Methodology for the United States Resident Population Estimates by Age, Sex, Race, and Hispanic Origin (Vintage 2009): April 1, 2000 to July 1, 2009

The U.S. Census Bureau produces monthly estimates of the resident population for the United States on an annual basis. We revise the estimates time series each year as final input data become available. These postcensal estimates from April 1, 2000 through July 1, 2009 supersede all previous estimates produced since Census 2000. For each month in the time series, the total national resident population is obtained by summing the estimates across all age, sex, race, and Hispanic origin categories. The following documentation describes the process by which we produce the April 1, 2000 through July 1, 2009 resident population estimates at the national level.

Population Universe

Estimates of the U.S. *resident population* include persons resident in the 50 states and the District of Columbia. These estimates exclude residents of the Commonwealth of Puerto Rico and residents of the Island areas under U.S. sovereignty or jurisdiction (principally American Samoa, Guam, United States Virgin Islands, and the Commonwealth of the Northern Mariana Islands). The definition of residence conforms to the criteria used in Census 2000, which defines a resident of a specified area as a person "usually resident" in that area. Estimates of the resident population exclude the U.S. Armed Forces overseas, as well as civilian U.S. citizens whose usual place of residence is outside the United States.

While this document focuses on the method we use to produce the U.S. resident population, we also produce and publish estimates of three other population universes: the *resident population plus Armed Forces overseas*, the *civilian population*, and the *civilian noninstitutionalized population*. In general, we calculate each of these populations by adding another population to or subtracting a subpopulation from the official resident population estimates.

Estimates of the *resident population plus Armed Forces overseas* include U.S. residents and members of the Armed Forces on active duty stationed outside the United States, but do not include military dependents and other U.S. citizens living abroad.

Civilian population estimates include only those U.S. residents not in the active duty Armed Forces. The difference between the resident population and the civilian population is the active duty Armed Forces population residing in the United States. The Defense Manpower Data Center (DMDC) and Census 2000 are the sources of the data on the active duty Armed Forces, reserves, and National Guard populations stationed abroad and within the United States.

Estimates of the *civilian noninstitutionalized population* exclude persons residing in institutions, primarily nursing homes, prisons and jails, mental hospitals, and juvenile facilities as well as the active duty Armed Forces population residing in the United States. To estimate this population universe, we use the DMDC data and data on the institutionalized population provided by members of the Federal-State Cooperative for Population Estimates (FSCPE), the Department of Veterans Affairs, the individual branches of service in the Department of Defense, and Census 2000.

Overview

We produce quarterly estimates of the U.S. population by age, sex, race, and Hispanic origin by updating the modified Census 2000 base population with data on the demographic components of population change. The basic procedure we use to estimate population is:

(1) base population
(2) plus births to U.S. resident women
(3) minus deaths to U.S. residents
(4) plus net international migrants

We calculate the population for each age, sex, race, and Hispanic origin group. We may also incorporate other changes due to corrections made since Census 2000. These corrections are the result of successful local challenges or special censuses.

Method

Base Population

The enumerated population from Census 2000 provides the starting point for these estimates and 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.
2. Additional modifications are made to the Census 2000 population counts as documented in the Count Question Resolution program and errata notes.[1]

We produce estimates for 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] However, 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, 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 race categories in two different ways. One group includes the five single-race categories and a sixth category that combines all categories with more than one race – referred to in our tables as "Two or More Races." The other group includes the five alone or in combination race groups. Each of the alone or in combination groups contains one of the single-race categories plus all the multiple-race categories that include

that single race. Alone or in combination groups do not sum to the 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.

Demographic Components of Population Change

Births

We estimate births to U.S. resident women based on data collected by the National Center for Health Statistics (NCHS).[4] NCHS ordinarily provides final data on births by month of birth, sex of child, age of mother, and race and Hispanic origin of both mother and father through the end of the calendar year two years prior to the vintage year of estimates (i.e., 2007 for Vintage 2009 estimates) and preliminary data on total births for the calendar year one year prior to the vintage year of estimates (i.e., 2008). For the Vintage 2009 estimates, NCHS was only able to provide final birth data through calendar year 2006 and provisional data for total births occurring in 2008. Instead of final 2007 birth data, NCHS provided control totals by state and preliminary detailed data on births occurring in 2007, which we proportionally adjusted to sum to the state control totals.

NCHS provides the birth certificate data in the 1977 OMB race categories of White; Black; American Indian, Eskimo or Aleut; and Asian or Pacific Islander, under the "mark one race" scenario. To produce post-2000 population estimates in the 1997 OMB race categories, we must model births by child's sex, race, and Hispanic origin to produce the full 31 possible single and multiple-race combinations.

Modeling the NCHS births by race and Hispanic origin is a three-step procedure. The first step is to produce a distribution of births by age of mother, sex, and race and Hispanic origin of both parents using 1997 OMB race categories. "Bridging" factors calculated by NCHS by age, sex, and county produce a formulation for a cross-classification of both the 1977 and 1997 OMB race taxonomies by county, age, sex, and Hispanic origin in the estimates base population.[5] We cross-classify the Census 2000 population by the two OMB race definitions, aggregate the distribution into geographic areas that are deemed to be distinct with respect to the cross-classification, and restrict the distribution to men age 20 to 49 years and women age 18 to 39 years. We use the resulting distribution to assign 1997 OMB joint race categories of mother and father in the birth file from the 1977 OMB joint categories provided by NCHS, and the assignment is carried out for every category of calendar month, age of mother, and sex of the birth.

The second step is to impute the distribution of births by race and Hispanic origin of child from the joint distribution by race and origin of both parents. We use information from Census 2000 on race and Hispanic origin reporting within family households for the population under 1 year of age and their parent(s) to impute the race and Hispanic origin composition of children for each parental joint category.

The third and final step is to benchmark the births by race and Hispanic origin to the composition of the estimates base population under 1 year of age by race and Hispanic origin. We calculate ratios of enumerated infants to births from April 1, 1999 to March 31, 2000, distributed according to steps 1 and 2. For most categories, these ratios are held constant and applied to each new cohort of births by race and Hispanic origin.[6]

To estimate the distribution of births for calendar year 2008, we distribute provisional 2008 births received from NCHS according to the 2007 births by age of mother, month of birth, sex, and modeled race and Hispanic origin.

To estimate the number and distribution of births by race and Hispanic origin of mother for the first half of 2009, we calculate calendar year, age-specific birth rates for women by race and Hispanic origin for 2008. We apply these rates to the projected populations of resident women by age, race, and Hispanic origin for each of the first two quarters of 2009 and then proportionally adjust for the observed seasonality in births by quarter in earlier years of birth data.

Deaths

We estimate deaths to U.S. residents based on data collected by NCHS. NCHS ordinarily provides final data on deaths by month of death, age, sex, race, and Hispanic origin through the end of the calendar year two years prior to the vintage year of estimates (i.e., 2007 for Vintage 2009 estimates) and preliminary data on total deaths for the calendar year one year prior to the vintage year of estimates (i.e., 2008) by Hispanic origin. For the Vintage 2009 estimates, NCHS was only able to provide final death data through calendar year 2006 and provisional data for total deaths occurring in 2008. Instead of final 2007 death data, NCHS provided control totals by state and preliminary detailed data on deaths occurring in 2007, which we proportionally adjusted to sum to the state control totals.

As with the birth certificate data, NCHS provides death certificate data in the 1977 OMB race categories of White; Black; American Indian, Eskimo or Aleut; and Asian or Pacific Islander, under the "mark one race" scenario. Therefore we model deaths by age, sex, race, and Hispanic origin to produce the full 31 possible single and multiple-race combinations.

To estimate the age, sex, race, and Hispanic origin distribution of deaths for calendar year 2007 and earlier years, we calculate separate death rates for the NCHS race categories by age, sex, and Hispanic origin using the 1998 deaths and 1998 population estimates from Vintage 2000, based on the 1990 Census.[7] For the non-Hispanic population, deaths are calculated by age-specific rates for aggregated race categories. We apply the death rates for the White; Black; and American Indian, Eskimo or Aleut groups to the corresponding White alone; Black alone; and American Indian and Alaska Native alone populations. We apply the Asian and Pacific Islander death rate to both the Asian alone population and the Native Hawaiian and Other Pacific Islander alone population. We estimate multiple-race deaths as the difference between total 2007 deaths reported by NCHS and the sum of deaths estimated for the single-race groups. We assume that age-specific death rates for the Hispanic population do not vary by race. From this first step in the process, we produce NCHS-based estimated deaths by quarter of occurrence, single years of

4

age (0-99, and 100+), sex, race (31 possible single and multiple-race combinations), and Hispanic origin.

Although this first step of the process produces the full age distribution of deaths, the data reflect the known discrepancies in the reporting of age between census counts and death registration, most notably in the oldest ages. To address this data issue, we utilize life table-based death rates to estimate deaths to the population ages 70 to 99 (single years of age) and 100 years and over, by sex, race, and Hispanic origin.[8] We apply these death rates to the population by age, sex, race, and Hispanic origin, and calendar quarter, as the population is being estimated, to obtain a quarterly matrix of deaths by age, sex, race, and Hispanic origin that would have occurred under the assumption that mortality levels did not vary by race or Hispanic origin. Then, we control these rate-based death estimates to the NCHS-based estimated deaths aggregated to quarterly deaths by age (single years of age from 0 to 69, and age 70 and older), sex, and the following combined race and Hispanic origin categories:

1. non-Hispanic White alone,
2. non-Hispanic Black alone,
3. non-Hispanic Asian alone and non-Hispanic Native Hawaiian and Other Pacific Islander alone,
4. non-Hispanic American Indian and Alaska Native alone and all non-Hispanic multiple races, and
5. Hispanic of all races.

From these final estimates of deaths and the estimates of the population, we derive implied central death rates and the quarterly proportion of total deaths in 2007. We use these implied rates to estimate preliminary deaths from January 1, 2008 to June 30, 2009 by quarter, as described above. We control the 2008 quarterly estimates to the NCHS provisional death total for 2008 and apply the quarterly proportions from 2007 to all six quarters from January 1, 2008 to June 30, 2009 to incorporate the known seasonal pattern of deaths.

Finally, we apply the NCHS distribution of total monthly deaths within each calendar year to the whole time series of full-detailed quarterly deaths to calculate the monthly deaths by age, sex, race, and Hispanic origin up to July 1, 2009.

Net International Migration

International migration, in its simplest form, is any change of residence across the borders of the United States (the 50 states and the District of Columbia). The net international migration component of the population estimates combines four parts: (a) net international migration of the foreign born, (b) net migration between the United States and Puerto Rico, (c) net migration of natives to and from the United States, and (d) net movement of the Armed Forces population to and from the United States.

Net international migration of the foreign-born population is estimated in two parts, immigration and emigration. The estimate of immigration utilizes information from the American Community Survey (ACS) on the reported residence of the foreign-born population in the prior year. The

foreign born who reported living abroad in the year prior to the survey are considered immigrants. Because this question is only asked of those age one and older, we make an additional assumption for foreign-born immigrants under the age of one. We assume that the number of foreign-born immigrants under the age of one is equal to half of the number of one-year-old foreign-born immigrants.

Emigration of the foreign born is estimated using a residual method. We age forward the foreign-born household population in Census 2000 using NCHS life tables to obtain the expected population in 2005, 2006, 2007, and 2008. We then compare the expected foreign-born population to the foreign-born population estimated by ACS 2005, ACS 2006, ACS 2007, and ACS 2008. Subtracting the estimated from the expected population produces a residual, which serves as the basis for emigration rates for the 2000 to 2005, 2000 to 2006, 2000 to 2007, and 2000 to 2008 time periods. We perform this calculation for two period-of-entry groups: the foreign born who entered the United States between 1990 and 1999; and the foreign born who entered before 1990. An average of the rates for each period-of-entry group are then applied to the population at risk of emigrating each year (i.e., the foreign-born population in the ACS who indicated that they lived in the United States one year ago) to obtain an annual estimate of emigration for 2000 to 2009. To estimate emigration from 2000 to 2008, we use an average of the rates developed from the 2000 to 2005, 2000 to 2006, and 2000 to 2007 residuals. To estimate emigration in 2009, we use an average of the rates developed from the 2000 to 2006, 2000 to 2007, and 2000 to 2008 residuals.

We estimate the net international migration of the foreign-born population by subtracting the number of emigrants from the number of immigrants. Age, sex, race and Hispanic origin information is estimated for foreign-born immigrants and emigrants separately using data from Census 2000 and the ACS three-year estimates for the 2005 to 2007 time period (ACS 2005-2007). Estimate year 2000 uses information from Census 2000, while 2005 and later years use information from the ACS 2005-2007. The incorporation of ACS data is phased in through linear interpolation between estimate years 2000 and 2005.

Foreign-born immigrants are assigned the age, sex, race, and Hispanic origin distribution of the foreign-born population who entered the United States within five years of the Census/survey year. Age is adjusted for foreign-born immigrants to represent age at arrival into the United States. Characteristics are applied to the estimates of foreign-born emigration by period of entry. The age, sex, race, and Hispanic origin distribution of the foreign born who entered the United States within ten years of the Census/survey year are applied to the estimate of emigrants who entered within ten years of the estimate year. The age, sex, race, and Hispanic origin of the foreign born who entered the United States more than ten years before the Census/survey year are applied to the estimate of emigrants who entered the United States more than ten years before the estimate year.

Net migration between the United States (the 50 states and the District of Columbia) and Puerto Rico is also estimated in two parts, immigration and emigration, when possible. For 2005 and later years, the ACS and the Puerto Rico Community Survey (PRCS) allow us to estimate annual migration flows directly using the question on place of residence one year ago.[9] People who indicated in the ACS that they lived in Puerto Rico one year ago are considered immigrants (i.e.,

they moved from Puerto Rico to one of the 50 states or the District of Columbia). People who indicated on the PRCS that they lived in the United States one year ago are considered emigrants (i.e., they moved from one of the 50 states or the District of Columbia to Puerto Rico). We assume the number of immigrants and emigrants under the age of one is equal to half of the number of one-year-old immigrants and emigrants, respectively.

For 2000 to 2004, we use prior research to establish a base estimate of net migration between the United States and Puerto Rico for 2000 and linearly interpolate between the 2000 net estimate and the 2005 net estimate to generate the estimates for 2001 to 2004.[10] For 2000, age, sex, race and Hispanic origin information is obtained from the Census 2000 population born in Puerto Rico who entered the United States in 1995 or later. For 2005 and later years, we base the age, sex, race, and Hispanic origin distribution of these net migrants on the demographic characteristics of the ACS 2005-2007 population born in Puerto Rico who entered the United States within ten years of the survey year. As described for foreign-born immigration and emigration, the incorporation of ACS data is phased in through linear interpolation between estimate years 2000 and 2005.

We estimate the net migration of natives to and from the United States using levels observed during the 1990s.[11] We apply the age, sex, race, and Hispanic origin distribution of natives residing in the United States to the estimate of net native migration. For 2000, the age, sex, race, and Hispanic origin distribution is obtained from Census 2000. For 2005 and later years, the distribution is obtained from ACS 2005-2007. From 2000 to 2005, information from the ACS is phased in using linear interpolation.

We derive the estimate of the net overseas movement of the Armed Forces population from data collected by the DMDC. DMDC provides data by age, sex, Hispanic origin, and individual branches of service in the Department of Defense. We apply the race distribution by Hispanic origin from the Census 2000 active military population to the Armed Forces movement overseas component.

Estimation of the Population by Quarter and Month

The calendar quarter is the basic time interval we employ for estimating the national population by age, sex, race, and Hispanic origin, but we also produce national population estimates by age, sex, race, and Hispanic origin for each month of the time series. To do this, we need components of change (births, deaths, and international migration data) by month.

NCHS and DMDC each supply their data by month. No seasonal information is available for net migration of the foreign-born, net emigration of natives, or net movement between Puerto Rico and the United States, so we assume that each part of the international migration component occurs at the same level throughout the year.

To estimate the population for the second and third months of each quarter, we first calculate a preliminary population estimate for each age, sex, race, and Hispanic origin combination by linearly interpolating between the population at the beginning of each quarter and the population

at the beginning of the next quarter. Then, we aggregate the population for each quarter and the monthly components of change for each quarter to each sex, race, and Hispanic origin combination and calculate the monthly population by sex, race, and Hispanic origin from the components of change using the basic balancing equation. Finally, we apply the age distribution of each month's preliminary estimate by sex, race, and Hispanic origin to the monthly population estimate (with no age detail) to produce the final monthly population estimates by age, sex, race, and Hispanic origin.

Estimation of the Population for Additional Universes

Once the monthly resident population estimates by age, sex, race, and Hispanic origin are complete, we produce estimates for three additional population universes: the resident population plus Armed Forces overseas, the civilian population, and the civilian noninstitutionalized population. Separate estimates are first produced by age, sex, race, and Hispanic origin for the overseas military population, the resident military population, and the institutionalized population.

As in the estimation of the net overseas movement of the Armed Forces population described above, we utilize the DMDC data on the Armed Forces population by age, sex, Hispanic origin, branch of service, and location. To estimate the full demographic detail of the overseas and resident military populations, we apply the race distribution by Hispanic origin from the Census 2000 active military population.

To estimate the institutionalized population, we first estimate the total group quarters (GQ) population aggregated by seven major types of facilities using GQ facility-level data provided by the FSCPE, the Department of Veterans Affairs, and the individual branches of service within the Department of Defense.[12] As these administrative record data include no information on the demographic detail of the GQ population, we apply the Census 2000 age, sex, race, and Hispanic origin distribution of the GQ population by the seven major facility types to the total GQ population estimates. Then, we aggregate the population in the seven major facility types into two populations by age, sex, race, and Hispanic origin: the institutionalized population and the noninstitutionalized population.

To produce the resident population plus Armed Forces overseas, we add the overseas military population by age, sex, race, and Hispanic origin to the resident population by age, sex, race, and Hispanic origin. To produce the civilian population, we again start with the resident population by age, sex, race, and Hispanic origin and subtract out the resident military population by age, sex, race, and Hispanic origin. Finally, to produce the civilian noninstitutionalized population, we subtract the institutionalized group quarters population by age, sex, race, and Hispanic origin from the civilian population by age, sex, race, and Hispanic origin.

Other Population Change

We incorporate data from other administrative sources into the estimates as necessary. These other sources include revisions from the population estimates review and update program (challenges) to the population estimates and the results of whole-entity special censuses.[13]

Data Notes Concerning the Population Estimates

In our ongoing evaluation of these population estimates, research indicates that the coverage of vital statistics administrative record data and decennial census data differ, with Census 2000 enumerating a smaller young child population (including children under 1 year of age) than expected from the birth data collected by NCHS.

As stated in the method described above, we base our estimates in large part on these Census 2000 population counts. As the population ages over the time series and we use the administrative records to update the census population, the level of the young child population is increasingly based on birth records. Therefore, in the annual estimates by age, there is a discontinuity between the population based mainly on administrative record birth data and the population based mainly on census data.

The method described above for benchmarking race and Hispanic categories of births to the base population distribution of infants addresses the serious disjunction between race of births and race of infants, but does not benchmark the total number of births, for which we rely solely on birth registration data.

---

[1] For more information on the Census 2000 Count Question Resolution program, see http://www.census.gov/dmd/www/CQR.htm.

[2] The OMB standards are detailed in Office of Management and Budget, "Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity" Notice, Vol. 62, No. 210, Thursday, October 30, 1997 http://www.whitehouse.gov/omb/fedreg/1997standards.html.

[3] Detailed information on the race modification is available at http://www.census.gov/popest/archives/files/MRSF-01-US1.html. The modified race data summary file is available at http://www.census.gov/popest/archives/files/MR-CO.txt.

[4] Birth and death data are released in three stages by NCHS: provisional, preliminary, and final. Data in each successive stage are more complete than in the previous stage. Provisional birth and death data are based on counts of events that occur in each state and the District of Columbia. Preliminary data are more complete and reflect full-record medical and demographic information about birth and death events. Final birth and death data are based on 100 percent of recorded events filed in all areas and reported to NCHS.

[5] For more information on the NCHS race bridging factors, see http://www.cdc.gov/nchs/nvss/bridged_race.htm.

[6] When we impute the race and Hispanic origin of infants from the race and Hispanic origin of their mother and father, we use an abbreviated distribution of race by Hispanic origin. This distribution consists of the largest 20 of the 62 cross-categories of Hispanic origin with race. The other 42 cross-categories, which represent less than half of one percent of the total population, are aggregated into two residual categories by Hispanic origin. In the case of fathers, we must also deal with special categories for non-reporting of either attribute. At the end of the process, we disaggregate the births in the two residual categories back into the 42 categories of births by race and Hispanic

origin based on the Census 2000 race and Hispanic origin distribution of children less than one year of age. In addition to this, there are two race-origin categories for which the benchmark ratios are not held constant, Hispanic Black (single race) and Hispanic Asian (single race). For these two categories the extraordinary growth in the number of births early in the decade suggested a substantial increase in the propensity to report these categories on birth certificates, possibly an indirect result of the gradual adoption of multiple-race reporting of births by states. Fortuitously, the reporting of these two categories prior to April 1, 2000 substantially understated the number of enumerated infants, so we attenuated the benchmarking ratios to unity from April 1, 2000 through December 31, 2005, producing a more gradual increase in the number of births and effectively ending the baseline adjustment for these two groups after 2005.

[7] We use death rates from 1998 estimates as the basis for mortality by race because the numerators from vital registration data and denominators from census-based estimates use the same 1977 OMB definition of race. Even prior to the revision of OMB guidelines, there were evident inconsistencies of reporting between census and vital registration data by race, especially for the American Indian, Eskimo and Aleut category. For this reason, we estimated mortality through the application of rates calculated for the sum of this category and the larger Asian and Pacific Islander category, proportionately adjusting deaths to match the registration data for these two categories combined. We assume that age-specific death rates for the Hispanic population do not vary by race. We determined non-Hispanic deaths by race by subtracting Hispanic origin deaths for each racial category from the total for the category.

[8] To derive the death rates for the age 70 and older population, we use a life table produced by the Social Security Administration. This life table and associated documentation are located at http://www.ssa.gov/OACT/STATS/table4c6.html.

[9] The Puerto Rico Community Survey was first fielded in 2005. See http://www.census.gov/acs/www/SBasics/FlyerPR.htm for more information.

[10] For more information on the estimate of net migration between the United States and Puerto Rico for 2000, see Christenson, M., 2002, "Evaluating Components of International Migration: Migration Between Puerto Rico and the United States," Population Division Working Paper No. 64 http://www.census.gov/population/www/documentation/twps0064.html.

[11] For information on estimates of net native migration, see Gibbs, J., G. Harper, M. Rubin, and H. Shin, 2003, "Evaluating Components of International Migration: Native-Born Emigrants," Population Division Working Paper No. 63 http://www.census.gov/population/www/documentation/twps0063.html.

[12] The seven major GQ facility types utilized in estimate production are: correctional institutions, juvenile institutions, nursing homes, other institutional facilities, college dormitories, military housing, and other noninstitutional facilities.

[13] For more information on the challenge program, see http://www.census.gov/popest/archives/challenges.html. For more information on the special census program, see http://www.census.gov/field/www/specialcensus/files/program_overview.htm. For more information on the incorporation of challenges, see http://www.census.gov/popest/topics/methodology/2008-est-relnotes.pdf.