Methodology for the State and County Population Estimates by Age, Sex, Race, and Hispanic Origin (Vintage 2007): April 1, 2000 to July 1, 2007

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[1] in the United States and the District of Columbia. The following documentation describes the process by which we produce the July 1, estimates of these demographic characteristics at the state and county level.

Overview

The Census Bureau develops estimates of demographic characteristics 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 2007. Thus, each vintage of estimates produces annual estimates for each year from the most recent Census to the current year, and the estimates for a given year may change somewhat as new administrative data are incorporated into the estimate. The vintage 2007 estimates contain the most current data available and supersede all previous vintages. The methodology used for the production of the July 1, 2007 vintage includes changes in the methods used to estimate international migration, domestic migration, and births and deaths. These components are further described in the sections below.

Race Categories Used in These Estimates

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.[2] 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 was edited to produce the estimates base, as described in the Specification of the Base Populations section below, respondents who selected the Some Other Race category alone were assigned to one of the OMB mandated categories.[3] 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 these categories two different ways. One of these ways is presenting the five single-race categories with a sixth group that combines all categories with more than one race – referred to in our tables as Two or More Races. The other method of presentation is to use five alone or in combination race groups. Each of these groups contains one of the single-race categories plus all the multiple-race categories that include that single race. Alone or in combination groups cannot be summed to create a 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.

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: 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 using data received annually from members of the Federal State Cooperative Program for Population Estimates.

The vintage 2007 state estimates differ from previous vintages in that this vintage treats military population not living in barracks as part of the GQ population during the production process. This change is introduced only into the state estimates because we do not have data on the military population at the county level. 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 base 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, as described in the Race Categories Used in These Estimates section above.

2. The April 1, 2000 base population reflects modifications to the Census 2000 population counts as documented in the Count Question Resolution program and errata notes.[4]

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.

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:

$P_1 = P_0 + B - D + NDM + NIM$
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. Births are only used to estimate the population of age 0 at the end of the year. To produce estimates of the July 1, 2007 household population, this technique is repeated for each year from April 1, 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, 2007. Consequently, as the base population is aged forward the youngest ages are estimated by the components of change, primarily births. The components of change must have age, race, sex and Hispanic origin 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 this further.

1. **Estimation of the July 1, 2000 Population**

   Annual population estimates are designed to reference 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 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 by county for calendar years 2000-2006. 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 2000-2005, and total registered births and deaths in 2006. The 2000-2005 NCHS data include sex, race, Hispanic origin, and age (for deaths) detail, as well as the month of occurrence. The 2000-2005 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 (again, for the corresponding year). Additionally, deaths receive the county-level age distribution of the NCHS data. Since, in vintage 2007, we receive no data for 2007 and only partial data for 2006, we require a method for generating the data needed to complete our time series. In previous vintages we used data from earlier years for this purpose; for vintage 2007 we used projections to supply the missing data. To create these projections we developed county-level age-specific fertility and mortality rates using birth, death, and population data from previous vintages and applied these rates to county population estimates for 2006 and projections for 2007. The result of the work just described is a complete county-level time-series of births and deaths for calendar years 2000-2007. However, since we produce estimates for July 1 of each year we require components of change for July1-June 30 intervals, which we refer to as *estimates years*. Consequently, the final step of this production is to convert the calendar year data into estimates years, using NCHS month-of-occurrence information, and control the results 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).

NCHS data differ from those we receive from many other agencies in that NCHS still uses the four race categories specified by OMB's 1977 directive.[5] Consequently these data must be converted into the new race categories. We do this using 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. The discussion also explains how this estimation has been improved for vintage 2007.

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'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 merge the IRS data for the two years in question. These merged records each 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 thus vary considerably 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. In previous vintages we were only able to perform matches on the tax filer while now we are able to perform matches on each exemption. This means that previously we assumed that all exemptions had the same migration behavior as the filer, while now we can detect when an exemption changes addresses independently of the filer.

The merged IRS records are then matched to the PCF, which enables us to identify the age, sex, race, and Hispanic origin for each exemption.[7] Previously we were only able to match the filer to the PCF, and we imputed the characteristics of the other exemptions based on those of the filer. We then tabulate the exemptions by these characteristics, state (or county) of residence in the first year, and state (or county) of residence in second year. For each state (or county), exemptions are classified as *out-migrants* if the first-year address is in that state (or county) and the second-year address is in a different state (or county). Similarly, exemptions are classified as *in-migrants* if the second-year address is in that state (or county) and the first-year address is in a different state (or county).

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 (or county) using these data we would take the ratio of the out-migrant exemptions to the total exemptions for that state (or county). This out-migration rate would then be multiplied by an estimate of the household population for that state (or county) to produce an estimate of that state's (or county's) domestic out-migration. We calculate and apply out-migration rates for each state (or county) 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. 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 (or county-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 distributions as the state-level in migration and out 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 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

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.

The Vintage 2007 net international migration estimates for the first two parts (net international migration of the foreign born and net migration between the United States and Puerto Rico) are created using a combination of the Vintage 2006

estimate and a new method, which we anticipate using for Vintage 2008 and later years. The estimate of emigration of natives from the United States remains unchanged from Vintage 2006.

a) Net migration of the foreign born. The Vintage 2007 net international migration estimate for the foreign-born population is created by averaging the Vintage 2006 estimate and the estimate generated using a new method for calculating foreign-born migration to and from the United States (50 states and District of Columbia). The Vintage 2006 estimate uses the change in the number of foreign born in two consecutive years of American Community Survey (ACS) data, with an adjustment for deaths to the foreign-born population. The new method utilizes information from the ACS on the reported residence of the foreign-born population in the prior year. Those who reported being abroad in the year prior to the survey are considered immigrants. We estimate the number of foreign-born emigrants separately using a residual method utilizing Census 2000 and ACS data. Subtracting the emigrants from the immigrants results in the new method's net international migration estimate for the foreign-born population. This estimate is then averaged with the Vintage 2006 estimate of the same population group to create the Vintage 2007 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) Net movement between Puerto Rico and the United States. The Vintage 2007 estimate of net migration between the United States and Puerto Rico is created by averaging the Vintage 2006 estimates and new estimates generated using American Community Survey (ACS) and Puerto Rico Community Survey (PRCS) data. For the 2000 to 2004 time periods, we estimate the net migration using levels observed during the 1990s.[8] In Vintage 2006, this estimate was held constant and carried forward to the end of the estimate period. In Vintage 2007, we average these constant estimates with new estimates for the 2004 through 2006 time periods. The new estimates utilize ACS and PRCS data on residence one year ago, subtracting the estimate of emigration from the United States to Puerto Rico from the estimate of immigration to the United States from Puerto Rico. 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.

c) Emigration of natives from the United States. For Vintage 2007, we estimate native emigration using levels observed during the 1990s.[10] We assume these emigrants are likely to have the same geographic and characteristics distributions as natives in the United States. We apply the Census 2000 age, sex, race, and Hispanic origin distribution of natives residing in the 50 states and the District of Columbia to the native emigrant population.

d) 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 (DMDC) and Census 2000. These national estimates are aggregated into annual changes and distributed to the states by age, race, sex and Hispanic origin using annual military population estimates received from DMDC. The race data used for this allocation is taken from the Census 2000 military state population; the remaining information is from the DMDC data. 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 state-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 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] Throughout this document, the term *county* includes county-equivalents such as parishes and independent cities.

[2] 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.

[3] 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.
[4] 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.
[5] Office of Management and Budget. Race and ethnic standards for Federal statistics and administrative reporting. Statistical Policy Directive 15. May 12,1977.
[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] Age is calculated as of the start of the estimation interval using date of birth information from the PCF file.
[8] 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.
[10] 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.