Noncitizen Coverage and Its Effects on U.S. Population Statistics

by

J. David Brown
U.S. Census Bureau

Misty L. Heggeness
U.S. Census Bureau

Marta Murray-Close
U.S. Census Bureau

CES 23-42 **August 2023**

Abstract

We produce population estimates with the same reference date, April 1, 2020, as the 2020 Census of Population and Housing by combining 31 types of administrative record (AR) and third-party sources, including several new to the Census Bureau with a focus on noncitizens. Our AR census national population estimate is higher than other Census Bureau official estimates: 1.8% greater than the 2020 Demographic Analysis high estimate, 3.0% more than the 2020 Census count, and 3.6% higher than the vintage-2020 Population Estimates Program estimate. Our analysis suggests that inclusion of more noncitizens, especially those with unknown legal status, explains the higher AR census estimate. About 19.8% of AR census noncitizens have addresses that cannot be linked to an address in the 2020 Census collection universe, compared to 5.7% of citizens, raising the possibility that the 2020 Census did not collect data for a significant fraction of noncitizens residing in the United States under the residency criteria used for the census. We show differences in estimates by age, sex, Hispanic origin, geography, and socioeconomic characteristics symptomatic of the differences in noncitizen coverage.

Keyword*:* Keywords: Administrative records, Population estimates, Immigration, Noncitizen coverage

* Brown: U.S. Census Bureau; j.david.brown@census.gov. Heggeness: University of Kansas and U.S. Census Bureau; misty.heggeness@ku.edu. Murray-Close: U.S. Census Bureau; marta.murray.close@census.gov. Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the U.S. Census Bureau. The Census Bureau has ensured appropriate access and use of confidential data and has reviewed these results for disclosure avoidance protection (Projects 7516813 and 7516814: CBDRB-FY23-0253, CBDRB-FY23-0255). We gratefully acknowledge the assistance of Samuel R. Cohen, Genevieve Denoeux, Suzanne Dorinski, Carl Lieberman, Linden McBride, Hongxun Qin, Allen E. Ross, Danielle H. Sandler, Lawrence Warren, and Moises Yi in preparing the data for this paper.

Introduction

The accuracy and completeness of noncitizen coverage in U.S. population statistics has wide-ranging effects. They include not only the size of the noncitizen population, but also the total population and its geographic, demographic, and socioeconomic distribution. This has implications for the study of immigration policy effects, Congressional apportionment, allocation of government funds across localities, and demographic and socioeconomic research.

We compare traditional Census Bureau population statistics to administrative record-based estimates (hereafter the AR census) with the same reference date of April 1, 2020. The AR census incorporates data from 31 federal and state government and third-party sources.[2] Though administrative record-based population estimates have been compared to counts in earlier decennial censuses, this collection of AR data includes several sources covering noncitizens that have not been used before for general population estimates.[3]

Traditional Census Bureau population estimates serve as benchmarks for each other. But in each set of counts or estimates, the data used for the foreign-born population (not a U.S. citizen at birth) come mainly or exclusively from survey-style data collection, so their coverage errors may be correlated. The 2020 Census counts are based primarily on survey-style enumeration for each housing structure in the United States thought to be potentially

[2] All data obtained with the assistance of Executive Order 13880 were incorporated into the analytic files by January 12, 2021. Brown et al. (2023) provide details on the integration process of these 31 data sources.
[3] Sailer and Weber (1998) compare Internal Revenue Service (IRS) data to the 1990 Census. Estimates using IRS, Medicare, Selective Service System, Indian Health Service, and Department of Housing and Urban Development data are compared to the 2000 Census (Farber and Leggieri 2002). Rastogi and O'Hara (2012) use the same sources as Farber and Leggieri, plus four other federal government sources and nine third-party sources to compare estimates to the 2010 Census. Our study is the first to produce AR-based U.S. population estimates by immigration status.

inhabited.[4] The Post-Enumeration Survey (PES) uses independently collected address lists and experienced enumerators to produce coverage error estimates for the decennial census (Khubba et al. 2022). However, the people not willing to respond to the PES may be similar to those not willing to respond to the decennial census, and unusual or hidden housing structures may be just as difficult to incorporate in the PES as the decennial census. Demographic Analysis (DA) estimates come from basic population accounting using Medicare data for those 75 and older, vital statistics for the U.S.-born aged 0 to 74, and the American Community Survey (ACS) for the foreign-born population aged 0 to 74 (Jensen et al. 2020). The Census Bureau's Population Estimates Program (PEP) vintage-2020 estimates for April 1, 2020, use 2010 Census, vital statistics, and ACS data (U.S. Census Bureau 2021b). The PEP foreign-born population estimates come from a combination of 2010 Census and ACS data, both of which are survey-style collections. The ACS uses PEP estimates as population controls, so any coverage error in the PEP estimates is reflected in the ACS estimates.

The undocumented population is particularly difficult to enumerate in surveys, so coverage error can be high for that group. Evans et al. (2019) report that focus group members said that undocumented immigrants had privacy concerns about 2020 Census participation, fearing that the data would be used for immigration enforcement.[5] Immigrant communities along the Texas-Mexico border, called *colonias*, are challenging to survey because of irregular housing and addressing, limited English proficiency, limited formal education, confidentiality concerns, complex households, and mobile household members (de la Puente and Stemper 2003). Census

[4] Administrative records were used to enumerate 4.59% of housing units (U.S. Census Bureau 2021a). Five of the 31 sources in the AR census were used for this.
[5] Title 13 of the U.S. Code prohibits use of Census Bureau data for enforcement purposes.

Bureau hard-to-count and low response scores include such characteristics.[6] Other than for housing tenure (renters or owners), the Census Bureau has not produced estimated undercounts by these characteristics, however.[7]

Based on case study evidence, Kissam (2017) posits that a key reason for decennial census undercounts of Mexican immigrants is that many of them live in unusual or concealed housing units not included in the Census Bureau's Master Address File (MAF). Generally, no attempts are made to collect decennial census responses from addresses omitted from the MAF.

Some recent studies suggest that the ACS underestimates the foreign-born and noncitizen population groups. Office of Immigration Statistics (OIS) noncitizen estimates are 3.4 to 4.7 million higher than the ACS estimates between 2007 and 2015 (Jasso and Rosenzweig 2020). By varying how item nonresponse to the questions used to identify the foreign-born in the ACS is handled, Mira and Bollinger (2021) estimate the noncitizen population to be between 19.7 and 38.7 million in 2019, compared to the official ACS estimate of 21.7 million. Their estimated range for the undocumented population is even wider, between 7.3 and 18.3 million.[8] Jensen et al. (2015) calculate coverage factors for the ACS foreign-born population using the 2000 Census and the 2000 Accuracy and Coverage Evaluation (ACE) Revision II (a post-enumeration survey). They find ACS foreign-born undercoverage of no more than 800,000. Undercoverage is greater

[6] Bruce et al.'s (2012) 12-factor hard-to-count score includes renters, in poverty, people aged 16 or over who are unemployed, linguistic isolation, and below high school education, among others. The Erdman and Bates (2017) 12-factor low response score includes renters, median household income, and college graduates.

[7] Khubba et al. (2022) report statistically significant estimated undercounts for renters of 1.09% and 1.48% in the 2010 and 2020 Census, respectively, based on the 2010 Census Coverage Measurement Survey and 2020 Post-Enumeration Survey.

[8] Van Hook et al. (2021) study plausible variability in undocumented population estimates based on the underlying assumptions about coverage error, emigration, and mortality. Their estimates are in a somewhat narrower range, with a 50% chance of being between 9.1 and 12.2 million and a 95% chance of being between 7.0 and 15.7 million.

for Hispanics under age 50, especially among males. Our contribution to this literature is to compare AR-based estimates to the ACS by immigration status.

Differences in coverage across groups by immigration status could affect sociodemographic statistics. According to Baker (2021a, 2021b), the demographic distributions by country of origin and age are very different for nonimmigrant visa holders and the unauthorized population. Baker (2021a) estimates that 68.7% of the 2018 unauthorized population came from Mexico, El Salvador, Guatemala, Honduras, Colombia, Brazil, and Venezuela, while Baker (2021b) reports that just 16% of nonimmigrant visa holders in 2019 came from Latin America.[9] Baker's reports estimate that 75.2% of the unauthorized population is in the 25 to 54 age group, compared to 60.5% of nonimmigrant visa holders. He reports that both groups have higher male shares than in the general population (51.4% for the unauthorized population and 57.4% for nonimmigrant visa holders compared to the AR census's 49.8% for the general population). Using the 2007 1-year ACS, Passel and Cohn (2009) report that 47% of likely undocumented immigrants aged 25 to 64 have less than a high school education, compared to 8% of U.S.-born residents. Among likely undocumented immigrant high school graduates aged 18 to 24, 49% have attended college, compared to 71% of U.S.-born residents. The median household income in 2007 is $50,000 for U.S.-born residents and $36,000 for undocumented immigrants. A fifth of likely undocumented immigrant adults live in poverty, compared to 10% of U.S.-born adults. Artiga and Diaz (2019) find that 8% of citizens, 23% of lawfully present immigrants, and 45% of likely undocumented immigrants aged 0 to 64 lack health insurance,

[9] Baker (2021b) reports that 17% of nonimmigrant visa holders were from North America, including 6% from Canada, as well as 5% from South America.

using the 2017 1-year ACS. They also report a higher incidence of having at least one full-time worker in undocumented immigrant households (86% compared to 84% for lawfully present immigrant households and 82% for citizen households).

The sources of coverage error are different in AR and survey-style collection, making a comparison of the two instructive. People who do not respond to surveys may still file taxes, enroll in a government program, or apply for a visa. The AR census has different potential errors, such as having out of date or incomplete information on some people.

Our comparison of the AR census to the DA by nativity shows that the AR census foreign-born estimate is at least 6.2 million higher than the DA's ACS-based estimate, fully accounting for the 6.1 million difference between the AR census total population estimate and the DA high estimate. The AR census noncitizen estimate exceeds the ACS estimate by at least 11.0 million. Analysis of linked and unlinked AR census and 2020 Census person records suggests that noncitizens, and especially those with unknown legal status, constitute a disproportionately high share of AR census people not in the 2020 Census.

We investigated why the AR census includes noncitizens not found in the 2020 Census. The housing structures where AR census noncitizens reside appear to have been omitted from the 2020 Census address universe at a much higher rate than for citizens. Inclusion of more noncitizens contributes not only to a higher AR census total population estimate, but especially estimates for Hispanic males aged 25 to 64 and in counties along the U.S.-Mexico border. Estimates for renters, those lacking health insurance, speakers of languages other than English at home, adults with less than a high school education, and employed people are also higher.

Data

To construct the 2020 AR census, we first compiled information on U.S. residents and their addresses from 29 federal and state administrative data sources and one third-party data source with reference dates in 2019 and 2020, listed in Appendix Table 1.[10] Hereafter, we refer to all these sources as administrative data sources.

The Census Bureau attempted to assign a unique person identifier, called a Protected Identification Key (PIK), to each administrative record. Only people with Social Security numbers (SSNs) or Individual Taxpayer Identification Numbers (ITINs) can be assigned a PIK, however.[11] To cover the noncitizen population more comprehensively, the Census Bureau conducted additional record linkage for people without SSNs or ITINs in this project. Records without PIKs, but with sufficient personally identifiable information (PII) to be linkable, were grouped into clusters that appeared to represent the same person. The clusters were assigned a unique Enhanced Protected Identification Key (EPIK). We linked records within and across sources and unduplicated them by PIK/EPIK.

We added children who were not found in those AR sources in two ways. The Census Household Composition Key (CHCK) file links the PIKs of children to the PIKs of their mother and father. We assigned children age 18 or under who were not in the AR sources to their primary

[10] We use a total of 31 sources. An additional AR source is used for young children, as described below. Here we use all data pertaining to 2019 and 2020, including late-arriving sources that were not received by the Census Bureau until 2021. Brown et al. (2023) provide a comprehensive description of the construction of the AR census.
[11] ITINs are nine-digit numbers in a publicly known range found in the SSN field of administrative records. They are issued by the Internal Revenue Service (IRS) to people needing to pay taxes, but who are ineligible for an SSN.

parent's location(s), provided that the parent was in one of the AR sources.[12] Children who according to the Social Security Administration (SSA) Numerical Identification file (NUMIDENT) were born in a U.S. state or the District of Columbia, were under the age of 2, and who were not found in the other 30 AR sources or included in the AR census through CHCK were added to the AR census in their state and city of birth.

We excluded people who had neither a PIK nor EPIK, did not have any AR with a state of residence, were not alive on April 1, 2020, according to AR sources, who were not a U.S. resident on April 1, 2020,[13] or who were linked to more than 12 addresses in 2019-2020.[14]

Some people had more than one address in AR in 2019-2020. We estimated the probability that each of the addresses was their residence on April 1, 2020, using a random forest model that was fit using linked 2018 AR and ACS data. We normalized the probabilities to sum to 1 for each person, so that they count as a single person in the aggregate statistics. When calculating population estimates for geographic areas, we weighted each person by their person-location probability in the respective geographic area.

For the 267,400,000 people who could be linked by PIK/EPIK to the 2020 Census, we used their 2020 Census values for age, sex, and race/ethnicity.[15] We used data from AR sources and

[12] We defined the primary parent as the mother, if present in the child's CHCK record, and the father otherwise.
[13] We excluded nonimmigrant visa holders who according to Arrival and Departure Information System (ADIS) data were outside the country on April 1, 2020, and did not return within 90 days, or they had an unexpired business or tourist visa. We also excluded people who had been deported before April 1, 2020, according to Immigration and Customs Enforcement (ICE) Enforcement and Removal Operations (ERO) data, and no other source indicated that they re-entered the United States. These groups do not meet the 2020 Census residency criteria.
[14] Linkage to many addresses in a short time is a symptom of record linkage error.
[15] This was done to abstract as much as possible from discrepancies between the 2020 Census reported and imputed values and those from other sources. Brown et al. (2023) document significant differences between 2020 Census values and those from other sources among people who could be linked to the 2020 Census. Changes to the race and ethnicity questions and their processing are thought to be a contributing factor (Jones et al. 2021).

previous Census Bureau decennial census and survey data, when available, for the people who could not be linked to the 2020 Census. We used modeled predicted probabilities for the remainder.

Citizenship, immigration status, and nativity variables were constructed using a combination of AR and past Census Bureau survey sources. Appendix Table 2 displays how many people are placed in each immigration status category by source. Modeled predictions were used to fill in missing values for citizenship. The appendix provides a detailed description of the construction of all the demographic variables.

The AR census has several limitations. Available AR sources do not cover every U.S. resident. We accessed driver's license data in just five states and Women, Infants, and Children (WIC) data in 14 states. Undocumented immigrants who entered the United States without inspection (EWIs) and had no interaction with the government may be missing, resulting in noncitizen omissions. On the other hand, more noncitizens, and especially nonimmigrant visa holders, may have left the country than citizens in the period between when they were last seen in 2019-2020 AR and April 1, 2020. Noncitizens could thus have elevated rates of both omissions and erroneous inclusions. AR contain incomplete death information for people without SSNs, and some of them may have died between their appearances in 2019-2020 AR sources and April 1, 2020. The demographic characteristics for the 74,200,000 AR census people who could not be linked to the 2020 Census vary from what they would have been if 2020 Census values were available. Record linkage error causes omissions, duplication, and incorrect characteristics assignment.

The traditional Census Bureau population estimate sources to which we compare the 2020 AR census contain error as well. The 2020 Census contains omissions, duplication, inclusion of people not alive on April 1, 2020, and inclusion of nonresidents, as documented by Khubba et al. (2022) and Brown et al. (2023). The PES and DA contain sampling and nonsampling error (Marra and Kennel 2022 and Jensen et al. 2020). The base for the 2020 PEP estimates is the 2010 Census, which contains omissions and erroneous enumerations (Mule 2012), and estimates of population change between 2010 and 2020 are based on similar sources to DA.

Results

We first report comparisons that can be made across all the sources. Table 1 shows that the 2020 AR census total population estimate is 6.1 million (1.8%) higher than the 2020 DA high estimate, 10.2 million (3.0%) higher than the 2020 Census count, and 12.1 million (3.6%) higher than the vintage-2020 PEP 2020 population estimate.[16]

The table breaks down the estimates by age-sex groups, allowing us to see where the gaps are wider. The difference between AR census estimates and the others is larger for males (4.1 to 7.9 million more in the AR census) than females (2.0 to 4.2 million more). Adults aged 25 to 64 account for most of the overall gap between the AR census and the other estimates. The AR census has between 3.5 and 6.4 million more males and 2.1 to 3.4 million more females aged 25 to 64 than the other sources, compared to between 251,000 and 1.3 million more in the AR census for males aged 0 to 17, from 191,000 to 1.1 million more for females aged 0 to 17,

[16] The 2020 PES estimates a 0.24% higher housing unit population than the 2020 Census (Khubba et al. 2022). It does not produce estimates including group quarters.

between 181,000 fewer and 130,000 more for males aged 65 and over, and from 690,000 to 850,000 fewer females aged 65 and over.

Hispanic estimate differences account for 48.7% of the gap between the AR census and 2020 Census for males aged 25 to 64 and 47.1% of the difference between the AR census and PEP for that group.[17] For females aged 25 to 64, differences in Hispanic estimates make up 62.4% of the difference between the AR census and the 2020 Census and 63.3% of the difference between the AR census and the PEP. These differences are much larger than the AR census shares of Hispanics in the age 25 to 64 male (20.4% Hispanic) and female (19.3% Hispanic) groups. However, they are similar to the AR census shares of Hispanics in the age 25 to 64 foreign-born male (50.5 to 50.9% Hispanic) and female (46.4 to 46.5% Hispanic) groups.[18]

The high Hispanic shares of the foreign-born aged 25 to 64 motivate our comparison of AR census and DA estimates for the foreign-born and non-foreign-born populations. The AR census and DA estimates for people aged 0 to 74 are divided into native-born, foreign-born, born abroad of U.S. parents, and born in Puerto Rico in Table 2.[19] The AR census estimates are higher overall for the latter three groups. The difference in the number of people is biggest for the foreign-born category (at least 6.2 million), which accounts for more than the entire gap between the AR census and the DA high estimates (6.1 million). The foreign-born difference

[17] DA does not produce estimates by Hispanic origin for the 25 to 64 age group. The 2020 PES estimates a 4.99% higher Hispanic population in housing units than the 2020 Census (Khubba et al. 2022).

[18] The AR census Hispanic shares among non-foreign-born people aged 25 to 64 are 11.7 to 11.8% for males and 12.1 to 12.2% for females. The AR census Hispanic percentages of the foreign-born and non-foreign-born are derived from the numbers in Tables 1 and 3.

[19] The DA estimates for age 75 and over come from Medicare data and are not divided into these groups. Table 1 shows that the DA estimates for age 75 and over are 1,000 higher than the AR census for males and 420,000 higher for females, so this age group is not one that can help explain the higher AR census estimates for the total population.

comes nearly entirely from males aged 25 to 64 (at least 3.7 million) and females aged 25 to 64 (at least 1.8 million).

Since the foreign-born population appears to be the primary source of the discrepancy between the AR census and other estimates, we break it into subcomponents to understand what segments of the foreign-born population may be more responsible than others. The ACS estimates the number of foreign-born who are naturalized citizens and noncitizens. Table 3 shows estimates using 2019 and 2021 ACS observations and the AR census by sex-age-Hispanic origin groups.[20] The AR census estimates a foreign-born population of 52.5 to 54.9 million, compared to 45.1 million in the ACS. For most groups, the ACS naturalized citizen estimate is above the range of AR census estimates, and the overall difference is at least 1.7 million. In contrast, the AR census noncitizen estimates are higher for every group, with an overall difference of at least 11.0 million. The widest gaps are among Hispanics aged 25 to 64 (at least 2.8 million for males and 1.6 million for females, compared to at least 1.8 million for non-Hispanic males and 1.2 million for non-Hispanic females).

We further disaggregate the population estimates by immigration status in Table 4, comparing the AR census to 2019 and 2021 ACS and the most recent Office of Immigration Statistics (OIS) and Pew Research Center estimates. The AR census and the ACS have virtually identical estimates for citizens, with higher AR census estimates for people born U.S. citizens and lower ones for naturalized citizens.

[20] We use the 2019 and 2021 ACS, because the 2020 ACS had an abnormally low self-response rate as a result of COVID-19, and it may be less representative of the population.

The AR census estimate for nonimmigrant visa holders is 3.8 million higher than OIS and 4.8 million higher than the Pew Research Center. OIS uses a statistical model to estimate the number of nonimmigrant visa holders in the country on any given day (Baker (2021b).[21]

The AR census noncitizen with unknown legal status estimate is 11,690,000, which is 300,000 higher than the OIS unauthorized population estimate of 11,390,000 for January 1, 2018 (Baker 2021a) and 1.19 million higher than the Pew Research Center estimate for 2017.[22] The gap between the AR census and Pew Research Center estimates is larger for all noncitizens (7.6 million) than for legal noncitizens (6.3 million), which results in the higher AR census noncitizen with unknown legal status estimate. The OIS-AR census patterns of differences appear similar to AR-Pew Research Center estimate differences, though a direct comparison cannot be made.[23]

We next study how well the demographic distributions of the AR census immigration status categories align with the gaps between the AR census and other estimates shown in Table 1. We leverage the differences in demographic distributions by immigration status reported in the literature reviewed in the introduction to understand how relative coverage of AR and traditional Census Bureau population estimates vary by immigration status. Appendix Table 3

[21] He notes that ADIS data capture most, but not all nonimmigrant visa holder exits from the country. The incomplete exit information is the reason for using a statistical model.

[22] Other recent studies raise the possibility that the undocumented population could be higher than any of these estimates. Based on Mexican administrative data, Bhandari et al. (2021) estimate an undocumented Latin American-born population between 2.2 and 3.7 million higher than the Pew Research Center estimates for that group in 2013-2017. Van Hook et al. (2021) suggest that there is significant uncertainty about the size of the unauthorized population, estimating a 95% confidence interval between 7.0 and 15.7 million.

[23] OIS decomposes the foreign-born population arriving in the United States in 1980 or later. We do not have year of entry data for the AR census foreign-born population. But it appears that the AR census estimates are higher than OIS for both all noncitizens and legal noncitizens.

shows how the AR census age-sex-Hispanic origin distribution varies across noncitizen categories, as well as how they compare to ACS noncitizen observations from 2019 and 2021. Much higher concentrations of noncitizens with unknown legal status are in the age-sex-Hispanic origin cells where the AR census has significantly higher estimates overall than the 2020 Census. Like Baker (2021a, 2021b), we find that the Hispanic share and the share aged 25-64 among nonimmigrant visa holders are much lower than those for noncitizens with unknown legal status (Hispanic shares of 36.2% for nonimmigrant visa holders and 81.8% for noncitizens with unknown legal status, and Hispanic aged 25 to 64 shares of 24.8% for nonimmigrant visa holders and 61.2% for noncitizens with unknown legal status). The male-female gap is much wider among Hispanics for noncitizens with unknown legal status and among non-Hispanics for nonimmigrant visa holders. When comparing shares for Hispanic males aged 25 to 64, only the noncitizen with unknown legal status share (37.2%) is higher than that of ACS noncitizens (22.9%). Hispanic females aged 25 to 64 exhibit this pattern as well, though it is more muted. In sum, the noncitizens with unknown legal status category are concentrated in the demographic groups where the gaps shown in Table 1 between the AR census and other population estimates are larger, unlike other immigration status categories.

We study coverage differences in the AR census and 2020 Census by examining the characteristics of person records that are included in both, those in the 2020 Census and not the AR census, and ones that may or may not be in both sources. The first column of Appendix Table 4 represents records in both sources. Some 2020 Census person records have the same unique identifier (PIK/EPIK), suggesting that they may have been enumerated multiple times. Column 2 shows the additional person records beyond the first one. They are not included in

the AR census, because the AR census unduplicates records by PIK/EPIK. The person records in column 3 are included in the 2020 Census, but they are filtered out of the AR census (e.g., they were not alive or not U.S. residents on April 1, 2020, according to AR information). Those 2020 Census records meeting all AR census eligibility requirements except having AR records in 2019 or 2020 (and thus not in the AR census) are in column 4. Column 5 is 2020 Census records that could not be assigned a PIK/EPIK, so it is not possible to uniquely link them to AR census person records. AR census person records with PIK/EPIKs not found in the 2020 Census are in column 6. People in columns 5 and 6 may overlap, but we are unable to directly link them.

Hispanics aged 25 to 64 make up a much higher share of the AR census records without links than those who can be linked to 2020 Census person records, 2020 Census duplicates, or 2020 Census records ineligible to be in the AR census. The Hispanic aged 25 to 64 share in column 6 is also higher than its share of 2020 Census person records without PIK/EPIKs (column 5), consistent with the AR census having a higher estimate for that group than the 2020 Census overall (Table 1). The Hispanic aged 25 to 64 share in 2020 Census person records lacking 2019 or 2020 AR data (column 4) is higher than its share of either linked or unlinked AR census records, suggesting that the AR census is undercovering them to a greater extent than other groups.

In Table 5, we divide AR census records into those linked by PIK/EPIK to a 2020 Census person record (column 1); unlinked people at an address where at least one 2020 Census person record lacks a PIK/EPIK and is in the same age group, sex, and Hispanic origin category (column 2); unlinked people at an address where at least one 2020 Census person record lacks a

PIK/EPIK, but none are in the same age group, sex, and Hispanic origin category (column 3); and unlinked people who are not in an address where at least one 2020 Census person lacks a PIK/EPIK. Our confidence that AR census people are in the 2020 Census is highest for those in column 1, followed by column 2 where linkage requirements are loosened. To be in the 2020 Census, a person in column 3 would either need to have discrepant demographic data or be in a different address in the 2020 Census, and a column 4 person would need to be in a different address.

Table 5 shows the share of AR census people within each immigration status group who are in each of the four linkage groups. The percentage of AR census people linked to a 2020 Census person record is about twice as high for citizens (82.3%) as noncitizens (40.9%). The share is 67.2% for lawful permanent residents, 43.5% for nonimmigrant visa holders, and 15.4% for noncitizens with unknown legal status. The share in the last two columns (where AR census people are less likely to be in the 2020 Census) is highest for noncitizens with unknown legal status (73.6%), followed by other legal noncitizens (55.0%) and nonimmigrant visa holders (53.1%). Just 14.9% of citizens are in the last two columns.

One would need to make very conservative assumptions to obtain a 2020 Census nonimmigrant visa holder population count lower than the OIS estimate of 3,090,000. If all the AR census nonimmigrant visa holders in the first column are also in the 2020 Census, none of those in the other columns are in the 2020 Census, and the number of nonimmigrant visa holders in the 2020 Census and not the AR census is zero, then the number in the 2020 Census would be 3,049,000. Just adding the people in column 2 would bring the 2020 Census count to 3,286,000,

exceeding the OIS one. In contrast, such assumptions applied to the noncitizen with unknown legal status category produce 2020 Census counts that are small fractions of the 11,390,000 OIS estimate (1.8 million using people in column 1 and 3.1 million using those in columns 1 and 2).

As mentioned above, there is some possibility that people in AR census unlinked records who do not appear to match a 2020 Census record within the AR census's housing structure may have different addresses in the 2020 Census and AR census. In the next exercise, we move from the housing structure to county level. As shown in Brown et al. (2023), 70.1% of people linked between the AR census and the 2020 Census who do not have a matching housing structure address match by county, while the share is just 21.7% for blocks and 32.2% for tracts. Also, 8.8 million AR census records contain county, but not lower levels of geography. We use sex-age group-Hispanic origin-county characteristics to match the two groups.[24] We do this by reweighting the unlinked AR census records to sum to the 2020 Census record without PIK/EPIK counts in each sex-age-Hispanic origin-county cell.[25] A caveat is that each immigration status group within a cell receives the same reweighting, though people in some groups may be more likely to be in the 2020 Census records without PIK/EPIKs in that cell than those in other groups are.

In Table 6 we compare the AR census distribution by immigration status in AR census records linked to 2020 Census person records, those not linked to 2020 Census person records, and reweighted unlinked AR census people. We also calculate the distribution for the difference in

[24] The age groups are 0-2, 3-5, 6-14, 15-17, 18-24, 25-34, 35-44, 45-54, 55-64, 65-74, and 75 and older.
[25] We collapse categories when a cell count/estimate is zero for the unlinked AR census or 2020 Census without PIK/EPIK group.

counts between the unlinked AR census records and the reweighted ones and show the implied estimated number of unlinked AR census people not among 2020 Census person records without a PIK/EPIK. The noncitizen share is much higher among unlinked records (25.6%) than among linked AR census records (5.0%). The increase in share across the first two columns among lawful permanent residents is smaller (from 2.6 to 4.6%) than it is for nonimmigrant visa holders (from 1.1 to 5.2%) and noncitizens with unknown legal status (from 0.7 to 13.2%). Comparing the fourth column to the third shows that the AR census noncitizen weights decrease much more than those of citizens to match the 2020 Census cell counts. Among noncitizens, those with unknown legal status experience the largest decrease in weights. The table implies that 6.4 million unlinked AR census noncitizens are absent from the 2020 Census. Of these missing noncitizens, 3.6 million have unknown legal status, 1.1 million are lawful permanent residents, and 1.1 million are nonimmigrant visa holders.

PEP estimates are not aggregated from individual person records like the 2020 Census, so we cannot study coverage differences through person record linkage. However, PEP estimates are used as population controls for the ACS. Table 7 compares 2016-2020 5-year ACS estimates for several characteristics likely to differ by citizenship and immigration status, where one set uses PEP controls and the other uses AR census population controls. The AR census controls are by sex-age group-Hispanic origin-citizenship-county cells.[26]

The AR census-controlled population is 4.1% higher than the PEP-controlled one. This is both because of the difference in reference dates (2016-2020 for PEP and April 1, 2020, for the AR

[26] The age groups are 0-2, 3-5, 6-14, 15-17, 18-24, 25-34, 35-44, 45-54, 55-64, 65-74, and 75 and older.

census) and a higher AR census population estimate for April 1, 2020 (Table 1). The estimates for all but two of the characteristic groups rise by more than 4.1% (column 4), suggesting that they make up larger shares of the AR census than PEP. The two exceptions are college graduates and out of the labor force. Those with less education, below-average income, in poverty, without health insurance, renters, speak a language other than English at home, and in the labor force (employed or unemployed) have higher shares when using AR census controls.

Regarding immigration status characteristics in the table, we include not only the share of noncitizens, but also variables designed to proxy for nonimmigrant visa holders and likely undocumented. The nonimmigrant visa holder proxy is noncitizens who came to the United States in the last four years.[27] The likely undocumented proxy is Hispanic noncitizens with less than high school education who arrived in the United States more than five years ago, among all people aged 25 and over.[28] The number of noncitizens increases by 37.6% when reweighting. The likely undocumented group increases more than the recently arrived noncitizen group (43.8% compared to 33.5%), suggesting that likely undocumented immigrants have a greater propensity to be excluded from the PEP estimates.

Some, but not all, of the overall ACS estimates are sensitive to the switch in population controls, including noncitizen (changes from 6.6 to 9.3%), likely undocumented (rises from 2.13

[27] This group should contain the bulk of nonimmigrant visa holders, since most people with temporary visas cannot legally stay in the country longer than this. Some people in other immigration status groups arrived recently, though.

[28] We follow Amuedo-Dorantes and Arenas-Arroyo (2018) and others here, though they do not apply the age filter. The less than high school education filter is intended to remove people on student visas and high skilled workers. The restriction to those arriving in the United States five or more years ago is designed to remove low-skilled migrants with nonimmigrant visas. We restrict the sample for this variable and the education variables to people aged 25 and over to focus on people who have completed their education.

to 3.16%), speaks another language at home (rises by 2.53 percentage points), has less than a high school education (1.09 percentage points higher), and lacks health insurance (increases by 0.61 percentage points). The employment and income estimates are relatively unaffected.

Noncitizen coverage differences could affect the geographic distribution of population estimates. We show differences in how AR census and 2020 Census and PEP population estimates vary by geography together with legal noncitizen and noncitizen with unknown legal status shares of the AR census population in Table 8. [29] The AR census population estimate is lower than both the 2020 Census and PEP in a few Western rural states, including Alaska, Idaho, Montana, North Dakota, and Wyoming, areas with well-below-average noncitizen shares. California, Nevada, and Texas have above-average differences between AR census and other population estimates as well as above-average shares of both noncitizen categories. The correlation between percent difference in the AR census and 2020 Census population and the AR census legal noncitizen share is .594, and the correlation with noncitizen with unknown legal status share is .775. For the AR census to PEP percent difference, the correlation with legal noncitizen share is .635, and the correlation with noncitizen with unknown legal status share is .556.

Population estimate gaps are much wider than average in states bordering Mexico. Counties along the border have bigger gaps than the remainder of those states, and counties with high

[29] These AR census state-level noncitizen with unknown legal status population shares are highly correlated with other estimates in the literature. The correlation with Warren and Warren's (2014) 2010 unauthorized population estimates as shares of the 2010 Census state population is .964, and the correlation with Passel and Cohn's (2018) 2016 estimates as a share of the 2016 Population Estimates Program state population is .931.

colonia concentration have larger gaps than other border counties (Table 8).[30] Border states, and particularly the high *colonia* concentration counties, are areas with high AR legal noncitizen and noncitizen with unknown legal status shares. These patterns are consistent with anecdotal evidence that immigrants near the border are especially difficult to enumerate.

Finally, we investigate whether missing housing structures from the 2020 Census address universe can help explain why the AR census includes more noncitizens than the traditional population estimate sources do. Appendix Table 5 compares the AR census immigration status distribution in housing structures included in the 2020 Census address universe (where responses were solicited) compared to other addresses. Consistent with the case studies, 19.8% of AR census noncitizens are in housing structures not in the 2020 Census universe compared to 5.7% of citizens. The differences are larger for nonimmigrant visa holders (9.1% of people in addresses outside of the 2020 Census universe compared to 1.5% of those with addresses in the universe) and noncitizens with unknown legal status (11.1% of people outside the universe and 2.7% of those in the universe). The noncitizen with unknown legal status share of people in addresses outside the 2020 Census universe (23.0%) is much higher than those for other noncitizen categories in border counties with high *colonia* concentration, whereas the nonimmigrant visa holder share is nearly as high as the noncitizen with unknown legal status share elsewhere.

[30] We follow Barton et al. (2015), who identify Cameron, El Paso, Hidalgo, Maverick, Starr, and Webb counties in Texas along the border with Mexico as having particularly high concentrations of *colonias*.

Conclusion

Facilitated by data sharing among government agencies, we produce AR-based statistics describing the population differently from Census Bureau survey-based population estimates. According to AR the U.S. population is more numerous, male, Hispanic, and middle-aged. It has more noncitizens, especially those with unknown legal status. Population shares are higher for people with less than a high school education, speaking another language at home, renters, and lacking health insurance. Counties along the U.S.-Mexico border are more populous, as are California and Texas in general. These results impact immigration policy, Congressional apportionment, government funding allocation, and socioeconomic research.

It is possible that erroneous inclusions in the AR data contribute to these differences. Counting each person only once is built into the design of the AR census, because only records with a unique identifier are used, and they are unduplicated by that identifier. The identifier may not actually be unique, however; some people could have more than one and appear multiple times in AR-based statistics.[31]

AR information about whether people were alive on April 1, 2020, may be incomplete, resulting in erroneous inclusion of some people who were not alive.[32] Such errors should be more prevalent in the estimates for the population aged 0 to 2 and 65 and over. The AR census

[31] We leave the study of duplication in Census Bureau person identifiers to future research. Khubba et al. (2022) estimate that the 2020 Census person records in housing units contain 5.2 million duplicates, and our Appendix Table 4 shows 5.8 million 2020 Census person record duplicates among those that could be assigned unique identifiers. The 54.3. million 2020 Census person records without unique identifiers surely also include duplicates.
[32] Among 2020 Census people who could be assigned unique identifiers, Brown et al. (2023) report that 2.1 million were not alive on April 1, 2020, according to AR.

estimates for the population aged 0 to 2 match the DA estimates, and the AR census estimates for those aged 65 and over are lower than those in any of the other sources, suggesting that inclusion of non-living people is not the cause of the differences between the AR census and other sources.

Information on the U.S. residency status of people in the AR census is incomplete. Some people with AR data reporting a U.S. address in 2019 or 2020 may have left the country by April 1, 2020. This is most likely to be an issue for nonimmigrant visa holders, who may be in the country for a short time and depart prior to the visa expiration date. ADIS nonimmigrant visa holder data miss some exits. Our results suggest, however, that noncitizens with unknown legal status make up a significantly larger share of people in the AR census not found in the 2020 Census than nonimmigrant visa holders do (3.6 million compared to 1.1 million in Table 6). Most undocumented immigrants are long-term residents who are unlikely to have emigrated.[33]

Though the AR census estimates are higher than those in other sources overall, they still may undercount the population. DA U.S.-born estimates are higher than the AR census by as much as 3.3 million. The 2020 Census includes 1.5 million people missing from the AR census who appear eligible to be enumerated. About 20.3% of the people only in the 2020 Census are Hispanics aged 25 to 64, so the AR census may undercount this group.[34]

[33] Passel and Cohn (2019) report that 66% of adult unauthorized immigrants in 2017 have been in the United States for more than 10 years. Massey et al. (2014) characterize them as long-term settlers as a result of increased border enforcement after 1986, while documented migrants circulate back and forth across the border with rising frequency. About 63.8% of AR census noncitizens with unknown legal status file IRS 1040 tax returns. An important reason that undocumented immigrants file taxes is to document U.S. residency, in case future immigration reform provides a path to citizenship (National Immigration Law Center 2022).

[34] Hispanics aged 25 to 64 represent just 7.8% of those known to be in both the 2020 Census and AR census.

Additional data and research are needed to refine AR-based population estimates. Including more sources covering noncitizens would help. They are in fewer AR sources on average (Brown et al. 2023), resulting in less accurate point-in-time location data. Another way to address the infrequent appearance of noncitizens in AR data is to widen the vintage window beyond two years and include the probability that a person is a U.S. resident on the reference date. That way people who appear infrequently are covered in the estimates, while accounting for the possibility that they disappeared from the data because they emigrated.[35]

As of this writing, several of the data sharing agreements are inactive. It will not be possible to improve the accuracy of the AR-based statistics for 2020, produce statistics for other years, or conduct further analysis with this rich set of AR sources unless agreements are renewed. This study demonstrates benefits of doing so.

References

Amuedo-Dorantes, C., & Arenas-Arroyo, E. (2018). Split families and the future of children: Immigration enforcement and foster care placements. *AEA Papers and Proceedings*, 108, 368-372.

Artiga, S., & Diaz, M. (2019). *Health coverage and care of undocumented immigrants (KFF report)*. San Francisco, CA: Kaiser Family Fund. Retrieved from

[35] Brown et al. (2023) provide a detailed list of recommendations for how to improve AR-based population statistics.

https://files.kff.org/attachment/Issue-Brief-Health-Coverage-and-Care-of-Undocumented-Immigrants.

Baker, B. (2021a). *Estimates of the unauthorized immigrant population residing in the United States: January 2015-January 2018* (DHS report). Washington, DC: U.S. Department of Homeland Security, Office of Immigration Statistics. Retrieved from http://dhs.gov/immigration-statistics/population-estimates/unauthorized-resident.

Baker, B. (2021b). *Population estimates of nonimmigrants residing in the United States: Fiscal years 2017-2019* (DHS report). Washington, DC: U.S. Department of Homeland Security, Office of Immigration Statistics. Retrieved from https://dhs.gov/sites/default/files/publications/immigration-statistics/Pop_Estimate/NI/ni_population_estimates_fiscal_years_2017_-_2019v2.pdf.

Baker, B. (2022). *Estimates of the lawful permanent resident population in the United States and the subpopulation eligible to naturalize: 2019-2021* (DHS report). U.S. Department of Homeland Security, Office of Immigration Statistics. Retrieved from https://dhs.gov/sites/default/files/2022-05/22_0405_plcy_lpr_population_estimates_2019_-_2021.pdf.

Barton, J., Perimeter, E. R., Blum, E. S., & Marquez, R. R. (2015). *Las colonias in the 21st century: Progress along the Texas-Mexico border* (Federal Reserve Bank of Dallas report). Dallas, TX: Federal Reserve Bank of Dallas. Retrieved from https://dallasfed.org/~/media/documents/cd/pubs/lascolonias.pdf.

Bhandari, R., Feigenberg, B., Lubotsky, D. & Medina-Cortina, E. (2021). *Projecting trends in undocumented and legal immigrant populations in the United States* (Working Paper NB21-16), Cambridge, MA: National Bureau of Economic Research.

Brown, J. D., Cohen, S. R., Denoeux, G., Dorinski, S., Heggeness, M. L., Lieberman, C., . . . Yi, M. (2023). *Real-time 2020 administrative record census simulation* (2020 Census Evaluations and Experiments final report). Washington, DC: U.S. Census Bureau.

Bruce, A., Robinson, J. G., & Devine, J. E. (2012). *A planning database to identify areas that are hard-to-enumerate and hard-to-survey in the United States*. New Orleans, LA: Proceedings of the International Conference on Methods for Surveying and Enumerating Hard-to-Reach Populations.

de la Puente, M., & Stemper, D. (2003). *The enumeration of colonias in Census 2000: Perspectives of ethnographers and census enumerators* (Census 2000 Evaluation final report). Washington, DC: U.S. Census Bureau.

Erdman, C., and Bates, N. (2017). The low response score (LRS): A metric to locate, predict, and manage hard-to-survey populations. *Public Opinion Quarterly, 81*(1), 114—156.

Evans, S., Levy, J., Miller-Gonzalez, J., Vines, M., Girón, A. S., Bates, N., & Trejo, Y. G. (2019). *2020 Census barriers, attitudes, and motivators (CBAMS) focus group final report* (U.S. Census Bureau report). Washington, DC: U.S. Census Bureau. Retrieved from https://becountedmi2020.com/wp-content/uploads/2020-report-cbams-focus-group.pdf.

Farber, J., & Leggieri, C. (2002). Building and validating a national administrative records database for the United States. *New Zealand Economic Papers, 36*(1), 65—67.

Jasso, G., & Rosenzweig, M. R. (2020). What is the size and legal composition of the US foreign-born population? *International Migration Review, 54*(3), 640—673.

Jensen, E. B., Bhaskar, R., & Scopilliti, M. (2015). *Demographic Analysis 2010: Estimates of coverage of the foreign-born population in the American Community Survey* (U.S. Census Bureau Population Division Working Paper No. 103). Washington, DC: U.S. Census Bureau. Retrieved from https://census.gov/library/working-papers/2015/demo/POP-twps0103.html.

Jensen, E. B., Knapp, A., King, H., Armstrong, D., Johnson, S. L., Sink, L., & Miller, E. (2020). *Methodology for the 2020 Demographic Analysis estimates* (U.S. Census Bureau Population Division technical documentation). Washington, DC: U.S. Census Bureau. Retrieved from https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/2020da_methodology.pdf.

Jones, N., Marks, R., Ramirez, R., & Rios-Vargas, M. (2021). *2020 Census Illuminates Racial and Ethnic Composition of the Country* (Blog in America Counts: Stories Behind the Numbers). Washington, DC: U.S. Census Bureau. Retrieved from https://www.census.gov/library/stories/2021/08/improved-race-ethnicity-measures-reveal-united-states-population-much-more-multiracial.html.

Kissam, E. (2017). Differential undercount of Mexican immigrant families in the U.S. Census. *Statistical Journal of the IAOS, 33*, 797-816.

Khubba, S., Heim, K., & Hong, J. (2022). *National census coverage estimates for people in the United States by demographic characteristics* (2020 Post-Enumeration Survey Estimation Report PES20-G-01). Washington, DC: U.S. Census Bureau.

Marra, E., & Kennel, T. (2022). *Source and accuracy of the 2020 Post-Enumeration Survey person estimates* (2020 Post-Enumeration Survey Methodology Report PES20-J-01). Washington, DC: U.S. Census Bureau.

Massey, D. S., Durand, J., & Pren, K. A. (2014). Border enforcement and return migration by documented and undocumented Mexicans. *Journal of Ethnic and Migration Studies*.

Mira, A. F., & Bollinger, C. R. (2021). *Credible interval estimates of the size and legal composition of the US foreign-born population* (Unpublished manuscript).

Mule, T. (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: U.S. Census Bureau.

National Immigration Law Center (2022). *Individual Taxpayer Identification Number (ITIN): A powerful tool for immigrant taxpayers* (Policy brief). Los Angeles, CA: National Immigration Law Center. Retrieved from nilc.org/wp-content/uploads/2022/12/Individual-Taxpayer-Identification-Number-OCT-2022PDF.pdf.

Passel, J., and Cohn, D. (2009). *A Portrait of unauthorized immigrants in the United States* (Pew Research Center Report). Washington, DC: Pew Research Center. Retrieved from https://www.pewresearch.org/hispanic/2009/04/14/a-portrait-of-unauthorized-immigrants-in-the-united-states/.

Passel, J. S., & Cohn, D. (2018). *U.S. unauthorized immigrant total dips to lowest level in a decade* (Pew Research Center Report). Washington, DC: Pew Research Center. Retrieved from https://pewresearch.org/hispanic/2018/11/27/u-s-unauthorized-immigrant-total-dips-to-lowest-level-in-a-decade/.

Passel, J. S. (2019). *Measuring illegal immigration: How Pew Research Center counts unauthorized immigrants in the U.S.* (Pew Research Center report). Washington, DC: Pew Research Center. Retrieved from https://pewresearch.org/short-reads/2019/07/12/how-pew-research-center-counts-unauthorized-immigrants-in-us/.

Rastogi, S., & O'Hara, A. (2012). *2010 Census Match Study* (2010 Census Planning Memoranda Series No. 247). Washington, DC: U.S. Census Bureau. Retrieved from https://www.census.gov/library/publications/2012/dec/2010_cpex_247.html.

Sailer, P., & Weber, M. (1998). *The IRS population count: An update* (Internal Revenue Service Statistics of Income working paper). Washington, DC: Internal Revenue Service. Retrieved from https://www.irs.gov/pub/irs-soi/1998-1999preprintar08.pdf.

U.S. Census Bureau (2021a). *2020 Census data quality operational metrics: Release 1* [Data set]. Washington, DC: U.S. Census Bureau. Retrieved from

https://www.census.gov/programs-surveys/decennial-census/decade/2020/planning-management/process/data-quality.html.

U.S. Census Bureau (2021b). *Methodology for the United States population estimates: Vintage 2020* (U.S. Census Bureau technical documentation). Washington, DC: U.S. Census Bureau. Retrieved from https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/2010-2020/methods-statement-v2020-final.pdf.

Warren, R., & Warren, J. R. (2013). Unauthorized immigration to the United States: Annual estimates and components of change, by state, 1990 to 2010. *International Migration Review, 47*(2), 296—329. https://journals.sagepub.com/doi/10.1111/imre.12022.

Table 1. Demographic Distributions

	2020 AR census	2020 Census	2020 PEP	2020 DA high estimates
U.S. total	341,600,000	331,400,000	329,500,000	335,500,000
Males	170,200,000	162,700,000	162,300,000	166,100,000
Age 0-2	5,842,000	5,453,000	5,803,000	5,842,000
Age 3-5	6,426,000	5,983,000	6,130,000	6,209,000
Age 6-14	19,630,000	19,310,000	18,870,000	19,590,000
Age 15-17	6,629,000	6,632,000	6,392,000	6,635,000
Age 18-24	15,930,000	15,870,000	15,340,000	15,470,000
Age 25-34	24,230,000	22,480,000	23,490,000	23,360,000
Age 35-44	22,580,000	20,890,000	21,050,000	21,750,000
Age 45-54	21,820,000	20,100,000	19,920,000	20,810,000
Age 55-64	22,170,000	20,980,000	20,490,000	21,350,000
Age 65-74	15,330,000	15,520,000	15,180,000	15,510,000
Age 75+	9,618,000	9,483,000	9,638,000	9,619,000
Females	171,400,000	168,700,000	167,200,000	169,400,000
Age 0-2	5,588,000	5,236,000	5,559,000	5,588,000
Age 3-5	6,056,000	5,736,000	5,864,000	5,937,000
Age 6-14	18,750,000	18,450,000	18,070,000	18,670,000
Age 15-17	6,337,000	6,305,000	6,137,000	6,345,000
Age 18-24	15,270,000	15,390,000	14,690,000	14,860,000
Age 25-34	23,440,000	22,280,000	22,600,000	22,730,000
Age 35-44	21,980,000	21,300,000	21,090,000	21,470,000
Age 45-54	21,400,000	20,770,000	20,440,000	20,850,000
Age 55-64	22,590,000	22,430,000	21,910,000	22,250,000
Age 65-74	17,070,000	17,600,000	17,370,000	17,340,000

	2020 AR census	2020 Census	2020 PEP	2020 DA high estimates
Age 75+	12,920,000	13,200,000	13,470,000	13,340,000
Non-Hispanic males	134,500,000	131,800,000	131,400,000	
Age 0-2	4,370,000	4,087,000	4,300,000	4,225,000
Age 3-5	4,638,000	4,475,000	4,560,000	4,495,000
Age 6-14	14,220,000	14,310,000	14,030,000	14,090,000
Age 15-17	4,836,000	4,938,000	4,818,000	4,847,000
Age 18-24	11,890,000	12,090,000	11,800,000	11,640,000
Age 25-34	18,590,000	17,596,000	18,480,000	
Age 35-44	17,490,000	16,548,000	16,540,000	
Age 45-54	17,400,000	16,511,000	16,270,000	
Age 55-64	18,820,000	18,390,000	17,920,000	
Age 65-74	13,710,000	14,130,000	13,810,000	
Age 75+	8,552,000	8,732,000	8,870,000	
Non-Hispanic females	137,200,000	137,500,000	136,800,000	
Age 0-2	4,138,000	3,917,000	4,110,000	4,034,000
Age 3-5	4,405,000	4,279,000	4,351,000	4,293,000
Age 6-14	13,570,000	13,650,000	13,410,000	13,420,000
Age 15-17	4,622,000	4,692,000	4,621,000	4,646,000
Age 18-24	11,430,000	11,800,000	11,310,000	11,260,000
Age 25-34	18,170,000	17,520,000	17,990,000	
Age 35-44	17,280,000	16,930,000	16,860,000	
Age 45-54	17,300,000	17,070,000	16,840,000	
Age 55-64	19,410,000	19,650,000	19,230,000	
Age 65-74	15,250,000	15,940,000	15,760,000	
Age 75+	11,580,000	12,080,000	12,340,000	

	2020 AR census	2020 Census	2020 PEP	2020 DA high estimates
Hispanic males	35,690,000	30,900,000	30,920,000	
Age 0-2	1,472,000	1,366,000	1,503,000	1,618,000
Age 3-5	1,788,000	1,508,000	1,570,000	1,715,000
Age 6-14	5,407,000	4,997,000	4,843,000	5,496,000
Age 15-17	1,793,000	1,694,000	1,574,000	1,788,000
Age 18-24	4,042,000	3,783,000	3,541,000	3,831,000
Age 25-34	5,637,000	4,884,000	5,013,000	
Age 35-44	5,092,000	4,342,000	4,515,000	
Age 45-54	4,423,000	3,589,000	3,653,000	
Age 55-64	3,351,000	2,594,000	2,574,000	
Age 65-74	1,619,000	1,392,000	1,368,000	
Age 75+	1,066,000	750,600	768,400	
Hispanic females	34,270,000	31,180,000	30,390,000	
Age 0-2	1,450,000	1,319,000	1,449,000	1,554,000
Age 3-5	1,651,000	1,457,000	1,513,000	1,645,000
Age 6-14	5,184,000	4,802,000	4,662,000	5,250,000
Age 15-17	1,715,000	1,613,000	1,516,000	1,699,000
Age 18-24	3,844,000	3,590,000	3,382,000	3,595,000
Age 25-34	5,275,000	4,759,000	4,614,000	
Age 35-44	4,697,000	4,373,000	4,228,000	
Age 45-54	4,105,000	3,702,000	3,598,000	
Age 55-64	3,184,000	2,781,000	2,682,000	
Age 65-74	1,821,000	1,663,000	1,611,000	
Age 75+	1,344,000	1,121,000	1,135,000	

Notes: Numbers for subgroups do not always add to the totals because of rounding. Source: U.S. Census Bureau, 2020 Demographic Household Characteristics file, Population Estimates Program vintage-2020 estimates for April 1, 2020, 2020 Demographic Analysis, and 2020 AR census.

Table 2. Comparison Between Demographic Analysis and AR Census by Nativity

	Native population	Foreign-born population	Born abroad of U.S. citizen parents	Net migration from Puerto Rico	Resident population
Male age 0-24					
DA high	50,050,000	2,985,000	585,100	121,900	53,750,000
AR maximum in category	49,890,000	3,736,000	568,700	270,900	
AR minimum in category	49,890,000	3,736,000	567,900	270,600	
AR total					54,460,000
Male age 25-64					
DA high	69,590,000	16,250,000	940,300	491,600	87,270,000
AR maximum in category	68,640,000	20,170,000	1,654,000	633,500	
AR minimum in category	68,630,000	19,920,000	1,399,000	631,400	
AR total					90,830,000
Male age 65-74					
DA high	13,150,000	2,127,000	98,570	132,100	15,500,000
AR maximum in category	12,510,000	2,046,000	810,700	97,610	
AR minimum in category	12,500,000	1,923,000	687,700	97,450	
AR total					15,330,000
Male 0-74					
DA high	132,800,000	21,360,000	1,624,000	745,700	156,500,000
AR maximum in category	131,000,000	25,960,000	3,034,000	1,002,000	
AR minimum in category	131,000,000	25,580,000	2,655,000	999,400	
AR total					160,600,000
Female age 0-24					
DA high	47,920,000	2,788,000	571,500	110,700	51,390,000
AR maximum in category	47,820,000	3,352,000	569,800	250,900	
AR minimum in category	47,820,000	3,351,000	569,000	250,700	
AR total					51,990,000
Female age 25-64					
DA high	69,090,000	16,850,000	920,900	448,800	87,310,000
AR maximum in category	68,500,000	18,760,000	1,656,000	623,100	
AR minimum in category	68,480,000	18,660,000	1,551,000	622,000	
AR total					89,410,000
Female age 65-74					
DA high	14,460,000	2,622,000	102,200	154,400	17,340,000

AR maximum in category	13,880,000	2,476,000	1,064,000	125,900	
AR minimum in category	13,700,000	2,180,000	768,700	125,800	
AR total					17,070,000
Female 0-74					
DA high	131,500,000	22,260,000	1,595,000	713,900	156,000,000
AR maximum in category	130,200,000	24,590,000	3,290,000	999,900	
AR minimum in category	130,000,000	24,190,000	2,889,000	998,500	
AR total					158,500,000

Notes: The nativity of some AR people is ambiguous. The AR maximum row allocates all the people in the ambiguous categories into the cell group, and the AR minimum row includes only people clearly belonging in the cell. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-0253). Source: U.S. Census Bureau, 2020 Demographic Analysis and 2020 AR census.

Table 3. Comparison Between ACS and AR Census Naturalized Citizen and Noncitizen Estimates

	ACS naturalized	AR minimum naturalized	AR maximum naturalized	ACS noncitizen	AR minimum noncitizen	AR maximum noncitizen
Male			Hispanic			
Age 0-24	242,900	168,900	180,200	971,000	1,744,000	1,753,000
Age 25-64	2,770,000	2,363,000	2,434,000	4,927,000	7,746,000	7,782,000
Age 65-74	476,900	320,900	352,500	264,700	511,400	531,500
Age 75+	284,300	184,200	235,000	106,700	408,900	433,600
All ages	3,774,000	3,037,000	3,202,000	6,269,000	10,410,000	10,500,000
Female						
Age 0-24	239,200	171,100	177,000	851,600	1,445,000	1,449,000
Age 25-64	2,956,000	2,778,000	2,828,000	4,302,000	5,868,000	5,897,000
Age 65-74	590,800	412,000	438,400	287,900	492,000	502,700
Age 75+	448,100	280,700	342,800	164,900	385,800	404,300
All ages	4,234,000	3,642,000	3,786,000	5,606,000	8,191,000	8,253,000
Male			Non-Hispanic			
Age 0-24	516,200	454,400	464,500	951,700	1,348,000	1,354,000
Age 25-64	4,824,000	4,606,000	4,831,000	3,338,000	5,159,000	5,199,000
Age 65-74	1,051,000	629,400	727,000	267,600	454,800	480,800
Age 75+	722,200	421,900	630,300	133,300	255,500	302,200
All ages	7,113,000	6,112,000	6,653,000	4,691,000	7,217,000	7,336,000
Female						
Age 0-24	556,100	471,000	479,100	901,300	1,251,000	1,255,000
Age 25-64	5,593,000	5,336,000	5,439,000	3,463,000	4,626,000	4,647,000
Age 65-74	1,266,000	789,100	1,066,000	334,900	479,700	539,000
Age 75+	1,077,000	597,700	1,250,000	221,700	331,100	455,700
All ages	8,492,000	7,194,000	8,234,000	4,921,000	6,688,000	6,897,000
Total	23,610,000	19,990,000	21,880,000	21,490,000	32,510,000	32,990,000

Notes: The immigration status is ambiguous for some AR people. The AR maximum columns include all the people who could potentially be in this category, and the AR minimum columns include only those clearly in the category. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-0253). Source: U.S. Census Bureau, 2019 and 2021 1-year American Community Survey, and 2020 AR census.

Table 4. Population Estimates by Immigration Status

Citizenship and legal status category	AR census	ACS	Office of Immigration Statistics	Pew Research Center
Total	341,600,000	330,100,000		
U.S. citizen	308,700,000	308,600,000		
Born U.S. citizen	287,300,000	285,000,000		
Naturalized U.S. citizen	19,980,000	23,610,000		20,700,000
Citizen of unknown type	1,411,000			
Noncitizen	32,510,000	21,490,000		24,900,000
Lawful permanent resident	10,460,000		13,350,000	12,300,000
Refugee	85,350			
Asylee	54,510			
Nonimmigrant visa holder	7,009,000		3,190,000	2,200,000
Other legal noncitizen	3,211,000			
Noncitizen with unknown legal status	11,690,000		11,390,000	10,500,000
Unknown citizenship and legal status	478,300			

Notes: The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-0253). Other legal noncitizens are people in the AR census who are known to be legal noncitizens, but their type of legal status is not known. The Office of Immigration Statistics (OIS) estimates for lawful permanent residents are for January 2020, their estimates for nonimmigrant visa holders are from October 1, 2018, to September 30, 2019, and their unauthorized population estimates (in the noncitizen with unknown legal status row) are for January 1, 2018. The Pew Research Center estimates are for 2017. Source: U.S. Census Bureau, 2020 AR census, 2019 and 2021 1-year American Community Survey; Baker (2021a, 2021b, 2022); Passel (2019).

Table 5. AR Census Linkage to 2020 Census by Immigration Status

	Linked to person in 2020 Census	Not linked to person in 2020 Census, address occupied in 2020 Census, at least one 2020 Census person record lacks PIK/EPIK, matched to 2020 Census person at address by age-sex-Hispanic origin at address	Not linked to person in 2020 Census, address occupied in 2020 Census, at least one 2020 Census person record lacks PIK/EPIK, not matched to 2020 Census person at address by age-sex-Hispanic origin at address	Not linked to person in 2020 Census, not in address where at least one 2020 Census person record lacks PIK/EPIK
U.S. citizen	82.27	2.79	5.46	9.48
Born U.S. citizen	82.38	2.77	5.39	9.46
Naturalized citizen	80.28	3.18	6.68	9.90
Citizen of unknown citizenship status at birth	86.04	2.09	3.24	8.65
Noncitizen	40.85	6.19	18.41	34.55
Lawful permanent resident	67.15	3.61	11.02	18.27
Refugee	58.86	2.78	10.36	27.98
Asylee	47.92	5.30	19.26	27.50
Nonimmigrant visa holder	43.50	3.38	9.61	43.51
Other legal noncitizen	41.33	3.63	14.87	40.17
Noncitizen of unknown legal status	15.43	10.92	31.33	42.28
Unknown citizenship and legal status	47.86	3.00	11.53	37.61
Total	78.28	3.12	6.70	11.91

Notes: The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-0253). Source: U.S. Census Bureau, 2020 Census Edited File and 2020 AR census.

Table 6. Unlinked AR Census Distribution by Immigration Status using 2020 Census Person Records without PIK/EPIK as Population Controls

	AR census people linked to 2020 Census	AR census people not linked to 2020 Census	AR census people not linked to 2020 Census, weighted by 2020 Census without PIK/EPIKs	Unlinked AR census – unlinked AR census weighted by 2020 Census without PIK/EPIKs	Estimated number of unlinked AR census people not in 2020 Census without PIK/EPIK group
Total people	266,900,000	74,170,000	54,310,000	19,860,000	19,860,000
U.S. citizen	94.96	74.05	76.45	67.52	13,410,000
Born U.S. citizen	88.50	68.46	71.11	61.28	12,170,000
Naturalized citizen	6.01	5.32	5.05	6.06	1,204,000
Citizen of unknown citizenship status at birth	0.45	0.26	0.29	0.18	35,750
Noncitizen	4.97	25.62	23.24	32.12	6,379,000
Lawful permanent resident	2.63	4.64	4.32	5.52	1,096,000
Refugee	0.02	0.05	0.05	0.04	7,944
Asylee	0.01	0.04	0.04	0.03	5,958
Nonimmigrant visa holder	1.14	5.23	5.13	5.50	1,092,000
Other legal noncitizen	0.49	2.51	2.29	3.11	617,600
Noncitizen of unknown legal status	0.68	13.15	11.41	17.92	3,559,000
Unknown citizenship and legal status	0.09	0.33	0.32	0.38	75,470

Notes: Only AR census people where county of residence is known are included here. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-0253). Source: U.S. Census Bureau, 2020 Census Edited File and 2020 AR census.

Table 7. ACS Characteristics with Population Estimates Program and AR Census Population Controls

	PEP-weighted percent	AR-weighted percent	Percent among additional people when AR-weighted	AR – PEP percent difference
Total population	100.00	100.00	100.00	4.08
Noncitizens	6.64	9.32	73.61	37.62
Came to United States in last four years	1.71	2.31	16.62	33.48
Likely undocumented among those age 25+	2.13	3.16	23.21	43.84
Less than high school among those age 25+	11.47	12.56	33.57	14.04
College graduate among those age 25+	32.93	32.51	24.52	3.77
Employed, age 16+	54.61	54.80	58.96	4.91
Unemployed, age 16+	3.05	3.07	3.37	5.02
Out of labor force, age 16+	42.34	42.13	37.67	4.07
In poverty	12.84	12.89	14.10	4.70
No health insurance	9.02	9.63	24.26	10.61
Rented or occupied without rent, housing unit residents	33.50	34.08	47.14	6.00
Speaks another language at home, age 5+	21.54	24.07	81.51	15.39
	PEP-weighted	AR-weighted	Among additional people when AR-weighted	
Mean total income among those age 15+	41,640	41,430	36,890	

Notes: The PEP weights refer to the final ACS weights, which use PEP population controls. The AR weights are by AR census sex-age group-Hispanic origin-citizenship-county cells. The AR – PEP percent difference is (AR-weighted estimate – PEP-weighted estimate) x 100/(AR-weighted estimate + PEP-weighted estimate)/2. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-0253).
Source: U.S. Census Bureau, 2016-2020 5-year ACS and AR census.

Table 8. Total Population and Noncitizen Differences by Geography

	Percent difference between AR census and 2020 Census total population	Percent difference between AR census and 2020 PEP total population	AR census percent legal noncitizens	AR census percent noncitizens with unknown legal status
U.S. total	3.03	3.61	6.12	3.42
Alabama	-0.30	1.75	1.60	1.49
Alaska	-0.57	-0.27	3.58	0.95
Arizona	3.41	-0.28	6.98	4.16
Arkansas	1.61	0.98	2.62	1.95
California	7.61	8.04	10.84	7.11
Colorado	2.19	1.61	4.84	3.69
Connecticut	1.70	3.07	6.13	2.62
Delaware	2.01	2.32	4.54	2.60
District of Columbia	2.14	-1.19	8.73	3.43
Florida	3.87	2.99	8.56	3.23
Georgia	1.85	1.85	4.33	3.64
Hawaii	0.68	4.04	9.77	1.76
Idaho	-1.15	-0.49	2.78	1.95
Illinois	3.15	4.88	5.84	4.14
Indiana	1.32	1.78	2.71	1.96
Iowa	2.63	3.45	3.57	1.28
Kansas	1.25	2.07	3.37	2.41
Kentucky	2.33	2.97	2.28	1.66
Louisiana	2.10	2.38	2.10	1.58
Maine	1.10	1.98	2.48	0.43
Maryland	4.03	6.01	6.33	4.62
Massachusetts	1.54	3.49	8.18	2.39
Michigan	1.09	2.21	3.40	1.06
Minnesota	1.17	2.03	3.68	1.64
Mississippi	2.01	1.80	1.21	1.00
Missouri	0.62	0.66	2.19	0.89
Montana	-1.96	-1.68	1.35	0.35
Nebraska	4.00	5.23	4.25	2.70
Nevada	5.18	4.12	7.57	6.04
New Hampshire	-0.07	0.80	3.00	0.56
New Jersey	4.95	9.42	9.12	3.77
New Mexico	1.64	2.21	6.10	3.98
New York	3.21	7.56	10.69	3.50

	Percent difference between AR census and 2020 Census total population	Percent difference between AR census and 2020 PEP total population	AR census percent legal noncitizens	AR census percent noncitizens with unknown legal status
North Carolina	0.76	-0.76	3.45	3.05
North Dakota	-2.57	-0.79	3.10	0.62
Ohio	1.26	2.20	2.43	0.84
Oklahoma	1.40	0.85	2.79	2.27
Oregon	2.59	2.47	4.28	2.65
Pennsylvania	0.15	1.86	3.54	0.91
Rhode Island	-1.01	2.71	6.31	1.84
South Carolina	4.19	2.25	2.23	1.74
South Dakota	1.22	0.55	2.39	0.92
Tennessee	1.55	1.90	2.34	2.04
Texas	4.95	4.23	8.17	5.74
Utah	1.18	1.86	3.99	3.44
Vermont	-2.73	0.40	2.65	0.35
Virginia	1.71	2.18	5.17	2.60
Washington	2.89	3.03	6.58	3.47
West Virginia	-0.17	0.34	0.87	0.25
Wisconsin	0.34	1.38	2.38	1.62
Wyoming	-2.14	-3.07	2.04	1.52
Border counties with high *colonia* concentration	8.44	8.73	11.93	5.92
Other border counties	8.10	6.75	9.80	4.44
Non-border counties in border states	5.92	5.57	9.39	6.31
Non-border states	2.03	2.93	5.10	2.48

Notes: The percent differences are calculated as 100 x (AR census – other estimate)/[(AR census + other estimate)/2]. The noncitizen probabilities for people with unknown citizenship or legal status are placed in the legal noncitizen category. The border counties with high *colonia* concentration are Cameron, El Paso, Hidalgo, Maverick, Starr, and Webb in Texas. The other border counties are Cochise, Pima, Santa Cruz, and Yuma in Arizona; San Diego and Imperial in California; Doña Ana, Hidalgo, and Luna in New Mexico; and Brewster, Hudspeth, Kinney, Presidio, Terrell, Val Verde, and Zapata in Texas. The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-0253 and CBDRB-FY23-0255).

Source: U.S. Census Bureau, 2020 Detailed Household Characteristics file, vintage-2020 Population Estimates Program estimates for April 1, 2020, and 2020 AR census.

Appendix

This appendix documents the sources used in the AR census and the methods for constructing citizenship, nativity, immigration status, age, sex, race, and ethnicity. It shows demographic distributions for AR census and ACS people by immigration status and for AR census and 2020 Census people by whether they can be linked or not. It also displays the immigration status distribution of people with addresses in the 2020 Census address universe compared to those at other addresses.

Appendix Table 1 shows the 31 sources used to build the 2020 AR census, the shares of people in each source, and the shares only in a particular source or group of sources. IRS sources cover most people (88.0%), though their coverage of noncitizens is somewhat lower (74.0%). VSGI is the second-most important source overall (41.3% of all people and 17.3% of noncitizens). U.S. Citizenship and Immigration Services (USCIS) sources cover 21.4% of noncitizens, Customs and Border Protection (CBP) data have 11.9%, and Immigration and Customs Enforcement data account for 5.0%. Noncitizens with unknown legal status are mainly in IRS (75.7%), CBP (11.7%), VSGI (9.9%), ICE (6.7%), and USCIS (6.5%) data. IRS is the only source for 60.8% of noncitizens with unknown legal status.

Citizenship

The rules we used to construct the citizenship variable are as follows. A person was classified as a U.S. citizen if they had information suggesting they were a citizen on April 1, 2020, in the NUMIDENT (either a U.S. citizen or missing citizenship but born in a place where birthright

citizenship was granted)[36], U.S. passport, and/or USCIS naturalization certificate data. People without evidence of being a citizen in the above sources and who are classified as noncitizens in at least one of the following sources were classified as non-U.S. citizens: NUMIDENT; USCIS lawful permanent residents, refugees and asylees; USCIS TPS; USCIS noncitizens receiving Deferred Action for Childhood Arrivals, classified as Special Immigrant Juveniles, or otherwise interacting with USCIS and appearing to lack a lawful immigration status; ITINs; BOP; USMS; ADIS; ERO; WRAPS; SEVIS; IMARS; LEMIS; DMDC; SNAP; TANF; state driver's licenses; and as-reported noncitizen responses in the ACS, American Housing Survey (AHS), Current Population Survey (CPS), and the Survey of Income and Program Participation (SIPP) obtained before April 1, 2020. People not meeting the criteria above and who had a citizen value from one of the following sources were classified as citizens: BOP; USMS; ADIS; DMDC; SNAP; TANF; state driver's licenses; and as-reported citizen responses in the ACS, AHS, CPS, and SIPP.[37]

Nativity

To make detailed comparisons to the DA estimates, we applied hierarchical rules to separate people into the different groups of U.S. residents used to construct DA estimates. We classified a person as part of the native population if their place of birth in the NUMIDENT is a U.S. state or the District of Columbia. The foreign-born category consists of people who have a USCIS naturalization certificate, have ever been a noncitizen in the NUMIDENT, are a noncitizen

[36] This included the 50 states, District of Columbia, Guam, Puerto Rico, the U.S. Virgin Islands, and the Commonwealth of the Northern Mariana Islands. It also included births prior to October 29, 2019 on a U.S. military base outside the United States. USCIS did not automatically grant U.S. citizenship to children born on overseas U.S. military bases between October 29, 2019 and September 17, 2020 (Britsky [2019] and USCIS [2020]).
[37] Abowd et al. (2020) conducted analysis supporting this set of rules.

according to the rules described above, or have an ITIN PIK or EPIK. Those who have always been a citizen in the NUMIDENT and not born in a U.S. state, the District of Columbia, or Puerto Rico were placed in the born abroad of U.S. citizen parents category.[38] Puerto Rico-born people in the NUMIDENT were put in the Puerto Rico-born group. We create four groups for people who do not clearly fit into one of the four DA categories. Two groups are for people whose NUMIDENT record does not show where they were born. One group is people who are NUMIDENT citizens who have never been noncitizens in the NUMIDENT. The second is those with blank NUMIDENT citizenship. The former group could be either in the native population or born abroad of U.S. citizen parents.[39] The latter group could be in the native population, born abroad of U.S. citizen parents, or foreign-born. Another group is people born somewhere other than the 50 states, District of Columbia, or Puerto Rico according to the NUMIDENT and have blank NUMIDENT citizenship. They could be born abroad of U.S. citizen parents or foreign-born. A residual category was made for people not fitting any of the above criteria. When comparing to the DA estimates, we allocate the ambiguous groups to different DA categories to calculate minimum and maximum values for the categories.

Immigration Status

We assigned detailed immigration status using the following hierarchical rules. If a person was classified as being born in a place with birthright citizenship or born abroad of U.S. citizen parents, the person was classified as born a U.S. citizen. Foreign-born citizens were placed in

[38] This may overestimate the born abroad of U.S. citizen parents category. Foreign-born people who did not have U.S. citizen parents could have received their SSN before or after being naturalized.
[39] This classification does not incorporate the possibility that they are foreign-born people who received their SSNs after being naturalized.

the naturalized citizen category. People were assigned to the citizen with unknown citizenship at birth group if they are a citizen, but it is not clear whether the person was born a citizen or not. The lawful permanent resident category includes noncitizens born in American Samoa, the Swain Islands, and lawful permanent residents in the USCIS data. People in the WRAPS data or classified as refugees in the USCIS data were classified as refugees. Those who are asylees in the USCIS data were classified as asylees. The nonimmigrant visa holder category includes people in TPS and those in ADIS or SEVIS with valid temporary visas as of April 1, 2020. Other noncitizens who appear to have lawful status (persons in legal noncitizen categories in the NUMIDENT who are not found in ADIS, ERO, IMARS, LEMIS, SEVIS, TPS, or other USCIS data) were placed in the other legal noncitizen category. Noncitizens with unknown legal status include people who are in ERO;[40] IMARS; LEMIS; USCIS data for people thought to be without legal status; in SEVIS with a visa start date after April 1, 2020; have an ITIN, or have a visa overstay in ADIS.[41] People not fitting any of the above criteria were placed in the unknown citizenship and legal status category. We used a modeled probability of being a citizen for them in the citizenship statistics.

[40] Those with an ERO final order date on or before April 1, 2020, and a removal date on April 1, 2020, or after were placed in the noncitizen with unknown legal status category even if they had information indicating being a lawful permanent resident, having a temporary visa, or otherwise appearing to have legal status. This is because noncitizens who had legal status could lose it in the ERO process. If there is evidence that a person obtained legal status (a lawful permanent resident, refugee, asylee, valid temporary visa, or TPS) after their ERO removal date, but by April 1, 2020, the legal status category was used in our classification. If a person in the ERO data had evidence of legal status and did not have a subsequent ERO removal date by April 1, 2020, they were placed in their particular noncitizen with legal status category rather than the unknown legal status category.

[41] We call this an unknown legal status category, because in most cases the data are not sufficiently detailed to classify people as definitely lacking legal status. For example, ITIN holders do not have the right to work, but some may still have legal status (e.g., as a dependent of someone with a work visa). A person who lacked legal status at one time may have gained it by April 1, 2020. We may not observe some people's evidence of legal status because of missing links to their visa records.

Appendix Table 2 shows how many people are in each detailed immigration status category. It also provides the number of people assigned to the category based on different kinds of AR source information.[42] The NUMIDENT provides evidence of citizenship status for most U.S. citizens.[43] Lawful permanent resident, refugee, and asylee status come mainly from USCIS, and ADIS provides most nonimmigrant visa holder information. The noncitizen with unknown legal status category is primarily people with ITINs and no other immigration status information. About 1.9 million are DACA, SIJ, or other immigrants without evidence of legal status in USCIS data, 1.7 million are visa overstayers in ADIS, and 1.1 million are in ERO.

Age, Sex, Race, and Ethnicity

For each person-location pair in the AR census, we assigned a probability that the given person in the given place was in each of the categories we defined for each of the following demographic characteristics: age (0-2, 3-5, 6-14, 15-17, 18-24, 25-34, 35-44, 45-54, 55-64, 65-74, or 75 and older), sex (male or female), race (American Indian or Alaska Native (AIAN), Asian, Black, Native Hawaiian or Pacific Islander (NHPI), White, Some Other Race, or Two or More Races), and ethnicity (Hispanic or non-Hispanic). As explained below, for people who could not be linked to the 2020 Census, we looked for information about each person's demographic characteristics in a set of AR and past Census Bureau survey files. If we could classify a person with respect to a given characteristic based on these files, we assigned the person to the

[42] The sources that inform immigration status are not necessarily the same ones that cause a person-location pair to be included in the AR census. For example, a person could be included in the AR census at a location based on information in a Tax Year 2019 IRS 1040 return, while the immigration status may be informed by a USCIS naturalization record, U.S. passport record, or another immigration source.

[43] The table is organized in a hierarchical way. We turn to naturalization certificates, passports, and other sources only when a person does not have evidence of U.S. citizenship in the NUMIDENT. Many people have evidence of U.S. citizenship in multiple sources.

selected category with certainty at all their locations in the AR census. Otherwise, we used a predictive model to give the person a full set of location-dependent demographic probabilities.

We used different data sources to obtain demographic information about people identified by an SSN PIK, ITIN PIK, or EPIK. For people with an SSN PIK not found in the 2020 Census, we computed their age on April 1, 2020, using their date of birth as recorded in the NUMIDENT.[44] We obtained their sex directly from the NUMIDENT. The race and ethnicity of people with an SSN PIK not found in the 2020 Census are from the Census Best Race File (Ennis et al., 2018). The Best Race File consolidates information from administrative records, household survey data, decennial census responses, and third-party data to assign people race and ethnicity categories.

We obtained the age and sex of people with an ITIN PIK not found in the 2020 Census from the following data sources, listed in order of priority: the administrative data sources where we found their person-location pairs, their most recent ACS response between 2010 and 2019, or their response in the 2010 Census.[45] If a person with an ITIN PIK did not appear in the 2020 Census or Best Race File, we looked for their race and ethnicity in the data sources contributing those person-location pairs. If an administrative source lacked data on race or ethnicity but included country of birth or country of nationality, we assigned people in that source a probability of being each race or ethnicity using country-specific race and ethnicity shares. We computed the shares by linking people in the NUMIDENT who were born in each country to

[44] ITIN PIKs and EPIKs are not in the NUMIDENT.

[45] Because of time constraints, we did not process demographic information for ITIN PIKs from administrative data sources that had fewer than 20,000 records with ITIN PIKs. We used demographic values from the ACS and CEF only if they are as-reported, edited for consistency, or taken from a previous survey.

their records in the Best Race File. If a person with an ITIN PIK had conflicting demographic information from multiple administrative data sources, we computed the probability that the person was in each demographic category, giving equal weight to the information from each source.

Neither the NUMIDENT nor the Best Race File covers EPIKs, so if the EPIK was not found in the 2020 Census, we obtained age, sex, race, and ethnicity from the administrative data sources that contributed EPIKs to the AR census, as well as the 2015-2019 ACS and the 2010 Census.[46] We used the same methodology to process the demographic information in these sources as we used for people with an ITIN PIK.

If we were able to classify a person with respect to a given demographic characteristic based on the data sources described above, we set the person's probability of being in the selected category equal to one and their probability of being in each other category equal to zero at all of their locations in the filtered person-location dataset.[47] If we were not able to classify a person with the data sources described above, we used logistic and multinomial logistic regression models to predict their probability of being in each category based on their own

[46] In addition to the administrative data sources that contributed EPIKs to the AR census, we looked for the demographic characteristics of people with an EPIK in several sources that were not sufficiently recent to contribute person-location pairs but that included demographic information about some of the people in more-recent sources. The additional sources were the AZ SNAP and TANF files from 2018, and Medicare file from 2017. We were not able to obtain demographic information about people with an EPIK from the 2010-2014 ACS files because records in these files were not assigned EPIKs.

[47] There was one exception to this rule: If a person had a value for ethnicity from the Best Race File (a "best ethnicity") that was derived exclusively from third-party data, we treated the value as missing and imputed a probability that the person was Hispanic and a probability that they were non-Hispanic. We did this because best ethnicities derived exclusively from third-party data were assigned using modeling rather than a business-rules approach, and we preferred to apply our own models.

characteristics and characteristics of their location.[48] In this second case, we allowed the person's demographic probabilities to vary across their locations in the filtered person-location dataset.[49] We scaled the predicted probabilities for each person-location pair so that they summed to one.[50]

Appendix Tables 3, 4, and 5 are discussed in the main text.

References

Abowd, J. M., Bell, W. R., Brown, J. D., Hawes, M. B., Heggeness, M. L., Keller, A. D., . . . Yi, M. (2020). *Determination of the 2020 U.S. Citizen Voting Age Population (CVAP) Using Administrative Records and Statistical Methodology* (Center for Economic Studies Working Paper CES 20-33). Washington, DC: U.S. Census Bureau.

Britsky, H. (2019). *Some children of US troops born overseas will no longer get automatic American citizenship* (News story). Military.com. Retrieved from https://www.military.com/daily-news/2019/08/28/children-us-troops-born-overseas-will-no-longer-get-automatic-american-citizenship.html.

[48] Brown et al. (2023) provides details about the models.

[49] We added a small number of records to the filtered person-location dataset after the demographic modeling process was complete. To assign demographic probabilities to the subset of these records that did not have "as-reported" demographics from the reference, composite, or administrative data files, we computed the probability that a randomly selected person from the target person's county was in each possible category for each demographic characteristic.

[50] Person-location records in Puerto Rico are included in this normalization. Thus, if a person had AR in both Puerto Rico and a U.S. state or the District of Columbia, their contribution to the AR census U.S. population estimates was less than 1.

Ennis, S. R., Porter, S. R., Noon, J., & Zapata, E. (2018). When race and Hispanic origin reporting are discrepant across administrative records and third party sources: Exploring methods to assign responses. *Statistical Journal of the IAOS*, 34, 179—189.

U.S. Citizenship and Immigration Services (2020). *USCIS implements new law related to citizenship for children of military members and U.S. government employees stationed overseas* (U.S. Citizenship and Immigration Services press release). Washington, DC; U.S. Citizenship and Immigration Services. Retrieved from https://www.uscis.gov/news/alerts/uscis-implements-new-law-related-to-citizenship-for-children-of-military-members-and-us-government.

Appendix Table 1. Percent of All People, Foreign-Born, Noncitizens, and Noncitizens with Unknown Legal Status in Each Administrative Record Source

Source	Percent of all people in this source	Percent of all people only in this source	Percent of foreign-born in this source	Percent of foreign-born only in this source	Percent of noncitizens in this source	Percent of noncitizens only in this source	Percent of noncitizens with unknown legal status in this source	Percent of noncitizens with unknown legal status only in this source
Internal Revenue Service (IRS) 1040	75.94	21.24	66.56	17.54	61.32	24.20	63.79	39.02
IRS 1099	59.95	4.33	53.29	6.28	41.89	7.80	32.99	9.43
IRS 1099R	13.98	0.00	5.48	0.00	2.34	0.00	0.86	0.00
Any IRS	88.00	36.77	79.26	35.07	73.98	43.73	75.68	60.76
Medicare	15.01	0.09	8.13	0.11	3.64	0.06	0.40	0.01
U.S. Federal Housing Administration mortgage insurance contracts	3.38	0.04	3.43	0.06	1.84	0.04	0.69	0.02
U.S. Department of Housing and Urban Development Public and Indian Housing Information Center, Tenant Rental Assistance Certification System, and Computerized Homes Underwriting Management System	0.91	0.08	0.83	0.06	0.34	0.04	0.05	0.01
Indian Health Service Patient Registration System	0.38	0.02	0.05	0.00	0.04	0.00	0.03	0.00
U.S. Social Security Administration (SSA) Master Beneficiary Record	14.88	0.04	7.12	0.02	3.37	0.01	0.39	0.00
SSA Supplemental Security Record and Special Veterans Benefits	2.54	0.14	2.71	0.07	1.31	0.06	0.27	0.02
Selective Service System Registration File	2.62	0.04	1.27	0.06	1.20	0.08	0.82	0.03
U.S. Postal Service National Change of Address	4.62	0.16	3.51	0.17	2.75	0.18	1.74	0.24
U.S. State Department Passports	5.20	0.25	4.42	0.16	0.05	0.01	0.01	0.00

Source	Percent of all people in this source	Percent of all people only in this source	Percent of foreign-born in this source	Percent of foreign-born only in this source	Percent of noncitizens in this source	Percent of noncitizens only in this source	Percent of noncitizens with unknown legal status in this source	Percent of noncitizens with unknown legal status only in this source
U.S. State Department Worldwide Refugee Admissions Processing System (WRAPS)	0.01	0.01	0.06	0.04	0.09	0.06	0.00	0.00
U.S. Customs and Border Protection Arrival and Departure Information System (ADIS)	1.15	0.96	7.44	6.19	11.92	9.94	11.72	11.50
U.S. Immigration and Customs Enforcement (ICE) Enforcement and Removal Operations (ERO)	0.24	0.19	1.55	1.25	2.49	2.01	6.68	5.53
ICE Student and Exchange Visitor Information System (SEVIS)	0.24	0.10	1.53	0.62	2.46	0.99	0.35	0.10
U.S. Citizenship and Immigration Service (USCIS) Naturalizations, Lawful Permanent Residents, Refugees, and Asylees	3.56	0.18	22.52	1.19	18.09	1.79	2.60	0.28
USCIS persons thought to be without lawful status on April 1, 2020 (e.g., Deferred Action for Childhood Arrivals (DACA), Special Immigrant Juveniles (SIJ))	0.27	0.05	1.74	0.30	2.77	0.48	6.50	1.12
USCIS Temporary Protected Status (TPS)	0.05	0.00	0.31	0.01	0.49	0.01	0.18	0.00
U.S. Department of Defense, Defense Manpower Data Center deployment data	0.03	0.00	0.01	0.00	0.00	0.00	0.00	0.00
U.S. Department of Interior (DOI) Law Enforcement Management Information System; DOI Incident Management Analysis Reporting System	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
U.S. Bureau of Prisons	0.11	0.06	0.13	0.09	0.16	0.12	0.32	0.28
U.S. Marshals Service	0.02	0.02	0.08	0.08	0.11	0.11	0.27	0.26

Source	Percent of all people in this source	Percent of all people only in this source	Percent of foreign-born in this source	Percent of foreign-born only in this source	Percent of noncitizens in this source	Percent of noncitizens only in this source	Percent of noncitizens with unknown legal status in this source	Percent of noncitizens with unknown legal status only in this source
Veteran Service Group of Illinois (VSGI)	41.33	4.83	30.52	5.73	17.28	6.35	9.92	3.37
Alaska Permanent Fund Dividend File	0.12	0.00	0.05	0.00	0.03	0.00	0.00	0.00
State Driver's Licenses	2.82	0.15	0.95	0.18	0.74	0.13	0.32	0.07
SNAP, TANF, WIC	3.74	0.84	2.56	0.51	2.00	0.50	1.11	0.42
Household Composition Key	1.05	1.05	0.06	0.06	0.08	0.08	0.01	0.01
2019 Medicaid	0.95	0.95	1.28	1.28	1.41	1.41	0.35	0.35
SSA Numerical Identification file (NUMIDENT) [ages 0-1]	0.43	0.43	0.00	0.00	0.00	0.00	0.00	0.00

Notes: The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-0253). Source: U.S. Census Bureau, 2020 AR census.

Appendix Table 2. AR Census Estimates by Immigration Status and Source

Citizenship and legal status category	Total people
Born Citizen	287,300,000
NUMIDENT, always citizen, with birthright citizenship by birthplace	259,300,000
NUMIDENT, missing citizenship variable, with birthright citizenship by birthplace	21,490,000
NUMIDENT, birthplace not one with birthright citizenship, always citizen	6,452,000
Naturalization	19,980,000
NUMIDENT, birthplace not one with birthright citizenship, switched from noncitizen to citizen	11,500,000
USCIS, naturalization certificate	7,813,000
U.S. passport, birthplace not one with birthright citizenship	588,300
ADIS, birthplace not one with birthright citizenship	38,770
Other citizenship source, birthplace not one with birthright citizenship	40,880
Citizen, unknown citizenship at birth	1,411,000
NUMIDENT, citizen, unknown birthplace, citizenship not always observed	32,600
U.S. passport, unknown birthplace	1,065,000
ADIS, unknown birthplace	136
Other citizenship source, unknown birthplace	313,200
Lawful Permanent Resident	10,460,000
USCIS	10,390,000
U.S. national	74,270
Refugee	85,350
USCIS refugee	63,760
WRAPS refugee	21,590
Asylee (USCIS asylee)	54,510
Nonimmigrant visa holder	7,009,000
ADIS	6,592,000
SEVIS	195,100
TPS	222,600
Other legal noncitizen (NUMIDENT, legal noncitizen)	3,211,000
Noncitizen of unknown legal status	11,690,000
ADIS, visa overstay	1,748,000
ERO	1,062,000
USCIS, DACA, SIJ, others without status	1,887,000
SEVIS	5,499
IMARS, LEMIS	74
Born in U.S. territories not automatically conferring lawful permanent residence at birth	470
Other SSN PIK (NUMIDENT, noncitizen of unknown type)	212,800
Other ITIN PIK (not linked to any other immigration status information)	6,666,000

Citizenship and legal status category	Total people
Other EPIK (not linked to any immigration status information)	104,800
Unknown citizenship and legal status (NUMIDENT, missing citizenship, missing birthplace or not born in place with birthright citizenship)	478,300
Total people	341,600,000

Notes: The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-0253). Source: U.S. Census Bureau, 2020 AR census.

Appendix Table 3. ACS and AR Census Noncitizen Percent Distribution by Demographic Characteristics

	ACS noncitizen	AR lawful permanent resident	AR refugee	AR asylee	AR nonimmigrant visa holder	AR other lawful noncitizen	AR noncitizens with unknown legal status	AR unknown citizenship or legal status
Male					Hispanic			
Age 0-24	4.52	2.60	1.30	10.53	4.36	3.75	8.89	2.06
Age 25-64	22.93	18.56	1.89	7.66	12.94	16.99	37.23	7.24
Age 65-74	1.23	2.46	0.23	0.18	0.83	1.60	1.23	4.21
Age 75+	0.50	1.33	0.10	0.06	0.39	0.90	1.82	5.16
All ages	29.17	24.94	3.52	18.44	18.52	23.24	49.17	18.67
Female								
Age 0-24	3.96	2.53	1.35	8.95	4.07	3.53	6.64	0.90
Age 25-64	20.02	16.79	1.27	7.27	11.81	14.89	23.97	5.83
Age 65-74	1.34	2.43	0.07	0.07	1.17	1.45	0.93	2.25
Age 75+	0.77	1.67	0.04	0.03	0.58	1.29	1.10	3.86
All ages	26.09	23.41	2.72	16.32	17.63	21.17	32.64	12.85
Male					Non-Hispanic			
Age 0-24	4.43	4.26	25.12	11.62	7.73	4.52	1.61	1.30
Age 25-64	15.53	16.71	24.90	25.98	24.90	21.54	8.03	8.38
Age 65-74	1.25	2.40	1.00	0.62	0.94	2.66	0.44	5.43
Age 75+	0.62	1.30	0.50	0.25	0.43	1.56	0.33	9.75
All ages	21.83	24.67	51.52	38.47	34.00	30.27	10.40	24.86
Female								
Age 0-24	4.19	4.09	22.05	9.30	7.36	4.12	1.30	0.88
Age 25-64	16.11	18.15	18.66	16.90	21.03	17.15	5.80	4.27
Age 65-74	1.56	2.83	0.97	0.36	1.01	2.16	0.36	12.40
Age 75+	1.03	1.94	0.56	0.21	0.46	1.88	0.29	26.07
All ages	22.90	27.01	42.24	26.77	29.86	25.31	7.76	43.63

	ACS noncitizen	AR lawful permanent resident	AR refugee	AR asylee	AR nonimmigrant visa holder	AR other lawful noncitizen	AR noncitizens with unknown legal status	AR unknown citizenship or legal status
Total people	21,490,000	10,460,000	85,350	54,510	7,009,000	3,211,000	11,690,000	478,300

Notes: The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-0253). Source: U.S. Census Bureau, 2019 and 2021 1-year American Community Survey, and 2020 AR census.

Appendix Table 4. Percent of People by Age-Sex-Hispanic Origin for Linked and Unlinked Records

	Linked 2020 Census-AR census	2020 Census with PIK/EPIKs, Duplicates	In 2020 Census with PIK/EPIKs, ineligible according to AR	In 2020 Census with PIK/EPIKs, Missing from AR census	2020 Census without PIK/EPIKs	Unlinked AR census
Total people	267,100,000	5,805,000	2,691,000	1,504,000	54,330,000	74,520,000
Male			Hispanic			
Age 0-24	3.63	4.77	1.18	5.88	5.97	6.43
Age 25-64	3.73	5.66	5.71	11.64	8.83	11.47
Age 65+	0.61	0.53	1.73	0.74	0.79	1.42
All ages	7.97	10.95	8.62	18.26	15.59	19.32
Female						
Age 0-24	3.53	4.32	0.91	4.71	5.53	5.92
Age 25-64	4.08	5.34	5.18	8.65	7.62	8.55
Age 65+	0.81	0.83	2.18	0.98	0.91	1.34
All ages	8.42	10.49	8.27	14.34	14.05	15.81
Male			Non-Hispanic			
Age 0-24	11.82	17.47	2.24	16.24	12.90	11.26
Age 25-64	21.41	15.71	21.39	19.74	18.50	20.29
Age 65+	7.40	5.88	15.02	2.75	4.26	3.36
All ages	40.62	39.05	38.65	38.73	35.65	34.92
Female						
Age 0-24	11.37	17.18	2.02	13.10	12.36	10.45
Age 25-64	22.59	14.89	21.42	12.41	17.11	15.87
Age 65+	9.03	7.45	21.04	3.19	5.24	3.64
All ages	42.98	39.52	44.48	28.70	34.71	29.95

Notes: The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-0253). Source: U.S. Census Bureau, 2020 Census Edited File and 2020 AR census.

Appendix Table 5. Among People in Addresses in 2020 Census Collection Universe and Others, Percent with Characteristics

	All		Border counties with high *colonia* concentration		Other border counties		Non-border counties in border states		Non-border states	
	In Census universe	Not in Census universe	In Census universe	Not in Census universe	In Census universe	Not in Census universe	In Census universe	Not in Census universe	In Census universe	Not in Census universe
Citizen	91.75	72.88	84.86	62.75	88.18	62.39	86.61	59.61	93.35	78.46
Born citizen	84.04	66.58	76.88	57.63	77.85	55.11	77.33	54.88	86.12	71.57
Naturalized citizen or citizen with unknown citizenship status at birth	6.20	4.72	7.86	4.76	10.22	6.90	9.20	4.51	5.25	4.73
Noncitizen	7.96	26.09	15.01	37.10	11.74	37.39	13.27	40.16	6.32	20.13
Lawful permanent resident, refugee, or asylee	3.00	3.24	8.48	6.32	5.29	6.95	4.74	3.13	2.40	3.10
Nonimmigrant visa holder	1.47	9.10	1.88	6.20	1.95	12.01	2.05	16.31	1.29	6.31
Other legal noncitizen	0.79	2.64	1.13	1.61	1.25	2.53	1.27	2.57	0.64	2.69
Noncitizen with unknown legal status	2.71	11.11	3.53	22.97	3.26	15.90	5.20	18.15	1.99	8.02
Unknown citizenship or legal status	0.12	0.26	0.20	0.49	0.13	0.56	0.12	0.25	0.13	0.24
Total people (millions)	317.7	23.97	2.431	0.342	5.143	0.5315	68.06	6.39	242.1	16.71

Notes: The data presented in this table are approved for dissemination by the DRB (CBDRB-FY23-0255). Source: U.S. Census Bureau, 2020 Census Unedited File and 2020 AR census.