



Family and Household Composition of the Population

Basic demographic information collected about household members—including the number of members; their relationship to each other; and each person’s sex, age, and marital status—is used to describe the composition of families and households. “Composition” describes the structure of families and households—the set of statuses and associated roles that are important for the functioning of society. American families and households have diverse and complex structures. For example, households can contain married couples, unmarried couples, single mothers, children, grandparents, other relatives (such as brothers, sisters, or in-laws), roommates, or simply one person living alone. Family and household composition is the result of demographic processes or family-related events such as marriage, divorce, and childbearing. Changes in the timing, number, or sequences of these events transform family and household composition. Examining the composition of households allows us to monitor how families and households are changing.

The household-based census is primarily designed to obtain social, demographic, and economic information about all the people living at a residence. Consequently, it can be used only to examine the family relationships of people living together at a given point in time. For example, census data can tell us how many families contain grandchildren living with their grandparents, but it cannot tell us the total number of grandparents in the United States because most grandparents do not live with their grandchildren.

Some Useful Definitions

U.S. households and families exhibit a wide variety of living arrangements: to discuss them, some definitions of key concepts are needed. A *household* can contain one or more people—everyone living in a housing unit makes up a household. In most cases, the person who owns or rents the residence is known as the *householder*. For the purposes of examining family and household composition using census data, the U.S. Census Bureau has defined two types of households: family and nonfamily. A *family household* has at least two members related by blood, marriage, or adoption, one of whom is the householder. *Families* consist of all related people in a family

household. A *nonfamily household* can be either a person living alone or a householder living only with nonrelatives.

Families can be maintained by married couples or by a man or woman with no spouse at home and may or may not contain children. In contrast, nonfamily households can only be maintained by a man or woman with no relatives at home. In censuses conducted through 1990, *children* included sons and daughters by birth, stepchildren, and adopted children of the householder regardless of the child’s age or marital status. *Own children* differ from *children* in that they are never married and are under the age of 18. Note that according to these definitions, nonfamily households cannot contain children or own children of the householder; all individuals under 18 years in nonfamily households are simply unrelated individuals.

When we want to know about the different types of families and households and how they have changed, we look at the composition of households and families. When we want to know about the relationships and characteristics of people in households, we examine the living arrangements of the individual. For example, if we wanted to know about children and families, we could ask, “How many families have children?” But we could also ask, “How many children live in families?” In the first case, we are interested in family composition; in the second, we are interested in living arrangements.

Trends in Family and Household Size and Composition

According to the 2000 census, there were 105.5 million households in the United States, up from 63.6 million in 1970. Traditionally, families have accounted for a large majority of all households. In 1940, only one out of ten households was a nonfamily household. Figure 1 shows that the proportion of nonfamily households increased steadily until 1970, increased steeply between 1970 and 1990, then remained steady in the 1990s. Part of the increase in nonfamily households up to 1990 was due to the growth in one-person households—people living alone. The proportion of households containing one person increased from 18 percent in 1970 to 25 percent in 1990 but did not increase between 1990 and 2000. The growth in one-person households, combined with declines in the proportion of households that contain five or more people, led to a reduction in household size. Between 1940 and 2000, the average number of people per household declined from 3.7 to 2.6.

Changes in marriage and divorce were important contributors to the increase in nonfamily households and to declines in the size of American families and households between 1970 and 1990. The postponement of marriage after 1960 led to a substantial increase in the percentage of young, never-married adults. Between 1970 and 1990, the percentage of women age 30 to 34 who had never married increased from 7 percent to 18 percent. For men this figure increased from 11 to 26 percent. The delay of marriage meant that young adults in 1990 were less likely than in the past to be living with their spouses and more likely to be living alone, with a cohabiting partner, or with roommates. Meanwhile, the proportion of divorced people more than doubled from 3 percent to 7 percent for men and from 4 percent to 10 percent for women from 1970 to 1990. Increases in divorce also reduced the size of households and families; divorce generally separates one household into two smaller households.

Figure 1 indicates that the growth in nonfamily households and decline in household size observed since 1940 stopped in 1990. Yet marriage continued to decline so that in 2000, 22 percent of women and 30 percent of men ages 30–34 had never married. The percentage of men and women divorced also increased, but only slightly, up to 9 percent for men and 11 percent for women. Continued delays in marriage should have led to a modest increase in the percentage of nonfamily households, but shifts in the age composition of the population counterbalanced this trend. Those ages 35–64 are substantially less likely than older or younger adults to live alone. As mentioned above, delays in marriage contribute to a high proportion of young adults living alone. Among households maintained by someone under 25, 51 percent do not include a family member. Additionally, significant improvements in the health and economic well-being of the elderly have increased life expectancy and the quality of life for both men and women. This, combined with the fact that women continued to outlive men by a significant number of years, contributes

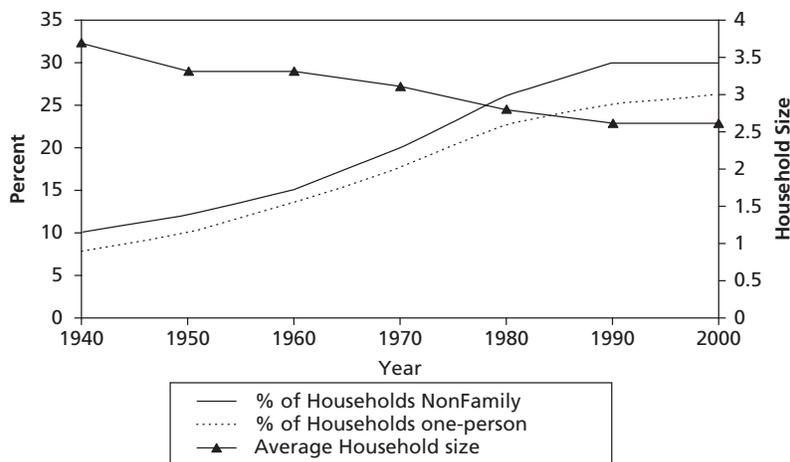
to a greater proportion of one-person nonfamily households among the elderly. Among households maintained by someone over age 75, 55 percent are nonfamily households. In contrast, only one in four households maintained by someone ages 35–64 does not include family members. During the 1990s the percentage of the population ages 35–64, the group most likely to live in family households, increased from 34 percent to over 38 percent.

Trends in fertility could also potentially influence trends in household size, but recently fertility rates have been relatively stable. In 2000, 48 percent of families included own children, the same as in 1990. Yet the context of fertility has changed substantially over time, and almost 40 percent of births 2010 are to unmarried women. Consequently, a decreasing proportion of children live with married parents. In 1970, 88 percent of families with children had married parents; this decreased to 76 percent in 1990 and to 72 percent in 2000. Trends in young children’s household contexts have changed in parallel. In 1990, 82 percent of children under age three lived in a family with married parents; in 2000 this had declined to 77 percent. Continued growth in nonmarital fertility since 2000 has likely contributed to further growth in the proportion of children living with an unmarried parent. Whereas divorce once was the primary factor contributing to changes in children’s family and household contexts, today delays in marriage and growing nonmarital fertility are increasingly important. High proportions of births to unmarried women, combined with low levels of marriage following a nonmarital birth, imply that children will spend fewer years living with their own mother and father and will likely experience a greater number of family and household transitions before they reach adulthood.

Same-Sex Households

In 1990 the census questionnaire added “unmarried partner” to the list of types of relationships a household member might have to the householder. This made it possible to identify

Figure 1. Trends in Household Size and Composition: 1940–2000



SOURCE: Decennial Censuses of Population, 1940–2000.

unmarried couples (also known as cohabiting couples) using census data. When combined with information on the sex of each household member, these data allow us to distinguish households maintained by heterosexual cohabiting couples from those maintained by gay and lesbian cohabiting couples. Additionally, some gay and lesbian couples identify themselves as married and their relationship to each other as husband or wife. These data raise many important methodological and political issues.

Until recently, the Census Bureau recoded married same-sex households as same-sex cohabiting households for its official data releases and reports because the federal government did not recognize same-sex marriage. In 2009, well after the Census Bureau had established editing procedures for the 2010 census, the Commerce Department determined that the Bureau no longer needed to follow this rule. Thus, for the 2010 census the bureau released two types of data. In the first set, it followed the original plan to continue to edit records so that same-sex individuals who reported that they were married were recoded as cohabiting. In the second set, classifications used the originally coded relationship.

Limitations of Census Data for Studying Family Composition

Census data provide a snapshot of the population at one point in time. They are excellent for describing changes in the structure of households and families because they are comparable across time and are large enough to allow analysis of extremely small groups in very small areas. However, because the 2010 census did not collect marital status and the census has never collected cohabitation status, these data miss married and cohabiting individuals who are not householders and thus are not a good source of information on the number of married or cohabiting couples, regardless of whether they are same-sex or opposite-sex. Additionally, census data are not adequate for describing the processes or events that change these structures. Standard measures of process include marriage, remarriage, divorce, fertility, and mortality rates; these measures are not currently available in the census. Administrative data collected by the National Center for Health Statistics are used to construct these vital statistics rates.

The American Community Survey (ACS) is another new source of information on family events. Starting in 2008, the ACS began collecting individual-level data on the year of most recent marriage, on divorce, and on childbearing. These data have the advantage of combining events as well as population bases, which will be valuable for calculating rates. The ACS can be pooled across multiple years for the calculation of rates for smaller population subgroups or geographic areas. As it accumulates more years of data, the ACS will come to represent an important resource for describing changes in family processes, yet even the ACS has limitations. Often family demographers are interested in information about relationships such as cohabitation and in exploring the linkages between social and economic factors and family processes. For these types of analyses, the census, the ACS, and

administrative data are limited. The National Survey of Family Growth, the Survey of Income and Program Participation, Fragile Families, and the National Longitudinal Surveys, are more appropriate data sources for investigating these complex issues.

See also *Federal household surveys.*

R. KELLY RALEY

BIBLIOGRAPHY

Carlson, Marcia, Sara McLanahan, and Paula England. "Union Formation in Fragile Families." *Demography* 41 (2004): 237–261.

Martin, Joyce A., Brady E. Hamilton, Paul D. Sutton, Stephanie J. Ventura, Fay Menacker, Sharon Kimeyer, and T. J. Mathews. *Births: Final Data for 2006*. National Vital Statistics Reports 57, no 7. Hyattsville, MD: National Center for Health Statistics, 2009.

Sweet, James A., and Larry L. Bumpass. *American Families and Households*. New York: Russell Sage Foundation, 1987.

Raley, R. Kelly, and Elizabeth Wildsmith. "Cohabitation and Children's Family Instability." *Journal of Marriage and the Family* 66 (2004): 210–219.

Federal Administrative Records

Records collected and maintained by federal agencies for the purpose of administering programs that affect broad segments of the population include some of the same kinds of information captured in the decennial census. The Census Bureau considered a number of potential uses of administrative records for the 2000 and 2010 censuses, and while none was adopted, research to evaluate prospective applications to future censuses continued. This article discusses the strengths and limitations of federal administrative records as an alternative or supplement to the information collected on a census form, reviews the Census Bureau's initial plans for administrative record use in the 2000 census, describes actual administrative record use in 2000 and 2010, and lays out several issues that must be addressed if administrative records are to play a role in the conduct of future censuses.

Major Federal Administrative Records

Each year the Internal Revenue Service (IRS) receives more than 140 million tax returns and more than a billion "information documents." Filers of tax returns and their dependents represent an estimated 92 percent of the entire U.S. population. Information documents—which include the statements that employers, financial institutions, and government agencies file to document wages, investment income, retirement income, and other benefits—cover an additional several percent of the population. With respect to items requested on the 2010 census form, both tax returns and information documents provide a current address, while tax returns also provide relationship to the filer. Both sets of documents also include Social Security numbers for nearly all persons represented, making it easy to link the tax information to demographic data maintained by the Social Security Administration (SSA). With these additional data, tax returns would approximate the content of the 2010 census form.

The SSA Numerical Identification System (Numident) file contains data collected from applications for Social Security numbers and replacement cards as well as applications to change the name associated with a previously issued number. Data recorded on the Numident besides the Social Security number include the date of birth, sex, race, Hispanic origin, country of birth, and all names associated with that number. The number of persons covered by the Numident file received a boost following the Tax Reform Act of 1986, which mandated the reporting of Social Security numbers for all dependents age five and older—a requirement that was soon expanded to age one and older. Most children now receive Social Security numbers through an application process that is coordinated with the registering of births.

The SSA Master Beneficiary Record contains benefit payment data for the 51 million people who receive Old-Age, Survivors, or Disability Insurance (OASDI) benefits. These data include current mailing addresses and benefit amounts. Because these data are used to issue payments, they are updated continuously as changes are reported.

The SSA Master Earnings File contains histories of annual earnings from jobs covered by Social Security—including self-employment. This file, which is updated from employers' annual reports of covered earnings, includes data for more than 180 million persons. The employer submissions provide the SSA with a source of current address for persons who are not yet beneficiaries.

The Health Insurance Master Record, maintained by the Centers for Medicare & Medicaid Services (CMS), is the primary administrative database for the Medicare program, which provides health insurance coverage to nearly all persons 65 and older and to younger persons with qualifying disabilities. The file covers 38 million aged beneficiaries and 8 million non-aged disabled beneficiaries and contains name, mailing address, sex, a limited race classification, age, and disability (if under 65).

The Delivery Sequence File (DSF) is a database of names and mailing addresses maintained by the U.S. Postal Service (USPS) to assist in the management of mail delivery. An authoritative source of accurate address information for mail delivery, the DSF is a valuable resource for both the USPS and its customers. While the DSF contains no demographic information on the persons associated with individual addresses, it provides a key input to the census Master Address File (MAF). Updated continuously with address changes and delivery applications submitted by USPS clients, the DSF has played a growing role in census enumeration.

Some federal agencies compile administrative records that are collected by the states and transmitted to Washington, D.C., to fulfill a legal obligation or as part of a federal-state cooperative effort. For example, the Medicaid program provides health insurance coverage to more than 60 million Americans annually through individual state programs, which receive federal reimbursements. As part of their reporting requirements, all states must now submit quarterly files of individual case record data directly to CMS. Another example is the vital statistics system. Following federal guidelines,

each state compiles records of births and deaths and submits these data to the National Center for Health Statistics, which constructs and publishes national statistics on the number and characteristics of births and deaths. The Supplemental Nutrition Assistance Program (SNAP, formerly the Food Stamp Program), administered nationally by the Food and Nutrition Service of the U.S. Department of Agriculture, serves more than 18 million low-income households throughout the nation. SNAP records are often cited as a potentially important resource for census use because they provide extensive data on households that often fall below the tax-filing threshold, but the states and localities that operate the program do not ordinarily share their administrative records with the federal government.

Limitations of Administrative Records

Despite the broad population coverage and census-relevant content of federal administrative records, these data have important limitations—particularly in the areas of data quality and timeliness. The addresses reported on tax returns are not always residential or consistent with the census concept of usual place of residence. An estimated 10 to 20 percent represent post office boxes, business addresses, or addresses of tax preparers. An unknown fraction of the remainder represents locations other than the census residence. In addition, children who are away at college may be reported on tax returns as living at home (that is, claimed as dependents), whereas the census would count them at their college addresses.

The tax-filing unit does not correspond, exactly, to a census household. Although most households contain a single filing unit, households with multiple filing units or with both filers and nonfilers who are not claimed as dependents are not uncommon. By sorting tax returns on the address field, filing units can be aggregated into households—subject to the completeness of the address information. Apartment unit numbers missing from some of the returns filed from multi-unit residences, for example, would result in separate households being combined erroneously. Furthermore, when filing units are aggregated in this manner, it may not be evident which unit contains the householder or how the members of one unit are related to a householder in another unit. Filing units may also include persons who reside in different households—as in the college student example cited above or when married couples live apart but file joint returns. In addition, filing-unit composition is based on the preceding calendar year (the tax year) rather than the filing date. Marital status is defined as of December 31, whereas dependents can satisfy the residency test in the first half of the tax year. Furthermore, dependents who died at any time during the previous year are still counted in the filing unit, and filers themselves may have died during the year, leaving a filing obligation for their surviving spouses or estates.

Perhaps the most important limitation of federal administrative records involves the race and Hispanic origin data on the Numident. The infrequency with which these data are collected—just once in a lifetime for most individuals—precludes a consistent and current measure of racial and ethnic identity across the population. The Office of Management and Budget

(OMB) revises its racial classification periodically, and the Census Bureau follows the OMB directives, but the Numident data generally reflect the classification used by the SSA at the time each individual applied for a Social Security number. Prior to 1980 the racial classification included only three categories: “black,” “white,” and “other.” Hispanic origin was not identified. In 1980 the SSA adopted the five-category classification mandated by OMB at the time, replacing the “other” category with “Hispanic origin” (intended to take precedence over “black” or “white”), “Asian or Pacific Islander,” and “North American Indian or Alaskan Native.” The current application for a Social Security number, reflecting the latest OMB directive, includes separate questions on race and Hispanic origin and allows the applicant to select multiple races from seven options.

The most serious limitation of Numident race data arises from the program initiated in 1990 to allow parents to obtain Social Security numbers for their newborn infants in conjunction with registering their births. Although the race of both biological parents (but not the newborn) is collected for the birth certificate, these data are requested in a confidential portion of the birth registration questionnaire that is not shared with the SSA. By the year 2010, well over a quarter of the records in the Numident lacked a race classification, and by 2020 this fraction is likely to exceed one-third.

The lag in data availability due to agency processing schedules is another important limitation of administrative records. Each year the IRS provides the Census Bureau with a file containing data extracted from all tax returns processed through late September. The file is received too late to be used in a census conducted in that year. Barring a change in what the IRS provides to the Census Bureau, and when, the Census Bureau would have to substitute the previous year’s file in any census application. With about 12 percent of the population moving in a given year, a comparable fraction of the tax return addresses could be expected to be out-of-date and, therefore, inappropriate for use in a census. Furthermore, the relationship data would be at least 15 months rather than 3 months out-of-date, close to 2 percent of the filers and dependents would be deceased, and all of the children born in the 15 months before an April 1 census day would be missing.

Yet another limitation of administrative records, at least potentially, is that their contents are controlled by the agencies that maintain the files and, therefore, are subject to change without regard to their census use. Just as tax reform added children’s Social Security numbers to the tax return, another tax law change could remove them—or, as happened in Canada, eliminate the reporting of dependents altogether.

With the addition of state records from federal programs that focus on the low-income population, it is possible that the coverage of administrative records could approach or even surpass that of a traditional census. Nevertheless, certain segments of the population may be systematically undercounted. Recent immigrants, who are precluded from participation in some of the major state and federal programs, are the most prominent example. Homeless persons are another. If administrative records were intended to replace the census enumeration, the attainment of complete or at least unbiased coverage would be

very important. With a more limited role, complete and unbiased coverage would be less critical.

Census 2000

In its April 1996 “Plan for Census 2000,” the Census Bureau announced that administrative records would contribute to the 2000 census in several ways. These applications included updating the MAF; assisting with the enumeration of special population groups (such as American Indians, Alaska Natives, people in group quarters, and people in remote areas); providing an alternative to in-person follow-up for about 5 percent of the households that do not return their census forms; augmenting the household rosters used in the coverage evaluation interviews; and filling in missing items on the long form. Ultimately, however, all of these applications were eliminated from the census plan, and, except for address listings from the U.S. Postal Service, military records on Armed Forces members stationed overseas, and administrative records for some group quarters, such as some prisons and nursing homes, administrative records were not used in Census 2000.

To support potential uses of administrative records in 2010, the Census Bureau focused its intercensal research on records from a small number of primarily federal sources. Building on efforts that were initiated prior to the 2000 census, the Census Bureau constructed a database of administrative records named StARS (Statistical Administrative Records System) and evaluated its coverage properties. The inputs to StARS were drawn from six sources: IRS tax returns and information documents, Department of Housing and Urban Development files, and Medicare, Selective Service, and Indian Health Service records. In creating the first version of StARS, the Census Bureau unduplicated 900 million administrative records, reducing them to 290 million unique individuals. The 1999 and 2000 versions of StARS were compared with records from the 2000 census. The results documented known or suspected weaknesses of administrative records—for example, underrepresentation of the young, mislocation of young adults and the institutionalized, and overestimation of the elderly. In subsequent years the Census Bureau continued to update and refine StARS as part of a growing program of administrative records research, which has included expanded efforts to link administrative records and survey data in order to improve understanding of measurement error in both sources.

Census 2010 and Beyond

Despite this ongoing research, administrative records played no greater role in the 2010 census than in its predecessor. A forthcoming evaluation of the latest administrative records database against the 2010 census data will document the advances made in coverage and content since 2000 and provide critical information on the viability of administrative records for population enumeration, as the Census Bureau plans for the 2020 census.

Several issues must be addressed if administrative records are to play an important role in the 2020 census. First, there must be an unambiguous determination that an enumeration from administrative records qualifies as an actual enumeration under the U.S. Constitution. Second, additional legal questions are

raised by the use of administrative records for purposes other than those for which they were collected; census applications may require legislation giving official recognition to statistical uses of such data and ensuring that the Census Bureau has continued access to key records and specific content. Third, in light of growing threats to the privacy and confidentiality of commercial and government records, the public attitude toward prospective census uses must be taken into consideration. Fourth, the processing cycles for key administrative records may have to be altered to meet the census schedule whenever the substitution of earlier data would seriously reduce data quality. Fifth, a mechanism for obtaining a contemporary measure of racial and ethnic identification across the population must be identified.

See also *Address list development; Coverage improvement procedures; Enumeration: group quarters.*

JOHN L. CZAJKA

BIBLIOGRAPHY

- Centers for Medicare & Medicaid Services. *Data Compendium*. 2009 ed. <http://www.cms.gov/datacompendium/>.
- Food and Nutrition Service. "Program Information Report: U.S. Summary, FY 2009 to FY 2010" [Excel spreadsheet]. Alexandria, Va.: U.S. Department of Agriculture, May 2010. http://www.fns.usda.gov/fns/key_data/may-2010.xls.
- Internal Revenue Service. *Statistics of Income, Individual Income Tax Returns, 2008*. Washington, D.C.: Internal Revenue Service, 2010.
- The National Research Council. Counting People in the Information Age. Panel to Evaluate Alternative Census Methods. Duane L. Steffey and Norman M. Bradburn, eds. Committee on National Statistics, Commission on Behavioral and Social Sciences and Education. Washington, D.C.: The National Academies Press, 1994. Available at http://www.nap.edu/catalog.php?record_id=4796.
- . Envisioning the 2020 Census. Panel on the Design of the 2010 Census Program of Evaluations and Experiments. Lawrence D. Brown, Michael L. Cohen, Daniel L. Cork, and Constance F. Citro, eds. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, D.C.: The National Academies Press, 2010. Available at http://www.nap.edu/catalog.php?record_id=12865.
- . Modernizing the U.S. Census. Panel on Census Requirements in the Year 2000 and Beyond. Barry Edmonston and Charles Schultze, eds. Committee on National Statistics, Commission on Behavioral and Social Sciences and Education. Washington, D.C.: The National Academies Press, 1995. Available at http://www.nap.edu/catalog.php?record_id=4805.
- . Preparing for the 2000 Census: Interim Report II. Panel on Alternative Census Methodologies.
- Andrew A. White and Keith F. Rust, eds. Committee on National Statistics, Commission on Behavioral and Social Sciences and Education. Washington, D.C.: The National Academies Press, 1997. Available at http://www.nap.edu/catalog.php?record_id=5886.
- Sailer, Peter, and Michael Weber. *The IRS Population Count: An Update*. Proceedings of the Section on Survey Research Methods. Alexandria, Va.: American Statistical Association, 1998.
- Social Security Administration. "Medicare Enrollment: National Trends, 1966–2008." In *Social Security Bulletin: Annual Statistical Supplement, 2009*. Baltimore, Md.: Social Security Administration, 2010. <http://www.cms.gov/MedicareEnRpts/Downloads/HISM108.pdf>.

Federal Household Surveys

Historically, the decennial census provided sufficient information on the population of the United States not only to satisfy the requirements of enumeration and apportionment but also

to support program planning at all levels of government and demographic and economic research in the private sector and academia. However, this information was updated only every ten years and thus was insufficient to answer all of the questions posed by Congress, policymakers, and researchers about the population, its characteristics, behaviors, and well-being. To acquire more timely information and to answer more in-depth questions, the Census Bureau and other federal agencies since the 1940s have relied principally on sample surveys that collect much more detailed information than is possible in a census but on considerably fewer households and people. Survey respondents are sampled in a manner that permits the results to be generalized to the full population with a known degree of sampling variability.

Beginning in 2010, the census is limited to collecting only basic demographic information for every household. The more detailed information that was previously collected in the full census (prior to 1940) or in a large sample of the population (on the "long form") is now being collected in the continuous American Community Survey (ACS), which is covered extensively in other articles in this volume.

In addition to the ACS, the federal government sponsors a number of household surveys that provide regularly updated information on the changing social, economic, and other characteristics of the U.S. population. Many of these surveys are conducted by the Census Bureau; others are conducted by the sponsoring agency or by private survey research firms. Survey information is used to allocate federal funds, evaluate programs, and inform policymakers. Some federal household surveys are narrowly focused on a specific topic or special subpopulation—for example, veterans—while others are broad, multipurpose surveys.

The decennial census, the ACS, and other federal household surveys complement each other, with each serving different purposes. The census, in addition to providing benchmark population counts every ten years, often provides the primary lists of households that will be sampled in a survey; the ACS provides a broad range of continually updated information for subnational geographic areas; and other household surveys provide detailed information in particular topic areas but with less geographic and demographic detail.

What Is a Survey, and How Is It Different from a Census?

Robert Groves and his coauthors define a sample survey as a "systematic method for gathering information from a sample of entities for the purposes of constructing quantitative descriptions of the . . . population of which the entities are members." A survey can be taken of virtually any population, provided that a list of objects or units in that population (that is, the "universe") and a way to find them are available. For example, households, businesses, or farms can be surveyed. Federal surveys of the U.S. population tend to sample households or other organizations that provide services to that population, such as schools or hospitals, and then ask specific questions about individuals within the household or

organization. For practical reasons, the universe may be more narrowly prescribed either to focus on the portion of the population that is of greatest interest or to enable the survey to be conducted in a more efficient manner. For example, the universe for many federal household surveys is the civilian noninstitutionalized population in the United States instead of the total resident population; however, some federal surveys include additional subpopulations in their universes. For example, a survey of the elderly may include people living in nursing homes.

Ideally, the subset of the in-universe population selected for a sample resembles the full in-universe population for the characteristics of interest to the survey so that conclusions drawn from a survey can be applied to the entire population. Asking questions of a subset of the population instead of the full population allows the collector to ask more questions, given a fixed budget, than could be asked of every unit.

The ACS and the Current Population Survey (CPS) represent examples of federal household surveys. The ACS is exceptional because of its large sample size (in 2010, the ACS obtained information from 1.9 million households) and the availability of information for small geographic areas. The other household surveys discussed in this article are more modest in size and may also be more ambitious in the amount of information collected. For example, the CPS is a key federal household survey that provides the monthly unemployment rate as well as providing the vehicle for supplemental surveys, such as the Annual Social and Economic (ASEC) Supplement. The ASEC provides annual estimates of the poverty rate and of people not covered by health insurance. The CPS has a monthly sample of approximately 60,000 households.

Why Use a Survey Instead of a Census?

The official census is limited in its content and timing, and thus it cannot be used to address many important political, economic, and social issues beyond enumeration and apportionment. Some issues require more in-depth sets of questions than those included in the decennial census to measure all facets of the problem. Consider, for example, an assessment of the impact of a change in taxes across the population at different income levels. Such a study of tax issues requires detailed information on income and population characteristics along with all the information included on federal income tax forms used to calculate tax liabilities before and after a proposed or actual change in the law. In addition, some population characteristics, such as the unemployment rate, can change rapidly and thus need more frequent measurement than could be collected from the whole population in a timely and regular manner. (In the case of the unemployment rate, the CPS is conducted monthly.)

Data collection efforts have fixed budgets and often do not have sufficient funds to collect much information on all people. Furthermore, the Paperwork Reduction Act directs the government to minimize the burden of data collection on the population as a whole, which can be done by questioning

the minimum number of people needed to accurately address an issue. Trade-offs exist among the accuracy of the results, the cost, the sample size, and the design. However, modern sampling and estimation techniques are employed to yield precise estimates needed to address many important issues, even with small to modest sample sizes. For example, the Consumer Expenditure Survey relies on a sample of fewer than 30,000 addresses to describe the spending patterns of over 310 million people in the United States.

Survey results produced by the Census Bureau, as well as other government agencies and private firms, are used in many ways by researchers, the media, nonprofit organizations, marketing firms, and all levels of government. Researchers often use survey data to assess the impact of policy changes implemented in the past, such as the major changes to social welfare programs that were introduced in 1996. The media, the federal government, and other institutions routinely use survey data to describe the population and its characteristics, such as the number of people and families living below the poverty level or the number of people who are disabled or who are college graduates. Researchers rely on analyses of underlying trends in the population covered by federal surveys to predict future patterns or to make decisions about future operations. Users of the data also simulate answers to “What if?” questions, such as “What would happen to the Social Security trust fund if the age requirement for receiving Social Security retirement benefits was increased?”

To support this statistical research, federal agencies release survey results to the public. Under the 2002 Confidential Information Protection and Statistical Efficiency Act (CIPSEA) and other legislation, such as Title 13 of the United States Code, which applies to the Census Bureau, federal agencies are required to strictly protect the confidentiality of individuals responding to the questions asked in surveys. Therefore, the data released can be in aggregate form (tables) or in the form of anonymous microdata (a file having one record per sample member where identifying information is suppressed). The Census Bureau provides access to publicly available data from the federal household surveys included in Table 1 through the Internet and through direct purchase arrangements.

How Does the Decennial Census Relate to Federal Household Surveys?

Federal surveys of the U.S. population often yield estimates about people and their characteristics, such as the number of people with health insurance coverage or the unemployment rate among different population groups. However, the underlying sampling method does not directly sample the population. Instead, the survey samples housing units, school districts, or other institutions providing services to the population. For many of the federal surveys in Table 1, the sample is based on addresses (such as 123 Main Street) obtained from the decennial census, and questions are asked of some or all of the people who reside at each of the chosen addresses.

The samples in Table 1, drawn from addresses in the decennial census, typically use the most recent decennial census supplemented with information on new housing construction subsequent to the census. The master list of addresses from which the sample is drawn is referred to as the frame. To minimize the cost of data collection while maximizing the precision of the survey estimates, the selection of addresses from the frame is not based on a simple random sample. Instead, the sample selection process involves several stages of selection at successively lower levels of geography (referred to as multistage sampling). First, the frame is often stratified by state or region or Metropolitan Statistical Area status to ensure adequate representation of these geographic areas or adequate sample sizes for these areas. Second, for household surveys conducted in person by interviewers, the addresses selected are typically clustered in smaller areas or segments based on geography to reduce travel between cases.

The decennial census also plays a key role in preparing estimates from survey results, a role described more fully in the last section of this entry.

How Long Has the Federal Government Been Conducting Surveys?

Federally sponsored surveys of the U.S. population date back to the nineteenth century. Early measures of trends in the cost of living were derived from annual surveys of income and expenditures conducted by the U.S. Department of Labor between 1888 and 1891. Periodic surveys of income and expenditures continued through the twentieth century and evolved into the Consumer Expenditure Survey, which was formalized in the early 1970s and was made continuous beginning in 1980. Until the formal survey was introduced, the expenditure surveys did not rely on the decennial census as the frame. The early expenditure surveys also did not rely on the same multistage probability sampling techniques currently in use.

The sampling methods now in use routinely by the federal government date back to the 1930s. The Financial Survey of Urban Housing began in 1934 using a sample of housing units drawn by randomly selecting blocks from large and small cities and then interviewing people in all units in the selected blocks. The Study of Consumer Purchases, conducted in 1935–1936, relied on a multistage design (then called double sampling). At the same time, the federal government conducted the National Health Survey using a sample of housing units selected from 83 cities and 23 rural counties.

In 1943 the Census Bureau conducted the Monthly Report on the Labor Force using the more sophisticated form of multistage sampling currently in use for the surveys in Table 1. This report was developed as a continuation of a series of surveys designed to study unemployment in the United States. It was renamed in 1947, in conjunction with the transfer of the project to the Census Bureau and the redesign of the sample, as the Current Population Survey, and its scope expanded beyond labor force issues.

What Do Federal Household Surveys Measure?

Most surveys have a specific focus (such as employment, health, crime victimization, etc.) guiding the content determination and the selection of the sample. Despite the wide variety of topics of federal household surveys, many of them also collect basic demographic and economic characteristics of people, such as age, race, sex, marital status, familial relationships, educational attainment, income, and labor force activity. The level of detail varies depending on the objectives. For example, surveys often will contain some measure of income of the household or of the people in the unit. This topic could be addressed using a single question (for example, “What is the total income you received from all sources last year?”) or using a series of highly detailed questions of each person on specific sources of income, amounts, and distribution within the year (as is true for the Survey of Income and Program Participation [SIPP]). The objective of the more detailed questions is to increase the accuracy of the estimate of total income or to provide details needed for analysis or both.

Most surveys query people on topics that the respondents know firsthand. Occasionally, and with permission from the respondent, data collectors supplement this personal information with information obtained directly from establishments that serve the respondents, such as health care providers. The surveys in Table 1 generally rely on self-reported information asked directly of respondents or information provided by another knowledgeable person serving as the sample member’s proxy.

The sample selection process generally restricts units to residential addresses that are occupied and whose occupants are in the civilian population. Thus, the resulting sample estimates apply to the civilian noninstitutionalized population of the United States. There are exceptions, however. For example, the New York City Housing Vacancy Survey covers unoccupied units and covers the targeted geographic area of New York City.

How Are the Data Collected?

The questions posed to respondents in federal surveys are contained within a questionnaire or instrument. The questions could be administered like the decennial census, where the questionnaire is mailed to the respondent who fills out the answers and mails it back. By their nature, these mail surveys are “self-administered”; that is, there is no interviewer to read the questions, and the questions are printed on paper. The National Survey of College Graduates is an example of a mail survey, although it also uses telephone interviews to follow up with sample members who do not respond initially.

Other surveys can be administered by an interviewer who reads each question and records the results. The interviewer can come in person to the respondent’s home or contact the respondent by telephone. Telephone interviewers may be centrally located at one or a few phone centers or dispersed throughout the community. The SIPP is conducted

by interviewers either in person or by telephone. When conducting the SIPP interview by phone, interviewers are dispersed throughout the community, most likely calling from their homes instead of calling from a central location.

The questions administered in surveys can be recorded on paper, as noted for mail surveys, or automated on a computer. The latter method is referred to as computer assisted interviewing (CAI). CAI surveys have become the norm for large, ongoing federal household surveys. CAI permits the use of the complex skip patterns that are often necessary to obtain the kind of detailed information requested in federal household surveys, while asking only questions that are relevant to the respondents. In addition, CAI instruments permit the customization of questions to the members and characteristics of the households and the use of consistency and edit checks to improve the quality of information that is collected.

CAI surveys can be interviewer administered, conducted either in person using a laptop computer or by telephone using a shared computer, such as those at central telephone centers. CAI surveys can also be self-administered, using instruments disseminated to respondents by telephone, by the Internet, or on a computer disk. The federal surveys discussed in Table 1 currently do not have a component administered through the Internet, although that option is under consideration by the Census Bureau.

Respondents to federal surveys are typically people age 15 years and older residing at selected addresses who meet the selection criteria unique to each survey. Depending on the objectives of the survey, the respondents may include just one person from the household reporting for the entire unit, all people meeting the age cutoff, or a sample of people from the household. Occasionally, and only when it is important to improve the accuracy of the responses, questions are directed to children under age 15. For example, the National Crime Victimization Survey (NCVS) administers questions to crime victims age 12 and older. The National Health Interview Survey allows any person age 18 or older to provide demographic characteristics for all household members and to provide characteristics of the family. One adult is then selected at random from each NHIS household to be administered a sample adult section of the instrument, and proxies for that respondent are not allowed. Finally, one sample child is selected at random from the NHIS household for information collected in the sample child section of the instrument, but in this case adults are allowed to provide the information on the selected child.

How Often Are Surveys Repeated?

Just as surveys can vary in their content and focus, they can vary in the number of repetitions and the interim time between repetitions. Surveys can occur one time and provide a single picture of a cross section of the population (cross-sectional), or they can be repeated over time, yielding a series of estimates that can be used to analyze trends (time series).

There are two types of repeated surveys: surveys that follow and reinterview the same unit over time (longitudinal) and

surveys that rely on newly selected samples of the full population each time they are administered (repeated cross-sectional). A number of federal surveys, such as the CPS, reflect a hybrid approach. In the CPS, addresses selected for interview are visited for 4 months in a row, retired for 8 months, and then revisited for 4 more months, potentially being surveyed a total of 8 times over a 16-month period. Although this sounds like a longitudinal survey, it is not because it does not follow people when they move. The SIPP, meanwhile, is a purely longitudinal study. It selects a sample of people for its first round of interviewing and then attempts to follow those people over time (as long as they remain alive and within the universe of the survey). In addition to being a longitudinal survey, the SIPP is refreshed every few years when the Census Bureau selects a new sample of people to follow.

How Does the Quality of Survey Data Compare to Census Data?

Statistical theory says that basing a study on a sample instead of a full population will provide results with known precision or error due to sampling. The error associated with the use of a sample instead of a census is referred to as sampling error, and the level of sampling error varies depending on how the sample is selected and the sample size.

Formulas determining variation of the sample estimates from the population estimates quantify the amount of sampling error in a given survey. For example, for each of its surveys, the Census Bureau publishes a Source and Accuracy Statement that documents the level of sampling error and its impact on survey-based estimates. Typically, estimates are provided with a variety of statistics that indicate the degree of uncertainty of the estimates, such as confidence intervals or standard errors, and are suppressed from publication if deemed unreliable based on these measures.

The quality of the data collection process is crucial to the reliability of the estimates ultimately produced from a survey. Aside from sampling error, all censuses and surveys have some amount of nonsampling error. Nonsampling error refers to the deviation of the results from the truth for reasons other than the use of a sample to collect the data. The sources of census and survey errors include coverage errors, nonresponse errors, measurement errors, and processing errors.

Coverage errors can result when the frame from which the sample is selected is not complete. Nonresponse errors may occur when there is nonresponse to the survey or census. Measurement errors may occur when respondents do not understand the question being asked or do not correctly recall the information. Measurement errors can also arise from interviewers, such as the interviewer incorrectly recording the respondent's answer or not asking the question as worded. Errors can arise in the flow of the instrument—that is, in determining the appropriate question to pose during the interview—and data processing errors can inadvertently change the answers after data collection.

Some of these forms of nonsampling error introduce bias into the survey estimates. Bias is a measure of the persistent

Table 1. Federal Surveys of U.S. Households Using Decennial Census Results and Conducted by the Census Bureau

CHARACTERISTIC	AMERICAN HOUSING SURVEY (AHS)		CONSUMER EXPENDITURE SURVEY (CE)	
	NATIONAL SURVEY	METROPOLITAN SURVEY	QUARTERLY INTERVIEW SURVEY	DIARY SURVEY
Purpose	Inform housing policy and housing program design and evaluation		Update Consumer Price Index, support poverty measurement, analyze consumer expenditures	
Frame	Decennial census + new construction		Decennial census + new construction	
Current Size	62,000 addresses (2009)	2,400 addresses (one metropolitan area, 2009)	60,000 addresses over four quarters	12,600 addresses over the year
Frequency	Biennially (odd years)	Biennially (odd years, beginning in 2007, previously even years)	Quarterly (continuous monthly interviewing); each address is in the survey for five quarters with first quarter used as <i>bounding</i> interview	Two weekly diaries (two-week reference period is designated throughout the year so all weeks are covered)
Interviewing Mode	Personal visit and telephone; automated instrument		Personal visit; automated instrument	Combination of personal visit and self-administered paper diary
Respondent	One adult resident age 16 and over (occupied unit) or owner or proxy (unoccupied unit)		One adult in the consumer unit age 16 and over (consumer units defined by household member relationships and sharing of expenses)	
Sponsor	U.S. Department of Housing and Urban Development		U.S. Bureau of Labor Statistics	
History	National survey began as Annual Housing Survey in 1973 based on 1970 census; metropolitan survey began in 1974 with 60 areas interviewed on a rotating basis (20 areas each year); biennial surveys began in 1982 (national survey) and 1996 (metropolitan survey); frame updated to 1980 census in 1985; sample size fluctuated with available funds; automation introduced in 1997; mobile home frame updated and assisted-living unit frame introduced from 2000 census		Began in 1979 based on 1970 census; frame updated to 1980 census in 1985, to 1990 census in 1995–1996, and to 2000 census in 2004–2005; automation (quarterly survey) introduced in 2003; major redesign currently under way (Bureau of Labor Statistics Gemini project; see http://www.bls.gov/cex/geminiproject.htm)	
Weighting Controls	Housing unit controls derived from the decennial census and changes in the housing stock within a decade		Population controls derived from the decennial census adjusted for births, deaths, immigration, and emigration within a decade	
CURRENT POPULATION SURVEY (CPS)				
CHARACTERISTIC	BASIC LABOR FORCE SURVEY	ANNUAL SOCIAL AND ECONOMIC SUPPLEMENT (ASEC)	OTHER SUPPLEMENTS	
Purpose	Provide monthly estimates of employment, unemployment, and labor force characteristics; disability questions added beginning in June 2008	Provide annual estimates of income, poverty, work experience, and health insurance coverage	Periodic estimates of other topics, such as child support, civic engagement, fertility, food security, housing vacancy (quarterly estimates), Internet use, school enrollment, tobacco use, veterans' characteristics, volunteering, voting	
Frame	Decennial census + new construction			
Current Size	72,000 addresses per month; about 112,000 adults per month	Hispanic oversample plus expanded sample (about 100,000 addresses and 210,000 adults); began in 2000 to support state-level estimates of children's health insurance coverage	Usually one month of sample	
Frequency	Monthly (week of the 19th); addresses are in sample for 4 months, out for 8 months, and in again for 4 months	Annually in February, March, April	Annually (with exceptions) in various months	
Interviewing Mode	Personal visit and telephone; automated instrument			
Respondent	One adult age 15 year and over responds for the unit and for each person in the unit (with some exceptions)			
Sponsor	Bureau of Labor Statistics and Census Bureau			Various agencies

(Continued)

Table 1. (Continued)

History	Monthly survey began in 1942 based on 1940 decennial census; underlying frame updated every ten years with most recent samples based on 2000 census; early samples drawn from an area frame with address sampling introduced with the 1960 census frame; sample size fluctuated over the years with maximum of 85,000 households per month (basic) in 1980; began with national sample design and switched to state-based sample in 1984		
Weighting Controls	Population controls derived from the decennial census adjusted for births, deaths, relocation, immigration, and emigration within a decade		
NATIONAL CRIME VICTIMIZATION SURVEY (NCVS)			
CHARACTERISTIC	BASIC	POLICE PUBLIC CONTACT SURVEY (PPCS)	SCHOOL CRIME SUPPLEMENT (SCS)
Purpose	Provide estimates of crime victimization	Provide estimates of interaction with police and police use of excessive force	Provide information on school-related victimization
Frame	Decennial census + new construction		
Current Size	51,000 addresses, about 100,000 people aged 12+ per year	56,000 addresses, about 96,000 people aged 16+	10,000 households, 11,000 students
Frequency	Semiannually (with interviewing spread evenly over six-month period); addresses are in sample for 7 interviews	Triennially; last survey in 2005	Biennially (with some exceptions)
Interviewing Mode	Personal interviews; automated instrument introduced in July 2006 (automated telephone instrument used in previous years)		
Respondent	Each person age 12 and over in interviewed unit	Designated sample member	Students age 12 to 18 enrolled in primary or secondary schools
Sponsor	Bureau of Justice Statistics		National Center for Education Statistics
History	Basic survey began in 1972 based on 1970 decennial census; sample frame updated every ten years with most recent frame based on 2000 census; sample size fluctuated in response to available funds; PPCS supplement began in 1996; SCS began in 1999 with predecessor surveys in 1989 and 1995; major redesign of basic NCVS is under way		
Weighting Controls	Population controls derived from the decennial census adjusted for births, deaths, immigration, and emigration within a decade		
CHARACTERISTIC	NATIONAL HEALTH INTERVIEW SURVEY (NHIS)	NATIONAL SURVEY OF COLLEGE GRADUATES (NSCG)	
Purpose	Provide estimates of the amount and distribution of illness and utilization of health care services; includes core and supplemental questions	Provide estimates of the numbers and characteristics of U.S. scientists and engineers	
Frame	Sample areas selected based on 2000 population	For the 2010 NSCG, samples from the 2009 American Community Survey, 2008 NSCG, and 2008 National Survey of Recent College Graduates	
Current Size	36,000 households (oversamples of Asians, blacks, Hispanics, and people age 65 and over in minority households)	100,000 adults (65,000 from the ACS)	
Frequency	Annually (with interviewing spread evenly over the year)	Biennially (longitudinal survey that will follow subsample of the 2010 NSCG with periodic refreshment of the sample from the ACS)	
Interviewing Mode	Personal interviews; automated instrument	Combination of mail, Internet, and computer-assisted telephone interviews	
Respondent	One adult age 18 and over responds for family and household questions; selected other adults respond to additional questions on designated adults and children	Designated sample members (noninstitutionalized adults under age 76 with bachelor's or higher degree living in the United States or its territories during the survey reference week); oversampling of ACS respondents with science and engineering bachelor's degrees	
Sponsor	National Center for Health Statistics	National Center for Science and Engineering Statistics	
History	Began in 1957 based on 1950 decennial census; frame updated in 1972, 1985, 1995, and 2006; automation introduced in 1996; various topical modules included over the years on topics of import for health services research (for example, Cancer Control Module, Child Health Mental Supplement, and Child Health Mental Services were among the topical modules in 2010)	Predecessor to the NS was the National Survey of Natural and Social Scientists and Engineers conducted in the 1970s and 1980s based on the 1970 and 1980 census long-form samples; the 1993 NSCG and the 2003 NSCG were the baseline surveys for the 1990s and 2000s—both were drawn from the census long-form sample, and subsamples were followed up over the decade	

(Continued)

Table 1. (Continued)

Weighting Controls	Population controls derived from the decennial census adjusted for births, deaths, immigration, and emigration within a decade	Weighted to the source of the sample case (ACS or 2008 surveys)
CHARACTERISTIC	NATIONAL SURVEY OF FISHING, HUNTING, AND WILDLIFE-ASSOCIATED RECREATION (FHWAR)	NEW YORK CITY HOUSING VACANCY SURVEY (NYCHVS)
Purpose	Provide estimates of fishing, hunting, and wildlife-related activities to aid in managing fish and wildlife resources	Estimate the vacancy rate for New York City's rental stock and the characteristics of housing and residents in the city
Frame	2010 decennial census	Decennial census, new construction lists, and other local information on the formation of rental units
Current Size	60,000 households for 2011 screener interview	18,000 units
Frequency	Household screener with three follow-up interviews, once every four months, of household members who participated or expected to participate in these activities to get reports of their actual participation; entire survey repeated approximately every five years	Triennially (with some exceptions)
Interviewing Mode	Personal and telephone interviews; automated instrument	Personal interview; paper instrument
Respondent	Adult household member for the screener, designated sample adult for follow-up	Adult resident (occupied units); owner, manager, or proxy (vacant units)
Sponsor	Fish and Wildlife Service of the U.S. Department of the Interior	New York City Department of Housing Preservation and Development
History	Began in 1955 and repeated at approximately five-year intervals; prior to 1996, the follow-up survey was administered once with recall period of five to sixteen months; wildlife-watching survey introduced in 1980	Began in 1962 based on 1960 decennial census; repeated approximately every three years; updated frame every decade with new decennial census results
Weighting Controls	Population controls derived from the decennial census adjusted for births, deaths, immigration, and emigration within a decade	Housing unit controls derived from the decennial census, adjusted for changes in the rental housing stock within a decade
SURVEY OF INCOME AND PROGRAM PARTICIPATION (SIPP)		
CHARACTERISTIC	1984–2008 PANELS	SURVEY OF PROGRAM DYNAMICS (SPD)
Purpose	Describe patterns of income and benefit receipt and amounts in the United States and other special topics	Estimate the impact of the welfare reform legislation of 1996
Frame	2000 decennial census + new construction	1992 and 1993 SIPP samples
Current Size	42,000 interviewed households in Wave 1 of 2008 panel	18,000 households
Frequency	Every four months for four years; includes core questions (income, program participation, and work experience by month) and topical modules (personal history, childcare, wealth, program eligibility, child support, use and cost of health care, disability, school enrollment, taxes, annual income)	Once per year for six years; similar topics as SIPP with a focus on welfare dependency
Interviewing Mode	Personal and telephone interviews; automated instrument	
Respondent	Household members age 15 and older, who are followed over the life of the panel or survey	
Sponsor	U.S. Census Bureau	
History	Predecessor surveys took place in 1978; 1979 based on 1970 decennial census frame; first SIPP panel introduced in 1984 based on 1980 census frame, followed sample adults for up to 2.5 years; subsequent panels of varying sizes and lengths introduced in 1985–1993 based on 1980 census frame; 1996 panel based on 1990 census frame followed sample adults for 4 years; 2001 panel followed sample adults for 3 years; 2004 panel followed sample adults for 4 years based on 2000 census frame; major redesign under way (see http://www.census.gov/sipp/dews.html).	Sample members from 1992 and 1993 panels of SIPP followed and interviewed once per year between 1997 and 2002; thereafter SPD funding was used to increase the size of the 2004 and 2008 SIPP panels
Weighting Controls	Population controls derived from the decennial census adjusted for births, deaths, relocation, immigration, and emigration within a decade	

SOURCE: U.S. Bureau of the Census, Demographic Surveys Division, *Survey Abstracts* (Washington, D.C., March 2010).

deviation of the estimate from the true value and can be measured by the comparison of the survey estimates with reliable independent estimates. A form of bias that is typical in surveys is underreporting of income relative to an independently derived benchmark. For example, John Czajka reports that the 2005 Current Population Survey ASEC estimate of the amount of Social Security income received by the U.S. civilian noninstitutionalized population was about 90 percent of an independently derived benchmark.

The Census Bureau and other data collection organizations employ a large number of quality assurance techniques to minimize errors in survey results and to ensure that high-quality data are produced. Typically, sample surveys devote more resources to minimizing nonsampling errors than would be practicable if the collection were a census, yielding higher-quality information.

How Are Estimates Derived from a Survey?

Regardless of the focus, federal surveys employ complex sampling algorithms; that is, the selection is not as simple as every tenth person in a row. Sample members are clustered in a way that minimizes the cost of data collection while maximizing the precision of the estimates. Thus, certain special techniques of estimation (primarily weighting) are required to use these data. Federal survey data typically contain weights for people, weights for the units sampled (that is, addresses), and weights for one or more groupings of people within the units sampled. Population estimates are derived by totaling these weights over the sample members. Following is some important information on weights.

- In a simple random sample, if the sampling process selects one out of every five units from the full universe, then each selected sample unit represents five members of the universe (the selected unit and four others). Hence, the unit can be said to have a weight of five (which is the inverse of the rate of selection), and whenever that sample member is counted in an analysis, it is counted five times. In the case of complex sample designs, the weight still represents the inverse of the rate of selection. However, the rate of selection—and hence the weight—varies across sample members.
- Weights are used to compensate for noninterviewed units. If all people eligible for an interview at a selected address either cannot or will not be interviewed, the result is a unit noninterview. If left uncorrected, the weighted estimates likely will be biased. To minimize the impact of unit noninterviews on estimates derived from surveys, the weights of interviewed units are adjusted to compensate for the missing noninterviewed units.
- Counting people in interviewed units based on the unit weight, even adjusted for nonresponse, does not always result in unbiased estimates of people because of undercoverage and the luck of the draw. Therefore, it is common for federal surveys to have person weights that are derived from unit weights but adjusted so that, when totaled over all people in the universe, they equal

the numbers in the estimates from the Census Bureau's Population Estimates Program. These estimates are derived from the decennial census updated to account for births, deaths, immigration, and emigration since the last census.

Aside from creating weights to support estimation, the Census Bureau and other survey organizations employ other processes to minimize nonsampling error in survey estimates. Typically, all new questions to be added to a survey are pretested unless already proven successful in another survey. Pretesting research is designed to address the issue of whether respondents interpret questions as intended and can provide appropriate answers. Additional techniques to test the validity of the questionnaire wording include cognitive techniques, focus groups, and expert review.

Some forms of nonsampling error can be minimized after the data are collected, such as the error associated with missing responses because of refusal, lack of knowledge, or an instrument flow problem. Just as the assignment of weights includes a correction for noninterview and errors in the original frame, the assignment of data to questions with missing responses reduces the impact of missing data on the resulting estimates. This assignment of answers to questions with missing responses is referred to as imputation. The success of imputation in reducing nonsampling error is a function of the approach used to impute the missing data, the amount of missing data to be imputed, and the extent to which and the way in which people with missing data differ from people with reported data.

In addition to imputing information not originally reported, the Census Bureau and other data collection agencies edit information if good reason exists to suspect it was reported or recorded in error. For example, if a respondent reported not being enrolled in school and then reported attending college, the answers are inconsistent, and one is likely to be incorrect. Data collectors often rely on experts in subject areas (such as an expert in education in this case) to make an assessment of which answer is more likely to be correct and then edit the data accordingly.

Summary

Many federally sponsored surveys of the U.S. population rely on the decennial census either directly or indirectly. The surveys in Table 1 are those that rely directly on the decennial census for selecting the sample and controlling the estimates produced from the surveys. Although Table 1 focuses on surveys administered by the Census Bureau, numerous other federally sponsored surveys are conducted by private firms. These firms often rely at least indirectly on published estimates from the decennial census to control their estimates and establish their sample frames.

See also *American Community Survey: methodology; American Community Survey: questionnaire content; Data products: evolution; Editing, imputation, and weighting; Long form.*

..... BRIAN HARRIS-KOJETIN

BIBLIOGRAPHY

- Bohme, Frederick G. "The Census Bureau's Current Programs: A History." Unpublished draft manuscript. U.S. Bureau of the Census, 1979–1981.
- Czajka, John C. "SIPP Data Quality." Appendix A in *Reengineering the Survey of Income and Program Participation*. Edited by Constance F. Citro and John Karl Scholz, Panel on the Census Bureau's Reengineered Survey of Income and Program Participation, Committee on National Statistics, National Research Council. Washington, D.C.: The National Academies Press, 2009. Available at http://www.nap.edu/catalog.php?record_id=12715.
- Duncan, Joseph W., and William C. Shelton. *Revolution in United States Government Statistics: 1926–1976*. Washington, D.C.: U.S. Department of Commerce, Office of Federal Statistical Policy and Standards, 1978.
- Groves, Robert M., Floyd J. Fowler Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. *Survey Methodology*. 2nd ed. Hoboken, N.J.: John Wiley and Sons, 2009.
- U.S. Bureau of the Census, Demographic Surveys Division. *Survey Abstracts*. Washington, D.C.: U.S. Bureau of the Census, 2010. Available at <http://www.census.gov/aboutus/surveyabstracts.pdf>.

Federal Statistical System Oversight and Policy: Intersection of OMB and the Decennial Census

The United States has what is frequently referred to as a "decentralized" statistical system. This decentralized system includes a total of nearly 100 agencies spread across virtually every department (and independent agency) of government. A substantial portion of the U.S. official statistics output is produced by 13 agencies that have statistical work as their principal mission. These agencies include the Census Bureau; the Bureau of Economic Analysis; and the Bureau of Labor Statistics; along with the National Centers for Education Statistics, Health Statistics, and Science and Engineering Statistics; the National Agricultural Statistics Service; the Economic Research Service; the Bureau of Justice and Transportation Statistics; the Energy Information Administration; the Internal Revenue Service's Statistics of Income Division; and the Social Security Administration's Office of Research, Evaluation, and Statistics. The heads of some of these agencies are appointed by the president with Senate confirmation, while others are career civil servants. Some of the appointed agency heads have fixed terms, while others do not. These agencies operate on the basis of separate statutes that authorize, or in some cases require, the secretaries of departments to collect and publish statistical data on particular subjects. Exclusive of funding for the decennial census, approximately 40 percent of resources dedicated to statistical work in the United States (or about \$2.5 billion annually) are expended by the principal statistical agencies. The remaining work is carried out by more than 80 agencies that conduct statistical activities in conjunction with another program mission, such as providing services (for example, medical care benefits for the elderly and the poor) or enforcing regulations (for example, with respect to the environment, transportation, or occupational safety).

With its decentralized system, the United States has had for more than 70 years a chief statistician located in the Executive Office of the President who has a number of key policy and

coordination responsibilities. The chief statistician is authorized by law and executive orders to carry out budget reviews, information collection approvals (including for the census and the American Community Survey), standard setting, and other statistical policy and coordination activities. The statutory authorities of the chief statistician currently derive from the Paperwork Reduction Act of 1995, the Budget and Accounting Procedures Act of 1950, and executive orders, as described below.

Statutory Authorities for the U.S. Chief Statistician

Sec. 3504(e) of the Paperwork Reduction Act of 1995 (44 U.S.C. 3504) assigns to the director of the Office of Management and Budget (OMB) nine "statistical policy and coordination" functions. These are to

"(1) coordinate the activities of the Federal statistical system to ensure—

"(A) the efficiency and effectiveness of the system; and

"(B) the integrity, objectivity, impartiality, utility, and confidentiality of information collected for statistical purposes;

"(2) ensure that budget proposals of agencies are consistent with system-wide priorities for maintaining and improving the quality of Federal statistics and prepare an annual report on statistical program funding;

"(3) develop and oversee the implementation of Government-wide policies, principles, standards, and guidelines concerning—

"(A) statistical collection procedures and methods;

"(B) statistical data classification;

"(C) statistical information presentation and dissemination;

"(D) timely release of statistical data; and

"(E) such statistical data sources as may be required for the administration of Federal programs;

"(4) evaluate statistical program performance and agency compliance with Government-wide policies, principles, standards and guidelines;

"(5) promote the sharing of information collected for statistical purposes consistent with privacy rights and confidentiality pledges;

"(6) coordinate the participation of the United States in international statistical activities, including the development of comparable statistics;

"(7) appoint a chief statistician who is a trained and experienced professional statistician to carry out the functions described under this subsection;

"(8) establish an Interagency Council on Statistical Policy to advise and assist the Director in carrying out the functions under this subsection that shall—

"(A) be headed by the chief statistician; and

"(B) consist of—

"(i) the heads of the major statistical programs; and

"(ii) representatives of other statistical agencies under rotating membership; and

"(9) provide opportunities for training in statistical policy functions to employees of the Federal Government under which—

“(A) each trainee shall be selected at the discretion of the Director based on agency requests and shall serve under the chief statistician for at least 6 months and not more than 1 year; and

“(B) all costs of the training shall be paid by the agency requesting training.”

Section 103 of the Budget and Accounting Procedures Act of 1950 (31 U.S.C. 1104[d]) directs the president to develop programs and prescribe regulations to improve the compilation, analysis, publication, and dissemination of statistical information by executive agencies. Executive Order No. 10253, as amended, delegates these functions to the director of OMB, to be redelegated to the administrator of the Office of Information and Regulatory Affairs. This agency houses the Statistical and Science Policy Office, which is headed by the chief statistician.

Executive Order No. 10033 of February 6, 1949, as amended, assigns the director of OMB responsibility for determining, with the concurrence of the secretary of state, what statistical information shall be provided in response to official requests received by the U.S. government from any international organization of which the United States is a member, and what agency is to provide the information.

These authorities initially were established administratively in the late 1930s and were subsequently mandated in legislation dating essentially from the 1940s. In addition, government-wide protections for the confidentiality of data collected for statistical purposes were set forth in the Confidential Information Protection and Statistical Efficiency Act of 2002.

In light of this highly decentralized system, the United States has essentially had a “culture of collaboration” among the statistical agencies for many decades. The primary role of the Office of Information and Regulatory Affairs’ Statistical and Science Policy Office is to provide oversight, coordination, and guidance for federal statistical activities. Staff members identify priorities for improving federal statistical programs, establish government-wide statistical policies and standards, and evaluate statistical programs for compliance with OMB guidance. In carrying out these responsibilities, the chief statistician has certain tools to enhance coordination, collaboration, and comparability. In particular, locating the statistical policy coordination function in OMB means that the chief statistician is directly involved in the budget development, legislative review, and information collection review processes.

Development of Budgets for Statistical Programs

One of the most important responsibilities of the Office of Management and Budget is its central role in the federal budget process. The director of OMB is responsible for preparing the president’s budget that is transmitted to the Congress each year. The process of creating the president’s budget involves not only examining agency budget proposals to ensure their economy and efficiency but also ensuring consistency with the policies and priorities of the current presidential administration. During OMB’s review, questions are raised about the

cost, utility, and feasibility of agency proposals. Using their knowledge of statistical agency programs and priorities, the statistical and science policy staff members formulate long-range plans to improve the performance of federal statistical programs so that robust measures are available for use by public and private decision makers. They also play a critical role in ensuring that budget proposals of the agencies are consistent and mutually supportive where appropriate.

In general, OMB tends to consider resources for statistical activities to be within the context of the responsibilities of individual departments—there is no overall “statistics budget” that is allocated to the statistics-producing agencies. However, the Statistical and Science Policy Office has the opportunity to recommend funding levels for each area of statistical activity and to advocate for resources that will address high-priority improvements for particular agency statistical programs or for related programs within and across agencies. Examples include the American Community Survey and the decennial census or the Bureau of Economic Analysis National Income and Products Accounts, which obtain input data from various agencies across the system.

Establishment and Enforcement of Standards

Statistical and science policy staff members ensure the quality, integrity, and accessibility of federal government statistical methodologies, activities, and products through the issuance of government-wide policies, guidelines, standards, and classifications that are developed in collaboration with the federal statistical agencies. These standards are of various types:

- **Core** standards and guidelines for statistical surveys focus on ensuring high-quality surveys and encourage agency best practices.
- **Guidelines** on protection of confidential statistical information include recommendations for a common pledge of confidentiality and promulgate best practices.
- **Classification** standards, such as industry and occupational classifications, standards for classification of data on race and ethnicity, and standards for the designation of geographic areas including Metropolitan and Micro-politan Statistical Areas, provide a “common language” for collecting and presenting data across the agencies of the decentralized statistical system.
- **Data-release** standards govern the release of official statistics. These include standards for the release of principal economic indicators (such as gross domestic product, trade data, construction statistics, and employment and unemployment statistics) that establish strict rules to separate data release by a statistical agency from policy interpretation of the information. They also include complementary standards for the release of other official statistics such as those on poverty, education, health, and crime.
- **Data access and pricing** standards provide, generally, that information produced by the federal government must be available to all on an equitable and timely basis.

User charges may be assessed to recover the cost of dissemination but *not* to cover costs associated with the original collection or processing of the data.

Approval of All National Data Collections

A key tool for monitoring and enforcing the government-wide use of the standards and classifications is the information collection review process. Under the Paperwork Reduction Act (PRA), all information that an agency proposes to collect from ten or more members of the public, whether from individuals, households, establishments, educational and nonprofit institutions, organizations, or other levels of government, must be approved by OMB. This review is not limited to surveys or to “statistical” agency data collections; rather, it extends to *all* collections of data, whether they originate for statistical, administrative, or regulatory uses. All agencies must submit all proposed information collections to OMB (about 8,000 are currently active). Furthermore, collections are approved for a maximum of three years and must be approved again if the sponsoring agency plans to continue them.

The intent of the PRA was, among other things, to minimize the “paperwork” burden that results from the collection of information by or for the federal government and to ensure the greatest possible public benefit from information created, collected, maintained, used, shared, and disseminated by or for the federal government. The law gave OMB broad authority to carry out this mandate, which applies regardless of medium—oral interview, paper form, or electronic submital—and whether or not the collection is voluntary or mandatory. In effect, this means that virtually every survey and census proposed by a federal agency must be submitted to OMB for review and approval. Central to this process is the requirement for consultation with members of the public and with other affected agencies. Public comment is solicited to evaluate whether the proposed collection of information is necessary for the proper performance of the functions of the agency, including determining whether the information will have practical utility; assessing the accuracy of the agency’s estimate of the burden of the proposed collection of information; enhancing the quality, utility, and clarity of the information to be collected; and minimizing the burden of the collection of information on respondents.

While most of the burdens of responding to government data collection requests—and indeed most of the government’s data collection requests—are not statistical in nature, statistical data collections are an important component of the government’s data collection activities and are subject to the requirements of the PRA. Surveys and other data collections are reviewed to ensure that they conform to proper statistical methodology, standards, and practices; to ensure that statistical methods are appropriate to intended uses; to monitor agencies’ use of classification standards; to coordinate collections carried out by various agencies; to prevent duplicative requests; and to reduce respondent burden.

The PRA also provides OMB with the authority to designate a central collection agency to obtain information needed

by two or more agencies and to direct an agency to make the information it collects available to another agency.

OMB and the Census

OMB’s interaction with the decennial census program is characterized by involvement in each of the roles described above—reviewing and recommending the budget resources that will be allocated for the decennial census and the American Community Survey (ACS); ensuring that classification standards are appropriately embraced in the census and ACS forms and that other standards for quality, confidentiality, and data release are honored; and reviewing and approving the forms before they are implemented by the Census Bureau. In addition, the Congress in the 1990s assigned to OMB a unique role with respect to the decennial census address lists. As part of the Local Update of Census Addresses (LUCA) program, the Congress called upon the OMB director, working through the chief statistician, to develop and implement a process whereby local jurisdictions could appeal the address lists compiled by the Census Bureau. For both the 2000 and 2010 decennial censuses, OMB designed and oversaw implementation of the Census Address List Appeals Office, a temporary activity associated with the conduct of each of these censuses.

With respect to the administration’s recommendations to the Congress for the budget for the decennial census program, the statistical and science policy staff work closely with the Census Bureau to determine the soundness of the estimates of the resources that will be needed. They also work with colleagues in the budget offices at OMB to ensure that the Census Bureau request is given appropriate attention during formulation of the president’s budget. For the decennial census and the ACS, these ongoing programs require both year-to-year discussions as well as longer-term multi-year assessments of the life-cycle costs for the activities envisioned.

With respect to implementation of government-wide statistical standards and to the review and approval of the forms ultimately used for the decennial census and the ACS, the involvement of OMB is ongoing and includes many dimensions. For example, it is often the case that census tests serve as the venue in which new terminology or categories are considered for occupations or for race and ethnicity. Another activity generally led by OMB, in collaboration with the Census Bureau, is the determination of content that ultimately will be incorporated into the decennial census and the ACS. For several decades it has been policy that the content for the decennial census (and more recently for the ACS, in its role as the replacement for the decennial census long form) would be limited to questions that were required by law or regulations. Given the limited “real estate” available on these forms and the role they play in serving needs across the government, such criteria for limiting the content were—and are today—essential. Through collaboration with an interagency committee, most recently cochaired by OMB and the Census Bureau, these criteria have guided the determination of the decennial census and the ACS content. Ultimately, OMB is responsible

for ensuring that the approved information collection forms meet quality standards, will provide data that have practical utility, respect privacy and confidentiality concerns, and minimize burden on the responding public.

See also *Content determination*.

..... KATHERINE K. WALLMAN

BIBLIOGRAPHY

The National Research Council, Committee on National Statistics. "Appendix A: Organization of the Federal Statistical System," and "Appendix B, Legislation and Regulations That Govern Federal Statistics." In *Principles and Practices for a Federal Statistical Agency*, 4th ed., ed. Constance F. Citro, Margaret E. Martin, and Miron L. Straf. Washington, D.C.: The National Academies Press, 2009. Available at http://www.nap.edu/catalog.php?record_id=12564.

U.S. Office of Management and Budget. "Chapter 19: Strengthening Federal Statistics." In *Budget of the U.S. Government: Fiscal Year 2012*, 341–344. Washington, D.C.: U.S. Government Printing Office, 2011. Available at <http://www.whitehouse.gov/sites/default/files/omb/budget/fy2012/assets/topics.pdf>.

U.S. Office of Management and Budget, Statistical and Science Policy Office, Office of Information and Regulatory Affairs. *Statistical Programs of the U.S. Government: Fiscal Year 2011*. Washington, D.C.: U.S. Government Printing Office, 2010. Available at http://www.whitehouse.gov/sites/default/files/omb/assets/information_and_regulatory_affairs/11statprog.pdf.

Foreign-Born Population of the United States

The foreign-born are U.S. residents who were born outside the United States and its outlying territories and whose parents were also foreign-born. They include people in a variety of immigration categories who fulfill the residency rules of the census, including permanent resident aliens, naturalized citizens, and some temporary migrants such as foreign students and exchange visitors. They also include persons illegally residing in the United States. Excluded from the foreign-born are persons born abroad of American parents.

A question on birthplace has been asked in each census since 1850, as well as in the American Community Survey (ACS), and has been used to distinguish U.S. natives from people of foreign birth.

Trends in Birthplace and Parentage

In 1850 the foreign-born numbered 2.2 million, accounting for just under 10 percent of the total population (see Table 1). The share of the foreign-born increased during subsequent decades, reaching a high of nearly 15 percent in 1890. The huge immigrant flows from southern and eastern Europe that started in the 1880s and continued into the 1920s dramatically swelled the foreign-born population. By 1930 the foreign-born population had peaked at 14.2 million, but due to the even greater increase in the native-born population, the foreign-born share had fallen to under 12 percent. The restrictive immigration laws of the 1920s and the economic hardships of the Great Depression caused immigration to slow to a trickle until

Table 1. U.S. Foreign-born Population: 1850–2008

YEAR	TOTAL POPULATION	FOREIGN-BORN	FOREIGN-BORN (%)
1850	23,191,876	2,244,602	9.7
1860	31,443,321	4,138,697	13.2
1870	38,558,371	5,567,229	14.4
1880	50,155,783	6,679,943	13.3
1890	62,622,250	9,249,547	14.8
1900	75,994,575	10,341,276	13.6
1910	91,972,266	13,515,886	14.7
1920	105,710,620	13,920,692	13.2
1930	122,775,046	14,204,149	11.6
1940	131,669,275	11,594,896	8.8
1950	150,216,110	10,347,395	6.9
1960	179,325,671	9,738,091	5.4
1970	203,210,158	9,619,302	4.7
1980	226,545,805	14,079,906	6.2
1990	248,709,873	19,767,316	7.9
2000	281,421,906	31,107,889	11.1
2008	304,059,728	37,960,935	12.5

SOURCES: Campbell Gibson and Kay Jung, *Historical Census Statistics on the Foreign-Born Population of the United States: 1850 to 2000*. Working Paper 81 (Washington, D.C., U.S. Bureau of the Census, 2006). 2008 American Community Survey.

about 1935. In fact, between 1930 and 1934 more people departed than entered the country. The total foreign-born population declined for the next four decades, falling to 9.6 million in 1970, or under 5 percent of the population. The Immigration Act of 1965, which opened up immigration from non-European countries, resulted in a dramatic resurgence in immigration. As a result, the foreign-born population climbed to 14.1 million in 1980 and more than doubled to 31.1 million in 2000. By 2008 it numbered nearly 38 million, an all-time high, comprising nearly 13 percent of the U.S. population.

A larger category, the foreign-stock, combines first- and second-generation immigrants—that is, the foreign-born and their U.S.-born children. Data on the total foreign-stock population are only available for the 1890–1930 and 1960–1970 censuses. Between 1890 and 1930 the foreign-stock population nearly doubled from 20.8 million to 40.3 million, accounting for approximately one-third of the U.S. population at both the start and end of the period (see Table 2). The decline in immigration after 1930 resulted in a corresponding drop in the foreign-stock population, which fell to 33.6 million by 1970, or just under 17 percent of the total population.

As noted, the foreign-stock population is composed of the foreign-born and their U.S.-born children, who could have both parents or just one parent who was foreign-born. During the early decades of the twentieth century, the foreign-born comprised between 35 and 42 percent of the foreign-stock population, and between 40 and 44 percent of

Table 2. Foreign-stock Population: 1890–1930 and 1960–1970

		DISTRIBUTION OF FOREIGN-STOCK (%)						
							NATIVE-BORN	
YEAR	TOTAL POPULATION	FOREIGN-STOCK	FOREIGN-STOCK (%)	FOREIGN-STOCK	FOREIGN-BORN	BOTH PARENTS FOREIGN-BORN	ONLY FATHER FOREIGN-BORN	ONLY MOTHER FOREIGN-BORN
1890	62,622,250	20,781,945	33.2	100.0	44.5	39.0	11.4	5.0
1900	75,994,575	26,038,397	34.3	100.0	39.7	40.9	12.9	6.4
1910	91,972,266	32,480,839	35.3	100.0	41.6	39.9	12.2	6.4
1920	105,710,620	36,715,938	34.7	100.0	37.9	42.9	12.4	6.7
1930	122,775,046	40,286,278	32.8	100.0	35.3	43.5	13.9	7.4
1960	179,325,675	34,050,442	19.0	100.0	28.6	41.4	18.9	11.0
1970	203,210,158	33,575,232	16.5	100.0	28.6	NA	NA	NA

NOTE: NA = Not available

SOURCE: Campbell Gibson and Kay Jung. *Historical Census Statistics on the Foreign-Born Population of the United States: 1850 to 2000*. Working Paper 81 (Washington, D.C., U.S. Bureau of the Census, 2006).

the foreign-stock population consisted of native-born individuals both of whose parents were born abroad. The balance included the native-born with only one foreign-born parent, and the immigrant parent was much more likely to be the father than the mother.

As immigration declined after 1930, the proportion of foreign-born people in the foreign-stock also fell. By 1970 it had dipped to just 29 percent, with the second generation comprising the remainder.

Geographic Origin

The continent of birth of the foreign-born population is available for the 1850–1930 and 1960–2000 censuses and the 2008 American Community Survey (see Table 3). In the 1800s, a period of generally open-door immigration to the United States, Europe accounted for the overwhelming share of the foreign-born. In 1850, for example, 92 percent of the foreign-born population was from Europe, while only 7 percent came from North America (essentially Canada). For the next five decades Europe and Canada contributed over 97 percent of the foreign-born population. The dominance of these two areas was assured when the United States banned the entry of Chinese and Japanese labor in 1882 and 1907, respectively, and virtually barred all Asian immigration in 1917. At the end of World War II these restrictions were slightly eased, but national-origin immigration quotas continued to favor Europe in general and countries in northern and western Europe in particular.

The Immigration Act of 1965 was a watershed in that the quota system was eliminated, creating opportunities for immigration from all countries. As a result, the proportion of Latin Americans and Asians in the immigration stream to the United States soared, while the share of Europeans declined. By 2008 Latin Americans and Asians composed 53 percent and 27 percent, respectively, of the total U.S. foreign-born population,

whereas Europeans made up just 13 percent. Thus, the dominance of European birthplace, evident among the foreign-born in the mid-nineteenth century, had waned by the start of the twenty-first.

The changing continents of origin were also reflected in the country of birth of the foreign-born (see Table 4). In 1850 the most frequently reported country of birth among the 2.2 million foreign-born was Ireland, with 962,000 people. This was followed by Germany (584,000), Great Britain (379,000), and Canada (148,000). Eight of the top ten source countries were European. By 1930 the major source countries of the foreign-born reflected the growing presence of immigrants from southern and eastern Europe. Italy was the top source country with 1.8 million people, followed by Germany, Canada, Poland, and Great Britain. By 2008 no European country was among the top ten source countries, and the ascendance of Latin America and Asia was evident in the rankings. Mexico, with 11.4 million people, was by far the most frequently reported country of birth, followed by China, the Philippines, India, and Vietnam.

Demographic Characteristics

Table 5 shows the age and gender characteristics of the foreign-born population since 1870. Immigrants disproportionately enter the United States in their prime working ages and tend to be younger than the general population. But their median age climbs as their length of stay in the country increases. The median age of the overall foreign-born population is thus influenced by both the level of immigration to the United States and the aging of earlier immigrant cohorts. The median age of the foreign-born population stood at 35 in 1870 and rose to 44 in 1930. With the precipitous drop in immigration in the 1930s and 1940s, the foreign-born population aged because it was not being replenished by younger immigrants. By 1960, with people age 65 and over comprising nearly one-third of the total foreign-born population,

Table 3. Continent of Birth of the Foreign-born Population: 1850–1930 and 1960–2008

YEAR	TOTAL FOREIGN-BORN*	DISTRIBUTION OF CONTINENT OF BIRTH (%)					
		EUROPE	ASIA	AFRICA	OCEANIA	LATIN AMERICA	NORTH AMERICA
1850	2,202,625	92.2	0.1	0.0	0.0	0.9	6.7
1860	4,134,809	92.1	0.9	0.0	0.1	0.9	6.0
1870	5,563,637	88.8	1.2	0.0	0.1	1.0	8.9
1880	6,675,875	86.2	1.6	0.0	0.1	1.3	10.7
1890	9,243,535	86.9	1.2	0.0	0.1	1.2	10.6
1900	10,330,534	86.0	1.2	0.0	0.1	1.3	11.4
1910	13,506,272	87.4	1.4	0.0	0.1	2.1	9.0
1920	13,911,767	85.7	1.7	0.1	0.1	4.2	8.2
1930	14,197,553	83.0	1.9	0.1	0.1	5.6	9.2
1960	9,678,201	75.0	5.1	0.4	0.4	9.4	9.8
1970	9,303,570	61.7	8.9	0.9	0.4	19.4	8.7
1980	13,192,563	39.0	19.3	1.5	0.6	33.1	6.5
1990	18,959,158	22.9	26.3	1.9	0.5	44.3	4.0
2000	31,107,573	15.8	26.4	2.8	0.5	51.7	2.7
2008	37,960,773	13.1	27.3	3.8	0.6	53.1	2.2

NOTE: * With a stated continent of birth

SOURCES: Campbell Gibson and Kay Jung. *Historical Census Statistics on the Foreign-Born Population of the United States: 1850 to 2000*. Working Paper 81 (Washington, D.C., U.S. Bureau of the Census, 2006). 2008 American Community Survey.**Table 4. Top 10 Source Countries of the Foreign-born: 1850, 1930, and 2008**

	1850			1930			2008	
1	Ireland	961,719	1	Italy	1,790,429	1	Mexico	11,412,668
2	Germany	583,774	2	Germany	1,608,814	2	China*	1,913,443
3	Great Britain	379,093	3	Canada	1,310,369	3	Philippines	1,684,802
4	Canada	147,711	4	Poland	1,268,583	4	India	1,622,522
5	France	54,069	5	Great Britain	1,224,091	5	Vietnam	1,138,039
6	Switzerland	13,358	6	Soviet Union	1,153,628	6	El Salvador	1,094,993
7	Mexico	13,317	7	Ireland	744,810	7	South Korea	1,030,691
8	Norway	12,678	8	Mexico	641,462	8	Cuba	974,657
9	The Netherlands	9,848	9	Sweden	595,250	9	Canada	818,920
10	Italy	3,679	10	Czechoslovakia	491,638	10	Dominican Republic	771,910

NOTES: * Includes the mainland, Taiwan, and Hong Kong. Data are based on political boundaries existing at the specified time points.

SOURCES: Campbell Gibson and Kay Jung. *Historical Census Statistics on the Foreign-Born Population of the United States: 1850 to 2000*. Working Paper 81 (Washington, D.C., U.S. Bureau of the Census, 2006). 2008 American Community Survey.

Table 5. Age Distribution and Sex Ratios of the Foreign-born Population: 1870–2008

YEAR	TOTAL	AGE DISTRIBUTION (%)				MEDIAN (YEARS)	SEX RATIO*
		UNDER 15	15 TO 44	45 TO 64	65 + AGE		
1870	5,567,229	8.4	65.8	21.8	4.0	34.6	117.4
1880	6,679,943	6.5	58.6	28.8	6.1	38.3	119.1
1890	9,249,547	8.0	57.0	27.5	7.5	37.1	121.2
1900	10,341,276	5.0	58.0	27.8	9.2	38.5	119.5
1910	13,515,886	5.7	59.9	25.5	8.9	37.2	131.1
1920	13,920,692	4.0	56.4	29.9	9.7	40.0	122.9
1930	14,204,149	2.5	50.2	35.3	12.0	43.9	116.6
1940	11,594,896	0.7	33.4	47.8	18.0	51.0	111.8
1950	10,347,395	1.9	23.1	48.6	26.3	55.9	103.3
1960	9,738,091	5.2	24.7	37.5	32.6	57.2	95.6
1970	9,619,302	6.3	34.7	27.0	32.0	52.0	84.4
1980	14,079,906	8.8	48.4	21.6	21.2	39.9	87.8
1990	19,767,316	7.5	56.8	22.0	13.6	37.3	95.8
2000	31,107,889	7.4	58.2	23.7	10.7	37.6	99.0
2008	37,960,935	5.3	53.5	29.0	12.3	40.8	100.4

NOTE: * Males per 100 females

SOURCE: Campbell Gibson and Kay Jung, *Historical Census Statistics on the Foreign-Born Population of the United States: 1850 to 2000*. Working Paper 81 (Washington, D.C., U.S. Bureau of the Census, 2006). 2008 American Community Survey.

the median age reached a high of 57. After passage of the Immigration Act of 1965, immigration once again increased, and the median age of the foreign-born declined. By 1990 the share of people age 65 and over declined to 14 percent of the total foreign-born, and the median age had dropped to 37. Despite continued immigration, the median age had increased to nearly 41 by 2008, partly a reflection of the aging of earlier immigrant cohorts.

Historically, males have accounted for a disproportionate share of immigrants entering the United States, and this disparity has skewed the sex ratio (number of males per 100 females) of the overall foreign-born population. In the 1880 foreign-born population, for example, males outnumbered females, with a sex ratio of 119 males per 100 females. This sex ratio became even more skewed as southern and eastern European immigrants entered the country. Their sex ratios were among the most lopsided of any period, and by 1910 the sex ratio for the foreign-born population had increased to 131. As immigration dipped in the 1930s and early 1940s, women outnumbered men in the immigration flow for the first time. For a few years after World War II, war brides entering the country made the immigrant stream even more disproportionately female. By 1950 the

sex ratio of the total foreign-born population had dropped to 103, and by 1970 it had reached a low of 84. Nearly one-third of the foreign-born in 1970 were elderly, and due to higher male mortality they were disproportionately female, thus skewing the overall sex ratio of the foreign-born. The Immigration Act of 1965, which made family reunification the main path of entry to the United States, witnessed the continued, though reduced, dominance of the immigrant stream by women. In 1980 the sex ratio of the foreign-born population stood at 88. It increased to 96 in 1990, and by 2008 there were an equal number of males and females among the foreign-born.

Region and State of Residence

Table 6 provides information on the foreign-born population's regional patterns of residence, as documented in censuses between 1850 and 2000 and in the 2008 American Community Survey. For the four major U.S. regions—the Northeast, Midwest, South, and West—population counts for the total and foreign-born populations are shown, as is the percentage of each region that is foreign-born.

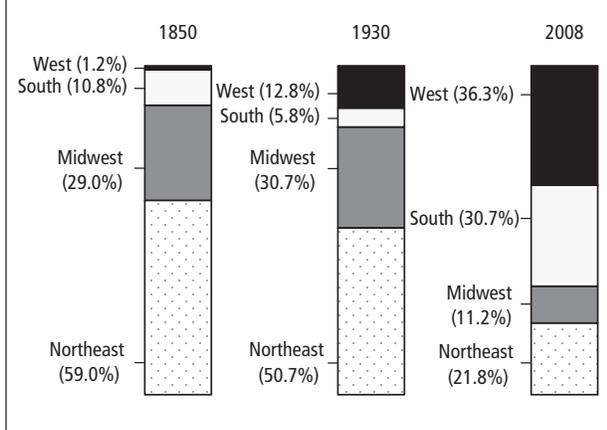
Historically, the South has had the smallest proportion of foreign-born. It was just 3 percent in 1850 and declined for

Table 6. Number and Percentage of Foreign-born for U.S. Regions: 1850–2008

YEAR	NORTHEAST			MIDWEST			SOUTH			WEST		
	TOTAL POPULATION	FOREIGN-BORN	FOREIGN-BORN (%)									
1850	8,626,851	1,325,543	15.4	5,403,595	650,375	12.0	8,982,612	241,665	2.7	178,818	27,019	15.1
1860	10,594,268	2,023,905	19.1	9,096,716	1,543,358	17.0	11,133,361	392,432	3.5	618,976	179,002	28.9
1870	12,298,730	2,520,606	20.5	12,981,111	2,333,285	18.0	12,288,020	399,975	3.3	990,510	313,363	31.6
1880	14,507,407	2,814,520	19.4	17,364,111	2,916,829	16.8	16,516,568	448,532	2.7	1,767,697	500,062	28.3
1890	17,401,545	3,888,177	22.3	22,362,279	4,060,114	18.2	19,830,813	530,346	2.7	3,027,613	770,910	25.5
1900	21,046,695	4,762,796	22.6	26,333,004	4,158,474	15.8	24,523,527	573,685	2.3	4,091,349	846,321	20.7
1910	25,868,573	6,676,283	25.8	29,888,542	4,690,461	15.7	29,389,330	740,011	2.5	6,825,821	1,409,131	20.6
1920	29,662,053	6,846,363	23.1	34,019,792	4,607,794	13.5	33,125,803	868,354	2.6	8,902,972	1,598,181	18.0
1930	34,427,091	7,201,674	20.9	38,594,100	4,359,876	11.3	37,857,633	818,614	2.2	11,896,222	1,823,985	15.3
1940	35,976,777	6,102,546	17.0	40,143,332	3,358,966	8.4	41,665,901	639,788	1.5	13,883,265	1,493,596	10.8
1950	39,341,610	5,287,165	13.4	44,281,175	2,707,390	6.1	47,085,880	767,320	1.6	19,507,445	1,585,520	8.1
1960	44,681,702	4,574,743	10.2	51,623,773	2,276,959	4.4	54,963,470	962,920	1.8	28,056,726	1,923,521	6.9
1970	49,044,015	4,119,681	8.4	56,564,917	1,873,561	3.3	62,792,882	1,316,205	2.1	34,808,344	2,309,855	6.6
1980	49,135,283	4,505,923	9.2	58,865,670	2,114,190	3.6	75,372,362	2,894,757	3.8	43,172,490	4,565,036	10.6
1990	50,809,229	5,231,024	10.3	59,668,632	2,131,293	3.6	85,445,930	4,582,293	5.4	52,786,082	7,822,706	14.8
2000	53,594,378	7,229,068	13.5	64,392,776	3,509,937	5.5	100,236,820	8,608,441	8.6	63,197,932	11,760,443	18.6
2008	54,924,779	8,264,367	15.0	66,561,452	4,255,979	6.4	111,718,549	11,674,124	10.4	70,854,948	13,766,465	19.4

SOURCES: Campbell Gibson and Kay Jung. *Historical Census Statistics on the Foreign-Born Population of the United States: 1850 to 2000*. Working Paper 81 (Washington, D.C., U.S. Bureau of the Census, 2006). 2008 American Community Survey.

Figure 1. Foreign-Born Population by Region of Residence: 1850, 1930, and 2008



SOURCE: Campbell Gibson and Kay Jung. *Historical Census Statistics on the Foreign-Born Population of the United States: 1850 to 2000*, Working Paper 81 (Washington, D.C., U.S. Bureau of the Census, 2006). 2008 American Community Survey.

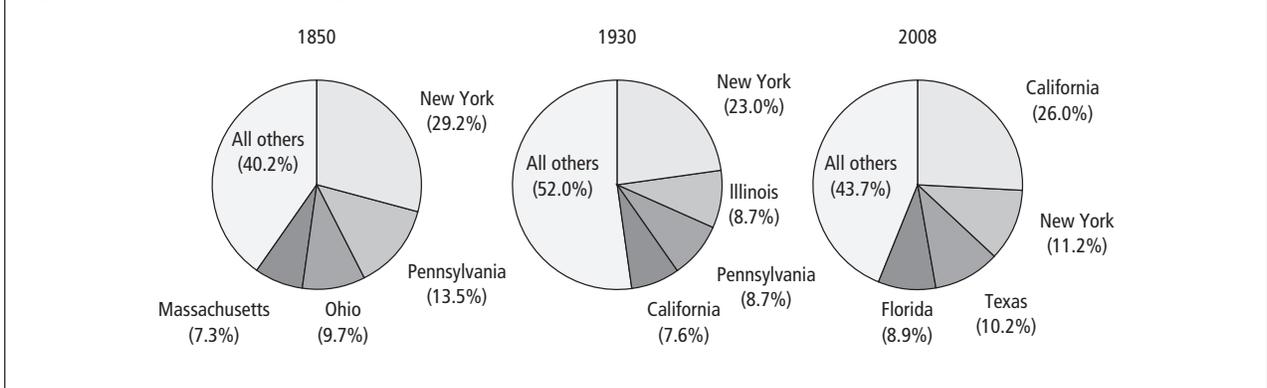
most of the early decades of the twentieth century, reaching a low of under 2 percent in 1940. In contrast, a large proportion of the population of the Northeast and West has been foreign-born. For example, in 1850 15 percent of the population of each of these regions was foreign-born. The absolute sizes of these foreign-born populations, of course, were quite different. In 1850, 1.3 million people of foreign birth lived in the Northeast, compared to just 27,000 in the western states. The proportion of foreign-born increased dramatically in both regions in the following decades. In the Northeast the huge flows from southern and eastern Europe boosted the foreign-born share to a high of 26 percent in 1910. In the West, high levels of immigration,

primarily recruited foreign workers, caused a dramatic increase in the proportion of foreign-born, which rose to 32 percent in 1870. Restricted Asian immigration precipitated a decline in the share of the foreign-born population in the following decades. The Midwest historically has also had a high foreign-born component. Its proportion of foreign-born increased during the last decades of the nineteenth century—reaching a high of 18 percent in 1890—but has largely fallen in subsequent decades.

The drop-off in immigration in the 1930s and 1940s brought large decreases in the proportion of foreign-born. By 1970 the percentages of foreign-born in the Northeast (8 percent), Midwest (3 percent), and West (7 percent) had reached all-time lows. As immigration bounced back after the Immigration Act of 1965, the proportion of foreign-born began to rise again. In 2008 the West had the highest percentage of foreign-born (19 percent), followed by the Northeast (15 percent), South (10 percent), and Midwest (6 percent).

Since 1850 there have also been shifts in the geographic distribution of the foreign-born population, reflecting changes in both the settlement patterns of immigrants and the population at large. In 1850, 59 percent of the foreign-born population lived in the Northeast, 29 percent in the Midwest, 11 percent in the South, and just 1 percent in the West (see Figure 1). By 1930 the share of the foreign-born population living in the Northeast had declined to 51 percent, with both the Midwest and the West increasing their share of the population to 31 percent and 13 percent, respectively. Changes in immigrant sources brought about by the Immigration Act of 1965—specifically, the increasing flows from Asia and Mexico—have resulted in a dramatic redistribution of the foreign-born population. Asians settle disproportionately in the western states, and Mexicans primarily immigrate to the West and South. By 2008 the West accounted for the largest share of the foreign-born (36 percent), and the rising share of the South (31 percent) exceeded that of the Northeast, which had fallen to 22 percent. Just 11 percent lived in the Midwest.

Figure 2. States with the Highest Share of the Foreign-Born Population: 1850, 1930, and 2008



SOURCE: Campbell Gibson and Kay Jung. *Historical Census Statistics on the Foreign-Born Population of the United States: 1850 to 2000*, Working Paper 81 (Washington, D.C., U.S. Bureau of the Census, 2006). 2008 American Community Survey.

Table 7. U.S. Foreign-born Population by Citizenship Status: 1920–1950 and 1970–2008

YEAR	TOTAL FOREIGN-BORN*	NATURALIZED	NATURALIZED (%)
1920	13,119,216	6,489,883	49.5
1930	13,704,296	7,919,536	57.8
1940	10,759,917	7,280,265	67.7
1950	9,615,610	7,562,970	78.7
1970	9,739,723	6,198,173	63.6
1980	14,079,906	7,110,475	50.5
1990	19,767,316	7,996,998	40.5
2000	31,107,889	12,542,626	40.3
2008	37,960,935	16,329,909	43.0

NOTE: *With citizenship status reported. Data for 1950–2000 are sample data.

SOURCES: Campbell Gibson and Kay Jung. *Historical Census Statistics on the Foreign-Born Population of the United States: 1850 to 2000*. Working Paper 81 (Washington, D.C., U.S. Bureau of the Census, 2006). 2008 American Community Survey.

These shifts in the regional distribution of the foreign-born population are further illustrated by data on the states with the largest share of that population in 1850, 1930, and 2008 (see Figure 2). In 1850 most of the foreign-born population lived in the northeastern and midwestern states of New York (29 percent), Pennsylvania (14 percent), Ohio (10 percent), and Massachusetts (7 percent). By 1930 the four top states included New York, with 23 percent of the foreign-born population, as well as Illinois and Pennsylvania (each with 9 percent), and, for the first time, California (8 percent) was ranked among the largest destinations of the foreign-born, a position it has not relinquished. Indeed, in 2008, 26 percent of the foreign-born population lived in California, while New York’s share had dropped to 11 percent. Texas (10 percent) and Florida (9 percent) rounded out the top four, a reflection of the growing immigrant presence in the South.

Citizenship Status

The foreign-born acquire U.S. citizenship through the process of naturalization. In recent decades a person wishing to

become naturalized had to be at least 18 years of age, have been lawfully admitted for permanent residence in the United States, and have continuously resided in the United States for at least five years (three years if naturalizing as a spouse of a U.S. citizen). The rate of naturalization tends to increase with more time spent in the United States, with the longest resident foreign-born having the highest rate of naturalization.

In 1920, the first census year for which data are available, one-half of the foreign-born were naturalized U.S. citizens, but by 1950 nearly four in five foreign-born residents were naturalized (see Table 7). Since immigration had slowed in the prior two decades, most of the foreign-born had arrived decades earlier, affording them ample time to naturalize. As the pace of immigration increased in subsequent decades, the overall percentage of the foreign-born population that was naturalized declined, reaching a low of 40 percent in 2000. Though large flows to the United States have continued in the first decade of the twenty-first century, the percent naturalized has increased, reaching 43 percent in 2008. This trend suggests that the inclination to naturalize has increased.

See also *Immigration*.

..... ARUN PETER LOBO
with CONTRIBUTIONS by ELLEN PERCY KRALY

BIBLIOGRAPHY

Gibson, Campbell, and Kay Jung. *Historical Census Statistics on the Foreign-Born Population of the United States: 1850 to 2000*. Working Paper No. 81. Washington, D.C.: U.S. Census Bureau, 2006. Available at <http://www.census.gov/population/www/documentation/twps0081/twps0081.pdf>.

Lobo, Arun Peter. “Unintended Consequences: Liberalized U.S. Immigration Law and the African Brain Drain.” In *The New African Diaspora in North America: Trends, Community Building, and Adaptation*, ed. Kwadwo Konadu-Agyemang, Baffour K. Takyi, and John Arthur, 189–208. Lanham, Md.: Lexington Books, 2006.

Lobo, Arun Peter, and Joseph J. Salvo. “Changing U.S. Immigration Law and the Occupational Selectivity of Asian Immigrants?” *International Migration Review* 32, no. 3 (1998): 737–760.

Salvo, Joseph J., and Arun Peter Lobo. “The Federal Statistical System: The Local Government Perspective.” *Annals of the American Academy of Political and Social Science* 631, no. 1 (2010): 75–88. doi:10.1177/0002716210374414.

U.S. Census Bureau. *A Compass for Understanding and Using American Community Survey Data: What State and Local Governments Need to Know*. Washington, D.C.: U.S. Government Printing Office, 2009.

Warren, Robert, and Ellen Percy Kraly. *The Elusive Exodus: Emigration from the United States*. Population Trends in Public Policy No. 8. Washington, D.C.: Population Reference Bureau, 1985.