

Methodology for State and County Population Estimates by Age, Sex, Race, and Hispanic Origin for Vintage 2006¹

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/>.

[PDF Version of this methodology](#)

The U.S. Census Bureau 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 on an annual basis.² The following documentation outlines the methods used in the production of the vintage 2006 estimates.

Overview

These estimates are produced by updating Census 2000. That is, we begin with population counts 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 2006. In addition to the changes introduced to account for the effects of hurricanes Katrina and Rita, the production of the current state-level estimates differs from that of earlier state-level estimates in one important respect. Previously we produced state estimates by age and sex and then, by a separate process, added race and Hispanic origin detail. We have now discontinued use of that method. For this set of estimates, we use one process to produce state estimates by age, sex, race, and Hispanic origin, which is the essentially the same as that used to

produce the county-level age-sex-race-Hispanic origin estimates. A detailed discussion of these methods is provided 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, and change in the GQ population is estimated by type using data received annually from members of the Federal State Cooperative Program for Population Estimates.

The vintage 2006 *state* estimates differ from previous vintages in that this vintage treats college students not living in dormitories as part of the GQ population during the production process.³ This change was introduced into the *county* estimates in the 2005 vintage (prior to 2005, college students living in dormitories were treated as part of the GQ population but all other college students were treated as part of the household population). The resulting household and GQ

estimates are added together to produce the new set of resident population estimates.

Specification of the Base Populations

The enumerated population from Census 2000 provides the starting point for these estimates. This population is modified in two ways to produce what we refer to as the estimates base.

1. The original race data from the Census are modified to eliminate the "some other race" category.⁴
2. The April 1, 2000 base populations reflect modifications to the Census 2000 population counts as documented in the Count Question Resolution program and errata notes.⁵

We also apply these modifications to the GQ population enumerated in Census 2000 to produce the GQ base population. The GQ base is subtracted from the estimates base to produce the household base population.

Estimating 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. Births are only used to estimate the population of age 0 at the end of the year. To produce estimates of the July 1, 2006 household population, this technique is used six successive times. 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 beginning population and apply the next year's components of change to produce an estimate for July 1, 2002, and so on, to July 1, 2006. These estimates have age, race, sex, and Hispanic origin detail, requiring that the components of change also have all this detail. Most of the work involved in the use of this method is the estimation of the components of change with age, sex, race, and Hispanic origin detail. The discussion below explains how this is done.

1. Estimation of the July 1, 2000 Population

Annual population estimates are designed to reference the midpoint of the year (July 1). Since Census 2000, and the base populations derived from it, reference April 1, 2000, the next step in the estimation process is to use the household base to develop estimates for July 1, 2000. We do this by controlling the household base to existing July 1, 2000 household

estimates 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 Program for Population Estimates (FSCPE) provide summary data on all registered births and deaths to residents of the members’ respective states for calendar years 2000-2005. The 2000-2004 data include county totals; the 2005 data provide state totals that are given the county total distributions from the 2004 data. The National Center for Health Statistics (NCHS) provides individual data on each registered birth and death occurring in the United States in calendar years 2000-2004 and total registered births and deaths in 2005. The 2000-2004 NCHS data include sex, race, Hispanic origin, and age (for deaths) detail, as well as month of occurrence. For each year, the county totals from the FSCPE data are controlled to the national total from the NCHS data and given the county-level sex-race-Hispanic origin distribution from the NCHS data. Additionally, deaths receive the county-level age distribution of the NCHS data. The 2005 data are given the county-level sex, race, Hispanic origin, and age (for deaths) distribution from the 2004 data. Because we need components of change by July-June intervals, the data are next converted from calendar year into these estimates intervals. We have no data for the last 6 months of the July 1, 2005 – June 30, 2006 estimates interval, so we produce preliminary estimates by assuming that the last six months are equal to the first six months of this period (July 1 – December 31, 2005). We produce the final estimates by controlling the preliminary estimates to national-level birth and death projections produced as part of the national population estimates process. No adjustments are made for

under coverage or differential coverage by state, sex, race, or Hispanic origin.

The processing of birth and death data for the vintage 2006 estimates differs from that of previous vintages in one important respect. NCHS data are received in race categories consistent with the 1990 Census, which permits respondents to choose from one of four race categories (White, Black, American Indian and Alaska Native, Asian and Pacific Islander). In Census 2000 the Asian and Pacific Islander category is split into two categories (Asian, Native Hawaiian and Other Pacific Islander) and respondents are permitted to choose more than one category, creating five single-race categories and 26 multiple race categories. In previous vintages, processing was done using the 1990 categories. For vintage 2006 processing, these data are converted into race categories consistent with Census 2000 by a procedure that employs race-bridging factors developed by NCHS.⁶

3. Estimation of Domestic Migration

The state- and county-level estimation methods for domestic migration differ substantially. The following discussion explains the features the two methods have in common, and then explains the features unique to the county method.

a) Features common to state and county. Both methods utilize data from two sources: annual individual-level extracts of 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 100 percent file, other administrative records data sources, and Census 2000. Preparation of migration estimates for the current set of population estimates began with the receipt from IRS of data extracted from tax returns filed in 2006, which was then merged with similar data

from 2005. These merged records each contain the filer's address in 2005 and 2006 and the number and type of the exemptions claimed, which enable us to estimate the size and composition of the household.⁷ The specific dates to which the 2005 and 2006 addresses pertain depend on when the respective tax returns were filed, and thus vary considerably from record to record. However, we assume that this information may be used to estimate migration between July 1, 2005 and June 30, 2006.

The merged IRS records are then matched to the PCF, which enables us to identify the age, sex, race, and Hispanic origin of the tax filers.⁸ Demographic characteristics are assigned to exemptions claimed for spouses and dependents using several simplifying assumptions. First, spouses are assigned the same age and the opposite sex as filers. Second, exemptions claimed for dependent children are assigned to the under-20 age group and exemptions claimed for dependent parents are assigned to the 65 and over age group. Third, the sex of dependent children and parents is assigned randomly. Fourth, the spouse and other dependents are assigned the same race and Hispanic origin as the filer. We then tabulate the exemptions by these characteristics, state of residence in 2005, and state of residence in 2006. For each state, exemptions are classified as *out-migrants* if the 2005 address is in that state and the 2006 address is in a different state. Similarly, exemptions are classified as *in-migrants* if the 2006 address is in that state and the 2005 address is in a different state.

We use exemptions 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 exemptions to the total exemptions 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 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 exemptions. This set of out-migration rates is combined with rates computed during previous years' processing to produce a time-series with a set of out-migration rates for each estimates year. 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 exemptions. The numerator of this ratio is the sum of the in-migration exemptions for the state in question and the denominator is the sum of the in-migration exemptions 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) Features unique to the county method. 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 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 exemptions (as defined above) to the state's out-migrant exemptions. Similarly, for state-level in-migration the county shares are computed as the ratio of each county's in-migrant exemptions to the state's in-migrant exemptions. These ratios are computed for each of seven race-ethnic groups: Hispanics, and non-Hispanics broken down into the six race groups. The use of this approach means that for both in-migration and out-migration the shares allocated to each county in a state have the same age-sex distribution as the state-level migration. We choose this approach for the following reasons. These data are not sufficient to reliably estimate migration at the county level with full demographic detail. Our research indicates that county-level migration flows can differ very 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 efforts at the county level on the race and Hispanic origin composition of the migration flows.

The procedure described in the previous paragraph enables us to estimate the numbers of migrants leaving each county for other states and entering each county from other states. However, for a complete estimate of domestic migration at the county level we also must estimate the migration between counties within a state. To do this, 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 exemptions that change counties within a state as the numerator, and the total exemptions 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 exemptions leaving that county for other counties within the state and the denominator is the number of exemptions 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 exemptions entering each county from other counties within the state and whose denominator is, again, the number of exemptions 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.

To summarize, for characteristics estimates we estimate domestic migration at the county level by allocating state-level migration to the counties within that state. This process utilizes state in- and out-migration estimated as part of the state characteristics estimates process and intra-state migration estimated specifically for this process. These state-level migration estimates are allocated to counties by county migration shares, which are the respective counties' proportions of the exemptions used in the numerators of the rates or proportions used to compute the state-level migration. Application of these county migration shares to the state-level migration estimates yields estimates of each of the four sub-components of county-level net domestic migration: 1) migration from each county to other states; 2) migration to each county from other states; 3) migration

from each county to other counties in the same state; and 4) migration to each county from other counties in the same state. Combining these four sets of migration estimates produces estimates of net domestic migration for each county by age, race, sex and Hispanic origin.

4. Estimation of Net International Migration

For the purposes of estimates production, 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 combines three parts: (a) net migration of the foreign born, (b) emigration of natives and net movement between Puerto Rico and the United States, and (c) net movement of the Armed Forces population.

a) Net migration of the foreign born. In an effort to maximize the use of available data, we use 2000-2005 American Community Survey (ACS) data as the basis for the level of net migration of the foreign born. After determining the net change of the foreign-born population, we account for deaths to the entire foreign-born population during the periods of interest. To account for variability due to small sample sizes of the foreign born, we use a moving average for the period changes to produce the final net foreign-born migration estimate. We apply the county-age-sex-race-Hispanic origin distribution of the non-citizen foreign-born population from Census 2000 who entered in 1995 or later to the national-level estimate of net migration of the foreign born.

b) Emigration of natives and net movement between Puerto Rico and the United States. We produce the net movement between Puerto Rico and the United States and the emigration of natives in similar ways.⁹ For both pieces, current annually-updated information is not available. Therefore,

we estimate net national levels of movement using levels observed during the 1990s, and apply the distributions from Census 2000 that are most similar to the population of interest.¹⁰ For the net movement between Puerto Rico and the United States, we base the distribution on the characteristics (age, sex, race, Hispanic origin) and geographic location (county) of the Census 2000 population born in Puerto Rico and who entered the United States in 1995 or later. For native emigration, we assume these emigrants are likely to have the same characteristics distributions as natives who currently reside in the United States. Therefore, we apply the age, sex, race, Hispanic origin, and county distribution of natives residing in the 50 states and the District of Columbia to the native emigrant population.

c) Net movement of the Armed Forces population. We estimate net movement of the Armed Forces at the state level by allocating the net movement estimated at the national level. The national-level net Armed Forces movement is estimated as the monthly change in armed forces overseas using data from the Defense Manpower Data Center and Census 2000. These national estimates are aggregated into annual changes and distributed to the states using annual station strength estimates received directly from the Armed Forces and the Department of Defense, with each state's estimate receiving the annual age-race-sex-Hispanic origin distribution of the aggregated national net Armed Forces movement estimates. The state estimates are allocated to counties using the county-level distribution of total Armed Forces population from Census 2000, with each county's estimate receiving the national-level age-race-sex-Hispanic origin distribution.

Finally, we combine the subcomponents of international migration (net migration of the foreign born plus net migration between Puerto Rico and

the United States minus emigration of natives plus net Armed Forces movement) to produce the net international migration component.

Estimating the GQ Population

Group Quarters (GQ) population is estimated separately from the household population because of the unique character of this subpopulation 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 as described in the “Specification of the Base Populations” section above. The next step is to estimate GQ change using data supplied by FSCPE members. The state FSCPE representatives have developed 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 send us 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 type 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 official 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. Rounding the results poses an additional problem, because no simple rounding procedure will ensure that consistency is maintained with both sets of controls. We have solved this problem by developing a rounding procedure that is specifically designed to maintain consistency with two independent sets of controls. Thus, the final step in state characteristics estimates production is controlling the estimates to be consistent with both the national characteristics estimates and the state total estimates by the use of iterative raking and our specialized rounding procedure.

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. When the county characteristics estimates become consistent with the state characteristics estimates, they also become consistent with the other estimates with which the state characteristics estimates are consistent. That is, making the county characteristics estimates consistent with the state characteristics estimates also makes the county characteristics estimates 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 The term *vintage* is used here to refer to the year in which we begin production on a set of estimates. Thus, the vintage 2006 estimates are those estimates whose production was begun in 2006.

² Throughout this document, the term *county* includes county-equivalents such as parishes and independent cities.

³ Though college students not living in dormitories are not part of the GQ population, we assume their demographic behavior is basically the same as that of dormitory students.

⁴ This modification has been accepted 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>.

⁵ 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>.

⁶ 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.

⁷ If the matched returns show different numbers of exemptions, we use the number from the earlier year.

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

⁹ For more information on the net movement from Puerto Rico and native emigration, see Kevin E. Deardorff and Lisa Blumberman, 2001, "Evaluating Components of International Migration: Estimates of the Foreign-Born Population by Migrant Status in 2000," Population Division Working Paper Series No. 58.

¹⁰ For more information on the estimate of 11,133 for the net movement from Puerto Rico see Christenson, M., "Evaluating Components of International Migration: Migration Between Puerto Rico and the United States," Population Division Technical Working Paper No. 64. For 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.