

## **Methodology for the State and County Resident Population Estimates by Age, Sex, Race, and Hispanic Origin (Vintage 2009): April 1, 2000 to July 1, 2009**

NOTE: These estimates include adjustments due to the effects of hurricanes Katrina and Rita. For a description of these adjustments, refer to Special Processing Procedures for the Areas Affected by Hurricanes Rita and Katrina at <http://www.census.gov/popest/topics/methodology/>.

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.<sup>1</sup> The following documentation describes the process by which we produce population estimates by 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 2000, which defines a resident of a specified area as a person “usually resident” in that area. The Census Bureau develops estimates by updating Census 2000. We begin with the population counts by age, sex, race, and Hispanic origin from Census 2000 and estimate the change that has occurred since that time. This change is measured annually to produce estimates of the population for July 1 of each year from 2000 to 2009. The vintage 2009 estimates contain the most current data available and supersede all previous estimates. The methodology used for the production of the 2009 vintage includes a change in the method used to estimate the age distribution of domestic migration at the county level. 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 2000 provides the starting point for these estimates. The census population is modified in two ways to produce the estimates base.

1. The original race data from the Census are modified to eliminate the "Some Other Race" category.
2. Additional modifications are made to the Census 2000 population counts as documented in the Count Question Resolution program and errata notes.<sup>2</sup>

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.<sup>3</sup> These race categories differ from those used in Census 2000 in one important respect. Census 2000 also allowed respondents to select the category referred to as Some Other Race. When Census 2000 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.<sup>4</sup> 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 2000, because respondents who selected Some Other Race and one of the OMB mandated races in Census 2000 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 2000 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:

$$P1 = P0 + B - D + NDM + NIM$$

where:

P1 = population at the end of the year

P0 = 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 P1 is always one year older than the population measured by P0. To produce estimates of the 2009 household population, this technique is repeated for each year following 2000. We begin with an estimate of the July 1, 2000 household population and apply the components of change for July 1, 2000 through June 30, 2001 to produce an estimate of the July 1, 2001 household population. We then use this estimate as our starting population and apply the next year's components of change to produce an estimate for July 1, 2002, and so on, to July 1, 2009.

## **1. Estimation of the July 1, 2000 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, 2000 household estimates base to develop estimates for July 1, 2000. We do this by controlling the household estimates base to existing July 1, 2000 household estimates. For the state-level estimates we control to the state total estimates and national estimates by characteristics, and for the county-level estimates we control to the county total estimates and the 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 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 years 2000-2008. The National Center for Health Statistics (NCHS) provides individual record data on each registered birth and death occurring in the United States in calendar years 2000-2007, and total registered births and deaths in 2008. The 2000-2007 NCHS data include detail by sex, race, Hispanic origin, age, and month of occurrence. The 2000-2007 county totals from the FSCPE data are controlled to the national total from

the NCHS data for the corresponding year and given the county-level sex-race-Hispanic origin distribution from the NCHS data. Deaths receive the county-level age distribution of the NCHS data.

For vintage 2009 we receive no data for 2009 and only partial data for 2008, so we create projections to complete our time series. These projections are obtained by applying county-level age-specific fertility and mortality rates to county population projections for 2008 and 2009. The result is a complete county-level time series of births and deaths for calendar years 2000-2009. We obtain state-level data by summing the county-level data. Because we produce estimates for July 1 of each year, we require components of change for July 1-June 30 intervals, which we refer to as *estimates years*. Calendar year data are converted into estimates years using NCHS month-of-occurrence information, and controlled to the corresponding data from the national estimates time series. No adjustments are made for undercoverage or differential coverage by state, sex, race, Hispanic origin, or age (for deaths).

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.<sup>5</sup> 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. 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.<sup>6</sup>

### **3. Estimation of Domestic Migration**

The state- and county-level estimation methods for domestic migration differ substantially. The next two paragraphs explain the features the two methods have in common. This is followed by two sections that explain the features unique to each method. The county-method section includes an explanation of the new method, introduced in this vintage, for estimating the age distribution of migration at the county level.

Both methods utilize 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.<sup>7</sup> We then tabulate the matched records by these characteristics, state of residence in the first year, and state of residence in second year.

**a) Estimation of state in- and out-migration.** The estimation of migration for state characteristics estimates begins with the tabulations described in the preceding paragraph. For each state, person-records are classified as *out-migrants* if the first-year address is in that state and the second-year address is in a different state. Similarly, person-records are classified as *in-migrants* if the second-year address is in that state and the first-year address is in a different state.

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. For example, to calculate an out-migration rate for a given state using these data we would take the ratio of the out-migrant records to the total records for that state. This out-migration rate would then be multiplied by an estimate of the household population for that state to produce an estimate of that state's domestic out-migration. We calculate and apply out-migration rates for each state by race, sex, Hispanic origin, and age category. Two precautions are taken to guard against the problems that can be caused by small denominators: 1) for Hispanics, all race groups are combined; 2) the age categories are constructed so that the denominator of the migration rate has at least 30 person-records. These rates are applied to estimates of the household population during the cohort-component process to produce estimates of domestic out-migration for each state by age, sex, race, and Hispanic origin.

State-level domestic in-migration is estimated by allocating out-migration to destination states 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 state in question and the denominator is the sum of the in-migration records for all states. These in-proportions are computed for all states by race, sex, Hispanic origin, and age group in the same fashion as the out-rates. 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 state.

**b) Estimation of migration at the county level.** Domestic migration is estimated at the county level by allocating state-level migration to the counties in that state. We use this approach because the population of many counties is too

small for direct estimation by demographic characteristics to be reliable. The following paragraphs explain first how we allocate the state-level in- and out-migration described previously and how we compute and allocate migration between counties within each state. The final paragraph explains how the final age distribution of migration to and from each county is estimated.

**(i) Allocation of inter-state migration to counties.** The first step in this procedure is to construct county-level migration shares. To allocate state-level out-migration, the county shares are computed as the ratio of each county's out-migrant records (as defined above) to the state's out-migrant records. Similarly, for state-level in-migration the county shares are computed as the ratio of each county's in-migrant records to the state's in-migrant records. These ratios are computed for each of seven race-ethnic groups: Hispanic; non-Hispanic White; non-Hispanic Black; non-Hispanic American Indian and Alaska Native; non-Hispanic Asian; non-Hispanic Native Hawaiian and Other Pacific Islander; non-Hispanic Multi-race. The use of this approach means that for both in-migration and out-migration the shares allocated to each county in a state initially have the same age-sex distributions as the state-level in-migration and out-migration. We choose this approach because we do not have enough data to reliably estimate county-level migration rates with full demographic detail. Our research indicates that county-level migration flows can differ greatly with respect to race and Hispanic origin, while differences with respect to age and sex are usually small by comparison. Consequently, we elect to focus our primary efforts at the county level on the race and Hispanic origin composition of the migration flows.

**(ii) Estimation and allocation of intra-state migration to counties.** To estimate the migration between counties within a state, we first produce state-level estimates of the number of migrants who change counties within each state and then allocate this state-level migration to the counties. Migration between the counties within a state, which we refer to as intra-state migration, is estimated by computing intra-state migration rates for each state and applying them to that state's population. We compute the intra-state migration rates using a method similar to that used to compute the state out-migration rates. These rates are computed for the same categories as the out-migration rates, using, for a given category, the number of person-records that change counties within a state as the numerator, and the total person-records in that category for that state as the denominator. We multiply these rates by the respective state populations to produce estimates of the number of migrants changing counties within each state by age, race, sex, and Hispanic origin.

Next we compute new county migration shares to allocate these intra-state migrants to the counties within each state. Using the same seven race-ethnic groups mentioned earlier, we construct ratios for each county where the numerator is the number of person-records leaving that county for other counties within the state and the denominator is the number of person-records that change counties within that state. Multiplying intra-state migration by these ratios

produces estimates of the migration from each county to other counties in the same state. Then, by constructing a second set of ratios whose numerator is the number of person-records entering each county from other counties within the state and whose denominator is, again, the number of person-records that change counties within that state, we are able to produce estimates of the migration to each county from other counties in the same state.

**(iii) Determining the age-distribution of migration to and from counties.**

Combining the results of (i) and (ii) permits us to produce preliminary estimates of the in- and out-migration for each county by age, sex, race, and Hispanic origin. The age and sex distributions of these estimates are derived from the migration estimated at the state level. The final age distributions are obtained by controlling these county-level in- and out-migration estimates to the age distributions of the in- and out-migrant exemptions, respectively, for each county.

#### **4. Estimation of Net International Migration**

International migration, in its simplest form, is any change of residence across United States (50 states and District of Columbia) borders. The net international migration component of the population estimates combines four parts: (a) the net international migration of the foreign born, (b) the net migration between the United States and Puerto Rico, (c) the net migration of natives to and from the United States, and (d) the net overseas movement of the Armed Forces population between the United States and overseas.

Net international migration of the foreign-born population at the national level is estimated in two parts, immigration and emigration. The estimate of immigration utilizes information from the American Community Survey (ACS) on the reported residence of the foreign-born population in the prior year. The foreign born who reported being abroad in the year prior to the survey are considered immigrants. Because this question is only asked of those ages one and older, we make an additional assumption for foreign-born immigrants under the age of one. We assume that the number of foreign-born immigrants under the age of one is equal to half of the number of one-year-old foreign-born immigrants.

At the national level, emigration of the foreign born is estimated using a residual method. We age forward the foreign-born household population in Census 2000 using NCHS life tables to obtain the expected population in 2006, 2007, and 2008. Then, we compare the expected foreign-born population to the foreign-born population estimated by ACS 2006, ACS 2007, and ACS 2008. Subtracting the estimated from the expected populations produces a residual, which serves as the basis for emigration rates for the 2000 to 2006, 2000 to 2007, and 2000 to 2008 time periods. We perform this calculation for two period-of-entry groups: those who entered the United States between 1990 and 1999; and those who entered

before 1990. The averaged rates for each period-of-entry group are then applied to the population at risk of emigrating each year (i.e., the foreign-born population in the ACS who indicated that they lived in the United States one year ago) to obtain annual estimates of emigrants for 2000 to 2008. To estimate immigration and emigration of the foreign born for later estimate years when ACS data are not available, we maintain the same level and distribution of these migration flows as estimated from the last year of available ACS data.

We produce the estimate of net international migration of the foreign-born population by subtracting the number of emigrants from the number of immigrants. The state distribution and state-level age, sex, race, and Hispanic origin information are estimated for foreign-born immigrants and emigrants separately using data from Census 2000 and the ACS three-year estimates for the 2005 to 2007 time period (ACS 2005-2007). Estimate year 2000 uses information from Census 2000, while 2005 and later years use information from the ACS 2005-2007. The incorporation of ACS data is phased in at the state level through linear interpolation between estimate years 2000 and 2005. The county distribution and county-level age, sex, race, and Hispanic origin information are applied to foreign-born immigrants and emigrants separately using information from Census 2000 for all estimate years. County-level data are controlled to state-level data to ensure the component data sum as required.

Foreign-born immigrants are assigned the state- and county-specific age, sex, race, and Hispanic origin distribution of the foreign-born population who entered the United States within five years before the Census/survey year. Age is adjusted for foreign-born immigrants to represent age at arrival to the United States. For foreign-born emigration, characteristics are applied by period of entry. The state- and county-specific age, sex, race, and Hispanic origin distributions of the foreign-born who entered the United States within ten years before the Census/survey year are applied to the estimate of emigrants who entered the United States within ten years before the estimate year. The state- and county-specific age, sex, race, and Hispanic origin distribution of the foreign born who entered the United States more than ten years before the Census/survey year is applied to the estimate of emigrants who entered the United States more than ten years before the estimate year.

Net migration between the United States (the 50 states and the District of Columbia) and Puerto Rico is also estimated at the national level in two parts, immigration and emigration, when possible. For 2005 and later years, the ACS and the Puerto Rico Community Survey (PRCS) allow us to estimate the annual migration flows directly, using the question on place of residence one year ago.<sup>8</sup> People who indicated on the ACS that they lived in Puerto Rico one year ago are considered immigrants (i.e., they moved from Puerto Rico to one of the 50 states or the District of Columbia). People who indicated on the PRCS that they lived in the United States one year ago are considered emigrants (i.e., they moved from one of the 50 states or the District of Columbia to Puerto Rico). We assume the

number of immigrants and emigrants under the age of one is equal to half of the number of one-year-old immigrants and emigrants, respectively. To estimate migration between the United States and Puerto Rico for later years when ACS and PRCS data are not available, we maintain the same level and distribution of these migration flows as estimated from the last year of available ACS and PRCS data.

For 2000 to 2004, we use prior research to establish a base estimate of net migration between the United States and Puerto Rico for 2000 and linearly interpolate between the 2000 net estimate and the 2005 net estimate to generate the estimates for 2001 to 2004.<sup>9</sup> For 2000, the state distribution and state-level age, sex, race, and Hispanic origin distribution are based on the Census 2000 population born in Puerto Rico who entered the United States in 1995 or later. For 2005 and later years, we base the state distribution and state-level age, sex, race, and Hispanic origin distribution of these net migrants on the demographic characteristics of the ACS 2005-2007 population born in Puerto Rico who entered the United States within 10 years before the survey year. As described for foreign-born immigration and emigration, the incorporation of ACS data is phased in at the state level through linear interpolation between estimate years 2000 and 2005. The county distribution and county-level age, sex, race, and Hispanic origin information from Census 2000 are applied to net Puerto Rico migrants for all estimate years using characteristics of the population born in Puerto Rico who entered the United States in 1995 or later. County-level data are controlled to state-level data to ensure the component data sum as required.

We estimate the net migration of natives to and from the United States at the national level using levels observed during the 1990s.<sup>10</sup> We apply the age, sex, race, and Hispanic origin distribution of natives residing in the United States to the estimate of net native migration. For 2000, the state distribution and state-level age, sex, race, and Hispanic origin information are obtained from Census 2000. For 2005 and later years, the state distribution and state-level age, sex, race, and Hispanic origin information are obtained from ACS 2005-2007. The incorporation of information on the state-distribution and state-level characteristics from ACS 2005-2007 is phased in through linear interpolation between estimate years 2000 and 2005. The county distribution and county-level age, sex, race, and Hispanic origin information from Census 2000 are applied to the net native migrants for all estimate years using characteristics of natives residing in the United States. County-level data are controlled to state-level data to ensure the component data sum as required.

We derive county-level data on net overseas movement of the Armed Forces population for the current estimates period using data originally supplied by the Defense Manpower Data Center (DMDC). DMDC provides data by age, sex, Hispanic origin, and individual branches of service in the Department of Defense. We apply the race distribution from the Census 2000 active military population to the Armed Forces movement overseas component. These results are distributed

to states using the state distribution in the DMDC data. Then, we distribute these state-level data to counties using the county distribution of the Armed Forces data from Census 2000. Finally, we control the county-level data to the original national-level data to ensure the component data sum as required.

### **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 2000. 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 2000. 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 come up with 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 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 state's (and DC'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 complicates the problem because raking to one set of controls can upset the consistency with the other set of controls. However, we have learned from experience that 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.

---

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

<sup>2</sup> Details about the Count Question Resolution Program can be found on the Census Bureau website at <http://www.census.gov/dmd/www/CQR.htm>. Errata notes can be found on the Census Bureau website at <http://www.census.gov/prod/cen2000/notes/errata.pdf>.

<sup>3</sup> 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>.

<sup>4</sup> 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/archives/files/MRSF-01-US1.html>.

<sup>5</sup> Office of Management and Budget. Race and ethnic standards for Federal statistics and administrative reporting. Statistical Policy Directive 15. May 12, 1977.

<sup>6</sup> 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.

<sup>7</sup> Age is calculated as of the start of the estimation interval using date of birth information from the PCF file.

<sup>8</sup> The Puerto Rico Community Survey was first fielded in 2005. See <http://www.census.gov/acs/www/SBasics/FlyerPR.htm> for more information.

<sup>9</sup> For more information on the net movement from Puerto Rico see Christenson, M., 2002, "Evaluating Components of International Migration: Migration Between Puerto Rico and the United States," Population Division Technical Working Paper No. 64.

<sup>10</sup> For more information on estimates of native emigration, see Gibbs, J., G. Harper, M. Rubin, and H. Shin, "Evaluating Components of International Migration: Native-Born Emigrants," Population Division Technical Working Paper No. 63.