Earlier this year the Office of Management and Budget (OMB) published a notice and request for comments about several aspects of the data that would be collected from the 2020 Census questionnaire. To restate the obvious for some readers who only focus on the federal census once a decade, at best, there is now only one questionnaire for the census: what used to be known as the short form. All the descriptive demographics that were formally collected via the long form are now collected on an ongoing basis through the American Community Survey (ACS).

The main topic here is whether the 1997 Standards for Maintaining, Collecting, and Presenting Federal Data on Race and Ethnicity as updated for the 2000 and 2010 Censuses should be changed with regard to the current two separate questions on Spanish origin (question 8) and race (question 9). Primarily the issues relate to 1) whether these two questions should be combined into one; and 2) whether there should be a new race category denominated as Middle Eastern and North African (MENA).

Role of the Census in the Federal Government. Let us first consider the focus of this information from the federal perspective. “Development of these Federal data standards stemmed in large measure from new responsibilities to enforce civil rights laws. Data were needed to monitor equal access to housing, education, employment opportunities, etc., for population groups that historically had experienced discrimination and differential treatment because of their race or ethnicity.”

Yet, as noted by several observers of the history of the census, the classifications that are determined to fulfill the requirements of federal agencies have non-government

---

1 Clark H. Bensen, B.A., J.D., consulting data analyst and attorney doing business as POLIDATA ® Polidata Data Analysis, a demographic and political research firm with offices in Washington, DC and VT. www.polidata.org
2 See Federal Register, Vol. 82, No. 39, Wednesday, March 1, 2017, 12242.
3 See https://www.whitehouse.gov/omb/fedreg_1997standards
implications\textsuperscript{4} and can also accomplish political objectives as Congress plays an integral role in the questions included in the census.

As stated in the OMB Notice the race “categories developed represent a socio-political construct designed to be used in the self-reported or observed collection of data on the race and ethnicity of major broad population groups in this country, and are not genetically-, anthropologically-, or scientifically-based.” Or, as paraphrased by one observer, a “socially constructed category freed from its origins in biology.\textsuperscript{5}”

The concerns of this note relate to these proposals with respect to:

a) Any effect on the so-called “PL” datasets that are published by the Bureau of the Census in response to the 1975 Act of Congress relating to “Tabulation of population for purposes of apportionment of state legislative bodies”;

b) Any impact on the apportionment process and litigation that inevitably accompanies the decennial redistricting cycle;

c) What appears to be a turn in the wrong direction, away from a society in which race would be less of a focus to one of increased focus; and

d) The lack of any demonstrated need for a new MENA category, either as a new race or ethnic category, as opposed to other persons similarly situated.

History of the Questions. Over recent decades the form of the basic census questionnaire has been modified with respect to the race and ethnicity questions. Since 1970, the first census following the reapportionment revolution, beginning after the decisions of the U.S. Supreme Court, notably \textit{Baker v. Carr} and \textit{Reynolds v. Sims}, and the 1965 Voting Rights Act (VRA), these questions have become more open ended in design\textsuperscript{6} and the data provided has increased in complexity\textsuperscript{7}.

The 1970 form had only one question in this regard, “4. COLOR OR RACE” with 9 basic choices (White; Negro or Black; Indian (Amer.); Japanese; Chinese; Filipino; Hawaiian; Korean; and Other) with a write-in blank (WIB) for either Indian (Amer.), “Print Tribe”, or Other, “Print Race”.

The 1980 form expanded in two ways: by the number of printed options for the unnamed race question and by having a separate question “Is this person of

\textsuperscript{4} Most recently: Schor, Paul; \textit{COUNTING AMERICANS: How the US Census Classified the Nation}; Oxford University Press, NY; 2017; translated from the original French version of 2009. See p.228.

\textsuperscript{5} See Schor at 3.

\textsuperscript{6} See https://www.census.gov/history/www/through_the_decades/index_of_questions/

\textsuperscript{7} Notably following the inclusion of multi-race responses which exploded the number of cells of data.
Spanish/Hispanic origin (SHO) or descent?”. This new question was two questions further in than the unnamed race question. The number of printed options increased from 9 to 15 but continued the WIB for the same two responses as in 1970. The printed options for the SHO question were No, plus Mexican, Mexican-Amer., Chicano; Puerto Rican; Cuban; and Other, but without any WIB option.

The 1990 form was largely along the lines of the 1980 form except that there were more WIB options for respondents. The race question incorporated the new Asian/Pacific Islander (API) race category and thus allowed two WIBs for three options: one for Indian (Amer.) and one for the Other API and Other race combined. The SHO question, again two questions after the now named “Race” question, was the same as in 1980 but it now included a WIB for other SHO groups with examples: “Argentinian, Columbian, Dominican, Nicaraguan, Salvadoran, Spaniard, and so on”.

The 2000 form was a slight modification over the 1990 form: the SHO question was basically the same, though now Spanish/Hispanic/Latino (SHL), still allowing one WIB but not listing any examples. The SHO question was now asked before the race question. The race question was basically the same in format aside from the new option for persons to “mark X one or more races”.

The 2010 form was almost identical to the 2000 form (with SHO before race) aside from a change in the SHO/SHL question to be Hispanic/Latino/Spanish origin (HLS) and the reintroduction of the 1990 examples for the WIB. The 2010 race question also included examples for Other Asian: “Hmong, Laotian, Thai, Pakistani, Cambodian, and so on” and for Other Pacific Islander: “Fijian, Tongan, and so on”.

The PL94-171 Datasets. Since its introduction following the 1980 Census the PL dataset has provided the basic census counts, at the census block and other geographic levels as designated, to state and local authorities with the responsibility of creating boundaries for representational districts. For 1980 and 1990 the datasets were a fairly simple and straightforward tabulation of the basic counts with breaks for the race and ethnicity questions. Beginning with the 2000 Census the PL dataset became much more complicated with tabulation of the hundreds of possible cross tabulations with the multi-race response options.

---

8 Public Law 94-171 (P.L. 94-171). Public Law (P.L.) 94-171, enacted in 1975, directs the Census Bureau to make special preparations to provide redistricting data needed by the fifty states. Within a year following Census Day, the Census Bureau must send the data agreed upon to redraw districts for the state legislature to each state's governor and majority and minority legislative leaders. To meet this legal requirement, the Census Bureau set up a voluntary program that enables participating states to receive data for voting districts (e.g., election precincts, wards, state house and senate districts) in addition to standard census geographic areas such as counties, cities, census tracts, and blocks. (https://factfinder.census.gov/help/en/public_law_94_171_p_l_94_171.htm) See the Census publication, https://www.census.gov/rdo/pdf/TheViewFromTheStates_2020.pdf.
For the 1980 and 1990 datasets the basic question confronting stakeholders in the districting phase of the apportionment process was simply how to choose the appropriate number between two options. That is, with two questions for race and ethnicity, the analyst could use, e.g., Black or Non-Hispanic Black. Using the first, inclusive form provided a higher number while the second, the exclusive form, having subtracted Hispanic Black, meant a lower number. The inclusive method meant that the sum of all groups could exceed 100% while the exclusive method meant that the sum of all groups would equal 100% of persons as each subgroup was mutually exclusive. This is a situation that has been exacerbated by the additional options for multi-race respondents.

It is no wonder that such a choice created confusion and the Department of Justice and OMB have weighed in on this at several points by attempting to clarify the issue through various iterations of guidance to end users of the census counts. Notwithstanding these attempts confusion can still be found in litigation as well as in how these numbers are used by the thousands of jurisdictions that use them for a variety of purposes in their governmental duties. Indeed, more clarification is needed without any changes.

Race, Ethnicity, or Ancestry. The first question posed in the Federal Register notice raises the bigger picture issue as to what the real purpose is for the race/ethnicity information already required of all census respondents. The question is, regardless of format, whether MENA should “be considered an ethnicity or a race?” A related word that does not appear in the notice is ancestry. A related concept has apparently already been addressed by some commenters with respect to the ethnoreligious groups.

---

9 See 28 CFR Ch. 1 (7-1-11 Edition), Part 51 with respect to section 5 of the VRA. See OMB Bulletin No. 00-02-Guidance on Aggregation and Allocation of Data on Race for Use in Civil Rights Monitoring and Enforcement, March 9, 2000.
10 Recent litigation involving a racial challenge for the Virginia legislative districts highlighted the dilemma facing those drafting district boundaries and the eventual litigants. There were competing numbers for the appropriate count of African Americans which revolved around the inclusion, or exclusion, of Hispanic Black persons. This situation was not an unusual one.
11 Middle Eastern and North African. “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.” Notice, op cit. at 12245.
12 Ancestry: Ancestry refers to a person’s ethnic origin or descent, “roots,” or heritage, or the place of birth of the person or the person’s parents or ancestors before their arrival in the United States. Some ethnic identities, such as “German” or “Jamaican,” can be traced to geographic areas outside the United States, while other ethnicities such as “Pennsylvania Dutch” or “Cajun” evolved in the United States. The intent of the ancestry question is not to measure the degree of attachment the respondent had to a particular ethnicity. For example, a response of “Irish” might reflect total involvement in an “Irish” community or only a memory of ancestors several generations removed from the individual. A person’s ancestry is not necessarily the same as his or her place of birth; i.e., not all people of German ancestry were born in Germany (in fact, most were not). (https://www.census.gov/population/ancestry/ )
13 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. Notice, op. cit. at 12245.
It is unclear what actual improvement in “the quality of race and ethnicity information collected and presented by Federal agencies” would be made by further revisions to the standards on behalf of the MENA proponents that could not also be made by singling out any other group of countries of origin. Why MENA? What would be gained by a more detailed set of questions that cannot already be provided by the ancestry questions, aside from having a more concrete number by having it on the actual census form as opposed to getting the information from the ongoing American Community Survey (ACS).

The notice raises this issue as well considering “the estimated size of the MENA group” and concerns about statistical reliability as well as confidentiality and privacy concerns. While the first concern calls into question the need for a new group, the latter concerns affect the implementation.

If there are concerns about confidentiality and privacy this detracts from the point of releasing it at the census block level, which is the lowest level for which the PL datasets include such information. Adding any privacy considerations through a redaction in the dataset would be quite problematic without adding an additional level of geography as all persons must be accounted for at the block level. It is the basic aim of the Bureau to account for all persons at their usual place of residence and to count them once and only once. Changing the standards so that the PL dataset would not allow this deserves much more consideration.

Some of the formats used in the 2015 National Content Test (NCT) seemed to be simply a simplified ancestry question for a preferred subset of all possible groups for which responses are tabulated. This could mean a reduction in the emphasis on the ancestry question in general. Moreover, just as the multi-race responses to the race question has exploded the PL dataset to unmanageable levels of complexity the addition of a new race category will explode it by another dimension.

Note should be made that the ancestry questions, which also allow for multiple responses, provide confusion in reporting as well. Tabulations allow for reporting of

---

14 The U.S. Census Bureau is committed to counting every person in the 2020 Census once, only once, and in the right place. The fundamental reason that the decennial census is conducted is to fulfill the Constitutional requirement (Article I, Section 2) to apportion the seats in the U.S. House of Representatives among the states. Thus, for a fair and equitable apportionment, it is crucial that the Census Bureau counts everyone in the right place during the decennial census. ([https://www.federalregister.gov/documents/2016/06/30/2016-15372/proposed-2020-census-residence-criteria-and-residence-situations](https://www.federalregister.gov/documents/2016/06/30/2016-15372/proposed-2020-census-residence-criteria-and-residence-situations))

15 This has several operational concerns as data analysts and re-publishers will pick and choose from the hundreds of variables offered. Importantly, when the Bureau produced corrections following the 2010 Census it did not provide corrected breakdowns for all the basis race and ethnicity counts.
only those who provided a single response, or those who reported more than one, and the question as to whether the denominator is the number of persons or the number of responses.

**Impacts.** It is likely that there would be some degree of confusion on the implementation of a new MENA category, especially if it is a new race category. Likewise, there may well be more opportunity for the judicial branch to try and deal with an already complex subject that is generally not well suited to the courtroom and best left to legislative bodies that are capable of dealing with multiple political concerns.

It remains unclear why the ACS data on ancestry cannot provide sufficient detail for what purposes are needed by the Federal agencies, at least with respect to MENA, in comparison to the cost and disruption to all redistricting stakeholders. Putting it on the 2020 Census form means that the data would be available at the level of the census block, which is the lowest level of the census for reporting the tabulations and the level in the PL dataset. Such data would thus be available during the redistricting following the 2020 Census.

This would lead to problems that will be encountered by redistricting stakeholders and those involved in litigation, either with the introduction of MENA as an enhanced group or with the combined questions. Litigation relating to districting operates on the cycle of the decennial census and comparisons are frequently made, even without the prospect for section 5 preclearance under the federal Voting Rights Act, census to census. If either of these proposals are adopted there must be some means by which the newly released data can be made comparable with previous census information.

In sum, the improvement in the quality of the data has not been demonstrated with respect to the MENA proposal. Such a proposal would increase confusion and thus increase the cost to all stakeholders as they enter into the next round of the redistricting process. Furthermore, it would go in the opposite direction from trying to become a color-blind society by enhancing the concept of Balkanization of political interests.

Enclosures:
[D:\PoliCorp\memo\wlhp\ncsl-2017-0806_census-proposals.doc~7/28/2017 2:28:00 PM]