Understanding the Quality of Alternative Citizenship Data Sources for the 2020 Census

by

J. David Brown
U.S. Census Bureau

Misty L. Heggeness
U.S. Census Bureau

Suzanne M. Dorinski
U.S. Census Bureau

Lawrence Warren
U.S. Census Bureau

Moises Yi
U.S. Census Bureau

CES 18-38 **August, 2018**

Abstract

This paper examines the quality of citizenship data in self-reported survey responses compared to administrative records and evaluates options for constructing an accurate count of resident U.S. citizens. Person-level discrepancies between survey-collected citizenship data and administrative records are more pervasive than previously reported in studies comparing survey and administrative data aggregates. Our results imply that survey-sourced citizenship data produce significantly lower estimates of the noncitizen share of the population than would be produced from currently available administrative records; both the survey-sourced and administrative data have shortcomings that could contribute to this difference. Our evidence is consistent with noncitizen respondents misreporting their own citizenship status and failing to report that of other household members. At the same time, currently available administrative records may miss some naturalizations and capture others with a delay. The evidence in this paper also suggests that adding a citizenship question to the 2020 Census would lead to lower self-response rates in households potentially containing noncitizens, resulting in higher fieldwork costs and a lower-quality population count.

[*] We thank career staff and statistical experts within the Bureau who graciously gave their time and effort to review, comment, edit, and make improvements to this document. The analysis, thoughts, opinions, and any errors presented here are solely those of the authors and do not necessarily reflect any official position of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. The Disclosure Review Board release numbers are DRB-B0093-CDAR-20180621, DRB-B0103-CDAR-20180712, and DRB-B0113-CDAR-20180806. Republication in whole or part must be cleared with the authors.

J. David Brown is a Senior Economist in the Center for Economic Studies at the U.S. Census Bureau and the corresponding author on this paper, 4600 Silver Hill Road, Washington, DC 20233, j.david.brown@census.gov. Misty L. Heggeness is Senior Advisor for Evaluations and Experiments in the Research and Methodology Directorate at the U.S. Census Bureau. Suzanne M. Dorinski is a Mathematical Statistician currently on detail with the Social, Economic, and Housing Statistics Division at the U.S. Census Bureau. Lawrence Warren is an Economist in the Center for Economic Studies at the U.S. Census Bureau. Moises Yi is an Economist in the Center for Economic Studies at the U.S. Census Bureau.

1. Introduction

National statistical agencies are charged with collecting and reporting accurate information about society, including individuals, households, and businesses. This information is used to produce official statistics about the demographic composition of persons living in the nation – including information about migration, citizenship, and mobility. For decades, the United States has relied on household survey questionnaires to collect data on migration and immigration status (Census Bureau 2002). Generally, the focus is on whether an individual has lived in that current location for more than one (or five) years, a date for their last move, citizenship status, and year of naturalization. To date, the collection of this information via survey vehicles has been sufficient for general statistical reporting on immigrants living in the U.S.; however, very few studies have examined the extent to which individuals answer these sensitive questions accurately, how inclusion of these questions affects overall response rates, or how item nonresponse on these questions compares to other questions.

In this paper, we study the quality of self-reported citizenship questions by comparing responses in the American Community Survey (ACS), the Census, the Survey of Income and Program Participation (SIPP), and administrative records on citizenship from the Social Security Administration. There are now multiple survey and administrative sources of data to study immigration and citizenship status. We examine the strengths and weaknesses of these sources for the development of future statistics on citizenship status. We focus on both the accuracy and completeness in all options. The alternatives we consider for constructing a count of resident citizens are the following: (A) no change in current data collection, combined with small area estimation using the ACS and administrative citizenship data sources, (B) add a citizenship question to the 2020 Census, (C) obtain citizenship status from administrative records for the entire 2020 Census population, and (D) combine alternatives (B) and (C). Factors to consider when evaluating these alternatives include the quality of the data sources, comprehensiveness and biases in data coverage, cost, and the effects on the quality of the 2020 full population count. We analyze each of these aspects.

We find that discrepancies between survey-collected citizenship data and administrative records are more extensive than discrepancy estimates from previous research. The degree to which persons who are noncitizens in administrative records self-report being citizens in surveys is greater for non-Hispanics than Hispanics. Most of the people with these discrepancies report being citizens from birth or naturalized long ago, regardless of ethnicity. The discrepancy patterns imply that the ACS estimate of the noncitizen share of the population is lower than comparable estimates based on currently available administrative records.

The remainder of the paper is structured as follows. Section 2 provides general background and history of the current issue. Section 3 documents the coverage of survey and administrative record citizenship data. The quality of the data from survey and administrative record sources is analyzed in Section 4. Section 5 contains regression analyses of item response and data quality. Section 6 estimates the effects of inclusion of a citizenship question on survey response rates. Estimates of the citizenship question's effects on the cost and quality of the 2020 Census in general are provided in Section 7. Forecasts of the number of people for whom citizenship is sourced by the 2020 Census

citizenship question, administrative records, and model imputation when using each of the alternatives are given in Section 8. Section 9 concludes.

2. Background

2.1 History of Citizenship Data Collection through Household Surveys and Censuses

The Census Bureau has collected and preserved citizenship data since 1820 via historical full count censuses, household surveys, and administrative records (AR), but the practice of asking citizenship and migration-related questions on censuses has varied over time. The 1820 and 1830 Censuses asked for a tally of the total number of non-naturalized foreigners in the household. The 1870 Census asked citizenship status of all male persons aged 21 and older (Census Bureau 2002). The federal government did not ask citizenship status during the 1880 Census, but reintroduced it in the 1890 Census, and the question stayed on full-count Census questionnaires through 1950. The 1950 Census was the last full-count Census to ask the citizenship status of every resident in the U.S. if he or she reported a foreign birthplace (Census Bureau 2002).

While the 1960 Census did not ask about citizenship throughout the country, it was reintroduced on the long form (which sampled approximately one-in-six households across the country) in the 1970 Census and remained on the long form until 2000 (Census Bureau 2002). The question never reappeared on the short form after 1950. After the 2000 Census, citizenship data collection moved to the American Community Survey (ACS), which replaced the Census long form. The ACS collects responses from approximately 1.6 percent of households annually (American Community Survey 2016a, American Community Survey 2016b).[2]

Since the advent of the long form and continuing with the ACS, the Census Bureau has focused Census enumeration on obtaining only the data necessary for a concise and condensed full-population count (Weinberg 2011). It also prioritizes the collection of data mandated by Public Law 94-171 (PL94), which instructs the Census Bureau to cooperate with state redistricting offices in support of their efforts to redraw legislative districts in compliance with the Constitution, Supreme Court, and the 1965 Voting Rights Act. The questionnaire asks only the core demographic, race, ethnicity, and housing questions, not including citizenship.

2.2 The Citizen Voting Age Population by Race and Ethnicity (CVAP) Table

On December 12, 2017, the Census Bureau received a request from the Department of Justice to include a citizenship question on the 2020 Census of Population and Housing (Department of Justice 2017). The request prompted the Census Bureau to conduct a study of the feasibility and best options for meeting this request. This paper summarizes the technical analysis conducted for alternative options for obtaining citizenship data for the entire population to produce the Citizen Voting Age Population by Race and Ethnicity (CVAP) table at the census block level. CVAP is

[2] We calculate this number using American Fact Finder (AFF) Tables B98001 and B25001.

currently produced at the census block-group level using estimates from the five-year American Community Survey (ACS) data.

Since 1975, the Census Bureau has provided population estimates by detailed geography to support redistricting under Public Law 94-171 (PL94). For the 2000 Census, the Citizen Voting Age Population (CVAP) estimates, tabulated at the block-group level, were produced from the long form citizenship question. Since 2011, the CVAP estimates have been tabulated annually at the block-group level from the most recent 5-year ACS data. The 2011 publication was based on the 2005-2009 ACS surveys. These data were released in the same time frame as the 2010 PL94 redistricting estimates.[3] The redistricting data must be released before April 1st of the year following a census under the authority of 13 U.S.C. Section 141.

The difficulty in integrating these two tables for redistricting and enforcement of the Voting Rights Act was cited by the Department of Justice in its December 12, 2017 letter. The Department of Justice requested block-level citizen voting-age population estimates by the U.S. Office of Management and Budget (OMB)-approved race and ethnicity categories[4] directly from the 2020 Census of Population and Housing, which would require the addition of a citizenship question directly onto the full count 2020 Census enumeration form.

2.3 Prior Research on Citizenship Data Quality

We build on past research on Census citizenship data quality. Prior studies have suggested that citizenship is inaccurately estimated in Census Bureau surveys. Passel and Clark (1997) document that the 1990 Census and 1996 Current Population Survey (CPS) estimates of the number of naturalized persons are much higher than the numbers from Immigration and Naturalization Services (INS) administrative data.[5] The study suggests that about 75 percent of those who report having lived in the U.S. fewer than five years and being naturalized citizens probably are not citizens, at least at the time of the survey. Furthermore, one-third of longer-resident Central American and Mexican origin individuals who self-reported naturalization were probably not citizens at the time of the survey. These discrepancies were attributed to incorrect reporting, possibly because respondents were confused about their status or had an incentive to misreport it to enumerators and interviewers.

Camarota and Capizzano (2004) conducted focus groups with over 50 field representatives (FRs) for the Census 2000 Supplemental Survey (a pilot for the ACS). FRs reported that foreign-born respondents living in the country illegally or from countries where there is distrust in government were less likely to participate. Some foreign-born respondents failed to list all household members. FRs suspected that some foreign-born respondents misreported citizenship status, and they

[3] For more information, see: https://www.census.gov/programs-surveys/decennial-census/about/voting-rights/cvap.html and https://www.census.gov/rdo/data/2010_census.html .
[4] See Office of Management and Budget (1997).
[5] This comes from Van Hook and Bachmeier's (2013) summary of Passel and Clark (1997).

believed this was due to "recall bias, a fear of the implications of certain responses or a desire to answer questions in a socially desirable way."

More recently, Van Hook and Bachmeier (2013) compared 2010 ACS and Office of Immigration Statistics (OIS) naturalizations data, finding that the ACS produced higher naturalization estimates than OIS for those residing in the U.S. less than five years, as well as for longer-resident Mexican-origin persons. Several papers have studied the effects of state immigration laws on the number and locational choices of immigrants (see, for example, Amuendo-Dorantes and Lozano 2014 and 2015, Bohn et al. 2014, Ellis et al. 2014, Good 2013, and Orrenius and Zavodny 2016). They have generally found reductions in the immigrant population after the introduction of these laws.[6] Deterioration in survey data quality during periods of stronger immigration enforcement could help explain the measured reductions. We contribute to the literature on Census citizenship data quality by directly linking Census and household survey data to administrative records. We not only examine the quality of survey-collected citizenship data, but also the effect of including a citizenship question on the quality of other data via their consequences for response rates and nonresponse follow-up.

3.1 Survey Coverage

In addition to the full count Census of Population and Housing that collects a limited amount of information on the entire population once every ten years, the Census Bureau also collects information on individuals and households in both legally-mandated and sponsored (reimbursable) surveys. These surveys collect more detailed demographic, social, and economic characteristics of people living in the United States, including information on citizenship status and migration variables.

The Census Bureau currently conducts four surveys that ask citizenship questions. The American Community Survey (ACS), the Current Population Survey (CPS), the American Housing Survey (AHS), and the Survey of Income and Program Participation (SIPP) all collect data on citizenship status. The universe for citizenship questions on these surveys is all persons living in the household. The ACS, CPS, SIPP, and AHS distinguish between citizens born in the United States, those born in U.S. territories, those born abroad to U.S. citizen parents, and those of foreign nativity but naturalized. Additionally, the SIPP asks about more nuanced naturalizations, including becoming a citizen through one's own or a spouse's military service or via adoption by U.S. citizen parents. [7]

To assess the citizenship coverage of existing Census Bureau survey data, we link all of the household surveys measuring citizenship status to the 2010 Census. The person-level linkage to

[6] For more information, see https://www.troutman.com/files/FileControl/89dad504-6be0-4335-aa1a-35a433102d63/7483b893-e478-44a4-8fed-f49aa917d8cf/Presentation/File/Survey%20of%20state%20and%20federal%20laws%20requiring%20E-Verify.pdf and table 1 in Orrenius and Zavodny (2016) for the list of states with mandatory E-Verify laws.

[7] This information is from the Master Demographic Pilot Feasibility Study.

the 2010 Census is based on the Protected Identification Key or PIK (the Census Bureau's internal unique person identifier) appended to person records using the Person Identification Validation System (PVS). To implement the record linkage, we first compiled an unduplicated list of individuals surveyed by the Census Bureau in Title 13 mandated surveys (ACS and SIPP) and reimbursable surveys (CPS[8] and AHS[9]). We link this unduplicated list of individuals to the 2010 Census (see Appendix Table A1).

Household surveys linked to the 2010 Census contain self-reported citizenship status for 44.6 million people, or 14.4 percent of the 2010 Census population. Of these, 43.1 million report being citizens (see Appendix Table A2). We conclude that the population coverage from existing survey data is a relatively small share of the total population, consistent with the sampling rates of these surveys.

Figure 1 Panel A shows item nonresponse in the 2016 ACS for sex, age, and citizenship.[10] We show nonresponse rates for the full sample, as well as for select subgroups by race/ethnicity and relationship to the householder.[11,12] Sex has the lowest nonresponse rates across the entire sample, as well as within subgroups with all recording less than 1 percent nonresponse, except for nonrelatives. Nonresponse rates for age are higher, and for some subgroups it has the highest level of nonresponse among the three items shown here. This is true for non-Hispanic white, non-Hispanic black, reference person, and relative of the reference person.[13] Hispanics and non-Hispanic other race[14] have higher rates of nonresponse for citizenship than for sex or age, providing some preliminary evidence that these groups could be disproportionately impacted by the addition of citizenship on the 2020 Census questionnaire.

[8] The CPS is sponsored by the Department of Labor's Bureau of Labor Statistics.

[9] The AHS is sponsored by the Department of Housing and Urban Development.

[10] Appendix Table A3 shows item nonresponse rates for questions on the 2000 Census short form and the 2010 Census. We choose sex and age as benchmarks, since they are on the Census questionnaire. As shown in Appendix Table A3, item allocation rates (including both nonresponses and responses that are edited) are higher for many ACS questions than for sex, age, or citizenship, but they are not being considered for inclusion on the Census questionnaire and are thus less relevant.

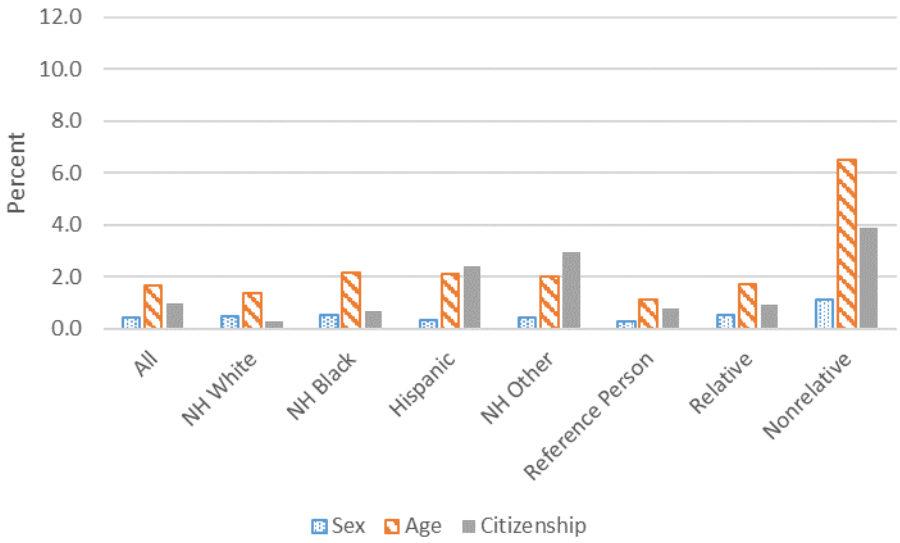
[11] Throughout the paper, we show results not only by citizenship, but also by race and ethnicity for two main reasons. The CVAP data provide counts not just by citizenship, but also race and ethnicity, so differential effects on race/ethnic groups from adding a citizenship question are relevant. In addition, our administrative record noncitizen measure has incomplete coverage (it does not cover noncitizens without SSNs), while a significant percentage of noncitizens without SSNs are Hispanic (Bond et al., 2014). Thus, to some extent the Hispanic category captures noncitizens excluded from the measured noncitizen category.

[12] The householder, also referred to as the reference person or person 1, is the first person listed on the household roster. The reference person typically is the primary or sole respondent to the survey. The relative and nonrelative categories are based on the person's relationship to the householder. The relative category includes husband or wife, biological son or daughter, adopted son or daughter, stepson or stepdaughter, brother or sister, father or mother, grandchild, parent-in-law, son-in-law or daughter-in-law, other relative, unmarried partner, and foster child. The nonrelative category includes roomer or boarder, housemate or roommate, and other nonrelative.

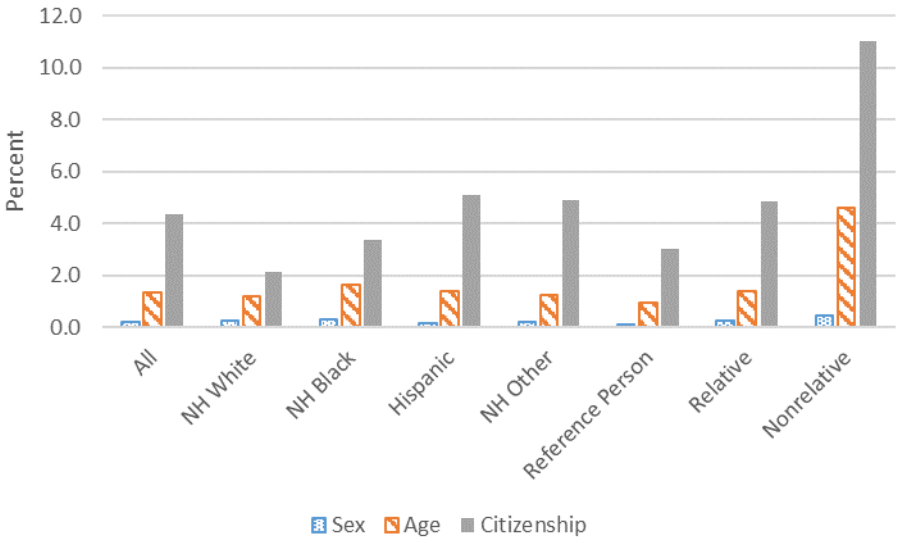
[13] We treat all persons in group quarters as reference persons. The results are qualitatively similar if group quarters are excluded.

[14] Non-Hispanic other race includes non-Hispanic Asian, non-Hispanic American Indian and Alaskan Native, non-Hispanic Native Hawaiian and Other Pacific Islander, and non-Hispanic two or more races.

Figure 1. American Community Survey (ACS) Nonresponse, 2016



Panel A. Item Nonresponse



Panel B. Item Nonresponse for Census Numident-Identified Noncitizens

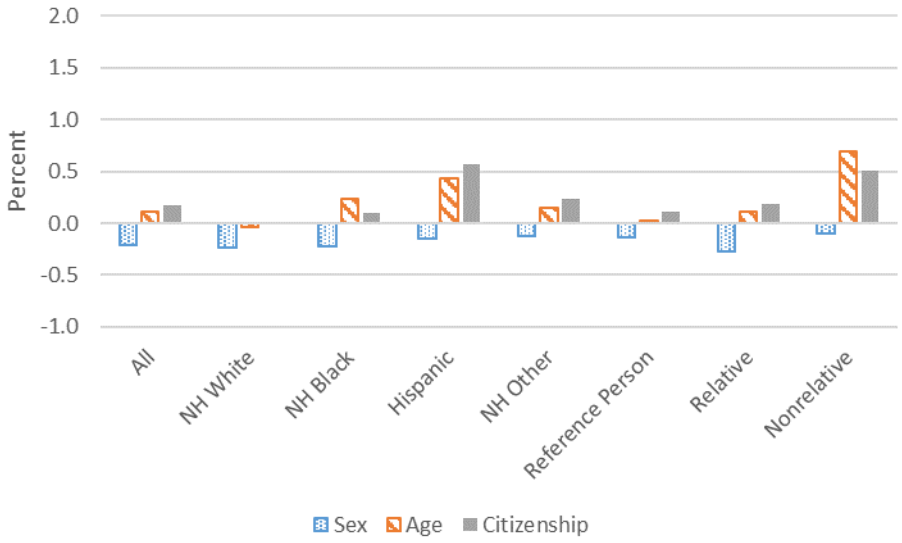
Source: American Community Survey (ACS) and Census Numident, 2016.

Given item nonresponse to the citizenship question as shown in Figure 1 Panel A, we are particularly interested in understanding the potential sensitivity of response specifically for noncitizens. Figure 1 Panel B shows the same information as Panel A, restricted to those

individuals who are identified as noncitizens in the Census Numident,[15] meaning that administrative records show their status as noncitizen. Panel B illustrates the heightened sensitivity associated with collecting citizenship data for noncitizens through surveys. Item nonresponse to the citizenship question is particularly high for nonrelative household members, where one-in-ten do not have a citizenship response in the ACS.

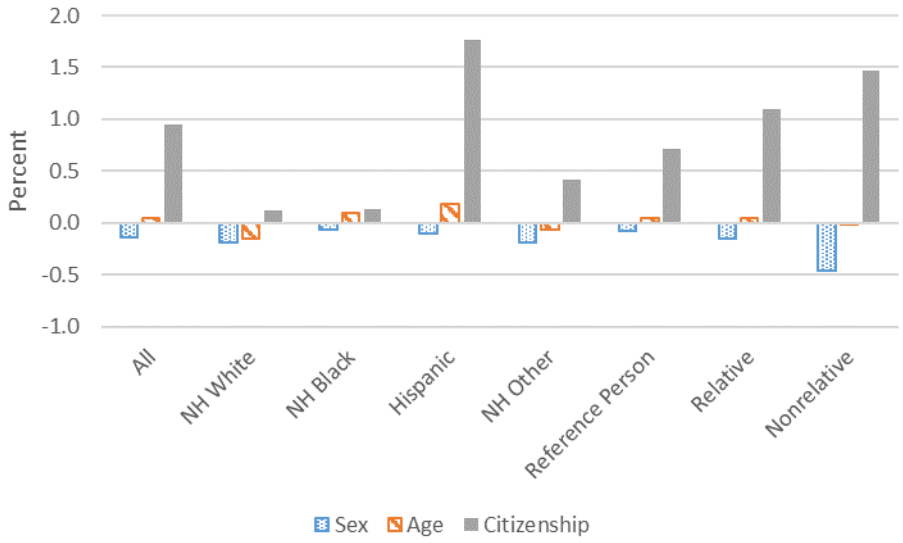
Next, we study whether nonresponse rates have been changing over time. Figure 2 has the same layout as Figure 1. It displays the difference in item nonresponse rates between the 2013 and 2016 ACS for the indicated variable.[16] A positive value indicates an increase in the item nonresponse rate, while a negative value indicates a decrease in the same rate. Figure 2, Panel A reports the difference in rates for the entire survey population as well as subgroups (see also Appendix Table A3 for the rates in the 2000 and 2010 Census short forms). Notice that item nonresponse rates for sex have gone down over time. However, item nonresponse for age and citizenship have increased, and, in particular, the increase in citizenship item nonresponse is largest for Hispanics and nonrelatives.

Figure 2. Difference in American Community Survey (ACS) Item Nonresponse between 2013 and 2016



Panel A. Difference in Item Nonresponse

[15] The Census Numident, which contains all Social Security card applications, is currently the Census Bureau's most complete and reliable administrative record source of citizenship data. For more details, see Section 3.2.
[16] Appendix Table A5 shows citizenship item nonresponse rates in 2013 and 2016 separately for mail-in and internet responses.



Panel B. Difference in Census Numident-Identified Noncitizen Item Nonresponse

Source: American Community Survey (ACS) and Census Numident, 2013 and 2016.

Note: Administrative record noncitizens make up 6.7 percent of the overall 2016 ACS sample.

Figure 2 Panel B shows the same differenced rates, but for those who are identified as noncitizens in the Census Numident. The trends over time are relatively similar for sex and age, with minimal changes. However, item nonresponse to the ACS citizenship question increased for all noncitizen groups, rising by 1.5 percentage points for nonrelatives and 1.8 percentage points for Hispanics. Hispanics, nonrelatives, and noncitizens are particularly sensitive to answering the citizenship question in the ACS, and that sensitivity has increased in recent years.

Table 1 shows break-off rates for the 2016 ACS internet self-responses (ISR) separately by question screen. Using this table, we examine which questions are subject to higher break-off rates. Higher break-off rates indicate potentially sensitive items. They are used as an indicator to inform when the respondent might stop answering the rest of the questions on a survey (Census Bureau 2013). A break-off is the moment in time during which a respondent decides not to continue with the survey and leaves the on-line survey. Break-off rates are highest for Hispanics and lowest for non-Hispanic whites in all question screens. Citizenship-related questions have the most heterogeneous rates across race/ethnicity groups: the ratio of break-off rates for Hispanics versus non-Hispanic whites is much higher for year of entry and citizenship than any of the other question screens in the ACS, except for English proficiency (included in Table 1 for reference purposes). In contrast, financial and work-related questions are sensitive for all groups. This again suggests that citizenship-related questions are more sensitive for Hispanics.

Table 1. 2016 ACS Internet Self-Response Break-off Rates (%) by Screen

	Non-Hispanic White		Non-Hispanic Other		Hispanic	
	(%)	S.E.	(%)	S.E.	(%)	S.E.
Work Location	0.642	0.011	1.045	0.032	1.246	0.038
Place of Birth	0.448	0.009	0.766	0.026	0.961	0.039
Wage Amount	0.589	0.006	0.691	0.029	0.751	0.032
Work Last Week	0.257	0.006	0.407	0.010	0.597	0.024
Work for Wages	0.365	0.009	0.459	0.019	0.590	0.028
Type of Employee	0.221	0.007	0.367	0.011	0.399	0.026
Verify Income	0.198	0.007	0.263	0.016	0.368	0.021
Citizenship	0.035	0.002	0.268	0.016	0.363	0.026
Health Insurance	0.188	0.006	0.331	0.015	0.336	0.019
Highest Level of Education	0.167	0.005	0.257	0.015	0.298	0.019
Work Duties	0.143	0.005	0.223	0.015	0.266	0.020
Year of Entry into U.S.	0.022	0.002	0.119	0.009	0.260	0.021
Taxes	0.164	0.005	0.182	0.014	0.259	0.019
Interest, Dividends Income	0.209	0.006	0.179	0.013	0.242	0.020
Residence Last Year	0.104	0.004	0.182	0.014	0.232	0.016
English Proficiency	0.003	0.001	0.020	0.005	0.036	0.007
Total Non-Breakoff	90.52	0.040	85.93	0.109	82.41	0.145

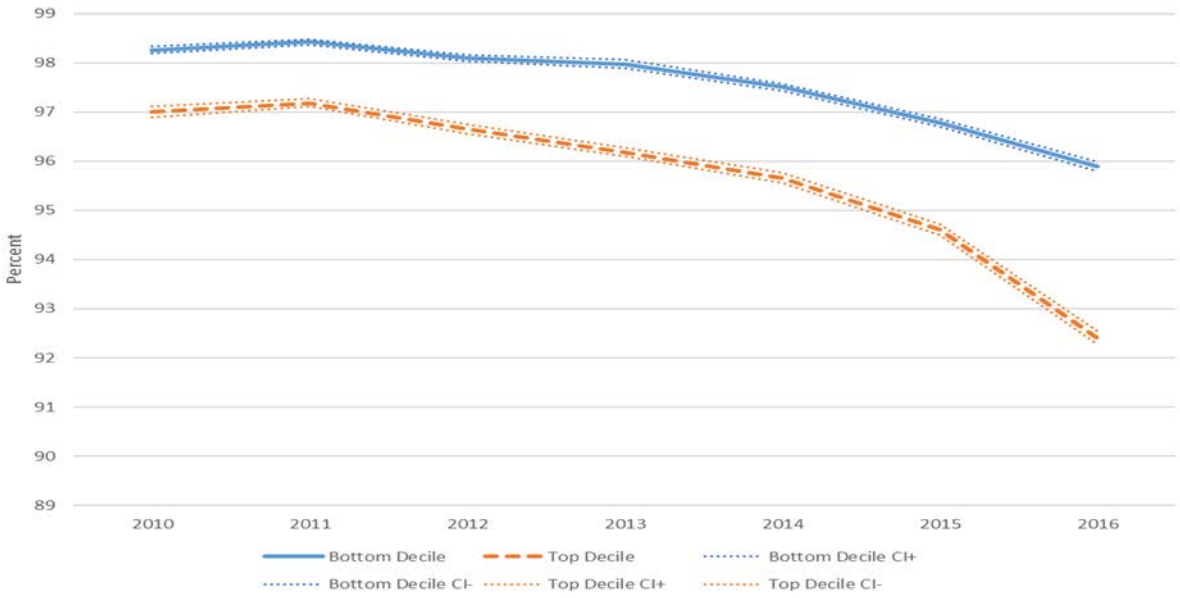
Source: 2016 ACS.

Notes: These are the top fifteen screens, sorted by Hispanic break-off rate. English proficiency and total non-breakoff are also included for reference. The rates are unweighted. The standard errors are calculated using Fay's balanced repeated replication variance estimation method, with 80 replicate weights, adjusting the original weights by a coefficient of 0.5.

Another alternative for measuring sensitivity of response is to examine the extent to which unit nonresponse changes. Unit nonresponse refers to a situation where no one in the household (or unit) responds to the survey. Figure 3 shows ACS unit nonresponse rates from 2010 to 2016 for housing units in the decile of tracts with the highest percent of noncitizens (25.5 percent noncitizens or more), and those in the decile of tracts that have the lowest percent of noncitizens (0.6 percent or less).[17] Tracts with noncitizen shares in the top decile have lower levels of unit response. In tracts with the highest concentrations of noncitizens, unit response rates have decreased over time and show a sharper drop between 2015 and 2016 than for units in tracts with the lowest concentrations of noncitizens.

[17] An internet response option was introduced to the ACS in 2013. Baumgardner, Griffin, and Raglin (2014) show that this was associated with an increase in self-response rates for economically advantaged groups and a decrease for economically disadvantaged groups, which could help explain the widening of the gap between these two tract groups in 2013. It cannot explain the further widening of the gap in 2016, however.

Figure 3. ACS Unit Response Rate by Tract-Level Share of Noncitizens



Source: American Community Survey (ACS), 2010-2016. The deciles of the distribution for noncitizen share of the tract population are 2011-2015 5-year ACS estimates.

Notes: The noncitizen share is 0.0 to 0.6 percent in the bottom decile and 25.5 to 100 percent in the top decile. The confidence intervals (CI) are at the 90 percent level, calculated via the successive differences replicate methodology, using 80 ACS replicate weights (see American Community Survey (2014)).

The data shown in this section provide preliminary evidence that unit nonresponse and citizenship item nonresponse rates are low in the population as a whole. The very low unit and item nonresponse rates among citizens and non-Hispanics mask increasingly higher noncitizen and Hispanic nonresponse rates, however.

3.2 Administrative Record Coverage

An alternative way to obtain citizenship information is to use data collected in the administration of government programs or by commercial data resellers. Respondent sensitivity to answering the question should be less of an issue with administrative sources, since proof of citizenship status is required to determine eligibility for a passport, a job, or government benefits. However, administrative data have incomplete coverage for other reasons, as discussed in this subsection.

Among the sources in Table 2, the Census Numident is the most complete and reliable administrative record source of citizenship data currently available to the Census Bureau. The Numident file is a record of individual applications for Social Security cards and certain subsequent transactions for those individuals. Unique, life-long Social Security Numbers (SSNs) are assigned to individuals based on these applications. In addition, a full record of all changes to the account information (such as change of name) is also maintained. To obtain an SSN, the

applicant must provide documented identifying information to the Social Security Administration (SSA). Through the "enumeration at birth" program, children can be issued an SSN when they are born.[18] Examples of data elements on a Numident record include name, date and place of birth, parents' names, and date of death. The SSA began requiring evidence of citizenship in 1972. Hence, citizenship data for more recently issued SSNs should be reliable as of the time of application.[19] SSA is not automatically notified when previously noncitizen SSN holders become naturalized citizens, however, so some naturalizations may be captured with a delay or not at all. To change citizenship status on an individual's SSN card, naturalized citizens must apply for a new card, showing proof of the naturalization (U.S. passport or certificate of naturalization).[20] Naturalized citizens wishing to work have an incentive to apply for a new card showing their U.S. citizenship, because noncitizen work permits expire, and the Numident is used in combination with U.S. Citizenship and Immigration Services (USCIS) data in the E-Verify program that confirms that job applicants are eligible to work.

Whether or not citizenship data are collected on the 2020 Census questionnaire, administrative records may be useful for editing and imputing the citizenship variable, when necessary.[21]

[18] A parent can apply for the infant's SSN at the hospital where the infant is born. Otherwise, applications for U.S.-born persons require an original or certified copy of a birth record (birth certificate, U.S. hospital record, or religious record before the age of five including the date of birth), which SSA verifies with the issuing agency, or a U.S. passport. Foreign-born U.S. citizen applications require certification of report of birth, consular report of birth abroad, a U.S. passport, a certificate of citizenship, or a certificate of naturalization. Noncitizen applications require a lawful permanent resident card, machine readable immigrant visa, arrival/departure record or admission stamp in an unexpired foreign passport, or an employment authorization document. See https://www.ssa.gov/ssnumber/ss5doc.htm. The enumeration at birth was rolled out starting in 1987, and 45 states, Puerto Rico, the District of Columbia, and New York City had signed agreements to offer it by 1991. Today over 90 percent of parents use this process in all 50 states plus Puerto Rico and the District of Columbia. See https://www.ssa.gov/policy/docs/ssb/v69n2/v69n2p55.html.

[19] A detailed history of the SSN is available at https://www.ssa.gov/policy/docs/ssb/v69n2/v69n2p55.html (Exhibit 1). For some categories of persons, the citizenship verification requirements started a few years later, but all were in place by 1978.

[20] For more information, see https://www.ssa.gov/ssnumber/ss5doc.htm.

[21] Data edits refer to updating data when there is a clear error either in data entry or in response. Imputations occur when the individual or household did not answer a survey or questions on a survey. They involve modeling a most likely response for that individual or household using other available data.

Table 2. Administrative Record (AR) Sources Currently Held and/or Under Negotiation for Acquisition

Administrative Records Data with Citizenship Info. Currently Held	Universe
Census Numident	National-level file of SSA transactions
HHS TANF	National Level (not full content for all states)
Alaska Permanent Fund	Alaska residents
Colorado Leap	Colorado low income energy assistance program
Some State SNAP/TANF	State-level program participants
Army	Active duty and retired soldiers and family members
Bureau of Prisons	Federal prison inmates
Commercial Files	Purchased data from data resellers
Administrative Records Data with Citizenship Info Under Negotiation for Acquisition	**Universe**
Department of Homeland Security United States Citizenship and Immigration Services	National-level file of Lawful Permanent Residents, Naturalizations
Department of Homeland Security United States Customs and Border Protection	National-level file of Customs and Border transaction data
Department of State Passport Services	National-level passport transaction data

Table 3 shows the coverage of the 2010 Census population by the 2010 Numident and ITINs.[22] Ninety-one percent of persons in the 2010 Census can be assigned a Protected Identification Key (PIK) by the Person Identification Validation System (PVS).[23] Once a PIK is assigned, the vast majority of records are matched to the 2010 Numident (98.2 percent in Table 3). Most of the PIKs associated with persons not in the 2010 Numident are derived from linkage to Individual Taxpayer Identification Numbers (ITIN), issued by the Internal Revenue Service to persons who do not have

[22] Rastogi and O'Hara (2012) used an earlier version of the crosswalk between the Numident and ITINs and the 2010 Census, and we show results using that version in Table A6. The enhanced crosswalk in Table 3 uses additional household and geospatial information to increase person linkage, and it has much greater coverage of ITINs. See Bond et al. (2014) for details.

[23] See NORC (2011) and Layne, Wagner and Rothhaas (2014) for details about the process used to assign and the quality of the PIKs used in data linkage at the Census Bureau.

and are not eligible to obtain SSNs, but are required to file a federal individual income tax return (4.3 million person links derived from ITINs vs. 804,000 person links that are not derived from ITINs). Among persons with nonmissing citizenship in the 2010 Numident, 91.3 percent are U.S. citizens.

Approximately 20.9 percent, or 57.6 million of the 2010 Numident records have missing citizenship status. Many older persons did not report citizenship when applying for an SSN, which was not required prior to 1972. Of these older persons with missing citizenship, 7.0 million have either passed away by 2017 or are likely to do so by 2020 (since they would be over 100 years old). Of the remaining 50.7 million persons with missing citizenship in the 2010 Numident, it becomes nonmissing for 5.8 million of them by 2017, nearly all switching to U.S. citizens. About 42.5 million of those still missing citizenship in 2017 were born in the U.S. We treat U.S.-born persons missing citizenship as administrative record citizens in our analysis.[24] This leaves just 2.5 million foreign-born persons with missing citizenship, some of whom could be noncitizens. In the analysis, we treat foreign-born persons with missing citizenship as having missing administrative record citizenship.

Appendix Table A7 shows that among persons who are missing citizenship, alive in 2017, and born after 1919, those who are foreign-born have a much lower propensity to be linked to the 2010 Census (36.3 percent vs. 74.5 percent for U.S.-born persons). Many of the foreign-born people missing citizenship in the Numident are presumably residing outside the U.S. and thus will not be counted in the 2020 Census.[25]

[24] Analysis in later sections of this paper labeled "initial assumptions" instead treats all persons with missing Numident citizenship values as AR citizens, whether they are U.S.- or foreign-born. This includes Table 6, Figures 10B, 11A, 12A, and 12C and Appendix Tables A8 and A9.

[25] An example is persons who received temporary work visas prior to when evidence of citizenship was required to receive an SSN and who have since returned to their home countries.

Table 3. Administrative Record (AR) Coverage of the 2010 Census

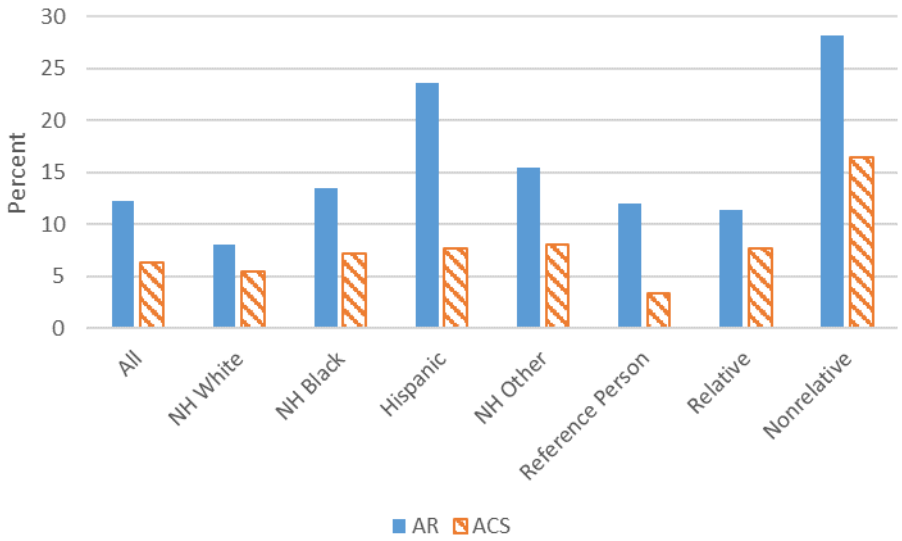
	Count	Percent of 2010 Census Population	Percent of Matched Sample
No PIK, not sent to PVS	10,260,000	3.3	
No PIK, failed in PVS	17,490,000	5.7	
PIK, but not in 2010 Numident, not an ITIN	804,000	0.3	
PIK, but not in 2010 Numident, is an ITIN	4,326,000	1.4	1.5
2010 Numident U.S. Citizen	199,300,000	64.6	71.1
2010 Numident Noncitizen	18,970,000	6.1	6.8
2010 Numident Missing Citizenship	57,620,000	18.7	20.6
Of which:			
Alive in 2017, born after 1919	50,670,000	16.4	18.1
Of which:			
2017 Numident U.S. Citizen	5,678,000	1.8	2.0
2017 Numident Noncitizen	70,500	0.0	0.0
2017 Numident Missing Citizenship	44,920,000	14.5	16.0
Of which:			
U.S.-born	42,460,000	13.8	15.2
Foreign-born	2,464,000	0.8	0.9
Total	308,745,538	100.00	100.00

Source: 2010 Census Unedited File (CUF) and 2010 and 2017 Census Numident Files.

Notes: The 2010 Census Numident File is used for all calculations with "Numident" in the label. The 2017 Census Numident File is used to calculate the number alive in 2017 and born after 1919 and the foreign-born share of them. PVS is the Person Identification Validation System used to assign PIKs. PIK is Protected Identification Key, which is a unique person identifier.

Figure 4 shows the share of persons in the 2016 ACS for whom administrative record citizenship status is not available, as well as the ACS citizenship allocation rate (including both item nonresponse and edits to original responses; i.e., the share of persons for whom the value tabulated is not the respondent's answer). The missing data rates are higher for administrative records (AR) than the ACS, and both sources' rates are higher for minorities and nonrelatives. The variability in coverage is higher for AR than the ACS.

Figure 4. Percent without Administrative Record or ACS Citizenship in 2016



Source: American Community Survey (ACS) and Census Numident, 2016.

Note: For the ACS this is the citizenship item allocation rate, which includes both item nonresponses and edited values.

As shown in Appendix Table A8, the percent of persons in the ACS who cannot be linked to citizenship in AR increases from 8.5 to 10.9 percent between 2010 and 2016. Note that 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 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 35 percent as large as in 2010,[26] suggesting that either fewer

[26] This percentage uses survey weights. Unweighted, it is 39 percent.

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.

There is an important caveat to the conclusion that survey-based citizenship data are more complete than currently held administrative records. The methods used to adjust the ACS weights for survey nonresponse and to allocate citizenship status for item nonresponse assume that the citizenship status distribution of the sampled non-respondents is statistically the same as that of respondents with similar related characteristics. They might not actually be similar, however, even when selecting the allocation of citizenship status using basic characteristics. For example, Hispanics who respond to the survey might be different from Hispanics who do not respond in various characteristics (including immigration status). Additionally, our unit and item nonresponse analysis in Section 3.1 above 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.

To reduce the AR coverage gaps, the Census Bureau is considering the possibility of acquiring access to several other national citizenship-related files listed in Table 2. United States Customs and Immigration Services (USCIS) administrative records on naturalizations and lawful permanent residents (LPR), and Customs and Border Protection transaction records on border entries can partially address the weaknesses of the Numident. Through preliminary project development discussions with USCIS, we were informed that USCIS records provide up-to-date information since 2001 (and possibly back to 1988, but with incomplete records prior to 2001). These will fill some gaps for naturalized citizens, lawful permanent residents, and persons with extended visa applications without SSNs, as well as naturalized citizens who did not inform SSA about their naturalization. These data do not cover naturalizations occurring before 1988, and they miss some between 1988 and 2000. USCIS records do not always cover children under 18 at the time a parent became a naturalized U.S. citizen. These children automatically become U.S. citizens under the Child Citizenship Act of 2000. The USCIS receives notification of some, but not all, of these child naturalizations. Others inform the U.S. government of their U.S. citizenship status by applying for U.S. passports, which are less expensive than the application to notify the USCIS. USCIS visa applications list people's children but the information may not be in electronic form.

U.S. passport administrative records available from the State Department can help plug the gaps for child naturalizations, missing status on the Numident, and out-of-date citizenship information on the Numident. Since U.S. citizens are not required to have a passport, however, these records will also have coverage gaps.

The acquisition of these sources would also improve record linkage for noncitizens by allowing the construction of a supplementary record linkage master list for such people, who are currently only in scope for receiving a PIK if they apply for and receive either an SSN or ITIN. Improved record linkage would not only facilitate greater use of administrative record citizenship data, but it could also permit other uses of these administrative records in 2020 Census operations to lower costs and raise quality. Noncitizens are a hard-to-count population (as evidenced by the lower ACS

unit response rates in tracts with more noncitizens in Figure 3), making having reliable administrative records on them particularly valuable.

If the Census Bureau were to obtain each of these files, the remaining AR citizenship data gaps would include the following categories:

1. U.S. citizens from birth with no SSN or U.S. passport. They will not be processed by the production record linkage system used for the 2020 Census, because their PII won't be found in the PVS reference files.

2. U.S. citizens from birth born outside the U.S., who do not have a U.S. passport, and either applied for an SSN prior to 1974 and were 18 or older, or applied before the age of 18 prior to 1978. These people will be assigned PIKs, but none of the administrative sources discussed above will reliably generate a U.S. citizenship variable.

3. U.S. citizens who were naturalized prior to 2001 and did not inform SSA of their naturalization, because they originally applied for an SSN after they were naturalized, and it was prior to when citizenship verification was required for those born outside the U.S. (1974). These people either already had an SSN when they were naturalized, and they didn't inform SSA about the naturalization, or they never applied for an SSN. The former group has inaccurate data in the Numident. The latter group will not be assigned a PIK.

4. U.S. citizens who were automatically naturalized if they were under the age of 18 when their parents became naturalized in 2000 or later, and they did not inform USCIS or receive a U.S. passport. Note that such persons would not be able to get an SSN with U.S. citizenship on the card without either a U.S. passport or a certificate from USCIS. These people will also not be assigned a PIK.

5. Lawful permanent residents (LPR) who received that status prior to 2001 and either do not have an SSN, or they applied for an SSN prior to when citizenship verification was required for those born outside the U.S. (1974). The former group will not be found in the PVS reference files. The latter group has inaccurate data in the Numident.

6. Noncitizen, non-LPR, residents who do not have an SSN or ITIN and who did not apply for a visa extension. These persons will not be found in PVS.

7. Persons with citizenship information in administrative data, but the administrative and Census data cannot be linked due to missing or discrepant PII.

It is uncertain whether Census Bureau household survey data could reliably fill the above gaps when their person record cannot be assigned a PIK or when they have a PIK but the administrative record lacks up-to-date citizenship information. Persons in Category 6 have a strong incentive to provide an incorrect survey answer, if they answer at all, due to concerns about the data being used for enforcement.[27] Presumably a significant, but unknown, fraction of persons without PIKs are in

[27] Title 13, U.S.C. prohibits the use of Census data for enforcement purposes, but respondents may still have this concern.

Category 6. Distinguishing these people from the other categories of persons without PIKs is inherently inexact, because there is no feasible method of independently verifying their citizenship status.

4. Data Reliability

To assess the reliability of citizenship data, we compare the responses to the 2000 Census long form and 2010 and 2016 ACS citizenship questions with the administrative record (AR) citizenship variable (from the 2002, 2010, and 2016 Numidents and ITINs for the latter two years).[28] Since previous studies suggest that Census survey-AR discrepancies are greater for Hispanics, and the CVAP tables show citizen counts by race/ethnicity and voting age, we show discrepancies separately by race/ethnicity and the voting-age population (age 18 and over). Appendix Tables A8 and A9 show a full set of results for all three years, while the discussion in this section focuses on the 2016 comparison.

Discrepancies between AR and ACS citizenship could be due to several causes: (1) Linkage errors result in the administrative records not matching to the right people in the ACS. The relative discrepancy rates would vary depending on whether AR citizens or noncitizens have more linkage errors. One might expect unrelated persons in the household to have more linkage errors than relatives of reference persons, since PII quality is likely to improve with familiarity. (2) AR incorrectly report that the person is a citizen. This would appear as AR citizen-ACS noncitizen discrepancies. (3) AR are out of date, missing some naturalizations captured by the ACS. This would show up as AR noncitizen-ACS citizen discrepancies.[29] (4) The respondent does not know the person's citizenship status and guesses wrong. This is most plausible for unrelated persons and least so for the reference person. (5) The respondent misunderstands the question and answers incorrectly, despite actually knowing the citizenship status. It is not clear whether this would lead to more AR citizen-ACS noncitizen or AR noncitizen-ACS citizen discrepancies, but it should not vary across reference person, related persons, and unrelated persons. (6) The respondent knows the person's citizenship status and misreports it. Here the reference person may have a harder time justifying item nonresponse (implying (s)he does not know her/his own citizenship), so the way to keep from attracting attention is to say (s)he is a U.S. citizen. When asked about others, the respondent can more easily say (s)he does not know. This factor is likely to be more relevant when people have heightened concerns that the data will be used for immigration enforcement.

Of the candidate reasons (1) through (3) relevant for administrative records, linkage errors (reason 1) would be the most difficult to overcome. If linked to the wrong people, even perfect administrative records will produce inaccurate statistics. Though improvements can be made to record linkage methods, the linkage quality also depends on the quality of PII supplied by the sources being linked. In contrast, the acquisition of more timely administrative record sources

[28] The 2002 Numident is the closest available Numident to the 2000 Census.

[29] Note that as the Census Bureau receives more administrative record sources of citizenship data, the probability that the administrative records are incorrect should fall.

should reduce missing naturalizations problems (reason 3). The use of additional administrative record sources can also help illuminate instances where currently held administrative records are more likely to be incorrect (reason 2).[30]

Guessing wrong (reason 4) and misunderstanding the question (reason 5) would reduce precision (i.e., increase statistical variability), but it is not clear that either would result in biased estimates. In contrast, intentional misreporting (reason 6) is likely to result in reduced accuracy (more bias), since citizens and noncitizens may have different incentives to misreport status. Of these three reasons, the extent of intentional misreporting is most likely to vary across geographical areas and over time, depending on the degree of concern about personal security.

Figure 5 Panel A shows that a remarkably high 99.6 percent of U.S. citizens (according to administrative records) report being U.S. citizens in the 2016 ACS.[31] This suggests that when AR report the person is a citizen, (s)he is actually a citizen, and reason (2) is not an important factor. The discrepancy rate is higher for Hispanics (2.0 percent) and other minorities (1.3 percent) than for non-Hispanic white individuals. The discrepancy rate is higher for nonrelatives than relatives of the respondent, and for relatives than reference persons, consistent with the reference person knowing other people's status less well than his/her own.

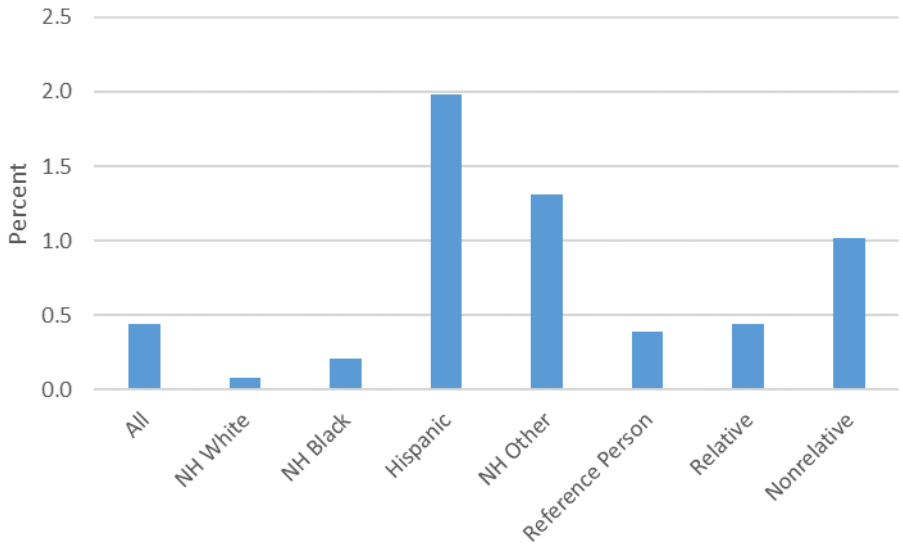
Discrepancy rates are higher for those individuals identified as U.S. noncitizens in administrative records: 37.6 percent report being U.S. citizens in the ACS, as shown in Figure 5 Panel B. This implies that ACS estimates of the U.S. citizen population are higher than they would be if one were to use currently available administrative records.[32] The ordering of rates across groups is reversed compared to the AR citizen-ACS noncitizen rates. Here non-Hispanic white individuals have the highest discrepancy rate and Hispanic individuals the lowest. This means that the difference between ACS citizen and AR citizen population estimates is greatest for non-Hispanic white individuals and lowest for Hispanic individuals. This contrasts with Van Hook and Bachmeier's (2013) conclusion based on aggregates that self-reported naturalizations by persons of Mexican origin are most likely to be incorrect.[33,34]

The AR noncitizen-ACS citizen discrepancy rate is highest for the reference person, followed by relatives and then nonrelatives. This pattern is not a clear outcome of out of date administrative

[30] For example, if a person is a foreign-born citizen in one administrative record source, but other administrative records and the survey response each say the person is a noncitizen, one might have more confidence in selecting noncitizen than when having only the first administrative record source and the survey response.
[31] This is even higher than the agreement rate for sex in the 2010 Census vs. the Numident, which is 99.4 percent. See Rastogi and O'Hara (2012).
[32] Note that since we are unable to compare records that are missing in one or both sources, the estimates provided in this section may understate the difference between the ACS estimate of the U.S. citizen population and the true value, especially since most unauthorized persons (other than the small fraction with ITINs) are missing AR citizenship data here.
[33] Hispanics make up the largest number of AR noncitizen-ACS citizen persons (2.6 million), compared to 2.5 million non-Hispanic other minorities, 1.7 million non-Hispanic whites, and 800,000 non-Hispanic blacks, which may be why previous studies' analysis of aggregated data find the largest administrative record-survey differences to be among Hispanics. But the discrepancy rate is more relevant for evaluating quality than the absolute number of discrepancies.
[34] According to 2016 1-year ACS data in American Factfinder Table S0201 (American Community Survey 2016c), 63.2 percent of Hispanics are of Mexican origin.

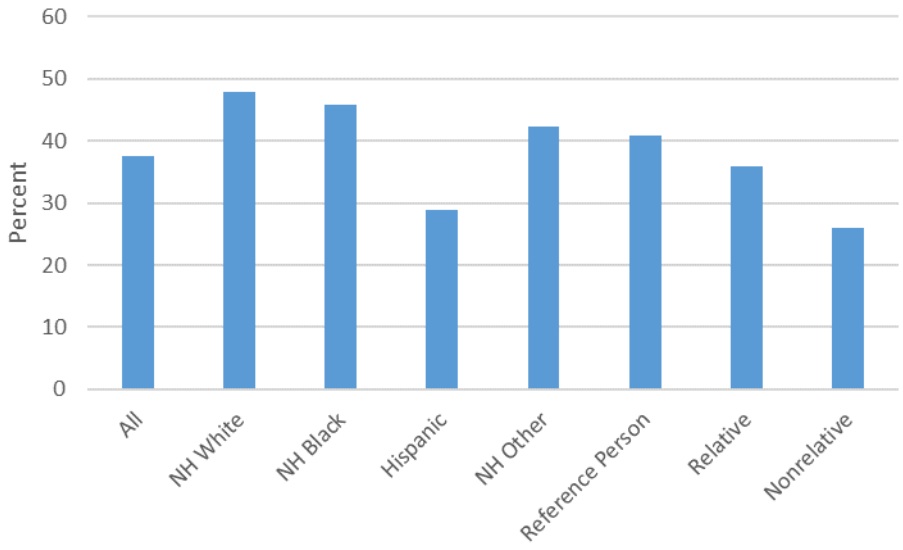
records (reason 3), lack of knowledge about others' status (reason 4), or misunderstanding the question (reason 5). Recall that citizenship item nonresponse is highest for nonrelatives and lowest for reference persons (see Figure 1). This suggests respondents behave differently when asked about their own status versus that of others. It may be easier for respondents to say they do not know the status of someone else (particularly a nonrelative) than their own status. They thus misreport their own status (reason 6), while they say they do not know the status of others.

Figure 5. Administrative Records-ACS Survey Response Citizenship Agreement



Panel A. Percent of Administrative Record Citizens who respond as 2016 ACS Noncitizens

Notes: Administrative record citizens make up 81.1 percent of the overall 2016 ACS sample, 90.1 percent for non-Hispanic white, 81.5 percent of non-Hispanic black, 60.2 percent of Hispanic, 62.5 percent of non-Hispanic other race, 81.1 percent of reference persons, 82.1 percent of relatives, and 64.8 percent of nonrelatives. See Appendix Table A10.



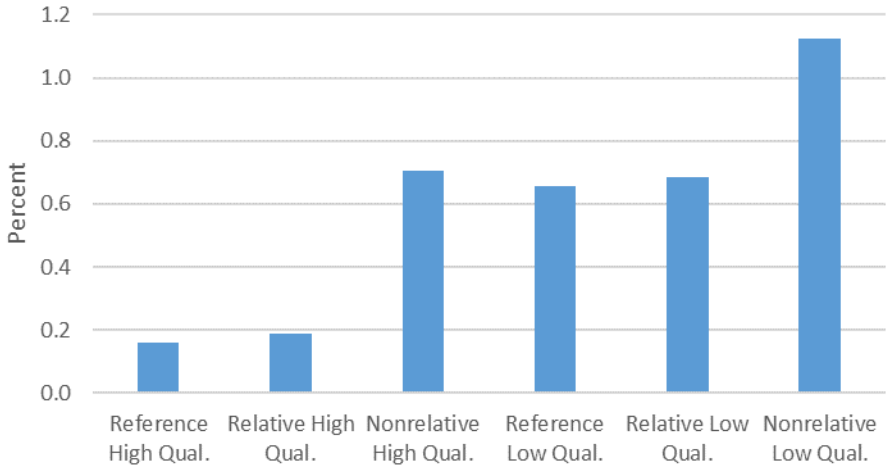
Panel B. Percentage of Administrative Record Noncitizens who respond as 2016 ACS Citizens

Notes: Administrative record noncitizens make up 6.7 percent of the overall 2016 ACS sample, 1.9 percent for non-Hispanic white, 5.1 percent of non-Hispanic black, 16.2 percent of Hispanic, 22.0 percent of non-Hispanic other race, 6.9 percent of reference persons, 6.5 percent of relatives, and 7.1 percent of nonrelatives. See Appendix Table A10.

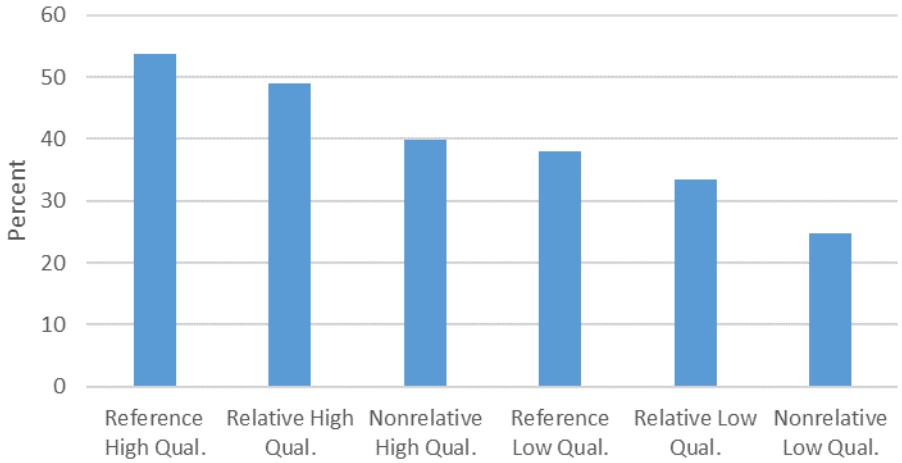
Source: American Community Survey (ACS) 1-year file and Census Numident, 2016.

We show the AR citizen-ACS noncitizen and AR noncitizen-ACS citizen discrepancies separately for higher- and lower-quality linkages and by reference person vs. relative vs. nonrelative categories in Figure 6. For AR citizen-ACS noncitizen discrepancies, the rates are lowest for the reference person and highest for nonrelatives, likely due to people being able to report their own PII more accurately than that of others. Records with high-quality links have lower discrepancy rates, consistent with linkage errors being a contributing factor to these discrepancies. The patterns reverse for AR noncitizen-ACS citizens. Higher-quality linked records actually have higher discrepancy rates, so linkage errors (reason 1) do not appear to explain the AR noncitizen-ACS citizen discrepancies. This pattern holds regardless of the type of person the reference person is responding about (oneself, a relative, or a nonrelative).

Figure 6. Quality of the Citizenship Question Responses by Relation to Reference Person and Higher- vs. Lower-Quality Linkage



Panel A. AR Identifies as a Citizen and 2016 ACS Identifies as a Noncitizen



Panel B. AR Identifies as a Noncitizen and 2016 ACS Identifies as a Citizen

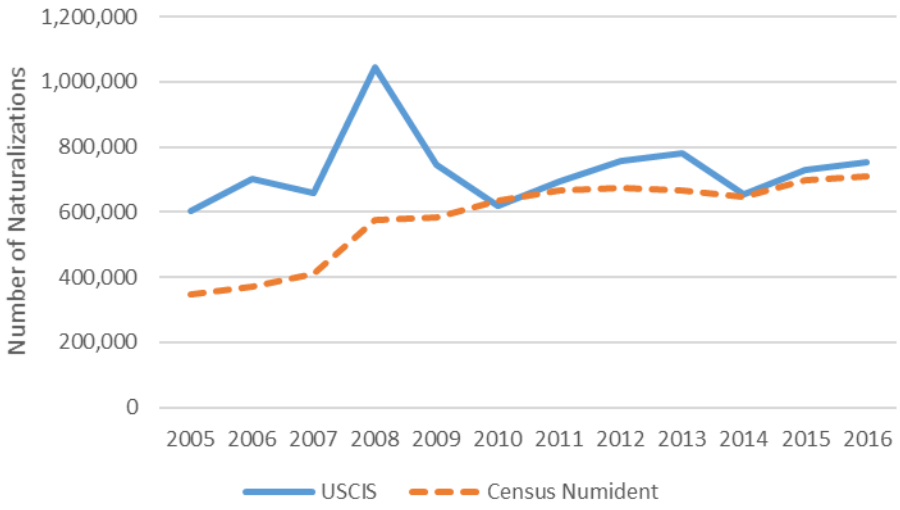
Source: American Community Survey (ACS) and Census Numident, 2016.

Notes: High-quality linkage is defined as having an above-median linkage confidence score on the first linking attempt (pass), and lower-quality is all others. The weighted sample shares of the ACS are 18.1 percent for reference person high-quality linkage, 23.9 percent for relative high-quality linkage, 0.6 percent for nonrelative high-quality linkage, 20.3 percent for reference person low-quality linkage, 33.8 percent for relative low-quality linkage, and 3.2 percent for nonrelative low-quality linkage. See Appendix Table A11.

To evaluate further the hypothesis that AR are out of date (reason 3), we make comparisons to USCIS statistics. In the AR-ACS citizenship status comparison above, we estimate 7,605,000 persons are AR noncitizens-ACS citizens. This is equivalent to the Numident missing all the

naturalizations reported by USCIS back to 2007, plus some of 2006. Figure 7 shows the annual number of persons who first entered the Numident as noncitizens and switch to U.S. citizenship in each particular year, as well as the number of naturalizations according to USCIS statistics.[35] USCIS reports significantly more naturalizations prior to 2010, but there is little difference subsequently. This suggests that if the main reason for the discrepancies were out-of-date Numident citizenship, the Numident would have to be missing many naturalizations that occurred long ago.

Figure 7. Estimated Annual Naturalizations in Census Numident Data versus USCIS Statistics

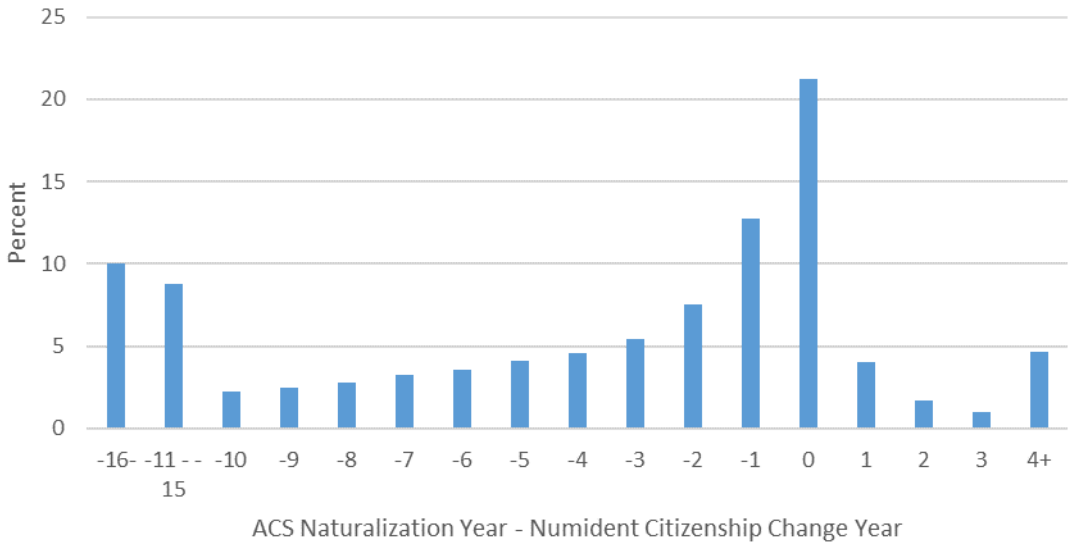


Source: USCIS Immigration Yearbooks and 2017 Census Numident.

We compare the ACS naturalization year and the year when citizenship switched to U.S. citizen in the Numident among persons with naturalized citizen status in both sources in Figure 8.[36] For 67.4 percent of these persons, the ACS naturalization year is earlier than the Numident citizenship change year, and 33.1 percent have an ACS naturalization year that is more than five years prior. Just 11.3 percent have a later ACS naturalization year. This is consistent with tardy notification to SSA about naturalizations.

[35] The Numident switches do not include persons who did not have an SSN prior to being naturalized. According to USCIS officials, the percentage of persons naturalized in 2014 who did not previously have an SSN is 0.33 percent, and it is 0.40 percent in 2015, suggesting that this type of Numident omission is negligible, at least recently.
[36] The Numident citizenship change year is the year when citizenship changed from noncitizen to citizen in the data.

Figure 8. Difference between ACS Naturalization and Numident Citizenship Change Years



Source: American Community Survey (ACS) and Census Numident, 2016. The sample is persons who are naturalized citizens in both sources, and the ACS citizenship value is as reported by the respondent.

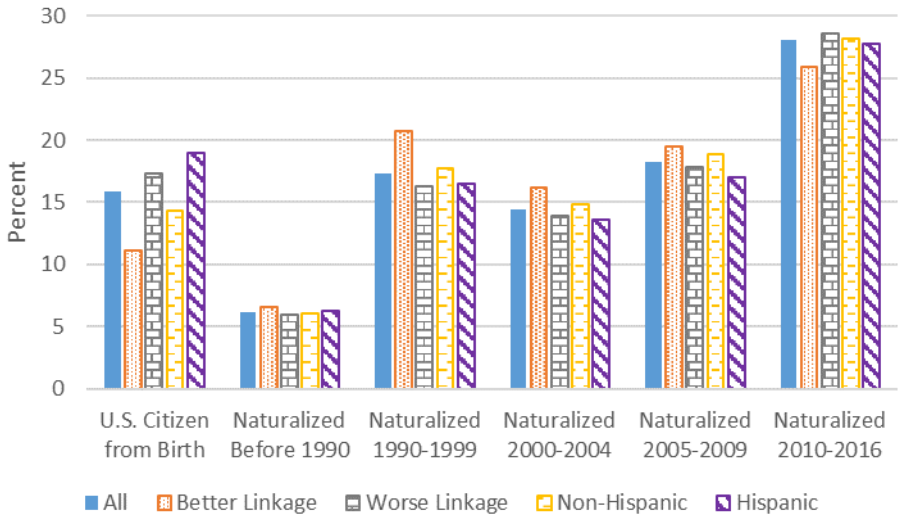
Figure 9 shows the distribution of AR noncitizen-ACS citizens by naturalization year. Approximately 15.9 percent report being citizens from birth, which, if true, would mean that the Numident is not out of date for these people, but incorrect from the first SSN application. This possibility seems unlikely, given that proof of citizenship status must be presented to SSA when applying for an SSN, whereas the ACS citizenship response is not checked. A third of the ACS-reported naturalizations (2.1 million) occurred between 2010 and 2016, while the total gap between USCIS naturalizations and Numident switches from noncitizen to citizen between 2010 and 2016 is several times less than that, at 288,000.

Figure 9 shows that the AR noncitizen-ACS citizen naturalization distributions are very similar for Hispanics and non-Hispanics. The results are contrary to Van Hook and Bachmeier's (2013) finding that citizenship misreporting by persons saying they were naturalized more than five years ago primarily occurs among persons of Mexican origin, and Passel and Clark's (1997) finding that it is among those of Mexican or Central American origin.

We also explore whether the AR noncitizen-ACS citizen naturalization distributions vary with linkage quality. One might expect that if linkage quality is driving the discrepancies, then persons with higher quality links would be recently naturalized, reflecting out-of-date Numident data. In contrast, more of the persons with low quality links would be ACS citizens from birth or naturalizations long ago, since the Numident and ACS records could be for different people, and the Numident should be less likely to be out of date for citizens from birth and earlier naturalizations. Figure 9 does show a higher share of ACS citizens from birth among those with

lower quality links, but also for more recent naturalizations. This is further evidence that linkage errors are probably not an important explanation for these discrepancies.

Figure 9. Distribution of ACS Citizenship Receipt Timing for Administrative Record Noncitizen-ACS Citizens by Linkage Quality and Ethnicity



Source: American Community Survey (ACS) and Census Numident, 2016.

As a final data quality check, we calculate the 2016 ACS citizenship distribution for persons with ITINs. Though only noncitizens may have ITINs, 6.6 percent say they were born citizens, and 11.1 percent report being citizens in the ACS (see Appendix Table A12).

5. Item Response and Data Quality Regression Analysis

We estimate multivariate regressions predicting item response in Table 4 and AR-ACS discrepancies in Table 5. The item response and citizenship status disagreement regressions test whether the associations shown above are statistically significant and robust to inclusion of controls. These analyses also provide an opportunity to study other potentially relevant factors. The item response regressions are estimated separately for AR citizens, AR noncitizens, and those missing AR citizenship. The item response variables are equal to one if there is a response for the item (whether it was later edited or not), and zero otherwise. The ACS noncitizen-AR citizen dependent variable is equal to one if the person is an as-reported noncitizen in the ACS and an AR citizen, and it is zero if both sources say the person is a citizen. Analogously, the ACS citizen-AR noncitizen dependent variable is equal to one if the person is an as-reported citizen in the ACS and an AR noncitizen, and it is zero if both sources say the person is a noncitizen. The last specification in Table 5 investigates determinants of the difference between the ACS naturalization year and the year in which the status changed to citizen in the Numident among persons who were noncitizens in their first SSN application.

Besides relationship to the reference person, we include several other factors that theoretically could drive differences observed in both survey response and data quality. These include demographic characteristics such as sex, race/ethnicity, log one plus age, and its square. We also include socioeconomic characteristics such as educational attainment, working in the last week, and searching for a job in the last four weeks. Educational attainment is classified as less than high school diploma (base category), at least high school but less than a bachelor's degree, bachelor's degree, and graduate degree. Time since entry to the U.S. and reference person English language variables are included, since these variables may influence item response and discrepancies in citizenship status reporting. For our analysis, those variables are log of one plus the number of years since entering the U.S. (or since birth if born in the U.S.) and its square[37] and English language ability for those speaking another language at home (speaking only English at home is the base category). We include an indicator for better or worse quality person linkage, since it may also drive differences in survey response and data quality. An indicator for whether the response is via mail or internet (i.e., without participation by an interviewer) vs. a personal or telephone interview. According to Camarota and Capizzano (2004), item nonresponse rates are lower in in-person interviews, and foreign-born persons are more likely to take the survey via personal interview, so controlling for mode could be particularly important when comparing the behavior of citizens and noncitizens.

The associations highlighted in Figures 1-6 above are robust to inclusion of other variables and are highly statistically significant.[38] Item nonresponse and ACS noncitizen-AR citizen discrepancy rates are higher for nonrelatives, but the ACS citizen-AR noncitizen propensity is much lower, again consistent with reference persons misreporting their own citizenship, but not reporting that of others at all, especially nonrelatives. Like nonrelatives, Hispanics have a lower propensity to provide citizenship, a higher propensity to have ACS noncitizen-AR citizen discrepancies, and a lower propensity to have ACS citizen-AR noncitizen discrepancies. Better linkage is strongly associated with ACS citizen-AR noncitizen discrepancies, inconsistent with the hypothesis that these discrepancies are driven by linkage errors.

Now turning to factors not investigated in previous sections, labor market activity is positively associated with having a citizenship answer; especially for AR noncitizens (see Table 4). However, as Table 5 shows, working is also associated with both types of citizenship status disagreements, particularly ACS citizen-AR noncitizen. Reference persons who speak another language at home have a higher propensity to respond about sex, especially when their English language ability is less strong. This is also true for AR citizens for the citizenship question, but when asked to report about AR noncitizens, those speaking another language at home have much lower citizenship item response rates. Those speaking English less well also have a higher propensity to report ACS noncitizen when the person they are responding about is an AR citizen, perhaps reflecting misunderstanding of the question. However, the reference person's English language ability is positively associated with ACS citizen-AR noncitizen discrepancies, again suggesting that

[37] In cases where the person came to live in the U.S. more than once, respondents are instructed to give the latest year.
[38] In results not shown here, we also estimate item response regressions with the full sample, regardless of AR citizenship status. The patterns are similar to those described in this paragraph, except that Hispanics have higher propensity to have item response for age in the full sample.

misunderstanding the question is an important factor behind ACS noncitizen-AR citizen, but not ACS citizen-AR noncitizen discrepancies. Responding without the participation of an interviewer results in lower item response (except for age for AR noncitizens), consistent with Camarota and Capizzano (2004), and this effect is particularly strong for citizenship item response among AR noncitizens. ACS noncitizen-AR citizen discrepancies are more prevalent with interviewer participation, but ACS citizen-AR noncitizen discrepancies are much less prevalent. Interviewers may develop a rapport that encourages noncitizens to truthfully respond to what is a sensitive question for them.[39] It could also be more difficult psychologically for a respondent to misreport to another person than when they fill out a questionnaire on their own.

As shown in Table 4, the associations with citizenship item response tend to be several times stronger for AR noncitizens than for citizens, with those missing AR citizenship falling in between the other two categories. Such differences are much more muted for sex and age. This again highlights the nonrandom nature of citizenship item nonresponse.

[39] This effect may be weaker in the Census than in the ACS, however, since ACS interviewers have much more experience than most Census enumerators.

Table 4. Item Response Regressions

	Sex Item Response			Age Item Response			Citizenship Item Response		
	AR Citizen	AR Noncitizen	AR Missing	AR Citizen	AR Noncitizen	AR Missing	AR Citizen	AR Noncitizen	AR Missing
Relative	-0.159	-0.079	-0.759	-0.234	-0.224	-4.446	-0.057	-0.480	-0.106
	(0.007)	(0.016)	(0.039)	(0.013)	(0.060)	(0.123)	(0.010)	(0.082)	(0.066)
Nonrelative	-0.455	-0.309	-1.146	-2.353	-3.509	-9.533	-1.141	-7.395	-4.808
	(0.035)	(0.072)	(0.084)	(0.080)	(0.307)	(0.300)	(0.047)	(0.390)	(0.200)
Non-Hispanic	-0.136	-0.160	-0.003	-0.142	-0.227	-0.225	-0.122	-3.092	-0.979
African Amer.	(0.014)	(0.050)	(0.082)	(0.029)	(0.143)	(0.247)	(0.012)	(0.171)	(0.078)
Hispanic	0.128	0.002	0.147	0.033	0.075	2.068	-0.391	-4.432	-1.692
	(0.013)	(0.030)	(0.069)	(0.032)	(0.103)	(0.210)	(0.024)	(0.140)	(0.119)
Other Non-	0.050	0.038	0.230	-0.100	-0.108	1.229	-0.177	-2.320	-1.885
Hispanic	(0.017)	(0.028)	(0.072)	(0.034)	(0.092)	(0.230)	(0.031)	(0.129)	(0.152)
Worked in	0.174	0.073	0.694	0.334	0.149	1.872	0.915	8.687	3.773
Last Week	(0.008)	(0.024)	(0.037)	(0.017)	(0.081)	(0.132)	(0.013)	(0.141)	(0.088)
Searched for	0.045	0.017	0.668	0.457	0.466	3.834	0.769	7.414	3.494
Job	(0.020)	(0.046)	(0.063)	(0.033)	(0.126)	(0.252)	(0.016)	(0.185)	(0.114)
English Very	0.116	0.101	0.690	0.084	0.068	1.823	0.087	-1.036	-0.580
Well	(0.014)	(0.029)	(0.064)	(0.035)	(0.088)	(0.224)	(0.028)	(0.133)	(0.129)
English Well	0.141	0.050	0.703	0.306	0.074	3.044	0.390	-1.688	-0.892
	(0.023)	(0.034)	(0.073)	(0.052)	(0.102)	(0.247)	(0.056)	(0.159)	(0.191)
English Not	0.125	-0.006	0.523	0.056	-0.148	1.728	0.475	-2.115	-0.441
Well	(0.024)	(0.041)	(0.082)	(0.073)	(0.128)	(0.270)	(0.070)	(0.191)	(0.190)
English Not	0.117	0.070	0.599	-0.179	-0.155	3.178	0.571	-1.241	0.846
At All	(0.035)	(0.036)	(0.072)	(0.143)	(0.189)	(0.272)	(0.122)	(0.236)	(0.178)
Better	1.022	0.338	2.502	1.384	1.193	9.002	0.127	1.766	2.078
Linkage	(0.010)	(0.019)	(0.061)	(0.015)	(0.040)	(0.122)	(0.008)	(0.115)	(0.125)
Mail or Internet	-0.967	-0.449	-2.703	-0.083	0.708	-3.527	-0.397	-5.923	-2.329
Response	(0.010)	(0.024)	(0.068)	(0.019)	(0.073)	(0.156)	(0.011)	(0.122)	(0.092)
Weighted Obs.	264,700,000	21,910,000	39,950,000	264,700,000	21,910,000	39,950,000	264,700,000	21,910,000	39,950,000
Unweighted Obs.	4,418,000	280,000	558,000	4,418,000	280,000	558,000	4,418,000	280,000	558,000

Source: American Community Survey (ACS) and Census Numident, 2016. Notes: These regressions are estimated by linear probability models (LPM), weighted by ACS person weights. Standard errors are clustered by household. The base categories are reference person for relationship, non-Hispanic white for race/ethnicity, speaks only English at home for English ability, and in-person or phone interview for response mode. We also include educational attainment (less than high school, high school but less than bachelor's degree, bachelor's degree, and graduate degree), log of one plus age and its square, and log of one plus the number of years in the U.S. and its square, but do not report them here.

Table 5. Citizenship Status and Naturalization Year Disagreement Regressions

	ACS Noncitizen-AR Citizen	ACS Citizen-AR Noncitizen	ACS – Numident Natural. Year
Relative	0.028	-0.753	-0.343
	(0.011)	(0.215)	(0.068)
Nonrelative	0.571	-5.461	-0.852
	(0.045)	(0.613)	(0.282)
Non-Hispanic	-0.137	2.744	0.683
African Amer.	(0.013)	(0.546)	(0.128)
Hispanic	0.621	-16.00	1.129
	(0.030)	(0.417)	(0.104)
Other Non-	-0.327	0.755	0.144
Hispanic	(0.034)	(0.376)	(0.093)
Worked in	0.398	1.992	0.631
Last Week	(0.015)	(0.260)	(0.095)
Searched for	0.302	-0.620	0.136
Job	(0.029)	(0.542)	(0.157)
English Very	-0.452	1.983	0.517
Well	(0.031)	(0.373)	(0.096)
English Well	0.114	1.063	0.712
	(0.081)	(0.426)	(0.107)
English Not	1.461	-4.927	0.997
Well	(0.113)	(0.480)	(0.129)
English Not	3.391	-8.282	1.656
At All	(0.260)	(0.592)	(0.210)
Better	0.060	4.586	0.006
Linkage	(0.009)	(0.308)	(0.067)
Mail or Internet	-0.262	3.810	0.365
Response	(0.012)	(0.285)	(0.077)
Weighted Obs.	250,300,000	20,220,000	6,407,000
Unweighted Obs.	4,165,000	254,000	89,000

Source: American Community Survey (ACS) and Census Numident, 2016.

Notes: These regressions are estimated by linear probability models (LPM), weighted by ACS person weights. Standard errors are clustered by household. The base categories are reference person for relationship, non-Hispanic white for race/ethnicity, speaks only English at home for English ability, and in-person or phone interview for response mode. We also include educational attainment (less than high school, high school but less than bachelor's degree, bachelor's degree, and graduate degree), log of one plus age and its square, and log of one plus the number of years in the U.S. and its square, but do not report them here.

The last specification of Table 5 shows that the ACS naturalization-Numident citizenship change gap is larger when reporting for a relative or especially a nonrelative, which could indicate lack of respondent knowledge about others' naturalization years.[40] Lack of English language ability is associated with a smaller gap between the ACS and Numident years, suggesting that misunderstanding the question is not an important explanatory factor. Employed people have smaller gaps, reflecting the incentive to promptly tell SSA about the naturalization to facilitate their employment eligibility verification.

[40] Since very few observations have Numident citizenship change years before the ACS naturalization year, a positive coefficient generally means a smaller gap.

6. Effect of Citizenship Question on Unit Self-Response Rates

To forecast the effect of adding a citizenship question to the 2020 Census, we compare mail response rates in the 2010 Census and the 2010 American Community Survey (ACS) for the same housing units. By comparing the self-response behavior of the same housing unit across two surveys, we control for the household's propensity to self-respond to mandatory Census Bureau household surveys in general.

The Census Bureau randomly selected a sample of households to receive the ACS questionnaire in 2010. The questionnaire included 75 questions and asked individuals to report their citizenship status. These households also received the full-count Census questionnaire in the same year, a list of 10 questions that did not include citizenship. We focus on Census housing units that received both questionnaires by mail. In the 2010 Census, these are the housing units from the initial mailing that did not have the questionnaire returned as Undeliverable as Addressed (UAA) and which were not classified as a vacant or delete (meaning uninhabitable or cannot be found). We define a 2010 Census self-response as a returned questionnaire from the first mailing that is not blank. For the 2010 ACS, a self-response is a mail response, also from the first contact mailing.

The presence of a citizenship question is not the only potential reason why a household may be less inclined to self-respond to the ACS than the Census. Census self-response is bolstered by a media campaign and intensive community advocacy group support, and the ACS questionnaire involves much greater respondent burden (OMB 2008, OMB 2009). To distinguish the citizenship question effect, we compare the actual ACS-Census difference in response rates for households that are likely to be more sensitive to the citizenship question to the ACS-Census difference for households less likely to be sensitive to the question. We assume that any reduction in self-response to the ACS vs. the Census for households unsensitive to the citizenship question is due to factors other than the presence of a citizenship question. We use two ways to divide the sample into sensitive and non-sensitive groups. The first is to define the sensitive group as households where at least one person is an AR noncitizen and has been assigned to this housing unit in Rastogi and O'Hara's (2012) administrative records person-address crosswalk (AR noncitizen households), and the less sensitive group is households where all of the persons assigned to the address are AR citizens (AR all-citizen households).[41] AR citizenship status is established using the 2010 Numident and ITINs, as described in Section 3.2.[42] The choice of noncitizens as the sensitive group is motivated by the results in Section 3.1 that AR noncitizens have much higher item nonresponse rates for the citizenship question, both relative to their nonresponse rates for other demographic questions and compared to other people for citizenship. The use of an independent source for where noncitizens are located avoids the potential problem that households with noncitizens may be less likely to provide PII on household members, preventing linkage to

[41] Here we impose a restriction that all household members have nonmissing AR citizenship for the less sensitive group, but we do not impose that restriction on the sensitive group.

[42] The initial definition of citizenship (treating all persons in the Numident but with missing citizenship as citizens) is used for this first set of groups. In the second set of groups, U.S.-born persons with missing citizenship in the Numident are treated as citizens, while foreign-born persons with missing citizenship in the Numident are treated as missing AR citizenship.

their AR citizenship data. The remaining noncitizen households where AR linkage is done may be relatively more cooperative, potentially biasing the results.

We examine a second set of groups for several reasons. We would like to project the citizenship self-response effect forward in time, since population characteristics associated with this effect may be changing. No administrative records person-place crosswalk is available after 2010, however, so we instead use the ACS household roster to define which people are living in the household.[43] AR noncitizens are probably not the people most sensitive to a citizenship question, since most of them are legal residents. Those lacking an SSN should presumably be even more sensitive to a citizenship question, so the AR noncitizen definition may exclude much of the sensitive population.[44] In our second dichotomy the less sensitive group is "AR & ACS all-citizen households", those households where all persons reported in the ACS to be living in the household at the time of the survey are AR citizens, and all are self-reported as being citizens in the ACS as well. The more sensitive group is "all other households", including those households where some residents are both AR citizens and self-reported citizens but at least one is not; there is a mismatch between the survey report and administrative record response; or citizenship status is not reported in one or both sources. We assume AR & ACS all-citizen households are less sensitive to a citizenship question than all other households, since they have demonstrated a willingness to provide citizenship status answers for all household members, those answers are consistent with administrative records and thus likely truthful responses,[45] and citizens presumably have less to fear about revealing their status than noncitizens. In comparison to others, more of this group's reluctance to self-respond to the ACS should be due to reasons other than the citizenship question, such as unwillingness to answer a longer questionnaire. Note that if some of the reluctance by AR & ACS all-citizens households to self-respond is due to the citizenship question in the ACS, then our analysis will underestimate the citizenship question unit self-response effect.

The sample size for the second set of groups is significantly larger than that for the first set of groups, because the first set excludes households where no persons are AR noncitizens at the address, but at least one person assigned to that address by administrative records cannot be linked to the Numident.

Table 6 displays unweighted 2010 Census and ACS response rates for the AR all-citizen households and AR noncitizen household groups. The self-response rate is higher for the 2010 Census than for the ACS for both citizenship categories, presumably reflecting the higher burden of the ACS. The all-citizen response rate is greater than the noncitizen rate in each survey, suggesting that noncitizen households have a lower participation rate in general. Most important for this study is understanding how the difference in self-response rate across groups varies

[43] Another reason to use the survey household roster rather than the AR crosswalk is that the AR crosswalk often places people in different locations. Rastogi and O'Hara (2012) report that among the 279.2 million persons in the 2010 Census who could be assigned a PIK, 27.2 percent are assigned to an address in the AR crosswalk that differs from their Census address.

[44] This is consistent with Camarota and Capizzano (2004), who say field representatives reported that illegal immigrants were less likely to respond than other foreign-born persons. Illegal immigrants are ineligible for SSNs.

[45] As shown in Section 4 above, when an administrative record shows that someone is a citizen, the ACS response is nearly always citizen as well, giving us a high degree of confidence that the person truly is a citizen.

between the 2010 Census and ACS. While the self-response rate for citizen households is 13.8 percentage points lower in the ACS than in the 2010 Census, the self-response rate for households with at least one noncitizen is 18.9 percentage points lower for the ACS than the self-response rate to the 2010 Census, which is a 5.1 percentage point difference between the two categories.

Table 6. Comparison of 2010 ACS to 2010 Census Response Rates with Initial Assumptions

	Self-Response Rate (%)		Difference
	2010 ACS	2010 Census	
Households with at least one AR noncitizen	52.6	71.5	-18.9
	(0.21)	(0.19)	(0.26)
AR all-citizen households	66.1	79.9	-13.8
	(0.05)	(0.04)	(0.06)
Difference-in-differences			-5.1
			(0.26)

Source: 2010 ACS 1-year file, 2010 Census Unedited File (CUF), and 2010 Numident.

Notes: 2010 CUF self-response is non-blank response to the first mailing, and only NRFU-eligible housing units are included. ACS self-response is mail response. All persons in the 2010 Numident that are missing citizenship are treated as citizens here. Robust standard errors are in parentheses, calculated from regressions. The estimates are unweighted. Around 5.9 percent of the households have at least one noncitizen. The sample size is 929,000. DRB clearance number CBDRB-2017-CDAR-001.

Using survey weights can facilitate comparisons of results across years, since sampling can change, and we would like to be able to project results forward in time. We thus display weighted response rates in Table 7, now both for the first and second sets of groups. As expected, the restriction to being a citizen in both the AR and ACS results in higher self-response rates in the AR & ACS all-citizen household group compared to the AR all-citizen household group. The response rates for the two noncitizen groups differ little from each other. The difference-in-differences estimate for the first set of groups increases to 8.9 percentage points compared to the unweighted gap in Table 6. It is three percentage points higher (11.9) across the second set of groups.

Table 7. Comparison of 2010 ACS to 2010 Census Response Rates (Weighted)

	Self-Response Rate (%)		Difference
	2010 ACS	2010 Census	
Households with at least one AR noncitizen	42.4	62.1	-19.7
	(0.32)	(0.18)	(0.26)
AR all-citizen households	62.0	72.8	-10.8
	(0.34)	(0.11)	(0.24)
Difference-in-differences			-8.9
			(0.35)
All other households	42.0	62.7	-20.7
	(0.32)	(0.14)	(0.25)
AR & ACS all-citizen households	65.6	74.4	-8.9
	(0.33)	(0.11)	(0.24)
Difference-in-differences			-11.9
			(0.34)

Source: 2010 ACS 1-year file, 2010 Census Unedited File (CUF), and 2010 Numident.

Notes: 2010 CUF self-response is non-blank response to the first mailing, and only NRFU-eligible housing units are included. ACS self-response is mail response. The standard errors are in parentheses. The standard errors for the self-response rates and differences are calculated using Fay's balanced repeated replication variance estimation method, with 80 replicate weights, adjusting the original weights by a coefficient of 0.5. The difference-in-differences (*DiD*) standard errors (*SE*) are calculated as $DiD\ SE = \sqrt{SE(Est_1)^2 + SE(Est_2)^2}$, where the two estimates (*Est*) are the 2010 Census – 2010 ACS differences for the two groups. The estimates use ACS housing unit weights. 88.2 percent of households are in the AR all-citizen household group vs. 11.8 percent in the households with at least one AR noncitizen group. 74.9 percent are in the AR & ACS all-citizen household group vs. 25.1 percent are in the all other households group. The number of observations is 1,418,000.

The larger decline in self-response rates for the AR noncitizen household and all other households groups may not actually be due to greater sensitivity. Other characteristics besides citizenship status could be associated with lower ACS self-response, and the AR noncitizen household and all other households groups could have a higher propensity to have such characteristics. To explore this possibility, we perform Blinder-Oaxaca decompositions (Blinder 1973 and Oaxaca 1973).[46]

Households may belong to one of two groups $G \in (S, U)$, where the *S* group is thought to be potentially sensitive to a citizenship question, while the *U* group is not. We set the self-responses $R_{G_iACS_t}$ and $R_{G_iCensus_t}$ equal to 100 if household *i* in group *G* self-responds in year *t* to the ACS and Census, respectively, and zero otherwise.[47] The difference between the survey responses is

$$\Delta R_{G_it} = R_{G_iACS_t} - R_{G_iCensus_t} \quad (1)$$

The vector of predictors *X* includes household size and reference person characteristics (sex, race/ethnicity, age, educational attainment, household income, working in the last week, job search

[46] This method was initially developed to study the extent to which the gender wage gap is due to different distributions of characteristics associated with wages by gender (explained variation) vs. differing behavior across gender for a given set of characteristics (unexplained variation). The unexplained variation is usually attributed to discrimination, but it also captures any effects of differences in unobserved variables.

[47] We use 100 for response so that the results are expressed in percentages.

in the last four weeks, and English language ability among those speaking a language other than English at home). β contains the slope parameters and intercept, and ε is an error term with mean zero.

We estimate OLS models for each household group

$$\Delta R_{S_{it}} = X'_{S_{it}}\beta_{S_t} + \varepsilon_{S_{it}} \quad (2)$$

$$\Delta R_{U_{it}} = X'_{U_{it}}\beta_{U_t} + \varepsilon_{U_{it}} \quad (3)$$

The difference-in-differences in expected self-response rates across the two surveys for the two groups *S* and *U* in year *t* is

$$\Delta\Delta R_{SU_t} = E\left(\Delta R_{S_t}\right) - E\left(\Delta R_{U_t}\right) \quad (4)$$

We decompose this as follows:

$$\Delta\Delta R_{SU_t} = \left[E\left(X_{S_t}\right) - E\left(X_{U_t}\right)\right]'\beta_{U_t} + \left[E\left(X_{S_t}\right)'\left(\beta_{S_t} - \beta_{U_t}\right)\right] \quad (5)$$

The first term (explained variation) applies the coefficients for the unsensitive group to the difference between the expected value of the sensitive group's predictors and those of the unsensitive group. The second (unexplained variation) is the difference between the expected value of the sensitive group's predictors applied to the sensitive group's coefficients and the same predictors applied to the unsensitive group's coefficients. The interpretation that the unexplained variation represents the citizenship question effect is dependent on the assumption that there are no unobserved variables relevant to the difference-in-differences in self-response across the two surveys.

Table 8 shows the results of the Blinder-Oaxaca decomposition for the two sets of groups. In the AR all-citizen vs. AR noncitizen comparison, virtually all the difference-in-differences is explained by differences in predictors across the two groups. Thus, it appears that the larger fall in self-response to the ACS vs. the Census for AR noncitizen households is not due to sensitivity to the citizenship question, but rather that AR noncitizen households have a greater propensity to have other characteristics that are associated with lower ACS self-response. In contrast, about half (6.1 percentage points) of the difference-in-differences for the AR & ACS all-citizen vs. all other household comparison is unexplained, suggesting that the larger drop-off in ACS self-response for all other households is partly due to sensitivity to the citizenship question. Appendix Table A13 shows the regression coefficients for equations (2) and (3), and the explained variation and unexplained variation coefficients for each predictor are shown in Appendix Table A14.

Table 8. Blinder-Oaxaca Decomposition of Comparison of Predicted 2010 ACS to 2010 Census to Response Rates by Households Citizenship Type

	2010 ACS – 2010 Census
Households with at least	-19.7
one AR noncitizen	(0.13)
AR all-citizen households	-10.8
	(0.12)
Difference-in-differences	-8.9
	(0.09)
Explained	-8.7
	(0.11)
Unexplained	-0.2
	(0.13)
All other households	-20.7
	(0.12)
AR & ACS all-citizen households	-8.9
	(0.12)
Difference-in-differences	-11.9
	(0.07)
Explained	-5.8
	(0.14)
Unexplained	-6.1
	(0.16)

Source: 2010 ACS 1-year file, 2010 Census Unedited File (CUF), and 2010 Numident.

Notes: 2010 CUF self-response is non-blank response to the first mailing, and only NRFU-eligible housing units are included. ACS self-response is mail response. The standard errors are in parentheses. The standard errors are bootstrapped using 80 ACS replicate weights. The number of observations is 1,418,000.

To see how changes in predictors over time affect the magnitude of the unexplained variation (*UV*) in the decomposition, we apply the coefficients from the 2010 models to the predictors in the 2016 ACS

$$UV_{2016} = E\left(X_{S_{2016}}\right)' \beta_{S_{2010}} - E\left(X_{S_{2016}}\right)' \beta_{U_{2010}} \quad (6)$$

Table 9 shows that the unexplained variation is still insignificant for the AR all-citizen vs. AR noncitizen comparison. It is of a similar magnitude in 2016 as in 2010 (5.8 percentage points vs. 6.1) for the AR & ACS all-citizen vs. all other household comparison. Note that this does not capture changes over time in the degree of sensitivity to a citizenship question for a housing unit with a fixed set of characteristics. That would require estimating models on fresher data of surveys with and without a citizenship question for the same households.

Table 9. Comparison of Predicted 2016 ACS to 2010 Census Response Rates for AR Noncitizen and All Other Households with Their Own vs. All-Citizen Models

	2016 ACS – 2010 Census
Model\Sample	AR noncitizen household sample
AR noncitizen household model	-19.7
	(0.47)
AR all-citizen household model	-20.5
	(0.34)
Difference-in-differences	0.8
	(0.58)
Model\Sample	All other household sample
All other household model	-21.7
	(0.33)
AR & ACS all-citizen household model	-15.9
	(0.39)
Difference-in-differences	-5.8
	(0.51)

Source: 2016 ACS 1-year file and 2016 Numident.

Notes: 2010 Census self-response is non-blank response to the first mailing, and only NRFU-eligible housing units are included. ACS self-response is mail response. The standard errors are in parentheses. The standard errors for the 2010 Census – 2016 ACS response differences are calculated using Fay's balanced repeated replication variance estimation method, with 80 replicate weights, adjusting the original weights by a coefficient of 0.5. The difference-in-differences (*DiD*) standard errors (*SE*) are calculated as $DiD\ SE = \sqrt{SE(Est_1)^2 + SE(Est_2)^2}$, where the two estimates (*Est*) are the 2010 Census – 2016 ACS differences for the two groups. The estimates use ACS housing unit weights. 28.6 percent are in the all other households group in 2016. The standard errors are in parentheses. They are the standard errors of the model predictions, based on the bootstrapped regressions in Appendix Table A12 that use 80 ACS replicate weights. The number of observations is 163,000 for the AR noncitizen household sample and 477,000 for the all other household sample.

Though suggestive, these exercises and the ones performed below are not perfect laboratories for studying the self-response effect of inclusion of a citizenship question on the 2020 Census. The ACS contains 75 questions, so any one question is unlikely to stand out, whereas an added question will be more visible in the 2020 Census questionnaire, which contains just 10 other questions.[48] Thus, we would ideally want to compare response rates on a short questionnaire without a citizenship question to one adding just the citizenship question. Second, the level of concern about using citizenship data for enforcement purposes may be very different in 2020 than it was in 2000 or 2010, so a more recent test would be preferable. These factors suggest the estimated effect on self-response from the exercise in Table 9 is conservative.

[48] A preferable test would be a randomized control trial (RCT) comparing self-response rates where some households are randomly chosen to have an 11-question Census questionnaire with a citizenship question (the treated group), and a randomly chosen set of control households receive a 10-question Census questionnaire without citizenship.

As robustness checks we do similar exercises below with the 2000 Census and the 2014 Survey of Income and Program Participation (SIPP), a longitudinal survey that follows the same individuals over time. Unlike the decennial census and the ACS, individuals respond for themselves in the SIPP. The 2000 Census long form (sent to one of every six housing units, selected randomly) contained a citizenship question among many other additional questions, while the short form (sent to the remaining housing units) did not. As in the first set of groups above, we divide housing units into those with all citizens and those with at least one noncitizen, based on citizenship data from the 2002 Numident for persons enumerated at those housing units in the 2000 Census.[49] As with the 2010 ACS and Census exercises, Table 10 shows that self-response rates are higher in the short form than the long form, and they are higher in households with all citizens. The short- vs. long-form difference in response rates is greater for households with at least one noncitizen by 3.3 percentage points, again consistent with the possibility that households with noncitizens are more sensitive to the inclusion of citizenship questions.

Table 10. 2000 Census Long Form and Short Form Analysis

	Self-response rate (%)		
Households by Citizen	Long Form	Short Form	Difference
At Least One Noncitizen	62.5	71.0	-8.5
	(0.017)	(0.016)	(0.023)
All Citizens	76.1	81.3	-5.2
	(0.005)	(0.004)	(0.006)
Difference	13.6	10.3	-3.3
	(0.017)	(0.016)	(0.024)

Source: 2000 Census short and long forms.

Notes: These are weighted using housing unit weights. The number of short forms is 105.5 million, and the number of long forms is 16.4 million. The definition of self-response is mail response here. Robust standard errors are in parentheses, generated from weighted regressions of response on an interaction of the household citizenship status with short form. The standard errors for the differences are calculated as $SE(Est_1 - Est_2) = \sqrt{SE(Est_1)^2 + SE(Est_2)^2}$.

Longitudinal data provide another means for understanding response sensitivity to questions of citizenship. Using the 2014 Survey of Income and Program Participation (SIPP) longitudinal panel waves 1 and 2, we show how nonresponse changes from Wave 1 to Wave 2 for noncitizen respondents, as well as for households with at least one noncitizen. The first row in Table 11 shows nonresponse rates for noncitizens from the 2014 Survey of Income and Program Participation (SIPP) Waves 1 and 2. Noncitizens made up around 6 percent of the 2014 SIPP survey in Wave 1. The proportion of noncitizens in Wave 2 decreased slightly, implying that noncitizens were more

[49] To be classified as a housing unit with all citizens in this exercise, all persons must be linked to the Numident. A housing unit can be classified as having at least one noncitizen if there is at least one person linked to the Numident who is a Numident noncitizen, whether or not all the other persons in the housing unit could be linked to the Numident or not.

likely to leave the survey due to attrition or other factors than citizens. In addition, the rate of nonresponse among those households with at least one noncitizen increased from Wave 1 to Wave 2, from 7.9 percent to 8.5 percent. While noncitizens were more likely to drop out of the survey, those who stayed were more likely to live in households where at least one member did not respond. These data provide additional hints of the potential future impact to nonresponse for noncitizens in surveys that ask about citizenship status.

Table 11. Noncitizens and Nonresponse in the 2014 Survey of Income and Program Participation

	Wave 1		Wave 2	
	(%)	(se)	(%)	(se)
Noncitizens	6.1	(0.144)	5.7	(0.174)
At least one member in the noncitizen household did not respond	7.9	(0.473)	8.5	(0.537)

Source: 2014 SIPP, Waves 1 and 2

Notes: Citizenship status refers to status in Wave 1. The standard errors are clustered in Wave 2. These estimates are run on the internal run 16 version of the 2014 SIPP.

7. Effects of Citizenship Question on Nonresponse Follow-up Costs and Enumeration Quality

A drop in the self-response rate from adding a citizenship question in Alternatives B (obtaining citizenship from the 2020 Census only) and D (obtaining citizenship from the 2020 Census and administrative records) results in increased costs in the Nonresponse Follow-up (NRFU) operation and affects the quality of the population count. Households deciding not to self-respond because of the citizenship question are likely to refuse to cooperate with enumerators coming to their door in NRFU, resulting in the use of neighbors as proxy respondents on their behalf. [50] As shown in Table 12, Mule (2012) reports that the correct enumeration rate is 27.1 percentage points lower for proxies than mail in self-responses based on data from the 2010 Census Coverage Measurement (CCM) survey. The person linkage rate is 62.9 percentage points lower for proxies than for mail in self-responses in the 2010 Census, according to Rastogi and O'Hara (2012). Both these studies provide suggestive evidence that proxies supply poor quality individual demographic and socioeconomic characteristic information about the person on behalf of whom they are responding.

[50] A proxy response is a response about the household by someone outside the household, such as a neighbor or property manager. The enumerator will seek a proxy response for households that don't mail back their Census questionnaire or give an in-person interview after several attempts.

Table 12. Enumeration Quality in Mailout/Mailback and Nonresponse Follow-up (NRFU) Proxy Responses

	Mailout/Mailback Response	NRFU Proxy
Correct Enumerations	97.3	70.2
Erroneous Enumerations	2.5	6.7
Whole-Person Census Imputations	0.3	23.1
Person Linkage Rate	96.7	33.8

Source: Mule (2012) for correct enumerations, erroneous enumerations, and whole-person Census imputations, and Rastogi and O'Hara (2012) for the person linkage rate.

We provide two sets of estimates, the first based on our initial assumptions (in parentheses), and a second based on revised assumptions. The main changes in the revised assumptions are an expansion of the group of housing units considered potentially sensitive to a citizenship question and the estimated percentage of them who will not respond to a questionnaire due to the presence of a citizenship question (5.8 percent in Table 9 vs. 5.1 percent in Table 6).

Using these estimates as well as the data in Table 12, we can develop cautious estimates of the data quality and cost consequences of adding the citizenship question to the enumeration form. We assume that all-citizen households are unaffected by the change and that an additional 5.8 percent (5.1 percent) of households that possibly have noncitizens go into NRFU because they do not self-respond.[51] We expect 320 million persons in 126 million occupied households in the 2020 Census.[52] Based on a combination of administrative records from the 2016 Numident and ITINs and the 2016 ACS, we estimate that 28.6 percent (9.8 percent) of all households could potentially contain at least one noncitizen. Combining these assumptions implies an additional 2,090,000 households (630,000 households) and 6.5 million persons (1.6 million persons) in NRFU.[53] 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 561,000 (139,000) fewer correct enumerations, of which 185,000 (46,000) are additional erroneous enumerations and 376,000 (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 percent of the time, and 21.0 percent receive proxy responses. If all of these new NRFU cases go to proxy responses instead,[54] the result would be 1,750,000

[51] Recall that the initial estimate is based on households with at least one AR noncitizen, which is only a fraction of the housing units in the all other households category, which also includes persons with missing citizenship in AR or the ACS or citizenship values that conflict between AR and the ACS.

[52] We assume 10 million residents of group quarters. Group quarters are not included in either mailout/mailback or NRFU operations, and here we assume no effect of a citizenship question on their enumeration.

[53] The initial assumption here is that average household size for households with at least one noncitizen is the same as the forecast for all households in the 2020 Census (2.54 persons). The revised assumption is that average household size for all other households is the same as its average in the 2016 ACS, 3.1 persons.

[54] If a household declines to self-respond due to the citizenship question, we suspect it would also refuse to cooperate with an enumerator coming to their door, resulting in a need to use a proxy.

(432,000) fewer correct enumerations, of which 272,000 (67,000) are erroneous enumerations, and 1,477,000 (365,000) are whole-person census imputations.[55] The number of persons who are linkable to administrative records would fall by 4.1 million (1 million).

Our estimate of the incremental cost proceeds as follows. Using the analysis in the paragraph above, the estimated NRFU workload will increase by approximately 2,090,000 households (630,000 households), or approximately 1.66 percentage points (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 $91.2 million ($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.

8. Distribution of 2020 Citizenship Data Sources by Collection Method

Figures 10-12 provide forecasts of how many U.S. residents in the 2020 Census acquire their citizenship data from survey responses, administrative records, and model-based imputation methods in Alternatives B, C, and D. Once again we provide forecasts based on initial and revised assumptions, with initial forecasts in parentheses.[56] A reduction in self-response rates and increase in proxy responses from adding the citizenship question in Alternatives B and D is likely to affect the number of persons with survey responses for citizenship. As shown above, reference persons are much less likely to answer the citizenship question for nonrelatives in the household than for themselves, so they may be even less likely to answer it for neighbors. In order to obtain a range of estimates based on best and worst case scenarios, Figure 10 Panel A and Figure 12 Panels A and B assume that proxies report citizenship at the same rate as they do in the 2010 ACS relative to all persons in the 2010 ACS,[57] while Figure 10 Panels B and C and Figure 12 Panels C and D assume none of the proxies report citizenship.

We begin with the estimated 2020 Population of 330 million, the total number of persons we expect to count in the 2020 Census. Under Alternative B with complete citizenship data from proxy

[55] These enumeration errors may not be avoidable simply by spending more money on fieldwork. Once a household decides not to cooperate, it may not be possible to obtain an accurate enumeration no matter how many times an enumerator knocks on their door.

[56] In addition to the differences between the initial and revised assumptions mentioned in Section 7, two others are relevant here. One is that the initial assumptions classify foreign-born persons with missing citizenship in the Numident and without an ITIN as AR citizens, while the revised assumptions classify them as having missing AR citizenship. A second is that instead of showing the difference in the AR linkage rate with and without a citizenship question in the 2020 Census as an increase in the AR linkage rate in Alternative C, the revised assumptions show it as a decrease in the AR linkage rate in Alternative D.

[57] Within 2010 ACS households that have NRFU proxy responses in the 2010 Census, the nonmissing citizenship rate is 96.7 percent, vs. 97.1 percent for all ACS households. We apply this proxy to total sample ratio to the 93.7 percent nonmissing citizenship rate in the 2016 ACS to get an estimated 92.9 percent nonmissing citizenship rate for proxies in 2020.

responses, 309.1 million citizenship responses are obtained from the Census. Applying the missing citizenship rate of 6.3 percent in the 2016 ACS, we expect 20.9 million to have missing data for the citizenship question, either because the respondent skipped the question, or because a proxy response in nonresponse follow-up (NRFU) did not deliver information on that question. Citizenship is imputed using models for these 20.9 million persons.[58] With no citizenship data from proxy responses, the number of citizenship responses drops to 290 million (294.6 million), with 40 million (35.4 million) modeled.[59] The accuracy of this imputation system is unknown at this time. As discussed above, the imputation will be challenging due to the fact that nonresponse is highly correlated with citizenship.

Under Alternative C, we expect to link 289.6 million (295.0 million) to administrative records containing citizenship data, applying the linkage rate for the 2016 ACS to currently available administrative records.[60] The remaining 40.4 million (35.0 million) will have citizenship imputed using models based on the variables common to the linked and non-linked portions of the data. At this time, the accuracy of that imputation system is not known, but it would be based on the administrative record citizenship variable, so it would not be subject to the biases caused by survey citizenship reporting issues.

Of the 309.1 million who provide valid responses to the Census citizenship question in Alternative D, we expect to link 269.6 million (272.5 million) records to the administrative data.[61] Of these, the vast majority, 260.9 million (263.0 million), will have administrative record and Census responses that agree (applying the 2016 ACS-AR agreement rate of 96.8 percent), and since the agreement is with the same administrative record system as in Alternative C, these people will have the same citizenship status under either alternative. Of the 269.6 million (277.4 million) linked Census responses with a valid answer to the 2020 Census question, we expect the administrative record and the Census response to disagree for 8.7 million (9.7 million). These are the persons for whom we have two choices: (1) accept the Census questionnaire answer or (2)

[58] General imputation models develop a response for those who did not respond using all available relevant data.

[59] Based on the analysis in Table 9, under our revised assumptions we project 6.5 million additional proxy responses due to the citizenship question, of which an estimated 840,000 already have missing citizenship (applying the allocation rate of 13.0 percent from the 2016 ACS among persons who do not both report being citizens and are AR citizens). This is in addition to an estimated 14.5 million proxy responses in 2020 without a citizenship question, of which an estimated 1,030,000 already have missing citizenship (applying the 2016 ACS citizenship item allocation rate of 6.3 percent among all ACS-AR citizenship groups, adjusted by the ratio of the 2010 ACS citizenship allocation rate for 2010 Census proxy respondents (3.3 percent) to the 2010 ACS citizenship allocation rate for the whole 2010 ACS sample (2.9 percent)). Note that the proxy responses that are anticipated to occur in 2020 regardless of presence of a citizenship question may happen in households containing people in any ACS-AR citizenship group, whereas the additional proxies due to the citizenship question are assumed to come from housing units where people are not in the group with both ACS and AR citizen responses.

[60] As discussed in Section 7, our initial estimate of the effect of a citizenship question on the number of linkable persons is 1 million, and the revised estimate is 4.1 million. Our initial estimate adds 1 million to the number of linked persons when no citizenship question is included in the questionnaire. We incorporate the change in the number of linkable persons as a reduction in AR linkage in Alternative D for our revised estimate, as discussed below.

[61] When applying the 2016 ACS linkage to administrative record citizenship rate, the estimate is 273.4 million persons with linked citizenship. Of the 4.1 million anticipated reduction in linkage due to the citizenship question in our revised estimate, about 3.9 million are applied to the group with observed 2020 citizenship, as 93.7 percent of persons are anticipated to have observed 2020 citizenship (applying the missing citizenship rate in the 2016 ACS).

replace the questionnaire answer with the administrative answer. If we do the former, all of these cases will differ from the Alternative C answer. The estimated direct response is U.S. citizen for 7.6 million (7.7 million) of these persons, compared to 1.1 million (2.0 million) in the administrative records. Use of direct responses for those with disagreement would result in a projected 6.5 million (5.7 million) more U.S. citizens than when using administrative records.[62]

Continuing with Alternative D, we would process the 20.9 million responses where we did not get a valid answer to the Census citizenship question as in Alternative C. This would result in 16.0 million (16.6 million) persons for whom we expect to find an answer in the administrative records, and 4.9 million (4.3 million) for whom we would use a modeled answer.[63] The models would be developed using the same methods as in Alternative C, but not the same input data, because of the change in response behavior associated with asking the citizenship question.

When 2020 citizenship is observed in Alternative D, but the record cannot be linked to administrative data, we would accept the survey response for an expected 39.5 million (31.7 million) people. The number of persons whose records can be linked to administrative data is lower by 4.1 million (10.7 million) in Alternative D than in Alternative C due to poorer linkage quality from proxy responses, which would have been self-responses without a citizenship question (see Table 10). This captures the negative effect of inclusion of the citizenship question on the ability to use administrative data for citizenship.

When we assume that none of the proxy responses report citizenship, the number where 2020 citizenship is observed falls to 289.5 million (294.6 million) in Alternative D, just as in Alternative B. 263.4 million (272.5 million) of these are linked to administrative record citizenship, 255.6 million (263.0 million) of those answers agree between sources, and 7.8 million (9.5 million) disagree. The direct response for the latter group is U.S. citizen for 6.8 million (7.5 million) vs. 1.0 million (2.0 million) U.S. citizens in administrative records, leading to a 5.8 million (5.6 million) higher count of U.S. citizens if direct responses are used.

Of the 26.6 million (22.2 million) persons for whom 2020 citizenship is observed, but the record cannot be linked to administrative data, we estimate that about 560,000 (500,000) noncitizens will respond as citizens, based on the AR noncitizens reporting as ACS citizens share of the 2016 ACS (2.3 percent in the initial estimates and 2.1 percent in the revised estimates).

These results show that there is a tendency for persons missing citizenship in one source to also be missing it in the other. Among persons with observed 2020 Census citizenship in Figure 12 Panel D, 90.8 percent have AR citizenship, while only 55.5 percent of those without 2020 Census citizenship have AR citizenship. Of those with AR citizenship, 92.2 percent have 2020 Census citizenship, but just 59.9 percent of those without AR citizenship have 2020 Census citizenship. The correlated missingness reduces the coverage gain from using multiple sources. Only 22.2 million persons' citizenship values can be covered by AR among those without 2020 Census

[62] To put this in context, the 2016 ACS estimates that 22.5 million U.S. residents are noncitizens, or 7.0 percent of the population.

[63] Here we apply the remainder of the anticipated 4.1 million reduction in linkage to administrative record citizenship due to increased proxy response to the group for which 2020 citizenship is not observed.

citizenship, whereas AR coverage would be 34.6 million if the missingness correlation were zero. Analogously, just 26.6 million persons missing AR citizenship have 2020 Census citizenship, vs. 39.0 million if the correlation were zero.

Across the three alternatives, the data for at least 255.6 million (263.0 million) persons would be identical, and it would be identical for at least 276.9 million (284.3 million) between alternatives C and D. If the administrative record response is used when the cases disagree, then the data for alternatives C and D would agree for 285.6 million (294.0 million) linked cases.

Alternative C results in more persons with modeled citizenship responses, while Alternative D has fewer imputations. If no proxy respondents report citizenship, then Alternative B has about the same number of imputations as Alternative C, but otherwise its level is in between that of Alternatives C and D.

As mentioned above, the estimated reduction in self-response due to the inclusion of a citizenship question is based on a comparison of a long 2010 ACS questionnaire to a short 2010 Census questionnaire. The visibility of the citizenship question may be more prominent when added to a short questionnaire, resulting in a larger reduction in self-response than what we have estimated here. If the assumption that all proxy responses result in citizenship item nonresponse is accurate, every additional person without Census citizenship will have to have modeled citizenship in Alternative B. With Alternative D, fewer of the additional nonresponses will be modeled, as some can be linked to administrative record citizenship data. The option to use administrative records in Alternative D thus partially mitigates the citizenship question self-response effect.

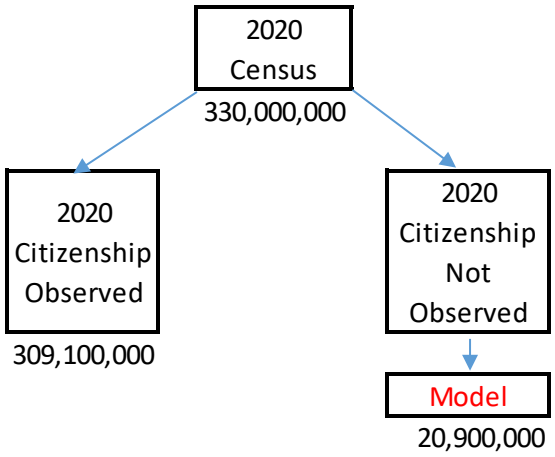
These estimates are based on currently available administrative record citizenship data and linkage capability. The Census Bureau may obtain several additional sources by 2020 and develop better linkage, in which case administrative record coverage may be higher than that shown here. This would lead to fewer imputations in Alternative D and especially Alternative C. The number of imputations in Alternative C is not much higher than in Alternative B, so even a small improvement in administrative record citizenship data coverage would lead to a lower imputation rate in Alternative C than B. Alternative D's advantage in coverage over Alternative C would shrink, though it is unlikely to vanish completely.

A key question when comparing Alternatives C and D is whether the data quality is higher for the 2020 Census or for imputed values for the persons with imputations in Alternative C and observed 2020 Census data in Alternative D. Survey citizenship data exhibit a markedly higher U.S. citizen share compared to administrative records for persons with both sources, but it is unknown whether that tendency also applies to persons without links to administrative records.

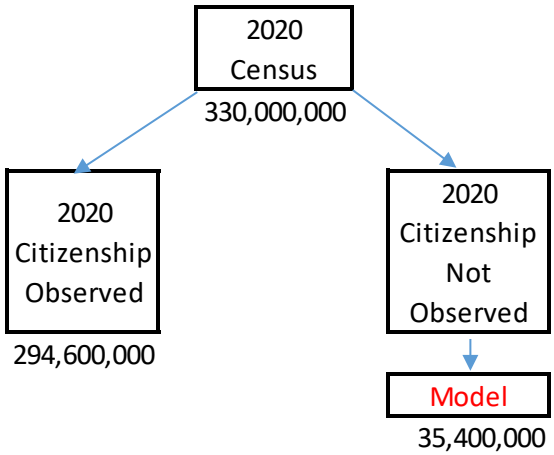
A second question is what data source(s) to use when administrative records and the survey response disagree in Alternative D. Citizenship status is verified via documentation from the issuing government agencies in the administrative records data, but not in the survey, and the analysis in Section 4 above exhibits patterns suggesting that the survey responses are more often inaccurate when they disagree. On the other hand, using administrative records when the sources disagree would mean that the survey response contribution to the citizenship statistics would be minor – it would only be necessary for persons without linked administrative record citizenship

data. The 2020 Census citizenship data is the sole source for 8.1 percent (6.7 percent) of persons in Figure 12 Panel D (Panel C), and this share could be smaller if administrative record coverage improves or survey coverage is lower than estimated. It could be difficult to justify burdening respondents with this question if needed for only a small fraction of the population.

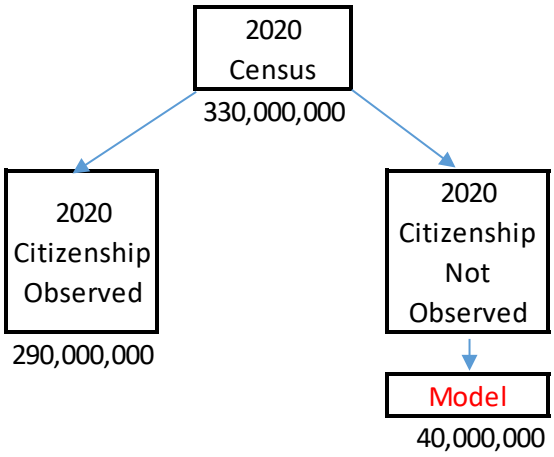
Figure 10. Alternative B



Panel A. Alternative B, Proxies Report Citizenship

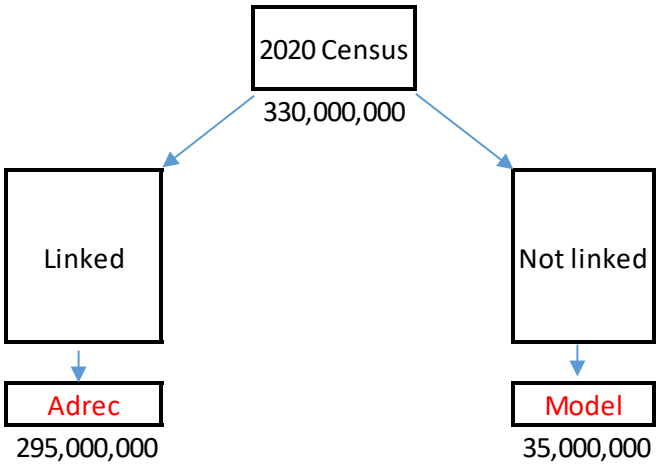


Panel B. Alternative B, Proxies Don’t Report Citizenship, Initial Assumptions

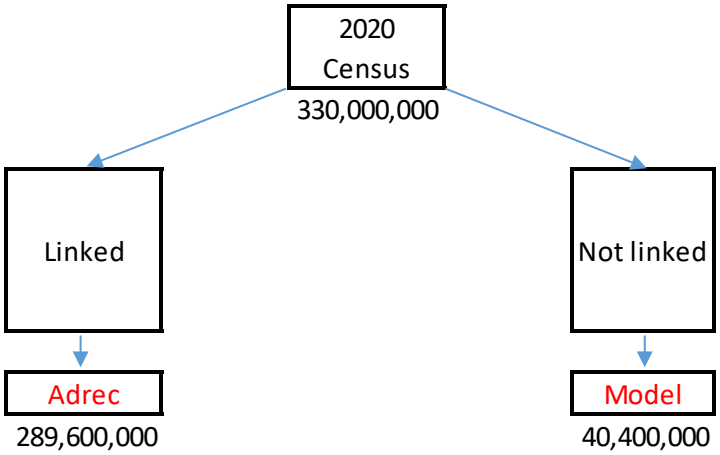


Panel C. Alternative B, Proxies Don’t Report Citizenship, Revised Assumptions

Figure 11. Alternative C

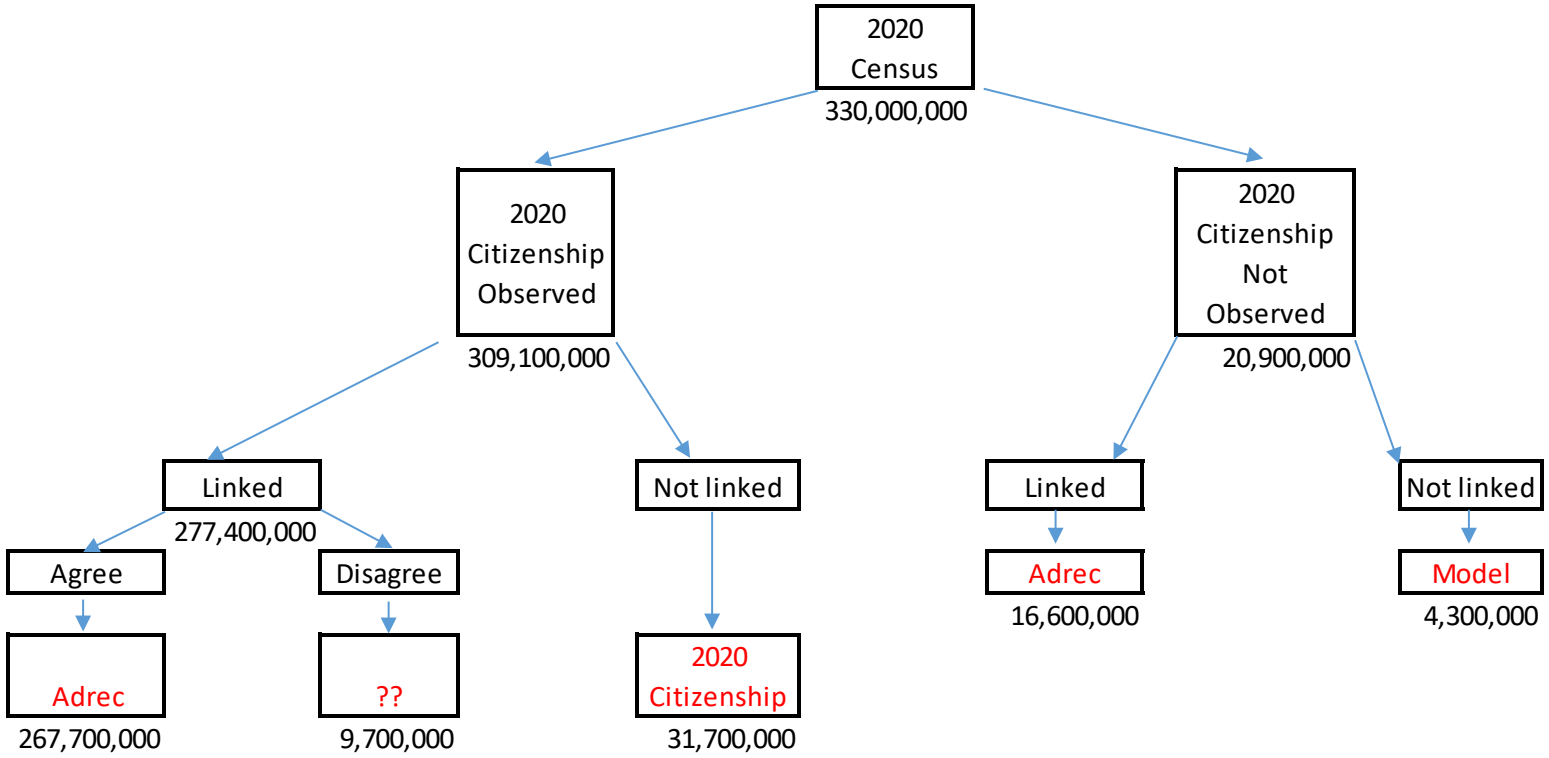


Panel A. Initial Assumptions

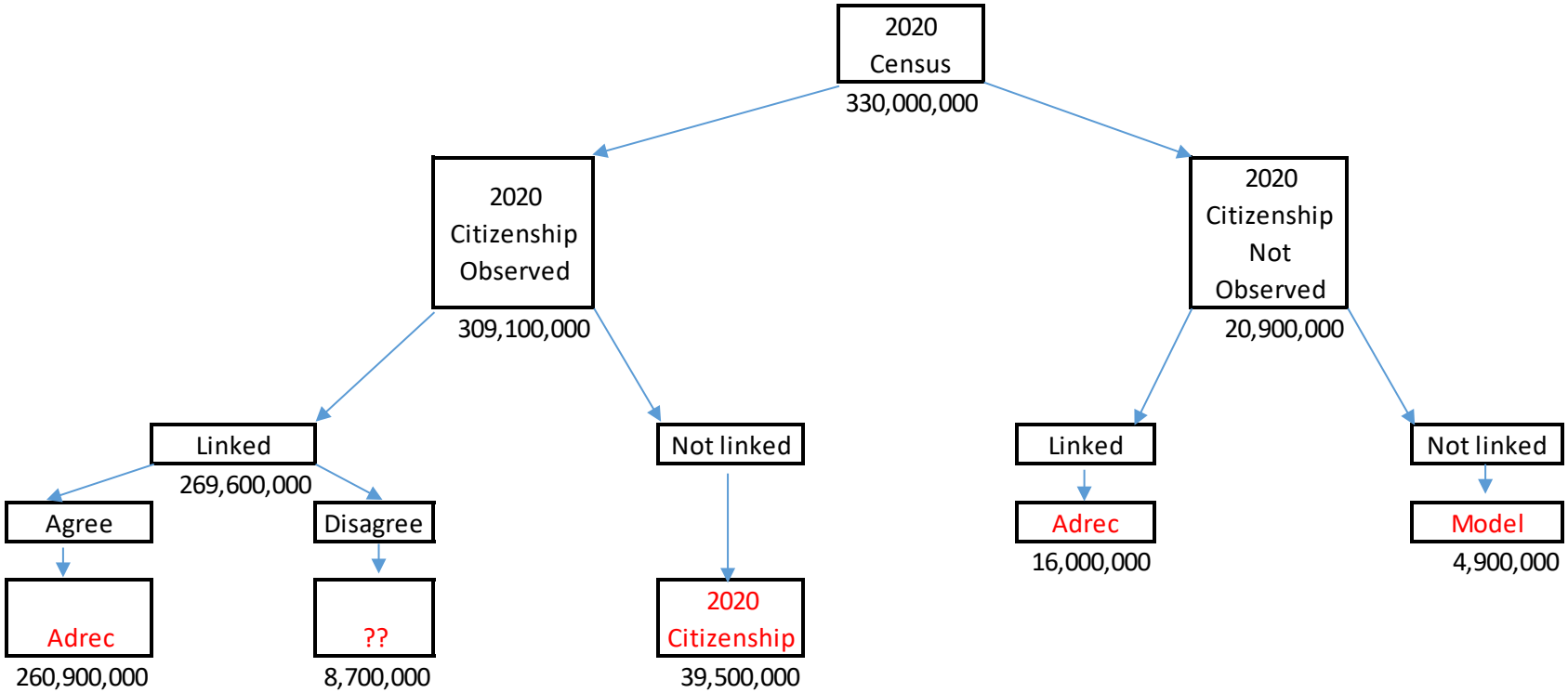


Panel B. Revised Assumptions

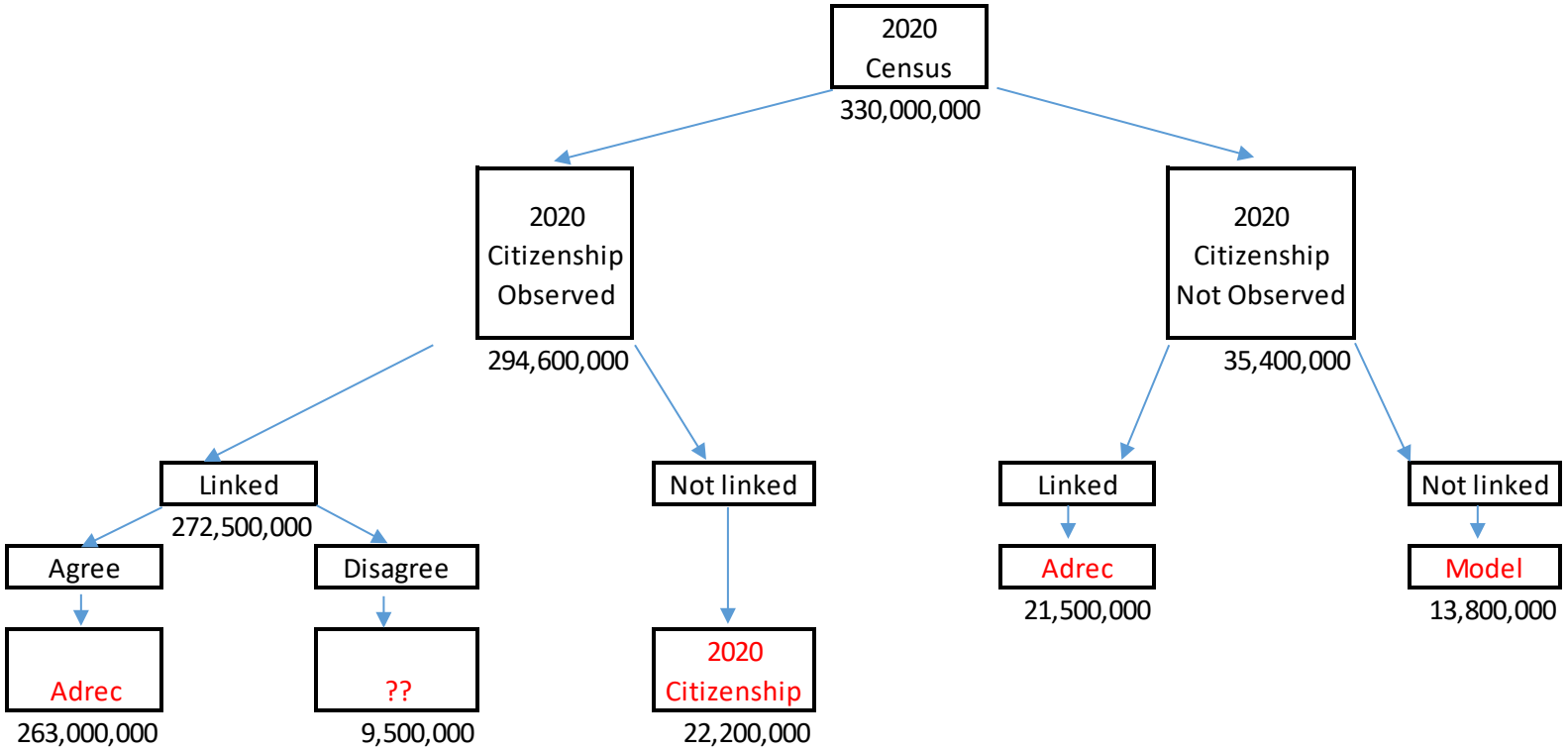
Figure 12. Alternative D



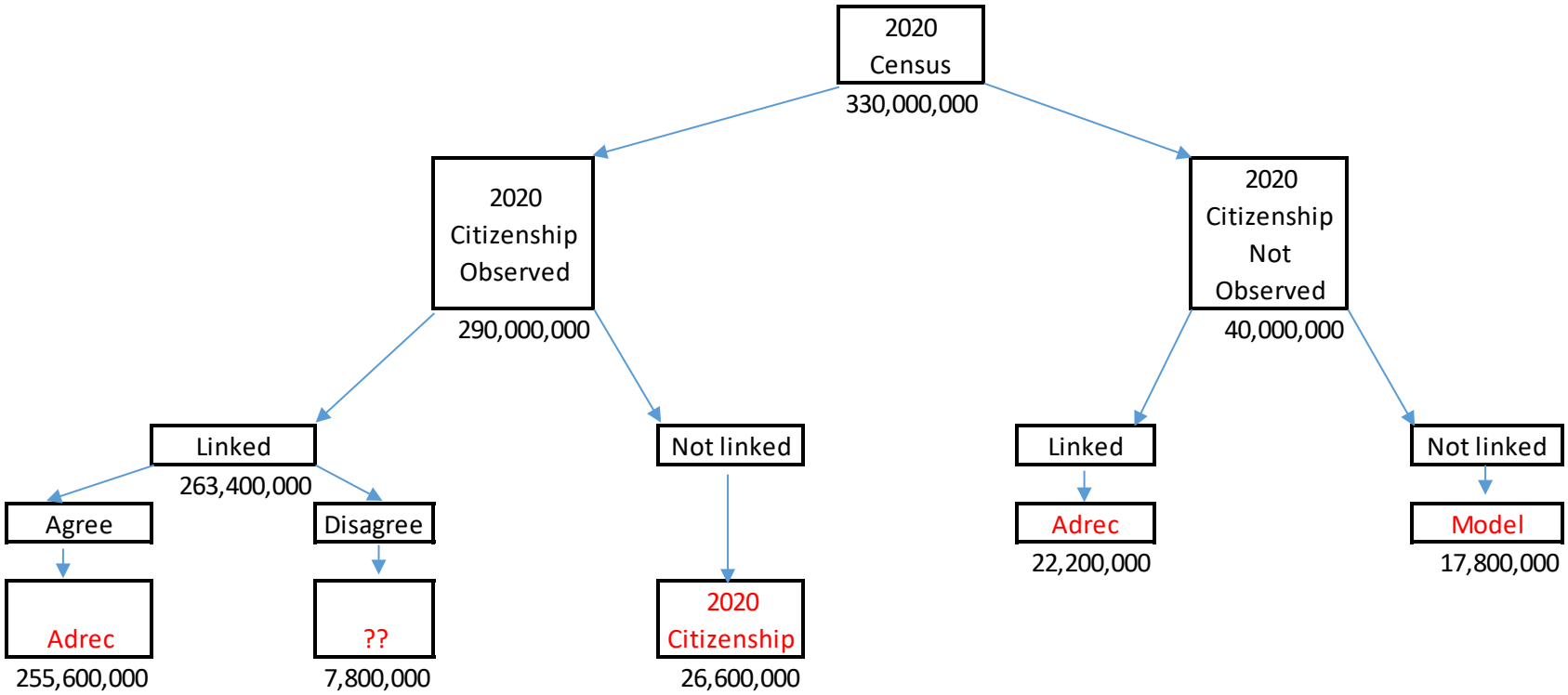
Panel A. Alternative D, Proxies Report Citizenship, Initial Assumptions



Panel B. Alternative D, Proxies Report Citizenship, Revised Assumptions



Panel C. Alternative D, Proxies Don't Report Citizenship, Initial Assumptions



Panel D. Alternative D, Proxies Don't Report Citizenship, Revised Assumptions

9. Conclusion

This paper analyzes general issues of data quality in self-reported citizenship data and examines the coverage and quality of survey-collected and administrative records data available to produce block-level estimates of the Citizen Voting Age Population (CVAP). Our descriptive and regression analyses suggest that many noncitizens misreport their own citizenship on the American Community Survey (ACS), and, in many cases, they do not provide it at all for other noncitizens in the household. The evidence also suggests some naturalized persons either do not notify the Social Security Administration (SSA) about their change in citizenship status or they do so with delay. This potential weakness in SSA data illustrates the desirability of obtaining more timely and complete citizenship data from the U.S. Customs and Immigration Services (USCIS), Customs and Border Protection (CBP), and the State Department. Addressing survey misreporting would be more difficult, however. In the absence of 100 percent complete, accurate, and up to date administrative records, one cannot rule out the possibility that the self-reported citizenship status is correct. Conceptually, it would be challenging to decide which answer to use when sources conflict. Asking respondents to provide proof of citizenship status could reduce misreporting, but this would significantly increase respondent burden and the cost of administering the survey, and it could result in additional unit nonresponse.

This paper's examination of several Census Bureau surveys with and without citizenship questions suggests that households that may contain noncitizens are more sensitive to the inclusion of citizenship in the questionnaire than all-citizen households. The implication is that adding a citizenship question to the 2020 Census would lead to lower self-response rates in households potentially containing noncitizens, resulting in more nonresponse follow-up (NRFU) fieldwork, more proxy responses, and a lower-quality population count.

References

American Community Survey, 2014, "Variance Estimation," in *American Community Survey Design Methodology (January 2014)* version 2.0, Chapter 12, https://www2.census.gov/programs-surveys/acs/methodology/design_and_methodology/acs_design_methodology_ch12_2014.pdf, accessed on June 27, 2018.

American Community Survey, 2016a, "Unweighted Housing Unit Sample," Table B98001 – American Community Survey 1-Year Estimates. U.S. Census Bureau, Washington DC, accessed on June 27, 2018, https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk.

American Community Survey, 2016b, "Housing Units," Table B25001 – American Community Survey 1-Year Estimates. U.S. Census Bureau, Washington DC, accessed on June 27, 2018, https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk.

American Community Survey, 2016c, "Selected Population Profile in the United States," Table S0201 – American Community Survey 1-Year Estimates. U.S. Census Bureau, Washington DC, accessed on July 31, 2018, https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_16_1YR_S0201&prodType=table.

American Community Survey, 2018a, "Item Allocation Rates," accessed July 9, 2018, https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/item-allocation-rates/.

American Community Survey, 2018b, "Item Allocation Rates Definitions," accessed July 9, 2018, https://www.census.gov/programs-surveys/acs/methodology/sample-size-and-data-quality/item-allocation-rates-definitions.html.

Amuedo-Dorantes, C., and F. Lozano, 2014, "Piecemeal Immigration Enforcement and the New Destinations of Interstate Undocumented Migrants: Evidence from Arizona," mimeo, Pomona College and San Diego State University.

Amuedo-Dorantes, C., and F. Lozano, 2015, "On the Effectiveness of SB1070 in Arizona," *Economic Inquiry*, 53, pp. 335-51. https://doi.org/10.1111/ecin.12138

Baumgardner, Stephanie K., Deborah H. Griffin, and David A. Raglin, 2014, "The Effects of Adding an Internet Response Option to the American Community Survey," 2014 American Community Survey Research and Evaluation Report Memorandum Series ACS14-RER-21.

Blinder, Alan S., 1973, "Wage Discrimination: Reduced Form and Structural Estimates," The Journal of Human Resources, 8, pp. 436-455. https://doi.org/10.2307/144855.

Bohn, Sarah, Magnus Lofstrom, and Steven Raphael, 2014, "Did the 2007 Legal Arizona Workers Act Reduce the State's Unauthorized Immigrant Population?" *Review of Economics and Statistics*, 96(2), pp. 258-269. https://doi.org/10.1162/rest_a_00429

Bond, Brittany, J. David Brown, Adela Luque, and Amy O'Hara, 2014, "The Nature of the Bias When Studying Only Linkable Person Records: Evidence from the American Community Survey," Proceedings of the 2013 Federal Committee on Statistical Methodology (FCSM) Research Conference, https://www.census.gov/library/working-papers/2014/adrm/carra-wp-2014-08.html (cited on July 3, 2018)

Camarota, Steven, and Jeffrey Capizzano, 2004, "Assessing the Quality of Data Collected on the Foreign-Born: An Evaluation of the American Community Survey," accessed July 31, 2018, http://www.copafs.org/seminars/evaluation_of_american_community_survey.aspx.

Census Bureau, 2002, "Measuring America: The Decennial Censuses from 1790 to 2000," Washington DC, accessed June 27, 2018, https://www.census.gov/history/pdf/measuringamerica.pdf.

Census Bureau, 2013, "Use of Paradata to Assess the Quality and Functionality of the American Community Survey Internet Instrument," 2012 American Community Survey Research and Evaluation Report Memorandum Series #ACS12-RER-26-R1/DSSD 2010 American Community Survey Memorandum Series #ACS12-MP-04-R1, Washington DC, accessed July 5, 2018, https://www.census.gov/content/dam/Census/library/working-papers/2013/acs/2013_Horwitz_01.pdf.

Department of Justice, 2017, "Request to Reinstate Citizenship Question on 2020 Census Questionnaire," letter accessed on June 27, 2018, https://www.documentcloud.org/documents/4340651-Text-of-Dec-2017-DOJ-letter-to-Census.html.

Ellis, M., R. Wright, M. Townley, and K. Copeland, 2014, "The Migration Response to the Legal Arizona Workers Act," *Political Geography*, 42, 46-56. https://doi.org/10.1016/j.polgeo.2014.06.001

Good, M., 2013, "Do Immigrant Outflows Lead to Native Inflows? An Empirical Analysis of the Migratory Responses to US State Immigration Legislation," *Applied Economics*, 45, pp. 4275-97. https://doi.org/10.1080/00036846.2013.786802

Layne, Mary, Deborah Wagner, and Cynthia Rothhaas, 2014, "Estimating Record Linkage False Match Rate for the Person Identification Validation System," CARRA Working Paper #2014-02.

Mule, Thomas, 2012, "2010 Census Coverage Measurement Estimation Report: Summary of Estimates of Coverage for Persons in the United States," DSSD 2010 Census Coverage Measurement Memorandum Series #2010-G-01, Washington, DC, accessed on July 6, 2018, https://www2.census.gov/programs-surveys/decennial/2010/technical-documentation/methodology/g-series/g01.pdf.

NORC, 2011, "Final Report: Assessment of the U.S. Census Bureau's Person Identification Validation System," University of Chicago: Bethesda, MD.

Oaxaca, Ronald, 1973, "Male-Female Wage Differentials in Urban Labor Markets," International Economic Review, 14, pp. 693-709. https://doi.org/10.2307/2525981.

Office of Management and Budget, 1997, "Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity," *Federal Register*, 62 FR 58782, Washington DC, accessed on June 28, 2018, https://www.federalregister.gov/documents/1997/10/30/97-28653/revisions-to-the-standards-for-the-classification-of-federal-data-on-race-and-ethnicity.

Office of Management and Budget, 2009, "The American Community Survey," OMB Control No. 0607-0810/ICR Reference No. 200910-0607-005, Washington DC, accessed on July 5, 2018, https://www.reginfo.gov/public/do/PRAViewICR?ref_nbr=200910-0607-005#.

Office of Management and Budget, 2008, "2010 Census," OMB Control No. 0607-0919/ICR Reference No. 200808-0607-003, Washington DC, accessed on July 5, 2018, https://www.reginfo.gov/public/do/PRAViewICR?ref_nbr=200808-0607-003#.

Orrenius, Pia, and Madeline Zavodny, 2016, "Do State Work Eligibility Verification Laws Reduce Unauthorized Immigration?" *IZA Journal of Migration*, 5(5). https://doi.org/10.1186/s40176-016-0053-3

Passel, Jeffrey S., and R.L. Clark, 1997, "How Many Naturalized Citizens Are There: An Assessment of Data Quality in the Decennial Census and the Current Population Survey," Paper presented at the 1997 Annual Meeting of the Population Association of America.

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

Rothhaas, C., Lestina, F. and Hill, J. (2012) "2010 Decennial Census Item Nonresponse and Imputation Assessment Report" 2010 Census Program for Evaluations and Experiments, Census Bureau: Washington DC, accessed July 6, 2018, https://www.census.gov/2010census/pdf/2010_Census_INR_Imputation_Assessment.pdf.

Van Hook, Jennifer, and James D. Bachmeier, 2013, "How Well Does the American Community Survey Count Naturalized Citizens?" *Demographic Research*, 29(1), pp. 1-32. https://doi.org/10.4054/DemRes.2013.29.1

Weinberg, Daniel H., 2011, "Management Challenges of the 2010 U.S. Census," Census Bureau Working Paper: Washington, DC, accessed on June 27, 2018, https://www.census.gov/history/pdf/weinberg-managing2010.pdf.

Appendix I: Tables & Figures

Table A1. Linkage Rates to the 2010 Census by Household Survey and Year, 2000 to 2015

	ACS			SIPP			CPS			AHS		
	Sample (N)	Linked (N)	(%)	Sample (N)	Linked (N)	(%)	Sample (N)	Linked (N)	(%)	Sample (N)	Linked (N)	(%)
2000							33,500	27,500	0.8114			
2001	1,301,000	1,097,000	0.8432	46,500	38,500	0.8328	44,500	37,500	0.8381			
2002	1,110,000	948,000	0.8539				71,000	61,000	0.8564			
2003	1,225,000	1,060,000	0.8652				68,500	59,000	0.8635			
2004	1,222,000	1,072,000	0.8770	86,000	76,000	0.8847	62,000	54,000	0.8716	6,700	5,700	0.8397
2005	4,068,000	3,609,000	0.8870				61,000	53,500	0.8783			
2006	4,259,000	3,792,000	0.8904				78,000	69,000	0.8853			
2007	4,159,000	3,754,000	0.9026				77,500	69,000	0.8909			
2008	4,123,000	3,774,000	0.9155	94,500	84,500	0.8927	75,500	68,000	0.9021			
2009	4,068,000	3,759,000	0.9241				77,500	70,500	0.9059			
2010	4,206,000	3,868,000	0.9195				77,000	70,500	0.9115			
2011	4,571,000	4,205,000	0.9200				77,000	70,000	0.9068	298,000	270,000	0.9047
2012	5,096,000	4,615,000	0.9055				75,500	67,000	0.8869			
2013	4,801,000	4,275,000	0.8903				75,000	65,500	0.8747	133,500	115,500	0.8683
2014	5,014,000	4,385,000	0.8745	66,000	57,500	0.8717						
2015	4,966,000	4,273,000	0.8605							173,500	146,500	0.8455

Source: 2010 Census and Master Demographics, U.S. Census Bureau.
Notes: Household survey data unweighted. All counts have been rounded.

Table A2. Citizenship in Household Surveys Linked to the 2010 Census by Demographics

	Household Surveys Linked to 2010 Census							2010 Census	
	Noncitizen		Citizen		Missing		Total		
	N	(%)	N	(%)	N	(%)	(%)	N	(%)
Total Population	1,523,000		43,090,000		1,192,000		**100.0**	308,745,538	**100.0**
								Coverage	***14.4***
Sex									
Female	785,000	1.7	22,380,000	48.9	613,000	1.3	**51.9**	157,000,000	**50.8**
Male	738,000	1.6	20,710,000	45.2	579,000	1.3	**48.1**	151,800,000	**49.2**
Race									
White	729,000	1.6	35,320,000	77.1	837,000	1.8	**80.5**	227,200,000	**73.6**
Black	127,000	0.3	4,157,000	9.1	173,000	0.4	**9.7**	40,400,000	**13.1**
American Indian, Aleut Eskimo	15,000	0.0	562,000	1.2	16,000	0.0	**1.3**	4,007,000	**1.3**
Asian or Pacific Islander	364,000	0.8	1,688,000	3.7	93,000	0.2	**4.7**	16,770,000	**5.4**
Other	287,000	0.6	1,358,000	3.0	74,500	0.2	**3.8**	20,400,000	**6.6**
Ethnicity									
Hispanic/Spanish	675,000	1.5	4,046,000	8.8	198,000	0.4	**10.7**	50,480,000	**16.4**
Non-Hispanic/Spanish	848,000	1.9	39,040,000	85.2	994,000	2.2	**89.3**	258,300,000	**83.7**

Source: 2010 Census and Master Demographics, U.S. Census Bureau.

Notes: The household survey data are unweighted. The reported population total is the official count from the 2010 Census. All other counts have been rounded. CBDRB-2017-CDAR-001.

Table A3. Item Nonresponse Rates for 2000 and 2010 Short Form Person Questionnaires

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

Source: Rothhaas, Lestina, and Hill (2012) Tables.

Notes: Rothhaas, Lestina, and Hill (2012) state "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."

Table A4. ACS Item Allocation Rates

Item	2010	2013	2016
Overall housing allocation rate occupied and vacant housing units	5.2	5.6	4.9
Overall person allocation rate total population	5.8	8.4	9.5
Vacancy status vacant housing units	2.9	3.5	3.9
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	3.4	3	3
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	16.2	17.1	18.2
Lot size occupied and vacant single family and mobile homes	4.2	3.9	3.9
Agricultural sales occupied and vacant single family and mobile homes with lot size greater than or equal to 1 acre	4.4	4.2	4
Business on property occupied and vacant single family and mobile homes	3	2.4	**
Number of rooms occupied and vacant housing units	5.2	5.5	5
Number of bedrooms occupied and vacant housing units	4.3	4.6	5.5
Running water occupied and vacant housing units	2	2.1	2.4
Flush toilet occupied and vacant housing units	2	2.2	**
Bathtub or shower occupied and vacant housing units	2	2.2	2.6
Sink with a faucet occupied and vacant housing units	2	2.2	2.6
Stove or range occupied and vacant housing units	2.5	2.8	3.1
Refrigerator occupied and vacant housing units	2.7	2.9	3.2
Telephone occupied housing units	1.1	1.2	1.5

Table A4. ACS Item Allocation Rates Continued

Item	2010	2013	2016
Number of vehicles occupied housing units	1.3	1.4	1.2
Heating fuel, occupied housing units	3.3	3.4	3.4
Monthly electricity cost occupied housing units	7.3	8.2	8.1
Monthly gas cost occupied housing units	9.8	9.9	9.6
Yearly water and sewer cost occupied housing units	8.1	8.8	8.5
Yearly other fuel cost occupied housing units	10.6	8.3	7.3
Yearly food stamp recipiency household occupied housing units	1.3	1.7	1.7
Yearly real estate taxes owner-occupied housing units	16.3	18.5	16.7
Yearly property insurance owner-occupied housing units	23.2	25.6	23.9
Mortgage status owner-occupied housing units	2.1	2.5	2.2
Monthly mortgage payment owner-occupied housing units with a mortgage payment	10.7	12.4	10.5
Mortgage payment incl. real estate taxes owner-occupied housing units with a mortgage	(X)	6.9	6.2
Mortgage payment incl. insurance owner-occupied housing units with a mortgage	(X)	7.4	6.8
Second mortgage owner-occupied housing units	3.4	3.7	3.2
Home equity loan owner-occupied housing units	4.2	4.3	3.7
Other monthly mortgage payment(s) owner-occupied housing units with second mortgage or home equity loan	17.9	21.7	23.3
Property value owner-occupied housing units and vacant housing units for sale	12.3	12.9	11.6
Yearly mobile home costs Occupied mobile homes and other units	19.9	21.5	21.7
Monthly condominium fee owner-occupied housing units	0.7	0.8	0.8
Monthly rent occupied housing units rented for cash rent and vacant housing units for rent	9.3	9.8	10.5

Table A4. ACS Item Allocation Rates Continued

Item	2010	2013	2016
Meals included in rent occupied housing units rented for cash rent and vacant housing units for rent	2	2.1	2.1
Desktop/laptop/notebook computer occupied housing units	**	3.2	1.3
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	**	3.7	1.7
Household has internet access occupied housing units	**	4.4	3.3
Dial-up internet service occupied housing units with internet access	**	5.7	3.8
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	**	26.7	7.6
Satellite internet service occupied housing units with internet access	**	5.7	3.8
High speed internet service occupied housing units with internet access	**	**	3.8
Some other internet service occupied housing units with internet access	**	5.7	3.8
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	1.3	1.6	1.7
Relationship total household population	1.2	1.1	1.2
Marital status total population 15 years and over	3	4.8	5.3

Table A4. ACS Item Allocation Rates Continued

Item	2010	2013	2016
Married past 12 months total population 15 years and over, except those never married	4.7	6.6	6.9
Widowed past 12 months total population 15 years and over, except those never married	4.5	7	7.4
Divorced past 12 months total population 15 years and over, except those never married	4.5	7	7.4
Times married Total population 15 years and over, except those never married	5.1	7.8	8.1
Year last married total population 15 years and over, except those never married	11.4	13.3	13.5
Place of birth total population	6.5	8.6	9.1
Citizenship total population	2.7	5.2	6
Year of naturalization total population naturalized citizens	16.6	22.5	22.5
Year of entry total population not born in U.S.	10.3	13.2	14.8
Speaks another language at home total population 5 years and over	3.4	5.9	6.8
Language spoken total population 5 years and over who speak another language at home	5.7	7	8.3
English ability total population 5 years and over who speak another language at home	4	5.9	7.1
School enrollment total population 3 years and over	3.7	6	6.7
Grade level attended total population 3 years and over enrolled	6	8.9	10.2
Educational attainment total population 3 years and over	5.6	8	8.5
Field of degree total population 25 years and over with a bachelor's degree or higher	9.8	12.4	13.5
Mobility status total population 1 year and over	4	6.5	7.2

Table A4. ACS Item Allocation Rates Continued

Item	2010	2013	2016
Migration state/foreign county total population 1 year and over movers	7.1	11.3	13.2
Migration county total population 1 year and over movers within U.S.	8.3	12.5	14.6
Migration minor civil division total population 1 year and over movers within U.S.	8.4	12.1	14.2
Migration place total population 1 years and over movers within U.S.	8.8	12.9	15
Health insurance through employer/union total population	6.2	9	10.7
Health insurance purchased directly total population	6.9	9.7	11.3
Health insurance through Medicare total population	5.2	8.1	9.5
Health insurance through Medicaid total population	7.9	10.5	12.2
Health insurance through TRICARE total population	8.1	10.8	12.5
Health insurance through VA total population	8.1	10.7	12.3
Health ins. Through Indian Health Service total population	8.5	11.1	12.8
Visual difficulty total population	3.4	6.1	7.1
Hearing difficulty total population	3.2	5.9	6.8
Physical difficulty total population 5 years and over	3.5	6.7	7.5
Difficulty remembering total population 5 years and over	3.5	6.7	7.5
Difficulty dressing total population 5 years and over	3.5	6.7	7.5
Difficulty going out total population 16 years and over	3.4	6.5	7.3
Grandchildren living in home noninstitutionalized population 30 years and over	0.9	1	1.1
Responsibility for grandchildren noninstitutionalized population 30 years and over who are grandparents with grandchildren in the home	12	15.7	17.7

Table A4. ACS Item Allocation Rates Continued

Item	2010	2013	2016
Months responsible for grandchildren noninstitutionalized population 30 years and over who are grandparents with grandchildren in the home that have responsibility	14.9	16.1	17.2
Fertility status female total population 15-50	3.7	6.7	7.8
Veteran status total population 17 years and over	3.8	6.8	7.3
Periods of military service total population 17 years and over on active duty now or previously	6.3	9.3	9.7
Service-connected disability rating total population 17 years and over, except those who never served in the Armed Forces	3.9	6.6	6.8
Employment status recode noninstitutionalized population 16 years and over	0.7	0.2	0.2
When last worked noninstitutionalized population 16 years and over	5.1	8.1	8.7
Weeks worked in the past 12 months noninstitutionalized population 16 years and over who worked in the past 12 months	6.9	9.7	10.6
Hours worked per week noninstitutionalized population 16 years and over who worked in the past 12 months	7.7	10.8	11.9
Place of work state/foreign county noninstitutionalized population 16 years and over at work last week	6.3	10.4	11.8
Place of work county noninstitutionalized population 16 years and over at work last week	7	11	12.5
Place of work minor civil division noninstitutionalized population 16 years and over at work last week	2.1	3.3	3.6
Place of work place noninstitutionalized population 16 years and over at work last week	7.6	11.6	13.1
Transportation to work noninstitutionalized population 16 years and over at work last week	5.7	8.8	9.6
Carpool size noninstitutionalized population 16 years and over at work last week who drive to work	6.8	9.9	10.9

Table A4. ACS Item Allocation Rates Continued

Item	2010	2013	2016
Time of departure noninstitutionalized population 16 years and over at work last week who don't work at home	12.8	18.5	20.2
Commuting time noninstitutionalized population 16 years and over at work last week who don't work at home	9.7	13.3	14.5
Class of worker total population 16 years and over who worked in the last 5 years	7.2	10.7	11.7
Industry total population 16 years and over who worked in the last 5 years	7.8	11.4	12.7
Occupation total population 16 years and over who worked in the last 5 years	8.1	11.8	13.4
Wages/salary income total population 15 years and over	16	19	19.1
Self-employment income total population 15 years and over	5.9	9.3	10.5
Interest, dividends, etc. income total population 15 years and over	8.8	12.6	15.2
Social security or railroad retirement total population 15 years and over	8.9	12.3	14.5
Supplemental security income total population 15 years and over	6.7	10.3	12.7
Public assistance total population 15 years and over	6.8	10.5	13.2
Retirement income total population 15 years and over	7.5	11.1	13.6
Other income total population 15 years and over	7.4	10.8	13.2
Some or all income allocated total population 15 years and over	22.4	25.3	28.4

Source: American Community Survey (ACS) 1-year files in 2010, 2013, and 2016.

Notes: Item allocation includes nonresponses and responses that were edited. See ACS (2018a and 2018b) for more information about ACS item allocation rates. ** Item was not asked in this year. (X) Some instances where no response to this question was required were incorrectly tallied as allocations, overstating the true level of item allocation required. The incorrect rates have been removed.

Table A5. Citizenship Item Allocation Rate by Response Mode, 2013-2016

	Mail-in Response				Internet Self-Response			
	2013		2016		2013		2016	
	(%)	s.e.	(%)	s.e.	(%)	s.e.	(%)	s.e.
NH White	6.1	(0.023)	6.3	(0.024)	6.2	(0.019)	6.2	(0.018)
NH Black	12.3	(0.090)	12.6	(0.100)	12.3	(0.100)	13.1	(0.091)
NH Asian/NHPI	10.3	(0.126)	12.7	(0.151)	9.4	(0.083)	9.6	(0.075)
NH Other	8.4	(0.143)	8.4	(0.154)	10.0	(0.128)	10.2	(0.114)
Hispanic/Latino	11.8	(0.080)	12.3	(0.088)	13.0	(0.078)	15.5	(0.071)

Source: 2013 & 2016 ACS 1-year files.

Note: Item allocation includes nonresponses and responses that were edited.

Table A6. Administrative Record (AR) Coverage of the 2010 Census, Using Initial AR-Census Crosswalk

	Count	Percent of Decennial Population	Percent of Matched Sample
No PIK, not sent to PVS	10,370,000	3.4	
No PIK, failed in PVS	19,200,000	6.2	
PIK, but not in Numident, not ITIN	8,900	0.0	
PIK, but not in Numident, is ITIN	1,567,000	0.5	
Blank Citizenship	57,910,000	18.8	20.9
U.S. Citizen	200,400,000	64.9	72.2
Noncitizen	19,270,000	6.2	6.9
Total	308,745,538	100.00	100.00

Source: 2010 Numident and initial administrative record-2010 Census crosswalk.

Note: This is the crosswalk used by Rastogi and O'Hara (2012).

Table A7. Percent Linked to 2010 Census among 2017 Numident Records with Missing Citizenship

	Foreign-Born	U.S.-Born
Percent Linked to 2010 Census	36.3	74.5
Total	6.8 million	57.0 million

Source: 2010 Census and 2017 Numident

Notes: These are persons in the 2017 Numident with missing citizenship, born after 1919, and with no date of death. Our preliminary analysis reported 6.6 million foreign-born persons, which excluded some relevant records.

Table A8 Panel A: Citizenship Agreement between 2000 Census Long Form and Administrative Records

	AR Citizen	AR Noncitizen	AR Missing	Percent by ACS Category
		All (N=42,580,000)		
Census Citizen	98.8	29.9	71.6	93.0
Census Noncitizen	0.9	66.4	23.8	6.2
Census Missing	0.3	3.7	4.6	0.8
Percent by AR Cat.	86.9	5.4	7.7	100.0
		Non-Hispanic White (N=31,690,000)		
Census Citizen	99.4	31.8	92.4	97.9
Census Noncitizen	0.4	65.8	5.9	1.8
Census Missing	0.2	2.4	1.7	0.3
Percent by AR Cat.	93.4	1.8	4.8	100.0
		Non-Hispanic Black (N=4,543,000)		
Census Citizen	99.3	36.3	92.5	96.1
Census Noncitizen	0.4	59.2	5.4	3.3
Census Missing	0.2	4.5	2.1	0.6
Percent by AR Cat.	85.3	4.0	10.8	100.0
		Hispanic (N=4,534,000)		
Census Citizen	94.3	25.6	35.9	69.3
Census Noncitizen	4.7	69.7	55.2	27.3
Census Missing	1.0	4.7	9.0	3.3
Percent by AR Cat.	60.6	19.1	20.3	100.0
		Non-Hispanic Other Race (N=1,821,000)		
Census Citizen	93.4	33.3	53.1	71.2
Census Noncitizen	5.1	63.7	37.0	26.0
Census Missing	1.4	3.0	9.9	2.9
Percent by AR Cat.	59.2	29.3	11.6	100.0
		Reference Person (N=16,450,000)		
Census Citizen	98.7	32.9	76.9	94.0
Census Noncitizen	0.9	63.3	19.3	5.3
Census Missing	0.4	3.9	3.8	0.8
Percent by AR Cat.	89.4	5.5	5.1	100.0
		Relative (N=24,980,000)		
Census Citizen	98.9	28.6	71.7	92.9
Census Noncitizen	0.8	68.0	24.3	6.4
Census Missing	0.2	3.4	4.1	0.7
Percent by AR Cat.	86.4	5.3	8.3	100.0
		Nonrelative (N=1,153,000)		
Census Citizen	97.2	20.4	58.6	80.4
Census Noncitizen	1.9	72.0	31.3	15.6
Census Missing	0.9	7.6	10.1	4.1
Percent by AR Cat.	63.7	7.4	28.9	100.0
		Age 18+ (N=31,260,000)		
Census Citizen	98.5	30.3	67.1	91.8
Census Noncitizen	1.1	65.5	26.8	7.2
Census Missing	0.4	4.1	6.1	1.1
Percent by AR Cat.	86.2	6.5	7.2	100.0

Source: 2000 Census Long Form and 2002 Census Numident.

Notes: These are weighted percentages. The first three rows of each panel contain percentages by survey group within the AR category, and the last row contains percentages by AR category of the sample as a whole. Here AR citizen includes Numident records with missing citizenship, regardless of their country of birth.

Table A8 Panel B: Citizenship Agreement between 2010 ACS and Administrative Records

	AR Citizen	AR Noncitizen	AR Missing	Percent by ACS Category
		All (N=4,520,000)		
ACS Citizen	96.9	32.7	74.8	90.2
ACS Noncitizen	0.8	63.2	17.5	6.9
ACS Missing	2.4	4.1	7.7	2.9
Percent by AR Cat.	83.9	7.5	8.5	100.0
		Non-Hispanic White (N=3,152,000)		
ACS Citizen	97.8	42.4	87.4	96.1
ACS Noncitizen	0.2	53.9	4.0	1.5
ACS Missing	2.0	3.7	8.6	2.4
Percent by AR Cat.	92.4	2.0	5.7	100.0
		Non-Hispanic Black (N=434,000)		
ACS Citizen	96.3	40.1	85.0	92.4
ACS Noncitizen	0.5	54.5	6.5	3.8
ACS Missing	3.2	5.4	8.6	3.8
Percent by AR Cat.	86.2	5.2	8.6	100.0
		Hispanic (N=609,000)		
ACS Citizen	93.9	23.8	61.1	72.7
ACS Noncitizen	2.9	72.4	32.9	23.4
ACS Missing	3.2	3.9	6.0	3.9
Percent by AR Cat.	59.8	21.5	18.8	100.0
		Non-Hispanic Other Race (N=326,000)		
ACS Citizen	93.7	41.3	59.0	77.3
ACS Noncitizen	2.8	54.3	31.8	18.5
ACS Missing	3.5	4.4	9.1	4.2
Percent by AR Cat.	65.7	25.4	8.8	100.0
		Reference Person (N=1,770,000)		
ACS Citizen	97.7	37.2	80.8	91.9
ACS Noncitizen	0.7	59.9	14.1	6.2
ACS Missing	1.6	2.9	5.1	1.9
Percent by AR Cat.	85.4	7.7	6.9	100.0
		Relative (N=2,504,000)		
ACS Citizen	96.5	30.6	75.3	89.7
ACS Noncitizen	0.7	64.7	16.4	6.9
ACS Missing	2.8	4.6	8.2	3.4
Percent by AR Cat.	84.0	7.5	8.5	100.0
		Nonrelative (N=102,000)		
ACS Citizen	94.1	20.0	53.0	77.0
ACS Noncitizen	1.9	72.2	34.8	16.6
ACS Missing	4.0	7.8	12.2	6.4
Percent by AR Cat.	66.0	9.4	24.6	100.0
		Age 18+ (N=3,505,000)		
ACS Citizen	97.0	33.1	71.7	89.1
ACS Noncitizen	0.9	62.9	20.9	8.2
ACS Missing	2.1	4.0	7.3	2.7
Percent by AR Cat.	82.6	9.0	8.4	100.0

Source: 2010 American Community Survey (ACS) and 2010 Census Numident.

Notes: These are weighted percentages. The first three rows of each panel contain percentages by survey group within the AR category, and the last row contains percentages by AR category of the sample as a whole. Here AR citizen includes Numident records with missing citizenship, regardless of their country of birth.

Table A8 Panel C: Citizenship Agreement between 2016 ACS and Administrative Records

	AR Citizen	AR Noncitizen	AR Missing	Percent by ACS Category
		All (N=5,255,000)		
ACS Citizen	93.8	34.7	70.4	87.3
ACS Noncitizen	0.7	57.6	17.7	6.4
ACS Missing	5.5	7.7	11.9	6.3
Percent by AR Cat.	82.4	6.7	10.9	100.0
		Non-Hispanic White (N=3,579,000)		
ACS Citizen	95.0	44.7	81.8	93.1
ACS Noncitizen	0.2	48.8	4.4	1.4
ACS Missing	4.8	6.6	13.7	5.5
Percent by AR Cat.	91.1	1.9	7.0	100.0
		Non-Hispanic Black (N=495,000)		
ACS Citizen	93.2	42.0	82.5	89.3
ACS Noncitizen	0.4	49.7	5.8	3.6
ACS Missing	6.4	8.4	11.7	7.2
Percent by AR Cat.	82.3	5.1	12.6	100.0
		Hispanic (N=732,000)		
ACS Citizen	90.5	26.6	58.0	73.1
ACS Noncitizen	2.7	65.2	32.3	19.3
ACS Missing	6.8	8.2	9.7	7.7
Percent by AR Cat.	62.0	16.2	21.8	100.0
		Non-Hispanic Other Race (N=449,000)		
ACS Citizen	90.3	39.1	54.2	74.6
ACS Noncitizen	2.4	53.5	32.2	17.3
ACS Missing	7.3	7.4	13.6	8.1
Percent by AR Cat.	65.8	22.0	12.2	100.0
		Reference Person (N=2,037,000)		
ACS Citizen	96.7	39.1	71.6	90.6
ACS Noncitizen	0.7	56.4	20.1	6.1
ACS Missing	2.7	4.5	8.3	3.2
Percent by AR Cat.	85.5	7.3	7.2	100.0
		Relative (N=2,789,000)		
ACS Citizen	92.3	32.5	68.4	86.0
ACS Noncitizen	0.7	58.2	18.6	6.3
ACS Missing	6.9	9.3	13.0	7.7
Percent by AR Cat.	83.5	6.5	10.0	100.0
		Nonrelative (N=135,000)		
ACS Citizen	85.3	21.5	52.4	71.8
ACS Noncitizen	1.5	61.5	23.6	11.8
ACS Missing	13.2	17.0	23.9	16.4
Percent by AR Cat.	65.6	7.1	27.3	100.0
		Age 18+ (N=4,178,000)		
ACS Citizen	94.3	34.8	68.0	86.6
ACS Noncitizen	0.9	57.7	20.3	7.6
ACS Missing	4.8	7.6	11.7	5.8
Percent by AR Cat.	81.0	8.2	10.7	100.0

Source: 2016 American Community Survey (ACS) and 2016 Census Numident.

Notes: These are weighted percentages. The first three rows of each panel contain percentages by survey group within the AR category, and the last row contains percentages by AR category of the sample as a whole. Here AR citizen includes Numident records with missing citizenship, regardless of their country of birth.

Table A9. Citizenship Agreement Rates (%) Between Census Surveys and Administrative Records

	Consistent	Inconsistent	Missing in One or Both Sources	Consistent, Conditional on Nonmissing	Inconsistent, Conditional on Nonmissing
2000 Census	89.4	2.4	8.2	97.4	2.6
2010 ACS	86.1	3.1	10.8	96.5	3.5
2016 ACS	81.1	2.9	15.9	96.5	3.5

Source: 2000 Census long form, 2002 Census Numident, 2010 American Community Survey (ACS), 2010 Census Numident, 2016 ACS, and 2016 Census Numident.

Notes: These are weighted percentages. Here AR citizen includes Numident records with missing citizenship, regardless of their country of birth. The original estimate for inconsistent in the 2000 Census was 2.3 percent, for consistent in the 2010 ACS was 86.0 percent, consistent conditional on nonmissing in the 2010 ACS was 96.4 percent, inconsistent conditional on nonmissing in the 2010 ACS was 3.6 percent, and consistent in the 2016 ACS was 81.2 percent.

Table A10. AR Citizen and Noncitizen Percentages of the 2016 ACS by Race/Ethnicity and Relationship to Reference Person

	AR Citizens	AR Noncitizens
All	81.1	6.7
Non-Hispanic White	90.1	1.9
Non-Hispanic Black	81.5	5.1
Hispanic	60.2	16.2
Non-Hispanic Other Race	62.5	22.0
Reference Person	81.1	6.9
Relative	82.1	6.5
Non-Relative	64.8	7.1

Source: 2010 ACS 1-year file

Notes: These are weighted percentages. The omitted category is persons missing AR citizenship.

Table A11. Percentages of the 2016 ACS Sample by Relationship to Reference Person and Record Linkage Quality

	High-Quality Linkage	Low-Quality Linkage
Reference Person	18.1	20.3
Relative	23.9	33.8
Non-Relative	0.6	3.2

Source: 2016 ACS 1-year file

Notes: These results are weighted. This excludes persons missing AR citizenship.

Table A12. 2016 ACS Citizenship Distribution for ITINs

	Percent of All ITINs
U.S. Citizens	11.1
Born Citizens	6.6

Source: 2016 ACS 1-year file

Note: These results are weighted.

Table A13. Comparison of 2010 ACS and 2010 Census Response Rates: Regressions by Household Citizenship Type

	AR all-citizen households	AR noncitizen households	AR & ACS all-citizen households	All other households
Log Household Size	-3.184	-8.237	-0.4762	-7.185
	(0.1476)	(0.5100)	(0.1737)	(0.3067)
Log Household Size Squared	-0.0998	1.565	-1.929	2.944
	(0.0899)	(0.2304)	(0.1159)	(0.1423)
Female	-6.665	-6.687	-6.263	-8.167
	(0.0557)	(0.1809)	(0.0578)	(0.1289)
Non-Hispanic African Amer.	-10.53	-8.422	-11.48	-3.573
	(0.1143)	(0.3952)	(0.1275)	(0.2064)
Hispanic	-7.532	-20.55	-7.145	-14.07
	(0.1585)	(0.2962)	(0.1640)	(0.2123)
Other Non-Hispanic	0.8338	-0.1256	0.4897	4.129
	(0.1809)	(0.2904)	(0.2162)	(0.2158)
Age 25-34	-4.052	-3.101	-4.658	-1.380
	(0.2078)	(0.5207)	(0.2602)	(0.3202)
Age 35-44	-9.122	-4.746	-9.582	-3.653
	(0.2117)	(0.5048)	(0.2704)	(0.3181)
Age 45-54	-11.83	-6.676	-12.26	-5.095
	(0.2418)	(0.5313)	(0.3082)	(0.3113)
Age 55-64	-12.78	-5.792	-13.20	-5.395
	(0.2715)	(0.5466)	(0.3530)	(0.3315)
Age 65+	-13.06	-4.225	-13.76	-3.617
	(0.3121)	(0.6672)	(0.4051)	(0.3857)
High School	0.7658	-1.195	1.641	-1.866
	(0.1055)	(0.2641)	(0.1097)	(0.1828)
Bachelor's Degree	3.864	2.383	5.116	0.1112
	(0.1197)	(0.3549)	(0.1262)	(0.2316)
Graduate Degree	7.685	7.098	8.448	6.310
	(0.1330)	(0.3923)	(0.1387)	(0.2661)
HH Income $1-$25,000	-1.854	-1.525	-2.249	-1.537
	(0.3130)	(0.9868)	(0.3665)	(0.5480)
HH Income $25,001-$50,000	-2.759	-1.995	-3.002	-3.304
	(0.3158)	(0.9549)	(0.3604)	(0.5348)
HH Income $50,001-$75,000	-3.093	-0.6062	-3.454	-2.555
	(0.3164)	(0.9907)	(0.3563)	(0.5494)
HH Income $75,001-$100,000	-3.037	-0.4054	-3.300	-2.435
	(0.3091)	(1.004)	(0.3590)	(0.5505)
HH Income $100,001+	-2.272	1.035	-2.499	-0.9051
	(0.3183)	(1.016)	(0.3672)	(0.5719)

Table A13. Continued

	AR all-citizen households	AR noncitizen households	AR & ACS all-citizen households	All other households
Worked in	2.741	1.204	3.027	-1.012
Last Week	(0.0644)	(0.2549)	(0.0742)	(0.1613)
Searched for	8.495	8.559	8.565	6.753
Job	(0.1282)	(0.3652)	(0.1357)	(0.2629)
Log Number of				
Years in	-11.17	-10.34	-14.09	-1.286
U.S.	(0.5538)	(0.5499)	(1.434)	(0.4572)
Log Number of				
Years in	2.845	1.997	3.315	-0.1304
U.S. Squared	(0.0904)	(0.1044)	(0.2063)	(0.0823)
English Very	0.9990	0.7404	0.7193	5.302
Well	(0.1669)	(0.2508)	(0.1760)	(0.1927)
English Well	3.823	0.4760	6.686	3.160
	(0.3037)	(0.3007)	(0.3449)	(0.2369)
English Not	-4.707	-7.014	-0.5008	-6.007
Well	(0.3595)	(0.3431)	(0.4334)	(0.3088)
English Not	-13.87	-14.00	-15.19	-13.50
At All	(0.6209)	(0.5355)	(1.070)	(0.4750)
Weighted Obs.	85,100,000	11,400,000	72,300,000	24,200,000
Unweighted Obs.	1,280,000	139,000	1,112,000	306,000

Source: 2010 ACS 1-year file, 2010 Census Unedited File (CUF), and 2010 Numident.

Notes: The 2010 Census self-response is non-blank response to the first mailing, and only NRFU-eligible housing units are included. ACS self-response is mail response. These regressions are estimated by linear probability models (LPM), weighted by ACS person weights. The standard errors are in parentheses. The standard errors are bootstrapped using 80 ACS replicate weights. The base categories are non-Hispanic white for race/ethnicity, less than high school for educational attainment, no income for household income, and speaks only English at home for English ability.

Table A14. Blinder-Oaxaca Decomposition Coefficients for Comparison of ACS 2010 and Census 2010 Self-Response Rates by Household Citizenship

	AR all-citizen vs. AR noncitizen households		AR & ACS all-citizen vs. all other households	
	Explained	Unexplained	Explained	Unexplained
Log Household Size	1.347	5.799	0.1479	6.758
	(0.0624)	(0.5861)	(0.0539)	(0.3311)
Log Household Size	0.0778	-2.687	1.108	-6.613
Squared	(0.0700)	(0.3702)	(0.0664)	(0.2305)
Female	-0.4591	0.0090	-0.2351	0.8355
	(0.0076)	(0.0761)	(0.0048)	(0.0619)
Non-Hispanic White	2.261	-0.6934	1.612	0.4970
	(0.0419)	(0.0480)	(0.0311)	(0.0566)
Non-Hispanic	-0.2343	-0.4469	-0.0307	-0.7966
African Amer.	(0.0042)	(0.0230)	(0.0030)	(0.0217)
Hispanic	1.196	4.487	0.6141	2.391
	(0.0420)	(0.0904)	(0.0280)	(0.0515)
Other Non-	-0.9869	-0.4652	-0.6266	-0.3870
Hispanic	(0.0258)	(0.0396)	(0.0202)	(0.0298)
Age Below 25	0.0184	0.1686	-0.0735	0.2663
	(0.0029)	(0.0160)	(0.0031)	(0.0156)
Age 25-34	-0.3132	0.7466	-0.1863	0.4588
	(0.0091)	(0.0536)	(0.0065)	(0.0402)
Age 35-44	0.0819	0.0026	0.0487	-0.0507
	(0.0107)	(0.0539)	(0.0068)	(0.0396)
Age 45-54	0.1199	-0.1936	0.0323	-0.3278
	(0.0036)	(0.0537)	(0.0017)	(0.0315)
Age 55-64	-0.2849	-0.3316	-0.1938	-0.3167
	(0.0069)	(0.0331)	(0.0059)	(0.0295)
Age 65+	-0.7552	-0.2910	-0.4319	-0.6361
	(0.0229)	(0.0218)	(0.0165)	(0.0354)
Below High School	0.4911	-0.2640	0.4301	-0.5491
	(0.0130)	(0.0609)	(0.0094)	(0.0338)
High School	-0.3682	0.4025	-0.1901	0.4193
	(0.0084)	(0.0735)	(0.0048)	(0.0557)
Bachelor's Degree	0.0169	0.0827	0.0341	0.4077
	(0.0013)	(0.0356)	(0.0018)	(0.0255)
Graduate Degree	-0.0978	-0.0594	-0.0036	-0.0646
	(0.0030)	(0.0320)	(0.0018)	(0.0217)
HH Income = $0	0.0017	0.0190	-0.0191	0.0117
	(0.0005)	(0.0099)	(0.0024)	(0.0096)
HH Income $1	0.0072	0.2648	0.0003	-0.0193
-$25,000	(0.0021)	(0.0567)	(0.0002)	(0.0453)
HH Income $25,001	0.0122	0.2196	0.0042	0.2364
-$50,000	(0.0017)	(0.0619)	(0.0007)	(0.0424)

Table A14. Continued

	AR all-citizen vs. AR noncitizen households		AR & ACS all-citizen vs. all other households	
	Explained	Unexplained	Explained	Unexplained
HH Income	0.0030	-0.1651	-0.0063	-0.0477
$50,000-$75,000	(0.0006)	(0.0470)	(0.0007)	(0.0288)
HH Income	-0.0032	-0.1245	-0.0060	-0.0278
$75,001-$100,000	(0.0005)	(0.0339)	(0.0007)	(0.0221)
HH Income	0.0003	-0.3579	-0.0001	-0.1981
$100,001+	(0.0003)	(0.0581)	(0.0007)	(0.0408)
Worked in	-0.3684	1.134	-0.0440	2.545
Last Week	(0.0089)	(0.1959)	(0.0024)	(0.1203)
Searched for	-0.1195	-0.0050	-0.0399	0.1230
Job	(0.0041)	(0.0286)	(0.0025)	(0.0198)
Log Number of.	-10.19	-2.434	-8.077	-42.62
Years in U.S	(0.5032)	(2.068)	(0.8219)	(4.835)
Log Years in	16.68	7.861	12.22	40.24
U.S. Squared	(0.5292)	(1.126)	(0.7610)	(2.490)
Only English	1.876	-0.2650	0.7321	-0.2700
	(0.1384)	(0.0675)	(0.1293)	(0.1763)
English Very	-0.9976	-0.3151	-0.4244	-1.183
Well	(0.0486)	(0.0873)	(0.0481)	(0.0699)
English Well	-1.270	0.4510	-1.010	0.3912
	(0.0491)	(0.0669)	(0.0360)	(0.0468)
English Not	0.3237	0.1948	-0.1211	0.5456
Well	(0.0444)	(0.0541)	(0.0393)	(0.0414)
English Not	0.6369	-0.0655	0.5055	-0.0864
At All	(0.0272)	(0.0377)	(0.0308)	(0.0358)

Source: 2010 ACS 1-year file, 2010 Census Unedited File (CUF), and 2010 Numident.

Notes: The 2010 Census self-response is non-blank response to the first mailing, and only NRFU-eligible housing units are included. ACS self-response is mail response. These regressions are estimated by linear probability models (LPM), weighted by ACS person weights. The standard errors are in parentheses. The standard errors are bootstrapped using 80 ACS replicate weights. The base categories are non-Hispanic white for race/ethnicity, less than high school for educational attainment, no income for household income, and speaks only English at home for English ability. The number of observations is 1,418,000 (unweighted) and 96,540,000 (weighted).