

**CSAC Administrative Records Working Group
Final Report
4 September 2018**

Overview

The Census Bureau is investigating the use of Administrative Records including data from federal and Third Party sources to improve the quality of demographic data collected for the 2020 Decennial Census and for ongoing American Community Surveys. Administrative Records are used to assign population counts and demographic characteristics for non-respondent households, thus reducing costs of physical follow-ups, as well as cutting back on respondent burden. Specific application of records from previous data collection or alternate data sources supports generation of a more comprehensive roster of households than can be collected by physical household visits alone, one that distinguishes vacant units from occupied units, and determines units that are not likely residential at the time of Census data collection.

Administrative Records are important for a number of reasons. Roughly 50 million non-respondent follow-up (NRFU) addresses must be checked and identified as Occupied, Vacant or Non-Existent. As of March 2018, Census Subject Matter Experts (SMEs) estimated that of the total 2010 NRFU addresses, 30 million are Occupied, 15 million are Vacant, and 5 million are Non-Existent. A particularly compelling need to explore the utility of Administrative Records to augment Census counts is an expectation that the number of NRFUs for upcoming Census data collections may rise given the American public's increasingly widespread skepticism, mistrust, or fear in government oversight activities.

Recent tests at Census indicate that the use of Administrative Records can distinguish among the three categories (Occupied, Vacant, Non-Existent) and significantly reduce the number of field staff required for repeated physical NRFU follow-up visits. SMEs argue that there is potential for as much as \$1.4 billion reduction in follow-up data collection costs. Very little time remains before the start of 2020 Decennial End-to-End testing. By the end of fiscal year 2018, the Census Bureau will make final decisions on the list of administrative record sources that may provide reliable information and improve the quality of data collection. Alternative sources can fill in missing information on population counts and characterization. Statistical models and methods can be refined with alternative sources as well, although testing is needed to understand which statistical methods will best inform unit counts as well as distinguishing among Vacant, Non-Existent and Occupied units.

Working Group Formation

Charter and Tasks

The Census Science Advisory Committee formed a Working Group in June 2017 to focus on these important considerations. A Working Group Charter was approved in July 2017 directing the working group attention to the following activities:

- Provide input on expanding and refining the use of administrative records to reduce respondent burden and improve statistical analyses;
- Review and assess current research, plans and procedures in what is being used, and on methods including ongoing statistical testing;
- Identify discrepancies and inefficiencies that can be improved in identifying occupied, vacant and non-existent addresses;
- Consider criteria in selecting what is used from the records and their impact on current disparities in reporting on households and individuals;

- Offer recommendations on additional data sets to determine vacant, non-existent and occupied housing units;
- Discuss if these additional data sets are a cost-effective means that will offer benefits to research and analysis beyond what is being currently used at the Bureau.
- Assist the Census Bureau in developing strategies in using various resources for data and increase awareness through an outreach plan amongst state-level organizations so they understand how all can benefit from the use of these types of records.

Timing and Deliverables

An initial conference call was held 14 June, 2017 with CSAC Chair Barbara Anderson, Working Group Convener Barbara Buttenfield, and Kim Brown and Tara Dunlop Jackson (both from Census) to discuss mechanics, tasks and to recommend members. The Working Group met monthly, with conference phone calls held 2 August, 28 August, 25 September, 23 October, 27 November, 18 December, as well as 29 January and 26 February 2018 to discuss the draft report with SMEs. Barbara Buttenfield gave a detailed presentation at the March 2018 CSAC meeting. One additional conference call occurred in August 2018 to hear a final briefing by SMEs.

The deliverables for the Working Group include a plan of work presented at the September 2017 Census Scientific Advisory Committee outlining activities from July 2017-July 2018. A report of Working Group activities was drafted with input from Census SMEs, and presented and discussed at the March 2018 CSAC meeting. A deck of PowerPoint slides for that presentation was delivered to Census in March 2018. Discussion and recommendations developed by the full CSAC Committee were incorporated into the final report to be submitted to the Census Bureau in September 2018.

Members and Subject Matter Experts

The Working Group consists of the following individuals (all members of CSAC):

Barbara Buttenfield, Working Group Convener

Professor of Geography, University of Colorado, Boulder CO

Allison Plyer, Working Group Member

Executive Director and Chief Demographer, The Data Center at Nonprofit Knowledge Works, New Orleans, LA

Ken Simonson, Working Group Member

Chief Economist, Associated General Contractors of America, Arlington VA

Jack Levis, Working Group Member

Senior Director of Process Management, United Parcel Service, Timonium, MD

Krishna Rao, Working Group Member

Director of Research and Products at Zillow Corporation, Seattle WA

Barbara Anderson, CSAC Chair and Ex-Officio Working Group Member

Professor of Sociology and Population Studies, University of Michigan, Ann Arbor MI

Census Subject Matter Experts (SMEs) who assisted the Working Group include:

Vincent Tom Mule (Decennial)

Quentin Brummet (Center for Administrative Records Research and Applications (CARRA))

Andrew Keller (Decennial)

Moises Yi (Center for Economic Studies)

Nicholas Jones (Population division)

Scott Konicki (Decennial)

Larry Warren (Center for Economic Studies)
Mark Leach (CARRA)
Deborah Wagner (Center for Statistical Research and Methodology)
Darcy Morris (Center for Statistical Research and Methodology)
Gary Chappell (Decennial)
Jennifer Ortman (American Community Survey Office)
Dave Raglin (American Community Survey Office)

Tommy Wright (Designated Federal Officer, Census) participated actively in the conference calls, also offering advice and clarification on procedures.

Selection of Sources for Administrative Records

By the end of fiscal 2018, Census will finalize the list of sources. The initial set of potential sources of Administrative Records under consideration was drawn from the following sources and agencies (in alphabetical order):

- Center for Administrative Records Research and Applications (CARRA)
- Centers for Medicare and Medicaid Services (CMS)
- Department of Housing and Urban Development (HUD)
- Indian Health Service (IHS)
- Internal Revenue Service (IRS)
- Parcel and Multiple Listing Service (MLS) data
- Selective Service System (SSS)
- Social Security Administration (SSA)
- Supplemental Nutrition Assistance Program (SNAP) and other state program data
- United States Postal Service (USPS)
- Other public and private data sources as available and appropriate, such as Third-party Veterans Service Groups

Core data sources used to corroborate information and/or used in models

- CARRA data such as KidLink
- CMS Medicare Enrollment Database
- IHS Patient Registration File
- IRS Form 1040 Individual Tax Returns
- IRS Form 1099 Informational Returns
- USPS Undeliverable-As-Addressed and Delivery Sequence File

Additional data sources used to corroborate information in core sources and/or used in models

- 2010 Decennial Census population counts and demographic characteristics
- American Community Survey 5-year Block Group estimates
- CARRA Best Race and Best Hispanic Research
- Census Bureau Master Address File (MAF)
- CMS Medicaid Statistical Information System
- HUD Computerized Homes Underwriting Management System
- HUD Public and Indian Housing Information Center
- HUD Tenant Rental Assistance Certification System
- SSA Numerical Identification (NUMIDENT) file
- SSS Registration File
- Third-party Veterans Service Groups of Illinois (VSGI) files

Working Group Activities

Briefings were provided by SMEs during the Working Group meetings on Administrative Records data sources, and roles of such data in statistical modeling to eliminate some NRFUs from subsequent attempts to interview as well as to characterize NRFUs deemed to be Occupied.

Administrative Records Data Sources

Discussions at early meetings focused upon the accuracy and reliability of alternative data sources that could be used to generate an Administrative Records roster, including state- and local-level data sources such as driver's license and car registrations, school enrollment data, birth and death certificates. The use of many suggested data sources is constrained by the Federal Right to Privacy Act of 1974 (such as school enrollment data), Title 13 (relating to individual privacy) and Title 26 (relating to financial privacy). Lack of addresses on birth certificates and unreliable addresses on drivers' licenses and car registrations caution against use of these administrative data sets.

A briefing on using state program data sets indicates higher costs and more variability than had been anticipated. Only 24 states have so far agreed to provide data to Census, in the form of Supplemental Nutrition Assistance Program SNAP, Temporary Assistance for Needy Families (TANF), or Special Supplemental Nutrition Program for Women, Infants, and Children (WIC). Delays can follow, in that it can take 2-3 years to get data after an agreement is signed.

A briefing on using CoreLogic MLS real estate listings to predict housing unit vacancy status reveals that the vast majority of vacant housing units do NOT have active listings in the MLS. Moreover, SMEs reported at the briefing that MLS data is extremely messy to work with, leading to the conclusion that the value derived from working with MLS data is not worth the cost of doing so.

Census workflow for handling NRFUs

As of 2016, SMEs utilized a hierarchical contact strategy to build an Administrative Records roster. The first stage uses administrative records to distinguish between Occupied and Vacant or Non-Existent addresses. In the second stage, addresses deemed Vacant or Non-Existent are handled with a subsequent mailing to solicit a self-response. The focus in this stage is placed initially on identifying Vacant units, rather than Occupied units, since (once identified) Vacant units do not require follow-up visits. Remaining NRFU addresses may fall into any of the three categories. An attempt is made to interview these addresses; and if the address is still unresolved, administrative records are used to identify occupied addresses, which are sent a mailing to solicit a self-response. If either solicitation for a self-response is unanswered, the first-stage characterization stands, and administrative records are used. For addresses from this second stage that cannot be resolved as Occupied, additional field contacts are utilized. It is the mailings and repeated attempts to interview (including additional field contacts) that drive up the cost of handling NRFUs.

Categorizing NRFUs as Occupied, Vacant or Non-Existent

An early briefing described tests to evaluate statistical models and strategies that rely upon administrative records to label NRFU addresses as Occupied, Vacant or Non-Existent, and to deal somewhat differently with each category. A multinomial logistic regression model uses USPS, IRS, CMS and IHS data in addition to 2010 Census NRFU categorization to assign probability that an address is Vacant or Non-Existent. American Community Survey (ACS) area-level estimates for tenure, race and income are also considered. Probabilities are used to develop cut-off workloads for NRFU workloads at vacant addresses. A slightly different logistic regression model is developed

for Occupied addresses to assign probabilities that a household is being enumerated at the correct address (informed by Administrative Records indicating a different address), and that the household composition is correct (informed by 2010 Administrative Records and ACS-area estimates of tenure, income and race). Core sources for Occupied units include IRS, CMS and IHS data as well as SSA, CARRA data, and USPS and VSGI data sources. Additional sources listed above are being tested, with Census finalizing its determination by September 2018.

The Working Group questioned whether models utilizing ACS area-level summaries as input included statistical controls based on area-level Margins of Error (MoEs) that are known to increase for smaller enumeration areas such as block group 5-year data. The models at present do not use MoEs, and this could reduce reliability of address determination in any category. Another reliability constraint that was questioned is that update schedules vary for each input data source and this could also impact the quality of address categorization.

After a concern was raised that there may be value in looking beyond the point estimates from the statistical model, a briefing was arranged on using the variance and covariance of parameter estimates in the determination of addresses to be removed. Covariates include ancillary independent variables such as mobility, poverty, income, and household composition, drawn from both the 2010 Census and Administrative Records as listed in an earlier section of this report. Covariates were entered into the logistic regression models and tested on 2015 ACS data to eliminate Vacant and Non-Existent addresses, allowing removal of roughly 5% (roughly 2,700) of the units originally identified as Occupied. Further analysis comparing prediction to the percentage of the samples that were reported as unoccupied in NRFU for the 2015 Census Test suggests that adjusting point estimates to incorporate the underlying variance of parameter estimates does not significantly improve the error rate of address removal.

Testing to Compare National with Sub-National Statistical Models

The goal of this test was to establish if determination of Occupied units for the Administrative Records roster could be improved by implementing multiple models that highlight either urban versus rural differences or regional ethnic or economic characteristics. A statistical analysis applying a single model for the entire nation (a “baseline”) was compared with three “subnational” models including an urban-rural (a “large county”) binary comparison, a four region model based on Census Geographic regions, an owner-renter binary comparison using a threshold of 20% renters, and a binary model that stratified 20% or higher Hispanic concentrations. Tests were performed to analyze occupancy as well as household composition, distinguishing true and false positives by field checking in a single study area.

A true positive would assign high probability that an address is Occupied to a field-checked address that proves to be Occupied. The comparison showed insufficient differences between the national model and any of the sub-national models to warrant adoption of an alternative for determining Occupied units, given that application of a single model for the entire nation will be computationally much simpler and faster to process. The testing did show however that the owner-renter model yielded better results for determining unoccupied units (that is, Vacant and/or Non-Existent taken as a single category).

The sub-national models were evaluated nationally, as follows. For the “large counties” example, models were fit over a national sample of the 2010 NRFU data in the 10 largest counties by population. Model parameters were scored over all the 2010 NRFU data on those 10 largest counties, with a similar procedure applied to the 2010 NRFU data for the remaining counties. The two scored sets of results (large & non-large counties) were combined to form one national set of results, which were compared against the single 2010 national “baseline” model. A separate test

was performed for data in Maricopa County Arizona in 2015, where modeling methods were applied, compared with a “baseline” and evaluated for the single county.

Discussion followed this briefing to clarify on the choice of specific percentages (between owners and renters, or for Hispanic concentrations) that do not match national averages. Working Group members asked for example would doubling the threshold for renters improve results? SMEs responded that this type of sensitivity analysis was not tested, given time restrictions. Concerns were raised by Working Group members about additional testing of the subnational models in only a single study area (Maricopa County) that may not fully represent the range of demographic, economic or urbanization conditions that vary regionally across the nation.

Assigning Demographic Characteristics to Units in the Administrative Records Roster

Discussions at later briefings focused on other possible uses for Administrative Records, including the option to expand on population counts by incorporating characterizations of age, sex, race or ethnicity, and household composition that might be collated or statistically estimated with good reliability. Characterizations of race, ethnicity, tenure and income have varying levels of reliability that are more difficult to judge, since the quality can be expected to vary with the data source, the geographic region, and temporality since various Administrative Records are collected at varying time periods. Working Group Members received a document reporting on the Census Person Identification Validation System (PVS) that assigns a unique identifier (Protection Identification Key, or PIK) to records in federal, state and third party data sources to facilitate record linkages among the files. The PVS is used to validate (for example) resident addresses and household composition between multiple data sources.

The Working Group expressed particular interest in using Administrative Records to assign characteristics (sex, age, race and ethnicity) as opposed to using the roster solely for population counts and household composition. For units enumerated as Occupied on the Administrative Records roster, an attempt is made to characterize as well as count residents. Characterizations for individuals include age, sex, race and ethnicity, and relationship to the Head of Household; and for Housing units, characterizations include tenure (for occupied units) and details on vacancy (for unoccupied units). SMEs examined the feasibility of using previously collected Census and Administrative Records data to assign characteristics on eight race groupings including American Indian/Alaska Native (AIAN), Native Hawaiian/Pacific Islander (NHPI), Middle Eastern – North African (MENA), Asian Alone, Black Alone, White Alone, 2 or More Races, and Some Other Race Alone; and on Hispanic origin.

A test simulating 2010 Census data assigned race and Hispanic origin to 16.7 million individuals placed into Administrative Records Occupied units, basing assignments on a hierarchy of five data sources. In order of priority, these sources included (first) ACS and 2000 Census data, (second) one of three sources (SSA Numident, SNAP data, and CARRA data for which a PIK identifier as well as country of origin or race had been collected). Absent information from these sources, only 4.4% (Race categories) and 4% (Hispanic origin) of the individuals remained to be characterized; and one of two methods was applied. A “Within Household” assignment designated race/ethnicity and Hispanic origin to everyone in the unit based upon any person in the unit whose designation was contained in the first five data sets. A “Nearest Neighbor Hot Deck” assignment based the designation to everyone in the unit upon the designation for the householder of a spatially proximal unit. The largest race differences between the simulation and the 2010 Census were for “White Alone” (0.57%), “Some other race alone” (0.41%) and “Two or more races” (0.16%). Differences for other race categories and for Hispanic origin were much smaller (0.01% -0.07%).

Working Group members suggest that a review and revision of these simulations might be useful given the Census Bureau's recent decision against using a combined question format on the census form for collecting race and ethnicity, or a separate "Middle Eastern or North African" category.

Other characteristics including Name, Age, Date of Birth, and Sex have been shown to resolve largely on the basis of 2010 Census data and SSA Numident information, with 97% or higher correct assignments if a PIK is available. For young persons in the 2010 Census whose parent in the KIDLINK files was the householder, 98% of the time that child was reported as the biological child of that census householder. Based on this finding, Census SMEs are comfortable assigning young persons in households to be "biological children" if the KIDLINK records show that the young person is the child of the census householder. Using KIDLINK to include children on the Administrative Records roster showed small improvements in count agreement over the Baseline approach, and with more noticeable improvements in high poverty areas. This is an important finding for improving the accuracy of the count of young children, particularly minority children, who have historically been undercounted at disproportionately high rates in the Decennial Census.

Tenure assignments are the focus an ongoing examination by SMEs, who are exploring HUD data sets such as the Tenant Rental Assistance Certification System (TRACS) and the Public and Indian Housing Inventory Management System (formerly PIC). CoreLogic data sets may be useful in addition to the federal sources, but this will require further analysis.

Testing ARs during the 2018 End-to-End Testing

Working Group members received an early August 2018 briefing summarizing the role of ARs in the 2018 End-to-End Census Test in Providence, Rhode Island. A reduced set of AR data sources supported AR Roster creation for the 2018 End-to-End Testing:

- IRS Form 1040 Individual Tax Returns
- IRS Form 1099 Informational Returns
- CMS Medicare Enrollment Database
- IHS Indian Health Service Patient Registration File
- Census KidLink File

These were combined with the following data for Administrative Record assignment of household units and persons as well as characteristics (age, sex, race, ethnicity and household composition):

- Census Master Address File (MAF)
- American Community Survey
- USPS Undeliverable as Addressed (UAA) files for 2018 End-to-End Testing

Three phases of effort included an initial phase, when a 2010 version of AR files was created for the 2010 NRFU person and housing unit files. A series of logistic regression models was based upon the 2010 data, and used on the 2018 NRFU End-to-End Test data for initial predictions categorizing addresses as Occupied, Vacant, or Non-Existent ("Delete"). The "Delete" addresses might for example be the result of a non-existent street address, by way of clarification. When IRS files became available in June, additional Occupied addresses were identified in the second phase as NRFU operations began. A third phase occurred after all NRFU attempts were exhausted; threshold parameters for the logistic regression models were adjusted to resolve remaining addresses that had not yet been categorized. Adjusted thresholds resolved some but not all of the remaining addresses. In some cases, the adjusted thresholds resulted in overlapping categorization of some addresses with high probability of being Vacant as well as Non-Existent. Some overlaps were also

encountered in categorization for the original thresholds. All overlapping addresses in the 2018 End-to-End Test were added to the NRFU list.

Three changes to the categorization were applied to improve results in the 2018 Test. First, a “corroboration requirement” was applied to Occupied units such that at least one person is identified from multiple federal sources as residing at that address. Second, for addresses initially categorized as Vacant or Delete, at least two attempts to deliver a postcard must be marked by the U.S. Postal Service as “Undeliverable as Addressed” (UAA), including the initial solicitation postcard delivered on Census Day. Addresses for which a subsequent postcard delivery was successful were added to the list for NRFU physical follow-ups. Third, a set of “supplemental” NRFU addresses that were not in the initial enumeration universe were sent a solicitation postcard in April, and a separate regression model was developed to predict one of the three categories.

The Providence End-to-End Test universe included 279,298 units, 167,971 NRFU addresses, and 667 supplemental addresses. In the NRFU universe, 39,870 units (23.7%) were categorized as Occupied, 12,256 units (7.3%) initially categorized as Vacant, 1,321 units (0.8%) as Delete, and 2,266 (1.3%) as overlapping (could be either Vacant or Delete); the remainder could not be determined. SMEs anticipate a smaller proportion of NRFU cases (possibly as low as 15%) will be identified as Occupied in the 2020 Decennial count because the End-to-End Test does not include outreach, promotion and advertising. A 20% sample of Occupied, Vacant, Delete, and overlapping cases were identified for full NRFU fieldwork in the End-to-End Test, to compare outcomes against the modelled AR determinations. As Phase 1 ended, projected final AR determinations for Vacant and Delete units requiring additional physical follow-ups was 5,295 (Vacant) and 359 (Delete).

Results reflect previous census tests with respect to household composition: following Phase 1, most AR Occupied households (85,068) are comprised of one adult with no children (38.5%), or 2 adults with or without children (roughly 50%). Age and sex characteristics of the AR occupied Units in the sample indicate that almost 23% are aged 17 years or younger, with roughly even proportions of males and females in older age groups.

Following receipt of information from IRS data files in Phase 2, it appeared that 126,412 addresses remained in the NRFU universe. Applying the corroboration requirement (multiple sources attesting to a person’s place of residence) along with two additional criteria (that 1-6 people reside in a household, and that 1-3 of those residents are adults), the number of Occupied units dropped to 47,656 units. Of these, only 5,185 units (4.1%) showed a change in person counts. Logistic regression models were run on these Occupied units identifying 852 addresses having a revised roster that met the Occupied criteria, which comprises only 0.7% of the IRS-provided addresses. A 20% sample of these (171 units) were selected for NRFU fieldwork. Working Group members suggested that informing NRFU fieldworkers what is the projected proportion of Vacant or Delete units as they go out in the field might reduce their workload in specific areas.

Compositions of these 171 households were summarized. It is notable that following Phase 2 (NRFU visits), proportions for the most frequent categories tend to change from initial household composition, and some changes are substantial. For example, the percent of households with 1 adult and no children rises from 38.5% to 53.3%; and the percentage of 2 Adult households with no children drops from 30.1% to 15.1%. Working group members advised SMEs to examine these results in more detail to assess if the discrepancies are meaningful. SMEs are working with the Census Population Division to summarize counts across race and Hispanic ethnicity.

Working group members questioned the SMEs about several factors related to the cost of multiple NRFU physical visits, relative to the initial estimate of a \$1.4 billion savings projected for the use of ARs in the 2020 Decennial Census count. One question referred to the cost reduction associated with 1 NRFU per household as opposed to 6 NRFU visits. Another question asked for a comparison

between using ARs to replace or supplement past estimations and imputed counts. A third asked if the AR records can adequately distinguish between (for example) Occupied households that require one versus many NRFU visits, and can the reductions in savings for all of these eventualities be projected against the \$1.4 billion? SMEs were cautiously optimistic that savings will be achieved, and they will explore whether the End-to-End Testing can inform or quantify potential savings.

Use of Administrative Records to Improve American Community Survey (ACS) counts

This topic was suggested at several points during the cycle of work, but the timetable for generating a final report was too short for the Working Group and SMEs to address it in detail. Some aspects of using ARs for ACS might differ, for example since the timeframe for ACS data collection is much shorter than for the Decennial Census, and the collection procedure is a sample rather than a complete count. As time and staffing permit in future months, CSAC might consider forming a new Working Group to address the use of ARs for ACS counts.

Challenges

The challenges associated with use of ARs for the purpose of reducing respondent burden and improving the quality of collected information are really no different than those for managing any enterprise project. Management issues include Scope (i.e., types of data sources), Time (in this case, time remaining prior to testing and actual Decennial Census data collection dates), Cost (of data processing and/or arranging for data access from state, local or private sources), Quality (potential for statistical bias or error), Risk (how much uncertainty is acceptable, and how much improvement warrants the extra processing and labor costs) and Dependencies (what are the critical paths, what sequencing could fail and thus impact other tasks). Each of these plays an important role in prioritizing and monitoring tasks, personnel, and workflows. The overarching issue is to maintain the important balance between effort and improvement in quality of the outcome. Therefore, it is essential to prioritize effort and deliverables based on benefit, cost, risk, and dependencies. This ensures projects produce the greatest value for the work performed.

Working group discussions have been informed about some of these management issues, especially regarding quality (reduction of bias and error), labor (reducing repeated number of physical field visits) and time as a limiting factor for undertaking more extensive testing. Cost is also an important component, especially with regard to third-party datasets, such as evidenced by the feasibility study on MLS data. The focus of attention for this Working Group may not extend to advising Census on how to prioritize tasks for using Administrative Records. The Working Group briefings and discussions with SMEs make it clear nonetheless that Census must examine these management criteria and establish clear priorities to insure that the use of Administrative Records can improve the quality of population counts and characterization in the 2020 Decennial Census.

Summary

As with all projects, resources are limited; and resources must be considered to include labor, expertise, and some aspects of time, as for example the time required for testing possible statistical methods or data collection strategies. We commend Census for developing an extensive suite of testing, cross validation, and analysis to demonstrate stability of results in identifying occupied, vacant and nonexistent addresses by multiple statistical techniques. We encourage continued attention to developing strategies for statistical testing, bias reduction, parameter selection, and validation in the use of Administrative Records.

It has been stated that the successful use of ARs can impact the 2020 Census cost by more than a billion dollars. This makes the prioritization of effort even more important. From the working group conference discussions however, it is unclear if the effort is properly prioritized. Is Census working on items which independently have low risk and significant benefits? Is initiation or completion of some testing, collection or data processing tasks dependent on other activities before they can add value? The work performed on AR collation and use is outstanding, but the Working Group cannot validate if it will lead to the savings projected. We encourage the Bureau to evaluate each effort regarding Administrative Records, and rate and rank each based on the criteria above (scope, time, cost, quality, risk, dependency). Ideally, the highest priority items would have the greatest benefit, with lowest cost, risk, and dependency.

A second and equally important issue relates to the use of Administrative Records not just for population and household counts and characteristics, but also for enumeration unit summaries and estimates. Given the Census Bureau will need to produce block level estimates, the Census Bureau must determine the reliability of integrating current data with administrative records to create estimates of citizenship by census block, for the Decennial Census as well as for ACS surveys. Due to differences in what is reported as well as where in the two processes Administrative Records are introduced, differing reliabilities could emerge. A recent Working Paper (Brown et al 2018) alludes to discrepancies between survey-collected citizenship and administrative records being higher than initial expectations, due to several contributing factors such as misreporting or nonreporting citizenship status of individuals or household members, or delays in reporting naturalizations. Authors conclude that adding a citizenship question to the 2020 Census could increase fieldwork costs and reduce the quality of the population count. The Working Group encourages Census SMEs to continue their vigilance about the citizenship question as well as to other potential sources of reduced quality in population counts and household characterizations.

Although the present Working Group completes its set of charter tasks and deliverables with this report, the full CSAC committee would like to be involved in the continued testing and analysis of the use of Administrative Records in upcoming census and surveys that Census undertakes. This includes requesting information on the timeline, relevant decision makers, and other information that will keep CSAC up to speed. As the current Working Group completes its tasks, we encourage Census to consider convening a new CSAC Working Group to consider for example the possibly expanded roles for Administrative Records in light of the inclusion of a citizenship question on the 2020 Decennial Census.

References

Brown JD, Hegeness ML, Dorinski SM, Warren L and Yi M. 2018. *Understanding the Quality of alternative Citizenship Data Sources for the 2020 Census*. Washington DC: U.S. Bureau of the Census, Center for Economic Studies CES 18-38, August 2018.
<https://www2.census.gov/ces/wp/2018/CES-WP-18-38.pdf>