

Vintage 2022 Release Notes

Vintage 2022 data products are associated with Data Management System projects P-6000042, P-7501659, and P-7527355. The U.S. Census Bureau reviewed these data products for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release (CBDRB FY23 0063).

The Vintage 2022 population estimates reflect the following methodological changes since the release of the Vintage 2021 estimates. For more detail, please refer to the full Vintage 2022 methodology statement: <http://www.census.gov/programs-surveys/popest/technical-documentation/methodology.html>.

Base Population

Whereas the latest decennial census typically serves as the population estimates base, the 2020 Census could not be adopted for this purpose due to unique challenges including the unavailability of specific data required for estimates processing and delays caused by the COVID-19 pandemic. As a result, the Population Estimates Program developed a process for integrating multiple data sources at varying levels of demographic and geographic detail to produce what we refer to as the blended base. For Vintage 2022, these sources were:

- **2020 Census Data:** 2020 Census data from the internal Census Edited File (CEF) tabulated into 2022 geographies at the subcounty level, infused with differentially private noise, and then aggregated to create county, state, and national total resident, household, and group quarters (by facility type)¹ population counts
- **2020 Demographic Analysis (DA)² Estimates:** National population estimates by age and sex
- **Vintage 2020 Population Estimates for April 1, 2020:** Nation, state, and county population estimates by age, sex, race, and Hispanic origin

In the Vintage 2022 blended base at the national level, resident, household, and group quarters population totals were derived from the 2020 Census, age and sex detail was drawn from 2020 DA, and race and Hispanic origin detail came from the Vintage 2020 estimates for April 1, 2020.

The development of the Vintage 2022 blended base for Puerto Rico followed a simplified version of this process. Since these estimates are produced for the Commonwealth and municipios by age and sex, and not for lower levels of geography, capturing geographic changes was not a factor. Additionally, there is no DA control available for Puerto Rico. The Vintage 2022 blended base for Puerto Rico consisted of the following:

¹ The seven major GQ facility types utilized in estimates production are: correctional institutions, juvenile institutions, nursing homes, other institutional facilities, college dormitories, military housing, and other noninstitutional facilities. While we do not release data on GQ by facility type, we do use them to calculate population universes such as “civilian noninstitutionalized.”

² The 2020 DA estimates of the national population by age, sex, race, and Hispanic origin on April 1, 2020 are developed from current and historical vital records, estimates of international migration, and Medicare records. The DA estimates are independent from the 2020 Census and are used to calculate net coverage error, one of the two main ways the U.S. Census Bureau uses population estimates to measure coverage of the census. For more information, see <https://www.census.gov/programs-surveys/decennial-census/about/coverage-measurement/da.html>

- **2020 Census Data:** Commonwealth and municipio total resident, household, and GQ population counts from the PL 94-171 Redistricting File
- **V2020 Population Estimates for April 1, 2020:** Commonwealth and municipio population estimates by age, sex, and universe (including GQ facility type)

Vital Statistics Estimates

For Vintage 2022, the data and methods were updated in two ways. First, we eliminated the race bridging process previously needed to convert the four 1977 race categories used in the vital records to the 31 race groups required for estimates processing. The National Center for Health Statistics (NCHS) now provides all race data in the categories defined by the Office of Management and Budget in 1997, so the race conversion is no longer necessary for births or deaths. Second, we utilized more recent NCHS provisional data to manage the lag in the available final data. This allowed us to account for recent trends in mortality and natality caused by the COVID-19 pandemic. The recency of provisional data varies by component. For births, we incorporated provisional monthly estimates by state for January 2021 through March 2022. For deaths, we incorporated national provisional monthly estimates by age, sex, and race along with monthly state-level totals for January 2021 through June 2022. We used characteristics data when available, but otherwise relied on distributions from the last year of final birth and death data available, 2020, to assign characteristics to the vital events in the provisional data. The availability of more recent data allowed us to simplify the short-term projection method at the national and subnational levels.

Net International Migration Estimates

Vintage 2022 features two major methodological changes to the net international migration (NIM) component. The combined impact from the following method changes increased national-level NIM totals compared to Vintage 2021 due to higher in-migration and lower out-migration:

1. NIM was updated to include migrants who were **born abroad of U.S. citizen parents**. The native-born migration component does not adequately measure this type of migration because our input data sources for most foreign countries collect country-of-birth (instead of country of citizenship) to identify Americans. Therefore, we used American Community Survey (ACS) migration (based on residence one year ago) and citizenship responses to include these migrants in the foreign-born immigration component. For the foreign-born emigration component, we expanded ACS input data to include born abroad of U.S. citizen parents. Since these migrants are part of the native-born (U.S. citizens at birth) population, we relabeled the “native-born” and “foreign-born” components as “U.S.-born” and “non-U.S.-born” respectively. This method change increased immigration levels.
2. The ACS-to-ACS residual method was updated to minimize the impact of higher ACS non-response on non-U.S.-born emigration estimates. Previous estimates were based on **variable rates** calculated from implied emigration trends in 2006-2019 ACS 1-year files. However, higher non-response can cause the residual method to inflate emigration. Our internal evaluation verified the increase in ACS non-response coincided with higher emigration estimates, especially after 2016. In addition, the ACS updated the migration (year of entry) question in 2015, which also caused the residual method to inflate emigration. Finally, input data from the 2020 ACS 1-year file were not

available due to disruptions to the survey caused by the COVID-19 pandemic. This would disrupt the residual method since processing requires the five most-recent consecutive ACS 1-year files to estimate emigration for the current year. For Vintage 2022, the residual method was updated to use **fixed rates** based on implied migration trends from ACS 1-year files averaged between the 2010-2014, 2010-2015, and 2010-2016 periods. This method change reduced non-U.S.-born emigration and will reduce annual volatility in the estimates.

In addition, Vintage 2022 includes the following adjustments to account for the COVID-19 pandemic:

3. The COVID-19 adjustments from Vintage 2021 were retained for the 2020 (April 1 - June 30, 2020) and 2021 (July 1, 2020 - June 30, 2021) estimates periods. Vintage 2022 estimates for this period are higher than Vintage 2021 for most states and counties due to the method changes described in the previous section.
4. Non-U.S.-born immigration was adjusted for the 2022 (July 1, 2021 - June 30, 2022) estimates period to account for the recovery in migration not shown in 2021 ACS input data. This adjustment projected immigration to pre-pandemic levels. Similar to Vintage 2021, this projection is based on current immigration trends seen in administrative data from the Department of State, U.S. Citizenship and Immigration Services, Institute of International Education, and Department of Justice. The adjustment also accounted for Afghan and Ukrainian humanitarian migrants reported by the Department of Homeland Security. The other NIM components did not receive an adjustment for 2022 because the current input data restored migration to pre-pandemic levels.

Vintage 2022 also includes the following update to NIM characteristics and geographic distributions in response to a lack of input data from the 2020 ACS 1-year file:

5. The 2020 (April 1 - June 30, 2020) national and state NIM characteristics used pooled 2017, 2018, and 2019 ACS 1-year files. The 2021 (July 1, 2021 - June 30, 2022) national and state NIM characteristics used pooled 2018, 2019, and 2021 ACS 1-year files. The 2021 characteristics were held constant for 2022 (July 1, 2021 – June 30, 2022). The 2022 characteristics will be updated next vintage when the 2022 ACS 1-year file is available. This update did not impact national NIM totals. However, characteristics and geographic distributions shifted as compared to Vintage 2021.

Puerto Rico Estimates

Domestic and international flight data were obtained from the Bureau of Transportation Statistics Air Carrier Statistics, T-100 Segment database. The flight-based methodology used to calculate net migration for the Commonwealth of Puerto Rico was robust, thus a COVID-19 adjustment was not necessary. Adjustments from previous vintages due to atypical seasonal trends in flight data were retained.

Previously, the expected population for each municipio on April 1, 2010 was projected from the Census 2000 count by accounting for change since that census due to births and deaths. The residual was converted into an annual average migration rate by age and sex for each municipio. These rates were then controlled to the national rate. Starting in Vintage 2022, these rates were calculated for the 2010 to 2020 period using 2010 Census counts and Vintage 2021 blended base estimates. This is not a method change,

but rather an incorporation of updated data. The new net migration rates had a small-to-medium impact on most municipios.

Despite moving to a flight-based methodology to calculate net migration, we still used ACS/Puerto Rico Community Survey (PRCS) files to calculate demographic characteristics for in- and out-migration flows. Net migration characteristics by sex came from residence one year ago estimates from the 2021 ACS/PRCS file. The 5-year ACS/PRCS file (proxy universe) was used to determine other demographic characteristics, which were then applied to the current year. The 5-year ACS/PRCS files represent an average over the 5-year period and not a specific point in time.³ In 2015, when the three-year ACS/PRCS files were discontinued, the Population Estimates Program began to use file “centering” or the midpoint year of the 5-year file for their estimates. This method has been discontinued and file “anchoring” of the 5-year ACS/PRCS files is now used, which “anchors” estimates at the last year of the file. This change improves data recency and accuracy, and aligns with the ACS Data User Handbook updated guidance.

Geographic Changes in Connecticut

In Vintage 2022, as a result of the formal request from the state, [Connecticut transitioned from eight counties to nine planning regions](#). Although household and group quarters (by major type) populations were available in the new Connecticut planning regions for use in estimates production, other input data and information on demographic characteristics were not available in the updated geographic boundaries. As a result, the Population Estimates Program applied data from the 2020 Census (featuring a small amount of differentially private noise as per agency guidelines) at the Minor Civil Division (MCD) level to derive a "crosswalk" from the old counties into the new planning regions: that is, MCDs were assigned the characteristics distribution of their old parent county, which was applied to their 2020 Census household or GQ populations, and then summed from the MCD level into the new planning regions.

For Core Based Statistical Areas (CBSAs) and Combined Statistical Areas (CSAs) in the state of Connecticut, the populations reflect the aggregation of minor civil divisions (MCDs) to the January 1, 2021 county boundaries.

³ U.S. Census Bureau, 2020. *Understanding and Using American Community Survey Data: What All Data Users Need to Know*, U.S. Government Publishing Office, Washington, DC. Accessible at <<https://www.census.gov/programs-surveys/acs/library/handbooks/general.html>>.