Version 2, May 2021

METHODOLOGY FOR THE UNITED STATES POPULATION ESTIMATES: VINTAGE 2020

Nation, States, Counties, and Puerto Rico – April 1, 2010 to July 1, 2020

Populations can change in three ways: people may be born (births), they may die (deaths), or they may move (domestic and international migration). The U.S. Census Bureau's Population Estimates Program measures this change and adds it to the last decennial census to produce updated population estimates every year.

OVERVIEW

Each year, the United States Census Bureau produces and publishes estimates of the population for the nation, states, counties, state/county equivalents, and Puerto Rico.[1] We estimate the resident population for each year since the most recent decennial census by using measures of population change. The resident population includes all people currently residing in the United States.

With each annual release of population estimates, the Population Estimates Program revises and updates the entire time series of estimates from April 1, 2010 to July 1 of the current year, which we refer to as the vintage year. We use the term "vintage" to denote an entire time series created with a consistent population starting point and methodology. The release of a new vintage of estimates supersedes any previous series and incorporates the most up-to-date input data and methodological improvements.

The population estimates are used for federal funding allocations, as controls for major surveys including the Current Population Survey and the American Community Survey, for community development, to aid business planning, and as denominators for statistical rates. Overall, our estimates time series from 2000 to 2010 was very accurate, even accounting for ten years of population change. The average absolute difference between the final total resident population estimates and 2010 Census counts was only about 3.1 percent across all counties.[2]

We produce estimates using a cohort-component method, which is derived from the demographic balancing equation:

Population Base + **Births** − **Deaths** + **Migration** = **Population Estimate**

The population estimate at any given time point starts with a population base (the last decennial census or the previous point in the time series), adds births, subtracts deaths, and adds net migration (both international and domestic).[3] The individual methods we use account for additional factors such as input data availability and the requirement that all estimates be consistent by geography and age, sex, race, and Hispanic origin.

This document describes the input data, methodology, and processes for the creation of population estimates for the nation, states, counties, state/county equivalents, and Puerto Rico. We begin with a short discussion on consistency in the estimates, describe the input data, and detail the processes by which we produce estimates.

[1] Population estimates for cities and towns and estimates of housing units are covered in a separate document.

[2] For more information on the accuracy of the population estimates, see http://www.census.gov/population/www/documentation/twps0100/E2%20County%20Totals_FINAL.pdf.

[3] Domestic migration sums to 0 at the national level and therefore has no effect on the estimates.

Estimates Consistency, Controlling, and the Residual

We produce the estimates using a "top-down" approach. Given that it is generally more reliable to estimate the change of a larger population, we begin by estimating the monthly population at the national level by age, sex, race, and Hispanic origin. We then produce estimates of the total annual populations of counties, which we sum to the state level. With the national characteristics, state total, and county total estimates created, we produce estimates of states and counties by age, race, sex, and Hispanic origin.

One of our key estimates principles is that all of the estimates we produce must be consistent across geography and demographic characteristics. For example, the sum of the county total populations must equal the total national population, and the sum of a particular race group within a state's counties must equal the total of that particular race group in the state. Since our various estimates products and processes use slightly different input data and methodology, they often do not generate this consistency automatically. Consequently, we adjust the final estimates to be consistent. As a result, the demographic components of change do not account for all of the year-to-year change in the estimates series. We call the difference between the result of the balancing equation and the final estimate the residual.

The national population estimates by characteristics do not contain a residual. This is because they are made first and are not required to sum to any pre-defined total. The balancing equations for the subnational processes initially produce what we call "uncontrolled" estimates. In order to ensure consistency, we use a process called controlling or raking. This involves calculating a rake factor as the control total (to which data must sum) divided by the sum of the numbers we wish to control (the initial estimated values).

$$Rake = \left(\frac{Control\,Total}{\sum(Uncontrolled\,Values)}\right)$$

We multiply this rake factor by the uncontrolled values to generate "controlled" estimates. In the simple case where the goal is to sum to a column total, this is fairly straightforward. However, deriving state and county population estimates by characteristics requires a slightly more complicated process. Since we produce national estimates by characteristics and state/county totals first, state and county characteristics need to use a two-way raking system. For example, state characteristics are required to be consistent with national characteristics and state total estimates (see the section on state and county characteristics).

The controlling process usually produces estimates that sum to a predefined total but are not integers. Because we require estimates in integer form, we round these data to remove the decimal values. Applying a simple rounding algorithm may upset the consistency established in the controlling process. To account for this, we use a variety of controlled rounding procedures (e.g., greatest mantissa or two-way controlled rounding).

Base Population

The enumerated resident population from the 2010 Census is the starting point for all post-2010 population estimates. We modify this enumerated population in three ways to produce the April 1, 2010 population estimates base.

First, we update the population estimates base to reflect changes to the 2010 Census population due to the Count Question Resolution (CQR) program.[4] This program allows legal entities to challenge their jurisdiction's

[4] For more information on CQR, see https://www.census.gov/programs-surveys/decennial-census/decade/2010/program-management/cqr.html.

decennial census value. Successful challenges are incorporated into the 2010 Census data used as the starting point for our population estimates.

Second, we incorporate legal boundary updates reported by January 1 of the vintage year, and other geographic program revisions.[5] While this generally does not impact national or state geography, it does occasionally change county, city, or town boundaries. Each vintage contains an entirely new time series that uses the most recent geographic boundaries for every estimates period.

Third, we modify the 2010 Census race categories to be consistent with the race categories that appear in our input data. The 2010 Census allowed for responses that included one or more of the race groups defined by the Office of Management and Budget (OMB) in 1997: White; Black or African American; American Indian and Alaska Native; Asian; and Native Hawaiian and Other Pacific Islander.[6] The 2010 Census also allowed responses that included “Some other race.” The Population Estimates Program produces estimates in five race categories and their combinations. "Some other race" responses have been “modified” to be consistent with these race categories.

This modification process re-categorizes responses that include “Some other race” alone or in combination with other races into one or more of the five OMB race categories. When a “Some other race” response appears in combination with one or more of the OMB races, we remove the “Some other race” category.[7] Responses that include “Some other race” alone are allocated to one or more of the five OMB categories listed above using information on the household when available or a hot decking procedure if necessary.

Note that this editing process produces tabulations for our estimates that show fewer people reporting two or more races than similar tabulations from the 2010 Census. This is primarily because responses of “Some other race” and one of the OMB defined races in the 2010 Census appear in the single OMB race category in the estimates base.

Group Quarters

We estimate the group quarters (GQ) population every year by single year of age, sex, race, Hispanic origin, and facility type.[8] The GQ method begins with an estimates base derived from the previous decennial census. We assume that the population in GQ remains constant throughout the decade unless we receive updated data on GQ population change.

Information on change to the base GQ population comes from our annual Group Quarters Report (GQR). The GQR consists of time series data from the branches of the military, the Department of Veterans Affairs, and our state partners in the Federal-State Cooperative for Population Estimates (FSCPE). Our data providers supply data at the facility level, which allows us to aggregate to all the other estimates geographies (e.g., counties and states). We use the submitted data to calculate a year-to-year change, which we then apply to the GQ population in the estimates base.

---

[5] For more information on census geography and boundary changes, see http://www.census.gov/programs-surveys/popest/guidance-geographies.html.

[6] The OMB standards are detailed in the OMB Notice, “Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity”, Vol. 62, No. 210, Thursday, October 30, 1997, available at https://obamawhitehouse.archives.gov/omb/fedreg_1997standards.

[7] To learn more about the “Modified Race” process, go to http://www.census.gov/programs-surveys/popest/technical-documentation/research/modified-race-data.html.

[8] 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.”

Once we have a times series of total GQ population at the facility level, we aggregate the facility-level data to the national level and apply the 2010 Census distribution of age, sex, race, and Hispanic origin detail by major facility type to generate estimates of the GQ population by demographic characteristics. We also apply the county distribution of age, sex, race, and Hispanic origin to the county level totals. To ensure consistency, we control the county characteristics to the national characteristics and the subcounty totals to the new county totals. Finally, we aggregate the data to the necessary levels for estimates production (e.g., three age groups for county totals production and full demographic detail for state characteristics production).

Vital Statistics

Vital statistics encompass two of the core components of the demographic equation: births and deaths. We receive data on vital statistics from the National Center for Health Statistics (NCHS) and the Federal-State Cooperative for Population Estimates (FSCPE). NCHS data are derived from birth and death certificates across the United States. Births data include date of birth, sex of child, residence and age of mother, and race and Hispanic origin of both mother and father. Deaths data include residence, age, sex, race, and Hispanic origin of each decedent, and the date each death occurred. The FSCPE contributes data on the geographic distribution of recent vital events within their respective states. Vital events data in the population estimates also include the results of our own short-term projections.

In general, the births and deaths data we receive from NCHS have a two-year lag. This means that the most recent data we have on births and deaths by geographic and demographic detail for each vintage of estimates refer to the calendar year two years prior to the vintage year. For example, the most current full-detail births and deaths data we used in Vintage 2020 were from calendar year 2018. We also receive preliminary or provisional NCHS total numbers of births and deaths at the national level for the year prior to the vintage (in this example, 2019). Using these data and the data received from the FSCPE, we create short-term projections that approximate the final NCHS data by characteristics, the preliminary NCHS national totals, and the FSCPE data (where available) by geographic distribution.

We also modify the NCHS births and deaths data to comply with our process. The births data require three changes. Since 2016, all 50 states and the District of Columbia have reported parents' race data to NCHS in the 1997 OMB race categories (non-Hispanic single-race White, non-Hispanic single-race Black or African American, non-Hispanic single-race American Indian and Alaska Native, non-Hispanic single-race Asian, non-Hispanic single-race Native Hawaiian and Other Pacific Islander, and Hispanic). NCHS also provides race data in the 1977 OMB race categories (White; Black; American Indian, Eskimo or Aleut; and Asian or Pacific Islander) where parents' race data are only classified into one race group. For our purposes, we first convert the race data from the 1977 standards into the newer 1997 classification utilizing a race bridging method designed by NCHS and the United States Census Bureau to make the multiple-race and single-race data comparable.[9]

Second, as birth certificates include only data on the race and Hispanic origin of the parents, not the child, we impute the race of the child through our "Kidlink" process.[10] This approach uses the combined distributions of mothers', fathers', and children's race and Hispanic origin from the 2010 Census to impute children's race and Hispanic origin.

Third, we adjust for inconsistencies between the imputed race and Hispanic origin distributions of births compared to the base population under age 1 in the 2010 Census. This benchmarking process allows us to adjust the overall race and Hispanic origin distribution of births to create a "census-consistent" time series of births.

[9] For more information on the NCHS race-bridging factors, see http://www.cdc.gov/nchs/nvss/bridged_race.htm.
[10] For more information on the Kidlink process, see https://nces.ed.gov/FCSM/pdf/Guarneri_2012FCSM_X-B.pdf.

We also make modifications to the NCHS deaths data. Although we often have direct information on the race and Hispanic origin of the decedent, deaths are still coded in many states according to the 1977 OMB race categories. We use the same race bridging process for deaths that we use to convert births into the 1997 race and Hispanic origin categories used in estimates production.

While we make no additional adjustments to deaths occurring to people under 70 years of age, we do modify death records for persons age 70 or over. Reporting of age at older ages is generally less reliable than at younger ages.[11] To address this issue, we redistribute all deaths occurring to the aggregate population 70 years and older by sex, race, and Hispanic origin to single year of age (70 to 99 and 100+ years) using life-table based death rates.[12]

We aggregate NCHS-based birth and death data for the production of national-level population estimates. When data by full geographic and characteristic detail are available, we use births directly as a base for the population under age 1. We apply death rates by characteristics and control to NCHS death data by aggregate characteristic groups (see the section on national estimates). For periods when full detail data are not available, we use available or estimated data on vital events to calculate characteristic-specific rates which are then controlled to preliminary or provisional total data from NCHS.

Distributing the projected national-level births and deaths to the subnational level requires additional computations. To do this, we use a combination of short-term projections of county-level population characteristic detail and FSCPE data on the geographic distribution of total county vital events (where available). The projections are derived by calculating county-level age-specific fertility and mortality rates. We then apply these rates to the county population projections from the prior vintage. The resulting projected data by demographic detail are then reconciled with FSCPE data on the geographic distribution of total county vital events. These values are then summed to the state level and controlled to national projections of characteristics described above. The final county data are then controlled to the resulting state values.

Net Domestic Migration

The third major component of the balancing equation is migration. Migration can be divided into net domestic migration (NDM) within the United States and net international migration (NIM) between the United States and elsewhere. The Population Estimates Program calculates domestic migration using several data sources and methods depending on the age group in question and the level of characteristic detail required.

For state and county total estimates, we calculate county-to-county net domestic migration based on four data sources:

1. Internal Revenue Service (IRS) tax return data for ages 0-64
2. Medicare enrollment data from Centers of Medicare and Medicaid Services (CMS) for ages 65+,
3. Social Security Administration's Numerical Identification File (NUMIDENT) for all ages
4. Change in the group quarters population (described in the "Group Quarters" section)

---

[11] For more information on age reporting at older ages, see http://www.cdc.gov/nchs/data/nvsr/nvsr62/nvsr62_07.pdf.

[12] To derive the death rates for the age-70-and-older population, we employ life tables based on annual 2000-2010 NCHS mortality files and 2000-2010 Intercensal Population Estimates prepared by the United States Census Bureau. The life tables are for males and females in five groups: Hispanic, non-Hispanic White, non-Hispanic Black, non-Hispanic American Indian and Alaska Native, and non-Hispanic Asian and Pacific Islander.

State and County Totals by Three Age Groups

We produce overall net rates of movement between counties for the total population estimates by three age groups: under 18, 18 to 64, and 65 and over. For the household population under age 18 and 18 to 64, we use person-level data on filers, spouses, and dependents from IRS tax return data. We match two years of IRS tax returns with age data from the NUMIDENT file to produce geographic data by age categories. The NUMIDENT is a database of all Social Security Numbers ever assigned, which is updated annually with new entries and any changes to a person's record.

Once tax returns are matched, we then compare the addresses between the two years of IRS data to identify the number of exemptions that moved from one county to another between tax filings. An IRS exemption is defined here as an individual who appears in the IRS tax return data, either as primary filer, spouse, or dependent.

Not all residents are represented in the tax exemption data, since not everyone files taxes. Therefore, the number of migrants in the IRS data is not equivalent to the number of migrants in the resident population. To overcome this coverage limitation, we calculate Net Domestic Migration (NDM) rates instead of using observed flows in the tax data. County specific net migration rates can be thought of as the ratio of net migrant exemptions to the number of exemptions present at the beginning of the migration period. Mathematically, the rate is first obtained by subtracting the number of out-migrants from the number of in-migrants for each county to produce the number net migrant exemptions. We then divide the net migrant exemptions by the sum of non- migrant and out-migrant exemptions for each county. We calculate these rates separately for each period by the two age groups (under 18 and 18 to 64), as follows:

$$NDM\ Rate_{0-17} = \frac{In\ migrants_{0-17} - Out\ migrants_{0-17}}{Non\ migrants_{0-17} + Out\ migrants_{0-17}}$$

$$NDM\ Rate_{18-64} = \frac{In\ migrants_{18-64} - Out\ migrants_{18-64}}{Non\ migrants_{18-64} + Out\ migrants_{18-64}}$$

Because the population aged 65 and over is more likely to enroll in Medicare than file taxes, we rely on Medicare enrollment data from CMS to account for movement of the older population. The process is similar to the under 18 and 18 to 64 age groups. Instead of tax exemptions, we match two years of Medicare enrollment data (address as of July 1) with age data from the NUMIDENT file. We then compare the addresses between the two years of Medicare data to identify the number of enrollees that moved from one county to another between the two years.

Similar to IRS filing, not everyone enrolls in Medicare. Therefore, the number of migrants in the Medicare data is not equivalent to the number of migrants in the resident population. For the same reason, we produce net rates based on Medicare enrollees for the 65 and older population. We calculate the net domestic migration (NDM) rate for the 65 and over population by subtracting the number of out-migrant enrollees from the in-migrant enrollees for each county to produce the number of Medicare-based net migrant enrollees. We then divide the number of Medicare-based net migrant enrollees by the sum of non-migrant enrollees and out-migrant enrollees for each county and period. The net rate is a ratio of the number of enrollees who moved in less those who moved out to the number of enrollees present at the beginning of the period, as given below:

$$NDM\ Rate_{65+} = \frac{In\ migrants_{65+} - Out\ migrants_{65+}}{Non\ migrants_{65+} + Out\ migrants_{65+}}$$

During the production of state and county total estimates, we apply these rates to the household population within the three age groups to produce a computed number of migrants for use in the balancing equation. We also treat change in GQ as an indirect measure of domestic migration. This methodology implicitly accounts for migration between GQ facilities as well as for household to GQ movement. To produce estimates of total migration for each of the three age groups, we combine age-specific domestic migration estimates from the application of these rates with the total amount of GQ population change in each age group. These total net domestic migration values are then controlled to sum to zero at the national level (as domestic migration must).

State and County Characteristics

The production of state and county characteristics estimates occurs after the production of state and county total estimates. The process for state and county total estimates only requires information on migration by age groups. However, to produce migration data by full characteristic detail, we need age, sex, race, and Hispanic origin.

To create net domestic migration estimates by full demographic detail, we use data from four sources:

1. Internal Revenue Service (IRS) tax return data for ages 0-64
2. Medicare enrollment data from Centers of Medicare and Medicaid Services (CMS) for ages 65+
3. Social Security Administration's (SSA) Numerical Identification File (NUMIDENT) for all ages
4. Demographic Characteristics File (DCF) for all ages

We use mailing address information from IRS tax return data for ages 0-64 to estimate migration. For ages 65 and older, we utilize address information from Medicare enrollment data to assign migration status. We use the NUMIDENT File to allocate age and sex to individuals in the migration universe.

The Population Estimates Program uses a Demographic Characteristics File (DCF) to allocate race and Hispanic origin to individuals in the migration universe with missing data. The DCF provides information on race and Hispanic origin. It is a dataset developed internally from a collection of person-level data derived from decennial census data, administrative records, and a set of imputation techniques when reported race and Hispanic origin are not available.

Because of known under coverage in the IRS and Medicare data (not everyone files taxes or claims benefits), we again calculate characteristic-specific out-rates and in-proportions and apply them to the population "at risk" of migrating. The population "at risk" is simply the population in each county in that particular age, sex, race, and Hispanic origin group.

We calculate domestic out migration rates by dividing the number of out-movers identified in the particular source data (IRS or Medicare, depending on age) by the total number of individuals at the beginning of the period. The total number of individuals at the beginning of the period is the sum of out movers and non-movers, as shown below:

$$Out\ Rate_{characteristic} = \frac{Out\ migrants}{Non\ migrants + Out\ migrants}$$

To distribute the pool of out movers by demographic characteristics to their destination counties, we use in-proportions. In-proportions are defined as the number of exemptions (for ages 0-64) or the number of enrollees (for ages 65+) moving into a county divided by the national total number of out mover exemptions/enrollees in a given demographic group. Though these can be very small proportions, this methodology creates in and out domestic migration rates that are consistent. It is important to note that no rounding is applied to these migration rates.

$$In\ Proportion_{characteristic} = \frac{In\ migrants}{\sum_{all\ counties} Out\ migrants}$$

In the production of state and county population estimates by characteristics, we apply the calculated out rates annually to each county's population "at risk" to produce estimated numbers of domestic out-migrants. Next, the national "pool" of out-migrants by demographic characteristics are then allocated to their destination counties with the in-proportions.

Net International Migration

The third major component of the balancing equation is migration. As noted, migration can be divided into NDM within the United States and NIM between the United States and abroad. We estimate international migration in several parts: immigration of the foreign born, emigration of the foreign born, net migration between the United States and Puerto Rico, net migration of natives to and from the United States, and net movement of the Armed Forces population to and from the United States. For each component, we first estimate the total migration flow for the nation.

For the sub-components of international movement other than the movement of the Armed Forces population, we use a proxy universe to distribute national total estimates by geography and demographic characteristic detail. A proxy universe is a geographic and characteristics distribution derived from a different population than the total estimate. We use a proxy universe because it allows us to utilize a larger sample (in the case of foreign- born immigration) and to produce characteristics for a population without direct observation (for native and foreign-born emigrants). We create total estimates for the United States either through direct or residual estimation, and then apply the distribution from the proxy universe in order to produce estimates of international migration for states and counties by demographic characteristics.

Again, excluding movement of the Armed Forces, state demographic characteristic distributions are based on three years of pooled American Community Survey (ACS) 1-year files, while county distributions are produced from ACS 5-year files. We control county-level data to state-level data to ensure the component data are consistent. For the net movement of the Armed Forces population, demographic characteristics and state distributions are developed based on a combination of data collected by the Defense Manpower Data Center (DMDC) and the most recent 5-year ACS file for each year of estimates.

We use the methodology described in the next sections to estimate the components of net international migration for April 2010 to June 2019 and develop the preliminary estimate for July 2019 to June 2020. To reflect the impact of the COVID-19 pandemic, we adjust the preliminary estimate to reflect lower migration levels for the last quarter of the 2020 estimates period. We use the 2019 ACS and 2020 monthly time series on international mobility from auxiliary data sources to estimate total net international migration. Auxiliary data sources include Airline Passenger Traffic (APT) data from the Bureau of Transportation Statistics (BTS), visas issued overseas from the Bureau of Consular Affairs, and internal-use entry data from Statistics Canada. We internally estimate alternative migration levels for March or April through June 2020 based on different assumptions from the auxiliary data. For all alternative estimates, we assume migration that occurred before the pandemic follow similar migration levels from the 2019 ACS input data. The final national net international migration total is an average of the alternative estimates. We use the proxy universe method to distribute characteristics for the nation, states, and counties. Because distributions from the ACS-based proxy universes lag by a few years, the geographic distribution and demographic composition of net international migration for 2020 will not reveal any localized COVID-19 effects.

Foreign-Born Immigration

We use the ACS residence one year ago (ROYA) question to estimate foreign-born immigration for the nation and states. We estimate foreign-born immigration separately for Mexico and All Other Countries since we expect these groups to exhibit different demographic and geographic patterns.

We use the 1-year ACS to estimate national- and state-level totals. These totals represent international movement occurring between the previous year and survey year. For example, we estimate movement between July 2012 and June 2013 from the 2013 ACS. Movement between April 2010 and June 2010 is one quarter of the 2010 ACS estimate. Movement for the final year of the time series is equal to the previous year's estimate because the latest ACS lags behind the vintage year. For example, in Vintage 2019, we used the 2018 ACS to estimate movement from July 2017 to June 2018 and held the estimate constant for July 2018 to June 2019. We revise the estimate the following vintage when more current ACS data become available.

We use a proxy universe to distribute national and state characteristics as well as county totals and characteristics. The proxy universe for foreign-born immigration is the foreign-born population who entered the United States within five years of the survey. We adjust age to reflect age at year of entry. There are separate proxy universes for the Mexican born and for those born in another foreign country. We apply proxy universe distributions from pooled ACS files to the state-level totals to derive state characteristics. We aggregate state characteristics to derive national characteristics. Next, we apply proxy universe distributions from the 5-year ACS to state characteristics to derive county totals and characteristics.

Foreign-Born Emigration

We use a residual method to estimate emigration of the foreign-born population at the national level. The residual method uses information on mortality and recent immigration to account for cohort change in the foreign-born population within a specific period. Mortality estimates come from NCHS Hispanic life tables by age and sex. Immigration estimates come from the ACS year of entry question. We develop an annual time series from consecutive 1-year ACS files to measure foreign-born population change. We attribute to emigration (residual) any part of foreign-born population change not explained by mortality or immigration. Next, we divide the residual by person years to create annualized emigration rates. Finally, we apply the rates to the population "at risk" of emigrating by sex, year of entry, and place of birth cohorts to calculate annual foreign-born emigration totals for the nation. We calculate emigration rates for seven mutually exclusive groups: 1) Mexican-born males who entered the United States within the past 10 years, 2) Mexican-born females who entered within the past 10 years, 3) Mexican born who entered more than 10 years ago, 4) Canadian and European born who entered within the past 10 years, 5) Asian born who entered within the past 5 years, 6) All other foreign born who entered within the past 10 years, and 7) Asian born who entered more than 5 years ago and non-Mexican born who entered more than 10 years ago. We calculate separate rates under the assumption that each group exhibits different propensities to emigrate, as well as different demographic compositions and geographic distributions.

Using the first group as an example, we tabulate the Mexican-born male population who entered the United States within the past 10 years. Next, we survive this population forward to obtain the expected population for a later year. Subtracting the observed population from the expected population for the later year yields a residual, which is assumed to represent total emigration occurring over the period. Next, we convert this residual into an annualized emigration rate. We calculate six rates based on three 2-year residuals, two 3-year residuals, and one 4-year residual. In order to reduce the effects of survey variability, we average the six rates. We apply the averaged rate to the population at risk of emigrating (tabulated from the 1-year ACS) to obtain annual estimates of emigration. We utilize one quarter of the 2010 estimate to calculate foreign-born emigration from April 2010 through June 2010.

We follow the same method for estimating emigration for the other six groups listed above. For groups (3) and (7), which represent non-recent arrivals, we average rates from multiple ACS files as an additional step to stabilize annual rates. These two groups have large "at risk" populations, and slight variability in emigration rates can cause improbably large fluctuations in the annual estimate of foreign-born emigration.

Migration between the United States and Puerto Rico

We use the ROYA question from the ACS and the Puerto Rico Community Survey (PRCS) to estimate annual migration flows between the United States and Puerto Rico. We classify ACS respondents who resided in Puerto Rico one year ago as in-migrants. We classify PRCS respondents who resided in the United States one year ago as out-migrants. We subtract out-migrants from in-migrants to calculate net migration. The proxy universe for net migration between the United States and Puerto Rico is the population born in Puerto Rico who entered the United States up to 10 years prior to the survey year. To account for the impact of Hurricane Maria on migration flows between Puerto Rico and the United States for the July 2017 to June 2018 and July 2018 to June 2019 estimates periods, we combine APT data from BTS with ROYA data from the 2017 and 2018 1-year ACS/PRCS respectively. To capture the impact of the COVID-19 pandemic, we combine APT data from BTS with the 2019 1-year ACS/PRCS for the July 2019 to June 2020 estimates period. We continue to use ROYA from the ACS and PRCS to estimate migration for all other years.[13]

Native-Born Migration

Estimates of net migration of the native-born population are produced using a method which utilizes data from approximately 80 countries. This work compares estimates of the United States-born or United States citizen population living overseas measured by population registers and censuses in other countries at two consecutive time periods. The residual is used to develop an average annual estimate of net native migration. The proxy universe for the net native migration component is the native-born civilian population whose residence one year ago was either in a different state or abroad (as this approximates the characteristics of people who migrate).

Movement of the Armed Forces Population to and from Overseas

We derive the estimate of the net overseas movement of the Armed Forces population from data collected by the Defense Manpower Data Center (DMDC). DMDC provides monthly tabulations of military personnel stationed or deployed outside the United States by age, sex, race, Hispanic origin, and individual branches of service within the Department of Defense. We assume that changes in the overseas military population, excluding deaths, indicate movement of personnel into and out of the United States. To derive estimates of net international movement of the armed forces at the county-level, we primarily use DMDC data by age, sex, race, Hispanic origin, and county. To improve the geographic distribution of military movement around certain domestic military installations, we use county grouping information derived from the most recent ACS five-year file.

National Population by Age, Sex, Race, and Hispanic Origin

The goal of the national population estimates process is to produce monthly resident population estimates by single year of age (0 to 100+), sex, Hispanic origin, and race (31 categories). We then divide these estimates into the following universes: household (HH), civilian (CIV), civilian noninstitutionalized (CNI), and resident plus armed forces overseas (RES+AFO). The core of the process is the demographic balancing equation. We take inputs on births, deaths, and net international migration by characteristics, and apply these components to the population at the beginning of the period.

[13] For more detail on this component, refer to the section "Puerto Rico Resident Population by Age and Sex".

The annual number of net international migrants is divided into monthly and quarterly values to use in the production of the estimates. The final year of available data (usually the year prior to the vintage year) is held constant to the end of the time series. Utilizing vital statistics (birth and death) information, however, is more complicated. Because we have limited final data by characteristics from NCHS, we use a combination of final and preliminary input data, populations at risk, rates, and controls. We use slightly different methods for three estimates periods based on data availability from NCHS: full-detail (up to two years prior to the vintage year), preliminary totals (the year prior to the vintage year), and no data (the vintage year to the end of the time series). Although the process across each of the three periods is essentially the same (apply rates to a population then control the result), the source of the rates and controls changes based on the level of detail available in the input data.

In the first period, from the base to two years prior to the vintage year, we have vital statistics by full characteristics. For births, we directly utilize the number of monthly events from NCHS data. For deaths, we multiply the starting population by life table death rates used in our Population Projections Program, and then control the result to NCHS deaths by sex and age (single-year-of-age under 70, and an aggregate of age 70 and over) for the following race and Hispanic origin groups: non-Hispanic White alone, non-Hispanic Black alone, non-Hispanic American Indian or Alaska Native alone, non-Hispanic Asian and Native Hawaiian or Other Pacific Islander alone, and Hispanic of any race.

In the second period, the year prior to the vintage year, we only have preliminary or provisional totals for births and deaths. Here we use the most recent year of final NCHS data (usually two years prior to the vintage year) to calculate characteristic-specific birth and death rates, apply those rates to the population, and control the result to the overall preliminary totals from NCHS.

In the third period, where NCHS data are not available (the vintage year), we use implied birth and death rates calculated from the most recent year for which preliminary data are available. We hold the preliminary totals constant, apply the new calculated rates to each period, and control to the annual total for every remaining year in the time series. From this point on, both the rates and the totals are held constant. Population size in each group provides characteristic-specific variation in the distribution of births and deaths.

There are three main steps in the production of monthly national population estimates (which include the vital statistics process above): estimate the quarterly national resident population; estimate the monthly population; and estimate the monthly population by the other four universes described earlier. The goal is to produce monthly resident population estimates by single year of age (0 to 100+), sex, Hispanic origin, and race (31 categories), then calculate the required population universes (e.g., household, civilian, etc.).

We create population estimates by quarter-years of age by applying final births, deaths, and international migration to the base, then aging the population forward one quarter-year of age. The process is repeated for every quarter in the time series. We round the final resident populations and components and assume any residual is part of international migration.

Once we have created final quarterly estimates of the population by characteristics, we estimate the population for the second and third month of each quarter. To do this, we assign the calculated monthly births and deaths for each quarter to specific months based on the monthly distribution of vital events in the most recent year of final NCHS data. Together with the international migration component, we use these vital statistics to estimate monthly values for population estimates by age, sex, race, and Hispanic origin.

The final step in the national estimates process is to calculate the additional population universes by demographic characteristics. To calculate the resident plus Armed Forces overseas population, we add the monthly overseas military population (from DMDC data) to the estimated resident population. The civilian population is the result of subtracting the monthly resident military population (also from DMDC) from the resident population. The civilian noninstitutionalized population is produced by subtracting the institutionalized group quarters population from the civilian population.[14] Finally, we estimate the household population by subtracting the total group quarters population from the resident population. In addition, we use linear interpolation to derive daily resident population estimates and monthly component settings (e.g., number of seconds per birth) for the Population Clock.[15]

State and County Total Resident Population

The goal of the state and county total population estimates process is to produce total population estimates and estimates of the state population aged 18 and over for all states, counties, and equivalents in the United States. We treat parishes in Louisiana, boroughs in Alaska, and several independent cities (in Maryland, Missouri, Nevada, and Virginia) as counties. The process focuses on the development of estimates for counties (and equivalents) only. State estimates exist only as a sum of the final estimates for counties.

Our process involves estimating the population separately for ages under 18, 18 to 64, and 65 and over. We estimate three age groups for this process for two reasons. First, we use different input data for domestic migration based on whether we are estimating a population under age 65 (IRS tax exemptions) or 65 and over (Medicare enrollment). Second, we produce estimates of the state population aged 18 and over to provide to the Federal Election Commission.

Producing state and county total population estimates is similar to the production of national estimates, as they are both based on the balancing equation. However, state and county estimates are produced for annual July 1 dates, and they incorporate domestic migration. Even though there are slight differences in the way we calculate the first three months (April to July) from the estimates base (using only one quarter of a year of migrants, for example), the process is very similar for all other points in the time series.

We first subtract the GQ population and "age" the population one year in order to produce an estimate of the household population at the start of each period. The aging process takes the proportion of the previous vintage county population age 17 and 64, applies that proportion to the current year, and moves that population into the next higher age group (e.g., the estimated number of 64-year-olds would "age" into the group aged 65 and over).

Net migration rates calculated from IRS and Medicare data are then applied to the aged household population at the start of the period to create estimates of net domestic migration. We then add net domestic migrants, add births (for the under 18 population), subtract deaths, and add international migrants to produce an uncontrolled estimate of the household population at the end of the period for each age group. The GQ population is then added to create uncontrolled resident population estimates for each age group.

The next step in the process ensures consistency with the national estimates. First, we control the calculated resident population numbers to equal the national numbers by the three age groups. Second, we add GQ change to the total household domestic net migration estimate for each age group and control that number to sum to zero at the national level by age group. We then round the final resident population by age group and allocate the remainder (usually very small) to the largest population value in the country. Finally, we

[14] The institutionalized population is defined as people under formally authorized, supervised care or custody in institutions including correctional institutions, juvenile institutions, nursing homes, skilled nursing facilities, psychiatric hospitals, and facilities for the disabled.
[15] The Population Clock is published on the Census Bureau website and is located at http://www.census.gov/popclock.

aggregate the three age groups into total estimates for counties and sum these estimates to create final estimates for states.

State and County Resident Population by Age, Sex, Race, and Hispanic Origin

The goal of the state and county resident population estimates by demographic characteristics process is to create population estimates by age, sex, race, and Hispanic origin for all states, counties, and equivalents in the United States. This process essentially follows the cohort-component approach, adding births, subtracting deaths, adding the effects of net domestic and international migration, and aging the population forward. An additional factor in this process is the requirement of consistency between population estimates for the multiple levels of geography and characteristic detail (see the section on estimates consistency). County characteristics, for example, are produced by single year of age (under age 85 and an aggregate of age 85 and over), sex, Hispanic origin, and race (31 categories). Accounting for all the cross-classifications, there could potentially be 10,664 possible combinations per geographic area.

The calculation of state and county estimates by characteristics uses a two-way raking process to ensure that the final estimates sum correctly by both geography and characteristics. The method involves iteratively controlling estimated values to the larger geography's characteristics and the smaller geography's total estimates. In other words, we control state characteristics to national characteristics and state totals then control county characteristics to state characteristics and county totals. After repeated rakings, changes in the data become progressively smaller, eventually allowing us to round the result.

The raking process produces population estimates that are not necessarily integers. We then apply a controlled rounding process which allows us to convert the estimates to whole numbers without changing the total values. For state estimates, we control to both the state totals and the national characteristics. For county estimates, we control to the county totals and the final state characteristics. Because the state characteristic estimates have already been controlled and rounded, creating consistency between county characteristics and state characteristics automatically makes counties consistent with the national values as well.

Puerto Rico Resident Population by Age and Sex

The U.S. Census Bureau produces annual estimates of the resident population for the Commonwealth of Puerto Rico and its municipios. The estimates are produced by age and sex using a cohort-component approach as described previously for the United States, states, and counties.

The cohort-component population estimation method starts with the 2010 Census base population by age and sex. It then follows each birth cohort as it ages and experiences mortality and migration. This procedure is repeated for each year of the estimation period, by age and sex.

The 2010 Census population for the Commonwealth and its municipios serves as the starting point for the estimates by sex and single year of age (0 to 99 and 100+). The population estimates base is updated to reflect changes to the 2010 Census population due to the CQR program, legal boundary updates reported by January 1, 2020, and other geographic program revisions.

Final births and deaths data for each month from January 2010 to December 2019, and provisional births and deaths data for each month from January 2020 to July 2020 were obtained from the Puerto Rico Institute of Statistics and originate from the Puerto Rico Department of Health vital registration system. The births data include month and year of birth, sex of the child, mother's age, and mother's country and municipio of residence. The deaths data include month and year of death, sex, age, and decedent's country and municipio of residence.

The migration estimates used in producing the annual Puerto Rico Commonwealth population estimates consist of two components: in-migrants from the United States to Puerto Rico and out-migrants from Puerto Rico to the United States. The United States/Puerto Rico migration was estimated using a respondent's current residence and the ROYA question in the ACS and PRCS. Respondents to the PRCS living in Puerto Rico and indicating residence in the United States during the previous year are categorized as in-migrants. Respondents to the ACS living in the United States and indicating Puerto Rico residence during the previous year are labeled out-migrants. Because children under 1 year of age would not have a residence one year ago, the value is assumed to be one-half of the estimate for the 1-year-olds. To estimate the United States/Puerto Rico in-migrants and out-migrants by sex for Puerto Rico, we use the 1-year ACS/PRCS files for 2015 to 2019 and the 5-year (2015-2019) ACS/PRCS files for age distributions within sex.

On September 20, 2017, Hurricane Maria made landfall on Puerto Rico. Due to this hurricane, we needed to modify the method used in prior vintages to generate the Puerto Rico estimates. Specifically, the hurricane halted PRCS data collection during the months of October, November, and December of 2017, while ACS data collection continued, and data regarding movement from Puerto Rico to the United States were collected for the same months. Because PRCS data collection was suspended following Hurricane Maria, and remained suspended for the last three months of the year, estimating Puerto Rico in-migration using the 2017 PRCS does not accurately reflect the impact of the hurricane on the island. Similarly, the annual 2017 ACS estimate did not reflect a substantial increase in out-migration from Puerto Rico to the United States, perhaps due to the late occurrence of the hurricane during the year and the ACS two-month residency rule for inclusion in the survey. This resulted in an inadequate measure of net migration between the United States and Puerto Rico for Vintage 2018. To address this issue, we modified the ACS/PRCS method by using the 2017 to 2018 monthly APT data from BTS for the July 2017 to June 2018 estimates period. For the July 2018 to June 2019 estimates period, we combined the 2018 to 2019 monthly APT data with ROYA data from the 2018 1-year ACS and PRCS.

The method used to adjust the Puerto Rico estimates for the effects of Hurricane Maria is as methodologically consistent as possible with previous ACS-based estimates. The revised method uses approximately the same time period (calendar year) and migration universe (United States to Puerto Rico). APT data shows consistently higher net out migration than the ACS/PRCS estimates for the period of interest. To account for this difference between data sources, the method "blends" the ACS/PRCS and APT data. This is accomplished by calculating the ratio of ACS-to-APT net migration for Puerto Rico for 2015 and 2016 and applying it to the net Puerto Rico-United States migration measured during the 2018 ACS/PRCS. The result of this method change is a July 1, 2018 estimate that takes into account the impact of Hurricane Maria and a July 1, 2019 estimate that takes into account return migration.

In Vintage 2020, modifications were needed to account for the impact of the COVID-19 pandemic on Puerto Rico's net migration estimates. The COVID-19 adjustment was applied directly to the ACS/PRCS estimates. Given that most movement to and from Puerto Rico is done via air travel, we assume that APT data are a reasonable proxy for measuring changes in net migration trends. Using APT data, the net passenger total was calculated for the months of March to June 2020. Next, a ratio was calculated using the historical average net passenger movement for March to June 2010-2019, excluding 2018 (Hurricane Maria), and the March to June 2020 net passenger total. We then reduced the ACS/PRCS seasonal total by multiplying one-third of the 2019 ACS/PRCS annual total by the net passenger movement ratio. For the final estimate, two-thirds of the 2019 ACS/PRCS estimates were applied to the ACS/PRCS reduced seasonal total.

Typically, March through May are net passenger outflow months, while June is a passenger inflow month. The March APT data showed a large increase in net passengers to the mainland United States; April and May showed small positive net gains to Puerto Rico; and June showed a large positive net gain to Puerto Rico. The overall impact of the COVID-19 adjustment was reduced net outmigration from Puerto Rico.

Net migration for the Puerto Rico municipios was estimated using a residual method. 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. Differences between the expected population and the population enumerated in the 2010 Census are assumed to represent net migration over the decade. The residual, which represents events over a period of 10 years, was converted into an annual average migration rate by age and sex for each municipio. These rates are then controlled to the national rate. Thus, whereas the municipios' net migration estimates include changes based on Hurricane Maria and the impacts of COVID-19, the resulting net migration distribution is similar to distributions in prior years.[16]

[16] This document reflects a correction to the description of the methodological adjustment applied to the Puerto Rico estimates to account for the COVID-19 pandemic.