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

July 29, 2024

To: Jay Breidt
Chair, Census Scientific Advisory Committee

From: Robert L. Santos
Director, U.S. Census Bureau

Subject: Recommendations to the Census Bureau from the Census Scientific Advisory Committee Spring 2024 Meeting

The U.S. Census Bureau thanks the Census Scientific Advisory Committee for its recommendations. We are responding to the committee recommendations submitted during its 2024 Spring Meeting on March 14-15, 2024.

Your feedback is welcomed 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: Robert Santos

Director, U.S. Census Bureau

From: Jay Breidt

Chair, Census Scientific Advisory Committee (CSAC)

March 15, 2024

Subject: Recommendations and Comments to the US Census Bureau from the Census Scientific Advisory Committee Spring 2024 Meeting

Introduction

The Census Scientific Advisory Committee (CSAC) thanks the members of the US Census Bureau for all their work in preparing the Spring 2024 CSAC meeting, especially the advisory staff who assured that the meeting ran smoothly despite the challenges of the hybrid format, and the Bureau's research teams who shared their ongoing work with the committee. The list of topics was varied, yet there were many interesting connections across agenda items. The Bureau's presentations were carefully prepared and delivered, resulting in lively discussion that helped shape the following observations and recommendations.

Several of the topics discussed in this meeting will be of ongoing interest to CSAC at future meetings and would be usefully followed up with one-way briefings to provide further information to CSAC. CSAC was pleased to see many new presenters and encourages the Bureau to set their expectations (in terms of the public requirements of CSAC's deliberation) in advance of the advisory meeting. We encourage the Bureau to use the mechanism of one-way briefings whenever such meetings might provide useful background or clarification on technical content for CSAC members, for discussion at future CSAC meetings (and to make sure Census presenters know that this may be possible in advance of a meeting). This allows CSAC to learn enough about new developments to make productive and targeted recommendations. From the Spring 2024 meeting, CSAC requests one-way briefings in the future on:

- Implementation of local differential privacy and the accompanying R package for the IRS demonstration product example.
- Outline of the technical details of the US-gridded data products.

Further, CSAC is interested in future presentations at CSAC meetings:

- The progress of pilot programs and tested approaches to improve the identification of hidden housing units, American Indian/Alaska Native (AI/AN) areas, and Puerto Rico.
- Updates on the new procedures for race and ethnicity in the Census and ACS,
- Updates on the disclosure avoidance systems for the American Community Survey (ACS) and the 2030 Census.
- A NEWS project presentation should focus on how under-counted population issues will be addressed in NEWS. The comments from the Coalition on Human Needs are worth noting here. CSAC notes that the "National Experimental WellBeing Statistics" or NEWS data focuses only on income and poverty, not well-being as widely conceptualized. Also, NEWS would be a good venue for testing brief and broad measures of well-being, beyond life satisfaction.
- Subnational long-term population projections undertaken within the International Programs Division.
  Updates on alternative approaches to data collection and address frame building (e.g. crowdsourcing imagery) and corresponding quality evaluations.

Statistical Product First Production Cycle

Statistical Product First (SPF) outlines a user-oriented framework by which the Bureau engages in projects in response to needs articulated by the data user community. This user orientation is guided by a set of eight guiding principles that articulate **ways of working** (collaboration, communication), **technical requirements** (privacy, scalability), and **design principles** (customizability, inclusivity).

Most census data users see the Bureau as “the expert” and are not likely to challenge or engage with the Bureau, thus working on behalf of “users” requires paying attention to this large and potentially silent group. Furthermore, the most vocal data users often make an outsized claim on attention and resources. It is important that the user oriented Statistical Product First framework maintains a balance between the users who actively seek support from the Bureau and the large community of users who are unlikely to ask for help.

1. **CSAC recommends that the Bureau proactively monitor the use of its data products (down to the level of individual variables) to better understand the needs of communities who do not engage directly with the Bureau. AI-supported search procedures are one approach that might be used to identify this larger group.**

   Census Bureau Response: The Census Bureau accepts this recommendation. The Bureau is working to address the ability to monitor third-party data product usage Bureau-wide.

One risk of a user-oriented approach is that it may privilege small incremental changes to data products, as opposed to more foundational investments that might have a large impact. This notion is captured by the phrase, "making faster horses" which is (allegedly) what Henry Ford said his customers would have requested if he asked people what they wanted.

2. **CSAC recommends that the Bureau maintain a list of high priority, high impact projects that are informed from and shaped by data user engagements.**

   Census Bureau Response: The Census Bureau accepts this recommendation. We are currently establishing a Community of Practice (CoP) for Statistical Product First (SPF) Engagement. The CoP for SPF Engagement aims to both uncover as well as prioritize the purpose and use needs elicited from data users across the user spectrum. This new CoP will allow us to better coordinate internally, leveraging information gathered from the different Census programs that already work in the engagement space. As we gather purpose and use needs collected through user engagements, we'll be able to uncover potential projects that would be of high priority and high impact.

The Bureau's mission is to be "the nation's leading provider of quality data about its people and economy" but the guiding principles of SPF do not include data quality. The Federal Committee on Statistical Methodology (FCSM) framework for data quality (https://nces.ed.gov/fcsm/pdf/FCSM.20.04_A_Framework_for_Data_Quality.pdf) outlines three broad domains (Utility, Objectivity, Integrity), each made up of multiple dimensions. The SPF guiding principles align well with the FCSM Utility domain (Relevance, Timeliness, Punctuality, Granularity, and Accessibility), and touch upon dimensions within other domains (Security and Confidentiality). Some other FCSM dimensions might be usefully called out in the SPF guiding principles. The dimensions of Accuracy and Reliability, Scientific Integrity, and Objectivity are standard for the Bureau, but might be explicitly stated in the SPF guiding principles. The FCSM Coherence dimension ("to maintain common definitions, classification, and methodological processes, to align with external statistical standards, and to maintain consistency and comparability with other relevant data") will require explicit consideration in the dynamic SPF workflow.

3. **CSAC recommends that the Bureau compare the SPF guiding principles to the FCSM framework for data quality and consider expanding the guiding principles so that the FCSM dimensions are fully covered.**

   Census Bureau Response: The Census Bureau accepts this recommendation. We will crosswalk the FCSM framework for data quality against our Statistical Product First guiding principles and make adjustments, as appropriate and consistent with OMB Statistical Policy Directive No. 1, as codified by the Foundations for Evidence Based Policymaking Act.

The SPF framework includes a segmentation of users and uses of Bureau data. The Bureau's segmentation of data users and data uses is a critical asset for the SPF framework. For

example, a complete segmentation is necessary to evaluate fairness and equity in Bureau activities.

4. **CSAC recommends that the Bureau formally evaluate SPF's user and data usage segmentation, given its importance for effective outreach to data user communities.**

   Census Bureau Response: The Census Bureau accepts this recommendation. The first part of this segmentation was to identify and categorize the 80+ different known Census user groups (informed through internal collaboration across 60+ Census colleagues who work in the engagement space.) The second part of this segmentation is to map the different user communities to the different purpose and use needs identified through engagement. We are still in the early stages of this user segmentation research.

It is of further importance to ensure that when user communities are grouped as a segment, these segments are appropriate, as exemplified in the AI/AN Demonstration Project.

5. **CSAC recommends that Native Hawaiian and Pacific Islanders be treated separately from AI/AN in the SPF demonstration product, because these groups are treated differently under federal law.**

   Census Bureau Response: The Census Bureau accepts this recommendation. The Census Bureau already distinguishes these groups and has different programs in place for them. We have a tribal affairs program within the Office of Congressional and Intergovernmental Affairs, dedicated to engaging with American Indian and Alaska Native communities (e.g., through Tribal Listening Sessions). We also have tribal relations specialists within the Field Directorate who coordinate with the AIAN communities across the country. However, there will be times when new statistical products or dissemination modalities informed by user engagement with one group may also be helpful for other communities. We aim to engage with other communities when appropriate, as we evaluate the utility and scalability of new products.

Having an appropriate segmentation would help the Bureau evaluate outreach activities. Not all tribal nations are able to attend the conferences, and many tribal nations are not members of the organizations, cited by the Bureau in the SPF material or prepare for meetings with the Bureau. As a result, the data interests of a few tribal nations drive the engagement between the Bureau and the AI/AN community. Meetings held with the primary purpose of identifying statistical needs and conducted in lay language may facilitate increased engagement.

6. **CSAC recommends the Bureau partner and engage with tribal nations beyond conferences (where tribal leaders and policy experts may encounter urgent competing policy agendas with other agencies and departments, such as emergency management, economic development, and Missing and Murdered**

Indigenous Women), and consider more back and forth collaborations rather than presentations and listening sessions with limited tribal leader conversation.

Census Bureau Response: The Census Bureau accepts this recommendation. Furthermore, the Census Bureau is already doing this. The Census Bureau has a robust tribal relations program that goes beyond conference participation and presentations. These tribal conferences often serve as a starting point. The conference presentations lead to side conversations about needed information and then subsequent meetings to continue the conversation. Additionally, the tribal relations specialists and coordinators go beyond conferences when scheduling tribal engagements. As a recent example, Census attended the Bureau of Indian Affairs Tribal Providers Conference in Alaska and presented the Statistical Product First initiative. The conference was only part of the engagement. Census used this opportunity to meet with multiple Alaska Native community organizations over multiple days outside of the actual conference. It was at these meetings that we received the most input about user needs.

The prior recommendations about validating segments and ensuring that outreach activities are sensitive to the needs of user groups are important to formally monitor.

7. **CSAC recommends that the SPF initiative track and publicly report the communities that are served by data concierge services as well as those that are not.**

Census Bureau Response: The Census Bureau accepts this recommendation. There is already an effort underway to capture the wide range of data concierge activities throughout the Bureau. We are cataloging these services and identifying how each service aligns with the user segments and personas served in recognition that the data concierge needs vary within each user segment. This endeavor will help identify gaps as well as opportunities for the possible expansion of data concierge services.

CSAC commends the connection between SPF and the Open Census initiative (https://www.census.gov/about/policies/open-gov.html).

8. **CSAC recommends the SPF initiative provide preferred citations for all statistical and data products as well as for open-source code products.**

Census Bureau Response: The Census Bureau accepts this recommendation. The recently introduced Open Census initiative aims to incorporate industry best practices into the Census Bureau's research and data ecosystems. Open Census will provide Census researchers, developers, and statisticians with repositories to develop, publish, and share their work conveniently and intuitively. It will also provide standards and best practices for cataloging and citing published products, including metadata, Persistent Identifiers, and Digital Object Identifiers. Of course, all of this will and must take place in accordance with the Census Bureau's 13 U.S.C. authorities, including 13 U.S.C.'s confidentiality requirements and requirements for Special Sworn Status, and any

required disclosure review prior to publication These measures should increase peer review, collaboration, citations, and the reuse of research performed by Census Bureau employees, as well as research resulting from the Census Bureau's joint statistical projects and cooperative agreements with outside entities and researchers.

CSAC commends the SPF initiative for its deliberations regarding data governance in parallel to the development of managing statistical products. This discussion is important and should be informed by the development, demonstration, and implementation of SPF principles and activities throughout Bureau activities.

9. **CSAC recommends the Bureau document the data governance principles associated with SPF, such as tiered access.**

   Census Bureau Response: The Census Bureau accepts this recommendation. We must develop statistical products to meet specific purpose and use needs in a way that is legal, ethical, transparent, and equitable. It will be essential that well-documented data governance principles and processes accompany the SPF workflow to ensure we fulfill these obligations. Additionally, new governance structures will be needed to accompany any new tier of access developed by the Census Bureau, and as we explore and develop these, they too will be aligned with these principles and will be well documented.

CSAC commends the Bureau's efforts in collaboration with the Internal Revenue Service (IRS) to compute aggregate statistics while respecting the confidentiality requirements for both agencies. The use of local differential privacy to share demographic data can help get reasonably accurate aggregates. This collaboration can serve as a model for sharing data to create cross-agency aggregates. There are several details that must be carefully worked out and clarified, e.g., by clearly spelling out what information is being protected and what information is not being protected in the data sharing. Selection of the privacy parameters in such an application is a challenging problem, as the legal framework of Title 13 does not immediately translate to an acceptable privacy loss budget. The choices made by the Bureau, and the process used to arrive at the privacy loss budget can form a valuable case study that may be useful for other statistical agencies, industry, and researchers as translating laws and/or regulations into specific privacy loss parameters is an open question.

10. **CSAC recommends that the Bureau provide a detailed explanation of their process for determining the privacy loss budget for the IRS use case.**

    Census Bureau Response: The Census Bureau accepts this recommendation. Local differential privacy, like other formal privacy frameworks, allows for the theoretical derivation of outcomes related to the risk of identifying attributes of respondents protected by the framework. This facilitates sharing such results openly, without further risks to respondent confidentiality. The Census Bureau will present such results and results related to accuracy performed on publicly available data, in one or more presentations.

Local differential privacy can often lead to larger utility loss compared to the central or data curator model when a single data curator has the full dataset. Cryptographic techniques such as Secure Multiparty Computation and homomorphic encryption can allow the two agencies to collaboratively compute noisy aggregates that are as accurate as the central setting, while ensuring that neither agency learns anything more than these noisy aggregates. These cryptographic techniques can thus lead to significantly more accurate statistics than would be permitted by a local differential privacy method, for the same privacy loss budget. These techniques have come a long way in the last several years, and have been deployed at scale. Some recent examples in this so-called cross-silo setting include: https://dl.acm.org/doi/10.1145/3209811.3212701 and various examples at mpc.cs.berkeley.edu.

11. **CSAC recommends that the Bureau evaluate cryptographic approaches to improve the accuracy of the statistics computed in the IRS application and other cross-silo settings.**

    Census Bureau Response: The Census Bureau accepts this recommendation. Interacting with data via such cryptographically secure systems could form another access tier to confidential data under various titles. The Census Bureau will evaluate the feasibility and legality of placing data protected under Title 13 into such a system and will consider options for producing outputs from such a system that adhere to the requirements of the same title.

As the Bureau increasingly publishes statistics that combine several sources of information beyond a survey (e.g., administrative data, imputations from a model, etc.), data users may find provenance information for data fields valuable in understanding the biases and the error bars associated with each data field. It would also enable survey respondents to more clearly see the various uses of their responses. Efforts along similar lines in other communities include *Datasheets for Datasets* and *Model Cards.* While the Bureau often publishes detailed documentation, a short summary with the most salient points accompanying each published table may be more appropriate.

12. **CSAC recommends the Bureau implement data governance principles, such as accessibility and transparency, by adding a short summary of data provenance and processing methods to each Bureau data product. For example, outlining the data sources for each variable, time reference, and the frequency of use of administrative and modeled records and responses.**

    Census Bureau Response: The Census Bureau accepts this recommendation and we will look for ways to include more detailed and user-friendly data provenance approaches to the SPF workflow and products developed as part of this workflow.

Address Frame Maintenance of the Future

CSAC commends the Bureau for advancing its research in address frame updates in preparation for the 2030 Census, and the use of in-office methods for much of the United States while maintaining additional approaches for unique geographies and populations, such as young children and low-income families and individuals.

Census Bureau Response: The Census Bureau appreciates all the recommendations and feedback that CSAC provided in response to the Address Maintenance of the Future presentation. As a general point of clarification, while we appreciate the CSAC's interest in reaching hard to count populations such as young children and low-income families and individuals, it is important to note that efforts at building a complete and accurate address frame do not specifically target populations based on demographic characteristics. While true that there may be some correlation between groups that are in the hard to count population and addresses that are hard to locate (e.g., American Indian Areas), the Master Address File (MAF) is person agnostic. Our primary directive in building and maintaining the MAF is to find everywhere someone could or does live, so that outreach and enumeration efforts can attempt to ensure that everyone in these housing units is included in the census count.

13. **CSAC recommends that the Bureau continue targeted field activities for unique and hard-to-count addresses, including but not limited to hidden housing, Tribal Lands, Alaska, and Puerto Rico. CSAC opposes completely eliminating field activities.**

    Census Bureau Response: The Census Bureau partially accepts this recommendation. As mentioned during the CSAC presentation, 97.3 percent of the addresses in the MAF have been categorized as gold plated, or stable. The Census Bureau's metadata shows that these addresses have received multiple validations over time, from sources such as the USPS Delivery Sequence File, tribal, state, and local governments, and decennial census enumeration. Unless the Census Bureau receives a strong signal indicating these addresses have changed (e.g., due to a natural disaster or demolition), it is not necessary to dedicate resources to continually review these addresses, as we have in the past.

    Instead, we will focus our efforts in areas with unique and hard-to-count addresses, including but not limited to hidden housing, Tribal Lands, Alaska, and Puerto Rico. The Census Bureau will continue to collaborate with stakeholders from these areas to identify and implement mutually beneficial approaches to capture addresses for these unique areas. When appropriate and where necessary, we may also use highly skilled professionals to conduct targeted field work to supplement our in-office methods prior to the 2030 Census.

    Address listing is complicated and requires specialized training to correctly implement. During the 2020 Census, field staff deleted 1.2 million existing MAF addresses that were

ultimately enumerated. Additional research in some areas, for example Los Angeles and Detroit, showed addresses deleted by field staff were twice as likely to be enumerated than addresses added by field staff. Lastly, during the 2030 Census, in-field enumeration operations conducted in these areas will include specialized training to support the capture of additional addresses found during enumeration.

Based on background information provided, the Bureau interpreted the trends in the Local Update of Census Addresses (LUCA) and Census Count Question Resolution (CQR) as demonstrating improvements in the address identification and resulting census counts.

Census Bureau Response: Just to clarify, as mentioned during the CSAC presentation, the Master Address File/Topologically Integrated Geographic Encoding and Referencing (MAF/TIGER) System is more accurate and complete than ever before, having been updated and validated by the USPS Delivery Sequence File, tribal, state, and local government partner files, and 2020 Census enumeration. A high-quality MAF/TIGER System lowers the number of adds generated from LUCA submissions and lowers the need for governments to challenge their boundaries and housing unit counts during CQR.

14. **CSAC recommends that the Bureau reconsider the interpretation of LUCA and CQR declines and perform additional analyses on the LUCA and CQR to assess whether the patterns of decline are from inaccuracy in the addresses and the count or from limited resources (e.g., time, people/technical expertise, software, funding) to update addresses for submission or count challenges.**

    Census Bureau Response: The Census Bureau partially accepts this recommendation. In responding to this recommendation, the Census Bureau would like to clarify that the analysis in the Address Maintenance of the Future paper related to LUCA refers the declining numbers of addresses added to the MAF, after processing from LUCA submissions, not to declining participation in the LUCA program. In fact, over the last two decades, participation was almost identical with 29.2% participation in 2010, and 29.4% participation in 2020. However, the number of adds generated from LUCA submissions dropped significantly from one decade to the next. This is attributed to the completeness and accuracy of the MAF over time.

    First, during the Geographic Support System Initiative Partnership Program, conducted from 2013-2018, the Census Bureau solicited address files from partners covering every county in the nation and received 134.1 million addresses. After reviewing and processing these addresses, 106.7 million addresses met our minimum address criteria, and 99.51 percent of those matched addresses already in the MAF. These partner-provided addresses were not only already included in the MAF, but were also part of the subsequent review of addresses in the 2020 LUCA Program.

    Next, during the 2020 LUCA Program, the Census Bureau received and processed a total of 22.15 million address records. Of those, 18.59 million addresses, or 83.9 percent, matched to Census Bureau addresses during processing, meaning the Census

Bureau already had those addresses in the MAF. When only considering "add" actions, participants specifically suggested that 9.2 million records (part of the 22.15 million) be added to the MAF/TIGER System. Of these, about 6 million records matched exactly to existing Census Bureau addresses. Of the 3.2 million records that did not match, approximately 1.2 million were enumerated.

Related to reasons for tribal, state, and local governments not participating in the 2020 LUCA Program, the 2020 Local Update of Census Addresses (LUCA) Operational Assessment Report contains additional details about participation by type of government and size of government, but not the specific reasons that a government declined to participate. As the Census Bureau prepares for the 2030 LUCA Program, we will continue to consider how to assess reasons for not participating in the 2020 LUCA Program through outreach with various stakeholder groups, specifically whether participation in the 2020 LUCA Program was dependent on resources, and/or whether satisfaction with the address lists provided was a factor in participating or declining to be in the 2020 LUCA Program.

Related to CQR, the assessment for the 2030 Count Question Resolution program will be available later this year. Until then, the Census Bureau will determine how best to assess whether resources, program design, and/or satisfaction with materials provided was a factor in whether a government submitted a Count Question Resolution challenge after the 2020 Census. However, as mentioned during the CSAC presentation, the Master Address File/Topologically Integrated Geographic Encoding and Referencing (MAF/TIGER) System is more accurate and complete than ever before, having been updated and validated by the USPS Delivery Sequence File, tribal, state, and local government partner files, and 2020 Census enumeration. A high-quality MAF/TIGER System negates the need for governments to challenge their boundaries and housing unit counts during CQR.

15. **CSAC recommends beginning LUCA earlier in the decade and enhancing outreach activities to engage tribes and smaller governments. This includes providing LUCA participants with tools (including but not limited to address standardizers and geocoders that output information in a format readily compatible with Bureau requirements) and expertise provided locally that will enable any new addresses to be added more readily to the Master Address File (MAF).**

Census Bureau Response: The Census Bureau accepts and is already making efforts to execute components of this recommendation.

The Census Bureau will continue to provide many no-cost tools and training to LUCA participants. For the 2030 LUCA Program the Census Bureau will provide:

- Access to the Current Address Count Listing Files and Current Census Address Count Listing Files Viewer to enable partners to determine how many addresses the Census Bureau currently has associated with each census block. Partners can use

these tools to compare the Census Bureau's count of addresses to their own count and to help identify where there may be discrepancies.

- Geographic Update Partnership Software, or "GUPS" software, that will enable participants to interactively review addresses and their associated geography on a map, or in tabular form.
- The Census Geocoder will enable participants to identify the geocode for any address, not only as part of the 2030 LUCA Program, but for any reason throughout the decade.
- And new this decade:
  - The Census Bureau will provide 2030 LUCA Program participants access to an Address Matching Service (AMS) that will provide access to the Census Bureau's address standardizer and matching software. This service will enable LUCA participants to standardize their address lists using the same method as the Census Bureau uses. The AMS will also provide participants with real-time feedback on which addresses in their submissions match to addresses in the MAF - all before the participants finalize their LUCA submission.
  - 2030 LUCA Program participants will be provided with 'areas of interest' or blocks that the Census Bureau believes represent new growth or coverage issues to enable participants to prioritize their review and use their resources more efficiently.
- The 2030 LUCA Program will provide trainings, a help desk, and online resources to support use of these tools.
- The 2030 LUCA Program is planning to start earlier than in past decades and will include a preview phase scheduled to begin in February 2027. The preview phase will allow LUCA participants to review their address lists in a Census Bureau provided software nearly a year earlier than they would have been able to in the 2020 LUCA Program. The 2030 LUCA Program will officially start in September of 2027 which is also earlier than the 2020 LUCA Program. Lastly, 2030 LUCA Program submissions will be accepted until March 31, 2028, providing a 6-month review period instead of the 120-day period allocated in the 2020 LUCA Program.

16. **CSAC recommends that future iterations of the Address Count Listing File include tallies that represent both the decennial census and ACS filters for each block to include the total number of housing units, basic street addresses, and group quarters facilities to assist local governments' understanding of changes to their address counts in support of the LUCA program.**

Census Bureau Response: The Census Bureau accepts this recommendation. The Census Bureau will consider the specifics of how, whether, and when we can implement these tallies into the Address Count Listing Files and associated interactive map viewer to enable all of our data users, not only LUCA participants, to gain a better understanding of the Census Bureau's address counts by block.

17. **CSAC recommends employing the LUCA program to identify hard to count areas, hidden housing units, multi-family buildings, and commercial addresses.**

Census Bureau Response: The Census Bureau has already undertaken some of the activities recommended and is making progress on some of the components of this recommendation. The LUCA Program is an opportunity for the tribal, state, and local governments to provide residential addresses, to the Census Bureau prior to the decennial census. For the 2030 LUCA Program, the Census Bureau plans to emphasize both the definition and need for local addressing authorities to identify all housing units in their LUCA submissions, including hidden housing units and accessory dwellings. Additionally, we encourage the use of the Current Address Count Listing Files and Current Census Address Count Listing Files Viewer to ensure that our partners understand how many addresses the Census Bureau currently has associated with each census block. Partners can use this count to compare it to their own address count by block to help identify whether known hidden housing units or accessory dwelling units are already reflected in this count. The Census Bureau will also continue to educate 2030 LUCA Program participants on the connection between identification of hidden housing units and the hard to count population, as at times these are related as described in this blog by Deborah Stempowski.

The Census Bureau would like to understand the CSAC's definition of multi-family buildings and is open to suggestions for how these would be identified. Presently, the MAF includes multi-unit buildings and contains information about the sub units at an address, for example "unit 1" or "garage apt.," but the MAF does not keep track of the number of families within each individual unit.

LUCA is the only opportunity offered to tribal, state, and local governments to review and comment on the U.S. Census Bureau's residential address list for their jurisdiction prior to the Decennial Census. Since the focus of LUCA is on residential addresses, commercial addresses and/or addresses that are "non-residential" are out of scope for the LUCA Program. However, if the Census Bureau receives an indication that housing units are part of a commercial structure, those housing units will be added to the MAF.

18. **CSAC recommends compiling the number of non-city-style street addresses by geography and comparing those results to MAF tabulations to target locations for in-field address updates where major discrepancies occur.**

Census Bureau Response: The Census Bureau accepts and has already undertaken the activities recommended and is making progress on this recommendation. The Census Bureau currently tallies both the number of city-style and non-city-style addresses by various levels of geography (e.g., county, city, and census block). We are prioritizing our research in areas that include non-city-style addresses, such as in remote and rural areas, tribal lands, and Puerto Rico. We have learned that many of the addresses being categorized as non-city-style addresses is largely because they do not receive mail delivery from the USPS, but do in fact have city-style addresses associated

with them. For other non-city-style addresses, we are assessing alternate techniques using satellite and aerial imagery to assign coordinates of latitude and longitude, or provide directions from major roads and landmarks. In areas where we are unable to populate this type of additional information, we may hire professionally trained geographers to capture in-field address updates as necessary.

19. **CSAC recommends publishing quality metrics and assessments (precision and recall) for identifying housing in different types of environments (e.g., urban/rural or heavily forested areas) that respond to geospatial imagery differently.**

    Census Bureau Response: The Census Bureau accepts and is already making efforts to execute components of this recommendation. The Census Bureau is using regional models within our automated change detection process to account for different types of environments across the nation. The Census Bureau is also using additional data, (e.g., parcel and permit data to identify housing, and land use/land cover datasets and NASA's Disturbance product (DIST)) to validate the change we identify on the landscape. We will continue to identify quality metrics to assess the effectiveness of our models.

20. **CSAC recommends expanding partnerships with tribal nations to better develop methodologies to identify addresses in tribal lands beyond reliance on the four tribal specialists for all 574 federally recognized tribal nations, all the tribal areas, and the state tribes. This is particularly important for building partnerships across these politically diverse nations with different regulations surrounding types of data collection (e.g., imagery of villages) and different forms of housing units (e.g., pueblos).**

    Census Bureau Response: The Census Bureau accepts and is already making efforts to execute components of this recommendation. The Census Bureau is using the Navajo Nation Address Pilot as a model for future partnerships with tribal lands. With that said, we understand that diversity across tribal nations may require variations in our approach. Statistics from the MAF/TIGER System indicate that the number of city-style addresses across American Indian Areas varies greatly. While 63 percent of the American Indian Areas recognized by the Census Bureau have 90 percent or more city-style addresses, 25 percent have fewer than 50 percent city-style addresses, and some very small American Indian Areas have no city-style addresses. The Navajo Nation Address Pilot is designed to build partnerships and understand the grassroots addressing efforts on tribal lands, and for the Census Bureau to provide subject matter expertise on addresses and addressing guidelines. However, with the differences in address types among American Indian Areas, we recognize that a one-size approach will not fit all tribal nations and are focusing on areas with high percentages of non-city-style addresses.

Frames Program Update

CSAC thanks the Bureau for the update on the Frames Program. This exciting, forward-looking program has the potential to bring many benefits, including improved data quality, increased efficiency, reduced burden, and the creation of new data products.

Census Bureau Response: The Census Bureau would like to correct the record regarding inclusion of records for individuals in Puerto Rico within the Demographic Frame. The Demographic Frame does not include Puerto Rico. Census Bureau staff are working to obtain and assess a more robust set of administrative records for Puerto Rico before we include them into the Demographic Frame. We will continue to coordinate with others within the Census Bureau as they continue their acquisition and assessment and will look forward to utilizing these sources and data within the Demographic Frame.

The Census Bureau would also like to correct what was stated regarding the relationship between the Demographic Frame and the 2020 Census Edited File (CEF). The 2020 CEF data have been loaded into the Demographic Frame database. In addition, 2020 CEF data may be used to update and/or enhance inputs to the Demographic Frame. For example, the Census Bureau's Best Race and Ethnicity Admin Records Modified File is used to assign race and ethnicity to people on the Demographic Frame. The Demographic Frame Team also plans to include 2020 CEF attributes as additional columns of information about people in a new series of extracts from the Demographic Frame.

21. **CSAC recommends that the Bureau develop and share a roadmap of its plans to enhance frames and for how the value added by the Frames Program will be measured and evaluated.**

    Census Bureau Response: The Census Bureau accepts this recommendation. While a roadmap for the Frames Program can be developed and shared, we note that measuring and evaluating the value added by the Frames Program to Census Bureau products and programs is more complex and depends largely on decisions made outside the Frames Program regarding use of the frames. The Frames Program is generating metrics related to address matching between frames (specifically, the Geospatial and Business Frames) as well as enhancements to completeness, accuracy, and utility of individual frames that will bring value to the programs and products that utilize the frames. The Frames Program will share metrics as they become available and are approved for dissemination.

Bureau programs and datasets rely heavily on the assignment of Protected Identification Keys (PIKs), but the methodology of the Person Identification Validation System (PVS) has evolved since it was first introduced. In addition, a small fraction of the population continues to be excluded from PIK assignment because members do not appear in the administrative databases on which PIK assignment is based.

22. **CSAC recommends that the Bureau prepare a working paper that describes improvements over the past decade in the PVS system and an updated report on PIK rates based on the ACS and also other data sources**.

    Census Bureau Response: The Census Bureau accepts this recommendation. The Census Bureau is working to increase the number of unique identifiers assigned to person records by exploring alternative record linkage approaches. Census Bureau staff are discussing the acquisition of additional data sources with other agencies. Research carried out by the 2020 Census program demonstrated improvements to coverage using these additional data sources; in particular, see Administrative Record Modeling in the 2020 Census and Administrative Data Used in the 2020 Census. The enhancements to processes for assigning unique person-level identifiers and acquisition of additional data sources will improve our ability to represent all population groups in our Frames and other data sources. The Frames Program will work with relevant staff within the Census Bureau regarding preparation of reports on improvements to processes for assigning unique identifiers to individuals.

As described by Genadek et al. (2021), parents authorize that birth certificate data be transferred from hospitals and birthing centers to the Social Security Administration (SSA), with social security numbers (SSNs) assigned to newborns in 95% of cases. These SSNs can be linked to parents via parental names, which are forwarded by the SSA to the Bureau on a quarterly basis. The Bureau uses this information to create the Census Household Composition Key (CHCK).

23. **CSAC recommends that the Bureau arrange to obtain from the SSA parental addresses from the birth certificates, in addition to parental names, and create the CHCK based on assigning PIKs using both names and addresses. Improving links between children and their parents will improve the Demographic Frame's coverage of young children, a group often undercounted.**

    Census Bureau Response: The Census Bureau appreciates this recommendation and agrees with the CSAC that obtaining SSA parental addresses from birth certificates would help improve coverage of young children. We can attempt to obtain these data from the SSA, but we cannot be certain SSA will agree to share these data. We need to consider any request to SSA for additional data in the context of existing data sharing agreements and SSA restrictions upon the Census Bureau's use of SSA data. In addition, the 2030 Program is actively researching ways to reduce the undercount of young children. This includes researching the feasibility of employing new methodology that uses existing administrative data sources to add young children to households where they have been previously omitted.

Households can be reconstructed by grouping persons who share a Master Address File ID (MAFID). The validity of this approach is yet to be fully investigated.

24. **CSAC recommends that the Bureau assess the strengths and weaknesses of household reconstruction by grouped MAFID through comparisons of the reconstructed households with decennial census data as well as selected national surveys such as the ACS.**

    Census Bureau Response: The Census Bureau accepts this recommendation. Work is planned to develop household and family relationships for people in the Demographic Frame. Those data will be assessed as the work progresses.

The Geospatial, Demographic, Jobs, and Business Frames are useful individually. Linking them provides additional information and insights.

25. **The quality of the links among the Geospatial, Demographic, Jobs, and Business Frames will depend on the degree to which the time periods mesh. For this reason, CSAC recommends that the temporal references of the frames be carefully documented.**

    Census Bureau Response: The Census Bureau accepts this recommendation. This work is underway and will continue. The Frames Program will continue to document temporal references for each frame as well as identify enhancements to temporal references in documentation for each frame. We note that the alignment of reference periods is one of the many things we are considering as we link the frames as well as when we link the frames to other data sources, such as the American Community Survey (ACS). The work with ACS and Non-Employer Statistics by Demographics (NES-D) shows that the closer the alignment of the reference dates, the better the person-level match rate.

26. **CSAC recommends that links between the Geospatial, Demographic, Jobs, and Business Frames be examined, compared, and assessed to improve the information included in individual frames and also to understand inclusion and exclusion of persons from the linked frames.**

    Census Bureau Response: The Census Bureau accepts this recommendation. This work is already underway and will continue.

Individuals do not need an identifiable residential address to exist in the real world, but under the current design of the demographic frame people must have an address (MAFID) to be included in the frame. Individuals may have zero, one, or many addresses at a given point in time.

27. **CSAC recommends removing any requirement that a person must have a MAFID to be included in the Demographic Frame.**

    Census Bureau Response: The Census Bureau accepts this recommendation. We note that we are in the process of incorporating methodology utilizing Individual Tax

Identification Numbers (ITINs) and geographic information in records that do not have a MAFID assigned. There are also Enterprise-wide collaborative efforts underway to improve PIK and MAFID assignment. Frames Program staff participate in and contribute to these efforts.

28. **CSAC recommends that MAFIDs be included in the Demographic Frame as supplementary information associated with a person (PIK), for example as a timestamped list [e.g., MAFID: Date1, MAFID: Date 2, ...].**

    ***Census Bureau Response*:** The Census Bureau accepts this recommendation. The MAFID(s) associated with a person (PIK) already are included in the Demographic Frame and are made available on Demographic Frame extracts. The Demographic Frame Person-Place Model identifies the level of probability for which each MAFID is the primary place of residence on a given date, generally either April 1 or July 1, depending on the preferred reference date. Additional dates and timestamps related to the MAFID(s) associated with a person derived from the administrative record(s) source(s) linked to the PIK-MAFID record are included on the extracts provided to customers.

The value of the Frames Program extends across the Bureau's directorates and is also of great interest to outside users. Questions about access through the Federal Statistical Research Data Centers (FSRDCs) arose at several points during the meeting.

29. **CSAC recommends that the Bureau develop and share a plan for when and how products from and evaluations of the Frames Program will be made available to users, including those participants in FSRDCs.**

    Census Bureau Response: The Census Bureau accepts this recommendation, but seeks clarification from the CSAC regarding its use of the word "products." The Frames contain protected microdata and, as such, products in the form of extracts from the frames cannot be released to public data users. Frames extracts can be made available for use in the FSRDCs. Some extracts—MAF extracts, quarterly Job Frame extracts—already are available in the FSRDCs.

    Census Bureau staff are already evaluating their use of the various Frames and presenting this information publicly. Please refer to the following website to see a full body of reports: Frames Program (census.gov).

Continuous Count Study Project

CSAC commends the Bureau for conducting this important project. A notable advantage of population estimates based on administrative records compared to a census is that the estimates based on administrative records can be repeated annually or even more often, with potentially the same quality as in a census year. The Bureau's presentation on the Continuous

Count Study Project highlights this feature and combined with Frames program enhancements provides great promise.

Administrative records (ADRECs) are not necessarily attributed to residential addresses and available ADREC addresses may not be in the same census tract or higher-level geography in which a person resides. The Continuous Count Study methodology recognizes this limitation by assigning probabilities to individual addresses when there is more than one address associated with a person. If there is only one address, however, the methodology treats that address as accurate.

30. **CSAC recommends that the Bureau:**
    - **Link the Business Frame and group quarters in the Continuous Count Study to modify the probability of those addresses providing a correct geographical assignment of an individual's residential address.**
    - **Conduct research to determine which administrative record sources provide the greatest accuracy and reflect the quality of specific sources when computing the probability of correct person-address assignment.**
    - **Provide information to CSAC on the number of records added to the frames extract by each administrative record source.**

    Census Bureau Response: The Census Bureau thanks the Committee for these research suggestions to improve residential address determination but cannot accept the recommendation. For persons with one address, the Census Bureau would like to clarify that one of the estimation methods (Method A) uses person-place probabilities from the Demographic Frame to account for that address being residence for the person for the given reference date.

    Differences in source accuracy are incorporated in the person-place models. They include the sources as explanatory factors. Table 43 on p. 101 of the Real-Time 2020 Administrative Record Census Simulation report (Real-Time 2020 Administrative Record Census Simulation) shows person-place match rates for each source separately, both unweighted and weighted by the modeled person-place probabilities.

The 2020 Post-Enumeration Survey (PES) state estimates of the household population were significantly different from the 2020 census estimates in 14 states. Understanding the reasons for these discrepancies would be helpful to the development of post-censal population estimates as well as planning for the 2030 Census and its coverage evaluation.

31. **CSAC recommends that the Continuous Count Study be employed to aid in understanding of state differences between the 2020 Decennial and 2020 PES and that the Bureau publish the results and explore the implementation of subsequent iterations of the Continuous Count Study to enhance the 2030 PES results**.

    Census Bureau Response: The Census Bureau accepts this recommendation. The Continuous Count Study (CCS) results for April 1, 2020 (Initial Results from the

Continuous Count Study Report for the Census Scientific Advisory Committee Spring Meeting March 14, 2024) can be another piece of information in addition to the Real-Time 2020 Administrative Record Census Simulation that can be used to possibly understand the differences.

This iteration of the Continuous Count Study provides substantial enhancements to those of the past and is supported by more complete methodological descriptions and more detailed statistics. CSAC commends the Bureau for this progress and appreciates the additional information. CSAC's remaining recommendations on the Continuous Count Study are grouped by theme: communication, quality improvements, evaluations, and enhanced data sources.

32. **On the subject of communication, CSAC recommends that the Bureau provide a report to CSAC on the planned use of the 2020 Decennial Census results in future iterations of this research, when these results will be available, and if they will be included in the 2030 Planning Database.**

Census Bureau Response: The Census Bureau accepts this recommendation. For our 2022 results, the CCS will continue to use race and Hispanic origin responses for a person from the 2020 Census. As we continue to refine and determine any additional usages of 2020 Decennial Census data, we will update the committee. Additionally, we will be posting updates on CCS progress on the CCS website (Continuous Count Study (census.gov)).

33. **Regarding quality improvements, CSAC recommends that the Bureau:**
    - **Explore the production of confidence intervals for specific geographic areas and demographic characteristics.**
    - **Develop quality measures for additional age categories, especially for individuals under the age of 18.**
    - **Explore the creation of household-based model(s) and compare those results to the current person-based models.**
    - **Assess the impact of respondent information as proxy responses for other household members versus information that has been provided directly by an individual in current surveys/censuses and administrative records.**

Census Bureau Response: The Census Bureau accepts this recommendation. We will pursue the four research topics and report on outcomes as the research matures.

34. **With respect to enhancing data sources, CSAC recommends that the Bureau:**
    - **Update the MAF to link non-city-style address information to MAFIDs and enhance the performance of the model by reducing geographic area imputation.**
    - **Determine the impact of adding other surveys to the demographic frame – such as the Current Population Survey and the Survey of Income and Program Participation.**

- **Collect additional data from state or local sources and provide incentives for those entities to participate (either through enhanced access to information or the return of new data products).**

Census Bureau Response: The Census Bureau accepts this recommendation. We will pursue the three research topics and report on outcomes as the research matures.

35. **CSAC recommends the Bureau expand and update its analyses of the adjusted measures for the continuous count, including but not limited to the following, and update the committee on the findings:**
    - **Compare tract-level statistics to estimates produced by state partners in the Federal-State Cooperative for Population Estimates, when such estimates are available.**
    - **Produce estimates for municipalities and compare these to the Bureau's subcounty population estimates.**

Census Bureau Response: The Census Bureau accepts this recommendation. We will pursue the two research topics and report on outcomes as the research mature.

SIPP SEAMLESS: Modernizing the Survey of Income and Program Participation

CSAC commends the Bureau for its efforts to revise the design of the Survey of Income and Program Participation (SIPP) to address issues that have troubled this unique and long-running survey.

The SIPP was launched in the Fall of 1983 following an extensive design study conducted under the Income Survey Development Program (ISDP). Major objectives of the SIPP included collecting data that would enable (1) the estimation of program eligibility based on income, family composition, and other characteristics observed in a given month, consistent with program rules and (2) measurement of short-term dynamics in income, family composition, program participation, and other characteristics. The initial design of SIPP had four rotation groups that conducted interviews in successive months, collecting data on the previous four months. New panels were introduced annually, and a panel length of 2.5 years was common. The core interviews were supplemented with topical modules, which varied by wave and collected data outside of the monthly time frame. A redesign implemented in 1996 increased the panel length to four years but replaced overlapping panels with abutting panels. Given the same resources, this revised design supported a larger panel sample size than the overlapping panel design. A "re-engineering" of SIPP, following a termination that proved temporary, replaced the four-month reference periods with a calendar year "plus" reference period, with data collection occurring in January through June and selected items being captured up to the time of the interview. An Event History Calendar was introduced to assist respondents with

recall over the extended reference period, and respondents in waves 2 and later were given the opportunity to revisit the prior wave's responses for the early months of the prior calendar year.

Since its inception, SIPP has experienced a number of measurement issues. Despite the short reference period, a pronounced "seam bias" has characterized SIPP data from the beginning. Reported transitions–particularly with respect to program participation–have occurred disproportionately between reference periods (with too few transitions within reference periods). And until the 2014 redesign, spell lengths derived from reported transitions were predominantly multiples of four months. Underreporting of program participation persists, albeit generally not as substantial as with other surveys. The quality of income data (initially one of SIPP's strengths) has deteriorated since SIPP's early days. Poverty estimates exhibit within-panel trends distinct from what is observed in cross-sectional surveys–a result that became more evident with the lengthened panels from 1996 forward. Attrition has grown over time as survey response rates in general have declined, jeopardizing the representativeness of both cross-sectional and longitudinal estimates from the SIPP. Overlapping, annual panels were reintroduced in recent years to improve cross-sectional estimates (which can be calculated by combining panels). Wave 1 response rates have plummeted, jeopardizing SIPP data collection altogether.

To address key challenges to SIPP's viability, the Bureau has proposed major revisions to the current design. Described as SEAMLESS SIPP, the proposed design resurrects elements of the original SIPP design, but with modifications. Like the original design, the proposed design will include rotation groups, but there will be six instead of four, and the reference period for each rotation group will be six instead of four months. With this design, data collection for a given panel will be distributed evenly across the year and field staff will be employed continuously. Panels will overlap, with a new panel starting each year. Selected content, presumably of the type once collected in the topical modules, will be collected only in odd-numbered waves.

SEAMLESS SIPP includes some innovations designed to improve response rates, reduce respondent burden (potentially reducing attrition as well), and enhance data quality. The survey will employ a two-phase data collection for each rotation group. Phase 1 will utilize Internet self response as the principal mode of data collection. Phase 2 will involve a subsampling of nonrespondents to phase 1, with contacts being made by field staff, who will collect data using Computer Assisted Personal Interviewing (CAPI). This is essentially the same approach used with considerable success in the ACS. In addition, the Bureau will utilize a combination of administrative records and imputation to replace some of the content. This will reduce respondent burden and, depending on what administrative sources are tapped, may also improve data quality. The Bureau will also make increased use of model-based imputation (which was introduced on a limited basis in the last redesign) in place of hot-deck imputation, which should improve the quality of the imputations and potentially reduce the incidence of anomalous results (such as high-income SNAP recipients) attributable to limitations in the number and form of hot-deck covariates.

CSAC notes some implications of the proposed design. As with the original design, staggered reference periods by rotation group mean that only one calendar month will be included in all rotation groups for a given wave. The common month is December for odd-numbered waves and June for even-numbered waves. Data from three waves will be required to produce estimates from all six rotation groups for any 12 calendar month period (the original design required data from four waves to cover a 12-month period). Seam bias will be eliminated from cross-sectional estimates because the estimates for any given calendar month will reflect the full distribution of recall periods (from one to six months), but seam bias in spell lengths will shift to multiples of six months. Finally, if the annual data collected in odd-numbered waves refer in some cases to a common calendar month (rather than, say, last month), responding to these items will present differential recall demands across the six rotation groups just as the calendar year reference period under the current design presents differential challenges by interview month. The Bureau's presentation provided a rationale for having December as the common month in those waves.

36. **CSAC recommends that the Bureau evaluate the comparative advantages of starting each panel in January versus February. One advantage of starting the panel in February is making January rather than December the common month for all six rotation groups in odd-numbered waves, consistent with most SIPP panels under the original design.**

    Census Bureau Response: The Census Bureau does not accept this recommendation. The December common reference month in the current SIPP provides the anchor for the collection of information about the entire prior year and is central to the inclusion of the suite of annual content that asks summary information about the prior calendar year reference period. Moving to a February collection start implies that we are no longer able to ask information with a reflection over the entire reference year (e.g., total miscellaneous annual income – "as of December, did you receive xxxx during reference year yyyy?"). While a January start means that the reference period from rotation group 1 is not part of the rectangular reference year, a February collection start implies that we are not including information from rotation group 6 in the very last wave that covers the reference year. The concerns of collecting data that is ultimately unused is still a concern. Information collected in the first rotation group is not all lost. We intend for this to be used to create left-censor summary information for the preceding year, to inform feedback passed into later interviews, to inform adaptive design in waves 2+, and to be incorporated into reference-year imputation.

Even if the two-phase sampling approach and other modifications markedly improve unit response rates, differential nonresponse is likely to remain a concern. The Bureau presentation does not mention prospective enhancements to unit nonresponse adjustments. Data sources being developed by the Bureau provide a possible avenue for improving such adjustments. These include Frames data and administrative records.

37. CSAC recommends that the Bureau explore whether the Frames data and administrative records can provide housing unit-level or at least small area-level covariates that can be incorporated into unit nonresponse adjustments.

Census Bureau Response: The Census Bureau accepts this recommendation and we are currently working on its implementation. We are researching the incorporation of administrative, and other, records to inform weighting. Census Bureau staff have been researching this topic, and we are additionally exploring Entropy Balancing Weights developed internal to the Census Bureau. We are evaluating these in the context of both sampling and stratification, as well as for weighting and Bayesian non-response adjustments.

The Bureau has indicated an intention to expand the use of model-based imputations for item nonresponse in the SIPP. CSAC endorses this goal, which has been recommended by other advisory groups in the past (https://nap.nationalacademies.org/catalog/2072/the-future-of-the-survey-of-income-and-program-participation and https://nap.nationalacademies.org/catalog/12715/reengineering-the-survey-of-income-and-program-participation. Among its advantages, model-based imputations make it possible to incorporate external data sources such as administrative records as covariates. The Bureau has used administrative records linked to individual respondents to edit reported Social Security and Supplemental Security Income (SSI) receipts in the SIPP.

38. CSAC recommends that the Bureau incorporate administrative records to the extent possible in its expanded use of model-based imputation in the SIPP.

Census Bureau Response: The Census Bureau accepts this recommendation. The current topic-model imputation already leverages Sequential Regression Multiple Imputation (SRMI) in combination with administrative records and is a specific subset of the suite of model-based imputation that we are referring to. We agree with this expansion of SRMI but are also researching other model-based imputation methods along with an expansion in the scope of administrative records used. This is part of planned work in the SIPP program to explore model-based imputation as a replacement for hot deck imputation.

The Bureau has proposed reordering sections of the SIPP questionnaire, providing as examples the movement of food security and disability questions from later in the instrument to sections that include the receipt of food assistance and disability income, respectively. In the examples presented for illustration, questions on food security would precede questions on SNAP and WIC and other food programs, and questions on disability would precede questions on the receipt of Social Security Disability Income and SSI. Given that question sequencing can affect item nonresponse as well as the responses themselves, it is important that the potential consequences of decisions to reorder sections within the instrument and questions within sections be understood.

39. CSAC recommends that the Bureau's decisions on reordering sections and questions incorporate findings from cognitive testing and field tests.

Census Bureau Response: The Census Bureau accepts this recommendation. Although resources for cognitive testing are currently limited, we are exploring this recommendation and would like to incorporate questions ordering into a cognitive test.

With the current SIPP design, the data for a given calendar year are collected in the first half of the next calendar year. With the proposed new design, the data for a given calendar year will be collected mostly during that calendar year. Thus, if data collection under the new design begins in January 2028, the first full calendar year of data will be for 2028. With the current design, data collection will need to occur in the first six months of 2028 in order to provide data for the 2027 calendar year. Since evaluating the impact of the design change requires production data from the two designs for the same 12 months, it would be necessary to continue the current design for an additional year–that is, to collect data in 2029, the second year of the new design. CSAC understands the Bureau's reluctance to do so. At the same time, CSAC believes that it is critical that the Bureau be able to compare the two designs for the first months of the calendar year, where there is evidence that the data collected under the current design are weakest.

40. CSAC recommends, as an alternative to comparing the two designs for the same calendar year, that the Bureau consider two complementary approaches and report back to CSAC. First, the Bureau could compare 2027 data collected under the current design and 2028 data collected under the new design to a third source for validation (e.g., administrative data for program participation and income). Second, the Bureau could directly compare estimates across the two years for other characteristics that can be presumed (or, ideally, demonstrated from other sources) to have minimal change between 2027 and 2028.

Census Bureau Response: The Census Bureau accepts this recommendation. We are working within the Census Bureau to identify resources to conduct evaluations. We are prioritizing the collection of overlapping, or at least adjacent data years with the current and new design. Administrative data will be used across topics to help identify both absolute differences from the external information source as well as the relative differences between survey data sources and the administrative sources.

CSAC recognizes that external factors (a recession, a pandemic, or simply the environment of a presidential election year) may affect comparability between the two years, potentially limiting the effectiveness of these two approaches to evaluating the new design. Consequently, CSAC does not offer a recommendation one way or the other regarding continuation of the current design for the additional year that would be necessary to obtain 12 months of overlapping data.

The Bureau proposes the use of business register data linked to SIPP respondents to replace survey questions as the source of employer characteristics. Because the business register is

itself a confidential source, the Bureau would produce synthetic estimates of the employer data for inclusion on SIPP public use files. The Bureau has asked if CSAC supports this approach.

CSAC appreciates the goal of reducing survey content and respondent burden by drawing on alternative sources. With regard to employer data specifically, the Bureau confirmed that the data obtained from the business register will match the SIPP with respect to providing information at the firm and/or establishment level. The Bureau has proposed synthesizing the data from the business register, as these data cannot be released unaltered on a public use file. CSAC recognizes that synthesis, while used for other Bureau products including the SIPP, is challenging. In addition, in other contexts, the release of synthetic data is often paired with access to a validation or verification process, whereby users can submit programs created from the synthetic data to be run on the true data. Users then receive a response indicating, in some manner, how closely the estimates from the true data match the estimates from the synthetic data. This process has been implemented with fully synthetic data. It is not clear if such a process would be cost effective for the limited synthesis envisioned here.

CSAC supports the Bureau's investigation of this approach but would need to see evidence that the proposed replacement of survey-collected data with synthetic estimates derived from matched employer data does not result in a significant degradation of the quality of the employer data provided on SIPP public-use files.

41. **CSAC recommends that the Bureau report back to CSAC (possibly in a one-way briefing or CSAC meeting) on its findings regarding the quality of the synthetic estimates that would replace survey responses on the SIPP public-use files.**

    Census Bureau Response: The Census Bureau accepts this recommendation. The potential value of including data from sources like the Census Bureau's Business Register is high, offering the ability to simultaneously reduce respondent burden and improve data quality. As we find opportunities to evaluate this design, both before and after production interviewing starts, we will keep CSAC informed of the results.

The Bureau also asks if there are other areas where CSAC would recommend this type of data integration or blending. While CSAC is not aware of findings suggesting serious quality issues with the employer data collected in the SIPP, program participation is an area where underreporting remains significant. The quality of data on program participation could potentially be improved by replacing survey responses with matched data from program administrative records. CSAC is concerned, however, about the ability of synthesis to replicate complex relationships between eligibility and participation and to adequately capture multiple program participation. CSAC wonders if an alternative form of disclosure limitation such as noise injection could meet the needs here.

42. **CSAC recommends that the Bureau explore the proposed data integration approach to program participation and report back to CSAC on its findings.**

Census Bureau Response: The Census Bureau accepts this recommendation. The interrelationships between variables are a critical component of SIPP data. There are meaningful limitations and protections that must be applied to administrative data sources, and balancing the valuable addition of this information into the data structure and necessary privacy protections is critical. One example in current SIPP of using administrative data to correct and improve estimates is the use of data from the Social Security Administration to correct reporting of SSI when the actual receipt was Old-Age, Survivors, and Disability Insurance (OASDI). We implement a model-based correction that is highly predictive, but not simply substitution of the administrative data into the survey data structure. This is not synthetic data in the traditional sense but is still protective of the administrative data used to correct this program reporting. We will continue to refine and expand the use of administrative data into additional SIPP areas.

Should the findings support this approach, CSAC notes that a slow roll-out accompanied by advance communication of the changes may be necessary to gain user trust.

The Bureau has also asked CSAC to weigh in on SIPP's unique contributions, identifying specific areas of SIPP's value added on which the Bureau should focus its redesign efforts, presumably ensuring that data quality in these areas is maintained or even enhanced.

43. **CSAC recommends that the Bureau focus on health insurance coverage dynamics; program participation rates, based on sufficient information to simulate eligibility; and household and family composition dynamics.**

Census Bureau Response: The Census Bureau accepts this recommendation. These are central to the SIPP vision, and are a focus for further developing intra-year dynamics that are part of SIPP's priority and will be part of the program's competitive advantage and unique place in the suite of household surveys.

The Bureau currently produces from SIPP an annual report and extensive supplemental tables on wealth, including detailed information on asset and debt holdings. The Bureau also produces occasional reports on a variety of other topics. The Bureau has asked CSAC what reports in addition to the wealth report it should publish on a regular basis.

CSAC acknowledges the value of the annual report on wealth–especially the extensive detail provided in the supplemental tables. Expanding the report to include discussion of selected data from the detailed tables would further enhance the value of the report. For example, comparing population subgroups with respect to the distribution of asset and debt types would provide a compelling story of cross-sectional differences in the composition of wealth and on the process of wealth accumulation over time.

44. **CSAC recommends that the Bureau expand the brief, annual report on wealth to include comparative analyses across subpopulations based on the extensive, supplemental tables.**

Census Bureau Response: The Census Bureau accepts this recommendation. As we review content to be retained in SIPP, we will continue to try to maximize the ability to support and include comparative analyses with the revised design and represent them in briefs and reports as time and resources allow.

CSAC notes that economic well-being, while important, does not encompass a broad definition of well-being and that wide and rich measures are generally in short supply among Census data efforts, such as Frames data. Given their high item response rates, including brief measures of adult and child well-being could increase respondent willingness to complete SIPP, and including questions about the family and children could improve reporting of young children. Substantively, including such measures would provide information on the implications of income, poverty, and wealth and support development of reports and briefs.

45. **CSAC recommends that the Bureau expand its SIPP products to include:**
    - **Reports on the topics listed as "value adds" above.**
    - **Expanded reports on topics covered in previous reports, including income sources of older households; benefits received by veterans; number, timing, and duration of marriages; and parental presence among children.**
    - **Reports on new topics including child care arrangements, adult and child well-being, and income dynamics of low-income families.**

Census Bureau Response: The Census Bureau thanks the Committee for this recommendation. As we review content to be retained in SIPP, we will continue to try to maximize the ability to support and include comparative analyses with the revised design and represent them in briefs and reports. The Census Bureau will evaluate the balance of burden and content scope, and it is likely that there will be some topics we will be better able to collect and report on than others.

The Bureau noted that it is in the process of rewriting in python much of the SAS code used in SIPP processing. Access to some of this code could benefit SIPP users.

46. **CSAC recommends that the Bureau release portions of its python code that are likely to benefit SIPP users.**

Census Bureau Response: The Census Bureau partially accepts this recommendation. We can release code that uses the public use data to generate table packages and reports. However, there will be some code that we will likely not be able to release, for example, code that edits the data from the instrument, as it may be disclosive.

Statistical Grids for the U.S. Census Bureau

Presentation on this topic to the CSAC is new, and we thank the Bureau for it. CSAC commends the Bureau for its continued work in global gridded population data products and its innovation in the development of US population and housing grids. CSAC separates its recommendations by global and US data products, noting that these developments address different sets of goals and likely different user groups.

Gridded data products allow for much wider dissemination of data than underlying small-area unit data, and micro-data records, where confidentiality concerns would arise. As such, such gridded data allows an initial form of 'tiered access' falling between current microdata and summaries to existing census geographies. Internationally, while there are many global data products on population grids, the Bureau's International Program fills a gap by producing grids in data-poor countries where gridded population estimation aims to use ancillary data and methods to update and spatially allocate population. In the US, production of population and housing grids is novel, and the Bureau stands to become a leader in this area. CSAC notes that gridded data products can help achieve privacy protection while not compromising analytical capacity of the data products. Because the gridded data products for the US are new, CSAC's recommendations below focus on that but where applicable can be generalized to the global program gridding program as well.

CSAC commends collaborations between the Bureau's Grids and SPF initiatives, and notes that user requests for statistical products are likely to include requests for statistical products for customized geographies.

47. **CSAC recommends the Grids initiative collaborate with the SPF initiative and report back to CSAC on the potential for delivering statistical products for novel, geographically-specific areas beyond existing Bureau geographies (e.g., school districts or areas impacted or likely impacted by hazards, such as coastal communities or populations within watersheds).**

    Census Bureau Response: The Census Bureau agrees with this recommendation. While we have not specifically collaborated with the SPF initiative yet, we will make sure our efforts are coordinated moving forward. We agree on the potential for using grids as a building block for novel geographies and have identified the creation of "user defined geographies" as an important use case for grids. The Grids effort is intended as an important first step towards realizing this goal.

48. **CSAC recommends that the Bureau work closely with the disclosure avoidance team in order to determine data structures that may be needed to inform aspects of gridded output (such as spatial resolution), and the production of an uninhabited grid.**

Census Bureau Response: The Census Bureau agrees with this recommendation. Disclosure concerns are a primary consideration for the Grids effort and are a major topic in all our discussions and planning. We are currently working to schedule an informational discussion with our Disclosure Review Board with a request to hold regular formal meetings moving forward. Informally, disclosure avoidance subject matter experts are members of our Content Team, and several members of this team regularly interface with our disclosure governance bodies.

CSAC was asked to comment on appropriate spatial resolution and appropriate grid system for its gridded data products. While global gridded data products of population tend to produce 1km quadrilateral products (though many now also produce grids at finer spatial resolutions ranging from 100m to 250m), there is no clear rationale for new grids for the US to be produced at that fairly moderate resolution. The US is geographically large, sparsely populated in most locations (such as rural areas), but densely populated in metropolitan statistical areas. Since the individual-level data can be used as the backbone for the production of new gridded data products, the resolution of the output data products should match user needs.

Census Bureau Response: The Census Bureau appreciates this feedback. Our plan is to produce an initial grid that supports the subdivision of cells into smaller nesting cells, without gaps or overlaps (which are potentially disclosive), as well as allow the aggregation into larger cells. As a base unit for a first product, $1km^2$ appears to be a good target, and would allow for this nesting hierarchy to support aggregation/disaggregation. As noted in our presentation to CSAC, while many national statistical offices use smaller grid cells, they generally also produce a base grid of $1km^2$. We do not intend for this to be the only resolution we produce, but it is important to issue an initial product as a proof of concept to solicit additional feedback from our stakeholders.

It is important to note that only basic geographic measures such as land and water area, and housing unit counts, may be possible at smaller grid cell sizes. Survey estimate products from the American Community Survey and other surveys, as well as Economic Census data, may be restricted to larger grid cell sizes. This will be determined based on guidance provided by our statistical methodology team and disclosure avoidance subject matter experts.

49. CSAC recommends that the Bureau produce variable-resolution data products for the US. If using a latitude-longitude grid, these grids should be high-resolution in urban areas, somewhat less fine in areas in the urban periphery and fairly coarse (perhaps coarser than 1 km) where the population is sparse. If the Bureau aims to produce a 1 km grid, as an initial data product, CSAC recommends also producing a grid no coarser than 250m for urban areas.

Census Bureau Response: The Census Bureau partially accepts this recommendation. As noted in responses to earlier items, we intend to pursue novel applications and use cases for our grid products. At this time, our focus is on producing a grid that can be used as base input for user defined areas, variable-resolution data products, and other similar forward leaning applications. We have had discussions with

colleagues of other national statistical offices that have experience with producing these types of geographies to learn about their experiences. At this time, we strongly believe that a standardized, equal area grid is an important first product and will be most accessible to the widest audience of data users.

Even within uniform grids, the geographic area of each grid cell depends on the location of that grid cell on the Earth's surface. Since many census data users are not familiar with using gridded data products, novice users should be made aware that even grid cells are not uniform in the surface (land and water) area in each grid cell.

Census Bureau Response: The Census Bureau appreciates this feedback. We strongly believe that once defined, the stability of a base grid layer is of primary importance. Area land and area water measurements will be provided for each unique grid cell, per our standard practice with all Bureau geospatial products. We recognize that while the grid may remain stable with respect to the Earth's surface, the percentage of land to water in each cell (particularly in coastal areas) may change from year to year and will necessitate at least annual updates. Again, this is standard practice for our geospatial products and we intend for grids to conform to this practice.

50. CSAC recommends that the Bureau produce a grid of land or surface (including water) area in each grid cell.

CSAC commends the Bureau on its experimentation with different grid types (including hexagonal ones). Because users' needs vary substantially, flexibility in grid size and transparency is important. While quadrilateral latitude-longitude grids are a popular and respectable choice, current business practices now employ other technologies for translating locational information into areal aggregations. New methods also allow incorporating features such as heights or altitudes, heretofore, a feature many users would want but don't have access to.

Census Bureau Response: The Census Bureau agrees with this recommendation. We agree that producing grid cells that include both land and water area measurements are necessary and intend to include these data in each geospatial grid product release.

51. CSAC recommends that the Bureau experiment with the use of geohashing or another modern method to translate latitude-longitude referenced data into spatially flexible data products. We emphasize that this would not preclude output that includes a 1 km grid.

Census Bureau Response: The Census Bureau agrees with this recommendation and will research geohashing as a potential solution.

52. CSAC recommends that the Bureau explore the attribution of a "Z" coordinate in its grid system to represent altitudes or structure heights. This addition is important for the integration and use of data for and by other agencies such as the USDA, NOAA, FEMA and others. Furthermore, this approach provides the

capability for the Bureau to project 3D images of characteristics such as housing and population density.

Census Bureau Response: The Census Bureau agrees with this recommendation. One of the major goals of the Grids effort is to produce a geographic product that will facilitate the integration of the Census Bureau's data with datasets from other federal agencies. We hope that this will result in a suite of interoperable federal geostatistical products that will allow users to join authoritative datasets across a variety of domains and create their own custom integrated datasets. Currently we have begun very preliminary research using our grid prototypes joined with parcel layers and building footprint layers. We agree that a "Z" coordinate is an important consideration and will explore the integration of altitude or elevation data.

53. **CSAC recommends that the Bureau allow for users to define the geographic specificity (extent and resolution) of grids, depending on a set of requirements about the underlying data (i.e., meeting certain thresholds). Testing would have to be done to compare top-down data products (such as those on the Bureau's geographic "spine" of nested regions) with user-generated gridded products.**

Census Bureau Response: The Census Bureau partially accepts this recommendation. We recognize that the production of gridded datasets can enable the creation of a variety of derivative products such as user-defined geographies and variable-resolution aggregated units. As noted, we are firmly committed to first producing a grid that is stable with equal area cells as a base unit that can be used as an input to these future projects. We look forward to future discussions with CSAC on developing these concepts.

54. **CSAC recommends that the Bureau compare estimates of spatial population counts from its new US gridded products with international grid statistical products applied to the US for validation and ground truthing. In addition, comparisons should not only be with respect to the predictions themselves, but also with respect to differences in inputs, methodologies, and calibration techniques.**

Census Bureau Response: The Census Bureau agrees with this recommendation. We intend to conduct this research and look forward to discussing our findings with CSAC. At this point we have reviewed a variety of global products to inform our own decision making and look forward to further research once we have finalized our plans.

While many geophysical, health and environmental scientists use raster data products (the usual format for grids), many statistical data users are familiar with tabular-formatted or vector-spatial data. Gridded data can be used in a wide variety of applications, but guidelines, metadata and perhaps even some example code on generating summary statistics will be needed to introduce non-spatial statistical data users to these new products. This includes description of underlying data and methods (including those involved in map projections or conversion between different coordinate systems that may be necessary), and assumptions used in modeling.

Census Bureau Response: The Census Bureau accepts this recommendation. We agree that these are all important lines of research and recognize the need to develop robust documentation and support materials for our data users. We agree on all points raised for consideration.

55. CSAC recommends that the Bureau produce guidelines for use of gridded data products, both those constructed for the US and other countries. This could include example programs (e.g., R, python) for users on how to summarize gridded data into tabular data or use with other types of data (e.g., how to calculate population or housing density and attach that information to user-supplied data of interest).

Census Bureau Response: The Census Bureau accepts this recommendation. As with the previous item, we agree that these are all important lines of research and recognize the need to develop robust documentation and support materials for our data users. We agree on all points raised for consideration.

There is a demand for population distributional data that represents daytime and seasonal movements. (Ambient population mapping is already produced by Oak Ridge National Laboratory in their LandScan USA product.) This could be a fruitful area of future research. Similarly, block (and block-group) level data are limited thematically in current data products, despite the need for small-area data on the population by fine age groups as well as sex, race and ethnicity, and other.

Census Bureau Response: The Census Bureau accepts this recommendation. We agree that grids are a novel product and could support the production of a variety of innovative geospatial and statistical products. We look forward to continuing to explore various options in our research and planning.

56. CSAC recommends the Bureau summarize and report back to CSAC on themes and temporal specifications that cannot easily be met by restricted data for the new US gridded data products.

Census Bureau Response: The Census Bureau accepts this recommendation. We agree to begin compiling a preliminary list of themes and specifications and look forward to discussing this topic further.

CSAC commends the Bureau on its relationship with Federal Emergency Management Administration (FEMA). Puerto Rico participates in the International Database program, and in National products, such as the ACS and Population Estimates. At the same time, Puerto Rico is in the path of potential hurricanes every year. As a result, there is an urgent need for gridded data products on Puerto Rico to support disaster recovery efforts, either through the National grid effort or as an International Database Program engagement. In addition, federal partners, such as FEMA, may find it useful to fund such an effort to leapfrog forward on the Puerto Rico grid agenda (should the Bureau need additional resources, especially in soliciting user feedback).

Census Bureau Response: The Census Bureau understand this recommendation. We fully recognize the significant data needs for Puerto Rico and all the other inhabited territories (specifically American Samoa, Guam, the Commonwealth of Northern Mariana Islands, and the U.S. Virgin Islands) particularly for disaster planning and response efforts. We recognize FEMA as a critical stakeholder in the US National Grids effort and have solicited and received substantial input regarding their requirements for Census Bureau products. We will continue to include and solicit input from FEMA and other federal data users as a core stakeholders in our planning efforts for Census Bureau products.

57. **CSAC recommends that the Bureau include Puerto Rico as a test case for its new gridded products, producing perhaps prototype grids than for other locations, and then providing those grids to FEMA and municipality authorities in Puerto Rico for feedback, and potential use in planning before the next storm arrives.**

    Census Bureau Response: The Census Bureau accepts this recommendation and will add Puerto Rico to our current area list for prototyping.

CSAC wishes to be kept apprised of the Bureau's efforts to develop and deliver new gridded data products.

Census Bureau Response: The Census Bureau looks forward to updating CSAC on our progress at an upcoming committee meeting.