

Estimates of the Population of Counties
by Age, Sex, Race and Hispanic Origin: 1990 to 1999

Source: U.S. Census Bureau
Internet Release date: August 30, 2000

These data are estimates of the resident population of the counties in the United States, by single year of age (0, 1, 2, ..., 84, 85 and over), sex (male, female), race (White; Black; American Indian and Alaska Native; Asian and Pacific Islander) for July 1 of each year from 1990 to 1999. In the discussion below, these data are referred to as the county detail estimates.

The county estimates included in this release are developed in a two-step procedure. First a set of state estimates by age, sex, race, and Hispanic origin (with the same categories as given above) are developed. These state estimates are developed using a cohort-component technique. A complete description of that methodology is included as Appendix A. The county detail estimates are produced in the second step using a ratio method. The ratio method is a mathematical technique for adjusting data to sum to a pre-determined total. It consists of multiplying each element of the data by the ratio formed by dividing the desired total by the sum of the data. When there are multiple totals to which we wish to adjust our data, as with the county estimates, we first partition the data into groups which correspond to the desired totals, then construct and apply ratios for each group using the same method as in the single-total situation. Applying the ratio method to a data set is referred to as raking.

The state detail estimates developed in step 1 along with the estimates of the total population of counties serve as the control totals for the county detail estimates. The methodology used to develop the estimates of the total population of counties is described in "Subnational Estimates of Total Population by the Tax Return Methodology" by Michael Batutis (available from the Population Division, US Bureau of the Census).

The April 1, 1990 modified census counts for counties by age, sex, race, and Hispanic origin released as computer tape file MARS, STF-S-3 are used as the starting point for each year of county detail estimates produced. These county-level MARS data are raked to the state estimates developed in step 1 and the county total population estimates for a given year to produce the county detail estimates for that year. An additional refinement has been added to this methodology, which is the separate estimation of the group quarters (GQ) population. The GQ population is those individuals residing in non-standard living arrangements such as correctional institutions and nursing homes. Their demographic characteristics are often very different from the rest of the population of the county in which they reside, which is why it is useful to estimate them separately. Estimates of the GQ population are drawn from a base of 1990 census data which is updated annually with data submitted by the members of the Federal State Cooperative Program for Population Estimates. These estimates are subtracted from our starting population and controls before performing the raking described above, so that the result of this process is non-GQ county detail estimates. The GQ and non-GQ county detail estimates are combined to produce the estimates presented here.

These data were developed as part of an ongoing project to develop postcensal population estimates of counties by age, sex, race, and Hispanic origin. Though the method we employed produces estimates which are fully

disaggregated with respect to these characteristics, the limitations of our methodology are such that we do not consider these data to be accurate for each individual cell. Although we do not have measures of error, we believe that aggregating the individual cells to larger groups reduces the level of error. Consequently, we offer the data here in three different formats: 1) complete age and sex detail with no race or Hispanic origin information; 2) complete race and Hispanic origin detail with no age or sex information; 3) 5-year age categories by sex with limited race/Hispanic origin categories (total Hispanic; White non-Hispanic; White Hispanic; total Black; total American Indian and Alaska Native; total Asian and Pacific Islander). Work is continuing on developing methods and data sets that can be used to more directly and accurately estimate the age, sex, race, and Hispanic origin distributions of county populations.

Technical Contact: Larry Sink
 Population Division
 301-457-2461

Appendix A

Estimates of the Population of States by Age, Sex, Race and Hispanic Origin: 1990 to 1999

These data are estimates of the resident population of the 50 States and the District of Columbia by single year of age (0, 1, 2, ..., 84, 85 and over), sex (male, female), race (White; Black; American Indian and Alaska Native; Asian and Pacific Islander) and Hispanic origin (Hispanic origin, non-Hispanic origin) for July 1, 1990 through 1999.

The State estimates included in this release are developed using a cohort-component method whereby each component of population change - births, deaths, domestic migration, and international migration is estimated separately for each birth cohort by sex, race, and Hispanic origin.

The cohort-component method is based on the traditional demographic accounting system:

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

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
NMA = net migration from abroad during the period

To generate population estimates with this model, we first developed separate data sets for each of these components. The procedures by which these data are developed by single year of age, sex, race, and Hispanic origin are described in the following sections. Once the data for each component were developed, the estimates could be produced simply by adding the components together, with the exception of internal migration. The reason for this exception and our procedure for dealing with internal migration are explained in the internal migration section below.

This overall approach is similar to that used in the development of the experimental set of state and metropolitan area estimates by race and, Hispanic origin. The Current Population Reports, Series P-25, No. 1040-RD-1, provides a detailed discussion of this general approach.

Starting Population

The April 1, 1990 Census data files which were used as the starting points in this methodology represent the modified age, race, sex, and Hispanic origin (MARS) census data released as computer tape file (MARS, STF-S-3). The modification methodology is outlined in Census Report, CPH-L-74. In order to develop the desired July 1, 1990 starting point from the April 1, 1990 data, we used the ratio method to make the April 1 data consistent with the July 1, 1990 national population estimates by age, race, sex, and Hispanic origin and consistent with the July 1, 1990 state population estimates by age and sex. The ratio method is a technique for adjusting data to sum to a pre-determined total, which consists of multiplying each element of the data by the ratio formed by dividing the desired total by the sum of the data. When there are multiple totals to which we wish to adjust our data, as with the state age-sex estimates, we first partition the data into groups which correspond to the desired totals, then construct and apply ratios for each group using the same method as in the single-total situation. Applying the ratio method to a data set is referred to as raking.

Vital Statistics

The data for births and deaths used in these estimates are based on 1) detailed data available from the National Center for Health Statistics (NCHS); 2) estimates of births and deaths for counties developed by the member agencies of the Federal State Cooperative Program for Population Estimates (FSCPPE); and 3) estimates of births and deaths by detailed demographic characteristics developed as part of the Population Division program for national population estimates.

Births-- Extracts of detailed individual-record data on births from NCHS for calendar years 1990 through 1997 are used to construct the state-level births. The race and Hispanic origin codes on the individual NCHS records were converted into our four race and two-Hispanic origin system as shown on the attached chart. The individual records for events occurring July 1, 1990 through June 30, 1997 are aggregated by year to the county level and adjusted to the county-level data provided by the FSCPPE member agencies. These results are further adjusted to agree with national-level race-Hispanic origin estimates developed as part of the national population estimates program (see ref. 4, p5). The data are then aggregated to the state level.

To estimate the July 1, 1997 through June 30, 1998 and July 1, 1998 through June 30, 1999 periods, we first aggregated the individual records for events occurring for the six month period July 1, 1997 through December 31, 1997. Because the 1998 and 1999 data were not available when we developed the original set of estimates, we used an alternative method to estimate the detailed data for the January 1 to June 30, 1998 and July 1, 1998 to June 30, 1999 periods. To do this, we adjusted the six month aggregations for July through December 1997 to: 1) preliminary estimates of births for July 1997 through June 1998 provided by the FSCPPE agencies; and 2) to preliminary estimates of national level births by race, sex and Hispanic origin for the July 1997 through June 1998 period, developed

as part of the national estimates program (see ref. 4, p5). Through this process, we obtained estimates of the July 1, 1997 through June 30, 1998 births, which were then adjusted to provide estimates of the July 1, 1998 through June 30, 1999 births in a similar fashion.

Deaths-- The estimates of deaths are developed in the same manner as that for births, except that the death data have the additional dimension of age. To develop the deaths by age, we examined the age at death and the date of birth information on the individual record. If the date of birth information was missing, age was set to the most recent valid value. The preliminary age at death value was computed using date of birth information. This preliminary value was compared to the age at death value on the individual record. If the difference between these two values was no greater than two years, the computed value was used. If not, the age at death value on the certificate was used.

Internal Migration

The values for internal migration used in these estimates are developed using a variant of the basic administrative records method. The development of the data rely upon two basic files - an annual extract of tax returns provided by the Internal Revenue Service (IRS), and a 20% sample of information on the Social Security Administration Application File (NUMIDENT) which includes Social Security Number (SSN), month and year of birth, race, sex, and 6 characters of the last name.

The basic administrative records method relies upon annual extracts of tax returns provided by the IRS. In this approach, using the SSN on the return, we are able to match the tax returns for two years and obtain state of residence for the two periods. By comparing the state of residence at the two points in time, we are able to develop annual measures of migration for states.

Because the standard tax return provides no demographic characteristics of the tax filer, the basic administrative records method provides data for the total population only. To obtain demographic characteristics, we rely upon an extract of the NUMIDENT file, which is merged with the tax returns file by SSN. Because the Census Bureau is able to receive only a 20-percent sample of this basic NUMIDENT file, we can only append the demographic characteristics of the primary filer to the same 20-percent sample of tax returns.

In addition to demographic characteristics of the primary filers, the model requires demographic characteristics of those persons claimed as exemptions on the tax return. The rules for assigning demographic characteristics to dependents are straightforward and rely on basic familial and demographic relationships.

1. Spouses on the tax return are given the age and race/Hispanic origin of the primary filer. They are assigned the opposite sex of the primary filer.
2. Dependent children are given the race/Hispanic origin of the primary filer and all assigned to the age group under 20.
3. Parent exemptions are assigned the race/Hispanic origin of the primary filer and all assigned to the age group 65 and over.
4. Other dependents are assigned the race/Hispanic origin of the

primary filer and all assigned the age group under 20.

In order to develop an estimate for July 1 of a given year using the cohort-component method, we need an estimate of the migration which took place between July 1 of the preceding year and June 30 of the year in question. However, the migration data we obtain using the administrative records method pertain to time periods determined by when the individual tax-payers file their returns, which, of course, varies from tax-payer to tax-payer. Since most tax returns are filed between January and April, it is roughly correct to say, for example, that the migration data obtained as a result of matching returns from tax years 1989 and 1990 pertains to a (one-year) period within the interval from January 1990 (the earliest most tax-payers would file their 1989 returns) to April 1991 (the latest most tax-payers would file their 1990 returns). We have assumed that this is a reasonable approximation to the interval needed for our 1991 estimates and, similarly, that the data from the tax years 1990-1991 through 1997-1998 match are appropriate for our 1992 through 1999 estimates, respectively. These assumptions are based on our research which indicates that state-to-state migration rates change on average only about 15% a year, and these changes tend to offset one another when out-rates and in-proportions are calculated.

The migration data yielded by the method described above consists of counts of those tax-filers whose SSNs were in the 20% sample plus the dependents claimed by those filers, disaggregated by state of origin, state of destination, age, race, sex, and, Hispanic origin. The first step in converting these counts into the statistics actually used in our estimates is to construct state-to-state migration rates by demographic characteristic (i.e. age, race, sex, and, Hispanic origin). This is done by summing all the counts for a given origin by characteristic and then dividing each count by the sum for that characteristic.

Because of the potentially large number of origin-destination-characteristic combinations, it is necessary in some cases to combine individual origin-destination-characteristic categories (which will be referred to as cells) to avoid stretching the data too thin. If a given cell has less than 30 individuals, then it is combined with adjacent age cells within the same origin-destination-ethnicity-race-sex group until the combined category contains at least 30 individuals. If it is not possible to create a combined category containing at least 30 individuals within an origin-destination- ethnicity-race-sex group, then cells are combined across sex. Because we do not know the ages of persons claimed as dependents, we have assumed that persons claimed as dependent children are under 20 and persons claimed as dependent parents are 65 or over. These assumptions require us to use under 20 and 65 and over as age categories when computing migration rates. When individual ages are combined to compute a migration rate, each of the ages is assigned that rate. Once the initial rates have been computed, they are smoothed across ages using a five-year moving average to reduce the impact of random error.

An additional problem is created by the fact that data collected by the SSA prior to 1980 have only three race categories: White, Black, and Other. In order to convert from this system into the four race system used in these estimates, it is necessary to split the "Other" category into American Indian and Alaska Native (AIAN) and Asian & Pacific Islander (API). This split is based on the relative sizes of the total, AIAN and API populations in the origin state for the out-rates and on the relative sizes in the destination state for the in-proportions. The racial composition of migration flows depends upon the racial composition of both the origin and the destination, so that in reality this "Other" group

probably has a different composition for each of the 2550 different state-to-state flows, but the numbers involved are too small to permit separate analysis for each flow. By combining the state-to-state rates into out-rates and in-proportions, with the method described below, we greatly increase the number of observations underlying each of our statistics and have the ability to base our rate calculations solely on origin characteristics and our proportions solely on destination characteristics. These separate rates and proportions for the four race groups were applied only to the non-Hispanic population. One set (by age and sex) of migration rates and proportions are computed for Hispanics without regard to race and applied to all Hispanic race groups.

The creation of out-rates and in-proportions from the state-to-state migration rates involves converting the origin-destination-characteristic-specific rates into origin-characteristic-specific rates and destination-characteristic-specific proportions. In this process, all calculations are performed separately for each combination of demographic characteristics. The state-to-state migration rates are multiplied by our starting population estimate for the appropriate group to obtain an estimate of the total migration flow between the states in question for this group. These flows are summed to get total out- and total in-migration by characteristic for each state. Each state's out-migration totals are divided by their respective populations to obtain out-migration rates and the out-migration totals are summed across states to obtain the national total of migration by demographic characteristic. These national totals are divided into each state's in-migration totals to obtain the in-migration proportions. The population figures used in these calculations are a preliminary set of population estimates which are produced using the ratio method. The out-rates and in-proportions are converted into actual estimates of migration within the process which produces the finished population estimates, since it is only at this point that we have the final population estimates needed for this calculation. This conversion is accomplished by multiplying each state's out-rates by the respective beginning-of-period population to obtain our estimate of that state's out-migration, which is then summed across states to obtain national-level migration. Finally, each state's in-proportions are multiplied by the national-level migration to get that state's in-migration estimates.

International Migration

The international migration component in these estimates is an aggregation of four separate parts: 1) alien immigration, refugees, and net undocumented migration; 2) legal emigrants; 3) net movement between Puerto Rico and the mainland; and 4) net movement of federal civilian citizens.

Immigration (including refugees and undocumented). We utilized legal immigration data developed from the Immigration and Naturalization Service public use microdata, refugee data drawn from unpublished reports of the Office of Refugee Resettlement, and net undocumented immigration files developed as part of the national estimates program (see ref. 5, pp24-29). These files all possess full demographic detail and state-level geography for all years in question.

Legal emigration. We utilized emigration data developed as a part of the national estimates program, as described in "U.S. Population Estimates by Age, Sex, Race, and Hispanic Origin: 1990 to 1996", U.S. Bureau of the Census PPL-57. These data possess full demographic detail and state-level geography.

Net Puerto Rican migration. We utilized a national-level file on net Puerto Rican migration with full demographic detail developed as part of the national estimates program (see ref. 3, p6). This net migration was distributed to the states based on their respective portions of the Puerto Rican migration developed from past research (see ref. 6).

Net federal citizen migration. We utilized a national-level file on net federal citizen migration with full demographic detail developed as part of the national estimates program (see ref. 3, pp6-7). State-level distributions were obtained using the IRS-SSA data employed in the internal migration estimates, which also contains data on movements to and from foreign countries. The Other races distribution was used for both, AIAN and API, and the below 20 and above 65 age distributions were taken from the national distribution. These state-level distributions were raked to the national-level distribution to yield the final data.

Consistency with other Census Bureau Estimates

The Census Bureau annually produces estimates of the population for various levels of geography and demographic detail. The estimates produced in this annual production cycle are collectively referred to as a round of estimates, and the current round of estimates is referred to as the 1999 round of estimates, since 1999 is the most recent year for which an estimate is included. All estimates produced in a given round are consistent with one another, which means that, for example, all of the estimates in the 1999 round will sum to the same national total for each estimate year and all of the state and sub-state estimates will sum to the same state totals. However, the estimates produced in a given round are not necessarily consistent with the estimates produced in previous rounds, and since each round produces estimates for each year from 1990 forward, each new round will contain estimates for each of the years in the previous round. What this means is that each new round of estimates replaces each year of the estimates in the previous round with a new set of estimates for that year which may not be consistent with the estimates in the previous round. For example, the current round of state and county detail estimates contains estimates for 1998 which in some cases differ substantially from the estimates for 1998 released in the 1998 round of estimates.

The reason that estimates for a given year can change significantly from one round to the next is that in each round the estimates for the latest year have been prepared using provisional data, since the final data for that year are not available at production time. When estimates for that year are prepared in subsequent rounds, the provisional data are replaced with final data, which can sometimes produce noticeable changes. Additionally, methodological changes are sometimes introduced between rounds, and these can affect every year in a round of estimates.

Limitations

These data were developed as part of an ongoing project to develop postcensal population estimates of states and counties by age, sex, race, and, Hispanic origin. These estimates represent an intermediate step in this overall project. Work is continuing on methods and data sets that can be used to more directly estimate the age, sex, race, and, Hispanic origin distributions of the state and county populations. As additional steps are completed, we plan to prepare new estimates for subsequent years and revise the existing series back to 1990.

This data set contains population estimates disaggregated by single year of age, sex, race, and, Hispanic origin for each state. However, the limitations of our methodology are such that we do not consider these data to be accurate for each individual cell. Although we do not have measures of error, we believe that aggregating the individual cells to larger groups will reduce the level of error. We include the separate data for your convenience in aggregating to various groups. Although the data included in this data set are unrounded, we do not consider these data to be accurate to the last digit.

Technical Contact: Larry Sink
 Population Division
 (301) 457-2461

References

1. Batutis, Michael J., "Subnational Estimates of Total Population by the Tax Return Methodology", Population Division, U.S. Bureau of the Census, Washington, DC, 1994.
2. Campbell, Paul R., Population Projections for States, by Age, Race, and Sex: 1993 to 2020, U.S. Bureau of the Census, Current Population Reports, P25-1111, U.S. Government Printing Office, Washington, DC, 1994.
3. Deardorff, Kevin E., Frederick W. Hollmann, and Patricia Montgomery, "U.S. Population Estimates by Age, Sex, Race, and Hispanic Origin: 1990 to 1994", U.S. Bureau of the Census, PPL-21, 1995.
4. Hollmann, Frederick W. United States Population Estimates, by Age, Sex, Race, and Hispanic Origin: 1980 to 1988, U.S. Bureau of the Census, Current Population Reports, Series P-25, No.1045, U.S. Government Printing Office, Washington, DC, 1990.
5. . "U.S. Population Estimates, by Age, Sex, Race, and Hispanic Origin: 1990 to 1993", U.S. Bureau of the Census, PPL-8, 1994.
6. Word, David L. "The Census Bureau Approach for Allocating Internal Migration to States, Counties and Places:1981-1991", U.S. Bureau of the Census, Technical Working Paper No. 1, 1992.

Conversion of National Center for Health Statistics (NCHS)
 Race, Ethnicity, and Age for State Estimates

RACE

NCHS	State Estimates
(1) White-----	White
(2) Black-----	Black
(3) American Indian or Alaska Native-----	American Indian or Alaska Native
(4) Chinese -----	
(5) Japanese	
(6) Hawaiian -----	Asian and Pacific Islander
(7) Filipino	
(8) Other API	

(9) Other Race-----

ETHNICITY

NCHS	State Estimates
(00) Non-Hispanic-----	Non-Hispanic
(01) Mexican -----	
(02) Puerto Rican	
(03) Cuban -----	Hispanic
(04) Central or South American	
(05) Other Hispanic -----	
(99) Unknown, not asked-----	allocated according to proportion Hispanic for appropriate sub-group in MARS file