Estimates and Projections Area Methodology
County Population Estimates by Age, Sex, Race, and Hispanic Origin for July 1, 2004

The U.S. Census Bureau produces estimates of the resident population by age, sex, race, and Hispanic origin for each county in the United States and the District of Columbia on an annual basis. The following documentation outlines the methods that were used in the production of the July 1, 2004 estimates.

Overview

For the July 1, 2004 county estimates of the resident population by age, sex, race, and Hispanic origin, the Census Bureau used a distributive cohort component method. This method was applied in the following manner. First, we started with previously developed resident county population estimates by age (0-64 and 65+) and resident state population estimates by age, sex, race, and Hispanic origin. Second, we estimated the age, sex, race, and Hispanic origin distributions for each county by estimating post-censal change in the corresponding populations with a cohort component model. Third, we applied these distributions to the original county estimates by age and state characteristics estimates. A detailed discussion of this method is provided below.

Estimating the Age, Sex, Race, and Hispanic Origin Distributions

The majority of the work that went into producing the county resident population estimates by age, sex, race, and Hispanic origin consisted of estimating the age, sex, race, and Hispanic origin distributions for each county age estimate. This was done by producing a preliminary set of age, sex, race, and Hispanic origin estimates for each county and then calculating the age-sex-race-Hispanic origin proportions from these.

The preliminary set of resident population estimates by age, sex, race, and Hispanic origin was produced by first splitting the Census population into two mutually exclusive universes: The household population and the group quarters (GQ) population. For the household population, a cohort component technique was then used to estimate change in this population. For the GQ population, GQ change was estimated through a data-collection effort conducted in conjunction with members of the Federal State Cooperative Program of Population Estimates (FSCPE). The resulting household and GQ estimates were then added together to produce the new set of preliminary resident population estimates.

The Preliminary Household Population Estimates

The cohort-component technique used to estimate the household population follows each birth cohort across time according to its exposure to mortality, fertility, and migration. This technique was applied using the following equation.

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

Where:

P1 = population at the end of the period
P0 = population at the beginning of the period
B = births during the period
D = deaths during the period
NDM = net domestic migration during the period
NIM = net international migration during the period
NMM = net military movement during the period

The actual application of this algorithm was quite simple. Not so simple was the preparation of the input data. Great care was taken to estimate each component of this equation as accurately as possible. The details of this work are outlined below.

One complicating factor in the estimation of the household population was that the administrative data used in these estimates continues to come to the Census Bureau with different race categories than used in Census 2000. In Census 2000, information was gathered using 6 races groups (White; Black; American Indian and Alaska Native; Asian; Native Hawaiian and Pacific Islander; and Some Other Race). In addition, individuals were allowed to report multiple races. Conversely, most administrative data available for the estimates still come to us in the 4 race categories consistent with the 1990 census (White; Black; American Indian, Eskimo, or Aleut; and Asian and Pacific Islander). For this reason, the household estimates were processed with the 4 race categories consistent with the 1990 census and the results were converted to race categories consistent with Census 2000. Details of all the conversions needed in order to carry out this procedure are presented below.

1. Specification of the Base Population

The enumerated population from Census 2000 was the base for the July 1, 2004 estimates. This population was modified in four ways to prepare it for inclusion in the cohort-component technique.

a. The original race data from the Census were modified to eliminate the "some other race" category.[1]
b. The April 1, 2000 population estimates base reflects modifications to the Census 2000 population as documented in the Count Question Resolution program and errata notes.[2]

[1] This modification has been accepted for all Census Bureau estimates produces 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 Hhttp://eire.census.gov/popest/data/national/tables/files/mod_race.phpH.

[2] Details about the Count Question Resolution Program can be found on the Census Bureau website at Hhttp://www.census.gov/dmd/www/CQR.htmH. Errata notes can be found on the Census Bureau website at Hhttp://www.census.gov/prod/cen2000/notes/errata.pdfH.

c. The Census 2000 base household population was converted from the 31 race categories to the four race groups consistent with the 1990 Census by "straight proportional allocation."
d. Finally, the results of step [c] were used to estimate the population on July 1, 2000.[3] This was done as follows:
    i. A set of age-sex-race-Hispanic origin proportions was calculated by summing the data by county and age (0 to 64 and 65+) and then dividing each age-sex-race-Hispanic origin cell by the corresponding sum.
    ii. These proportions were then applied to the previously produced county age estimates for July 1, 2000.
    iii. Finally, the results were controlled to the July 1, 2000 state estimates by age, sex, race, and Hispanic origin.

2. Specification of Births

The birth component was calculated with data from two sources. The Federal State Cooperative Program for Population Estimates (FSCPE) provided data on all registered births that occurred in each county of the members' respective state for calendar years 2000-2002. In addition, they provided births for their state as a whole for calendar year 2003, which were subsequently distributed by county using the distribution from the FSCPE 2002 births. After this, the FSCPE births for each year were adjusted and distributed by sex, race (consistent with the 1990 Census), and Hispanic origin using data from The National Center for Health Statistics (NCHS). Finally, data for the last 6 months of the estimate period (January 1 – June 30, 2004) were assumed to be equal to the number of births from July 1 – December 31, 2003. These final birth counts were considered to be complete for the resident population. No adjustments were made for under coverage or differential coverage by county, sex, race, or Hispanic origin.

3. Specification of Deaths

The death component was also calculated with data from two sources. The Federal State Cooperative Program for Population Estimates (FSCPE) provided data on all registered deaths that occurred in each county of the members' respective state for calendar years 2000-2002. In addition, they provided deaths for their state as a whole for calendar year 2003, which were subsequently distributed by county using the distribution from the FSCPE 2002 deaths. After this, the FSCPE deaths for each year were adjusted and distributed by age, sex, race (consistent with the 1990 Census), and Hispanic origin using data from the National Center for Health Statistics (NCHS). Finally, data for the last 6 months of the estimate period (January 1 – June 30, 2004) were assumed to be equal to the number of deaths from July 1 – December 31, 2003. These final deaths counts were considered to be complete for the resident population. No adjustments were made for under coverage or differential coverage by county, age, sex, race, or Hispanic origin.

[3] This step was needed since the estimates procedure produces annual estimates and since the target reference date for each estimate year is July 1.

4. Specification of Net International Migration

The net international migration component consisted of three migration flows: (1) net migration of the foreign-born, (2) emigration of natives, and (3) net movement from Puerto Rico to the United States.

To measure the net migration of the foreign born, we used the American Community Survey (ACS) because it provided annually updated data. We first determined the national level of migration for the foreign born by calculating the net difference in the estimates of these surveys from 2000 to 2001, 2001 to 2002, and 2002 to 2003. We then accounted for deaths to the entire foreign-born population during the periods of interest to arrive at the final national estimate of net migration of the foreign born. In order to account for variability due to small sample sizes of the foreign born, we used a moving average for the period changes to get the final net foreign-born migration estimate. Then, to ascribe county of destination, age, sex, race, and Hispanic origin to these estimates, we applied the distribution of the non-citizen foreign born from Census 2000 who entered in 1995 or later to the national-level estimate. Finally, we assumed the net migration of the foreign born between 2003 and 2004 to be the same as net migration between 2002 and 2003.

The national net movement from Puerto Rico to the United States by age and sex was measured using levels observed during the 1990s.[4] To assign characteristics to these flows, we applied the age-sex-race-Hispanic origin-destination county distributions from Census 2000 from those who indicated that their place of birth was Puerto Rico and who had entered the United States in 1995 or later.

The emigration of natives was produced in a similar way to the net movement from Puerto Rico. Again, the national levels of movement were measured using levels observed during the 1990s.[5] Then, to assign characteristics to these flows, we applied the age-sex-race-Hispanic origin-destination county distributions from Census 2000 of all natives who currently reside in the United States. Therefore, the characteristics of natives who emigrated were assumed to be the same age-sex-race Hispanic origin-destination county distribution as natives residing in the 50 states and the District of Columbia in Census 2000.

Once the net migration of the foreign born, net movement from Puerto Rico, and the emigration of natives were estimated, all three parts were combined to estimate a final net international migration component.

5. Specification of Net Domestic Migration

The input data used for the estimation of county net domestic migration came from two administrative sources: annual individual-level extracts of tax returns provided by the

[4] A description of the methodology used to produce these estimates can be found on the Census Bureau website at Hhttp://www.census.gov/population/www/documentation/twps0064.htmlH.
[5] A description of the methodology used to produce these estimates can be found on the Census Bureau website Hhttp://www.census.gov/population/www/documentation/twps0063.htmlH.

Internal Revenue Service (IRS); and the Census Numident file derived from the Social Security Administration 100 percent file (SSA). The IRS 1040 tax return records were matched to the SSA data to identify the age, sex, race, and Hispanic origin of each tax filer, their spouse, and their dependents. After this, two years of records were matched and a migration status was assigned to each filer. Filers with a change in residence across county boundaries between the two periods along with their spouses and dependents were identified as "domestic migrants." If there was no change in the county of residence, the filers and their spouses and dependents were identified as non-migrants.

While it would have been theoretically possible to estimate county domestic in-migration and out-migration by calculating county-to-county flows, we decided against this based on the nature of the input data and the calculations required to transform them into the estimates of net migration. In short, since the IRS exemption data reflected only a portion of the population, we needed to convert these data into rates and then apply them to an estimate of the entire population. At the county level, the creation of rates for each demographic-characteristic group (age, sex, and race-origin) within all 3,141 counties would have resulted in millions of cells of data, most with values either equaling zero or too small to produce robust estimates of migration. Therefore, all calculations of the number of domestic migrants for the July 1, 2004 county characteristics estimates occurred at the state level and then were distributed to the counties by proportional allocation.

The process of estimating the net number of domestic migrants for each county was comprised of four steps. First, the total number of domestic migrants crossing county borders within each state was estimated. This was done with a process parallel to the one used to estimate the number of inter-state migrants for each state and occurred in the following manner.[6] To begin, an inter-county migration rate for each age-sex-race-Hispanic origin group within a state was computed by dividing the total number of exemptions associated with filers who moved across county borders during the period by the number of exemptions associated with filers in the state as a whole at the beginning of the period.[7] [8] Next, the rates were smoothed across ages using a moving average. Finally, the rates were applied to an estimate of the state population by characteristics at the beginning of the period from the state characteristics estimates to generate the number of domestic within-state county migrants by age, sex, race, and Hispanic origin.

---

[6] For a description of this method, see the following URL:
Hhttp://www.census.gov/popest/topics/methodology/2004_st_char_meth.htmlH

[7] In the process, when the number of exemptions for any age-sex-race-Hispanic origin category (which will be referred to as a cell) was low, the exemptions were combined with those of other cells in order to improve the robustness of the resulting migration rates. Subsequently, each of the individual groups that were combined to make up the aggregation was assigned the rate or proportion of the aggregated group.

[8] Technically, the quotient produced by this procedure is the probability of moving. However, we make the simplifying assumption that the migration-rate is the same as the probability of moving.

The second step in the process of estimating the net number of domestic migrants for each county was to distribute these estimates to the counties within each state. To do this, two ratios were calculated for each county from the original IRS data; (1) an out-migration ratio, or the number of exemptions associated with filers who moved out of a given county and into all other counties within the state to the total number of exemptions associated with filers who moved across county boundaries within the state, and (2) an in-migration ratio, or the number of exemptions associated with filers who moved into a given county from all other counties within the state to the total number of exemptions associated with filers who moved across county boundaries within the state. After this, these ratios were applied to the total number of domestic migrants crossing county borders within each state to arrive at the number of domestic within-state out-migrants and in-migrants for each county.

The third step in the process of estimating the net number of domestic migrants for each county was to distribute the total number of domestic migrants going out of and coming into each state to each county within the state.[9] This was done in a way that paralleled the distribution of the total number of domestic migrants crossing county borders within each state. To begin, two ratios were calculated for each county from the original IRS data; (1) an out-migration ratio, or the number of exemptions associated with filers who moved out of a given county and into all other *states* to the total number of exemptions associated with filers who moved out of the state, and (2) an in-migration ratio, or the number of exemptions associated with filers who moved into a given county from all other *states* to the total number of exemptions associated with filers who moved into the state. After this, these ratios were applied to the total number of domestic migrants exiting and entering the state, respectively, to arrive at the number of inter-state out-migrants and inter-state in-migrants for each county.

The final step in the process of estimating the net number of domestic migrants was simply to sum the numbers of domestic within-state out-migrants and inter-state out-migrants by county to get the total number of out migrants for each county, and to sum the numbers of domestic within-state in-migrants and inter-state in-migrants by county to get the total number of in migrants for each county.

6. Specification of Net Military Movement

The net movement of the military, both foreign and domestic, into each county for each year was estimated using data received directly from the armed forces and the Department of Defense. They provided yearly estimates of the station strength in each state from 2000 to 2004. The net-movement was then calculated as the difference in the station strength from one year to the next. Finally, the county, age, sex, race, and Hispanic origin distribution was assigned using those who reported being employed by the military in Census 2000.

The Preliminary GQ Population Estimates

[9] Developed during the production of state characteristics estimates.

Group Quarters (GQ) population change is estimated separately from the household population because of the unique character of this subpopulation and the ability to acquire direct data that reflects change in this population. The technique for estimating the GQ population for the vintage 2004 estimates started with the Census 2000 enumerated GQ population. As with the household population, the race breakdowns of the GQ data were converted from categories consistent with Census 2000 to categories consistent with 1990 through proportional allocation (see above).

Next, the state representatives who participate in the Federal–State Cooperative Program for Population Estimates (FSCPE) developed an independent list of GQ facilities in their state by county with the populations typically associated with them at the time of Census 2000 and annually from 2000 to 2004 from the sources available to them in their state. In turn, the Census Bureau calculated the implied change in the GQ population from the numbers provided by the FSCPE members. This change was then applied to the Census GQ base to come up with the estimate of the total GQ population in each county. Finally, these county totals were distributed by age, sex, race, and Hispanic origin using the distribution of the GQ population within each GQ type from the base GQ population.

The Final Population Estimates by Demographic Characteristic

The final steps in the production of the vintage 2004 county characteristics estimates consisted of 1) adding together the household and GQ population estimates for each year, 2) calculating the necessary proportions from the preliminary estimates and applying them to both the previously created county population estimates by age (0-64 and 65+) and the State estimates by age, sex, race, and Hispanic origin, and 3) converting the data from the 4 race categories consistent with the 1990 census to the categories consistent with Census 2000.

In combining the household and GQ population estimates, the only caveat that should be noted is that it was assumed that there was no change to the GQ population between April 1, 2000 and July 1, 2000.

The next step in the production of the county estimates by age, sex, race, and Hispanic origin was to control the preliminary resident county estimates by age, sex, race, and Hispanic origin to both the previously created resident county population estimates by age (0-64 and 65+) and the state resident estimates by age, sex, race, and Hispanic origin. This was done through a two-step iterative process.

In the first step, the preliminary county characteristic estimates were summed to the state level by age, sex, race, and Hispanic origin. Next, proportions were calculated by dividing each county-age-sex-race-Hispanic origin cell by the corresponding sum. Finally, these proportions were applied to the original state characteristics estimates in order to calculate an intermediate set of county estimates by age, sex, race, and Hispanic origin.

In the second step, a similar procedure was used to calculate new proportions from the transformed data so that the county age estimates (0 to 64 and 65+) could be distributed by age,

race, and Hispanic origin. First, the preliminary county characteristic estimates were summed by the appropriate age groups. Next, proportions were calculated by dividing each county-age-sex-race-Hispanic origin cell by the corresponding sum. Finally, these proportions were applied to the original county age estimates in order to calculate an intermediate set of county estimates by age, sex, race, and Hispanic origin.

When attempting to make estimates consistent with two different sets of data using the procedure described above, performing the second step of the iteration tends to distort the fit achieved in the first step. Likewise, repeating the first step again tends to distort the fit achieved in the second step. However, these distortions can be minimized by repeating the two-step procedure multiple times. For this reason, the iterative procedure was repeated five times. After this, the resulting set of estimates was rounded to integers.

The final step in the estimates process was to convert the estimates from the 4-race categories consistent with the 1990 census in which the estimates were processed to the 31-race categories consistent with Census 2000. To do this, the procedure used to go from 31 to 4 races was essentially reversed. First, the proportion of each 4-race category associated with the 31 race categories was calculated for both universes using Census 2000 household and GQ data by county, age, sex, and Hispanic origin. Next, the GQ population for each estimate year was subtracted from the corresponding resident population in order to arrive at the household population for that estimate year. Then, the proportions for both the household and GQ data were applied to the respective 4-race estimates for each year to arrive at the 31-category race household and GQ estimates.[10] Finally, the 31-category household and GQ estimates by age, sex, and Hispanic origin were added together to arrive at the final resident county population estimates by age, sex, race, and Hispanic origin.

[10] The reason for separating the process of converting the household and GQ data from 4 to 31 races is that the GQ data are processed by GQ type so that they may be used in estimates not described in this document.