

Methodology for the State and County Resident Population Estimates by Age, Sex, Race, and Hispanic Origin (Vintage 2011): April 1, 2010 to July 1, 2011

The U.S. Census Bureau annually produces estimates of the resident population by age, sex, race, and Hispanic origin for each state and county in the United States and the District of Columbia.¹ The following documentation describes the process by which we produce population estimates for these demographic characteristics at the state and county level.

Overview

Resident population includes all residents (both civilian and Armed Forces) living in the United States and is based on the concept of residence used in Census 2010, which defines a resident of a specified area as a person “usually resident” in that area. The Census Bureau develops estimates by updating Census 2010. We begin with modified population counts by age, sex, race, and Hispanic origin from Census 2010 (see Specification of the Base Population, below) and estimate the change that has occurred since that time. This change is measured annually to produce estimates of the population for July 1, 2010 and July 1, 2011. The Vintage 2011 estimates contain the most current data available and supersede all previous estimates. These components are further described in the sections below.

Estimating Population Change

Population can change as a result of births, deaths, or migration, which are known collectively as the components of change. In the United States, births and deaths are recorded with relative accuracy and completeness, and these data are readily available. Migration, on the other hand, can be very difficult to estimate accurately and is the largest source of population change for many areas. For these estimates, migration is divided into two independently estimated sub-components: domestic and international.

We produce separate estimates of the population living in special housing arrangements known as group quarters (for example, college dormitories) because movement into and out of these facilities is unlikely to be captured by our migration estimates, and because we receive data to estimate this population separately. Consequently, our estimation procedure begins by splitting the Census population into two mutually exclusive universes: the group quarters (GQ) population, and the non-GQ or household population. We estimate change in the household population by estimating the components of change mentioned above. Change in the GQ population is estimated using data received annually from members of the Federal-State Cooperative for Population Estimates. The resulting household and GQ estimates are added together to produce the new set of resident population estimates.

Specification of the Base Population

The enumerated population from Census 2010 provides the starting point for these estimates. The census population is modified in two ways to produce the estimates base.

1. Responses of "Some Other Race" from the 2010 Census are modified. This results in differences between the population for specific race categories produced for the 2010 Census population estimates base versus those in the original 2010 Census data.²
2. The April 1, 2010 population estimates base reflects changes to the 2010 Census population from the Boundary and Annexation Survey (BAS) and other geographic program revisions. It does not reflect changes from the Count Question Resolution program. All geographic boundaries for the 2011 population estimates series are defined as of January 1, 2011.

These estimates use the race categories mandated by the Office of Management and Budget's (OMB) 1997 standards: White; Black or African American; American Indian and Alaska Native; Asian; Native Hawaiian and Other Pacific Islander.³ These race categories differ from those used in Census 2010 in one important respect. Census 2010 also allowed respondents to select the category referred to as Some Other Race. When Census 2010 data were edited to produce the estimates base, respondents who selected the Some Other Race category alone were assigned to one of the OMB mandated categories.⁴ For those respondents who selected the Some Other Race category and one or more of the other race categories, the edits ignored the Some Other Race selection. This editing process produced tabulations from our estimates that show fewer people reporting two or more races than similar tabulations from Census 2010, because respondents who selected Some Other Race and one of the OMB mandated races in Census 2010 appear in the single OMB race category in the estimates base.

In the tables created from these estimates, we group race categories in two different ways. One group includes the five single-race categories and a sixth category that combines all categories with more than one race – referred to in our tables as Two or More Races. The other group includes the five alone-or-in-combination race groups. Each of the alone-or-in-combination groups contains one of the single-race categories plus all the multiple-race categories that include that single race. Alone-or-in-combination groups do not sum to the population total, because each multiple-race person is included in more than one of these groups. For example, people who are White *and* Asian would be included in both the White alone-or-in-combination group and the Asian alone-or-in-combination group. We also apply these modifications to the Census 2010 GQ population to produce the GQ estimates base. The GQ estimates base is subtracted from the total estimates base to produce the household estimates base population.

Estimation of the Household Population

The household population is estimated using a technique known as the cohort-component method. In this context, the term *cohort* refers to a group of individuals born in the same time period. The cohort-component method applies the components of population change to groups of individuals based on when they were born.

The following equation illustrates how our application of this technique treats annual population change:

$$P_1 = P_0 + B - D + NDM + NIM$$

where:

P_1 = population at the end of the year

P_0 = population at the beginning of the year

B = births during the year

D = deaths during the year

NDM = net domestic migration during the year

NIM = net international migration during the year

We apply this equation to our beginning population by single year of age, with the result that the population measured by P_1 is always one year older than the population measured by P_0 . To produce estimates of the 2011 household population, this technique is repeated for the year following 2010. We begin with an estimate of the July 1, 2010 household population (as described below) and apply the components of change for July 1, 2010 through June 30, 2011 to produce an estimate of the July 1, 2011 household population.

1. Estimation of the July 1, 2010 Population

Annual population estimates refer to the midpoint of the year (July 1). The first step in the estimation process is to use the April 1, 2010 household estimates base to develop estimates for July 1, 2010. We do this by controlling the household estimates base to existing July 1, 2010 household estimates. For the state-level characteristics estimates we control to the July 1 state total estimates and July 1 national estimates by characteristics, and for the county-level estimates we control to the July 1 county total estimates and the July 1 state estimates by characteristics. This is done using the process described below in the section entitled, Ensuring Consistency with Other Estimates.

2. Estimation of Births and Deaths

The birth and death components are estimated using vital records data from two sources. Members of the Federal-State Cooperative for Population Estimates (FSCPE) provide summary data on all registered births and deaths to residents of their respective states by county for calendar year 2010. The National Center for Health Statistics (NCHS) provides birth and death data which include sex, race, Hispanic origin, age, and month of occurrence detail for years prior to 2010.

The 2009 NCHS data are combined with the Census Bureau's county-level 2000-2010 intercensal estimates for 2009 to create county-level age-specific fertility and mortality rates.⁵ These rates are multiplied by population data from the estimates base to produce preliminary county-level birth and death estimates with sex, race, Hispanic origin, and age (for deaths) detail. These preliminary estimates are used to create full-detail county-level birth and death estimates for the July 1, 2010 to June 30, 2011 interval by a two-step process. First, these estimates are controlled to the county totals provided by the FSCPE members. Second, the results of the first step are controlled to national-level birth and deaths estimates for the July 1, 2010 to June 30, 2011 interval from the national estimates process. State-level estimates are obtained by summing these county estimates.

Data from NCHS on births and deaths differ from inputs that we receive from other agencies in that they are still provided in the four single-race categories specified by OMB's 1977 directive.⁶ Since 2003, NCHS has received birth and death data by the revised OMB categories, but not from all states. Consequently, data collected under the old definitions must be converted into the new race categories using race-bridging factors. In the case of births, race-bridging factors are used to first convert the single-race of the mothers and fathers in the birth data to the revised OMB categories.⁷ Then, data from Census 2000 on the race reported for children when the parents are of different races are used to obtain the race of each birth based on the revised race of the mother and father.

3. Estimation of Domestic Migration

Estimates of domestic migration at the state and county levels are produced by essentially the same method. This method utilizes data from two sources: annual person-level data from tax returns provided by the Internal Revenue Service (IRS); and the Census Bureau's Person Characteristics File (PCF), which is derived from the Social Security Administration's 100 percent file, other administrative records data sources, and Census 2000. Keeping in mind that we estimate components of change for *estimates years* that begin with July 1 of one year and continue to June 30 of the next, the first step is to match the person-level IRS data for the two years in question. These matched records contain the addresses from which the returns were filed in both years. The specific dates to which the addresses pertain depend on when the respective tax returns were filed, and may vary from record to record. However, we assume that this information may be used to estimate migration between July 1 of the first year and June 30 of the second.

The matched person-records are then matched to the PCF, which enables us to identify the age, sex, race, and Hispanic origin for each individual.⁸ We then tabulate the matched records by these characteristics, place (i.e., state or county) of residence in the first year, and place of residence in second year. For each place, person-records are classified as *out-migrants* if the first-year address is in

that place and the second-year address is in a different place. Similarly, person-records are classified as *in-migrants* if the second-year address is in that place and the first-year address is in a different place.

We use person-records to calculate migration rates and proportions, and we assume they may be applied to the full household population to produce migration estimates even though the tax filers and their dependents do not represent the entire population. An out-migration rate for a given place can be calculated using these data by taking the ratio of the out-migrant records to the total matched records for that place. Multiplying this out-migration rate by an estimate of the household population for that place will produce an estimate of that place's domestic out-migration. We calculate and apply out-migration rates for each place by race, sex, Hispanic origin, and age. These rates are applied to estimates of the household population during the cohort-component process to produce estimates of domestic out-migration for each place by age, sex, race, and Hispanic origin.

In-migration is estimated by allocating out-migration to destination places using migration in-proportions. Like the migration rates, the migration proportions are computed as the ratio of two sets of person-records. The numerator of this ratio is the sum of the in-migration records for the place in question and the denominator is the sum of the in-migration records for all places. These in-proportions are computed for all places by race, sex, Hispanic origin, and age. During the cohort-component process these proportions are applied to the national sum of out-migration by age, race, sex, and Hispanic origin to produce estimates of domestic in-migration for each place.

4. Estimation of Net International Migration

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. To determine the state- and county-level age, sex, race, and Hispanic origin distribution of each component, proxy universes are developed that are assumed to be representative of the different components. The demographic characteristics and geographic distribution of these proxy universes are then applied to the totals for each component. For all components except net movement of the Armed Forces population to and from the United States, state-level characteristics are based on the American Community Survey (ACS) three-year 2007-2009 file. County-level characteristics are based on data from the ACS five-year 2005-2009 file. County-level data are controlled to state-level data to ensure the component data sum as required. For the net movement of the Armed Forces population, demographic characteristics and state distributions are based on data collected by the Defense Manpower Data Center (DMDC) and Census 2000.

Immigration of the foreign born is estimated using the ACS question on residence in the prior year. The foreign-born population who indicated that they lived abroad in the prior year are considered immigrants. The number of foreign-born migrants who entered the United States from April 2010 through June 2010 is estimated as one quarter of the foreign-born population in the 2010 ACS who reported living abroad one year ago. Because this question is asked only of those aged one and higher, the estimate of foreign-born immigrants under the age of one is assumed to be equal to half the number of immigrants aged one. Information from the 2010 ACS is used to estimate migration for the July 2010 to June 2011 period because more recent data were not available. The foreign-born population whose year of entry was within five years of the survey year is used as the proxy universe to estimate the state- and county-level characteristics of foreign-born immigrants. Age in the ACS is modified for foreign-born immigrants to represent age at arrival to the United States.

Emigration of the foreign born is estimated using a residual method. The foreign-born household population in Census 2000 is aged forward (using NCHS life tables) to obtain the expected population in 2007 (2008, 2009, and 2010). The expected population is then compared to the population estimated in ACS 2007 (ACS 2008, ACS 2009, and ACS 2010). Subtracting the estimated from the expected populations provides us with the residual, which serves as the basis for our emigration rates for the 2000 to 2007, 2000 to 2008, 2000 to 2009, and 2000 to 2010 time periods. This calculation is performed for two period-of-entry groups: the foreign born who entered the United States between 1990 and 1999, and the foreign born who entered before 1990.

We then calculate three-year averaged rates for each period of entry group and apply the rates to the population at risk of emigrating each year to obtain estimates of emigration of the foreign-born population who entered the United States within the last ten years and of those who entered more than ten years ago. The average of the rates from the 2000 to 2007, 2000 to 2008, and 2000 to 2009 residuals are applied by period of entry to ACS 2009. The annual estimates are then divided by four to obtain estimates of foreign-born emigration by period of entry from April 2010 through June 2010. The average of the rates from the 2000 to 2008, 2000 to 2009, and 2000 to 2010 residuals are applied to ACS 2010 to obtain estimates of emigration by period of entry from July 2010 through June 2011. The proxy universe we use for foreign-born emigrants who entered the United States within ten years of the estimate year is the foreign-born population in the ACS who entered the United States within ten years of the survey year. The proxy universe we use for foreign-born emigrants who entered the United States more than ten years before the estimate year is the foreign-born population in the ACS who entered the United States more than ten years before the survey year.

Data from the American Community Survey and the Puerto Rico Community Survey (PRCS) allow us to estimate the annual migration flows between the United States and Puerto Rico directly using the question on place of prior

residence.⁹ People who indicated on the ACS that they lived in Puerto Rico one year ago are considered immigrants. People who indicated on the PRCS that they lived in the United States one year ago are considered emigrants. The proxy universe for the net migration between the United States and Puerto Rico is the population born in Puerto Rico whose year of entry was five or fewer years before the survey year.

The net migration of natives is based on research by Schachter (2008) using data from over 80 countries.¹⁰ This work compared estimates of the U.S. born or U.S. citizen population living overseas measured at two consecutive time periods and used the difference to develop estimates of net native migration. The proxy universe we use for the net native migrant component is the native household population residing in the United States.

We derive the estimate of the net overseas movement of the Armed Forces population from data collected by DMDC. DMDC provides monthly tabulations of military personnel stationed or deployed outside the United States by age, sex, Hispanic origin, and individual branches of service within the Department of Defense. We assume that change in the overseas military population, excluding deaths, indicates movement of personnel in and out of the United States. To derive the estimates of net movement by race, we apply the race and geographic distribution of the active-duty military population from Census 2000 to DMDC estimates by age, sex, Hispanic origin, and branch of service.

Estimation of the Group Quarters Population

Group Quarters (GQ) population is estimated separately from the household population because of the unique character of this population and our ability to acquire direct data that reflects change in this population. The technique for estimating the GQ population begins with the GQ base population derived from Census 2010. The next step is to estimate GQ change using data supplied by FSCPE members. FSCPE representatives provide independent lists of GQ facilities in their respective states at the county level with the populations typically associated with them at the time of Census 2010. They also provide annual updates to this list that we use to calculate the change in the GQ by type of GQ facility. This change is applied to the GQ base to derive annual estimates of the total GQ by type for each county. In states where no GQ data are submitted by the FSCPE, we hold the GQ base data constant. Finally, we distribute these totals by age, sex, race, and Hispanic origin using the distribution of the GQ population by seven major types from the GQ base.

Ensuring Consistency with Other Estimates

The Census Bureau produces a variety of population estimates for different levels of geography and in differing degrees of demographic detail. There can be minor inconsistencies among them because these different estimates utilize different data and processing techniques. For example, when the initial state characteristics estimates are

summed to state totals, these totals may differ slightly from the estimates produced by our state totals estimation process. Consequently, the final step in estimates production is to control the estimates to previously produced estimates to ensure consistency. We do this by a technique called *raking*, which involves calculating a *rake factor* as the control total divided by the sum of the numbers we wish to control and then multiplying the numbers we wish to control by the rake factor. In the case of the example just mentioned, we would calculate a rake factor for each state and the District of Columbia and then multiply each area's characteristics estimates by their respective rake factor. This process would produce a set of state characteristics estimates whose totals were consistent with the state totals estimates, but it is likely that many of the new estimates would not be integers. Thus, the final step in this process is to apply a technique we refer to as *controlled rounding*, which enables us to convert the estimates to integers without changing the totals.

The state characteristics estimates must be consistent with both the state totals estimates and the national characteristics estimates. The existence of two independent sets of controls is a complication because raking to one set of controls can upset the consistency with the other set of controls. However, by raking first to one set of controls and then to the other for five iterations, the results are approximately consistent with both sets. A specialized rounding procedure is then applied to maintain consistency with two independent sets of controls.

The situation for county characteristics estimates is similar to that for state characteristics estimates. The county characteristics estimates must be consistent with the county totals estimates and the state characteristics estimates. We accomplish this by iterative raking and our specialized rounding, in the same fashion as we do for the state characteristics estimates. By making the county characteristics estimates consistent with the state characteristics estimates, county characteristics estimates become consistent with the state totals and national characteristics, because the state characteristics are consistent with these estimates. Thus, by controlling the state characteristics estimates to the state totals and national characteristics, and then controlling the county characteristics estimates to the county totals and state characteristics, we ensure consistency among all these estimates.

¹ Throughout this document, the term *county* includes county-equivalents such as parishes and independent cities.

² For more information, see <http://www.census.gov/popest/data/historical/files/MRSF-01-US1.pdf>.

³ Office of Management and Budget. *Revisions to the standards for the classification of Federal data on race and ethnicity*. Federal Register 62FR58781-58790, October 30, 1997. Available from: <http://www.whitehouse.gov/omb/fedreg/1997standards.html>.

⁴ This modification is used for all Census Bureau estimates products and is explained in the document entitled "Modified Race Data Summary File Technical Documentation and ASCII Layout" that can be found on the Census Bureau website at <http://www.census.gov/popest/data/historical/files/MRSF-01-US1.pdf>.

⁵ The national intercensal estimates data and documentation are available at <http://www.census.gov/popest/data/intercensal/index.html>.

⁶ Office of Management and Budget. *Race and ethnic standards for Federal statistics and administrative reporting*. Statistical Policy Directive 15. May 12, 1977.

⁷ For a description of the development of NCHS's race-bridging factors, see: Ingram DD, Parker JD, Schenker N, Weed JA, Hamilton B, Arias E, Madans JH. *United States Census 2000 population with bridged race categories*. National Center for Health Statistics. *Vital Health Stat* 2(135). 2003.

⁸ Age is calculated as of the start of the estimation interval using date of birth information from the PCF file.

⁹ The Puerto Rico Community Survey was first fielded in 2005. See <http://www.census.gov/acs/www/SBasics/FlyerPR.htm> for more information.

¹⁰ Schachter, Jason. 2008. "Estimating Native Emigration from the United States," Memorandum.