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

January 19, 2018

MEMORANDUM FOR: Wilbur L. Ross, Jr.
Secretary of Commerce

Through: Karen Dunn Kelley
Performing the Non-Exclusive Functions and Duties of the Deputy Secretary

Ron S. Jarmin
Performing the Non-Exclusive Functions and Duties of the Director

Enrique Lamas
Performing the Non-Exclusive Functions and Duties of the Deputy Director

From: John M. Abowd
Chief Scientist and Associate Director for Research and Methodology

Subject: Technical Review of the Department of Justice Request to Add Citizenship Question to the 2020 Census

The Department of Justice has requested block-level citizen voting-age population estimates by OMB-approved race and ethnicity categories from the 2020 Census of Population and Housing. These estimates are currently provided in two related data products: the PL94-171 redistricting data, produced by April 1st of the year following a decennial census under the authority of 13 U.S.C. Section 141, and the Citizen Voting Age Population by Race and Ethnicity (CVAP) tables produced every February from the most recent five-year American Community Survey data. The PL94-171 data are released at the census block level. The CVAP data are released at the census block group level.

We consider three alternatives in response to the request: (A) no change in data collection, (B) adding a citizenship question to the 2020 Census, and (C) obtaining citizenship status from administrative records for the whole 2020 Census population.

We recommend either Alternative A or C. Alternative C best meets DoJ's stated uses, is comparatively far less costly than Alternative B, does not increase response burden, and does not harm the quality of the census count. Alternative A is not very costly and also does not harm the quality of the census count. Alternative B better addresses DoJ's stated uses than Alternative A. However, Alternative B is very costly, harms the quality of the census count, and would use substantially less accurate citizenship status data than are available from administrative sources.

Summary of Alternatives			
	Alternative A	***Alternative B***	***Alternative C***
Description	No change in data collection	Add citizenship question to the 2020 Census (i.e., the DoJ request), all 2020 Census microdata remain within the Census Bureau	Leave 2020 Census questionnaire as designed and add citizenship from administrative records, all 2020 Census microdata and any linked citizenship data remain within the Census Bureau
Impact on 2020 Census	None	Major potential quality and cost disruptions	None
Quality of Citizen Voting-Age Population Data	Status quo	Block-level data improved, but with serious quality issues remaining	Best option for block-level citizenship data, quality much improved
Other Advantages	Lowest cost alternative	Direct measure of self-reported citizenship for the whole population	Administrative citizenship records more accurate than self-reports, incremental cost is very likely to be less than $2M, USCIS data would permit record linkage for many more legal resident noncitizens
Shortcomings	Citizen voting-age population data remain the same or are improved by using small-area modeling methods	Citizenship status is misreported at a very high rate for noncitizens, citizenship status is missing at a high rate for citizens and noncitizens due to reduced self-response and increased item nonresponse, nonresponse followup costs increase by at least $27.5M, erroneous enumerations increase, whole-person census imputations increase	Citizenship variable integrated into 2020 Census microdata outside the production system, Memorandum of Understanding with United States Citizen and Immigration Services required to acquire most up-to-date naturalization data

Approved: _____ Date: _____

John M. Abowd, Chief Scientist
and Associate Director for Research and Methodology

Detailed Analysis of Alternatives

The statistics in this memorandum have been released by the Census Bureau Disclosure Review Board with approval number CBDRB-2018-CDAR-014.

Alternative A: Make no changes

Under this alternative, we would not change the current 2020 Census questionnaire nor the planned publications from the 2020 Census and the American Community Survey (ACS). Under this alternative, the PL94-171 redistricting data and the citizen voting-age population (CVAP) data would be released on the current schedule and with the current specifications. The redistricting and CVAP data are used by the Department of Justice to enforce the Voting Rights Act. They are also used by state redistricting offices to draw congressional and legislative districts that conform to constitutional equal-population and Voting Rights Act nondiscrimination requirements. Because the block-group-level CVAP tables have associated margins of error, their use in combination with the much more precise block-level census counts in the redistricting data requires sophisticated modeling. For these purposes, most analysts and the DoJ use statistical modeling methods to produce the block-level eligible voter data that become one of the inputs to their processes.

If the DoJ requests the assistance of Census Bureau statistical experts in developing model-based statistical methods to better facilitate the DoJ's uses of these data in performing its Voting Rights Act duties, a small team of Census Bureau experts similar in size and capabilities to the teams used to provide the Voting Rights Act Section 203 language determinations would be deployed.

We estimate that this alternative would have no impact on the quality of the 2020 Census because there would be no change to any of the parameters underling the Secretary's revised life-cycle cost estimates. The estimated cost is about $350,000 because that is approximately the cost of resources that would be used to do the modeling for the DoJ.

Alternative B: Add the question on citizenship to the 2020 Census questionnaire

Under this alternative, we would add the ACS question on citizenship to the 2020 Census questionnaire and ISR instrument. We would then produce the block-level citizen voting-age population by race and ethnicity tables during the 2020 Census publication phase.

Since the question is already asked on the American Community Survey, we would accept the cognitive research and questionnaire testing from the ACS instead of independently retesting the citizenship question. This means that the cost of preparing the new question would be minimal. We did not prepare an estimate of the impact of adding the citizenship question on the cost of reprogramming the Internet Self-Response (ISR) instrument, revising the Census Questionnaire Assistance (CQA), or redesigning the printed questionnaire because those components will not be finalized until after the March 2018 submission of the final questions. Adding the citizenship question is similar in scope and cost to recasting the race and ethnicity questions again, should that become necessary, and would be done at the same time. After the 2020 Census ISR, CQA and printed questionnaire are in final form, adding the citizenship question would be much more expensive and would depend on exactly when the implementation decision was made during the production cycle.

For these reasons, we analyzed Alternative B in terms of its adverse impact on the rate of voluntary cooperation via self-response, the resulting increase in nonresponse followup (NRFU), and the consequent effects on the quality of the self-reported citizenship data. Three distinct analyses support the conclusion of an adverse impact on self-response and, as a result, on the accuracy and quality of the 2020 Census. We assess the costs of increased NRFU in light of the results of these analyses.

B.1. Quality of citizenship responses

We considered the quality of the citizenship responses on the ACS. In this analysis we estimated item nonresponse rates for the citizenship question on the ACS from 2013 through 2016. When item nonresponse occurs, the ACS edit and imputation modules are used to allocate an answer to replace the missing data item. This results in lower quality data because of the statistical errors in these allocation models. The analysis of the self-responses responses is done using ACS data from 2013-2016 because of operational changes in 2013, including the introduction of the ISR option and changes in the followup operations for mail-in questionnaires.

In the period from 2013 to 2016, item nonresponse rates for the citizenship question on the mail-in questionnaires for non-Hispanic whites (NHW) ranged from 6.0% to 6.3%, non-Hispanic blacks (NHB) ranged from 12.0% to 12.6%, and Hispanics ranged from 11.6 to 12.3%. In that same period, the ISR item nonresponse rates for citizenship were greater than those for mail-in questionnaires. In 2013, the item nonresponse rates for the citizenship variable on the ISR instrument were NHW: 6.2%, NHB: 12.3% and Hispanic: 13.0%. By 2016 the rates increased for NHB and especially Hispanics. They were NHW: 6.2%, NHB: 13.1%, and Hispanic: 15.5% (a 2.5 percentage point increase). Whether the response is by mail-in questionnaire or ISR instrument, item nonresponse rates for the citizenship question are much greater than the comparable rates for other demographic variables like sex, birthdate/age, and race/ethnicity (data not shown).

B.2. Self-response rate analyses

We directly compared the self-response rate in the 2000 Census for the short and long forms, separately for citizen and noncitizen households. In all cases, citizenship status of the individuals in the household was determined from administrative record sources, not from the response on the long form. A noncitizen household contains at least one noncitizen. Both citizen and noncitizen households have lower self-response rates on the long form compared to the short form; however, the decline in self-response for noncitizen households was 3.3 percentage points greater than the decline for citizen households. This analysis compared short and long form respondents, categories which were randomly assigned in the design of the 2000 Census.

We compared the self-response rates for the same household address on the 2010 Census and the 2010 American Community Survey, separately for citizen and noncitizen households. Again, all citizenship data were taken from administrative records, not the ACS, and noncitizen households contain at least one noncitizen resident. In this case, the randomization is over the selection of household addresses to receive the 2010 ACS. Because the ACS is an ongoing survey sampling fresh households each month, many of the residents of sampled households completed the 2010 ACS with the same reference address as they used for the 2010 Census. Once again, the self-response rates were lower in the ACS than in the 2010 Census for both citizen and noncitizen households. In this 2010 comparison, moreover, the decline in self-response was 5.1 percentage points greater for noncitizen households than for citizen households.

In both the 2000 and 2010 analyses, only the long-form or ACS questionnaire contained a citizenship question. Both the long form and the ACS questionnaires are more burdensome than the shortform. Survey methodologists consider burden to include both the direct time costs of responding and the indirect costs arising from nonresponse due to perceived sensitivity of the topic. There are, consequently, many explanations for the lower self-response rates among all household types on these longer questionnaires. However, the only difference between citizen and noncitizen households in our studies was the presence of at least one noncitizen in noncitizen households. It is therefore a reasonable inference that a question on citizenship would lead to some decline in overall self-response because it would make the 2020 Census modestly more burdensome in the direct sense, and potentially much more burdensome in the indirect sense that it would lead to a larger decline in self-response for noncitizen households.

B.3. Breakoff rate analysis

We examined the response breakoff paradata for the 2016 ACS. We looked at all breakoff screens on the ISR instrument, and specifically at the breakoffs that occurred on the screens with the citizenship and related questions like place of birth and year of entry to the U.S. Breakoff paradata isolate the point in answering the questionnaire where a respondent discontinues entering data—breaks off—rather than finishing. A breakoff is different from failure to self-respond. The respondent started the survey and was prepared to provide the data on the Internet Self-Response instrument, but changed his or her mind during the interview.

Hispanics and non-Hispanic non-whites (NHNW) have greater breakoff rates than non-Hispanic whites (NHW). In the 2016 ACS data, breakoffs were NHW: 9.5% of cases while NHNW: 14.1% and Hispanics: 17.6%. The paradata show the question on which the breakoff occurred. Only 0.04% of NHW broke off on the citizenship question, whereas NHNW broke off 0.27% and Hispanics broke off 0.36%. There are three related questions on immigrant status on the ACS: citizenship, place of birth, and year of entry to the United States. Considering all three questions Hispanics broke off on 1.6% of all ISR cases, NHNW: 1.2% and NHW: 0.5%. A breakoff on the ISR instrument can result in follow-up costs, imputation of missing data, or both. Because Hispanics and non-Hispanic non-whites breakoff much more often than non-Hispanic whites, especially on the citizenship-related questions, their survey response quality is differentially affected.

B.4. Cost analysis

Lower self-response rates would raise the cost of conducting the 2020 Census. We discuss those increased costs below. They also reduce the quality of the resulting data. Lower self-response rates degrade data quality because data obtained from NRFU have greater erroneous enumeration and whole-person imputation rates. An erroneous enumeration means a census person enumeration that should not have been counted for any of several reasons, such as, that the person (1) is a duplicate of a correct enumeration; (2) is inappropriate (e.g., the person died before Census Day); or (3) is enumerated in the wrong location for the relevant tabulation (https://www.census.gov/coverage_measurement/definitions/). A whole-person census imputation is a census microdata record for a person for which all characteristics are imputed.

Our analysis of the 2010 Census coverage errors (Census Coverage Measurement Estimation Report: Summary of Estimates of Coverage for Persons in the United States, Memo G-01) contains the relevant data. That study found that when the 2010 Census obtained a valid self-response (219 million persons),

the correct enumeration rate was 97.3%, erroneous enumerations were 2.5%, and whole-person census imputations were 0.3%. All erroneous enumeration and whole-person imputation rates are much greater for responses collected in NRFU. The vast majority of NRFU responses to the 2010 Census (59 million persons) were collected in May. During that month, the rate of correct enumerations was only 90.2%, the rate of incorrect enumeration was 4.8%, and the rate of whole-person census imputations was 5.0%. June NRFU accounted for 15 million persons, of whom only 84.6% were correctly enumerated, with erroneous enumerations of 5.7%, and whole-person census imputations of 9.6%. (See Table 19 of 2010 Census Memorandum G-01. That table does not provide statistics for all NRFU cases in aggregate.)

One reason that the erroneous enumeration and whole-person imputation rates are so much greater during NRFU is that the data are much more likely to be collected from a proxy rather than a household member, and, when they do come from a household member, that person has less accurate information than self-responders. The correct enumeration rate for NRFU household member interviews is 93.4% (see Table 21 of 2010 Census Memorandum G-01), compared to 97.3% for non-NRFU households (see Table 19). The information for 21.0% of the persons whose data were collected during NRFU is based on proxy responses. For these 16 million persons, the correct enumeration rate is only 70.1%. Among proxy responses, erroneous enumerations are 6.7% and whole-person census imputations are 23.1% (see Table 21).

Using these data, we can develop a cautious estimate of the data quality consequences of adding the citizenship question. We assume that citizens are unaffected by the change and that an additional 5.1% of households with at least one noncitizen go into NRFU because they do not self-respond. We expect about 126 million occupied households in the 2020 Census. From the 2016 ACS, we estimate that 9.8% of all households contain at least one noncitizen. Combining these assumptions implies an additional 630,000 households in NRFU. If the NRFU data for those households have the same quality as the average NRFU data in the 2010 Census, then the result would be 139,000 fewer correct enumerations, of which 46,000 are additional erroneous enumerations and 93,000 are additional whole-person census imputations. This analysis assumes that, during the NRFU operations, a cooperative member of the household supplies data 79.0% of the time and 21.0% receive proxy responses. If all of these new NRFU cases go to proxy responses instead, the result would be 432,000 fewer correct enumerations, of which 67,000 are erroneous enumerations and 365,000 are whole-person census imputations.

For Alternative B, our estimate of the incremental cost proceeds as follows. Using the analysis in the paragraph above, the estimated NRFU workload will increase by approximately 630,000 households, or approximately 0.5 percentage points. We currently estimate that for each percentage point increase in NRFU, the cost of the 2020 Census increases by approximately $55 million. Accordingly, the addition of a question on citizenship could increase the cost of the 2020 Census by at least $27.5 million. It is worth stressing that this cost estimate is a lower bound. Our estimate of $55 million for each percentage point increase in NRFU is based on an average of three visits per household. We expect that many more of these noncitizen households would receive six NRFU visits.

We believe that $27.5 million is a conservative estimate because the other evidence cited in this report suggests that the differences between citizen and noncitizen response rates and data quality will be amplified during the 2020 Census compared to historical levels. Hence, the decrease in self-response for citizen households in 2020 could be much greater than the 5.1 percentage points we observed during the 2010 Census.

Alternative C: Use administrative data on citizenship instead of add the question to the 2020 Census

Under this alternative, we would add the capability to link an accurate, edited citizenship variable from administrative records to the final 2020 Census microdata files. We would then produce block-level tables of citizen voting age population by race and ethnicity during the publication phase of the 2020 Census using the enhanced 2020 Census microdata.

The Census Bureau has conducted tests of its ability to link administrative data to supplement the decennial census and the ACS since the 1990s. Administrative record studies were performed for the 1990, 2000 and 2010 Censuses. We discuss some of the implications of the 2010 study below. We have used administrative data extensively in the production of the economic censuses for decades. Administrative business data from multiple sources are a key component of the production Business Register, which provides the frames for the economic censuses, annual, quarterly, and monthly business surveys. Administrative business data are also directly tabulated in many of our products.

In support of the 2020 Census, we moved the administrative data linking facility for households and individuals from research to production. This means that the ability to integrate administrative data at the record level is already part of the 2020 Census production environment. In addition, we began regularly ingesting and loading administrative data from the Social Security Administration, Internal Revenue Service and other federal and state sources into the 2020 Census data systems. In assessing the expected quality and cost of Alternative C, we assume the availability of these record linkage systems and the associated administrative data during the 2020 Census production cycle.

C.1. Quality of administrate record versus self-report citizenship status

We performed a detailed study of the responses to the citizenship question compared to the administrative record citizenship variable for the 2000 Census, 2010 ACS and 2016 ACS. These analyses confirm that the vast majority of citizens, as determined by reliable federal administrative records that require proof of citizenship, correctly report their status when asked a survey question. These analyses also demonstrate that when the administrative record source indicates an individual is not a citizen, the self-report is "citizen" for no less than 23.8% of the cases, and often more than 30%.

For all of these analyses, we linked the Census Bureau's enhanced version of the SSA Numident data using the production individual record linkage system to append an administrative citizenship variable to the relevant census and ACS microdata. The Numident data contain information on every person who has ever been issued a Social Security Number or an Individual Taxpayer Identification Number. Since 1972, SSA has required proof of citizenship or legal resident alien status from applicants. We use this verified citizenship status as our administrative citizenship variable. Because noncitizens must interact with SSA if they become naturalized citizens, these data reflect current citizenship status albeit with a lag for some noncitizens.

For our analysis of the 2000 Census long-form data, we linked the 2002 version of the Census Numident data, which is the version closest to the April 1, 2000 Census date. For 92.3% of the 2000 Census long-form respondents, we successfully linked the administrative citizenship variable. The 7.7% of persons for whom the administrative data are missing is comparable to the item non-response for self-responders in the mail-in pre-ISR-option ACS. When the administrative data indicated that the 2000 Census respondent was a citizen, the self-response was citizen: 98.8%. For this same group, the long-form response was

noncitizen: 0.9% and missing: 0.3%. By contrast, when the administrative data indicated that the respondent was not a citizen, the self-report was citizen: 29.9%, noncitizen: 66.4%, and missing: 3.7%.

In the same analysis of 2000 Census data, we consider three categories of individuals: the reference person (the individual who completed the census form for the household), relatives of the reference person, and individuals unrelated to the reference person. When the administrative data show that the individual is a citizen, the reference person, relatives of the reference person, and nonrelatives of the reference person have self-reported citizenship status of 98.7%, 98.9% and 97.2%, respectively. On the other hand, when the administrative data report that the individual was a noncitizen, the long-form response was citizen for 32.9% of the reference persons; that is, reference persons who are not citizens according to the administrative data self-report that they are not citizens in only 63.3% of the long-form responses. When they are reporting for a relative who is not a citizen according to the administrative data, reference persons list that individual as a citizen in 28.6% of the long-form responses. When they are reporting for a nonrelative who is not a citizen according to the administrative data, reference persons list that individual as a citizen in 20.4% of the long-form responses.

We analyzed the 2010 and 2016 ACS citizenship responses using the same methodology. The 2010 ACS respondents were linked to the 2010 version of the Census Numident. The 2016 ACS respondents were linked to the 2016 Census Numident. In 2010, 8.5% of the respondents could not be linked, or had missing citizenship status on the administrative data. In 2016, 10.9% could not be linked or had missing administrative data. We reached the same conclusions using 2010 and 2016 ACS data with the following exceptions. When the administrative data report that the individual is a citizen, the self-response is citizen on 96.9% of the 2010 ACS questionnaires and 93.8% of the 2016 questionnaires. These lower self-reported citizenship rates are due to missing responses on the ACS, not misclassification. As we noted above, the item nonresponse rate for the citizenship question has been increasing. These item nonresponse data show that some citizens are not reporting their status on the ACS at all. In 2010 and 2016, individuals for whom the administrative data indicate noncitizen respond citizen in 32.7% and 34.7% of the ACS questionnaires, respectively. The rates of missing ACS citizenship response are also greater for individuals who are noncitizens in the administrative data (2010: 4.1%, 2016: 7.7%). The analysis of reference persons, relatives, and nonrelatives is qualitatively identical to the 2000 Census analysis.

In all three analyses, the results for racial and ethnic groups and for voting age individuals are similar to the results for the whole population with one important exception. If the administrative data indicate that the person is a citizen, the self-report is citizen at a very high rate with the remainder being predominately missing self-reports for all groups. If the administrative data indicate noncitizen, the self-report is citizen at a very high rate (never less than 23.8% for any racial, ethnic or voting age group in any year we studied). The exception is the missing data rate for Hispanics, who are missing administrative data about twice as often as non-Hispanic blacks and three times as often as non-Hispanic whites.

C.2. Analysis of coverage differences between administrative and survey citizenship data

Our analysis suggests that the ACS and 2000 long form survey data have more complete coverage of citizenship than administrative record data, but the relative advantage of the survey data is diminishing. Citizenship status is missing for 10.9 percent of persons in the 2016 administrative records, and it is missing for 6.3 percent of persons in the 2016 ACS. This 4.6 percentage point gap between administrative and survey missing data rates is smaller than the gap in 2000 (6.9 percentage points) and 2010 (5.6

percentage points). Incomplete (through November) pre-production ACS data indicate that citizenship item nonresponse has again increased in 2017.

There is an important caveat to the conclusion that survey-based citizenship data are more complete than administrative records, albeit less so now than in 2000. The methods used to adjust the ACS weights for survey nonresponse and to allocate citizenship status for item nonresponse assume that the predicted answers of the sampled non-respondents are statistically the same as those of respondents. Our analysis casts serious doubt on this assumption, suggesting that those who do not respond to either the entire ACS or the citizenship question on the ACS are not statistically similar to those who do; in particular, their responses to the citizenship question would not be well-predicted by the answers of those who did respond.

The consequences of missing citizenship data in the administrative records are asymmetric. In the Census Numident, citizenship data may be missing for older citizens who obtained SSNs before the 1972 requirement to verify citizenship, naturalized citizens who have not confirmed their naturalization to SSA, and noncitizens who do not have an SSN or ITIN. All three of these shortcomings are addressed by adding data from the United States Citizen and Immigration Services (USCIS). Those data would complement the Census Numident data for older citizens and update those data for naturalized citizens. A less obvious, but equally important benefit, is that they would permit record linkage for legal resident aliens by allowing the construction of a supplementary record linkage master list for such people, who are only in scope for the Numident if they apply for and receive an SSN or ITIN. Consequently, the administrative records citizenship data would most likely have both more accurate citizen status and fewer missing individuals than would be the case for any survey-based collection method. Finally, having two sources of administrative citizenship data permits a detailed verification of the accuracy of those sources as well.

C.3. Cost of administrative record data production

For Alternative C, we estimate that the incremental cost, except for new MOUs, is $450,000. This cost estimate includes the time to develop an MOU with USCIS, estimated ingestion and curation costs for USCIS data, incremental costs of other administrative data already in use in the 2020 Census but for which continued acquisition is now a requirement, and staff time to do the required statistical work for integration of the administrative-data citizenship status onto the 2020 Census microdata. This cost estimate is necessarily incomplete because we have not had adequate time to develop a draft MOU with USCIS, which is a requirement for getting a firm delivery cost estimate from the agency. Acquisition costs for other administrative data acquired or proposed for the 2020 Census varied from zero to $1.5M. Thus the realistic range of cost estimates, including the cost of USCIS data, is between $500,000 and $2.0M

Questions on the Jan 19 Draft Census Memo on the DoJ Citizenship Question Reinstatement Request

1. **With respect to Alternatives B and C, what is the difference, if any, between the time when the data collected under each alternative would be available to the public?**

   Since the collection of this data, whether from administrative records or from an enumerated question, occurs prior to the creation of the Microdata Detail File (MDF) from which all tabulations will be performed, there is no difference in the timing of when the data collected under either alternative B or C could be made available to the public. The exact date for completion of the MDF is still being determined as the 2020 Census schedule is matured. However, the 2020 Census is working towards publishing the first post-apportionment tabulation data products as early as the first week of February 2021.

2. **What is the "2020 Census publication phase" (page 1 of the Detailed Analysis for Alternative B) versus Alternative C? Would there be any difference?**

   The 2020 Census publication phase is a broad window stretching from the release of the apportionment counts by December 31, 2020 through the last data product or report published in FY 2023, the final year of decennial funding for the 2020 Census. However, as stated in the answer to question 1, these data could be made available to the public on the same schedule as any other post-apportionment tabulated data product regardless of whether alternative B or C is used in its collection.

3. **What is the non-response rate for: (A) each question on the 2000 and 2010 Decennial Census short form and (B) each question on the 2010 ACS and most recent ACS?**

   The table below shows the item non-response (INR) rate for each question on the 2000 and 2010 Decennial Census short form. This is the percentage of respondents who did not provide an answer to an item.

   Item Nonresponse Rates for 2000 and 2010 Short Form Person Questions

	Relationship	Sex	Age	Hispanic Origin	Race	Tenure
2010	1.5	1.5	3.5	3.9	3.3	4.5
2000	1.3	1.1	3.7	3.1	2.9	4.1

   Source: Rothhaas, Lestina and Hill (2012) Tables

   Notes and Soucre:
   Rothhaas, C., Lestina, F. and Hill, J. (2012) "2010 Decennial Census Item Nonresponse and Imputation Assessment Report" 2010 Census Program for Evaluations and Experiments, January 24, 2012.

From report:

The INR rate is essentially the proportion of missing responses before pre-editing or imputation procedures for a given item (i.e., the respondent did not provide an answer to the item). For INR, missing values are included in the rates, but inconsistent responses (i.e., incompatible with other responses) are considered non-missing responses.

Online link to 2010 report that has 2000 information as well.
https://www.census.gov/2010census/pdf/2010_Census_INR_Imputation_Assessment.pdf

See attached spreadsheet for the item allocation rates by questions for the ACS for 2010, 2013, and 2016.

4. **What was the total survey response rate (i.e., percentage of complete questionnaires) for the 2000 long form and the 2000 short form? Of the incomplete long forms, what percentage left the citizenship question blank? Of the completed long forms, what percentage (if known) contained incorrect responses to the citizenship question?**

We do not have measures of total survey response rates from the 2000 long form and 2000 short form available at this time. The mail response rate in 2000 was 66.4 percent for short forms and 53.9 percent for long forms. No analysis that we were aware of was conducted on the incomplete long forms that left the citizenship question blank. The Census 2000 Content Reinterview Survey showed low inconsistency of the responses to the citizenship question. Only 1.8 percent of the respondents changed answers in the reinterview.

Source for 2000 mail response rates:
https://www.census.gov/pred/www/rpts/A.7.a.pdf

Source for 2000 Content Reinterview Survey. Page 32 source.
https://www.census.gov/pred/www/rpts/B.5FR_RI.PDF

5. **For the 2000 long and short forms, what was the percentage unanswered (left blank) for each question (i.e., what percentage of the responses for each question (sex, race, ethnicity, income, citizenship, etc.) were left blank)?**

For the 2000 shortform, the table in question 3a provides the percentage unanswered for each question.

For the 2000 longform, Griffin, Love and Obenski (2003) summarized the Census 2000 longform responses. Allocation rates for individual items in Census 2000 were computed, but because of the magnitude of these data, summary allocation measures were derived.

These rates summarize completeness across all data items for occupied units (households) and are the ratio of all population and housing items that had values allocated to the total number of population and housing items required to have a response. These composite measures provide a summary picture of the completeness of all data. Fifty-four population items and 29 housing items are included in these summary measures. The analysis showed that 9.9 percent of the population question items and 12.5 percent of the housing unit question items required allocation. Allocation involves using statistical procedures, such as within-household or nearest neighbor matrices, to impute missing values.

https://ww2.amstat.org/sections/srms/Proceedings/y2003/Files/JSM2003-000596.pdf

6. **What was the incorrect response rate for the citizenship question that was asked on the Long Form during the 2000 Decennial Census? Does the response rate on the 2000 Long Form differ from the incorrect response rate on the citizenship question for the ACS?**

   In the 2000 long form, 2.3 percent of persons have inconsistent answers, 89.4 percent have consistent answers, and 8.2 percent have missing citizenship data in the SSA Numident and/or the 2000 long form. Among persons with nonmissing citizenship data in the SSA Numident and/or the 2000 long form, 2.6 percent have inconsistent answers and 97.4 percent have consistent answers.

   In the 2010 ACS, 3.1 percent of persons have inconsistent answers, 86.0 percent have consistent answers, and 10.8 percent have missing citizenship data in the SSA Numident and/or the 2010 ACS. Among persons with nonmissing citizenship data in the SSA Numident and/or the 2010 ACS, 3.6 percent have inconsistent answers and 96.4 percent have consistent answers.

   In the 2016 ACS, 2.9 percent of persons have inconsistent answers, 81.2 percent have consistent answers, and 15.9 percent have missing citizenship data in the SSA Numident and/or the 2016 ACS. Among persons with nonmissing citizenship data in the SSA Numident and/or the 2016 ACS, 3.5 percent have inconsistent answers and 96.5 percent have consistent answers.

   These ACS and 2000 Census long form rates are based on weighted data.

   This shows that inconsistent response rates are higher in the 2010 and 2016 ACS than in the 2000 long form.

7. **What is the incorrect response rate on other Decennial or ACS questions for which Census has administrative records available (for example, age, sex or income)?**

   Table 7a shows the agreement rates between the 2010 Census response and the SSA Numident for persons who could be linked and had nonmissing values, and Table 7b shows

the agreement rates between the 2010 ACS and the SSA Numident. Gender has low disagreement (0.4-0.5 percent), and white alone (0.9 percent), black alone (1.7-2 percent), and age (2.1 percent) also have low disagreement rates. Disagreement rates are greater for other races (e.g., 46.4-48.6 percent for American Indian or Alaska Native alone). Hispanic origin is not well measured in the Numident, because it contains a single race response, one of which is Hispanic.

Table 7a. Demographic Variable Agreement Rates Between the 2010 Census and the SSA Numident

2010 Census Response	Percent Agreement with SSA Numident
Hispanic	54.2
Not Hispanic	99.7
White Alone	99.1
Black Alone	98.3
American Indian or Alaska Native Alone	51.4
Asian Alone	84.3
Native Hawaiian or Other Pacific Islander Alone	74.4
Some Other Race Alone	17.7
Age	97.9
Gender	99.4

Source: Rastogi, Sonya, and Amy O'Hara, 2012, "2010 Census Match Study," 2010 Census Planning Memoranda Series No. 247.

Table 7b. Demographic Variable Agreement Rates Between the 2010 Census and the SSA Numident

2010 ACS Response	Percent Agreement with SSA Numident
White Alone	99.1
Black Alone	98.0
American Indian or Alaska Native Alone	53.6
Asian Alone	82.9
Native Hawaiian or Other Pacific Islander Alone	72.9
Some Other Race Alone	17.2
Age 0-2 Date of Birth	95.2
Age 3-17 Date of Birth	95.6
Age 18-24 Date of Birth	95.2
Age 25-44 Date of Birth	95.8
Age 45-64 Date of Birth	95.9
Age 65-74 Date of Birth	96.5
Age 75 and older Date of Birth	92.7
Male	99.5
Female	99.5

Source: Bhaskar, Renuka, Adela Luque, Sonya Rastogi, and James Noon, 2014, "Coverage and Agreement of Administrative Records and 2010 American Community Survey Demographic Data," CARRA Working Paper #2014-14.

Abowd and Stinson (2013) find correlations of 0.75-0.89 between Survey of Income and Program Participation (SIPP) and SSA Detailed Earnings Record annual earnings between 1990-1999.[1]

8. How does the Census presently handle responses on the (A) Decennial Census and (B) the ACS when administrative records available to the Census confirm that the response on the Decennial Census or ACS is incorrect? Is the present Census approach to incorrect responses based on practice/policy or law (statute or regulation)?

We have always based the short form Decennial Census and the ACS on self-response, and while we have procedures in place to address duplicate or fraudulent responses, we do not check the accuracy of the answers provided to the specific questions on the Census questionnaire. This is a long established practice at the Census Bureau that has been thoroughly tested and in place since 1970, when the Census Bureau moved to a mail-out/respond approach to the Decennial Census. Title 13 of the U.S. Code allows the Census Bureau to use alternative data sources, like administrative records, for a variety of purposes, and we are using data in new ways in the 2020 Census. While this includes the use of administrative records data to fill in areas where a respondent does not provide an answer, we have not explored the possibility of checking or changing responses that a responding household has provided in response to the questionnaire.

9. Please explain the differences between the self-response rate analysis and the breakoff rate analysis. The range of breakoff rates between groups was far smaller than the range of self-response rates between groups.

Self-response means that a household responded to the survey by mailing back a questionnaire or by internet, and a sufficient number of core questions were answered so that an additional field interview was not required.

A breakoff occurs when an internet respondent stops answering questions prior to the end of the questionnaire. In most cases the respondent answers the core questions before breaking off, and additional fieldwork is not required. The breakoff rates are calculated separately by which question screen was the last one reached before the respondent stopped answering altogether.

The share of Hispanic respondents who broke off at some point before the end of the questionnaire (17.6 percent) is much higher than for non-Hispanic whites (9.5 percent).

[1] Abowd, John M., and Martha H. Stinson, 2013, "Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data," Review of Economics and Statistics, Vol. 95(55), pp. 1451-1467.

Spreading the overall breakoff rates over 134 screens in the questionnaire works out to quite small rates per screen. It works out to an average breakoff rate of 0.131 percent per screen for Hispanics and 0.066 percent for non-Hispanic whites.

10. The NRFU numbers are comparatively small – approximately one additional household for NRFU per Census enumerator. Is this really a significant source of concern?

Yes, this is a significant concern. First, it gives rise to incremental NRFU cost of at least $27.5 million. This is a lower bound becaues it assumes the households that do not self-respond because we added a question on citizenship have the same follow-up costs as an average U.S. household. They won't because these households overwhelmingly contain at least one noncitzen, and that is one of our acknowledged hard-to-count subpopulations.

11. Given that the breakoff rate difference was approximately 1 percent, why did Census choose to use the 5.1 percent number for assessing the cost of Alternative B?

If a household breaks off an internet response at the citizenship, place of birth, or year of entry screens, this means it would have already responded to the core questions. This would not trigger follow-up fieldwork and thus would not involve additional fieldwork costs. In contrast, if a household does not mail back a questionnaire or give an internet response, fieldwork will be necessary and additional costs will be incurred. Thus, the 5.1 percent number for differential self-response is more appropriate for estimating the additional fieldwork cost of adding a citizenship question.

12. Alternative C states that Census would use administrative data from the Social Security Administration, Internal Revenue Service, and "other federal and state sources." What are the other sources?

In addition to continuing the acquisition of the Social Security Administration and Internal Revenue Service data, the Census Bureau is in discussion with the U.S. Citizen and Immigration Services (USCIS) staff to acquire additional citizenship data.

13. Is Census confident that administrative data will be able to be used to determine citizenship for all persons (e.g., not all citizens have social security numbers)?

We are confident that Alternative C is viable and that we have already ingested enough high-quality citizenship administrative data from SSA and IRS. The USCIS data are not required. They would, however, make the citizenship voting age tabulations better, but the administrative data we've got are very good and better than the data from the 2000 Census and current ACS. The type of activities required for Alternative C already occur daily and routinely at the Census Bureau. We have been doing this for business data products,

including the Economic Censuses, for decades. We designed the 2020 Census to use this technology too.

14. For Alternative C, the memo says, "we assume the availability of these record linkage systems and associated administrative data" – does Census already have in place access to this data or would this need to be negotiated? If negotiated, for which data sets specifically?

The Census Bureau has longstanding contractual relationships with the Social Security Administration and the Internal Revenue Service that authorize the use of data for this project. For new data acquired for this project (i.e., USCIS) we would estimate a six-month development period to put a data acquisition agreement in place. That agreement would also include terms specifying the authorized use of data for this project.

15. Are there any privacy issues / sensitive information prohibitions that might prevent other agencies from providing such data?

There are no new privacy or sensitivity issues associated with other agencies providing citizenship data. We have received such information in the past from USCIS. We are currently authorized to receive and use the data from SSA and IRS that are discussed in Alternative C.

16. How long would Census expect any negotiation for access to data take? How likely is it that negotiations would be successful? Are MOA's needed/required?

Current data available to the Census Bureau provide the quality and authority to use that are required to support this project. Additional information potentially available from USCIS would serve to supplement/validate those existing data. We are in early discussions with USCIS to develop a data acquisition agreement and at this time have no indications that this acquisition would not be successful.

17. What limitations would exist in working with other agencies like IRS, Homeland Security, etc. to share data?

The context for sharing of data for this project is for a one-way sharing of data from these agencies to the Census Bureau. Secure file transfer protocols are in-place to ingest these data into our Title 13 protected systems. For those data already in-place at the Census Bureau to support this project, provisions for sharing included in the interagency agreement restrict the Census Bureau from sharing person-level microdata outside the Census Bureau's Title 13 protections. Aggregates that have been processed through the Bureau's disclosure avoidance procedures can be released for public use.

18. If Alternative C is selected, what is Census's backup plan if the administrative data cannot be completely collected and utilized as proposed?

The backup plan is to use all of the administrative data that we currently have, which is the same set that the analyses of Alternative C used. We have verified that this use is consistent with the existing MOUs. We would then use estimation and modeling techniques similar to those used for the Small Area Income and Poverty Estimates (SAIPE) to impute missing citizenship status for those persons for whom we do not have administrative records. These models would also include estimates of naturalizations that occurred since the administrative data were ingested.

19. Does Census have any reason to believe that access to existing data sets would be curtailed if Alternative C is pursued?

No we do not believe that any access to existing data sets would be curtailed if we pursue Alternative C.

20. Has the proposed Alternative C approach ever been tried before on other data collection projects, or is this an experimental approach? If this has been done before, what was the result and what were lessons learned?

The approach in Alternative C has been routinely used in processing the economic censuses for several decades. The Bureau's Business Register was specifically redesigned for the 2002 Economic Census in order to enhance the ingestion and use of administrative records from the IRS and other sources. The data in these administrative records are used to substitute for direct responses in the economic censuses for the unsampled entities. They are also used as part of the review, edit, and imputation systems for economic censuses and surveys. On the household side, the approach in Alternative C was used extensively to build the residential characteristics for OnTheMap and OnTheMap for Emergency Management.

21. Is using sample data and administrative records sufficient for DOJ's request?

The 2020 Census data combined with Alternative C are sufficient to meet DoJ's request. We do not anticipate using any ACS data under Alternative C.

22. Under Alternative C, If Census is able to secure interagency agreements to provide needed data sets, do we know how long it would take to receive the data transmission from other agencies and the length of time to integrate all that data, or is that unknown?

With the exception of the USCIS data, the data used for this project are already integrated into the 2020 Census production schema. In mid-to late 2018, we plan to acquire the USCIS data and with those data and our existing data begin to develop models and business rules to select citizenship status from the composite of sources and attach that characteristic to

each U.S. person. We expect the development and refinement of this process to continue into 2019 and to be completed by third quarter calendar year 2019.

23. Cross referencing Census decennial responses with numerous governmental data sets stored in various databases with differing formats and storage qualities sounds like it could be complicated. Does Census have an algorithm in place to efficiently combine and cross reference such large quantities of data coming from many different sources? What cost is associated with Alternative C, and what technology/plan does Census have in place to execute?

Yes, the 2018 Census End-to-End test will be implementing processing steps to be able to match Census responses to administrative record information from numerous governmental data sets. The Census Bureau has in place the Person Identification Validation System to assign Protected Identification Keys to 2020 Census responses. The required technology for linking in the administrative records is therefore part of the 2020 Census technology. This incremental cost factored into the estimate for Alternative C is for integrating the citizenship variable specifically, since that variable is not currently part of the 2020 Census design. No changes are required to the production Person Identification Validation system to integrate the administrative citizenship data.

24. For section C-1 of the memo, when did Census do the analyses of the incorrect response rates for non-citizen answers to the long form and ACS citizenship question? Were any of the analyses published?

The comparisons of ACS, 2000 Decennial Census longform and SSA Numident citizenship were conducted in January 2018. This analysis has not been published.

25. Has Census corrected the incorrect responses it found when examining non-citizen responses? If not, why not?

In the American Community Survey (ACS), and the short form Decennial Census, we do not change self-reported answers. The Decennial Census and the ACS are based on self-response and we accept the responses provided by households as they are given. While we have procedures in place to address duplicate or fraudulent responses, we do not check the accuracy of the answers provided to the specific questions on the Census questionnaires. This is a long established process at the Census Bureau that has been thoroughly tested and in place since 1970, when the Census Bureau moved to a mail-out/respond approach to the Decennial Census.

26. Has the Department of Justice ever been made aware of inaccurate reporting of ACS data on citizenship, so that they may take this into consideration when using the data?

Not exactly. The Census Bureau is in close, regular contact with the Department of Justice (DOJ) regarding their data requirements. Our counterparts at DOJ have a solid understanding of survey methodology and the quality of survey data, and they are aware of the public documentation on sampling and accuracy surrounding the ACS. However, the specific rate of accuracy regarding responses to the ACS question on citizenship has never been discussed.

27. Why has the number of persons who cannot be linked increased from 2010 to 2016?

The linkage between the ACS and administrative data from the SSA Numident and IRS ITIN tax filings depends on two factors: (a) the quality of the personally identifiable information (PII) on the ACS response and (b) whether the ACS respondent is in the SSN/ITIN universe.

With respect to the quality of the PII on the ACS, there may be insufficient information on the ACS due to item nonresponse or proxy response for the person to allow a successful match using the production record linkage system. There may also be more than one record in the Numident or ITIN IRS tax filings that matches the person's PII. Finally, there may be a discrepancy between the PII provided to the ACS and the PII in the administrative records.

Alternatively, the person may not be in the Numident or ITIN IRS tax filing databases because they are out of the universe for those administrative systems. This happens when the person is a citizen without an SSN, or when the person is a noncitizen who has not obtained an SSN or ITIN.

Very few of the unlinked cases are due to insufficient PII in the ACS or multiple matches with administrative records. The vast majority of unlinked ACS persons have sufficient PII, but fail to match any administrative records sufficiently closely. This means that most of the nonmatches are because the ACS respondent is not in the administrative record universe.

The incidence of ACS persons with sufficient PII but no match with administrative records increased between 2010 and 2016. One contributing factor is that the number of persons linked to ITIN IRS tax filings in 2016 was only 39 percent as large as in 2010, suggesting that either fewer of the noncitizens in the 2016 ACS had ITINs, or more of them provided PII in the ACS that was inconsistent with their PII in IRS records.

28. Independent of this memo, what action does Census plan to take in response to the analyses showing that non-citizens have been incorrectly responding to the citizenship question?

The Census Bureau does not have plans to make any changes to procedures in the ACS. However, we will continue to conduct thorough evaluations and review of census and survey data. The ACS is focusing our research on the potential use of administrative records

in the survey. For instance, we are exploring whether we can use IRS data on income to reduce the burden of asking questions on income on the ACS. We are concentrating initially on questions that are high burden, e.g., questions that are difficult to answer or questions that are seen as intrusive.

29. Did Census make recommendations the last time a question was added?

Since the short form Decennial Census was established in 2010, the only requests for new questions we have received have been for the ACS. And, in fact, requests for questions prior to 2010 were usually related to the Decennial Census Long Form. We always work collaboratively with Federal agencies that request a new question or a change to a question. The first step is to review the data needs and the legal justification for the new question or requested changes. If, through this process, we determine that the request is justified, we work with the other agencies to test the question (cognitive testing and field testing). We also work collaboratively on the analysis of the results from the test which inform the final recommendation about whether or not to make changes or add the question.

30. Does not answering truthfully have a separate data standard than not participating at all?

We're not sure what you're asking here. Please clarify the question.

31. What was the process that was used in the past to get questions added to the decennial Census or do we have something similar where a precedent was established?

Because no new questions have been added to the Decennial Census (for nearly 20 years), the Census Bureau did not feed bound by past precedent when considering the Department of Justices' request. Rather, the Census Bureau is working with all relevant stakeholders to ensure that legal and regulatory requirements are filled and that questions will produce quality, useful information for the nation. As you are aware, that process is ongoing at your direction.

32. Has another agency ever requested that a question be asked of the entire population in order to get block or individual level data?

Not to our knowledge. However, it is worth pointing out that prior to 1980 the short form of the Decennial Census included more than just the 10 questions that have been on the short form since 1990.

33. Would Census linking of its internal data sets, with other data sets from places like IRS and Homeland Security, have an impact on participation as well (i.e., privacy concerns)?

The potential that concerns about the use of administrative records could have an impact on participation has always been a concern of ours, and it's a risk that we're managing on our risk register. We've worked closely with the privacy community throughout the decade, and we established a working group on our National Advisory Committee to explore this issue. We've also regularly briefed the Congress about our plans. At this stage in the decade there does not appear to be extensive concerns among the general public about our approach to using administrative records in the Nonresponse Operation or otherwise. We will continue to monitor this issue.

34. Would Alternative C require any legislation? If so, what is the estimated time frame for approval of such legislation?

No.

35. Census publications and old decennial surveys available on the Census website show that citizenship questions were frequently asked of the entire population in the past. Citizenship is also a question on the ACS. What was the justification provided for citizenship questions on the (A) short form, (B) long form, and (C) ACS?

In 1940, the Census Bureau introduced the use of a short form to collect basic characteristics from all respondents, and a long form to collect more detailed questions from only a sample of respondents. Prior to 1940, census questions were asked of everyone, though in some cases only for those with certain characteristics. For example, in 1870, a citizenship question was asked, but only for respondents who were male and over the age of 21.

Beginning in 2005, all the long-form questions – including a question on citizenship -- were moved to the ACS. 2010 was the first time we conducted a short-form only census. The citizenship question is included in the ACS to fulfill the data requirements of the Department of Justice, as well as many other agencies including the Equal Employment Opportunities Commission, the Department of Health and Human Services, and the Social Security Administration.

ACS Item Allocation Rates for United States: 2016, 2013, 2010

Title	2016	2013	2010
Overall housing allocation rate occupied and vacant housing units	4.9	5.6	5.2
Overall person allocation rate total population	9.5	8.4	5.8
Vacancy status vacant housing units	3.9	3.5	2.9
Tenure occupied housing units	1.2	1.3	1.2
Units in structure occupied and vacant housing units	1.5	1.5	1.5
Year moved in occupied housing units	3	3	3.4
Month moved in occupied housing units into which households move in the last two years	0.7	0.7	0.7
Year built occupied and vacant housing units	18.2	17.1	16.2
Lot size occupied and vacant single family and mobile homes	3.9	3.9	4.2
Agricultural sales occupied and vacant single family and mobile homes with lot size greater than or equal to 1 acre	4	4.2	4.4
Business on property occupied and vacant single family and mobile homes	**	2.4	3
Number of rooms occupied and vacant housing units	5	5.5	5.2
Number of bedrooms occupied and vacant housing units	5.5	4.6	4.3
Running water occupied and vacant housing units	2.4	2.1	2
Flush toilet occupied and vacant housing units	**	2.2	2
Bathtub or shower occupied and vacant housing units	2.6	2.2	2
Sink with a faucet occupied and vacant housing units	2.6	2.2	2
Stove or range occupied and vacant housing units	3.1	2.8	2.5
Refrigerator occupied and vacant housing units	3.2	2.9	2.7
Telephone occupied housing units	1.5	1.2	1.1
Number of vehicles occupied housing units	1.2	1.4	1.3
Heating fuel occupied housing units	3.4	3.4	3.3
Monthly electricity cost occupied housing units	8.1	8.2	7.3
Monthly gas cost occupied housing units	9.6	9.9	9.8

Yearly water and sewer cost occupied housing units	8.5	8.8	8.1
Yearly other fuel cost occupied housing units	7.3	8.3	10.6
Yearly food stamp recipiency - household occupied housing units	1.7	1.7	1.3
Yearly real estate taxes owner-occupied housing units	16.7	18.5	16.3
Yearly property insurance owner-occupied housing units	23.9	25.6	23.2
Mortgage status owner-occupied housing units	2.2	2.5	2.1
Monthly mortgage payment owner-occupied housing units with a mortgage	10.5	12.4	10.7
Mortgage payment incl. real estate taxes owner-occupied housing units with a mortgage	6.2	6.9	(X)
Mortgage payment incl. insurance owner-occupied housing units with a mortgage	6.8	7.4	(X)
Second mortgage owner-occupied housing units	3.2	3.7	3.4
Home equity loan owner-occupied housing units	3.7	4.3	4.2
Other monthly mortgage payment(s) owner-occupied housing units with second mortgage or home equity loan	23.3	21.7	17.9
Property value owner-occupied housing units and vacant housing units for sale	11.6	12.9	12.3
Yearly mobile home costs occupied mobile homes and other units	21.7	21.5	19.9
Monthly condominium fee owner-occupied housing units	0.8	0.8	0.7
Monthly rent occupied housing units rented for cash rent and vacant housing units for rent	10.5	9.8	9.3
Meals included in rent occupied housing units rented for cash rent and vacant housing units for rent	2.1	2.1	2
Desktop/laptop/notebook computer occupied housing units	1.3	3.2	**
Handheld computer/smart mobile phone occupied housing units	**	3.3	**
Tablet or other portable wireless computer occupied housing units	1.6	**	**
Smartphone occupied housing units	1.6	**	**
Other computer occupied housing units	1.7	3.7	**
Household has internet access occupied housing units	3.3	4.4	**
Dial-up internet service occupied housing units with internet access	3.8	5.7	**
DSL internet service occupied housing units with internet access	**	5.7	**

Cable modem internet service occupied housing units with internet access	**	5.7	**
Fiber-optic internet service occupied housing units with internet access	**	5.7	**
Cellular data plan (formerly mobile broadband) occupied housing units with internet access	7.6	26.7	**
Satellite internet service occupied housing units with internet access	3.8	5.7	**
High speed internet service occupied housing units with internet accesss	3.8	**	**
Some other internet service occupied housing units with internet access	3.8	5.7	**
Race total population	1.5	1.6	1.5
Hispanic origin total population	1.8	2.1	1.8
Sex total population	0.1	0.1	0.1
Age total population	1.7	1.6	1.3
Relationship total household population	1.2	1.1	1.2
Marital status total population 15 years and over	5.3	4.8	3
Married past 12 months total population 15 years and over, except those never married	6.9	6.6	4.7
Widowed past 12 months total population 15 years and over, except those never married	7.4	7	4.5
Divorced past 12 months total population 15 years and over, except those never married	7.4	7	4.5
Times married total population 15 years and over, except those never married	8.1	7.8	5.1
Year last married total population 15 years and over, except those never married	13.5	13.3	11.4
Place of birth total population	9.1	8.6	6.5
Citizenship total population	6	5.2	2.7
Year of naturalization total population naturalized citizens	22.5	22.5	16.6
Year of entry total population not born in US	14.8	13.2	10.3
Speaks another language at home total population 5 years and over	6.8	5.9	3.4
Language spoken total population 5 years and over who speak another language at home	8.3	7	5.7
English ability total population 5 years and over who speak another language at home	7.1	5.9	4
School enrollment total population 3 years and over	6.7	6	3.7
Grade level attending			

total population 3 years and over enrolled	10.2	8.9	6
Educational attainment total population 3 years and over	8.5	8	5.6
Field of degree total population 25 years and over with a bachelor's degree or higher	13.5	12.4	9.8
Mobility status total population 1 years and over	7.2	6.5	4
Migration state/foreign county total population 1 years and over movers	13.2	11.3	7.1
Migration county total population 1 years and over movers within US	14.6	12.5	8.3
Migration minor civil division total population 1 years and over movers within US	14.2	12.1	8.4
Migration place total population 1 years and over movers within US	15	12.9	8.8
Health insurance thru employer/union total population	10.7	9	6.2
Health insurance purchased directly total population	11.3	9.7	6.9
Health insurance through Medicare total population	9.5	8.1	5.2
Health insurance through Medicaid total population	12.2	10.5	7.9
Health insurance through TRICARE total population	12.5	10.8	8.1
Health insurance through VA total population	12.3	10.7	8.1
Health ins. thru Indian Health Service total population	12.8	11.1	8.5
Visual difficulty total population	7.1	6.1	3.4
Hearing difficulty total population	6.8	5.9	3.2
Physical difficulty total population 5 years and over	7.5	6.7	3.5
Difficulty remembering total population 5 years and over	7.5	6.7	3.5
Difficulty dressing total population 5 years and over	7.5	6.7	3.5
Difficulty going out total population 16 years and over	7.3	6.5	3.4
Grandchildren living in home noninstitutionalized population 30 years and over	1.1	1	0.9
Responsibility for grandchildren noninstitutionalized population 30 years and over who are grandparents with grandchildren in the home	17.7	15.7	12
Months responsible for grandchildren noninstitutionalized population 30 years and over who are grandparents with grandchildren in the home that have responsibility	17.2	16.1	14.9
Fertility status female total population 15-50	7.8	6.7	3.7

Veteran status total population 17 years and over	7.3	6.8	3.8
Periods of military service total population 17 years and over on active duty now or previously	9.7	9.3	6.3
Service-connected disability rating total population 17 years and over, except those who never served in the Armed Forces	6.8	6.6	3.9
Service-connected disability rating value total population 17 years and over with a service-connected disability	0.2	0.2	0.7
Employment status recode noninstitutionalized population 16 years and over	8.7	8.1	5.1
When last worked noninstitutionalized population 16 years and over	9.6	9.1	5.7
Weeks worked in the past 12 months noninstitutionalized population 16 years and over who worked in the past 12 months	10.6	9.7	6.9
Hours worked per week noninstitutionalized population 16 years and over who worked in the past 12 months	11.9	10.8	7.7
Place of work state/foreign county noninstitutionalized population 16 years and over at work last week	11.8	10.4	6.3
Place of work county noninstitutionalized population 16 years and over at work last week	12.5	11	7
Place of work minor civil division noninstitutionalized population 16 years and over at work last week	3.6	3.3	2.1
Place of work place noninstitutionalized population 16 years and over at work last week	13.1	11.6	7.6
Transportation to work noninstitutionalized population 16 years and over at work last week	9.6	8.8	5.7
Carpool size noninstitutionalized population 16 years and over at work last week who drive to work	10.9	9.9	6.8
Time of departure noninstitutionalized population 16 years and over at work last week who don't work at home	20.2	18.5	12.8
Commuting time noninstitutionalized population 16 years and over at work last week who don't work at home	14.5	13.3	9.7
Class of worker total population 16 years and over who worked in the last 5 years	11.7	10.7	7.2
Industry total population 16 years and over who worked in the last 5 years	12.7	11.4	7.8
Occupation total population 16 years and over who worked in the last 5 years	13.4	11.8	8.1
Wages/salary income total population 15 years and over	19.1	19	16
Self-employment income			

total population 15 years and over	10.5	9.3	5.9
Interest, dividends, etc. income total population 15 years and over	15.2	12.6	8.8
Social security or railroad retirement total population 15 years and over	14.5	12.3	8.9
Supplemental security income total population 15 years and over	12.7	10.3	6.7
Public assistance total population 15 years and over	13.2	10.5	6.8
Retirement income total population 15 years and over	13.6	11.1	7.5
Other income total population 15 years and over	13.2	10.8	7.4
Some or all income allocated total population 15 years and over	28.4	25.3	22.4

Source: ACS 1-year data. See following links for more information:

https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/item-allocation-rates/

https://www.census.gov/programs-surveys/acs/methodology/sample-size-and-data-quality/item-allocation-rates-definitions.html

Note:

** This item was not asked in this year.

Summary Analysis of the Key Differences Between Alternative C and Alternative D

This short note describes the Census Bureau's current assumptions about two alternatives to address the need for block level data on citizen voting age populations. The goal is to measure the citizenship status of all people enumerated in the 2020 Decennial Census. Both alternatives utilize administrative data on the citizenship status of individuals, however one option, Alternative D, proposes to also include the current American Community Survey (ACS) question on citizenship status on the 2020 Decennial Census short form.

In both alternatives described here, the methodology requires linking 2020 census response data and administrative records. However, as illustrated both alternatives would also need to assign/impute citizenship for a portion of the population. The Census Bureau will have to assign citizenship in cases of questionnaire non-response and item non-response. Additionally, it is important to note, that even when a self-response is available it is not always possible to link response data with administrative records data. Poor data quality (e.g., name and age) and nonresponse or incomplete 2020 Census responses mean that we will not have a direct measure of citizenship status for all residents enumerated in 2020. The Census Bureau will to need employ an imputation model for these cases.

One of the key differences between to the two alternatives described below is the number of cases requiring imputation. The other key difference is the impact of errors in the citizenship status reported on the 2020 Census.

In the most recent version of the 2020 Decennial Life Cycle Cost Estimate, the Census Bureau projects counting 330 million residents in 2020. Figure 1 summarizes how citizenship status will be measured under Alternative C that does not employ a citizenship question on the 2020 Census. Figure 2 summarizes how this will be done using both administrative records and a 2020 citizenship question under Alternative D.

Alternative C is a simplified process for assigning citizenship through direct linkage and modelling, without including the question on the 2020 Census. The Census Bureau will link the responses for the 330 million census records to administrative records that contain information on the citizenship status of individuals. The Census Bureau expects to successfully link and observe this status for approximately 295 million people. The Census Bureau would need to impute this status for approximately 35 million people under Alternative C whose 2020 responses cannot be linked to administrative data. Although the Census Bureau has fully developed and tested the imputation model, it has high confidence that an accurate model can be developed and deployed for this purpose. Further, we will most likely never possess a fully adequate truth deck to benchmark it to.

Measuring citizenship status is slightly more complex under Alternative D where all U.S. households will be given the opportunity to provide the citizenship status of each household member. Based on response data for the ACS citizenship and other response data research, we know that not all households that respond to the 2020 Census will answer this question, leaving the question blank or with otherwise invalid responses. Additionally, Alternative D, must also account for those households that do not respond at all or will have proxy responses. Due to these reasons, we estimate that we will get 2020 citizenship status responses for approximately 294.6 million people, a slightly higher estimate

than Alternative C. For the 35.4 million people without a 2020 citizenship response, the Census Bureau will employ the same methodology as in Alternative C, linking the 2020 Census responses to the administrative records. The Census Bureau estimates that it will be able to link these cases to administrative records where we observe citizenship status for approximately 21.5 million people. For the remaining 13.8 million will be imputed through a model as described above. Thus, there will be a need for imputing many cases across either alternative.

The Census Bureau will link the 294.6 million records from the 2020 Census with the administrative records. This will be done both for potential quality assurance purposes and to improve the quality of future modeling uses. Based on the current research from the ACS, the Census Bureau expects to successfully link approximately 272.5 million of these cases. Of these, 263 million will have citizenship statuses that agree across the 2020 response and administrative record. The Census Bureau estimates there will be 9.5 million cases where there is disagreement across the two sources. Historic Census Bureau practice is to use self-reported data in these situations. However, the Census Bureau now knows from linking ACS responses on citizenship to administrative data that nearly one third of noncitizens in the administrative data respond to the questionnaire indicating they are citizens, indicating that this practice should be revisited in the case of measuring citizenship. Finally, for those 22.2 million cases that do not link to administrative records (non-linkage occurs for the same data quality reasons discussed above), the Census Bureau will use the observed 2020 responses. Again, Census Bureau expect some quality issues with these responses. Namely, the Census Bureau estimates that just under 500 thousand noncitizens will respond as citizens.

The relative quality of Alternative C versus Alternative D will depend on the relative importance of the errors in administrative data, response data, and imputations. To be slightly more but not fully precise consider the following description of errors under both alternatives. First note that all possible measurement methods will have errors. Under Alternative C, there will be error in the administrative records, but we believe these to be relatively limited dues to the procedure following by SSA, USCIS and State. In both Alternative, the modeled cases will be subject to prediction error. Prediction error occur when the model returns the incorrect status of a case. As there are more models cases in Alternative C, prediction error will be a bigger issue there. Alternative D has an additional source or error, response error. This is where 2020 respondent give the incorrect status. Statisticians often hope these error are random and cancel out. However, we know from prior research that citizenship status responses are systematically biased for a subset of noncitizens. Response error is only an issue in alternative D. Unfortunately, the Census Bureau cannot quantify the relative magnitude of the errors across the alternatives at this time.

Figure 1

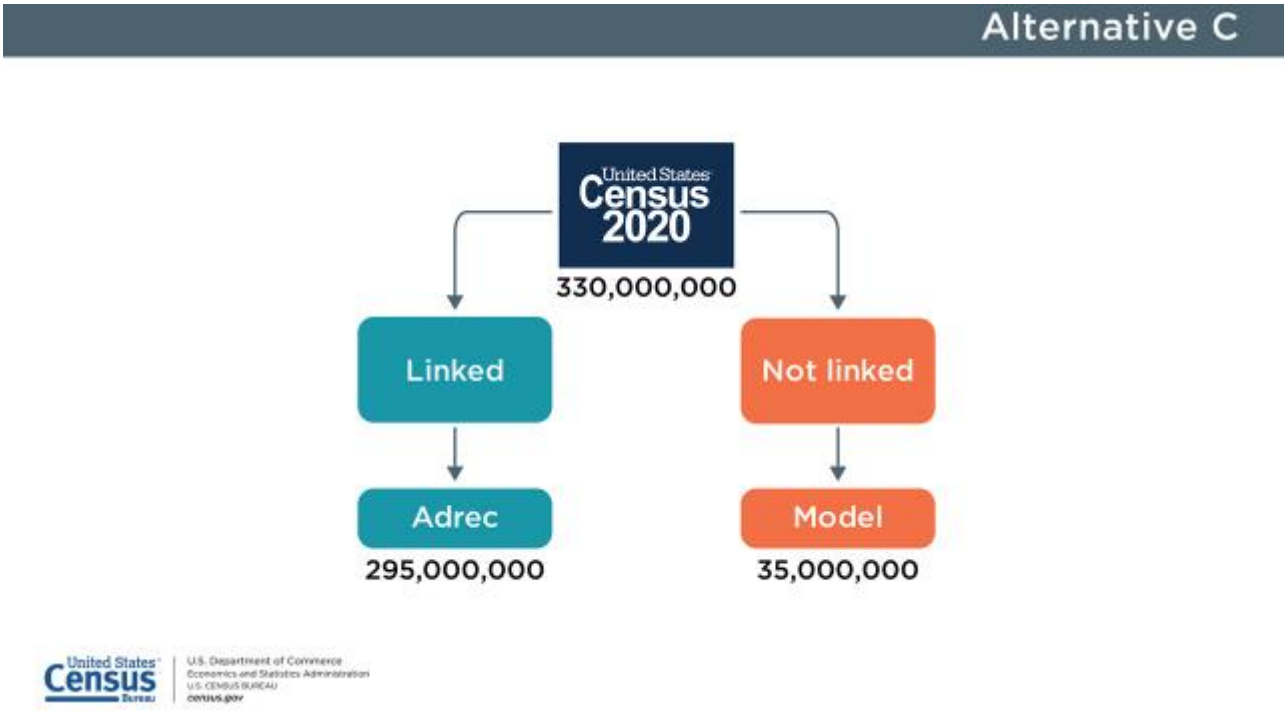
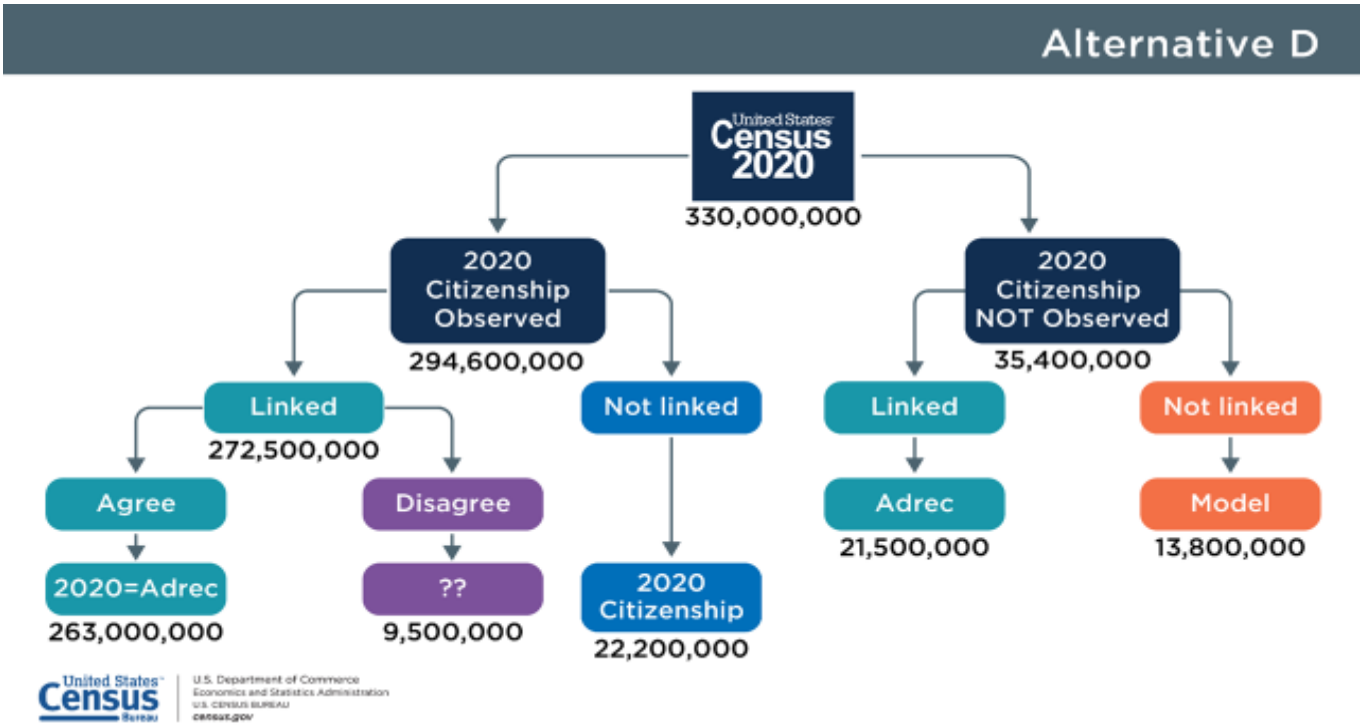

Alternative C
United States Census 2020
330,000,000
Linked
Not linked
Adrec
Model
295,000,000
35,000,000
United States Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Figure 2


Alternative D
United States Census 2020
330,000,000
2020 Citizenship Observed
294,600,000
2020 Citizenship NOT Observed
35,400,000
Linked
272,500,000
Not linked
Linked
Not linked
Agree
Disagree
Adrec
Model
21,500,000
13,800,000
2020=Adrec
263,000,000
??
9,500,000
2020 Citizenship
22,200,000
United States Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov