



UNITED STATES DEPARTMENT OF COMMERCE
Economics and Statistics Administration
U.S. Census Bureau
Office of the Director
Washington, DC 20233-0001

AUG 15 2017

MEMORANDUM FOR: Barbara Anderson
Chair
Census Scientific Advisory Committee

From: Ron Jarmin
Performing the Nonexclusive Functions
and Duties of the Director
U.S. Census Bureau 

Subject: U.S. Census Bureau Responses to Census Scientific Advisory
Committee Spring 2017 Recommendations

The U.S. Census Bureau thanks the Census Scientific Advisory Committee for its recommendations. We are responding to the committee recommendations submitted during its March 30-31, 2017, meeting.

Your expertise is necessary to ensure that the Census Bureau continues to provide relevant and timely statistics used by federal, state, and local governments, as well as business and industry, in an increasingly technologically-oriented society.

Attachment

To: Census Bureau
From: Census Scientific Advisory Committee
Topic: Comments and Recommendations from Spring 2017 CSAC Meeting
Date: March 31, 2017

SECTION I: 2020 Census Update and 2020 Systems Update with Census Questionnaire Assistance and Internet Self-Response

There were two interesting, timely and interrelated presentations on the current situation in 2020 Census planning. We will discuss them together.

1. CSAC appreciates the update on the 2020 Census planning. Although we were reassured by much of the presentation, we are concerned that budgetary uncertainties could have implications for the success of the 2020 Census. Many of the innovations undertaken for 2020 seem to be motivated by reducing the costs compared to 2010, but we hope the Census Bureau will better articulate a more complete set of success metrics. We are especially interested in the metrics of accuracy (overcount and undercount) and attitudes toward the Census Bureau. For example, will self-response increase the overcount, and among which population subgroups? How might various decisions (e.g. collecting email addresses in online form; using IRS records) impact attitudes towards the Census?

CENSUS BUREAU RESPONSE: The Census Bureau agrees with CSAC's concerns and is working to achieve quality results.

For purposes of the communications campaign, we plan to understand attitudes toward the Census Bureau by way of the Census Barriers, Attitudes and Motivators Survey (CBAMS). However, the instrument does not contain attitudes about using administrative or 3rd party data in lieu of, or to supplement, self-response. That information is available, however, from the ongoing Trust in Federal Statistics module of the Gallup Overnight Tracking Poll. This research is conducted in ADRM in the Center for Survey Measurement.

2. We want to emphasize the importance of the Internet self-response—it will be the way most Americans interact with the 2020 Census. Although we could tell that improvements to the user experience had been made since the presentation about PRIMUS, the online interface has a clunky, outdated look that is not very user-friendly. It clearly needs a great deal of further work. We hope that the standard for evaluating the success of Internet self-response will be not only a high response rate, but also accurate responses and a positive user experience. It is not sufficient to follow a list of recommendations. The final version of the Internet Self-Response interface would be much improved by committing to a user-centered design process. Specifically, progress will continue to be slow unless the frequency of usability tests is increased, other types of user-centered activities like contextual inquiry sessions are employed, and crowdsourced usability tests are employed. They could use the participation of a graphic designer and survey expert in the design. Census should look at other online survey interfaces and try to match their quality.

CENSUS BUREAU RESPONSE: The Census Bureau agrees that Internet self-response is critical and is working to further improve the user experience and increase the likelihood of accurate responses.

3. Have there been tests to assure that very high usage will not crash the system at the time of the Census? We are interested in what measures are being taken to protect against sophisticated cyber attacks.

CENSUS BUREAU RESPONSE: The Census Bureau shares CSACs concerns about high usage crashing the system and the possibility of cyber attacks. Over the next three years, the Census Bureau will concentrate on developing solutions to mitigate the main cybersecurity threats associated with the 2020 Census that can be reused at an enterprise level across the agency. The Census Bureau will work with federal partners and industry providers to develop solutions and plans to address cybersecurity threats outside and within its direct control.

Continuing to build on the approach taken in the 2016 and 2017 Census Tests, and including the results of 2018 End-to-End Census Test, the Census Bureau has scheduled further tests of systems that will support the 2020 Census. It is also defining the systems' workload based on the results of internal and external demand models.

Performance testing of ECaSE Internet Self-Response capability with 2020 census workloads will be conducted by Technical Integrator in July/August 2017 to confirm system scalability. CEDCaP staff are working with the 2020 staff and the Technical Integrator to ensure the design and architecture of the systems meets 2020 Census demand models to confirm system scalability.

Over the next three years, the Census Bureau will concentrate on developing solutions to mitigate the main cyber security threats associated with the 2020 Decennial Census that can be reused at an Enterprise level across the Census Bureau. We will work with our Federal partners and Industry providers to develop solutions and plans to address Cybersecurity threats outside and within our direct control

Cyber threats are either External to the Census Bureau, on the Perimeter of the Census systems, and/or Internal to the Census Systems. The Census does not have direct control of mitigations for external cyber threats to the Census and will rely on industry and other federal agencies to help provide services that can take actions to resolve those threats. The Census Bureau can detect some of these External threats but cannot take direct actions to resolve them on our own and examples of these are bad actors impersonating the Census via fake websites or Phishing scams or even compromising external network access. For the threats in the perimeter and internal to the Census Systems, the Census has the ability to take direct actions to prevent and resolve these threats such as Data Breaches, Compromising User Devices and Disrupting the Internet Self Response Website.

4. If the race/ethnicity question is changed to a single question which allows the respondent to indicate their "race" and related nationality/subcategories, the related nationalities/subcategories appearing on a separate/later screen could negatively affect the results that have been improved by combining the race/ethnicity questions into one. Specifically it's possible the proximity of the subcategories could help the user to

understand the definition of each category. Consider the possibility of allowing that question to continue past "the fold" so that users see the category and subcategories together (as they do on the paper version). More, generally, this highlights an issue that needs to be researched to develop the best presentation.

CENSUS BUREAU RESPONSE: The initial screen for the combined race/ethnicity question that was used in the 2017 Census Test, and planned for the 2018 End-to-End Census Test, includes checkboxes for each of the major categories (White, Black, etc.) alongside detailed examples for these categories. For instance, next to the White checkbox is: "For example, German, Irish, English, Italian, Polish, French, etc." If a respondent selects the White checkbox, the next screen elicits details on the specific White ethnicities and those six examples are included as checkboxes (along with a write-in field to provide additional details). The Census Bureau's research and testing program has found this approach to work well, both in quantitative and qualitative testing opportunities. The Census Bureau does continually assess each of the census questions, including race/ethnicity, and appreciates this suggestion for future consideration for 2020.

5. There was an impressive list of Test Systems and Support Systems presented. Prioritization of the importance of these would be helpful. Especially which are most urgent for future decisions made by Census? Where can CSAC be most helpful?

CENSUS BUREAU RESPONSE: The primary 2020 IIP milestones are still in draft and decisions must be made in the development of core data collection systems from Census Enterprise Data Collection and Processing (CEDCaP), including the Enterprise Census and Survey Enabling (ECaSE) platform's Internet Self-Response, enumeration, Operational Control System (OCS), and Field OCS instruments, in addition to the Listing and Mapping Application (LiMA). CSAC can be most helpful in assessing the validity of the technical solutions. It is imperative that the Census Bureau ensures the solution architecture and the technology approach, including the cloud, is in alignment with industry standards.

6. What are the provisions made for the visually impaired to fill out the Internet questionnaire? Are there protocols for accessibility that must be observed, or are all those with visual impairments expected to use an alternative means of completing the Census?

CENSUS BUREAU RESPONSE: As with the 2017 Census Test instrument, the 2018 End-to-End Census Internet Self-Response instrument follows the GSA government-wide Section 508 Requirements and Standards, ensuring that it is accessible to all users, including those with disabilities. Section 508 compliance of the Internet Self-Response instrument was and will continue to be reviewed and tested by the Census Bureau's Center for Survey Measurement Usability Lab.

7. Is it common practice to ask for an email address on Census surveys? Does this require additional protocols regarding privacy or discourage completion? Perhaps it could be split tested in the 2018 end-to-end test.

CENSUS BUREAU RESPONSE: No. As standard practice, the Census Bureau doesn't ask for email addresses on its surveys. We've done some testing on this issue for 2020 and we

found requesting email addresses didn't prove to be helpful so we have no plans to do so in either the 2018 End-to-End Census Test or the 2020 Census.

ACS tested collecting email address. They found no difference in unit response rates. See report here: https://www.census.gov/content/dam/Census/library/working-papers/2016/acs/2016_Zelenak_01.pdf

8. We would like to know how the FAQs in the census operator interface are developed and vetted.

CENSUS BUREAU RESPONSE: Census Bureau staff works with internal subject-matter experts to develop the FAQs, which are then reviewed by leadership.

9. This presentation addressed the use of simulations to ascertain Census 2020's system readiness. This is an important approach that needs to be done carefully. Simulations need to express the "observing system" as realistically as possible. Presumably there will be a geographic component (both physical geography and more generally network-defined geography), a temporal component, and of course there is a "household" component that implies correlation between individuals in the same (strong correlation) household and in nearby (less-strong correlation) households. Systems can fail when the underlying state moves into new regimes. A good simulation experiment uses random variation to capture some of this. CSAC suggests that the random variables be simulated to reflect the sort of variability seen in the field. In that case normal distributions should not be the distributions of first choice; long-tailed distributions (e.g., extreme-value distributions) should be considered where appropriate. We suggest that the resulting simulation be considered a best-case scenario. Then, based on the simulated data, the analysis should stress-test the system in as many ways as possible, including the possibility of malicious attacks from third parties.

CENSUS BUREAU RESPONSE: For the 2020 Census, the Census Bureau is working with the TI to define a framework that addresses areas of security, fraud, scalability, availability, disaster recovery, continuity of operations, etc., under which the 2020 Census must operate. The Census Bureau is defining the key aspects that determine the feasibility of the system of systems to function as expected under varying simulations. For the approach to work successfully, the Census Bureau has given the TI a directive to acquire and provision simulated, response, para, and meta data. The Census Bureau is simulating the tests to contain geographic, temporal, and household components as defined by CSAC. The simulations take into consideration the results of internal and external data models. The Census Bureau will work with the TI to consider variation in simulations to reflect what would be seen in the field. Additionally, it will proactively identify simulations that represent extreme value distributions.

10. We would like to know the process for developing the language on the Internet Self-response form around homelessness.

CENSUS BUREAU RESPONSE: During the 2020 Census, the Enumerating Transitory Locations operation will enumerate individuals in transitory locations who do not have a usual home elsewhere, and the Group Quarters operation will enumerate people living or

staying in group quarters, and provide an opportunity for people experiencing homelessness and receiving service at service-based locations to be counted in the census.

With respect to the process for developing the language on the Internet instrument, subject matter experts -- representing the Group Quarters, Non-ID Processing, Content and Forms Design, and Internet Self-Response operations -- collaborated on developing the language around homelessness. The language was developed in a way that if a respondent does not have a street address or rural route address when filing out the Internet instrument, he or she is asked, "Were you experiencing homelessness on April 1?" The questionnaire then continues as the respondent is guided to provide a city and state, or ZIP code and physical location description.

The intent of this language was to ensure that those who may be experiencing homelessness on April 1 are not dissuaded from responding on the Internet instrument, but rather understand that they are still eligible respondents even if they are homeless on Census Day.

11. Some examples of Internet response issues are, if you find a typo in a name, you can't go back and change it once you move to demographics; it is not clear what the age constraints are, but a value of 123 was accepted; there might be an explicit NA category. Currently, you can't distinguish item nonresponse from NA (e.g., on the detailed ethnicity/tribe categories); in the online self-complete, it would be helpful to have a link as to why the question is being asked; it would be helpful to have a text box at end in case people have comments/feedback; if someone doesn't have pin, you might give option to say that they had partially completed previously.

CENSUS BUREAU RESPONSE: The Census Bureau appreciates CSACs concerns over these issues and is exploring ways to address them.

SECTION II: Recommendations

1. CSAC strongly recommends a redesign of the internet self-response option, relying on a user-centered process.

CENSUS BUREAU RESPONSE: The Internet Self-Response Integrated Project Team continues to work with the Content and Forms Design Integrated Project Team, Language Services Integrated Project Team, the Center for Survey Measurement, and the commercial vendor User Experience Team, focusing on a user-centered design to improve the user experience. Additionally, the Center for Survey Measurement Usability Lab conducts numerous usability tests with users both in the lab and in the field.

2. A Working Group on Census User Experience should be established. The Working Group on Census User Experience could assist with the user-centered design as well as with communications and outreach.

CENSUS BUREAU RESPONSE: The Census Bureau appreciates this suggestion and will explore creating such a working group.

3. CSAC would appreciate receiving frequent updates (such as every three months) about the status and planned activities related to the development of the user interface.

CENSUS BUREAU RESPONSE: The Census Bureau will find a way to update CSAC about the development of the Internet Self-Response instrument in addition to conducting our quarterly Program Management Reviews, which provides such updates.

4. We would like to know about planning and implementation of protection against cyber-attacks and social media disinformation attacks.

CENSUS BUREAU RESPONSE: Cyber threats are either External to the Census Bureau, on the Perimeter of the Census systems, and/or Internal to the Census Systems. The Census Bureau does not have direct control of mitigations for external cyber threats to the Census and will rely on industry and other federal agencies to help provide services that can take actions to resolve those threats. The Census Bureau can detect some of these External threats but cannot take direct actions to resolve them on our own and examples of these are bad actors impersonating the Census via fake websites or social media disinformation attacks. For external threats, we will have to work closely with our federal and our internal operations and communications partners to respond to these incidents. For the threats in the perimeter and internal to the Census Systems, the Census has the ability to take direct actions to prevent and resolve these threats such as Data Breaches, Compromising User Devices and Disrupting the Internet Self Response Website.

SECTION III: Technical Research on Adaptive Design

This is creative, exciting work, with substantial spillover benefits. It emphasizes the importance of human factors in design. However, there are several important issues.

1. There is a need to engage with field representatives to better implement adaptive design, as a part of participatory design. For this to work, the Incentives-->compensation model needs to change from the perspective of the field representative, given government constraints on incentive policies. Verification of quality needs to be part of compensation.

CENSUS BUREAU RESPONSE: The Census Bureau agrees with this recommendation. Although we referenced it only briefly in our CSAC presentation, work on this has already begun. In October of 2016, we established an Adaptive Design Community of Practice (CoP) with the specific goal of engaging field representatives and better aligning incentives with procedures. To do this we asked the six Census Regional Directors to nominate one representative from each Regional Office. This group has helped us to understand issues involved in attempting to test new field procedures in the context of a number of ongoing surveys.

Engagement with the field on these challenges also includes collaboration between the Center for Adaptive Design and the Field Directorate's Office of Survey and Census Analytics (OSCA). OSCA provides research, analysis, and recommendations on Field Representative (FR) performance metrics and is, therefore, in a key position to help

improve the alignment of incentives with field procedures. OSCA and CAD meet bi-weekly to ensure coordination of research and field efforts in the area of adaptive design.

We believe the most efficacious way to ensure faithful implementation of adaptive design interventions is to make changes in case assignment systems for the demographic surveys. Rather than prioritizing cases and relying on FRs to follow priorities, we should be able to provide only those cases that should be worked to FRs for a given period (day, week, etc.). That way, there is no FR discretion in whether to follow priorities or not. This sort of system has been developed for the 2020 Non-response Follow-up (NRFU) operation.

2. There are concerns about appropriateness for certain surveys. How do you determine which types of surveys would benefit the most?

CENSUS BUREAU RESPONSE: In order to determine the appropriateness for adaptive design approaches, the Center for Adaptive Design works through the following six questions with Survey Sponsors at the beginning of discussions. These questions help teams plan the right approach to data collection and processing improvements before attempting to put an adaptive treatment in place.

- 1) What are the survey's pain points?

The answer to this question often results in a relatively short list of problems that can be prioritized. We then discuss whether the ability to intervene during data collection (to change collection mode or adjust case priority, for example) addresses pain points. In this way, we may find that adaptive design is not the most appropriate approach.

- 2) What is most important?

This question helps provide the information we need to construct the best adaptive approach. We often restate the question this way: "On what aspects of this survey do we want to optimize?" The answers could be cost, timeliness, response rate, total survey error, etc., but we find that having a clear understanding of the survey's key estimates is critical.

- 3) What trade-offs can be made?

This question is the mirror image of question two. In the context of a particular survey, is reducing nonresponse bias more important than boosting response rate? Does increasing representativeness win out over reducing variance? In short, can the sponsor prioritize survey goals? Adaptive survey design is about enabling trade-off decisions. The key question inevitably becomes quality vs. cost.

- 4) What data are available?

The Center for Adaptive Design has found that a persistent obstacle to implementing an adaptive approach is the difficulty in establishing dependable access to data. Even when there is sufficient knowledge about what data exist and where they are located, processes to control data ownership, security and confidentiality can severely slow down new and unfamiliar applications. Knowing what data are available and accessible is critical for the

purposes of adaptive design. Further, it is important to understand the time commitment involved in working through data issues.

5) What is possible?

Adaptive Design is in its infancy and much remains to be proven, both in the field and through scientific evaluation. This question deals primarily with the time available to make effective changes within the context of a survey in production. In other words, what can adaptive design do to help the situation sponsors find themselves in right now? This question also deals with things like length of data collection period. Surveys with very short data collection periods, for example, may not provide enough time for adaptive treatments to provide benefits.

6) What will data users understand and accept?

The challenges survey managers are facing make the effectiveness of adaptive design techniques an increasingly vital discussion in the statistical community. First, we must provide the proof that adaptive design delivers quality, robust results comparable to traditional methods. Second, we must communicate adaptive methods and results in a way that leads to data user acceptance.

The Census Bureau can only do this through continued and expanding scientific experimentation combined with a sustained communication effort. We are always looking for new places and situations to test techniques that will save time and money or improve data quality, and we welcome continued queries and ideas from the CSAC.

3. Given that we have to weight data, what is the preferred balance between adaptive design and weighting? The Census should look retrospectively at the biases created when using adaptive design. By changing the contact approach, you change the characteristics of the population that responds. Although it is a good idea to look at survey data as they come in, adaptive design cannot guarantee cost savings. The examples where the R statistic was improved seemed to result from less overrepresentation of some groups, rather than better representation of underrepresented groups. What is gained in this as opposed to weighting?

CENSUS BUREAU RESPONSE: The Census Bureau accepts the call for additional scientific analysis embodied in this recommendation.

In the 2013 NSCG proof of concept test, we achieved greater sample balance at the cost of a lower response rate and we were more successful in reducing the overrepresentation of groups than reducing underrepresentation of groups. However, in the 2015 NSCG, we were able to increase representativeness without a negative effect on response rates. We had improvements in both under and over representative groups. In both studies, we controlled cost.

The issue of whether to attempt to correct sample representativeness during data collection or afterward is an active area of research in adaptive design. While post survey adjustment is standard practice, it is also not without downsides. It adds time to the production process. Weights developed to correct sample imbalances can add significant variance to estimates. Research by Barry Schouten and others has shown

both empirically and theoretically that even after nonresponse adjustments increased sample balance can reduce nonresponse bias. Additionally, Särndal and Lundquist have shown that when using calibration for post data collection nonresponse adjustments, increased sample balance reduces nonresponse bias if the auxiliary variables that are balanced on are used in the calibration adjustment and the auxiliary variables are correlated with outcome variables.

4. When both representativeness and likelihood to respond are considered, those members of underrepresented groups with a high likelihood to respond will be pursued. Those who are likely and less likely to respond within a group are almost certainly different. This approach could lead to coming closer to targets, such as by race and ethnicity, but could result in less representative samples within each group.

CENSUS BUREAU RESPONSE: In general, the adaptive design perspective takes as a starting point the recognition that time and budgetary constraints make it impossible to interview every sampled case. Rather than seeking simply to maximize the number of cases interviewed, adaptive design attempts to direct effort in line with what is achievable during the field period and the representativeness of the sample. Attempting to interview cases that will improve sample balance and are more likely to respond than others like them is the operational balancing of costs and data quality. Cases that are never interviewed may still be different from those that are captured, but the bias overall is reasonably expected to be less than the bias resulting from a collection procedure that ignores the achieved sample composition.

5. There should be clear documentation for users of the data product on the methodology, including data snapshots/methodology. The Census Bureau should study methodology for storing data/metadata in order to share with users.

CENSUS BUREAU RESPONSE: The Census Bureau accepts this recommendation. As a first step the Center for Adaptive Design has begun working to create general instructions and a template for adaptive design experiments and production data collections to be used for OMB clearance packages. We will be sharing this with OMB in the fall. This will include data snapshots/methodology and recommendations for data/metadata use and storage. This work will make it easier for federal programs to gain OMB approval for adaptive design experiments and production data collections. It will also make it easier for OMB to evaluate these methods. After this first step is complete and approved by OMB, we will extend it to provide appropriate documentation for our data users.

6. Given the cost of SAS, the adaptive design group should consider moving toward open source alternatives. Open sourcing the analysis code could allow quicker development and easier identification of bugs.

CENSUS BUREAU RESPONSE: The Census Bureau accepts this recommendation. The Center for Adaptive Design built the Concurrent Analysis and Estimation System (CAES) as a loosely coupled, component-based, multi-layered platform designed with the architecturally significant requirements of horizontal scalability (i.e., it can easily add or remove hardware to increase or decrease processing power and resources to

match the Bureau's current needs), analytical flexibility (i.e., it can support models and analyses in several programming languages and easily adapt to new analytical requirements), and solution sustainability (i.e., it can easily adapt to emerging technologies and functionalities to meet future requirements).

SAS is only one of the analytical tools that CAES can support. At this time, given the large number of systems and employees who use SAS at the Census Bureau, we feel that support of SAS is critical for the adoption of CAES. However, CAES currently runs open-source analytical software (R, Python, Spark, etc.) and is based on an open source big-data platform (Hadoop, Hortonworks). Further, a new service we are developing in the CAES staff is translation of legacy analytical models from SAS to open source software languages.

SECTION IV: Recommendations

1. Do a cost study of continuing to use closed source products as compared to using open source products.

CENSUS BUREAU RESPONSE: The Census Bureau accepts this recommendation. Open source software can provide cost advantages. We are already working to move in that direction, and will continue to analyze the costs and benefits involved. Transition and support costs associated with open source packages are material, as is our ability to attract and retain new talent. We recognize the fact that most new graduates entering the workforce in our areas of discipline are proficient in open source software tools, and are far less proficient in closed-source languages like SAS.

2. The code used in whatever methodology the Census adaptive design group develops for analysis should also be open source so the community can understand what is done.

CENSUS BUREAU RESPONSE: The Census Bureau accepts this recommendation. We are currently working on providing open source code to allow the survey community to run what we are calling cross survey R-indicators. One of the drawbacks of using R-indicators is that particular surveys can only compare R-indicators if the variables in the model that is used are common between the two surveys. Cross survey R-indicators are an attempt to provide a standard list of variables from the Census Planning Database as well as code to run R-indicators that can be compared across surveys. Surveys will only need to geocode cases to be able to run an R-indicator that can be compared across surveys. This work was presented at AAPOR this spring and we hope to have documentation and code to share in the near future. We are planning to provide additional open-source tools in the future, both for collection and analysis purposes.

SECTION V: Algorithms for Including Administrative Data to Address NRFU Efforts

There was an interesting presentation on research on administrative data.

1. More recognition of the variances and co-variances of the estimates could be incorporated into the research. For example, the Euclidean distance could become a standardized distance, often called a Wald statistic. Also, the results are usually visualized on a simplex (an equilateral triangle), rather than a right-angled triangle. Finally, the covariates have errors in them, and the attempt to smooth them out is one way to go. The other possibility is to fit an errors-in-variables model. The two approaches might be compared in a small simulation experiment.

CENSUS BUREAU RESPONSE: The Census Bureau appreciates this recommendation. The Census Bureau is always looking for ways that its current approach might be improved. Thanks go to Dr. Noel Cressie for already providing a reference on his suggestion to account for analysis of outcomes when the sum is bounded.

The Census Bureau accepts this recommendation. Ongoing research will address these concerns. The proposed production uses of the Euclidean distance measure have been reviewed by the standing committee of the National Academy of Sciences, which thought that they represented a substantial improvement over decision-making based on earlier prototypes.

2. We would like to reinforce the urgency to collect and use state-specific data sets that will be useful in determining vacant, non-existent, and occupied addresses in NRFU, as well as in determining the specific characteristics of individuals and households. Key examples include utility records, driver's licenses, birth records, death records, state identification cards, and Medicaid files.

CENSUS BUREAU RESPONSE: The Census Bureau would be happy to work with the new Working Group on Administrative Records to discuss sources.

(Ingestion of these data is an ADRM effort through CARRA)

The Census Bureau currently receives Medicaid enrollment data, however, we receive these data at a lag and we do not receive address information. The Centers for Medicare and Medicaid (CMS) are in the process of updating their data collection systems, which will allow the Census Bureau to obtain more timely data and also potentially receive address information. We are in ongoing discussions with CMS to receive more timely Medicaid data with addresses when these data become available.

The Census Bureau is also acquiring other state public assistance administrative data to help cover hard to count populations. We are acquiring Supplemental Nutrition Assistance Program (SNAP), Women, Infants, and Children (WIC), and Temporary Assistance for Needy Families (TANF) data.

3. We suggest the Census focus data collection and use for administrative records research in key states that are both large and where counts are likely to affect representation: e.g., CA, TX, FL, NY, IL, NJ, AZ. Moreover, the statistical algorithm doesn't have to be one-size-fits-all for the nation as a whole; it could and perhaps should vary across states based on the available data in each.

CENSUS BUREAU RESPONSE: Thank you for this recommendation. We will discuss with the new working group on administrative records.

4. How is the cutoff drawn in the administrative records research? Can it potentially be more effective if it varies across states?

CENSUS BUREAU RESPONSE: The Census Bureau neither accepts nor disputes this recommendation. Many factors affect which states can be used for administrative record research including the willingness of the state to enter into the required data use agreement with the Census Bureau. Broadly speaking, data collection and administrative records research is already focused on the suggested key states. The 2015 and 2016 tests covered locations in AZ, CA, and TX. Recent work by ADDC/ADRM researchers (not yet approved for public distribution) examines model estimation (“training”) on different geographies in the ACS and subsequent predictions in the 2015 Census Test and finds little effect on match rates in counts and compositions among addresses that were removed from the NRFU workload.

Nevertheless, should additional state-specific data become available, it would be worthwhile to produce state-specific model estimates. Currently, research in administrative record modeling has been expanded to use 2010 Decennial as well as more recent ACS data in the event state-specific administrative data becomes available. The binding constraint is availability of timely state-specific administrative data.

5. How is the cutoff drawn in the administrative records research? Can it potentially be more effective if it varies across states?

CENSUS BUREAU RESPONSE: The Census Bureau can work with the new Working Group on Administrative Records to provide more information about how the cutoffs have been researched and determined.

The current cutoff used in research represents a 10% workload removal rate. It can be varied depending on the situation in the field and across states. Since the size of the NRFU workload varies across locations, a 10% removal rate requires different distance cut-off values. The research cited in the answer to the question above has also examined how the match rate varies across cutoffs.

6. We urge a data content comparison (e.g., age, sex, race, ethnicity, household relationships, housing) between AR predictions for occupied households and fieldwork at those same households. This exercise would compare actual cases that are estimated using AR versus data collected from actual household respondents at the same address.

CENSUS BUREAU RESPONSE: The Census Bureau has done this analysis using data from the 2010 Census and past census tests. The plan is to do it again for the 2018 End-to-End Census Test. Past results can be shared with the new Working Group on Administrative Records, and the Census Bureau would be interested in hearing the group’s reactions.

SECTION VI: Recommendations

1. The Administrative Records researchers at the Census should focus on state data, as described in #2 above.

CENSUS BUREAU RESPONSE: The Census Bureau can work with the new Working Group as it assesses the possible usage of state data sources.

The Census Bureau neither accepts nor disputes this recommendation. As explained above many factors exogenous to the Bureau affect our ability to use state administrative data. Census Bureau researchers are assessing the quality and coverage of Medicaid, SNAP, WIC, and TANF state data.

2. We strongly recommend a Working Group on Administrative Records to help with these issues, as recommended from the last meeting and as supported by the Census Bureau response.

CENSUS BUREAU RESPONSE: The Census Bureau is setting this up.

SECTION VII: Edit Reduction Research

1. The presentation from the Economic Editing Reduction Team was much appreciated. The report from the edit reduction research and the paper outline suggested next steps for implementing improvements related to edit reduction. The work and research that was done by the group was impressive with the ACES survey. The suggestions/recommendations outlined in the paper are reasonable next steps for implementing progress related to the research.

CENSUS BUREAU RESPONSE: See Section VII: item #3

2. Of these suggestions, changes to the questionnaires to reduce the 'data slides' appears to be a simple change that will result in the greatest reduction of edits required. As stated in the presentation and report, data slide edits are by far the most common edits required.

CENSUS BUREAU RESPONSE: See Section VII: item #3

3. The Census is encouraged to make these same types of proactive changes in the questionnaire(s) to control the input for data fields that tend to have errors. As far as the edit process can be moved to the front of the workflow the better. Front end control of input, utilizing rule based entry, will reduce the editing required further along in the workflow.

CENSUS BUREAU RESPONSE: We appreciate the Committee's support of our work on edit reduction and the support of our next steps. We agree that moving as many edits into the collection instrument as possible is the best way to go. We currently have several "soft" edits in the collection instruments, but we limit the number of "hard" edits. We try to balance editing during data collection and the possibility of provoking a nonresponse if the edits are

perceived as burdensome. We appreciate your suggestion to examine our collection instruments edits, and we would like to look at the paradata to see if we can make any recommendations for the “built-in” edits as well.

SECTION VIII: Big Data Working Group Report

1. This working group gave a report on its activities. A question is whether this working group will continue to exist.

CENSUS BUREAU RESPONSE: The Census Bureau agrees that it would be beneficial to us for the Big Data Working Group to continue. We value their external perspective on methodology and data sources. In particular, with their vast knowledge from diverse organizations, the group, through the CSAC, may be able to help us access some types of administrative records and commercial data that can meet our needs.

SECTION IX: Recommendation

1. We recommend the continued existence of a CSAC Big Data Working Group. It interacts productively with the Census Bureau and provides external perspectives and expertise on Census Big Data activities. It is also helpful in keeping CSAC informed about such activities.

CENSUS BUREAU RESPONSE: See response in Section VIII.1.

SECTION X: CEDCaP Working Group Report

1. This working group gave a report on its activities. There is a question of whether the scope of the working group should be modified. Many aspects of CEDCaP are now in the execution phase, so there are recommendations relating to execution.

CENSUS BUREAU RESPONSE: The CEDCaP/ROCKiT Working Group is closed out and we will establish a new working group, in consultation with Decennial, to address Census needs for the 2020 systems. The 2020 Census staff will present at the CSAC fall meeting.

SECTION XI: Recommendations

The following are things to be done:

1. Validate that the learnings from the Prototype have been properly transferred into execution, including completing simulation and evaluating insights. It is not clear whether this has happened, but it could be extremely important. To facilitate this the people involved in the Prototype should also be involved in the execution.

CENSUS BUREAU RESPONSE: Accept - The Census Bureau agrees that the learnings from ROCKiT prototyping should be utilized for the execution of the 2020 Census. The key staff from the prototyping have been involved in the planning for the 2020 Census to date. Over the past few months especially we have ensured that the refinement of the low level scope for re-engineering field operations are based on the concepts of optimization and adaptive design formed through the previous census tests and the prototyping efforts. Staff from the prototyping effort are very involved in this effort.

2. Manage the scope through following change control procedures.

CENSUS BUREAU RESPONSE: Accept - CEDCAP has implemented and currently follows program change management process that aligns with the Census Bureau Enterprise policies which is integrated with the 2020 Census change management process. Any changes to the established program or project scope are submitted as change requests which are reviewed, impacted, analyzed and then evaluated by the CEDCaP Change Control Board. We believe no additional action is needed for this recommendation.

3. Manage risk, including having a list of risks and contingency plans for each in case the risk becomes real.

CENSUS BUREAU RESPONSE: Accept - CEDCaP had already implemented and currently follows a program risk and issue management process that aligns with the Census Bureau Enterprise policies and that is integrated with the 2020 Census change management process. CEDCaP maintains mitigation and contingency plans for all high program and project risks. All CEDCaP risks are maintained in the enterprise risk management tool. We believe no additional action is needed for this recommendation.

4. Evaluate and measure execution status, including consideration and tracking of earned value.

CENSUS BUREAU RESPONSE: Accept - the CEDCaP program already partners with 2020 to evaluate and measure execution status as part of the overall work to ensure readiness of systems for the 2020 Census. We believe no further action is needed for this recommendation.

5. Embrace change management.

CENSUS BUREAU RESPONSE: Accept - as noted above, the CEDCaP program has implemented a change management process. This process allows the program to review and address change over the life of the program. No additional action is needed for this recommendation.

6. Modification of the scope of the working group in order to continue to be maximally helpful to the Census Bureau would be welcome.

CENSUS BUREAU RESPONSE: The Census Bureau recommends closing out the CEDCaP-ROCKiT Working Group and working with the 2020 Census program to propose the formation of a new working group focused on 2020 Census IT Systems. This proposal would be presented at the Fall 2017 CSAC meeting.

SECTION XII: Update on 2015 National Content Test Report

1. This was an interesting and important presentation. CSAC appreciates the careful analysis of the 2015 National Content Test and the detailed material made available online.

CENSUS BUREAU RESPONSE: The Census Bureau's 2015 National Content Test (NCT) Analysis Team, Census Bureau Director, and Executive Staff appreciate the support and encouragement of the Census Scientific Advisory Committee with this important research.

SECTION XIII: Recommendations

1. Proceed with the combined question, but implement testing of whether ordering matters when using the internet.

CENSUS BUREAU RESPONSE: Regarding the ordering of the response categories, the Census Bureau's 2015 NCT research on race/ethnicity has shown that listing response categories and examples in population size order has yielded reasonable and expected racial/ethnic distributions, comparable to previous census results, including the reporting of smaller population groups. Previous census research demonstrated that alphabetical ordering, which places "American Indian or Alaska Native" as the first category, resulted in erroneous reporting as non-American Indian respondents focused on the "American" wording in the category name. However, the Census Bureau continually assesses census questions and is considering the possibility of studying the effect of category ordering, such as random or alphabetical, in the future.

2. Design the internet version better

CENSUS BUREAU RESPONSE: The Census Bureau continues to develop and refine the Internet Self-Response instrument in preparation for the 2018 End-to-End Census Test and the 2020 Census.

3. Integrate information on why this question is asked directly, rather than just under Help.

CENSUS BUREAU RESPONSE: The Census Bureau appreciates the recommendation to include information about why each question is asked directly on the screen for online respondents. Research and testing has found that the current design of the question is working well and easy to follow for most respondents. In particular, the Census Bureau has conducted extensive usability and cognitive testing of the internet instrument with a diverse variety of participants in order to assess whether the question is understood by respondents.

4. Use 2017 test to evaluate further potential mode effects.

CENSUS BUREAU RESPONSE: The Census Bureau will continue to evaluate each of the census questions, through quantitative and qualitative means, to ensure that there is a full understanding of the response distributions by mode.

5. Continue to assess the political climate related to the MENA category – use 2017 test and follow-up with local groups in areas where the test was conducted. This work should continue up to the point when the final questions have to be given to Congress.

CENSUS BUREAU RESPONSE: The Census Bureau will continue to research the race and ethnicity question optimal design in the 2018 Census Test based on the 2015 National Content Test results. Meanwhile, the Census Bureau will continue its outreach with the Middle Eastern and North African community and all other population groups to continue to assess any stakeholder concerns and needs. The Census Bureau will continue to engage with CSAC and our other advisers and stakeholders as these discussions continue.

6. Post answers to likely FAQs online.

CENSUS BUREAU RESPONSE: The Census Bureau will consider this recommendation for future tests and the 2020 Census.

7. Keep in mind the relevance and value of asking parental place of birth in the ACS.

CENSUS BUREAU RESPONSE: The Census Bureau has been investigating the legislative need to add parental place of birth to the American Community Survey and has reached out to other federal agencies. No other federal agency has identified a legislative need for that question, and at this time, the Census Bureau does not have any plans to add it.

8. The data linkages across census and related to the analysis of the change in the race/ethnicity question should be available in some form to the research community – this would benefit the census.

CENSUS BUREAU RESPONSE: The Census Bureau appreciates your suggestion to release the linked data on race and ethnicity to the research community. The Census Bureau will investigate the feasibility of providing access to this data through the Research Data Centers.

9. Continue work on bridging with past censuses and with vital statistics data.

CENSUS BUREAU RESPONSE: The Census Bureau bridges race data based on 1997 OMB Standards with “some other race” and vital statistics data without “some other race.” If OMB decides to allow the use of the combined question, the Census Bureau will perform the research, evaluation, and processing in order to provide a high-quality conversion system between the two classification schemes. Initial plans include developing bridging factors from 2020 Census data; linking records with missing race detail to previous census or American Community Survey (ACS) data; developing socio-economic modeling techniques from ACS data; and using the Demographic Characteristics File method used to assign race and Hispanic origin to all new entrants into the domestic migration universe who could not be found in previous data (e.g., children or new immigrants since 2010). Combinations of

these techniques would not only allow conversion (and therefore compatibility among agency reporting, historical series, and research data), but also it would allow an opportunity to update the conversion factors over time as race reporting evolves.

SECTION XV: Recommendation

1. We recommend that there be a presentation and discussion of privacy issues regarding published data, as discussed in the Census Bureau response to the CSAC recommendations from the Fall 2016 meeting in section 5.3.

CENSUS BUREAU RESPONSE: The Census Bureau accepts this recommendation. Recommendation 5.3 from Fall 2016 is an ongoing area of concern and research. As noted at that time, some statistical methods do work when the Census Bureau is not the custodian of the confidential data. We will schedule an update on our privacy-preserving data publication research for the Fall 2017 CSAC.

SECTION XVI: Availability of Material for CSAC Meetings

1. CSAC very much appreciates improvement in the organization of CSAC meetings. We appreciate the much more timely provision of PowerPoints and background papers and the greater amount of time for discussion and commentary. When it is not possible for background materials to be available before the CSAC meeting, it would be desirable for them to be available at the meeting.

CENSUS BUREAU RESPONSE: The Census Bureau appreciates CSAC's observation regarding improvement in the organization of CSAC meetings and the much more timely provision of PowerPoints and background papers, along with the greater amount of time for discussion and commentary. The Census Bureau will continue to strive to make the meetings and CSAC's involvement as efficient, effective and productive as possible.