

# Census Bureau Citizenship Data Research and Product Development

**John M. Abowd**

**Chief Scientist and Associate Director for Research and Methodology, U.S. Census Bureau**

September 6, 2019

Council of Professional Associations on Federal  
Statistics

The opinions in this presentation are those of the author and not the official position of the U.S. Census Bureau, except as explicitly noted. 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-B0035-CED-20190322, CBDRB-FY19-CMS-7901, and CBDRB-FY19-CMS-7917.

**Shape  
your future  
START HERE >**

United States<sup>®</sup>  
**Census  
2020**

# Topics for Today

1. 2020 Census data products and the President's Executive Order
2. Research on citizenship data quality, including Brown, Heggeness, Dorinski, Warren, and Yi (*Demography* 2019)
3. Sources of administrative citizenship data
4. Producing block-level Citizenship Voting-age Population (CVAP) using administrative and survey data with the 2020 Census frame

# Apportionment

The Census Bureau's primary data product from the 2020 Census is apportionment counts that will be delivered to the President by December 31, 2020.

The counts are used to reapportion the U.S. House of Representatives.

The apportionment population count for each of the 50 states includes the state's total resident population (citizens and non-citizens), plus a count of overseas federal employees (and their dependents living with them) who are allocated to their home states.

This is identical to the method used for the 2010 Census, but based on the 2020 Census residence criteria.

The apportionment counts are calculated using the Census Unedited File (CUF), which is produced by November 30, 2020.

The CUF does not contain any citizenship data.

# Redistricting (PL94-171)

The Census Bureau is required, under Public Law 94-171, to make data available to the states to assist in redistricting.

These data are produced from the Census Edited File (CEF), which is produced from the CUF by imputing item missing data using administrative records and statistical models.

The CEF is produced by January 25, 2021.

The CEF is sent to the 2020 Census Disclosure Avoidance System, which releases the Micro-data Detail File (MDF) to the tabulation system.

Redistricting data at the block level are produced from the MDF, and will be released by state from February 18 through March 31, 2021.

# Redistricting (PL94-171) Format

Total population by the 63 detailed race categories – Table P1;

Total population by Hispanic origin (across all races) and for the non-Hispanic origin population by the 63 detailed race categories – Table P2;

Total voting-age population by the 63 detailed race categories – Table P3;

Total voting age population by Hispanic origin (across all races) and for the non-Hispanic origin population by the 63 detailed race categories – Table P4.

Total Population only - Group Quarters Population by Group Quarters Type – Table P5.

Housing Unit Counts - Occupancy Status – Table H1.

# Citizen Voting-Age Population Data

The Paperwork Reduction Act clearance package for the 2020 Census and the President's Executive Order 13880 commit the Census Bureau to releasing Citizen Voting-Age Population (CVAP) data by March 31, 2021.

These data will be produced by combining administrative data from a number of federal, and possibly state, agencies into a separate micro-data file that will contain a "best citizenship" variable for every person in the 2020 Census.

The citizenship micro-data file and the CEF will be simultaneously sent through the 2020 Disclosure Avoidance System, which will do the final record linkage and place a confidentiality protected citizenship variable on the same MDF as will be used to produce the redistricting data.

CVAP data will be produced at the block-level from the MDF and released to the public by March 31, 2021.

# CVAP Data Format

No final decisions have been made regarding the methodology and format of the block-level CVAP data.

No decisions have been made regarding the future of the American Community Survey-based CVAP data that have been produced annually since 2011.

The Census Bureau's internal working group has set March 31, 2020 as the final date for determining the viability of each potential administrative data source on citizenship.

March 31, 2020 is also the final date for releasing the specifications of the CVAP data to be released by March 31, 2021.

The Census Bureau is considering the release of demonstration products based on historical data using the proposed methodology for the 2020 CVAP data.

# Data Sources, Brown et al. (2019)

**American Community Survey (ACS) in 2010, 2017**

**2010 Census**

**2010, 2017 Social Security Administration (SSA) Numident**

Misses persons without Social Security Numbers (SSNs)

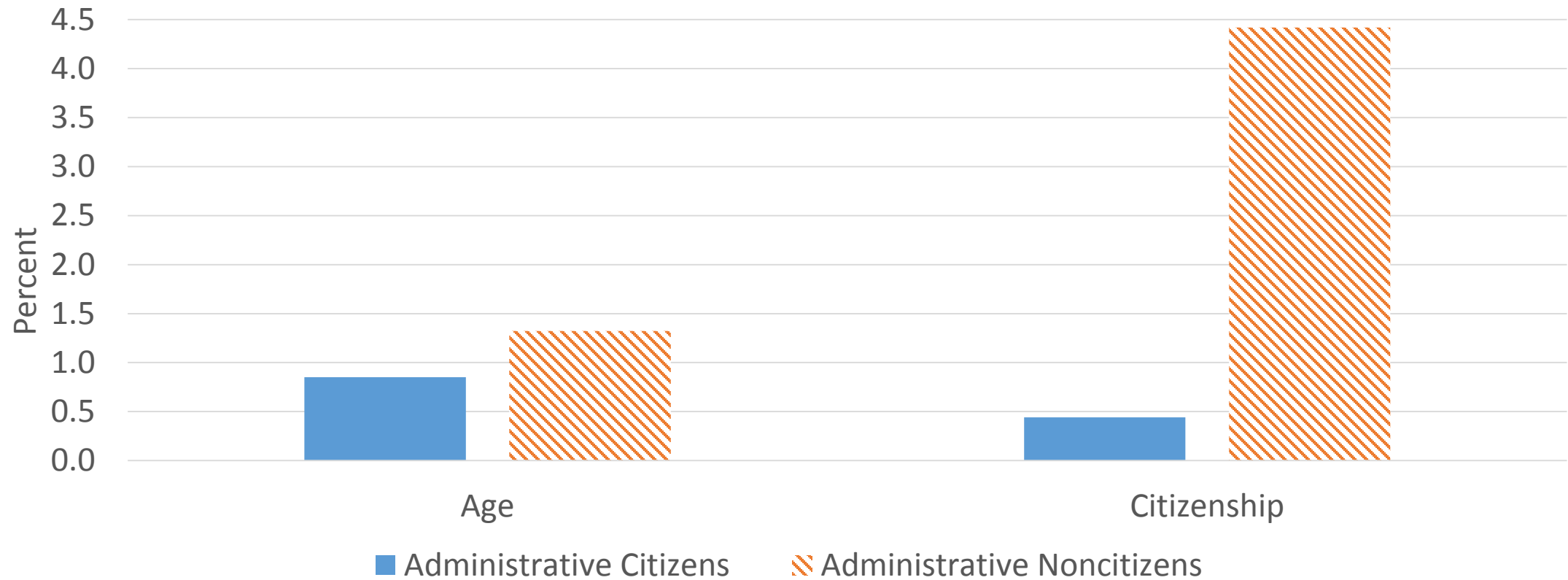
Not all naturalized persons report their status change to SSA, or they do so with delay

**Individual Tax Identification Numbers (ITINs)**

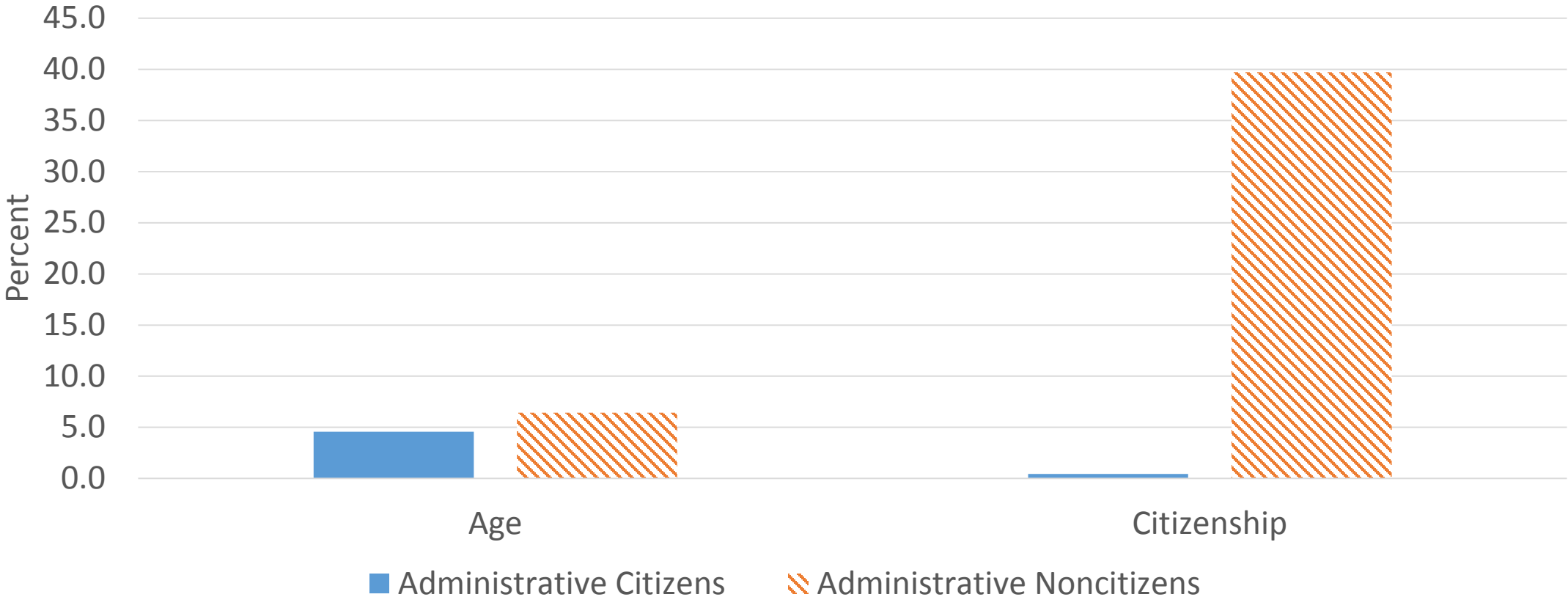
Persons who need to pay taxes, but do not have work authorization



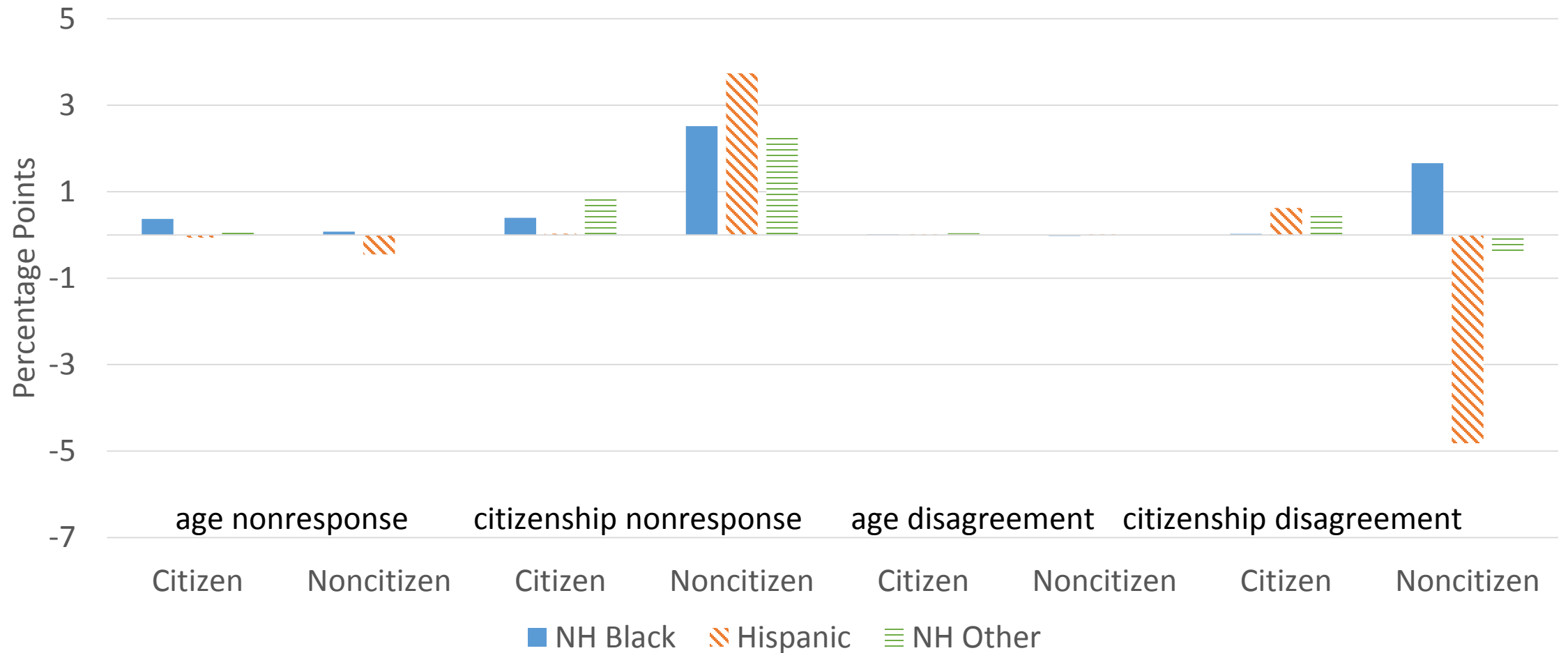
## 2017 ACS Item Nonresponse: Administrative Record Citizens and Noncitizens



# 2017 ACS-Administrative Record Disagreement: Administrative Record Citizens and Noncitizens



## Race/Ethnicity vs. Non-Hispanic White



# Measuring Effect of Citizenship Question on Self-Response Rate

Natural experiment: random sample of 1,418,000 households receiving both ACS (with citizenship question) and Census (without) in 2010

Households may be less willing to respond to one survey than the other for reasons other than citizenship question

Divide households into ones likely more vs. less sensitive to citizenship question

Less sensitive: everyone in household is citizen in ACS and administrative data

More sensitive: all other households

Difference between self-response rate across surveys for less sensitive group represents general difference in propensity to self-respond across surveys

Difference-in-differences can isolate citizenship question effect

# Blinder-Oaxaca Decomposition of Comparison of Predicted 2010 ACS to 2010 Census to Self-Response Rates by All-Citizen vs. All Other Households

	2010 ACS – 2010 Census
All other households	-20.7
AR & ACS all-citizen households	-8.9
Difference-in-differences	-11.9
Explained	-3.1
Unexplained	-8.8

## Blinder-Oaxaca Unexplained Component Using 2017 ACS Characteristics

$$UV_{2017} = E(X_{S_{2017}})' \beta_{S_{2010}} - E(X_{S_{2017}})' \beta_{U_{2010}}$$

---

	2017 ACS – 2010 Census
All other household model ( $\beta_{U_{2010}}$ )	-19.9
AR & ACS all-citizen household model ( $\beta_{S_{2010}}$ )	-11.9
Difference-in-differences	-8.0

---

N=755,000 households

# Effect on Overall Self-Response Rate

Apply 8.0 percentage point reduction in self-response to 28.1% of housing units potentially having at least one noncitizen (estimated in 2017 ACS)

Implies 2.2 percentage point reduction in housing unit self-response for the universe

At a cost of \$55 million per percentage point, implies an increase in NRFU fieldwork costs of \$121 million

# Caveats

Assumes self-response rate of all-citizen households will be unaffected by citizenship question

Some households in group potentially containing at least one noncitizen likely contain only citizens, which may understate the citizenship question effect on households actually containing at least one noncitizen

Does not capture change in degree of sensitivity to citizenship question since 2010

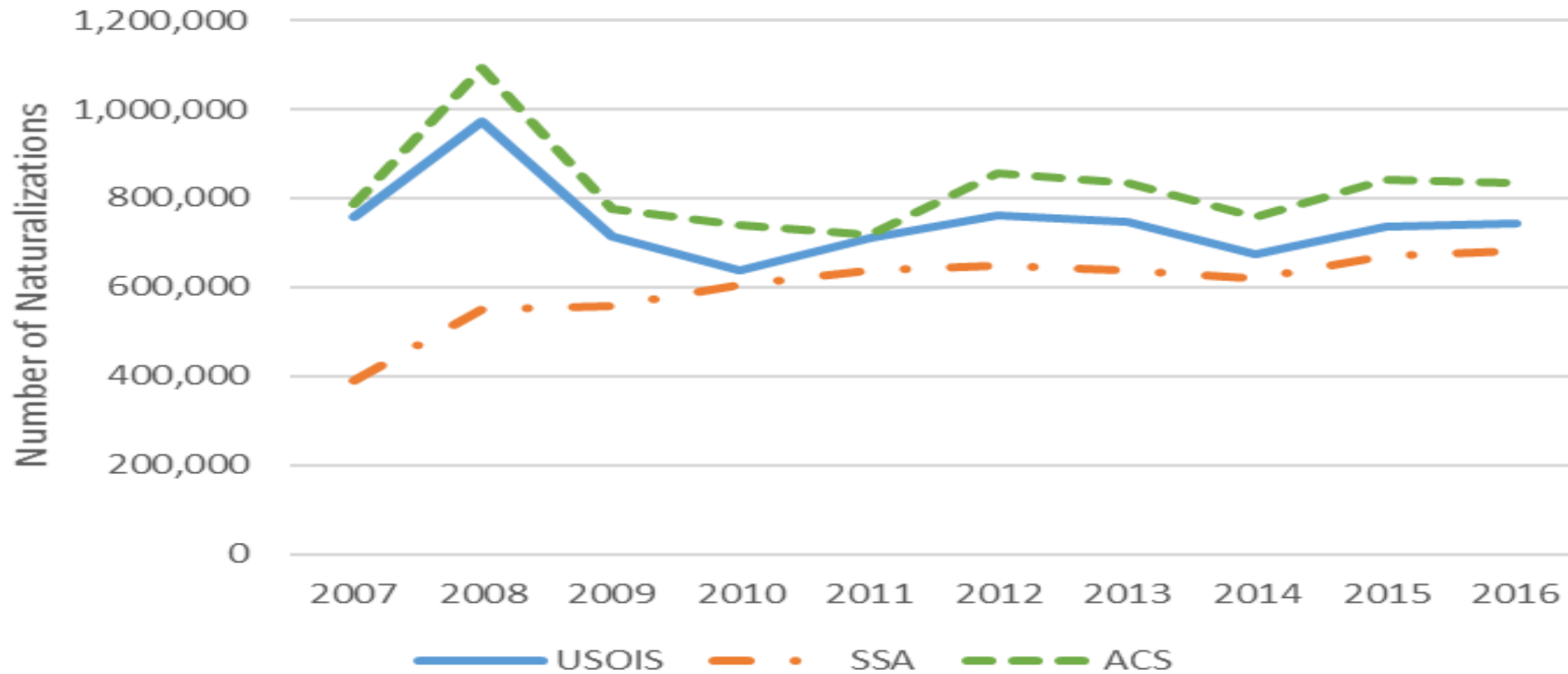


## Enumeration Quality in Mailout/Mailback and Nonresponse Follow-up (NRFU) Proxy Responses

	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

\$55 million estimated fieldwork cost for each percentage point drop in self-response rate

# Estimated Annual Naturalizations in ACS, USOIS Statistics, and SSA NUMIDENT



# Conclusions

Households potentially containing at least one noncitizen have an estimated 11.9 percentage point larger reduction in self-response to the 2010 ACS vs. the 2010 Census compared to all-citizen households

6.3-8.8 percentage points of the difference-in-differences is unexplained, which is attributed to sensitivity to the ACS citizenship question

Implies an estimated 2.2 percentage point decline in self-response overall, increasing NRFU cost by \$121 million and lowering data quality

# Using Administrative Data

Following the Secretary's March 26, 2018 instructions, modeling efforts focused on using survey responses (to the question on the 2020 Census) and administrative records

When the Supreme Court upheld the injunction on asking the question, and the President issued Executive Order 13880, modeling efforts focused on using more administrative record sources

The Director convened the Interagency Working Group, which consists of high-level executives in federal agencies that have person-level data relevant to estimating citizenship

Primarily two uses of administrative data for estimating citizenship:

- (1) Keeping the names, addresses, and other PII in the record linkage system current
- (2) Determining citizenship status from variables on the files and eligibility conditions

# Current Sources of Citizenship Data

## **Social Security Administration NUMIDENT**

Contains place of birth and citizenship status for approximately 94% of its universe

## **Individual Taxpayer Identification Numbers**

NOTE: the Census Bureau does not receive, and has not requested, application for ITIN data  
ITINs can be identified when they are used in the SSN field of a form the Census Bureau does receive

## **IRS 1040 and 1099 forms**

Primarily used to keep the record linkage system current

## **CMS Medicare and Medicaid/CHIP**

Contain some citizenship information but are primarily used to keep the record linkage system current

## **Housing and Urban Development**

Federal Housing Administration, Public and Indian Housing Information Center, Tenant and Rental Assistance Certification System, Low-income Housing Tax Credits, Computerized Homes Underwriting Management System used to keep the linkage system current

# Additional Federal Citizenship Data

## Department of Homeland Security USCIS/CBP/ICE

Lawful permanent residents and naturalization data (CIS), visas (ICE), arrival/departure (CBP)

## Department of State (Passport Services)

Passport data

## Social Security Administration

Master beneficiary record

## Indian Health Service

Patient registration

## Department of Justice

US Marshals and Citizenship and Immigration Data Collection

# Research Program

## Citizenship modeling

Develop statistical models that efficiently and accurately combine multiple sources of administrative citizenship data to estimate “best citizenship” for each person known to the Person Identification Validation System (PVS), which is the production record linkage system for the 2020 Census

Use these models to prepare a micro-data file outside the 2020 Census production system that can be combined with the 2020 Census Edited File to provide the 2020 Disclosure Avoidance System with the “best citizenship” variable to tabulate block-level CVAP tables

This research began in April 2018, final specifications and modeling details are planned for release before March 31, 2020, which is the internal deadline for finalizing the input administrative record sources

# Research Program II

## Enhanced record linkage capabilities

The production PVS can link persons found in the SSA NUMIDENT and ITIN universes; about 90% of the U.S. resident population

Many of the requested files from DHS, State, and others, are expected to provide the PII that enables record linkage for much of the balance of the resident population, provided that the PII on the 2020 Census is as reliable as it was in 2010



# Confidentiality Protection

As with all administrative data ingested by the Census Bureau, the citizenship data will be used only for statistical purposes

As with all administrative data ingested by the Census Bureau, the confidentiality of the citizenship data will be fully protected by Title 13, Section 9, which prohibits:

“... mak[ing] any publication whereby the data furnished by any particular establishment or individual under this title can be identified”

The CVAP tables will be produced using the 2020 Census Disclosure Avoidance System, which implements differential privacy using the TopDown algorithm

The CVAP tables will share the privacy-loss budget determined by the Data Stewardship Executive Policy Committee for the 2020 Census publications

# Reference

Brown, J. David, Misty Heggeness, Suzanne Dorinski, Lawrence Warren, and Moises Yi. "Predicting the Effect of Adding a Citizenship Question to the 2020 Census," *Demography* (2019) 56: 1173.

<https://doi.org/10.1007/s13524-019-00803-4>

# Questions?

Shape  
your future  
START HERE >

United States<sup>®</sup>  
Census  
2020



Thank you  
[John.Maron.Abowd@Census.Gov](mailto:John.Maron.Abowd@Census.Gov)