natality Statistics Program*. Hyattsville, Maryland.

III. INTERIM FINDINGS

Each Subgroup considered public comments received following the September 30, 2016’s Federal Register Notice and reviewed the current standards and its implementation. Taking these into account, each Subgroup then posed specific research questions to inform an assessment of the resulting information quality, should specific proposed revisions occur. The results of these analyses are presented here, organized by review area. Initial proposals and requests for further public comment appear in the last section of this report.

A. Question Format and Nonresponse

1. Current Standard

According to the current standards,¹¹ "Self-reporting or self-identification using two separate questions is the preferred method for collecting data on race and ethnicity. In situations where self-reporting is not practicable or feasible, the combined format may be used." When using two Separate Questions, ethnicity (i.e., Hispanic or Latino; Not Hispanic or Latino) should be collected before race (i.e., American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White). When using a Combined Question, limited to cases such as observer-collected data on race and ethnicity—"both race (including multiple responses) and ethnicity can be collected when appropriate and feasible," according to the current standards. The standards also state that selecting one category is acceptable if using a Combined Question (i.e., American Indian or Alaska Native, Asian, Black or African American, Hispanic or Latino, Native Hawaiian or Other Pacific Islander, White). In such cases, if a respondent is observed as Hispanic, a “race” does not need to be imputed. See Tables 1 and 2, below.

Table 1. Current Separate Questions Format

ETHNICITY	Hispanic or Latino
	Not Hispanic or Latino
RACE	American Indian or Alaska Native
	Asian
	Black or African American
	Native Hawaiian or Other Pacific Islander
	White

Table 2. Current Combined Question Format

RACE OR ETHNICITY	American Indian or Alaska Native
	Asian
	Black or African American
	Hispanic or Latino
	Native Hawaiian or Other Pacific Islander
White	

¹¹ The “two, separate questions” approach established in the 1997 OMB standards was informed by quantitative research conducted by the former Interagency Committee for the Review of the Racial and Ethnic Standards as part of its systematic review of the 1977 standards. At that stage of review, the option for respondents to identify more than one race/ethnicity was not offered. The former Interagency Committee concluded that a combined race and Hispanic origin question would result in lower reporting in the Hispanic origin category alone compared to when a separate Hispanic origin question was employed. Since the implementation of the 1997 revisions, more research has been conducted to understand the quality of data on race and ethnicity, and much experience has been gained about the functioning of the two Separate Questions approach and the Combined Question approach. Thus, the current Working Group agreed that further examination of the two Separate Questions approach versus the Combined Question approach was warranted.

Federal statistical standards not only provide guidance on the collection of race and ethnicity information; these standards also provide guidance on the calculation of nonresponse in information collections. However, the coding procedures of race and ethnicity data that are collected is not standard across Federal data collections. The Working Group has observed that different approaches to coding item nonresponse may exist across Federal agencies. The coding of data on race and ethnicity may impact the distribution of groups in the specific categories which are used to make population estimates. A review of different coding processes across the Federal statistical system has not been fully explored, but is of interest to the Working Group for future consideration.

Implementation of the Current Standards in Statistical Agency Collections

Appendix 1 describes the implementation of the current standards as conducted for the Census Bureau, National Center for Health Statistics, Bureau of Justice Statistics, and National Center for Education Statistics data collections, and concludes with a summary of the implementations. Of particular interest is that in addition to the minimum race categories set forth by OMB, the Census Bureau is required by Congress¹² to allow respondents to select a category called “Some Other Race.” Currently, to make data comparable across other Federal data collections which do not have this category, the Census Bureau effectively assigns the “Some Other Race” response to another race such as White, Black or African American, Asian, etc. If a Combined Question format for self-reported race were permitted when the category of “Some Other Race” is required, it is anticipated that fewer cases would be coded as “Some Other Race” where ethnicity is reported but race is not.

Summary of the Implementation of the Current Standards

The current standard of a Separate Questions format for **self-reported** information is more often followed in the collection of Federal survey data than in either the use of non-Federal administrative data collections or in the reporting of Federal statistics. The barriers to fully implementing the standards seem to be the agency-determined acceptability of imputation to address nonresponse (in which case, “other” categories are offered), and the limitation of non-Federal administrative data systems in accommodating a “mark all that apply” approach. Regarding reporting of data, the barriers to full implementation of the current standards seem to be related to sample size and the associated risk to respondent confidentiality and statistical reliability. Federal agencies have explored various means to achieve the spirit of the current standards despite these barriers, but challenges remain. Understanding agency-specific coding practices will be an important component of evaluating potential solutions, particularly regarding conformance to reporting standards. However, the wider use of a Combined Question format could also be considered to address nonresponse challenges to information collection, particularly when “Some Other Race” is a required reporting category.

2. Public Comments

September 30, 2016’s Federal Register Notice requested public comment on the overall merit of revising the current standards to require the use of a combined race/ethnicity question, rather than the current requirement to use separate race/ethnicity questions except in circumstances where collection must occur through observation (for example, funeral parlors, school administrators, or criminal justice service providers). Comments were also requested regarding optimal question phrasing.

¹² H.R. 2862, 109th Cong. (2005-2006).

THEME 1. Combined Question support—67 Comments

These comments noted the following about the Combined Question:

- Represents an equitable approach for obtaining data on all racial and ethnic groups, ensuring that data on the Hispanic or Latino populations would be presented as having equal value as data regarding the other five race groups;
- Better reflects the OMB principle for “respecting individual dignity” by recognizing the right to self-identify as the respondent sees fit;
- Allows for the most accurate and detailed data possible;
- The Census Bureau’s 2010 AQE research and 2015 NCT research represent sound methodological research to support the revision; and
- Comparability and bridging of data over time, consistent tabulations of data by Federal agencies, and clear protocols and guidance for data users to follow are concerns.

Comments in support of the Combined Question format also noted the conceptual complexities of race and ethnicity and that combining the Hispanic ethnicity into the racial category is a better reflection of the experience of some populations, and therefore, will be more sensitive to those persons who struggle with such self-identification.

THEME 2. Accuracy and data concerns—11 comments

These comments did not support either the Combined Question format or the two Separate Questions format. Rather, the commenters expressed:

- Support for whichever format could increase the response and “participation” of their constituency;
- Noted the complexity of race and ethnicity and urged for continued testing and feedback of question formats, especially with other population groups like Native Hawaiians, Pacific Islanders, American Indians, and Alaska Natives;
- Comparability and bridging of data are concerns; and
- OMB should establish clear guidance if a change in the question format is implemented. Specifically, the issue of teasing apart actual demographic shifts versus artifacts of survey methodology were raised.

THEME 3. Combined Question non-support—10 Comments

These comments did not support the Combined Question format. The comments indicated that:

- Race and ethnicity were distinct constructs and that “lumping” all Hispanics into one group was not reflective of OMB’s guiding principle of “respecting individual dignity”;
- The Hispanic community has no consensus when it comes to self-identification and that moving to a Combined Question would lose valuable details that are available with the Separate Question format; and
- Changing the question format would result in substantial costs to Federal agencies and would take years to implement. This additional cost may not be offset by gains in accuracy as there were still be an amount of item nonresponse.

There was also discussion of how to report ethnicity if a Combined Question were used, as several Federal agencies currently issue statistics regarding ethnicity, regardless of race (e.g., Hispanic alone rather than Hispanic, White and Hispanic African American or

Black, etc.).

THEME 4. Civil rights/divisive and confusing categories—Six Comments

The remaining comments touched on issues of civil rights enforcement and the confusing and general divisive nature of categories. One comment provided detail about a concept of “ethical accuracy for civil rights” in which the question format should shed light on structural/systematic discrimination and inequities. Another comment provided detail about “racialization,” the visual dimension that relates how meanings are assigned to physical characteristics, and the importance of ancestry in addition to race and ethnicity. The other four comments all addressed that disentangling ethnicity from race is problematic and is not beneficial to society. Commenters felt that classification and categorization is divisive at a time when our country needs to come together. There was also confusion as to why “Hispanic or Latino” is the only ethnic group and why there appears to be no interest to collect information on other ethnicities and ancestries.

3. Research Questions

The Subgroup examined two research questions related to question format and response:

Research Question 1:

- For Federal **self-response** data collections that include a "Some Other Race" response option, including censuses and surveys, how does question format affect the distribution of collected race/ethnicity data? How does question format affect the distribution in the reporting of self-response race/ethnicity data?
- Do these effects persist in **self-response** data collections that do not offer a "Some Other Race" response option?

For Federal **administrative** data collections, how does question format affect the collection and reporting of race/ethnicity data?

Research Question 2:

- In **self-reported** surveys where “some other race” is offered as a response category, does a combined race/ethnicity question provide more reliable responses than the current two separate questions format?
- In **self-reported** surveys where “some other race” is not offered as a response category, does a combined race/ethnicity question provide more reliable responses than the current two separate questions format?

4. Overview of Extant Research

Item nonresponse to questions on race and ethnicity was analyzed across several current Federal agency data collections. Overall, these analyses found that item nonresponse to a single combined race/ethnicity question and to two Separate Questions on race and Hispanic origin varies widely across Federal statistical agencies, and even within agencies across data collections and types of data collection (e.g., self-report vs. administrative collections). However, the opportunity to make direct comparisons between the two Separate Questions approach and the single Combined Question approach with Federal surveys is limited, as relatively few agencies collect the data with both design approaches. Recent research from the Census Bureau found that a Combined Question approach in the decennial census may result in lower item nonresponse for race/ethnicity data collections. However, the format for the census differs from that of other Federal agencies in that it includes the category “some other race,” and this format may also play a contributing role to the effectiveness of a

Separate versus Combined Question approach to reducing nonresponse. A detailed look into existing research across Federal agencies is included in Appendix 2.

The Federal Statistical System is highly decentralized. To achieve its goals of producing accurate, objective, and relevant official statistics, the 127 agencies of this system must coordinate closely. As mentioned previously, the Census Bureau uses vital record information from States, supplied through NCHS, as components in the production of intercensal population estimates. But Census Bureau also uses administrative data from the Immigration and Naturalization Service on legal immigrants, and information on persons admitted to the U.S. as refugees from the Office of Refugee Resettlement. For internal migration statistics, the Census Bureau relies upon information provided by the Internal Revenue Service and the Social Security Administration. To the extent that nonresponse to race/ethnicity questions occurs in these information collections, estimates produced by many Federal agencies can be affected.

The Subgroup is interested in examining the impact of non-response on Federal surveys and administrative data collections, such as whether and how non-response for race and ethnicity may impact an agency's ability to perform its mission or carry out programs or policies. These impacts have not been fully explored beyond the Census Bureau's data collections.

5. Data Sources and Methodology

As previously mentioned, the primary source of quantitative data for the analyses described in this Federal Register Notice is the 2015 National Content Test (NCT) conducted by the Census Bureau on 1.2 million households in all 50 states and Puerto Rico. The sample units for the NCT were weighted to reflect the complex sample design and to adjust for nonresponse. In the 2015 NCT tables shown below, pairwise t-tests were performed to compare the rates of consistency, reporting, or item nonresponse between the question formats except where noted. Additionally, multiple comparison corrections were used to maintain the family-wise error rate at $\alpha = 0.05$.

We briefly describe how "Hispanic" race responses were coded in the NCT in the Separate Questions approach and how they were coded in the Combined Question approach. This coding reflects procedures used at the Census Bureau, and may differ from those used by other Federal agencies.

Hispanics Who Do Not Report a "Race"

Coding in a Separate Question Format: If a respondent self-reports as Hispanic to the separate Hispanic question, and does not provide an answer to the separate race question, his/her response is coded as Hispanic for ethnicity and race is imputed. If a respondent selects Hispanic to the Hispanic question, and writes in "Hispanic" to the race question, the response is coded as Hispanic and Some Other Race, per the Census Bureau's unique provision for such a category. Both ethnicity and race responses are required for the two Separate Questions approach, per the 1997 standards, for the Census Bureau's reporting needs. (The Census Bureau must produce a count to provide needed denominators for other statistics; therefore, allowing "missing"

responses cannot be permitted.¹³⁾

Coding in a Combined Question Format: For the purposes of this test, the Census Bureau applied the current reporting standards for a combined race/ethnicity question (currently, only permitted which is when self-report is not feasible). That is, if a respondent selects Hispanic and a “race” is not observed, the respondent’s race/ethnicity is coded as Hispanic *only* and a race is *not* imputed; a response regarding race OR ethnicity is considered complete.

Hispanics Who Also Report a “Race”

Coding in a Separate Questions Format: If a respondent selects Hispanic and one or more specific “race” groups (e.g., White, Black, Asian) within the separate race question, his/her response is coded as Hispanic for ethnicity *and* all selected races are reported (e.g., White, Black, Asian). So, for example, if a respondent selects as Hispanic and Black, the response is coded as both Hispanic and Black.

Coding in a Combined Question Format: The same is true within a Combined Question approach. If a respondent identifies as Hispanic and also reports a specific “race,” the response is coded as Hispanic, and all selected races are reported. So, for example, if a respondent selects as Hispanic and Black in the Combined Question, the response is coded as both Hispanic and Black.

In addition to reviewing the 2015 NCT research, the Subgroup also reviewed item nonresponse rates for race and ethnicity across Federal surveys. For example, the Subgroup conducted analyses using administrative data from the Bureau of Justice Statistics and the National Center for Education Statistics.

6. Result of Analyses

The detailed analyses for both Research Question 1 and Research Question 2 are included in Appendix 3.

Research Question 1:

a) For Federal **self-response** data collections that include a “some other race” response option, including censuses and surveys, how does question format affect the distribution in the reporting race/ethnicity data? How does question format affect the distribution in the reporting of self-response race/ethnicity data?

b) Do these effects persist in **self-response** data collections that do not offer a “some other race” response option?

c) For Federal **administrative** data collections, how does question format affect the collection and reporting of race/ethnicity data?

The Subgroup considered whether proposed solutions for different self-response question formats would improve, have no effect, or reduce observed accuracy of race and ethnicity data collection and reporting. The anticipated impact of proposed solutions on data flows within the

¹³ For details on the modification procedures used in this process for 2010 Census and Census 2000, see <http://www2.census.gov/programs-surveys/popst/technical-documentation/methodology/modified-race-summary-file-method/mrsf2010.pdf>.

Federal statistical system were also considered. These considerations include the potential impact on race/ethnicity population estimates, which are used by nationally representative surveys for sample design, sampling fractions, and post-stratification weights. For the purposes of the present inquiry, the Subgroup focused its analysis on data accuracy, including (by association) nonresponse. The Subgroup also noted the importance of considering operational, burden, and cost impacts when proposing its recommendation to OMB; the Subgroup plans to review this aspect more fully upon receipt of requested public comments.

The Subgroup also examined administrative data collections. Federal agencies are required to collect and report persons of Hispanic origin as a separate category regardless of race, but this does not impede Federal agencies from also reporting Hispanic origin by identified race. The Separate Questions format, even if, as administered by a non-Federal database, would only allow for a single response to the racial question, would still allow statistical agencies to capture and report race by ethnicity. Particularly for small racial groups that may show high rates of Hispanic ethnicity (such as persons identifying as American Indian/Alaska Native), the racial information would be lost in a database that forced a single response to a combined race/ethnicity question. In addition, based on their knowledge of the population or understanding of the database, statistical agencies could make reasonable imputations for missing race or ethnicity data under the two Separate Questions format if either of the questions is missing.

Research Question 2:

*a) In **self-reported** surveys where “Some Other Race” is offered as a response category, does a combined race/ethnicity question provide more reliable responses than the current two Separate Questions format?*

*b) In **self-reported** surveys where “Some Other Race” is not offered as a response category, does a combined race/ethnicity question provide more reliable responses than the current two Separate Questions format?*

The Subgroup examined Federal self-response data collections that include a “Some Other Race” response option, including censuses and surveys. To examine whether survey self-responses are more reliable when using a Separate Questions format or a Combined Question format where a “Some Other Race” category is offered, the Subgroup compared rates of consistency between responses to the 2015 NCT self-response survey and the 2015 NCT re-interview. Consistency was compared across all current minimum OMB reporting categories, the “Some Other Race” category, and a distinct reporting category for Middle Eastern or North African. The Subgroup also examined distributions of consistency in reported race among Hispanics only when a Separate Questions approach or a Combined Question approach is used.

The Subgroup also examined rates of item nonresponse by the Separate Questions and the Combined Question format. The Combined Question format has one item nonresponse rate for the race/ethnicity question, whereas the Separate Questions approach has three different item nonresponse rates that can be analyzed: one for the Hispanic origin question, one for the race question, and one that considers whether there is a response to both questions.

The analyses conducted to date of self-response data collections have not included collections that do not feature a “Some Other Race” category. It is not clear that the patterns observed when a “Some Other Race” option is offered extend to those collections where it is not offered.

Further, given that two of 127 Federal agencies in the Federal statistical system are permitted to offer a “Some Other Race” response category through special, limited provisions, it would be important to better understand the magnitude and scope of the issue when considering anticipated benefits of changes to the existing standard.

Therefore, the Subgroup plans to extend its current research findings by examining the accuracy of self-reported race/ethnicity collected using a Separate Questions format versus a Combined Question format in the context of surveys that do not feature a “Some Other Race” response option.

7. Summary of Initial Evaluation

The Subgroup’s review of current agency practices to collect and report data on race/ethnicity has identified challenges faced by some agencies with the implementation of the current standards. The Subgroup also identified challenges anticipated if the current standards were revised from a Separate Questions format to a Combined Question format. The public comments received to date also articulated both of these concerns, with the public generally noting that a Combined Question approach resonates with personal conceptions of race/ethnicity. (That is, most commentators thought there was no basis to distinguish between race and ethnicity.) However, concerns were also raised regarding the anticipated operational feasibility and cost for implementing this change across all agencies and data collections, particularly among Federal commentators.

Analyses to date suggested that collecting these data using a Combined Question may improve information quality for some respondents in some information collections, such as the decennial census, but the magnitude and scope of that improvement is still under investigation. For respondents selecting Hispanic, the improvements to response consistency and accuracy are statistically significant, albeit in some tests comparatively small and not observed for other minimum reporting groups. However, these results may apply most readily to self-reported collections offering a “Some Other Race” response category—limited to certain collections conducted by the Census Bureau and NCHS. Further, higher accuracy for both the Separate Questions and the Combined Question format was observed when the survey was conducted by Internet compared to paper. And, the results do not seem to generalize easily to the collection of race/ethnicity through administrative data collections—a method on which many Federal agencies, including the Bureau of Justice Statistics, the National Center for Health Statistics, and the National Center for Education Statistics, rely heavily, and which is increasingly common as administrative records are used more routinely to generate Federal statistics. It is clear, however, that both the magnitude and scope of anticipated benefits and costs of a Combined Question format must be considered, as well as possible alternatives that would not require a revision to the question format for every agency in order to address the issues identified.

a) Initial Plans: The Subgroup plans to continue its review of current Federal agency practices to determine whether or how a revised question format might improve the collection, tabulation, and utility of racial/ethnic statistics for Federal programs and policies. The Subgroup Group also plans to review current and potential solutions to nonresponse and the reporting of race and ethnicity for surveys that do not provide a “Some Other Race” response option.

b) Request for Public Comment: To assist in its deliberations, the Subgroup requests public comment on the following questions. Thinking about how information is collected:

1. What factors should be considered when evaluating anticipated information quality? Should both magnitude and scope (that is, the majority of collections) be considered? Should magnitude of the improved information outweigh the scope of the improved change, or vice versa? What amount of improvement would be considered meaningful?
2. What factors should be considered when evaluating anticipated feasibility? Should burden to local, State, and Federal agencies be considered? What amount of burden would caution against implementing a change? How should potential lags in data delivery be weighed?
3. What factors should be considered when evaluating anticipated cost of implementing a change? Should costs be weighed differently when experienced at a local, State, or Federal level? How should the costs of improving or failing to improve information quality be considered?
4. When considering information quality, feasibility, and cost, how should benefits and costs be weighed? In which cases would information quality outweigh feasibility and cost concerns? In which cases would feasibility and cost concerns outweigh information quality?

B. Classification of Middle Eastern or North African (MENA) Race/Ethnicity

1. Current Standard

Since the Federal standards on race and ethnicity first were issued in 1977, the Federal government has classified White as people having origins in any of the original peoples of Europe, the Middle East, or North Africa. As part of periodic review of the standards during the mid-1990s, OMB considered several suggestions to improve the accuracy and reliability of Federal race and ethnicity statistics. Among the suggestions proposed was the inclusion of an additional, distinct minimum reporting category for respondents identifying as “Arabs or Middle Easterners.” At the conclusion of that review, agreement could not be reached among public stakeholders on the intended measurement concept (i.e., whether the category should be based on language, geography, etc.) nor on a definition for this category. As a result, an additional, minimum reporting category for this group was not created. Instead, the *1997 Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity* advised that further research be done to determine the best way to improve data for “Arabs/Middle Easterners.”

2. Public Comments

THEME 1: MENA as a racially diverse ethnicity – 1,042 comments

- people from the MENA region identify with many racial backgrounds, but no designation currently exists for them, and they remain undercounted and underrepresented; and
- MENA is an ethnic classification because it would include people of ethnic groups that are not country-specific, specifically groups such as Amazighs, Arabs, Assyrians, Berbers, Chaldeans, Coptics, Kurds, and others.

THEME 2: Disagreement with classification of MENA as White – 19 comments

- current OMB standards classifying people of Middle Eastern or North African descent as White have led to a significant undercount of the Middle Eastern or North African population; and
- undercounts prevent access to basic services such as language assistance, educational grants, funding for cultural competency training, as well as more effective implementation, monitoring, and enforcement of civil rights protections.

THEME 3: Classification of countries or ethnic groups as MENA – 1,347 comments

Several of the comments dealt with what should be included in the definition as follows:

- Nationalities included in the League of Arab States (74 comments);¹⁴
- Turkish (76 comments);
- Iranian (1,094 comments);
- Israelis (87 comments).
- Pan-national ethnic communities from this region;
- Groups such as Afghan, Azerbaijani, Ethiopian, Persian, Portuguese, Georgian, Greek, Italian, Somali, Spaniard, and Sudanese;
- Ethnoreligious groups, such as Assyrian, Chaldean, and Coptic, should be included in the MENA definition. Comments also were received that related to religious identities, i.e., "Jewish";
- The MENA classification should be strictly geographic;
- Armenians should be allowed to self-identify as MENA (2 comments);
- The World Bank's classification of MENA should be used to define this population.

THEME 4: Support for adding MENA minimum reporting category – 2,235 comments

Almost all comments received supported an additional, minimum reporting category for this population. These comments provided both personal and societal reasons for their support.

- The category would alleviate respondent confusion about how to respond to the current racial categories;
- Introduction of the MENA category would be an improvement to the current standards; and
- Accurate data on the MENA population are necessary to inform the community, researchers, and policy-makers in order to help in health care research, law enforcement, civil rights enforcement, and the development of anti-discrimination policies.

THEME 5: Non-support for adding a MENA minimum reporting category – 38 comments

- There is no scientific basis for a MENA category;
- Those who are MENA are not "distinctly different" than those who are White;
- Creating a new category is divisive and the proposed MENA category creates another special interest group and is a way for this group to get special funding;

¹⁴ Nationalities belonging to the League of Arab States are: Algerian, Bahraini, Comoran, Djiboutian, Egyptian, Iraqi, Jordanian, Kuwaiti, Lebanese, Libyan, Mauritanian, Moroccan, Omani, Palestinian, Qatari, Saudi Arabian, Somali, Sudanese, Syrian, Tunisian, Emirati, and Yemeni. See <http://www.lasportal.org/en/aboutlas/Pages/CountryData.aspx> and <http://www.britannica.com/topic/Arab-League>.

- It is inappropriate given current world events; and
- European ethnic origins are not represented on the form. The responses stated they would like to report their race or origin as European (e.g., Scandinavian, German, etc.) or American in Federal censuses and surveys.

3. Research Questions

The Subgroup posed three specific questions for its review:

- *Which nationalities, geographic areas, and ethnicities should be included in a classification of MENA?*
- *Does the Middle Eastern or North African population of interest identify with its inclusion in the White category?*
- *Should the OMB standards be revised to include an additional minimum reporting category for MENA distinct from other races or ethnicities (White, Black, or Hispanic, for example)?*

4. Overview of Extant Research

The Middle Eastern and North African population is one of the most diverse panethnic groups in the Nation, including those of many different linguistic, religious, national, and ethnic backgrounds.¹⁵ In recent years, there has been increasing interest among the public, academia, community organizations, and within the Federal government to address the concern that many people of Arab, Middle Eastern, or North African origin do not identify with the current Federal classification of White.¹⁶ This was demonstrated during the 2010 Census when a number of write-in campaigns were launched encouraging Middle Eastern and North African respondents to mark the “Some Other Race” checkbox and write their detailed Arab, Iranian, Assyrian, and other Middle Eastern and North African origins in the write-in space provided on the question on race.¹⁷

The current classification of groups from the Middle East and North Africa as White when aggregated for comparison across data sources may have contributed to sparse and inconsistent statistics on this population.¹⁸ In response, during the decade leading up to the 2010 Census, the Census Bureau initiated a number of research studies examining the development of a Middle Eastern or North African response category. In the 2010 Census Race and Hispanic Origin Alternative Questionnaire Experiment (AQE), focus groups found that respondents of Middle Eastern or North African origin did not identify with the White response category. Many participants also felt that the inclusion of the examples of

¹⁵ Bozorgmehr, M, Ong, P., & Tosh, S. (2015). Panethnicity revisited: contested group boundaries in the post-9/11 era. *Ethnic and Racial Studies*, 39(5), 727-745. Retrieved from: <http://dx.doi.org/10.1080/01419870.2015.1081959>

¹⁶ Ajrouch, K. J. & Jamal, A. (2007). Assimilating to a White Identity: The Case of Arab Americans. *International Migration Review*, 41(4), 860-79.

¹⁷ Kayyali, R. (2013). US Census Classifications and Arab Americans: Contestations and Definitions of Identity Markers. *Journal of Ethnic and Migration Studies*, 39(8), 1299-1318. Retrieved from: <http://dx.doi.org/10.1080/1369183X.2013.778150>

¹⁸ Kayyali, R. (2013). US Census Classifications and Arab Americans: Contestations and Definitions of Identity Markers. *Journal of Ethnic and Migration Studies*, 39(8), 1299-1318. Retrieved from: <http://dx.doi.org/10.1080/1369183X.2013.778150>

“Egyptian” and “Lebanese” within the White racial category was “wrong” or “inaccurate.” Comments like these were often connected to recommendations by focus group participants for a separate Middle Eastern, North African, or Arab category.¹⁹

These comments also were conveyed through more recent outreach activities. In May 2015, the Census Bureau held a forum of community members, academics, organization leaders, and other experts to gather feedback on testing a Middle Eastern or North African category in the 2015 National Content Test. This proposal was supported by nearly all participants at the forum, and most participants expressed that responses to this category should not be classified as White. Additionally, many forum respondents agreed with the groups included in the Census Bureau’s working classification of MENA but thought that additional groups, such as Turkish, Sudanese, and Somali, should be added.²⁰

5. Data Sources and Methodology

The 2015 National Content Test (NCT) conducted by the Census Bureau was used to empirically examine the MENA classification and category. The NCT’s split sample design allowed comparison of responses when a distinct MENA category was offered and when it was not.

A particular component of the NCT gathered race/ethnicity response patterns for persons identifying as Algerian, Amazigh or Berber, Arab, Assyrian, Bahraini, Bedouin, Chaldean, Copt, Druze, Egyptian, Emirati, Iraqi, Iranian, Israeli, Jordanian, Kurdish, Kuwaiti, Lebanese, Libyan, Moroccan, Omani, Palestinian, Qatari, Saudi Arabian, Syriac, Syrian, Tunisian, and Yemeni (the Census Bureau’s working classification of a Middle Eastern or North African reporting category), as well as persons identifying as Afghan, Armenian, Azerbaijani, Cypriot, Djiboutian, Georgian Commonwealth of the Independent States, Mauritanian, Somali, South Sudanese, Sudanese, Turkish, and Turkish Cypriot (groups sometimes classified within a Middle Eastern or North African category).

6. Result of Analyses

The detailed analyses of existing research is contained in Appendix 4.

The 2015 NCT research examined the degree to which respondents identified with a distinct MENA category. Responses were compared when a distinct MENA category was offered and when it was not. The NCT found that respondents select the distinct MENA category when it is available. When the distinct MENA category is not available, MENA respondents tend to identify with White or “Some Other Race.”

Analysis also was conducted to evaluate the Census Bureau’s working classification of MENA.

¹⁹ U.S. Census Bureau. (2012). *2010 Census Race and Hispanic Origin Alternative Questionnaire Experiment* (DSSD 2010 CPEX Memorandum Series #B-05-R, 2010 Census Planning Memoranda Series #211).

Washington, DC: Compton, E., Bentley, M., Ennis, S. & Rastogi, S.. Retrieved from:

https://www.census.gov/2010census/pdf/2010_Census_Race_HO_AQE.pdf

www.census.gov/2010census/pdf/2010_Census_Race_HO_AQE.pdf

²⁰ U.S. Census Bureau. (2016). 2015 Forum on Ethnic Groups from the Middle East and North Africa.

Washington, DC: Buchanan, A., Marks, R., Alvarez, M. Retrieved from:

<http://www.census.gov/library/working-papers/2015/demo/2015-MENA-Experts.html>

Respondents indicating these groups tended to identify as MENA when presented with the distinct MENA category. When no MENA category was available, groups in the Census Bureau's working classification of MENA most frequently reported in the White category. However, when a MENA category was available, all of the groups in the Census Bureau's working classification, except for Israeli, most frequently reported their detailed origins in the MENA category (see Table 23 in Appendix 4).

This analysis also showed that groups not in the NCT MENA working classification generally did not identify as MENA. Groups not in the NCT MENA working classification were generally more likely to identify with other categories such as White, Black, Asian, etc. than with the MENA category. An exception to this trend is Afghan. The NCT data were then analyzed to examine how offering a distinct MENA category affects reporting of other race/ethnicity groups. Initial analyses indicate that the inclusion of a distinct MENA category does not appear to affect the reporting of other major racial/ethnic groups, but further statistical analyses are underway. The percent of respondents who report as MENA was not affected by the inclusion of a MENA response category and remained under 1% of the total population that self-reported.

7. Summary of Initial Evaluation

Many people have signaled that an ethnic classification for the MENA population should be added because those with roots, origins, and heritage from the Middle East and North Africa are racially diverse. Further, results from the 2015 NCT show that those in the Census Bureau's working classification of MENA largely identify with the MENA response category when it is available. Last, findings from the Census Bureau's 2015 Forum on Ethnic Groups from the Middle East and North Africa and a review public comments on *Proposed Information Collection; Comment Request; 2015 National Content Test (12/2/2014)* found that some experts and stakeholders believe that a classification of this population should be geographically based.

a) Initial Proposal: The Subgroup proposes that a Middle Eastern or North African classification be added to the standards. The classification for the Middle Eastern and North African population should be geographically based. The MENA classification should be defined as: A person having origins in any of the original peoples of the Middle East and North Africa. This includes, for example, Lebanese, Iranian, Egyptian, Syrian, Moroccan, Israeli, Iraqi, Algerian, and Kurdish.²¹ The Subgroup bases this initial recommendation on public comment and analyses to date.

However, some questions remain. Some of the groups proposed for inclusion under a MENA classification were also ethnoreligious groups. A challenge to ethnicity measurement can be the intersection of ethnicity with religious affiliation. The race/ethnicity standards are not intended to measure religion (see P.L. 94-521), and it is unclear how to address inclusion of

²¹ The rationale for using these examples is to include the two largest Middle Eastern Arab nationalities (Lebanese and Syrian), the two largest North African Arab nationalities (Egyptian and Moroccan), and the two largest non-Arab nationalities within the Middle Eastern / North African region (Iranian and Israeli) as the first six examples. This is followed by the next largest Middle Eastern Arab nationality (Iraqi), the next largest North African Arab nationality (Algerian); as well as an example of a transnational, non-Arab group (Kurdish).

ethnoreligious groups while clearly maintaining the intent and use of the resulting measure as not indicating religion. Further, although the great majority of public comments received on the measurement of MENA supported an additional, required minimum reporting category, the cost and burden of requiring this additional reporting category when race/ethnicity is measured across the Federal government is unclear.

b) Request for Public Comment: To assist in its deliberations, the Subgroup requests public comment on the following questions. Thinking about how information is collected:

1. If MENA were collected as a separate reporting category, assuming that separate race/ethnicity questions continue to be the standard, should MENA be considered an ethnicity or a race? [Note that, in either case, respondents will still be encouraged to mark all that apply.]
2. Should the MENA category be a required minimum reporting category that is separate from the White minimum reporting category?
3. Outreach conducted with the Israeli American Council and Jewish American organizations indicates that persons of Ashkenazi, Mizrahi, and Sephardi origin do not wish to be included in the MENA category, as these ethnicities directly identify persons as Jewish. Moreover, experts at the Census Bureau's 2015 Forum on Ethnic Groups from the Middle East and North Africa expressed that those who identify as Assyrian, Chaldean, Coptic, or Druze would like to be included in a MENA category. Which, if any, specific ethnoreligious groups should be included in a MENA classification?
4. We have also observed from initial feedback that the definition of MENA may be misunderstood to include only persons who are foreign born. Our intention is that a MENA category, should it be adopted, would include persons of MENA origins, regardless of country of birth. How can this best be communicated to respondents?
5. What is the estimated cost and public burden associated with requiring an additional reporting category for MENA across Federal information collections? Given the small estimated size of the MENA group, would a separate reporting category allow reporting of statistically reliable estimates? Would the size of the MENA group present confidentiality or privacy concerns?

C. Additional Minimum Reporting Categories

1. Current Practice

In the 1990s, OMB conducted a systematic review of the Federal standards on collecting and reporting data on race/ethnicity. During this review, the then Interagency Committee for the Review of the Racial and Ethnic Standards received approximately 10,000 letters and postcards on the issue of classifying data on Native Hawaiians. The Interagency Committee recommended to OMB that data on Native Hawaiians continue to be classified in the Asian or Pacific Islander category. This recommendation was opposed by the Hawaiian congressional delegation, the 7,000 individuals who signed and sent preprinted yellow postcards, the State of Hawaii departments and legislature, Hawaiian organizations, and other individuals who commented on this recommendation. However, OMB did not accept the Interagency Committee recommendation. Based largely on evidence presented at public hearings in

Hawaii. OMB decided to break apart the Asian or Pacific Islander category into two categories—one called “Asian” and the other called “Native Hawaiian or Other Pacific Islander.”

In addition, Federal agencies were encouraged to collect more detailed race and ethnicity data. Specifically, the 1997 standards state:

“In no case shall the provisions of the standards be construed to limit the collection of data to the categories described above. The collection of greater detail is encouraged; however, any collection that uses more detail shall be organized in such a way that the additional categories can be aggregated into these minimum categories for data on race and ethnicity.”

There are numerous examples of Federal agencies collecting detailed race and ethnicity data in their statistical reporting; these are not limited to decennial censuses or extremely large surveys, such as the American Community Survey (ACS). Nonetheless, OMB has learned that the minimum reporting categories as described in the current standards are often misinterpreted as the only permissible reporting categories rather than allowing more detailed categories, provided that detail group data can be aggregated in a systematic way to allow for comparison across data sources.

2. Public Comments

THEME 1: Collect more than the minimum categories – 1,274 comments

- Support the collection of detailed race and ethnicity data by agencies beyond the minimum reporting categories;
- Disaggregate the Asian category and Native Hawaiian or Pacific Islander category (1,234 comments);and
- Disaggregate the “Black or African American” category (10 comments).

THEME 2: Change the minimum reporting categories – 7 comments

- Support a requirement to collect detailed race and ethnicity data beyond the minimum reporting categories (7 comments);
- Require subcategories under Asian and Native Hawaiian or Other Pacific Islander, and aggregate to Asian and Native Hawaiian or Other Pacific Islander” in reporting only when necessary (4 comments).
- Required two categories under Asian: South Asian and Asian (1 comment);and
- Native Hawaiian should be a separate minimum reporting category (2 comments).

Lists were provided of specific detailed groups, but were intended to be descriptive but not standardized nor exhaustive. Many comments cited health disparities, and language and cultural differences among these detailed groups, such as Chinese, Asian Indian, or Samoan.

THEME 3: Create a uniform item for detailed collection– 9 comments

- Support a universal race and ethnicity question and set of response options for collecting detailed race and ethnicity data as a solution to the confusion about the minimum response options. Of those nine, one offered a specific plan for the universal standard.

THEME 4: Do not change the minimum reporting category standard– 2 comments

- OMB should not change the existing standards for minimum reporting categories.

THEME 5: Report as many categories as possible–6 comments

- Disapprove rolling categories into “All Other Races” in reporting data. Commenters specifically advocated for the American Indian or Alaska Native, Asian, and Native Hawaiian or Other Pacific Islander categories be reported whenever collected.

3. Research Questions

In conducting its review of the functioning of the current standards, the Subgroup identified four specific questions related to proposals for more detailed race/ethnicity data, and the intent and use of minimum reporting categories.

- *What are the current needs for and uses of detailed race and ethnicity data?*
- *Should Federal guidelines be established for the collection of detailed race and ethnicity data? If so, would the guidelines apply to collections that are not self-reported?*
- *Should the current standards be revised to strongly encourage, but not require, the collection of detailed race and ethnicity data? If so, would this language on its own be sufficient to increase the likelihood of agencies opting to collect detailed race and ethnicity data? Or, would such encouragement in the standards need to be coupled with guidelines for the collection of detailed race and ethnicity data?*
- *Should the current standards be revised to require the collection of detailed race and ethnicity data across all Federal agencies? If so, should the requirement be coupled with guidelines for the collection of detailed race and ethnicity data?*

4. Data Sources and Methodology

The Subgroup considered input received through the public comment process and input from Federal agency experiences. The Subgroup is interested to learn more regarding the needs and considerations of Federal agencies when determining the level of detail on race/ethnicity they will collect. The Subgroup plans to also consider whether strong encouragement, but not a requirement, for detailed race/ethnicity collection will improve information quality, or if a requirement would be necessary (and under what conditions).

5. Summary of Initial Evaluation

The initial review of the 1997 standards did not identify additional, minimum reporting categories for detailed race/ethnicity groups as an element for evaluation. However, during the public comment period for September 30, 2016’s Federal Register Notice, the Working Group received more than 1,200 comments expressing the need for further disaggregated data for Asian communities and Native Hawaiian or Other Pacific Islander communities. Other comments express a similar need for disaggregated data, including 10 comments advocating for the disaggregation of the Black or African American category.

a) Initial Proposal: Based on public comment and Federal agency input received to date, the Subgroup proposes that OMB issue specific guidelines for the collection of detailed data for American Indian or Alaska Native, Asian, Black or African American, Hispanic or Latino, Native Hawaiian or Other Pacific Islander, and White groups for *self-reported* race and ethnicity collections. By providing these guidelines, consistent collection of detailed race and ethnicity data will be supported across Federal agencies. Such direction would not be applied to the

collection of observed race/ethnicity, since the accuracy such reporting at a detailed level would be a concern. Further, the Subgroup plans to consider under what conditions detailed data *should not* be collected. However, the Subgroup plans to continue its deliberations as to whether OMB should require or, alternatively, strongly support but not require Federal agencies to collect detailed data.

b) Request for Public Comment: The Subgroup requests public comments on the guidelines that should be provided for collecting detailed race and ethnicity data. Additionally, to evaluate whether or not the reporting of detailed categories should be required, or if such reporting should be strongly encouraged but not required, additional information is needed. The Subgroup recognizes that collecting detailed race and ethnicity data likely would impose a substantial cost on Federal agencies, State and local agencies, and businesses and a burden on the public. Therefore, the Subgroup requests public comment on the consideration that should be given to evaluate the value of improved information quality taking into account anticipated cost and public burden. Specifically, the Subgroup seeks public comment on the following questions:

1. If issuing specific guidelines for the collection of detailed American Indian or Alaska Native race and ethnicity data, should OMB adopt the NCT format, which includes separately Navajo Nation, Blackfeet Tribe, Mayan, Aztec, Native Village or Barrow Inupiat Traditional Government, and Nome Eskimo Community? If not, how should OMB select the detailed race and ethnicity categories?
2. If issuing specific guidelines for the collection of detailed Asian race and ethnicity data, should OMB adopt the 2010 Decennial Census and NCT format, which includes separately Chinese, Filipino, Asian Indian, Vietnamese, Korean, Japanese, and an “other Asian” category²²? If not, how should OMB select the detailed Asian race and ethnicity categories?
3. If issuing specific guidelines for the collection of detailed Black or African American race and ethnicity data, should OMB adopt the NCT format, which includes separately African American, Jamaican, Haitian, Nigerian, Ethiopian, and Somali? If not, how should OMB select the detailed race and ethnicity categories?
4. If issuing specific guidelines for the collection of detailed Hispanic or Latino race and ethnicity data, should OMB adopt the NCT format, which includes separately Mexican or Mexican American, Puerto Rican, Cuban, Salvadoran, Dominican, and Colombian? If not, how should OMB select the detailed race and ethnicity categories?
5. If issuing specific guidelines for the collection of detailed Native Hawaiian or Other Pacific Islanders race and ethnicity data, should OMB adopt the 2010 Decennial Census format, which includes separately Native Hawaiian, Chamorro,²³ Samoan, and an “other Pacific

²² The checkboxes used in Census 2010 were Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, and Other Asian with five additional examples of Hmong, Laotian, Thai, Pakistani, and Cambodian.

²³ In the 1997 standards, the actual OMB standards used the term Guam, not Guamanian. Census 2010 featured the following checkboxes: Native Hawaiian, Guamanian or Chamorro, Samoan; and provided the

Islander” category? Should it use the NCT format, which includes separately Native Hawaiian, Samoan, Chamorro, Tongan, Fijian, and Marshallese? If neither of these, how should OMB select the detailed Native Hawaiian or Other Pacific Islander race and ethnicity categories?

6. If issuing specific guidelines for the collection of detailed data for White race and ethnicity data, should OMB adopt the NCT format, which includes separately German, Irish, English, Italian, Polish, and French?²⁴ If not, how should OMB select the detailed race and ethnicity categories?
7. What burden and cost would a Federal requirement to collect detailed race and ethnicity data place on Federal agencies, State and local agencies, private sector entities and the public? How should this burden and cost be weighed against any anticipated improvement in information quality?
8. Should Federal agencies be required to collect detailed race and ethnicity data even when such data could not be responsibly reported due to statistical reliability and confidentiality concerns? If so, in which cases? What factors should be considered?
9. If OMB were to strongly encourage, but not require, collection of detailed race and ethnicity data by Federal agencies, how likely are Federal agencies to adopt collection of detailed race and ethnicity data?
10. If OMB were to strongly encourage, but not require, collection of detailed race and ethnicity data by Federal agencies, what criteria should be used to encourage and evaluate conformance with such guidance?

D. Relevance of Terminology

1. Current Standard

Although many respondents report within the race and ethnicity categories specified by the 1997 standards, the standards themselves may not be well understood by the public. Over time, in some cases, some terms may be preferred over others. For example, some respondents do not identify themselves within current Federal definitions of the race and ethnicity minimum categories. Additionally, use of references to geographic location in descriptions of race and ethnicity categories may be incomplete, unclear, inconsistent, or confusing. In addition, some respondents may not know that they can report more than one race and/or ethnicity category. Finally, respondents may find certain terms to be outdated and/or offensive. Collectively, these issues point to the importance of periodic review of terms used in the OMB standards to ensure continued relevance and clarity.

following examples listed for other NHPI: Fijian and Tongan. Since Census 2010, based on feedback received by members of the NHPI community, Census no longer includes the term Guamanian in its collections.

²⁴ These are the examples used when MENA was included in NCT questionnaires. When MENA was not included in NCT questionnaires, the examples are as follows: German, Irish, English, Italian, Lebanese, and Egyptian.

2. Public Comments

THEME 1: Use of the terms “race or origin,” “race or ethnicity,” or “categories.” – 14 comments

- Voiced uncertainty regarding the intended definitions of the terms “race” and “ethnicity;” and
- Disagreed with the current usage of one of these terms, or presented an understanding of these terms that is not consistent with their intended conceptualization as socio-political constructs (rather than based on biology) as described in the 1997 revision to the standards.

THEME 2: Need versus no need to change terminology – 8 comments

- Support changing terminology used in the standards, or specifically changing the name of a race/ethnicity category or its detailed group listing (6 comments);
- No need to change the terminology used in the standards (1 comment); and
- Remove the term “Negro”(1 comment).

THEME 3: Discontinue use of term “principal minority race” – 24 comments

- No longer use the term “principal minority race” (18 comments); and
- Either change the term’s definition or continue review of whether the term is appropriate in its current usage (6 comments).

THEME 4: Use of term “multiracial” in reporting – 6 comments

- Use a “multiracial” group in reporting; and
- Revise the instructions used in the collection of race and ethnicity to be worded “Mark two or more” (1 comment).

THEME 5: Categorize all race/ethnicity by geographic region— 2 comments

- Do not use geographic locations in race definitions (e.g., the use of “African American”) given that geographic origin may not reflect race (e.g., someone from an African nation may be of the White, not Black, race); and
- Change how a Geographic region is classified within a given race/ethnicity category.

3. Research Questions

In conducting its review, the Subgroup identified three specific areas for research: a) instructions and race/ethnicity terms; b) clarification of classifications; and c) presentation of and terms used for detailed groups.

a) Instructions and Race/Ethnicity Terms

From comments received, the Subgroup noted some confusion about the opportunity of respondents to select more than one category, and some confusion about the use of the terms of race, origin, and ethnicity. Two sets of proposed race/ethnicity instruction and question-wording terminology were considered: 1) instruction terminology of “Mark [X] one or more boxes” versus “Mark all boxes that apply,” and 2) question-wording terminology of “race or origin” versus “race or ethnicity” versus “categories.” The Subgroup examined the following questions:

- *Do the proposed revisions to instructions lower respondent nonresponse?*
- *Do the proposed revisions to question wordings change the distribution of respondents*

reporting either single or multiple races or ethnicities? Does this vary based on respondent age, respondent racial group, inclusion of the MENA category, or use of the combined versus separate Hispanic question?

- *Which question wording results in the fewest instances of write-in responses (which may indicate respondent confusion on the intent of the item or the terms used)? Further, what are the top 10 most common written-in responses for each instruction wording? Which groups have the highest frequencies of write-in responses for each instruction wording?*

b) Clarification of Classifications

The Subgroup observed a lack of clarity in the classification of several groups in the current standard. To improve guidance, the Subgroup pursued two research questions:

- *How should the standards classify the following responses: Australian (including the original people of Australia/the Aborigines), Brazilian, Cape Verdean, New Zealander, Papua New Guinean?*
- *Do South and Central American Indians identify as American Indian or Alaska Native (AIAN) on the decennial census and other Federal data collections?*

c) Presentation of and Terms Used for Detailed Groups

There is an error in 1997 OMB classification standards with “Cuban” being listed twice in the reporting category of “Hispanic or Latino.” To correct this, Cuban would appear once in the order of population size, according to the current method governing the ordering of detailed groups listed within the Hispanic/Latino reporting category. However, the Subgroup noted that population size currently does not govern the listing of detailed groups across all minimum reporting categories. Accordingly, the Subgroup examined three questions:

- *Should the current ordering of detailed lists for Hispanic/Latino be changed to reflect current population size?*
- *Should detailed lists within minimum reporting categories be ordered by population size, across all minimum reporting categories?*
- *Should the same number of detailed groups be listed within each minimum reporting category (e.g., only listing the top six largest detailed groups)?*

Note that if the detailed lists are changed, some new detailed groups may be included within the existing minimum reporting categories. In addition, some detailed groups that are currently listed may be removed (e.g., if they are no longer one of the largest detailed groups for a given minimum reporting category).

The Subgroup also carefully considered the use of “Principal Minority Race” in the current standard, taking particular note of: (1) PL 114-157, which amends the Department of Energy Organization Act to revise the definition of “minority” to mean any U.S. citizen who is an Asian American, Native Hawaiian, Pacific Islander, African American, Hispanic, Puerto Rican, Native American, or Alaska Native; and (2) the Local Public Works Capital Development and Investment Act of 1976 to revise the definition of “minority group members” (who own minority group enterprises) to mean U.S. citizens who are Asian American, Native Hawaiian, Pacific Islanders, African American, Hispanic, Native American, or Alaska Natives. Although PL 114-157 only impacts the Department of Energy, it serves as a resource for how some people in the Federal statistical system view the term “Hispanic” and “minority group members.”

Given the changes in the US population and the language in PL 114-157, the Subgroup examined:

- *Should the current standards still refer to and provide guidance on reporting “principal minority race?”*
- *If the term “principal minority race” is still meaningful, should the designation be guided by minority group population size, historical disadvantage, or some other principle? Further, should “principal minority race” be expanded to include “ethnicity”?*

The Subgroup also noted that some terms used in the current standards to describe race/ethnicity may not (or no longer) resonate with the public. Accordingly, it examined:

- *Should the term “Negro” be removed from the standards?*
- *Should the term “Far East” be removed from the standards?*

4. Data Sources and Methodology

Empirical research, (i.e., qualitative and quantitative analyses), and input from stakeholders were considered when formulating recommendations concerning the research questions listed above. Data used include results from the 2015 National Content Test (NCT) and the results of focus group research, as described further below.

5. Result of Analyses

The detailed results of the analyses on these questions are found in Appendix 6.

a) Instructions and Race/Ethnicity Terms

The 2015 NCT examined two alternative wordings of race/ethnicity item instructions: 1) “Mark [X] one or more boxes” and 2) “Mark all boxes that apply.” The Subgroup identified two criteria on which to base its evaluation: 1) which wording produces lowest nonresponse; and 2) which working produces greatest reporting of multiple race/ethnicities. Although nonresponse for these two options was similar (both for respondents reporting using paper or internet versions of the item), the “Mark all boxes that apply” wording was associated with more respondents reporting multiple races/ethnicities across both modes of data collection.

The Subgroup also examined three alternative wordings in items collecting data on race/ethnicity: 1) “race or origin,” 2) “race or ethnicity,” and 3) “categories.” The Subgroup investigated which of these wording options produces the lowest nonresponse and which produces the largest reporting of multiple race/ethnicities. Patterns of nonresponse and selection of multiple races/ethnicities by terminology options differed by mode of administration (i.e., internet versus paper).

b) Clarification of Classifications

To guide its recommendations to clarify classification for persons identifying as Australian (including the original people of Australia/the Aborigines), Brazilian, Cape Verdean, New Zealander, and Papua New Guinean, the Subgroup examined the extent to which these groups were provided as write-in responses to the “some other race” category. Analyses were conducted both when the Middle Eastern or North African (MENA) category was included and excluded. Additionally, analyses were conducted using both the combined race/ethnicity question and the separate race and ethnicity questions.

None of these groups appeared in the top 10 write-in responses in any of these analyses. The Subgroup then reviewed the proportion of write-in responses to the “Some Other Race” category by geographic location. Of the five classifications under the current review, Brazilian was the most common write-in response (approximately 1 percent, weighted).

The Subgroup also examined reporting patterns of South and Central American Indian respondents to determine if self-identification and reporting behavior is consistent with the current standard, which includes them in the American Indian or Alaska Native reporting category. Focus group data initially collected to inform the design of the 2015 NCT were examined. A review of those data suggest that Central and South American Indians do not tend to self-identify as American Indian or Alaska Native.

c) Presentation of and Terms Used for Detailed Groups

To determine which duplicate initial category mention of “Cuban” should remain, the Subgroup recommends this be guided by population size. The Subgroup also considered whether the current ordering of the classification list be updated to reflect current population size, taking note of the rationale provided for Subgroup ordering decisions related to the NCT. As a next step, the Subgroup plans to apply this rationale to the classification listing and determine the magnitude and benefit of any resulting changes.

The Subgroup also considered the relevance of terms used in the current standards. The Subgroup recommends that the term “Negro” be removed from the standards. This is in accordance with PL 114-157 and with the Census Bureau’s decision to remove the term “Negro” in its information collections beginning in 2014. Further, the Subgroup recommends that the term “Far East” be removed from the current standards. The term has never been used by the Census Bureau when collecting data²⁵, and appears to be out of date with common usage, and may be offensive to some groups.

Given the inclusion of “Hispanic” in these definitions of “minority,” and the relative prevalence of the Hispanic or Latino versus Black or African American populations, the Subgroup also considered whether referring to Black or African American as the “principal minority race” is still relevant, meaningful, accurate, and acceptable. The Subgroup plans to also consider if “Hispanic” should be among the groups considered, and the salience of alternative terms (e.g., “principal minority race/ethnicity”). “Hispanic” usually is considered an ethnicity while “minority” is usually used when referencing race. Changing “principal minority race” to “principal minority race/ethnicity” supports the changing demographics of the U.S. population. Many consider it important to continue identifying the principal minority group in Federal data collections and reporting systems.

6. Summary of Initial Evaluation

a) Initial Proposal: The Subgroup proposes no changes be made to the current standards to specifically incorporate the following geographic locations into any existing race or ethnicity category: Australian (including the original people of Australia/the Aborigines), Brazilian, Cape Verdean, New Zealander, Papua New Guinean. This proposal takes into account the low

²⁵ Pratt, B., Hixson, L., and N. Jones (2015). “Measuring Race and Ethnicity Across the Decades: 1790-2010.” U.S. Census Bureau. Retrieved from: http://www.census.gov/population/race/data/MREAD_1790_2010.html/

prevalence of these geographic locations appearing as write-in responses according to the research presented above, and public comments to the September 30, 2016 FRN advocating against use of geographic regions in race/ethnicity group definitions.

Based on its analyses to date, the Subgroup proposes more research and public input be conducted to enable a more complete consideration of adding more specific South or Central American subgroups to the current description of the AIAN category in order to improve identification with the reporting category.

The Subgroup proposes that the duplicate, initial category mention of “Cuban” be deleted so that listing is presented according to population size. The Subgroup also considered whether the current ordering of the classification list be updated to reflect current population size. As a next step, the Subgroup plans to apply this rationale to the classification listing and determine the magnitude and benefit of any resulting changes. The Subgroup plans to share the results of this analysis with the public.

The Subgroup proposes that the term “Negro” be removed from the standards. Further, the Subgroup recommends that the term “Far East” be removed from the current standards. The Subgroup plans to consider if the term “Guamanian” should also be removed.²⁶ The Subgroup also proposes that race/ethnicity coding procedures be documented and publicly available, as this would allow greater transparency and promote further consistency in Federal data collections.

The Subgroup proposes clarifying the standards to indicate the classification is not intended to be genetically based, nor based on skin color. Rather, the goal of standards is to provide guidelines for the Federal measurement of race/ethnicity as a social construct and therefore inform public policy decisions.

b) Request for Public Comment: The Subgroup also considered whether referring to Black or African American as the “principal minority race” is still relevant, meaningful, accurate, and acceptable. However, it is not clear if the referent groups should change given changing demographics. The Subgroup intends to examine further whether “Hispanic” should be among the groups considered among “principal minorities” and whether alternative terms are salient (e.g., “principal minority race/ethnicity”). “Hispanic” usually is considered an ethnicity while “minority” is usually used when referencing race.

1. Should Hispanic or Latino be among the groups considered among “principal minorities”? Would alternative terms be more salient (e.g., “principal minority race/ethnicity”)? Hispanic or Latino usually is considered an ethnicity while “minority” is usually used when referencing race.

²⁶ In the 1997 standards, the actual OMB standards used the term Guam, not Guamanian. Census 2010 featured the following checkboxes: Native Hawaiian, Guamanian or Chamorro, Samoan; and provided the following examples listed for other NHPI: Fijian, and Tongan. Since Census 2010, based on feedback received by members of the NHPI community, Census no longer includes the term Guamanian in its collections.

IV. CONCLUSION

This Interim Report describes the Working Group's progress to date and requests for further public comment. During the comment period requested, the Working Group plans to continue its research as described in the Interim Report. The conclusions of that research and responses to this second request for public comment will inform the proposals of each Subgroup and the discussion and subsequent preparation of the Working Group's recommendations to OMB. OMB plans to make the Working Group's recommendations public and, once it has concluded its review, OMB plans issue a notice of final decision.

Appendix 1: Implementation of Current Standards on the Separate Questions

US Census Bureau

This section describes the implementation of the current standards as conducted for the US Census Bureau, National Center for Health Statistics, Bureau of Justice Statistics, and National Center for Education Statistics data collections, and concludes with a summary of the implementations.

In addition to the minimum race categories set forth by OMB, the Census Bureau is required by Congress²⁷ to allow respondents to select a category called “Some Other Race.”

Write-in responses to the separate format race question in the 2010 Census were coded according to one or more of the five minimum reporting categories provided in the 1997 OMB standards. For example, a 2010 Census response of “Japanese” was coded as “Asian”; a response of “German” was coded as “White”; and a response of “Haitian” was coded as “Black” following the standards’ definitions. If it was not possible to determine which minimum reporting category the write-in response should be assigned, the response was included in the “Some Other Race” category. For example, respondents reporting entries such as multiracial, mixed, interracial, or a Hispanic, Latino, or Spanish group (for example, Mexican, Puerto Rican, Cuban, or Spanish) in response to the race question are included in this category. Therefore, the category “Some Other Race” includes all other responses not included in the “White,” “Black or African American,” “American Indian or Alaska Native,” “Asian,” and “Native Hawaiian or Other Pacific Islander” minimum reporting categories.^{28 29}

However, in order for race and ethnicity reporting categories to be comparable to those of other agencies which do not tabulate a “Some Other Race” category, the Census Bureau creates a Modified Race Summary File wherein a specified race (one of the minimum OMB reporting categories) is assigned to replace the “Some Other Race” response that was tabulated in the 2010 Census. This procedure effectively changes the “Some Other Race” response to a response of White or Black or Asian, etc. (See *Overview of Extant Research*.) If a Combined Question format for self-reported race were permitted, a “Some Other Race” category would not be needed to account for cases where ethnicity is reported in census collections, but race is not.

National Center for Health Statistics (NCHS)

NCHS collects race and ethnicity data on many of its surveys. NCHS’s National Health Interview Survey (NHIS) is one of the Nation’s largest health surveys and is often used to report health data by race and ethnicity. The NHIS currently uses the Separate Questions format. If there is not a response to the race or ethnicity questions, NCHS employs statistical methods adapted from the Census Bureau for imputing missing information. NCHS’s National Health and Nutrition Examination Survey (NHANES) is another large survey often used to assess race differences in health and health behaviors. NHANES currently uses the Separate Questions format to collect race and ethnicity data.

²⁷ H.R. 2862, 109th Cong. (2005-2006).

²⁸ With the current two Separate Questions approach, when a respondent provides a write-in response of Hispanic to the race question, such as Mexican American or Peruvian, the Census Bureau codes these Hispanic responses as “Some Other Race,” rather than classifying the reported Hispanic responses as a specific race group, such as White or Asian or Black or coding them as “missing.”

²⁹ U.S. Census Bureau (2011). 2010 Census Redistricting Data (Public Law 94-171) Summary File — Technical Documentation. Washington, DC. Retrieved from: www.census.gov/prod/cen2010/doc/pl94-171.pdf

However, NHANES uses a combined race and ethnicity variable for reporting purposes. Publicly available microdata categories are Mexican American, Other Hispanic, Non-Hispanic Black, Non-Hispanic White, Non-Hispanic Asian, Non-Hispanic Other (including Multi-Racial), and Missing. Race data are not reported for persons identifying their ethnicity as Hispanic.

Bureau of Justice Statistics (BJS)

The data sources used by Federal agencies to produce Federal data on race/ethnicity affect their ability to fully conform to the Federal standards. The Bureau of Justice Statistics (BJS) produces Federal statistics from several administrative data sources, including Federal, State, and local law enforcement and correctional agencies. In many of these collections, agencies are reporting aggregate counts of individuals in the justice system. Several of these non-Federal collections allow respondents (typically, administrators on behalf of the population of interest) to report observed race using an additional category of “other race” for those individuals who are not perceived as identifying with one of the minimum race/ethnicity reporting categories. Further, many non-Federal administrative data systems do not have IT systems in place to allow for selecting more than one race category. In these cases, BJS asks responding agencies to report persons of more than one race as “Multiracial” or “Two or more races.” Nonetheless, few offender management software systems offer these categories as an option which has led to many jurisdictions not reporting complete race/ethnicity information.

On BJS surveys of jail and prison inmates whereby race/ethnicity is collected by self-report (both of which have Separate Questions for race/ethnicity), respondents are asked to identify their race by selecting one or more categories: Black, White, or “All Other Races.” If respondents select “All Other Races,” they are prompted to specify a specific race category that aligns with the minimum reporting categories specified in the current standards. Due to small sample sizes and therefore concerns regarding statistical reliability of estimates and confidentiality of respondents, BJS typically combines American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, Asian, Two or More Races, and other races into a residual category for reporting purposes. It is important to note that the Federal standards do not offer “additional or other race,” “all other races,” “multiracial,” or “two or more races” as response categories.

The National Crime Victimization Survey (NCVS) uses the Separate Question format for race and ethnicity. Respondents are asked to “mark all that apply” for race and are presented with the minimum reporting categories. Respondents are also given the option of marking “other – specify” and asking the Field Representative (interviewer) to write in their race. BJS statisticians then combine the race and Hispanic origin variables and publish estimates on White (non-Hispanic), Black (non-Hispanic), Hispanic, and persons of other races (non-Hispanic; includes American Indian or Alaska Native, Asian, Native Hawaiian and Other Pacific Islander, and persons of two or more races). This last category, persons of other races, is typically created due to small sample sizes for some of the racial categories. These approaches suggest that while the race/ethnicity data are collected reflecting a two Separate Questions approach, they are tabulated in a manner reflecting a Combined Question approach.

National Center for Education Statistics (NCES)

NCES collects race and ethnicity data through a range of survey and administrative data collections. In surveys of individuals, respondents are presented with the Separate Question format on race and Hispanic origin. When the data are tabulated for publications or prepared for online data tools, the responses are coded to fall into mutually exclusive categories. Public-use data sets generally have a

constructed race/ethnicity variable in which multiple race responses are coded as “Two or More Races,” unless ethnicity is reported, in which case the responded is coded as Hispanic regardless of which race categories have been selected. Frequently, situations arise where the sample size for the Native Hawaiian or Other Pacific Islander respondents or American Indian or Alaska Native respondents may become too small to present reliable results. In statistical indicator reports, the reporting preference is to show each of the groups distinctly, with a “not meeting statistical standards symbol” where appropriate, rather than by aggregating them into a residual category. This allows the presentation of the smaller groups to the extent feasible and avoids the potential concern of data users ascribing the characteristics of the overall residual category to specific groups not presented. For many NCEs surveys, the race/ethnicity variable is considered a key variable, so values are generally imputed for non-respondents on this specific question. It is important to note that “Two or more races” is not a response or presentation category offered by the Federal standards.

NCEs collects administrative data from elementary and secondary schools and postsecondary institutions through aggregate counts. Student race/ethnicity data are collected through an aggregated tabular structure in conjunction with other student data (e.g., enrollment by grade; high school graduates; suspensions; expulsions; etc.) and reported to the Department of Education through reporting systems coordinated through State education agencies. Although local schools are expected to use the Separate Question format to gather the student race and ethnicity data from parents or guardians, the data submitted to NCEs are compiled into the minimum reporting categories with Hispanic origin superseding racial identity. This applies to postsecondary institutions as well. The objective of this approach is to keep the scope of data collection manageable and to keep the results consistent with the standard NCEs display of a combined race/ethnicity variable. Thus, these tabulations also seem to demonstrate that while the race/ethnicity data are collected using a two Separate Questions approach, they are reported in a manner reflecting a Combined Question approach.

Appendix 2: Combined vs Separate Question: Existing Research on Non-response in Surveys and Administrative Data Collections

Item Nonresponse to Race/Ethnicity in US Decennial Census

Specifically, as mentioned previously the Census Bureau, unlike other Federal agencies, is required by law to offer the reporting category of “Some Other Race.” This response option is not an element of the standards on race/ethnicity issued by OMB, and, indeed, OMB does not approve the use of “Some Other Race” for any other Federal agency, despite the apparent “legacy” use of categories such as “other,” “multiracial,” or “Two or more races” in the collection and presentation of race/ethnicity data as described above.

Specifically, analysis of the decennial censuses in 2000 and 2010 conducted by the Census Bureau indicates that the number of individuals choosing to report “Some Other Race” instead of a specific race group grew by about 25 percent. In fact, the “Some Other Race” population was the third largest racial group in 2010, representing 6 percent of all respondents. Further, 97 percent of those selecting “Some Other Race” also reported a Hispanic origin ethnicity, suggesting that these persons viewed “Hispanic or Latino” as a racial identity as opposed to an ethnic identity.³⁰ The Census Bureau found that when participants were asked a combined race/ethnicity question rather than the current standard of Separate Questions, the rate of overall nonresponse dropped from 7 percent to under 0.5 percent, largely due to Hispanics selecting only “Hispanic” in the Combined Question format. In the scenarios tested, the “Some Other Race” response option was provided as required by law for the Census Bureau.³¹

In related focus group research conducted by the Census Bureau, respondents indicated that “only Hispanic” resonates more closely with how Hispanics self-identify. Further, participants commented that the current racial and ethnic question format seemed to treat groups unequally: the separate Hispanic origin question was sometimes perceived as potentially identifying Hispanics for discriminatory reasons, while others felt that Hispanics received preferential treatment with their own question. Participants also commented that the Separate Questions format did not provide Whites and Blacks with an opportunity to indicate their specific detailed identities (e.g., German; Irish; Jamaican; Nigerian). The Subgroup may consider exploring how the current practices of coding and editing race/ethnicity data compare across by agencies, to better understand their application to this discussion.

With the projected steady growth of the Hispanic or Latino population, the number of respondents who select Hispanic and either 1) do not select a race category or 2) write in Hispanic in response to the race question is expected to increase (Compton et al. 2012; Rios et al. 2014). As described more fully under Data Sources and Methodology, the current Census procedure for coding these cases requires significant effort to produce intercensal population estimates and projections—the foundation (or “denominator”) for race and ethnicity measured in other Federal surveys. Although

³⁰ Rios, M., Romero, F, and Ramirez, R. (2014). “Race Reporting Among Hispanics: 2010.” U.S. Census Bureau Working Paper Series Number 102. Retrieved from: www.census.gov/population/www/documentation/twps0102/twps0102.pdf

³¹ Compton, E., Bentley, M., Ennis, S., and Rastogi, S. (2012). 2010 Race and Hispanic Origin Alternative Questionnaire Experiment. Decennial Statistical Studies Division and Population Division, U.S. Census Bureau. Retrieved from: https://www.census.gov/2010census/pdf/2010_Census_Race_HO_AQE.pdf

this adjustment process has been used for some time, the magnitude of the adjustment began to increase exponentially in recent decades. Over time, it has presented a wider disconnect between the full racial/ethnic enumeration of the U.S. population and the baseline foundation for other demographic surveys, and a growing issue of concern for Federal information quality.

Nonresponse to Race/Ethnicity in Other Federal Surveys

Among other Federal agencies (which are not permitted to offer the “Some Other Race” response category),³² survey data generally had nonresponse rates for combined or separate race and ethnicity questions below 6 percent. When Separate Questions were asked, the Hispanic origin question had lower rates of nonresponse than did race. Taking the example of surveys conducted by the National Center for Education Statistics, the National Household Education Surveys Program reported an overall nonresponse rate of 5.3 percent. Nonresponse varied between 1.9 percent and 4 percent for selection of race, and 1.5 percent and 3.4 percent for response to ethnicity on the National Survey of College Graduates, depending on the mode of survey. The most recent Current Population Survey, conducted for the Bureau of Labor Statistics, had a nonresponse rate of 0.7 percent for the question on ethnicity and 4.2 percent for selection of race. Race and ethnicity questions on the Medicare Current Beneficiary Survey, conducted by the Centers for Medicare and Medicaid Services, had nonresponse rates of less than 1 percent, as did the National Crime Victimization Survey, conducted by the Bureau of Justice Statistics. Recent data from another survey conducted by the Bureau of Justice Statistics, show that among persons interviewed in person while under the custody of State and Federal correctional officials, 0.1 percent did not respond to the ethnicity question, and 0.5 percent did not know or refused to answer the race question (2004). Of those inmates not responding to the race question (0.5 percent), 90 percent reported Hispanic ethnicity. In another survey conducted by the Bureau of Justice Statistics in 2012, which asked prisoners asking about sexual victimization while in custody, 10.1 percent of respondents did not respond to the race question, of which 89 percent self-identified as Hispanic.

In addition, the Subgroup reviewed extant cognitive testing research on question format and wording regarding race and ethnicity questions. The National Center for Health Statistics conducted cognitive interviews to improve the National Health Interview Survey question on race.³³ Four variants of question wording were tested: cultural, social, administrative, and ancestral. Results favored the cultural version of question wording, “Do you feel like you belong to or relate to one of these groups [READ GROUPS] more than the other?” as providing more specific responses and therefore more likely to reduce both response burden and non-response. Lastly, the General Social Survey (GSS) changed its race and ethnicity questions in 2001 to follow the 1997 standards. During the change, they analyzed the improvement on measurement and the effect it had on data continuity. The 1997 standards were compared to the standard GSS interview observation item on race and the GSS respondent racial self-identification item. Smith (2001) concluded that making the change would have a minimal impact on the GSS time series analyses.³⁴ Proportions of White and

³² Two Federal agencies have been granted exceptions to the current Federal standards to include a response category for “Some Other Race:” 1) the U.S. Census Bureau in response to action by Congress, and 2) the National Center for Health Statistics on the U.S. standard birth and death certificates in order to maintain comparability between the demographically related data systems of vital statistics and the decennial census.

³³ Miller, Kristen and Stephanie Wilson. (2002). “Cognitive Testing of NCHS Race Questions Interviews Conducted July 8 – September 23, 2002 in the Questionnaire Design Research Lab.” Hyattsville, MD. National Center for Health Statistics.

³⁴ Smith, W. Tom. (2001). “Aspects of Measuring Race: Race by Observation vs. Self-Reporting and Multiple Mentions of Race and Ethnicity.” University of Chicago. NORC.

Black respondents may decline slightly along with a slight increase in the “Other” race category, but the improvement in Hispanic ethnicity measurement would allow for more detailed analyses to be conducted.

Nonresponse in Federal Administrative Records

Administrative data collections tend to have higher rates of item nonresponse to both the Combined Question and the Separate Questions compared to self-response surveys. A complicating factor involved in administrative data collections is the reporting level. Often a Combined Question is used for gathering aggregated data from reporting organizational units; however, the organizational units contributing the data are expected to use the Separate Questions format for obtaining initial data. Local organization data may be “filled in” through observational data, imputations, or even guesses by individuals asked to provide data on behalf of others. This makes interpretation of item nonresponse very difficult, as it is not always clear whether or to what extent some of the original data were estimated. For example, in several sources of education data, item nonresponse frequently comes in the form of administrative nonresponse when entire schools, districts, or States have missing data. The impact of using a Combined Question on this type of organizational nonresponse is not known. It is noted here as an unresolved source of item nonresponse.

BJS found that rates of item nonresponse generally improved over time in its collection of both individual-level and aggregate counts of race and ethnicity in U.S. correctional populations, as local, State, and Federal authorities’ offender management systems were improved and began to capture race and ethnicity data according to the current standards. Nonetheless, in 2014, no racial category was assigned for 6 percent of State prisoners, and 20 percent were missing data for Hispanic origin. Further, in 2014, 31 percent of probationers and 7 percent of parolees (individuals supervised in the community at the county and State/Federal levels, respectively) had no response recorded for the combined measure of race/ethnicity.

Appendix 3: Detailed Analysis of Subgroup 1 Research Questions on Combined Question

Research Question 1:

a) For Federal self-response data collections that include a “some other race” response option, including censuses and surveys, how does question format affect the accuracy of collected race/ethnicity data? How does question format affect accuracy in the reporting of self-response race/ethnicity data?

b) Do these effects persist in self-response data collections that do not offer a “some other race” response option?

c) For Federal administrative data collections, how does question format affect the accuracy of collected race/ethnicity data?

The Subgroup considered whether proposed solutions for different question formats would improve, have no effect, or reduce observed accuracy of race and ethnicity data collection and reporting. The anticipated impact of proposed solutions on data flows within the Federal statistical system were also considered. These considerations include the potential impact on race/ethnicity population estimates, which are used by nationally representative surveys for sample design, sampling fractions, and post-stratification weights. For the purposes of the present inquiry, the Subgroup focused its analysis on data accuracy, including (by association) nonresponse. The Subgroup also noted the importance of considering operational, burden, and cost impacts when proposing its recommendation to OMB; the Subgroup plans to review this aspect more fully upon receipt of requested public comments.

Federal self-response data collections that include a “some other race” response option, including censuses and surveys

The Subgroup examined the patterns of self-reported race and ethnicity by question format in the NCT—a data source that includes a “Some Other Race” response option. (See Tables 3 and 4 below.) The observed difference in the distribution of race and ethnicity reporting between the Separate Questions format and the Combined Question with Write-In Areas format was statistically significant. The observed difference in the distribution of race and ethnicity reporting between the Separate Questions format and the Combined Question with Detailed Checkboxes format was also statistically significant. These results held for both Internet and paper responses. A comparison of race and ethnicity distributions between the Separate Questions and the Combined Question, without either write-in areas and/or detailed checkboxes, was not conducted. In other words, all versions tested in the NCT, for both the Separate Questions and the Combined Question approaches, included write-in lines and/or detailed checkboxes (for the Combined Question approach). The NCT was designed primarily to inform procedures for the 2020 Census, and not necessarily inform procedures used more customarily by other Federal agencies, which accounts for the absence of testing for this form of the Combined Question.

Table 3. Distribution of Hispanic Responses and Non-Hispanic Responses by Question Format – Internet Responses

		Separate Question		Combined Question with Write-In Response Areas		Combined Question with Detailed Checkboxes	
Hispanic	Hispanic alone	0.4	(0.02)	8.9	(0.29)	8.9	(0.29)
	Hispanic + SOR alone	4.4	(0.19)	0.1	(0.01)	0.1	(0.01)
	Hispanic + Other Major Group(s)	6.6	(0.11)	3.6	(0.07)	3.3	(0.07)
Not Hispanic	White alone	66.3	(0.49)	65.4	(0.46)	66.7	(0.47)
	Black alone	6.7	(0.34)	6.7	(0.33)	6.9	(0.34)
	Asian alone	6.1	(0.15)	6.3	(0.15)	6.2	(0.13)
	AIAN alone	0.3	(0.02)	0.3	(0.02)	0.3	(0.02)
	MENA alone	0.3	(0.02)	0.3	(0.02)	0.3	(0.02)
	NHPI alone	0.1	(0.01)	0.1	(0.01)	0.1	(0.01)
	SOR alone	0.2	(0.01)	0.1	(0.01)	0.1	(0.01)
	Multiple Responses	7.1	(0.1)	7.2	(0.09)	6.0	(0.1)
	Invalid	0.6	(0.03)	0.3	(0.02)	0.3	(0.02)
	Missing	1.0	(0.04)	0.8	(0.04)	0.7	(0.04)

Source: U.S. Census Bureau, 2015 National Content Test data. Note: Estimates are weighted with standard errors in parentheses.

For Internet respondents, persons self-identifying as being of Hispanic origin were less likely to report that they were Hispanic and “Some Other Race” when the Combined Question format was used featuring either write in or detailed check boxes than when the Separate Questions format was used (0.1 percent versus 4.4 percent). A similar pattern for the Separate Questions versus the Combined Question was observed for persons answering the questions in the NCT by paper (Table 4).

Table 4. Distribution of Hispanic Responses and Non-Hispanic Responses by Question Format – Paper Responses

		Separate Question		Combined Question with Write-In Response Areas		Combined Question with Detailed Checkboxes	
Hispanic	Hispanic alone	5.0	(0.21)	14.7	(0.29)	13.5	(0.46)
	Hispanic + SOR alone	3.8	(0.17)	0.1	(0.01)	0.1	(0.03)
	Hispanic + Other Specific Group(s)	8.9	(0.25)	2.6	(0.08)	7.0	(0.23)
Not Hispanic	White alone	56.6	(0.7)	58.5	(0.67)	55.3	(0.8)
	Black alone	14.3	(0.71)	14.1	(0.65)	12.5	(0.65)
	Asian alone	3.9	(0.21)	3.9	(0.15)	3.6	(0.23)
	AIAN alone	0.6	(0.04)	0.6	(0.03)	0.5	(0.06)
	MENA alone	0.2	(0.04)	0.2	(0.03)	0.4	(0.07)
	NHPI alone	0.1	(0.02)	0.2	(0.02)	0.1	(0.03)

		Separate Question		Combined Question with Write-In Response Areas		Combined Question with Detailed Checkboxes	
	SOR alone	0.2	(0.02)	0.2	(0.02)	0.1	(0.02)
	Multiple Responses	4.3	(0.14)	3.7	(0.1)	5.7	(0.22)
	Invalid	0.3	(0.04)	0.4	(0.03)	0.3	(0.05)
	Missing	1.9	(0.09)	0.8	(0.04)	0.8	(0.08)

Source: U.S. Census Bureau, 2015 National Content Test data. Note: Estimates are weighted with standard errors in parentheses.

Federal self-response data collections that do not offer a “Some Other Race” response option, including censuses and surveys

The analyses conducted to date of self-response data collections have not included collections that do not feature a “Some Other Race” category. It is not clear that the patterns observed when a “Some Other Race” option is offered extend to those collections where it is not offered. Further, given that two of 127 Federal agencies in the Federal statistical system are permitted to offer a “Some Other Race” response category through special, limited provision, it would be important to better understand the magnitude and scope of the issue when considering anticipated benefits of changes to the existing standard.

Therefore, the Subgroup plans to extend its current research findings by examining the accuracy of self-reported race/ethnicity collected using a Separate Questions versus Combined Question format in the context of surveys that do not feature a “Some Other Race” response option.

Federal administrative data collections

The Subgroup considered the effect of question format on the accuracy of race/ethnicity data reported through administrative data systems. Administrative records are used increasingly by Federal agencies to produce relevant Federal statistics in a more efficient, cost-effective manner than is possible with survey or census information collections. In fact, in several cases, Federal agencies rely on administrative records as principal data sources for their key statistics. Some of these administrative record data sets are Federal, but many (if not most) are managed at the local and State level. And, although localities and States may choose to conform to Federal statistical standards, they are not required to do so. The discrepancies observed between collection and reporting at local and State levels, and Federal standards can mean that Federal statistics produced through administrative records do not conform to the Federal standards in all cases.

For example, in many local and State administrative databases, the functioning of the information system limits respondents to choosing 1) a single racial group and 2) a binary question on Hispanic origin, rather than providing the opportunity for the respondent to “mark all that apply.” Such databases will undercount persons identifying with more than one race. However, if a Combined Question format were adopted at a local or State level, without a change to the information system functionality, respondents would be limited to choosing only one response to report their race/ethnicity—the result of which may be even less accurate information than is currently provided through many local and State administrative record data sets and subsequently used to produce Federal statistics. See the example in Table 5.

Table 5. Example: Reporting of Race/Ethnicity by a Person Identifying as Black, White and Hispanic, by Question Format and Data Source

	Separate Format in Typical Federal Survey or Census (Some Other Race not permitted)	Separate Format in Typical Administrative Record System (forces selection of one category)	Combined Format in Typical Administrative Record System (forces selection of one category)
Q1: Race	Black AND White	Black OR White	Black OR White OR Hispanic
Q2: Ethnicity	Hispanic	Hispanic	--

For respondents whose most salient race/ethnicity identity is Hispanic, the possible loss of additional detail if a combined format were used to produce Federal statistics from local and State administrative records may not be a great concern; the additional detail was not providing additional relevant information from the perspective of those respondents. However, for respondents who identify not only as Hispanic but also with a specific race group, using a Combined Question format when producing Federal statistics from administrative records could result in the loss of relevant information due to the functioning of information technology systems undergirding record systems outside the authority of the Federal government.

As a further example of this issue, Table 6 shows distributions of race/ethnicity statistics when reported using a 1) Combined Question format and a 2) Separate Questions format. In all cases, the statistics were produced drawing from State department of corrections (DOC) individual-level records collected using separate race/ethnicity questions.

Table 6. Race Distribution for State Prisoners by Question Format

Race/Hispanic origin ³⁵	(A) Combined race/ethnicity question (State prisoners, National Prisoner Statistics program 2014)	(B) 2-question format (State prisoners, National Corrections Reporting Program 2014, 20 percent of Hispanic origin missing)	(C) 2-question format (State prisoners, National Corrections Reporting Program 2014, 9 percent of Hispanic origin missing)	(D) 2-question format (State prisoners, National Corrections Reporting Program 2014, assuming all data missing for Hispanic origin means non-Hispanic)
White, non-Hispanic	537,043	409,767	487,587	534,271
Black, non-Hispanic	537,200	379,915	439,170	507,530
Hispanic	201,075	232,374	232,394	232,394
American Indian/Alaska Native, non-Hispanic	18,564	10,941	14,667	19,572
Asian, non-Hispanic	7,141	4,834	5,479	5,674
Native Hawaiian/Pacific Islander, non-Hispanic	2,992	219	332	447
Two or more races,	778	674	874	896

³⁵The category, "Other" is provided by the State administrative record dataset; it is not a Federal collection or reporting category per the Federal standards on race/ethnicity.

Race/Hispanic origin³⁵	(A) Combined race/ethnicity question (State prisoners, National Prisoner Statistics program 2014)	(B) 2-question format (State prisoners, National Corrections Reporting Program 2014, 20 percent of Hispanic origin missing)	(C) 2-question format (State prisoners, National Corrections Reporting Program 2014, 9 percent of Hispanic origin missing)	(D) 2-question format (State prisoners, National Corrections Reporting Program 2014, assuming all data missing for Hispanic origin means non-Hispanic)
non-Hispanic				
Other, non-Hispanic	9,342	8,033	12,060	12,415
Missing/unknown	2,273	269,631	123,844	3,209
Total	1,316,407	1,316,407	1,316,407	1,316,407

Sources: Bureau of Justice Statistics, National Prisoner Statistics Program and National Corrections Reporting Program, 2014.

This table describes several challenges faced by Federal agencies when using local and State administrative records to produce Federal statistics. In several States, Hispanic origin is an optional variable resulting in missing data on approximately 20 percent of National Corrections Reporting Program (NCRP) records. Localities and States are not required to conform to the Federal standards requiring the collection of both race and (Hispanic) ethnicity. Columns B-D show variations of the two, Separate Questions format from NCRP, making assumptions about the reclassification of missing Hispanic origin. When State DOCs analyze their data to obtain an aggregate count to report through the National Prisoner Statistics (NPS) program (column A), it appears that they are assuming all individuals with missing data for Hispanic origin are non-Hispanic; the counts in column A most closely resemble those of column D, which makes the same assumption in the two separate questions NCRP data.^{36 37} The validity of this assumption can be questioned, and columns B and C provide alternate treatment of the missing data, including listwise deletion of all cases with missing data (column B). By making reasonable assumptions based on State DOC reporting methods, the Bureau of Justice Statistics estimates that about half of the 20 percent of missing records for Hispanic origin can be classified as ‘likely non-Hispanic’ (column C).³⁸

Further, from this table, it appears that there tend to be fewer reports of more than one race, non-Hispanic when a Separate Questions format is used to calculate Federal statistics using State administrative record data.

³⁶ Not all states submit NCRP data in a given year. In 2014, 46 states gave year-end prison records to BJS through NCRP. The distributions of race/ethnicity of submitting states were weighted to the national total for the state prison population, under the assumption that the states not in the data file did not differ substantially from the states that did submit data.

³⁷ NPS measures people under the jurisdiction, or legal authority, of state correctional authorities. NCRP is a collection of persons under the physical custody of state correctional authorities, plus those housed in privately operated facilities. The difference between the two measures are prisoners held in local jail facilities, which are included in jurisdictional counts but not in custody counts. In 2014, this represented 81,000 or 6% of the total state prison population. Therefore, the distributions will vary slightly between NCRP and NPS, but BJS routinely uses the detail in the NCRP to represent the total state prison population by weighting up to national counts.

³⁸ In 11 percent of cases with missing Hispanic origin, the state departments of corrections only submitted data for Hispanic origin if the person was of Hispanic origin (“yes” on the yes/no question). BJS made the assumption all non-yes responses for these states were indicative of non-Hispanic ethnicity.

Another complexity arises when considering administrative record data sources that collect race/ethnicity information by observation, rather than self-report. Third-person observers are less likely to correctly describe the racial background of persons who would self-identify as two or more races. Similarly, the accuracy of observed race/ethnicity information may differ in cases where the (initial) intended use of the data is to facilitate rapid, rather than precise, visual identification of a person (for example, in an institutional setting).

Federal agencies are required to collect and report persons of Hispanic origin as a separate category regardless of race, but this does not impede Federal agencies from also reporting Hispanic origin by identified race. The Separate Questions format, even if, as administered by a non-Federal database, would only allow for a single response to the racial question, would still allow statistical agencies to capture and report race by ethnicity. Particularly for small racial groups that may show high rates of Hispanic ethnicity (such as persons identifying as American Indian/Alaska Native), the racial information would be lost in a database that forced a single response to a combined race/ethnicity question. In addition, based on their knowledge of the population or understanding of the database, statistical agencies could make reasonable imputations for missing race or ethnicity data under the two Separate Questions format if either of the questions is missing.

Research Question 2:

a) In self-reported surveys where “Some Other Race” is offered as a response category, does a combined race/ethnicity question provide more reliable responses than the current two Separate Questions format?

b) In self-reported surveys where “Some Other Race” is not offered as a response category, does a combined race/ethnicity question provide more reliable responses than the current two Separate Questions format?

Federal self-response data collections that include a “Some Other Race” response option, including censuses and surveys

To examine whether survey self-responses are more reliable when using a Separate Questions format or a Combined Question format where a “Some Other Race” category is offered, the Subgroup compared rates of consistency between responses to the 2015 NCT self-response survey and the 2015 NCT re-interview. Consistency was compared across all current minimum OMB reporting categories, the “Some Other Race” category, and a distinct reporting category for Middle Eastern or North African. The rates of consistency are shown in Table 7. As in the current standard, in both the Separate Questions format and the Combined Question format, respondents were permitted to select one or more categories. Two different approaches were tested with the Combined Question format: one approach utilized write-in response areas to collect detailed data, and the other approach utilized a series of detailed checkboxes and write-in areas to collect detailed data. A comparison of consistency in race and ethnicity distributions between the Separate Questions and the Combined Question, without either write-in areas and/or detailed checkboxes, was not conducted. In other words, all versions tested in the NCT, for both the Separate Questions and the Combined Question approaches, included write-in lines and/or detailed checkboxes (for the Combined Question approach). The NCT was designed primarily to inform procedures for the 2020 Census, and not necessarily inform procedures used more customarily by other Federal agencies, which accounts for the absence of testing for this form of the Combined Question.

Table 7. Consistency Between Self-Response and Re-interview for Major Race/Ethnic Groups by Question Format³⁹
(In percent)

	Alone or In Combination Groups							
	White	Hispanic	Black	Asian	AIAN	MENA	NHPI	SOR
Separate Questions Approach	96.9 (0.24)	90.3 (0.88)	97.2 (0.54)	96.2 (1.02)	56.7 (2.84)	68.9 (7.35)	44.5 (9.40)	31.0 (7.82)
Combined Question with Write-In Response Areas	94.8 (0.29)	93.4 (0.83)	96.2 (0.81)	95.1 (1.17)	48.6 (2.38)	63.0 (5.66)	55.2 (7.05)	13.1 (4.26)
Combined Question with Detailed Checkboxes	95.8 (0.26)	94.9 (0.78)	95.9 (0.78)	96.5 (0.76)	55.0 (2.83)	75.9 (6.15)	63.2 (9.18)	12.8 (6.19)

Source: U.S. Census Bureau, 2015 National Content Test data. Note: Estimates are weighted with standard errors in parentheses.

When comparing the Combined Question with Detailed Checkboxes format to the Separate Questions format, respondents that identified as Hispanic had a statistically higher rate of consistency in the Combined Question with Detailed Checkboxes approach. However, the increase in consistency is about 1 percent when standard errors are considered. By comparison, the Separate Questions format had a significantly higher consistency rate for respondents that identified as White when compared to either of the Combined Question formats tested.

The Subgroup also examined distributions of consistency in reported race among Hispanics only when a Separate Questions approach or a Combined Question approach is used. See Table 8.

Table 8. Race Distribution for Hispanic Respondents by Question Format – Internet Responses
(In percent)

	White alone	Black alone	Asian alone	AIAN alone	MENA alone	NHPI alone	SOR Alone	Two or More	Invalid	No Other Major Categories Reported*
Separate Question	16.5 (0.46)	1.4 (0.12)	0.7 (0.07)	1.0 (0.09)	0.1 (0.02)	0.1 (0.03)	39.0 (0.92)	38.2 (0.73)	1.1 (0.11)	2.0 (0.14)
Combined Question with Write-In Response Areas	19.4 (0.76)	1.3 (0.09)	0.8 (0.07)	1.3 (0.11)	0.1 (0.03)	0.1 (0.02)	0.6 (0.07)	5.5 (0.25)	0.8 (0.07)	70.1 (0.93)
Combined Question with Detailed Checkboxes	18.4 (0.74)	1.6 (0.12)	0.9 (0.08)	0.8 (0.07)	0.1 (0.02)	0.1 (0.01)	0.7 (0.09)	4.8 (0.24)	0.6 (0.06)	72.1 (0.91)

*Includes responses of Hispanic when no other major category (White, Black, Asian, AIAN, MENA, NHPI, or SOR) is reported. Source: U.S. Census Bureau, 2015 National Content Test data.

³⁹ Please note regarding Table 7 and Table 8: In the current 1997 OMB Standards, Middle Eastern and North African (MENA) responses are classified within the "White" racial category, following the OMB definition of "White" as "people with origins in the original peoples of Europe, the Middle East, or North Africa." In order to analyze MENA responses in the 2015 NCT, a separate MENA analytical category was developed. The inclusion of this MENA category does not mean that it has been defined as a distinct racial or ethnic category. Such a determination is a policy decision, and will be the purview of the current OMB review of the 1997 OMB Standards.

Note: Estimates are weighted with standard errors in parentheses.

When comparing the Separate Questions format to the Combined Question with Write-In Response Area format, the percentage of Hispanics that provided a “Some Other Race” response was statistically lower, and substantial, for the Combined Question with Write-In Response Area format (about 0.6 percent compared with about 39 percent in the Separate Questions format). The percentage of Hispanics that reported “Two or More categories” was statistically lower for the Combined Question formats tested (about 5.5 percent, compared with about 38 percent in the separate questions approach). When comparing the Separate Questions design to the Combined Question with Detailed Checkboxes design, the percentage of Hispanics that selected no other specific race category was significantly, and substantially, higher for the Combined Question formats tests compared to the Separate Questions format (about 70 percent compared with 2.0 percent).

The Subgroup also examined rates of item nonresponse by the Separate Questions and the Combined Question format. The Combined Question format has one item nonresponse rate for the race/ethnicity question, whereas the Separate Questions approach has three different item nonresponse rates that can be analyzed: one for the Hispanic origin question, one for the race question, and one that considers whether there is a response to both questions. See Table 9 for the pattern of nonresponse for the Internet response and Table 10 for the pattern of nonresponse for the paper responses.

Table 9. Pattern of Nonresponse by Question Format – Internet Responses

What percentage of respondents provide...?	Hispanic Origin Question			Race/Ethnicity Question			Both Questions		
	No Response	Invalid Response	No Valid Response	No Response	Invalid Response	No Valid Response	No Response	Invalid Response	No Valid Response
Separate Questions	0.8 (0.04)	0.1 (0.01)	1.0 (0.04)	1.2 (0.05)	0.7 (0.03)	2.0 (0.06)	0.8 (0.04)	0.0 (0.01)	0.8 (0.04)
Combined Question with Write-In Response Areas	N/A	N/A	N/A	0.8 (0.04)	0.3 (0.02)	1.1 (0.04)	N/A	N/A	N/A
Combined Question with Detailed Checkboxes	N/A	N/A	N/A	0.7 (0.04)	0.3 (0.02)	1.0 (0.04)	N/A	N/A	N/A

Source: U.S. Census Bureau, 2015 National Content Test data. Note: Estimates are weighted with standard errors in parentheses.

Table 10. Pattern of Nonresponse by Question Format– Paper Responses

What percentage of respondents provide...?	Hispanic Origin Question			Race/Ethnicity Question			Both Questions		
	No Response	Invalid Response	No Valid Response	No Response	Invalid Response	No Valid Response	No Response	Invalid Response	No Valid Response
Separate Questions	7.8 (0.25)	0.2 (0.03)	8.0 (0.25)	6.8 (0.22)	0.4 (0.04)	7.2 (0.23)	0.8 (0.06)	0.0 (0.01)	1.0 (0.06)
Combined Question with Write-In Response Areas	N/A	N/A	N/A	0.8 (0.04)	0.4 (0.03)	1.2 (0.05)	N/A	N/A	N/A
Combined Question with Detailed Checkboxes	N/A	N/A	N/A	0.8 (0.08)	0.3 (0.05)	1.2 (0.11)	N/A	N/A	N/A

Source: U.S. Census Bureau, 2015 National Content Test data. Note: Estimates are weighted with standard errors in parentheses.

For both Internet and paper responses, the percentage of no valid response for the race/ethnicity question was significantly higher with the Separate Questions format compared to the Combined Question formats tested. However, the size of the difference observed in the Internet administration was very small—most likely attributed to the electronic edit message that prompted a respondent to answer a question if a question was left blank.

Federal self-response data collections that do not include a “Some Other Race” response option, including censuses and surveys

The analyses conducted to date of self-response data collections have not included collections that do not feature a “Some Other Race” category. It is not clear that the patterns observed when a “Some Other Race” option is offered extend to those collections where it is not offered. Further, given that two of 127 Federal agencies in the Federal statistical system are permitted to offer a “Some Other Race” response category through special, limited provisions, it would be important to better understand the magnitude and scope of the issue when considering anticipated benefits of changes to the existing standard.

Therefore, the Subgroup plans to extend its current research findings by examining the accuracy of self-reported race/ethnicity collected using a Separate Questions format versus a Combined Question format in the context of surveys that do not feature a “Some Other Race” response option.

Appendix 4: Detailed Analyses on a New MENA Category

The 2015 NCT research examined the degree to which respondents identified with a distinct MENA category. Responses were compared when a distinct MENA category was offered and when it was not. The NCT found that respondents select the distinct MENA category when it is available. When the distinct MENA category is not available, MENA respondents tend to identify with White or “Some Other Race.” See Table 11.

Table 11. Reporting of MENA Responses by Presence of Distinct MENA Category
(Total percent of population identifying as MENA⁴⁰ in self-response data)

Question Format	Total Identified as MENA	In which category was MENA response provided?				
		White	Black	MENA	SOR	Another category
Question with Distinct MENA Category	0.9(0.03)	20.0(0.93)	0.6(0.21)	78.8(1.03)	3.3(0.43)	1.0(0.27)
Question with NO MENA Category	0.9(0.03)	85.5(0.99)	1.3(0.26)	N/A	11.5(0.88)	2.9(0.52)

Source: 2015 National Content Test data; U.S. Census Bureau. Note: Estimates are weighted with standard errors in parentheses.

The NCT re-interviewed respondents after receiving their initial responses to better understand how participants interpreted category meanings. The sample of respondents who were re-interviewed included not only respondents within the working definition of MENA, but also included respondents who identified in the initial survey as White, Hispanic or Latino, Black or African American, Asian, American Indian or Alaska Native, Native Hawaiian or Pacific Islander, and “Some Other Race.” Findings indicate that MENA respondents had difficulty identifying as only MENA when no distinct category was available. However, some respondents who identified as MENA in the re-interview did not identify as MENA in any category in the self-response, even when a distinct MENA category was provided. See Table 12.

⁴⁰ Identifying as MENA per the working classification used by the US Census Bureau in the NCT.

Table 12. Reporting Patterns of the MENA Re-interview Population to Survey Question Formats
(Total percent of population identifying as MENA⁴¹ in the re-interview)

Re-interview Response	MENA Category	No MENA Category
Total	100.0	100.0
Identified as MENA Only	33.9 (4.92)	6.6 (2.63)
Identified as MENA AND White	24.7 (4.10)	60.0 (4.90)
Identified as MENA AND Another Group(s)	6.7 (1.91)	7.8 (2.68)
Did Not Identify as MENA	33.5 (5.24)	25.6 (3.81)
Missing/Invalid	-	-

Source: 2015 National Content Test data; U.S. Census Bureau. Note: Estimates are weighted with standard errors in parentheses. A dash (-) indicates there is a cell count of ten or fewer.

Analysis also was conducted to evaluate the Census Bureau’s working classification of MENA. Respondents indicating these groups tended to identify as MENA when presented with the distinct MENA category. When no MENA category was available, groups in the Census Bureau’s working classification of MENA most frequently reported in the White category. However, when a MENA category was available, all of the groups in the Census Bureau’s working classification, except for Israeli, most frequently reported their detailed origins in the MENA category.

From this analysis, groups not in the NCT MENA working classification generally did not identify as MENA. Groups not in the NCT MENA working classification were generally more likely to identify with other categories such as White, Black, Asian, etc. than with the MENA category. An exception to this trend is Afghan.

Table 13. Reporting of Detailed MENA Groups in Category Response Areas by Presence of Distinct MENA Category
(In Percent of Population)

Detailed MENA Group	In which category was the detailed MENA response provided?									
	White		Black		MENA		Some Other Race		Another category	
	MENA Category	No MENA Category	MENA Category	No MENA Category	MENA Category	No MENA Category	MENA Category	No MENA Category	MENA Category	No MENA Category
Lebanese	34.0 (2.59)	97.7 (0.66)	-	-	69.0 (2.63)	N/A	0.4 (0.20)	1.9 (0.63)	1.1 (0.64)	0.5 (0.25)
Iranian	14.7 (1.80)	80.7 (2.42)	-	-	85.0 (1.92)	N/A	2.1 (0.70)	13.8 (2.19)	1.0 (0.36)	8.3 (1.80)
Egyptian	8.4 (1.76)	91.0 (2.04)	-	2.3 (0.71)	91.5 (1.77)	N/A	2.3 (1.10)	7.1 (1.99)	-	-
Syrian	28.9 (3.71)	91.1 (3.11)	-	-	71.9 (3.78)	N/A	-	9.3 (3.17)	-	-

⁴¹ Identifying as MENA per the working classification used by the US Census Bureau in the NCT.

Detailed MENA Group	In which category was the detailed MENA response provided?									
	White		Black		MENA		Some Other Race		Another category	
	MENA Category	No MENA Category	MENA Category	No MENA Category	MENA Category	No MENA Category	MENA Category	No MENA Category	MENA Category	No MENA Category
Moroccan	13.0 (3.56)	66.7 (6.98)	3.3 (1.50)	10.7 (3.86)	75.2 (4.81)	N/A	12.8 (3.93)	21.6 (6.58)	-	-
Algerian	7.9 (4.50)	96.6 (3.58)	-	-	83.8 (9.71)	N/A	-	-	-	-
Arab	38.1 (9.19)	52.8 (5.79)	-	-	35.4 (8.85)	N/A	35.3 (10.15)	38.0 (5.27)	-	12.6 (6.21)
Assyrian	34.0 (10.08)	72.9 (6.82)	-	-	56.4 (9.59)	N/A	12.7 (5.61)	27.7 (6.81)	-	-
Chaldean	10.2 (3.84)	77.7 (9.02)	-	-	86.8 (4.35)	N/A	5.7 (2.29)	24.8 (9.09)	-	-
Iraqi	9.1 (2.69)	90.5 (4.34)	-	-	91.2 (2.69)	N/A	-	9.3 (4.34)	-	-
Israeli	49.6 (4.08)	90.9 (2.97)	-	-	38.9 (4.41)	N/A	10.6 (3.39)	4.8 (1.74)	-	-
Jordanian	10.0 (5.49)	94.2 (2.55)	-	-	94.4 (2.61)	N/A	-	-	-	-
Kurdish	-	99.7 (0.37)	-	-	86.6 (10.21)	N/A	-	-	-	-
Libyan	-	84.7 (16.55)	-	-	99.4 (0.70)	N/A	-	-	-	-
Middle Eastern	60.2 (8.38)	46.2 (7.20)	-	-	25.8 (8.52)	N/A	13.6 (3.98)	52.3 (7.07)	-	-
North African	35.3 (15.33)	51.2 (17.54)	-	-	31.4 (17.45)	N/A	-	37.0 (19.37)	-	-
Palestinian	8.7 (4.42)	96.5 (1.23)	-	-	90.2 (4.42)	N/A	1.5 (0.86)	3.5 (1.37)	-	1.1 (0.56)
Saudi Arabian	-	70.6 (12.84)	-	-	99.1 (0.87)	N/A	-	23.4 (11.92)	-	6.0 (3.22)
Tunisian	-	54.4 (30.03)	-	-	70.2 (13.87)	N/A	-	-	-	-
Yemeni	-	84.6 (12.57)	-	-	99.0 (0.61)	N/A	-	15.2 (12.58)	-	-
Afghan*	18.9 (5.88)	32.5 (8.65)	-	-	30.0 (6.88)	N/A	35.1 (8.14)	26.8 (7.69)	28.5 (7.14)	42.9 (9.51)
Armenian*	79.0 (1.87)	90.8 (1.71)	-	-	12.6 (1.58)	N/A	9.3 (1.13)	9.6 (1.72)	0.5 (0.30)	0.3 (0.14)
Azerbaijani*	82.9 (13.82)	69.9 (35.80)	-	-	-	N/A	-	-	-	-
Cypriot*	70.6 (30.57)	99.7 (0.30)	-	-	-	N/A	-	-	-	-
Georgian CIS*	86.6 (14.23)	100.0 (0.00)	-	-	-	N/A	-	-	-	-

Detailed MENA Group	In which category was the detailed MENA response provided?									
	White		Black		MENA		Some Other Race		Another category	
	MENA Category	No MENA Category	MENA Category	No MENA Category	MENA Category	No MENA Category	MENA Category	No MENA Category	MENA Category	No MENA Category
Somali*	-	-	94.2 (4.81)	96.2 (3.64)	-	N/A	4.8 (4.73)	-	-	-
South Sudanese*	-	-	100.0 (0.00)	-	-	N/A	-	-	-	-
Sudanese*	-	-	87.2 (7.58)	98.4 (1.09)	8.0 (6.70)	N/A	-	-	-	-
Turkish*	74.1 (4.16)	93.0 (1.95)	-	-	16.2 (3.53)	N/A	9.6 (2.50)	5.4 (1.81)	3.6 (1.93)	-

* Indicates the nationality or ethnicity is not included in the 2015 NCT working classification of MENA.

Note: The following groups were included for evaluation in the 2015 NCT, but because of the small populations sizes in the United States, the data are too sparse to show in this table: Bahraini, Bedouin, Berber, Druze, Emirati, Kuwaiti, Omani, "Other Middle Eastern," Qatari, Syriac, Western Saharan, Djiboutian, * Mauritanian, * and Turkish Cypriot. Note that Some Other Race (SOR) is not a reporting category according to the OMB race and ethnicity standard. Source: 2015 National Content Test data; U.S. Census Bureau. Note: Estimates are weighted with standard errors in parentheses. A dash (-) indicates there is a cell count of ten or fewer.

The NCT data were then analyzed to examine how offering a distinct MENA category affects reporting of other race/ethnicity groups. Initial analyses indicate that the inclusion of a distinct MENA category does not appear to affect the reporting of other major racial/ethnic groups, but further statistical analyses are underway. The percent of respondents who report as MENA was not affected by the inclusion of a MENA response category. See Table 14.

Table 14. Race/Ethnicity Group Distribution by Presence of Distinct MENA Category
(Percent of total population, self-response data)

Alone or in Combination Groups	Distinct MENA Category Included	No MENA Category
White	73.3 (0.66)	74.5 (0.64)
Hispanic	13.5 (0.25)	12.9 (0.29)
Black	10.6 (0.53)	10.1 (0.49)
Asian	6.6 (0.19)	6.7 (0.18)
AIAN	3.9 (0.05)	3.9 (0.05)
MENA	0.9 (0.03)	0.9 (0.03)
NHPI	0.4 (0.01)	0.3 (0.01)
SOR	3.9 (0.06)	4.1 (0.09)
Invalid	0.4 (0.02)	0.4 (0.02)
Missing	0.9 (0.03)	0.9 (0.03)

Note: Some Other Race (SOR) is not a reporting category according to the OMB race and ethnicity standard. Source: 2015 National Content Test data; U.S. Census Bureau. Note: Estimates are weighted with standard errors in parentheses.

Appendix 5: Detailed Results of Analyses on Relevance of Terminology

a) Instructions and Race/Ethnicity Terms

The 2015 NCT examined two alternative wordings of race/ethnicity item instructions: 1) “Mark [X] one or more boxes” and 2) “Mark all boxes that apply.” The Subgroup identified two criteria on which to base its evaluation: 1) which wording produces lowest nonresponse; and 2) which working produces greatest reporting of multiple race/ethnicities. Although nonresponse for these two options was similar (both for respondents reporting using paper or internet versions of the item), the “Mark all boxes that apply” wording was associated with more respondents reporting multiple races/ethnicities across both modes of data collection. See Tables 15 and 16.

Table 15. Percentage of Respondents Selecting Each Race/Ethnicity Category by Wording of Instructions/Terminology by Respondent Race and Mode of Administration-Internet

Race	“Mark [X] one or more boxes”	“Mark all boxes that apply”	“Race or Origin”	“Race or Ethnicity”	“Categories”
White only	69.4 (0.45)	68.7 (0.51)	69.3 (0.47)	69.0 (0.51)	68.9 (0.48)
Hispanic only	6.2 (0.18)	6.1 (0.20)	6.2 (0.19)	6.2 (0.21)	6.0 (0.19)
Black only	6.8 (0.33)	6.8 (0.34)	6.7 (0.32)	6.8 (0.35)	6.8 (0.34)
Asian only	6.1 (0.14)	6.1 (0.13)	6.1 (0.15)	6.3 (0.14)	5.9 (0.14)
AIAN only	0.3 (0.02)	0.3 (0.02)	0.4 (0.02)	0.3 (0.02)	0.3 (0.02)
MENA only	0.3 (0.02)	0.2 (0.01)	0.2 (0.02)	0.2 (0.02)	0.3 (0.02)
NHPI only	0.1 (0.01)	0.1 (0.01)	0.1 (0.01)	0.1 (0.01)	0.1 (0.01)
SOR only	1.0 (0.03)	0.9 (0.03)	0.9 (0.03)	1.0 (0.04)	1.0 (0.04)
Multiple races/ethnicities selected	9.0 (0.10)	9.9 (0.12)	9.1 (0.13)	9.2 (0.13)	10.0 (0.13)
Blank or Invalid ¹	0.8 (0.03)	0.8 (0.03)	0.8 (0.04)	0.9 (0.04)	0.8 (0.04)

¹ Invalid includes those respondents who selected every possible race/ethnicity category and did not provide a valid write-in response. Columns may not add to 100 due to rounding. Source: 2015 National Content Test data; U.S. Census Bureau.

Table 16. Percentage of Respondents Selecting Each Race/Ethnicity Category by Wording of Instructions/Terminology by Respondent Race and Mode of Administration- Paper

Race	“Mark [X] one or more boxes”	“Mark all boxes that apply”	“Race or Origin”	“Race or Ethnicity”	“Categories”
White only	58.5 (0.78)	57.5 (0.63)	58.5 (0.76)	57.3 (0.72)	57.1 (0.63)
Hispanic only	11.6 (0.31)	12.1 (0.23)	12.2 (0.31)	9.6 (0.31)	14.0 (0.29)
Black only	14.5 (0.62)	13.4 (0.65)	14.4 (0.62)	13.8 (0.71)	12.7 (0.67)
Asian only	3.7 (0.16)	3.7 (0.16)	3.7 (0.15)	3.7 (0.20)	3.7 (0.20)
AIAN only	0.6 (0.04)	0.5 (0.02)	0.6 (0.03)	0.6 (0.04)	0.5 (0.04)
MENA only	N/A	0.4 (0.04)	0.1 (0.02)	0.4 (0.05)	0.4 (0.06)
NHPI only	0.1 (0.02)	0.1 (0.02)	0.1 (0.02)	0.1 (0.03)	0.1 (0.02)
SOR only	0.9 (0.06)	0.7 (0.04)	0.9 (0.05)	0.6 (0.05)	0.8 (0.06)
Multiple races/ethnicities selected	9.0 (0.16)	10.3 (0.16)	8.4 (0.13)	12.5 (0.28)	9.8 (0.21)
Blank or Invalid	1.1 (0.05)	1.1 (0.04)	1.0 (0.05)	1.4 (0.07)	0.9 (0.07)

¹ Invalid includes those respondents who selected every possible race/ethnicity category and did not provide a valid write-in response. Columns may not add to 100 due to rounding. Source: 2015 National Content Test data; U.S. Census Bureau.

The Subgroup also examined three alternative wordings in items collecting data on race/ethnicity: 1) “race or origin,” 2) “race or ethnicity,” and 3) “categories.” The Subgroup investigated which of these wording options produces the lowest nonresponse and which produces the largest reporting of multiple race/ethnicities. See Tables 15 and 16. Patterns of nonresponse and selection of multiple races/ethnicities by terminology options differed by mode of administration (i.e., internet versus paper).

b) Clarification of Classifications

To guide its recommendations to clarify classification for persons identifying as Australian (including the original people of Australia/the Aborigines), Brazilian, Cape Verdean, New Zealander, and Papua New Guinean, the Subgroup examined the extent to which these groups were provided as write-in responses to the “some other race” category. Analyses were conducted both when the Middle Eastern or North African (MENA) category was included and excluded. Additionally, analyses were conducted using both the combined race/ethnicity question and the separate race and ethnicity questions.

None of these groups appeared in the top 10 write-in responses in any of these analyses. The Subgroup then reviewed the proportion of write-in responses to the “Some Other Race” category by geographic location. Of the five classifications under the current review, Brazilian was the most common write-in response (approximately 1 percent, weighted).

Table 17. Overall Top 10 “Some Other Race” Write-in Responses and Rankings for Classifications of Interest (MENA Category not Present)

Rank	Write-in response (Sorted by Weighted Percentage)	Weighted Percentage (Standard Error)	Unweighted Frequency (Unweighted Percent)
1	273 – Hispanic	15.6 (0.67)	4,476 (16.4)
2	211 – Mexican American	8.0 (0.39)	2,333 (8.6)
3	210 – Mexican	6.9 (0.36)	2,079 (7.6)
4	995 – American	6.4 (0.37)	1,617 (5.9)
5	996 – Uncodeable ¹	5.7 (0.36)	1,309 (4.8)
6	008 – White	4.7 (0.26)	1,023 (3.8)
7	272 – Latino	2.9 (0.24)	846 (3.1)
8	723 – Mixed	2.4 (0.19)	621 (2.3)
9	245 – Puerto Rican	1.9 (0.15)	638 (2.3)
10	313 – Black	1.9 (0.15)	384 (1.4)
...
26	705 – Brazilian	0.8 (0.12)	188 (0.7)
68	706 – Cape Verdean	0.2 (0.07)	40 (0.1)
122	187 – Australian	0.1 (0.04)	13 (0.0)
227	190 – New Zealander	0.0 (0.01)	-
241	701 – Aborigine	0.0 (0.01)	-
--	662 – Papua New Guinean	0 (N/A)	-

¹ Refers to cases when a respondent responded to the question but provided a response that was either coded as “invalid” or the reporting pattern was determined to be “invalid” based on reporting requirements. Examples include reporting “Human Being,” objections such as “None of your business.” Source: 2015 National Content Test data; U.S. Census Bureau. A dash (-) indicates there is a cell count of ten or fewer.

Table 18. Overall Top 10 “Some Other Race” Write-in Responses and Rankings for Classifications of Interest (MENA Category Present)

Rank	Write-in response (Sorted by Weighted Percentage)	Weighted Percentage (Standard Error)	Unweighted Frequency (Unweighted Percentage)
1	273 – Hispanic	15.8 (0.62)	4,959 (16.9)
2	211 – Mexican American	7.9 (0.34)	2,457 (8.4)
3	210 – Mexican	7.3 (0.35)	2,216 (7.6)
4	995 – American	6.8 (0.30)	1,776 (6.1)
5	996 – Uncodeable ¹	5.9 (0.32)	1,462 (5.0)
6	008 – White	4.1 (0.23)	984 (3.4)
7	272 – Latino	3.4 (0.23)	1,078 (3.7)
8	723 – Mixed	2.5 (0.17)	705 (2.4)
9	245 – Puerto Rican	2.5 (0.20)	975 (3.3)
10	313 – Black	1.9 (0.17)	440 (1.5)
...
22	705 – Brazilian	0.9 (0.11)	194 (0.7)
51	706 – Cape Verdean	0.3 (0.09)	68 (0.2)
152	187 – Australian	0.0 (0.02)	12 (0.0)
188	701 – Aborigine	0.0 (0.02)	-
321	190 – New Zealander	0.0 (0.00)	-
--	662 – Papua New Guinean	0 (N/A)	-

¹ Refers to cases when a respondent responded to the question but provided a response that was either coded as “invalid” or the reporting pattern was determined to be “invalid” based on reporting requirements. Examples include reporting “Human Being,” objections such as “This is none of your business.” Source: 2015 National Content Test data; U.S. Census Bureau. A dash (-) indicates there is a cell count of ten or fewer.

Table 19. Top 10 “Some Other Race” Write-in Responses and Rankings for Classifications of Interest, by Separate Questions vs. Combined Question Formats

Rank	Write-in Response (Sorted by Weighted Percentage)	Weighted Percentage (Standard Error)	Unweighted Frequency (Unweighted Percentage)
1	273 – Hispanic	24.2 (0.69)	9,170 (24.9)
2	211 – Mexican American	12.4 (0.40)	4,667 (12.7)
3	210 – Mexican	10.6 (0.39)	4,014 (10.9)
4	995 – American	4.9 (0.27)	1,622 (4.4)
5	272 – Latino	4.9 (0.25)	1,868 (5.1)
6	996 – Uncodeable ¹	3.4 (0.22)	1,048 (2.9)
7	245 – Puerto Rican	3.2 (0.21)	1,482 (4.0)
8	008 – White	2.7 (0.19)	794 (2.2)
9	270 – Latin American	2.2 (0.17)	875 (2.4)
10	274 – Spanish	1.6 (0.13)	610 (1.7)
...
18	705 – Brazilian	0.7 (0.10)	188 (0.5)
83	706 – Cape Verdean	0.1 (0.05)	22 (0.1)
100	187 – Australian	0.1 (0.03)	16 (0.0)
223	701 – Aborigine	0.0 (0.01)	-
--	190 – New Zealander	0 (N/A)	-
--	662 – Papua New Guinean	0 (N/A)	-

¹ Refers to cases when a respondent responded to the question but provided a response that was either coded as “invalid” or the reporting pattern was determined to be “invalid” based on reporting requirements. Examples include reporting “Human Being,” objections such as “This is none of your business.” Source: 2015 National Content Test data; U.S. Census Bureau. A dash (-) indicates there is a cell count of ten or fewer.

Table 20. Combined Question with Write-In Response Areas

Rank	Write-in Response (Sorted by Weighted Percentage)	Weighted Percentage (Standard Error)	Unweighted Frequency (Unweighted Percentage)
1	995 – American	9.9 (0.61)	1,043 (9.1)
2	996 – Uncodeable ¹	9.0 (0.65)	914 (7.9)
3	008 – White	7.0 (0.40)	689 (6.0)
4	723 – Mixed	4.3 (0.34)	510 (4.4)
5	313 – Black	3.8 (0.31)	332 (2.9)
6	078 – Irish	2.9 (0.29)	265 (2.3)
7	065 – German	2.5 (0.28)	271 (2.4)
8	109 – Portuguese	2.4 (0.28)	281 (2.4)
9	719 – Biracial	2.0 (0.26)	155 (1.4)
10	079 – Italian	1.9 (0.21)	211 (1.8)
...
24	705 – Brazilian	1.0 (0.20)	110 (1.0)
41	706 – Cape Verdean	0.6 (0.15)	50 (0.4)
181	187 – Australian	0.0 (0.02)	-
155	701 – Aborigine	0.0 (0.05)	-
287	190 – New Zealander	0.0 (0.00)	-
--	662 – Papua New Guinean	0 (N/A)	-

¹ Refers to cases when a respondent responded to the question but provided a response that was either coded as “invalid” or the reporting pattern was determined to be “invalid” based on reporting requirements. Examples include reporting “Human Being,” objections such as “This is none of your business.” Source: 2015 National Content Test data; U.S. Census Bureau. A dash (-) indicates there is a cell count of ten or fewer.

Table 21. Combined Question with Detailed Checkboxes

Rank	Write-in Response (Sorted by Weighted Percentage)	Weighted Percentage (Standard Error)	Unweighted Frequency (Unweighted Percentage)
1	996 – Uncodeable ¹	10.8 (0.78)	809 (9.8)
2	995 – American	8.8 (0.69)	728 (8.8)
3	008 – White	7.3 (0.56)	524 (6.3)
4	723 – Mixed	4.0 (0.37)	325 (3.9)
5	313 – Black	3.2 (0.43)	209 (2.5)
6	998 – Religious Responses ²	2.7 (0.42)	242 (2.9)
7	109 – Portuguese	2.6 (0.36)	183 (2.2)
8	311 – African American	2.1 (0.30)	154 (1.9)
9	M41 – AIAN, Tribe Not Specified	2.1 (0.32)	156 (1.9)
10	078 – Irish	2.0 (0.29)	149 (1.8)
...
24	705 – Brazilian	1.1 (0.23)	84 (1.0)
36	706 – Cape Verdean	0.6 (0.19)	36 (0.4)
173	187 – Australian	0.0 (0.04)	-

Rank	Write-in Response (Sorted by Weighted Percentage)	Weighted Percentage (Standard Error)	Unweighted Frequency (Unweighted Percentage)
190	190 – New Zealander	0.0 (0.03)	-
--	662 – Papua New Guinean	0 (N/A)	-
--	701 – Aborigine	0 (N/A)	-

Source: 2015 National Content Test data; U.S. Census Bureau.

¹ Refers to cases when a respondent responded to the question but provided a response that was either coded as “invalid” or the reporting pattern was determined to be “invalid” based on reporting requirements. Examples include reporting “Human Being,” objections such as “This is none of your business.”² Examples include “Jewish”, “Christian”, or “Hindu”. For the complete list refer to the “2015 NCT Race, Ethnicity, or Origin Code List” in the appendix of the 2015 National Content Test Study Plan Race and Ethnicity. A dash (-) indicates there is a cell count of ten or fewer.

Given the low prevalence of these geographic locations appearing as write-in responses and public comments to the September 30, 2016 FRN advocating against use of geographic regions in race/ethnicity group definitions, the Subgroup recommends no changes be made to the current standards to specifically incorporate the following geographic locations into any existing race or ethnicity category: Australian (including the original people of Australia/the Aborigines), Brazilian, Cape Verdean, New Zealander, Papua New Guinean.

The Subgroup also examined reporting patterns of South and Central American Indian respondents to determine if self-identification and reporting behavior is consistent with the current standard, which includes them in the American Indian or Alaska Native reporting category. Focus group data initially collected to inform the design of the 2015 NCT were examined. A review of those data suggest that Central and South American Indians do not tend to self-identify as American Indian or Alaska Native. However, the Subgroup noted more research and public input would be necessary before a more complete consideration of adding more specific South or Central American subgroups to the current description of the AIAN category in order to improve identification with the reporting category.

c) Presentation of and Terms Used for Detailed Groups

To determine which duplicate initial category mention of “Cuban” should remain, the Subgroup recommends this be guided by population size. The Subgroup also considered whether the current ordering of the classification list be updated to reflect current population size, taking note of the rationale provided for Subgroup ordering decisions related to the NCT. As a next step, the Subgroup plans to apply this rationale to the classification listing and determine the magnitude and benefit of any resulting changes.

The Subgroup also considered the relevance of terms used in the current standards. The Subgroup recommends that the term “Negro” be removed from the standards. This is in accordance with H.R. 4238, which was signed by President Obama on May 20, 2016 and with the Census Bureau’s decision to remove the term “Negro” in its information collections beginning in 2014. Further, the Subgroup recommends that the term “Far East” be removed from the current standards. The term has never been used by the Census Bureau when collecting data⁴²), and appears to be out of date with common

⁴² Pratt, B., Hixson, L., and N. Jones (2015). “Measuring Race and Ethnicity Across the Decades: 1790-2010.” U.S. Census Bureau. Retrieved from: http://www.census.gov/population/race/data/MREAD_1790_2010.html/

usage, and may be offensive to some groups.

Given the inclusion of “Hispanic” in these definitions of “minority,” and the relative prevalence of the Hispanic or Latino versus Black or African American populations, the Subgroup also considered whether referring to Black or African American as the “principal minority race” is still relevant, meaningful, accurate, and acceptable. The Subgroup plans to also consider if “Hispanic” should be among the groups considered, and the salience of alternative terms (e.g., “principal minority race/ethnicity”). “Hispanic” usually is considered an ethnicity while “minority” is usually used when referencing race. Changing “principal minority race” to “principal minority race/ethnicity supports the changing demographics of the U.S. population. Given that many of the groups classified as racial and ethnic minorities have experienced institutionalized or State-sanctioned discrimination as well as social disadvantage and oppression, many consider it to be important to continue identifying the principal minority group in Federal data collections and reporting systems.