

ESTIMATES AND PROJECTIONS AREA METHODOLOGY

STATE POPULATION ESTIMATES BY AGE, SEX, RACE, AND HISPANIC ORIGIN FOR JULY 1, 2002

BACKGROUND

The U.S. Census Bureau produces estimates of the resident population by age, sex, race and Hispanic origin for each state in the United States on an annual basis. The following documentation outlines the methodology that was used in the production of the July 1, 2002 resident population estimates by age, sex, race, and Hispanic origin for the 50 states in the United States and the District of Columbia.

OVERVIEW

The Census Bureau develops state population estimates with a demographic procedure called a cohort-component method. This method follows each birth cohort across time according to its exposure to mortality, fertility, and migration. In order to apply the model, Census Bureau demographers estimate each component of population change separately. For the population residing in households the components of population change are births, deaths, and net migration, including net international migration. For the non-household population, change is represented by the net change in the population living in group-quarters facilities. A more detailed discussion of the methodology is provided below.

METHOD

The cohort-component method is based on the traditional demographic accounting system. Starting with a base population, deaths are subtracted from the population and births are added to the population, forming new cohorts. Estimates of net international migration and net internal migration are added to or subtracted from the population. The components of change are measured separately by age, sex, race, and Hispanic origin for each state and added to the base population as follows:

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

Where:

- P_1 = population at the end of the period
- P_0 = population at the beginning of the period
- B = births during the period
- D = deaths during the period
- NDM = net internal migration during the period
- NIM = net international migration during the period

In the process of developing the July 1, 2002 estimates, revised estimates of the July 1, 2001, and July 1, 2000 state population with demographic detail were produced. The revised estimate for 2001 and 2000 incorporates actual data for the demographic components that were not previously available and includes updates or corrections to the data previously used. In cases where we do not have data for all states for the current estimate year (2002), we estimate the components of population change based on one or more simplifying assumptions.

One of the guiding principles in the Census Bureau's subnational methodology is that all of our population estimates are consistent. This means that the sum of the state estimates must be

equal to the independently produced national population estimates. This consistency is required for all demographic characteristics produced. While this consistency is essential in the production and interpretation of the population estimates it does add an additional layer of complexity to their development.

The methodology used to produce the July 1, 2002 estimates is described next.

STEP 1: SPECIFICATION OF THE BASE POPULATIONS

The enumerated resident population in Census 2000 is the base for the post-2000 population estimates. The enumerated population was modified in two ways for purposes of developing these estimates. First, the race data were modified to eliminate the "Some other race" category in order to be more consistent with race categories that appear on the administrative records used to produce the population estimates. Second, the April 1, 2000 population estimates base reflects modifications to the Census 2000 population as documented in the Count Question Resolution program.

The race modification conforms to the Office of Management and Budget's (OMB) 1997 revised standards for collecting and presenting data on race and ethnicity. The revised OMB standards identified five minimum race categories: White; Black or African American; American Indian and Alaska Native; Asian; and, Native Hawaiian and Other Pacific Islander. Additionally, the OMB recommended that respondents be given the option of marking or selecting one or more races to indicate their racial identity. Finally, for respondents unable to identify with any of the five race categories, the OMB approved including a sixth category - "Some other race" - on the Census 2000 questionnaire.

No modification was necessary for responses indicating only an OMB race alone or in combination with another OMB race. However at the national level, about 18.5 million people checked "Some other race" alone or in combination with another race. These people were primarily of Hispanic origin and many wrote in their Hispanic origin or Hispanic origin type (such as Mexican or Puerto Rican) as their race. For purposes of estimates production, responses of "Some other race" alone were modified by blanking the "Some other race" response and imputing an OMB race alone or in combination with another race response. The responses were imputed from a donor, who matched on response to the question on Hispanic origin. Responses of both "Some other race" and an OMB race were modified by blanking the "Some other race" response and keeping the OMB race response.

The resulting race categories (White; Black; American Indian and Alaska Native; Asian; and, Native Hawaiian and Other Pacific Islander) conform with OMB's 1997 revised standards for the collection of data on race and ethnicity and are more consistent with the race categories in other administrative sources, such as vital statistics.

Because the group quarters (GQ) population experiences somewhat different demographic processes, the first step in the estimates process is to subtract the GQ population from the Census 2000 resident population to develop a base population that consists of two pieces (the household population and the population residing in group quarters). Both pieces of the base population contain full demographic detail (age, sex, race, and Hispanic origin) for each state and the District of Columbia.

1A. Base Household Population

1A.i. The Census 2000 household population (obtained by subtracting the GQ population from the resident population) is the starting point for the July 1, 2002 state

population estimates. The inclusion of demographic detail in the development of the state population estimates adds an additional layer of complexity to the estimation method. Although the Census 2000 population data were available for the full set of race categories described above, the administrative data sources used to estimate the components of change (births, deaths, and migration) were not available for all 31 races. Because the administrative data were available only in the 4 race categories consistent with the 1990 census (White; Black; American Indian, Eskimo, Aleut; Asian and Pacific Islander), the Census 2000 base household population was converted from the 31 race categories to the four race groups consistent with the 1990 Census. Then the July 1, 2002 population estimates were produced for the four race categories consistent with the 1990 Census. Finally, the July 1, 2002 estimates were converted to 31 races to be consistent with Census 2000 (see Step 6).

The conversion of Census 2000 categories to 1990 Census categories was based on a "straight proportional allocation." This uses the simplifying assumption that multiple race responses in Census 2000 would be evenly distributed between the comparable single race responses allowed in the 1990 Census. For example, the Census 2000 population in the three race categories of "White Alone," "Black Alone," and "White and Black" were converted into two of the 1990 Census race categories, "White" and "Black". The entire White Alone population enumerated in Census 2000 was assigned to White category and the entire Black Alone population was assigned to the Black category. Based on the assumption of straight proportion allocation, half of the Census 2000 "White and Black" population were assigned to the White race category and half of the population were assigned to the Black race category. These assignments are done at the state level, by age groups, sex, and Hispanic origin.

The assumption of proportional allocation is the best available assumption at this time, though future estimates may not require this conversion or may be based on different distributions to the single races. See Step 6 for the conversion of the July 1, 2002 estimates for four races back to the 31 race categories.

1A.ii. Because the Census 2000 reference date is April 1, 2000 and the estimate periods are July 1 to June 30, it was first necessary to develop a July 1, 2000 base population. This base population was calculated using the July 1, 2000 national estimates with full demographic detail (by age, sex, race, and Hispanic origin) and July 1, 2000 state estimates by age and sex. A ratio method was used to calculate July 1, 2000 state population estimates by age, sex, race, and Hispanic origin. This method applied the age, sex, race, and Hispanic origin distribution for states from Census 2000 to the July 1, 2000 state population estimates by age and sex to develop initial July 1, 2000 estimates with demographic detail. The national estimates by age, sex, race, and Hispanic origin were applied as controls to the initial estimates to generate the July 1, 2000 set of base population estimates that sum to equal the July 1, 2000 national population estimates by age, sex, race, and Hispanic origin and the state populations by age and sex.

The July 1, 2000 estimates then serve as the base population for the July 1, 2001 and July 1, 2002 estimates produced using the cohort-component method. The July 1, 2000 estimates could not be calculated using the cohort-component method because the administrative records used in the cohort-component method are available for calendar years and not the three-month time period from April 1, 2000 to July 1, 2000.

1B. Base Group-quarters Population

Examples of types of group-quarters (GQ) populations are: military personnel living in barracks, college students living in dormitories, and persons residing in institutions. Inmates of correctional facilities, persons in health care facilities, persons in Job Corps

Centers, and persons residing in nursing homes are also included in this category.

The Census 2000 group quarters population (obtained by subtracting the household population from the resident population) is the starting point for the July 1, 2002 state population estimates. First, the Census 2000 GQ data for the 31 race groups were converted to the four race groups consistent with the 1990 Census as described above.

STEP 2: SPECIFICATION OF BIRTHS AND DEATHS (VITAL STATISTICS) COMPONENTS

2A. The birth and death components are calculated from three sources of data. Files containing all registered births and deaths that occurred to U.S. residents during the estimate period are obtained from the National Center for Health Statistics (NCHS). The birth files contain the total numbers of birth in a calendar year by state and county of mothers' residence, sex, race, and Hispanic origin. The NCHS death files contain the total numbers of deaths by sex, race, Hispanic origin, age at death, and state and county of residence at death. The Federal State Cooperative Program for Population Estimates (FSCPE) also report annual numbers of registered births and deaths by sex, race, Hispanic origin, age at death, and state and county of residence at death or county of mothers residence at birth. A reconciliation process occurs between the NCHS and FSCPE vital statistics. In general, we believe that the total demographic characteristics distribution of data from the NCHS file is more accurate due to its national coverage, while the geographic distribution of data from FSCPE files is more accurate due to more specific local knowledge.

It is assumed that the vital statistics files represent complete counts of births and deaths for the resident population. No adjustments are made for undercoverage or differential coverage by states, age, race, or Hispanic origin.

2B. After the NCHS and FSCPE figures are reconciled, they are controlled to the national estimates of the numbers of births and deaths by sex, race, Hispanic origin, and age at death developed as part of the national population estimates for the same time period.

2C. Finally, the births are added to the base population for each year (July 1, 2000 and July 1, 2001) and the deaths are subtracted from the base population.

STEP 3: SPECIFICATION OF NET INTERNATIONAL MIGRATION

We estimate the net international migration to/from the United States as several sub-components: net foreign-born international migration, net movement to/from Puerto Rico, net federal and civilian citizen movement, and native emigration. In this last vintage, we did not have current state-level data in the detail we needed to be able to directly estimate net international migration at the state-level by characteristics. Instead, for each of the sub-components of net international migration we used the national characteristics net international migration data and distributed by characteristics to the states consistent with the state total population estimates.

STEP 4: SPECIFICATION OF NET INTERNAL MIGRATION

Step 4A. Match of Tax Returns to create counts of exemptions (filers and dependents) who migrate by demographic characteristics

4A.i. For the July 1, 2002 estimates the component of internal migration was developed using data from two administrative record sources: annual extracts of tax returns provided by the Internal Revenue Service (IRS) linked by Social Security Number across successive years; and the Census Numident file, derived from the Social Security Administration 100 percent file (SSA). In order to ensure confidentiality and privacy, these data sets are matched by SSN/PIK (Protected Identification Key) and are referred to jointly as IRS-SSA data. The IRS 1040 tax return records were matched to the SSA data to identify the age, sex, race, and Hispanic origin of the tax filers. A number of assumptions were made to assign demographic characteristics to spouses and dependents. Exemptions claimed for children were assigned to the under 20 age group and exemptions claimed for parents were assigned to the age category 65 and over. Sex was assigned randomly for exemptions. Spouses were assigned the same age and the opposite sex as filers. All spouses and exemptions were assigned the same race and Hispanic origin as filers.

4A.ii. After the demographic characteristics are added to the IRS tax return records, two years of records are matched by SSN/PIK to determine migration status. Filers (and their dependents) with a change in the state of residence between the two periods were identified as "Inter-State" migrants. Otherwise, if there was no change in the state of residence, the filers (and dependents) were identified as non-migrants.

Step 4B. Calculate State Out-Migration Rates and Number of State Out-Migrants
Migration rates are computed using the number of exemptions with addresses in different states in the second period as the numerator and the total number of exemptions in the state in the first period as the denominator. The rates are applied to the July 1, 2000 and July 1, 2001 base populations by age, sex, race, and Hispanic origin to generate a the number of state out migrants by age, sex, race, and Hispanic origin.

Because of the potentially large number of origin-characteristic combinations, a few simplifying assumptions were required in the production of the July 1, 2002 estimates. It was necessary in some cases to combine individual origin-characteristic categories (which will be referred to as cells) to improve the robustness of the data. If a given cell had less than 30 exemptions, then it was combined with adjacent age cells within the same origin-ethnicity-race-sex group until the combined category contained at least 30 exemptions. If it was not possible to create a combined category containing at least 30 exemptions within an origin-ethnicity-race-sex group, then cells were combined for both sexes. When individual ages were combined to compute a migration probability, each of the ages was assigned the probability for the aggregated age group.

Step 4C. Calculate In-Migration Proportion and In-Migrant Population
From the matched records the destinations of migrants by demographic characteristics can be determined. The numbers of out-migrants calculated from Step 4b were distributed as in-migrants to states by applying proportions of total in-migrant exemptions who moved to each state. These proportions were calculated by age, sex, race, and Hispanic origin. The numbers of out-migrants by state and characteristics were subtracted from the base population for each time period and the numbers of in-migrants were added to the base population.

STEP 5: PROCESSING OF GROUP QUARTERS POPULATION

GQ population change was estimated separately from the demographic accounting procedure described above. This was done primarily because of the uniqueness of this subpopulation and the special difficulties of estimating the GQ population.

The July 1, 2002 GQ estimates were calculated starting with the Census 2000 GQ population by

age, sex, race, Hispanic origin, and seven GQ types for each states. States provide updated information on the total GQ population by GQ type to the Census Bureau each year.¹ The Census 2000 age, sex, race, and Hispanic origin distributions of the GQ population by state and type were applied to the July 1, 2002 GQ populations by type reported by states to produce GQ population estimates for states with demographic detail.

STEP 6: GENERATE RESIDENT POPULATION ESTIMATES BY DEMOGRAPHIC CHARACTERISTIC

6A. Prior to combining the July 1, 2000 (revised), July 1, 2001 (revised) and July 1, 2002 group quarters and household population estimates, each set of estimates was converted from the four race groups consistent with the 1990 Census to the 31 race groups consistent with Census 2000 by applying conversion factors. Continuing the example from Step 1A.i., the estimated "White" population was apportioned to the "White Alone" and "White and Black" populations. The estimated July 1, 2002 White population was multiplied by the ratio of the White Alone population from Census 2000 to the sum of the White Alone and half of the "White and Black" population from Census 2000 to produce the July 1, 2002 estimate for the White Alone population. The estimated July 1, 2002 White population was multiplied by the ratio of the "White and Black" population from Census 2000 to the sum of the White Alone and half of the "White and Black" population from Census 2000 to produce part of the July 1, 2002 estimates for the White and Black population. The remaining part of the July 1, 2002 "White and Black" population was obtained by applying comparable ratios to the July 1, 2002 estimates Black population.

6B. The group quarters and household estimates of the population were summed by state and by age group and sex within states. The preliminary estimates were compared with independently calculated state total population estimates, state estimates by age and sex, and national estimates by age, sex, race and Hispanic origin. The final resident population estimates were adjusted to equal the independent totals by multiplying each estimate by the ratio of the independent totals to the sum of the relevant estimates. These adjusted estimates were rounded to whole numbers for each combination of demographic characteristics within states and compared with the independent totals.

¹ This data is collected annually by state agencies through the Federal-State Cooperative Program for Population Estimates (FSCPE).