

# Appendix I. Data Sources

---

*Health, United States* consolidates the most current data on the health of the population of the United States, the availability and use of health care resources, and health care expenditures. Information was obtained from the data files and published reports of many federal government, private, and global agencies and organizations. In each case, the sponsoring agency or organization collected data using its own methods and procedures. Therefore, data in this report may vary considerably with respect to source, method of collection, definitions, and reference period.

Although a detailed description and comprehensive evaluation of each data source are beyond the scope of this appendix, readers should be aware of the general strengths and weaknesses of the different data collection systems shown in *Health, United States*. For example, population-based surveys are able to collect socioeconomic data and information on the impact of an illness, such as limitation of activity. These data are limited by the amount of information a respondent remembers or is willing to report. For example, a respondent may not know detailed medical information, such as a precise diagnosis or the type of medical procedure performed, and therefore cannot report that information. In contrast, records-based surveys, which collect data from physician and hospital records, usually contain good diagnostic information but little or no information about the socioeconomic characteristics of individuals or the impact of illnesses on individuals.

Different data collection systems may cover different populations, and understanding these differences is critical to interpreting the resulting data. Data on vital statistics and national expenditures cover the entire population. However, most data on morbidity cover only the civilian noninstitutionalized population, so may not include data for military personnel, who are usually young; for institutionalized people, including the prison population, who may be of any age; or for nursing home residents, who are usually older.

All data collection systems are subject to error, and records may be incomplete or contain inaccurate information. Respondents may not remember essential information, a question may not mean the same thing to different respondents, and some institutions or individuals may not respond at all. It is not always possible to measure the magnitude of these errors or their effect on the data. Where possible, table notes describe the universe and method of data collection, to assist users in evaluating data quality.

Some information is collected in more than one survey, and estimates of the same statistic may vary among surveys because of different survey methodologies, sampling frames, questionnaires, definitions, and tabulation categories. For example, cigarette use is measured by the National Health Interview Survey, the National Survey on Drug Use and

Health and the National Youth Tobacco Survey. These surveys use slightly different questions, cover persons of differing ages, and interview in diverse settings (e.g., at school compared with at home), so estimates may differ.

Overall estimates generally have relatively small sampling errors, but estimates for certain population subgroups may be based on a small sample size and have relatively large sampling errors. Numbers of births and deaths from the National Vital Statistics System represent complete counts (except for births in those states where data are based on a 50% sample for certain years). Therefore, these data are not subject to sampling error. However, when the figures are used for analytical purposes, such as the comparison of rates over a period, the number of events that actually occurred may be considered as one of a large series of possible results that could have arisen under the same circumstances. When the number of events is small and the probability of such an event is rare, estimates may be unstable, and considerable caution must be used in interpreting the statistics. Estimates that are unreliable because of large sampling errors or small numbers of events are noted with asterisks in tables, and the criteria used to determine unreliable estimates are indicated in an accompanying footnote.

In this appendix, government data sources are listed alphabetically by data set name, and private and global sources are listed separately. To the extent possible, government data systems are described using a standard format. The “Overview” section is a brief, general statement about the purpose or objectives of the data system. “Coverage” describes the population or events that the data system covers: for example, residents of the United States, the noninstitutionalized population, persons in specific population groups, or other entities that are included in the survey or data system. “Methodology” presents a short description of the methods used to collect the data. “Sample Size and Response Rate” provides these statistics for surveys. “Issues Affecting Interpretation” describes major changes in the data collection methodology or other factors that must be considered when analyzing trends shown in *Health, United States*: for example, a major survey redesign that may introduce a discontinuity in the trend. For additional information about the methodology, data files, and history of a data source, consult the “References” and “For More Information” sections that follow each summary.

# Government Sources

## National Health and Nutrition Examination Survey (NHANES)

### National Center for Health Statistics (NCHS)

*Overview.* NHANES is designed to assess the health and nutritional status of adults and children in the United States. The survey is unique in that it combines interviews and physical examinations. NHANES collects data on the prevalence of chronic diseases and conditions (including undiagnosed conditions) and on risk factors such as obesity, elevated serum cholesterol levels, hypertension, diet and nutritional status, and numerous other measures.

*Coverage.* NHANES III, conducted during 1988–1994, and the continuous NHANES, which began in 1999, target the civilian noninstitutionalized U.S. population.

*Methodology.* NHANES includes clinical examinations, selected medical and laboratory tests, and self-reported data. NHANES interviews persons in their homes and conducts medical examinations in a mobile examination center (MEC), including laboratory analysis of blood, urine, and other tissue samples. Medical examinations and laboratory tests follow very specific protocols and are standardized as much as possible to ensure comparability across sites and providers. During 1988–1994, as a substitute for the MEC examinations, a small number of survey participants received an abbreviated health examination in their homes if they were unable to come to the MEC.

The survey for NHANES III was conducted from 1988 to 1994 using a stratified, multistage probability design to sample the civilian noninstitutionalized U.S. population. About 40,000 persons aged 2 months and over were selected and asked to complete an extensive interview and a physical examination. Participants were selected from households in 81 survey units across the United States. Children aged 2 months to 5 years, persons aged 60 and over, black persons, and persons of Mexican origin were oversampled to provide precise descriptive information on the health status of selected population groups in the United States.

Beginning in 1999, NHANES became a continuous annual survey, collecting data every year from a representative sample of the civilian noninstitutionalized U.S. population, newborns and older, through in-home personal interviews and physical examinations in the MEC. The sample design is a complex, multistage, clustered design using unequal probabilities of selection. The first-stage sample frame for continuous NHANES during 1999–2001 was the list of primary sampling units (PSUs) selected for the design of the National Health Interview Survey. Typically, an NHANES PSU is a county. For 2002, an independent sample of PSUs (based on current census data) was selected. This independent design was used for the period 2002–2006. In 2007–2010 and 2011–2014, the sample was redesigned. For 1999, because of a delay in the start of data collection, 12 distinct PSUs were in the annual sample. For each year during 2000–2016, 15

PSUs were selected. The within-PSU design involves forming secondary sampling units that are nested within census tracts, selecting dwelling units within secondary units, and then selecting sample persons within dwelling units. Selection of the final sample person involves differential probabilities of selection according to the demographic variables of sex (male or female), race and ethnicity, and age. Because of the differential probabilities of selection, dwelling units are screened for potential sample persons.

Beginning in 1999, NHANES oversampled low-income persons, adolescents aged 12–19, persons aged 60 and over, black or African American persons, and persons of Mexican origin. The sample for data years 1999–2006 was not designed to give a nationally representative sample for the total Hispanic population residing in the United States. Starting with 2007–2010 data collection, all Hispanic persons were oversampled, not just persons of Mexican origin, and adolescents were no longer oversampled. In 2011–2014, the sampling design was changed and the following groups were oversampled: Hispanic persons; non-Hispanic black persons; non-Hispanic Asian persons; non-Hispanic white and other persons at or below 130% of poverty; and non-Hispanic white and other persons aged 80 and over. In 2015–2016, the sampling design was revised again, changing the cut-point for low-income oversampling from at or below 130% of poverty to at or below 185% of poverty. For more information on the sample design for 1999–2006, see: [https://www.cdc.gov/nchs/data/series/sr\\_02/sr02\\_155.pdf](https://www.cdc.gov/nchs/data/series/sr_02/sr02_155.pdf); for 2007–2010, see: [https://www.cdc.gov/nchs/data/series/sr\\_02/sr02\\_160.pdf](https://www.cdc.gov/nchs/data/series/sr_02/sr02_160.pdf); for 2011–2014, see: [https://www.cdc.gov/nchs/data/series/sr\\_02/sr02\\_162.pdf](https://www.cdc.gov/nchs/data/series/sr_02/sr02_162.pdf); and for 2015–2016, see: <https://www.cdc.gov/nchs/nhanes/continuousnhanes/Overview.aspx?BeginYear=2015>.

The estimation procedure used to produce national statistics for all NHANES involved inflation by the reciprocal of the probability of selection, adjustment for nonresponse, and poststratified ratio adjustment to population totals. Sampling errors also were estimated, to measure the reliability of the statistics.

*Sample Size and Response Rate.* Over the 6-year survey period of NHANES III, 39,695 persons were selected; the household interview response rate was 86% (33,994), and the medical examination response rate was 78% (30,818). For NHANES 1999–2000 to NHANES 2011–2012, the number of persons selected ranged from 12,160 to 13,431. The percentage who were interviewed ranged from 73% to 84%, while the percentage who were examined ranged from 70% to 80%. For NHANES 2013–2014, a total of 14,332 persons were eligible, of which 71% (10,175) were interviewed and 68% (9,813) completed the health examination component. For NHANES 2015–2016, a total of 15,327 persons were eligible, of which 61% (9,971) were interviewed and 59% (9,544) completed the health examination component. For more detailed information on unweighted NHANES response rates and response weights using sample size weighted to Current Population Survey population totals, see: <https://www.cdc.gov/nchs/nhanes/ResponseRates.aspx>.

*Issues Affecting Interpretation.* Data elements, laboratory tests performed, and the technological sophistication of medical examination and laboratory equipment have changed over time. Therefore, trend analyses should carefully examine how specific data elements were collected across the various survey years. Data files are revised periodically. If the file changes are minor and the impact on estimates is small, then the data are not revised in *Health, United States*. Major data changes are incorporated.

Periodically, NHANES changes its sampling design to oversample different groups. Because the total sample size in any year is fixed due to operational constraints, sample sizes for the other oversampled groups (including Hispanic persons and nonlow-income white and other persons) were decreased. Therefore, trend analyses on demographic subpopulations should be carefully evaluated to determine if the sample sizes meet the NHANES Analytic Guidelines. In general, any 2-year data cycle in NHANES can be combined with adjacent 2-year data cycles to create analytic data files based on 4 or more years of data, in order to improve precision. However, because of the sample design change in 2011–2012, the data user should be aware of the implications if these data are combined with data from earlier survey cycles. Users are advised to examine their estimates carefully to see if the 4-year estimates (and sampling errors) are consistent with each set of 2-year estimates.

### References

Ezzati TM, Massey JT, Waksberg J, Chu A, Maurer KR. Sample design: Third National Health and Nutrition Examination Survey. National Center for Health Statistics. *Vital Health Stat* 2(113). 1992. Available from: [https://www.cdc.gov/nchs/data/series/sr\\_02/sr02\\_113.pdf](https://www.cdc.gov/nchs/data/series/sr_02/sr02_113.pdf).

National Center for Health Statistics. Plan and operation of the Third National Health and Nutrition Examination Survey, 1988–94. National Center for Health Statistics. *Vital Health Stat* 1(32). 1994. Available from: [https://www.cdc.gov/nchs/data/series/sr\\_01/sr01\\_032.pdf](https://www.cdc.gov/nchs/data/series/sr_01/sr01_032.pdf).

Johnson CL, Paulose-Ram R, Ogden CL, Carroll MD, Kruszon-Moran D, Dohrmann SM, Curtin LR. National Health and Nutrition Examination Survey: Analytic guidelines, 1999–2010. National Center for Health Statistics. *Vital Health Stat* 2(161). 2013. Available from: [https://www.cdc.gov/nchs/data/series/sr\\_02/sr02\\_161.pdf](https://www.cdc.gov/nchs/data/series/sr_02/sr02_161.pdf).

Johnson CL, Dohrmann SM, Burt VL, Mohadjer LK. National Health and Nutrition Examination Survey: Sample design, 2011–2014. National Center for Health Statistics. *Vital Health Stat* 2(162). 2014. Available from: [https://www.cdc.gov/nchs/data/series/sr\\_02/sr02\\_162.pdf](https://www.cdc.gov/nchs/data/series/sr_02/sr02_162.pdf).

For more information, see the NHANES website at: <https://www.cdc.gov/nchs/nhanes/index.htm>.

## National Health Expenditure Accounts (NHEA)

### Centers for Medicare & Medicaid Services (CMS)

*Overview.* NHEA provides estimates of aggregate health care expenditures in the United States from 1960 onward. NHEA contains all of the main components of the health care system within a unified, mutually exclusive, and exhaustive structure. The accounts measure spending for health care in the United States by type of good or service delivered (e.g., hospital care, physician and clinical services, or retail prescription drugs) and by the source of funds that pay for that care (e.g., private health insurance, Medicare, Medicaid, or out of pocket). A consistent set of definitions is used for health care goods and services and for sources of funds that finance health care expenditures, allowing for comparisons over time.

*Methodology.* NHEA estimates health care spending using an expenditures approach to national economic accounting. NHEA includes all of the main components of the health care system within a comprehensive and mutually exclusive structure. Expenditures are estimated for the payers, as well as the categories of medical goods and services. A common set of definitions allows comparison among categories and over time. In addition, estimates are benchmarked to revenue estimates from the Census Bureau's quinquennial Economic Census.

An assortment of government and private sources are used to create NHEA. In addition to the Economic Census, government sources include data from the Census Bureau's Services Annual Survey, the Bureau of Economic Analysis, the National Income and Product Accounts, and Medicare claims data. Private data sources include the American Hospital Association's Annual Survey and the Kaiser Health Research and Educational Trust Employer Health Benefits Survey.

For example, private health insurance spending for health care goods and services is derived using data from the Census Bureau, the American Medical Association, the American Hospital Association, IQVIA (formerly IMS Health), and the Medical Expenditure Panel Surveys (MEPS) data from the Agency for Healthcare Research and Quality. For a matrix of data sources used for NHEA, see Exhibit 4 of [National Health Expenditure Accounts: Methodology Paper, 2017](#).

*Issues Affecting Interpretation.* Every 5 years, NHEA undergoes a comprehensive revision that includes the incorporation of newly available source data, methodological and definitional changes, and benchmark estimates from the Economic Census. During these comprehensive revisions, the entire NHEA time series is opened for revision.

### References

Martin AB, Hartman M, Washington B, Catlin A, The National Health Expenditure Accounts Team. National health care spending in 2017: Growth slows to post–

Great Recession rates; share of GDP stabilizes. *Health Aff (Millwood)* 38(1):96–106. 2019.

Centers for Medicare & Medicaid Services. National Health Expenditure Accounts: Methodology Paper, 2017. Definitions, sources, and methods. 2017. Available from: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/DSM-17.pdf>.

Centers for Medicare & Medicaid Services. Summary of 2014 comprehensive revision to the National Health Expenditure Accounts. 2015. Available from: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/benchmark2014.pdf>.

For more information, see the CMS National Health Expenditure Accounts website at: <https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsHistorical.html>.

## National Health Interview Survey (NHIS)

### National Center for Health Statistics (NCHS)

*Overview.* Data from the NHIS are used to monitor the health of the U.S. population on a broad range of health topics by many demographic and socioeconomic characteristics. During household interviews, NHIS collects information on the demographic and socioeconomic characteristics of respondents, in addition to information on activity limitation, illnesses, injuries, chronic conditions, health insurance coverage (or lack thereof), utilization of health care, and other health topics.

*Coverage.* The survey covers the civilian noninstitutionalized population of the United States. Among those excluded are patients in long-term care facilities, persons on active duty with the Armed Forces (although their civilian family members are included), incarcerated persons, and U.S. nationals living in foreign countries.

*Methodology.* NHIS is a cross-sectional household interview survey. Sampling and interviewing are continuous throughout each year. The sample design follows a multistage area probability design that permits the representative sampling of households and noninstitutional group quarters (e.g., college dormitories). The sample design for NHIS is redesigned approximately every 10 years to better measure the changing U.S. population and to meet new survey objectives. A new sample design was implemented in 2016.

The current 2016 sample design has many similarities to the design that was in place from 2006 to 2015, but there are some key differences. Sample areas were reselected to take into account changes in the distribution of the U.S. population since 2006, when the previous sample design was first implemented. Commercial address lists were used

as the main source of addresses, rather than field listing; and the oversampling procedures for black, Hispanic, and Asian persons that were a feature of the previous sample design were not implemented in 2016. However, persons aged 65 or over who are black, Hispanic, or Asian continue to have a higher chance of being selected for the sample adult selection stage.

The first stage of the current 2016 sample design consists of a sample of 319 PSUs drawn from approximately 1,700 geographically defined PSUs, with some PSUs in each of the 50 states and the District of Columbia.

In the current 2016 sample design, PSUs with the largest populations (e.g., the New York City metropolitan area), also called self-representing (SR) PSUs, are sampled with certainty. The set of PSUs with smaller populations, called nonself-representing (NSR) PSUs, is stratified geographically by state. Independently within each state, a systematic sample of address clusters was selected. The NSR PSUs where these address clusters were located were then considered to be in the sample. Similarly, independently within each state, a systematic sample of address clusters was selected from the state's SR PSUs.

The 2016 NHIS sampling frame consists of three nonoverlapping parts: the unit frame (a list of addresses); the area frame (geographic areas without city-style addresses or where the unit frame was not considered to be a sufficient sampling resource); and the college dormitory frame.

The total NHIS sample is subdivided into four separate panels such that each panel (and any combination of the panels) is representative of the U.S. civilian noninstitutionalized population. This design feature has a number of advantages, including flexibility for the total sample size.

The current NHIS questionnaire, implemented in 1997, has two basic parts: a Core and one or more supplements that vary by year. The Core remains largely unchanged from year to year and allows for trend analysis and for data from more than one year to be pooled to increase the sample size for analytic purposes.

NHIS consists of four components: the Household Composition, the Family Core, the Sample Adult, and the Sample Child. The Household Composition section collects basic demographic and relationship information about everyone in the household. The Family Core, which is administered separately for each family in the household, collects information on everyone in the family. From each family participating in NHIS, one adult is randomly selected to participate in the Sample Adult questionnaire. For families with children under age 18, one child is randomly selected to participate in the Sample Child questionnaire. For children, information is provided by a knowledgeable family member aged 18 or over residing in the household. Because some health issues are different for children and adults, these two questionnaires differ in some items, but both collect basic information on health status, use of health care services, health conditions, and health behaviors.

NHIS implemented a redesigned survey in January 2019. The redesign is intended to improve the measurement

of covered health topics, reduce respondent burden by shortening the length of the questionnaire, harmonize overlapping content with other federal health surveys, establish a long-term structure of ongoing and periodic topics, and incorporate advances in survey methodology and measurement.

**Sample Size and Response Rate.** The NHIS sample size varies from year to year. It may be reduced for budgetary reasons or may be augmented if supplementary funding is available. The normal annual sample size (i.e., the number of households or persons for whom data are collected and publicly released) for the previous 2006–2015 sample design and for the new 2016 sample design is about 35,000 households containing about 87,500 persons.

In 2011–2016, the NHIS sample size was augmented in 32 states and the District of Columbia to increase the number of states for which reliable state-level estimates can be produced. Each year during 2011–2016, the sample size was augmented between 13% and 27%. In contrast to previous years, there was no sample size augmentation in 2017. In 2017, the NHIS sample included 78,132 persons, with 26,742 persons participating in the Sample Adult questionnaire and 8,845 participating in the Sample Child questionnaire. In 2017, the total household response rate was 66.5%. The final response rate in 2017 was 53.0% for the Sample Adult file and 60.6% for the Sample Child file.

**Issues Affecting Interpretation.** As part of the 1997 questionnaire redesign, some basic concepts were changed, and other concepts were measured in different ways. For some questions, there was a change in the reference period. Also in 1997, the collection methodology changed from paper-and-pencil questionnaires to computer-assisted personal interviewing. Some indicators presented in *Health, United States* begin with 1997 data because the redesign caused a break in the trend. Also, starting with *Health, United States, 2005*, estimates for 2000–2002 were revised to use 2000-based weights and differ from previous editions of *Health, United States* that used 1990-based weights for those data years. The weights available in the public-use NHIS files for 2000–2002 are 1990-based. Data for 2003–2011 use weights derived from the 2000 Census. Data for 2012 and beyond use weights derived from the 2010 Census. In 2006–2010, the sample size was reduced, and this is associated with slightly larger variance estimates than in other years when a larger sample was fielded. Starting in 2010, a geographic nonresponse adjustment was made to both the sample adult weight and the sample child weight; see Moriarity.

## References

Moriarity C. 2009 National Health Interview Survey sample adult and sample child nonresponse bias analysis. 2010. Available from: [https://www.cdc.gov/nchs/data/nhis/nr\\_bias\\_analysis\\_report\\_2009\\_NHIS.pdf](https://www.cdc.gov/nchs/data/nhis/nr_bias_analysis_report_2009_NHIS.pdf).

National Center for Health Statistics. 2017 National Health Interview Survey (NHIS) public use data release. Hyattsville, MD. Available from: <ftp://ftp.cdc.gov/pub/>

[Health\\_Statistics/NCHS/Dataset\\_Documentation/NHIS/2017/srvydesc.pdf](https://www.cdc.gov/nchs/dataset_documentation/nhis/2017/srvydesc.pdf)

For more information, see the NHIS website at: <https://www.cdc.gov/nchs/nhis.htm>.

## National HIV Surveillance System

### Centers for Disease Control and Prevention (CDC), National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention (NCHHSTP)

**Overview.** Human immunodeficiency virus (HIV) surveillance data are used to detect and monitor cases of HIV infection in the United States, evaluate epidemiologic trends, identify unusual cases requiring follow-up, and inform public health efforts to prevent and control the disease. Data collected on persons with diagnosed HIV infection include age, sex, race, ethnicity, transmission category, and geographic region.

**Coverage.** All 50 states, the District of Columbia (D.C.), and 6 U.S.-dependent areas (American Samoa, Guam, Northern Mariana Islands, Puerto Rico, Republic of Palau, and the U.S. Virgin Islands) report confirmed diagnoses of HIV infection to CDC using a uniform surveillance case definition and case report form. As of April 2008, all reporting areas had implemented confidential, name-based HIV infection reporting and agreed to participate in CDC's National HIV Surveillance System. *Health, United States* only presents data for the 50 states and D.C.

**Methodology.** The surveillance programs of local, territorial, and state health departments collect information, on demographic, transmission risk, and clinical information that also includes routine reporting of laboratory tests for HIV infection, using a standard confidential case report form. This data is then transmitted electronically, without personal identifiers, to the CDC National HIV Surveillance System (NHSS). HIV surveillance includes de-identified case report data from 50 states, D.C., and six dependent areas.

**Issues Affecting Interpretation.** Although the completeness of reporting of cases of HIV infection to state and local health departments differs by geographic region and patient population, studies conducted by state and local health departments indicate that the reporting of cases of HIV infection in most areas of the United States is at least 85% complete.

In 2014, the criteria used to define HIV diagnoses were updated. Cases diagnosed through 2013 were classified according to the 2008 HIV case definition and cases diagnosed after 2013 were classified according to the new updated definition. The 2014 case definition is similar to the 2008 definition except for the following: (1) inclusion of criteria for stage 0, (2) the inclusion of CD4 lymphocyte testing for stage 3 in children, and (3) changes in the cutoffs for CD4 percentage of total lymphocytes used for classification of stages 1 and 2 in persons aged 6 years and older. See [Appendix II](#), Human immunodeficiency virus (HIV)

disease and Acquired immunodeficiency syndrome (AIDS) for discussion of HIV diagnoses reporting definitions and other issues affecting interpretation of trends.

### Reference

Centers for Disease Control and Prevention. HIV surveillance reports. Published annually. Available from: <https://www.cdc.gov/hiv/library/reports/hiv-surveillance.html>.

For more information, see the NCHSTP website at: <https://www.cdc.gov/nchstp>.

## National Immunization Surveys (NIS)

### Centers for Disease Control and Prevention (CDC), National Center for Immunization and Respiratory Diseases (NCIRD)

*Overview.* NIS is continuing nationwide telephone sample surveys to monitor vaccination coverage rates among children aged 19–35 months (NIS–Child) and among teenagers aged 13–17 years (NIS–Teen). Data collection for children aged 19–35 months started in 1994, and data collection for teenagers aged 13–17 years started in 2006.

*Coverage.* Children aged 19–35 months and adolescents aged 13–17 years in the civilian noninstitutionalized population are represented in this survey. Estimates of vaccine-specific coverage are available for the country, the 50 states, the District of Columbia, and some U.S. territories.

*Methodology.* NIS is a nationwide telephone sample survey of households with age-eligible children. The survey uses a two-phase sample design. First, a random-digit-dialing sample of telephone numbers is drawn. When households with at least one age-eligible child are contacted, the interviewer collects demographic and access-related information on all age-eligible children, the mother, and the household, and obtains permission to contact the children's vaccination providers. Second, identified providers are sent vaccination history questionnaires by mail. Final weighted estimates are adjusted for households without telephones and for nonresponse. All vaccination coverage estimates are based on provider-reported vaccination histories. NIS–Teen followed the same sample design and data collection procedures as NIS, except that only one age-eligible adolescent was selected from each screened household for data collection.

Starting in 2011, the NIS sampling frame was expanded from a single-landline frame to dual-landline and cellular telephone sampling frames. This change increased the representativeness of the sample characteristics but had little effect on the final 2011 NIS and NIS–Teen national estimates of vaccination coverage overall and when stratified by poverty status. See details of the dual-frame sample design in the annual NIS data user's guide on the NIS website, available from: <https://www.cdc.gov/vaccines/imz-managers/nis/datasets.html>.

*Sample Size and Response Rate.* In 2017, the overall Council of American Survey Research Organizations (CASRO) response rate for NIS was 26.1%. Response rates for the landline and cellular telephone samples were 51.9% and 25.0%, respectively. Of the 2,235 age-eligible children with completed household interviews from the landline sample, 1,279 (57.2%) had adequate provider data. From the cellular telephone sample, 14,054 of the 26,230 eligible children with completed household interviews had adequate provider data (53.6%).

The overall CASRO response rate for the 2017 NIS–Teen was 25.7%. Response rates for the landline and cellular telephone samples were 51.5% and 23.5%, respectively. Of the 6,663 age-eligible adolescents with completed household interviews from the landline sample, 3,572 (53.6%) had adequate provider data. From the cellular telephone sample, 17,377 of the 36,928 (47.1%) eligible adolescents with completed household interviews had adequate provider data.

*Issues Affecting Interpretation.* The findings in recent years are subject to several limitations. Data year 2011 was the first year that NIS and NIS–Teen used a dual-frame sampling scheme that included landline and cellular telephone households. Estimates from 2011 and subsequent years might not be comparable with those from prior to 2011, when surveys were conducted via landline telephone only. NIS is a telephone survey, and statistical adjustments might not compensate fully for nonresponse and for households without landline telephones prior to 2011. Underestimates of vaccination coverage might have resulted in exclusive use of provider-reported vaccination histories because completeness of records is unknown.

Before January 2009, NIS did not distinguish between Hib vaccine production types; therefore, children who received three doses of a vaccine product that requires four doses were misclassified as fully vaccinated. For more information, see: CDC. Changes in measurement of *Haemophilus influenzae* serotype b (Hib) vaccination coverage—National Immunization Survey, United States, 2009. *MMWR Morb Mortal Wkly Rep* 59(33):1069–72. 2010. Available from: <https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5933a3.htm>.

Starting in 2014, NIS–Teen defined an adolescent's vaccination record as having adequate provider data if that adolescent had vaccination history data from one or more of the named vaccination providers, or if the parent reported that the adolescent was completely unvaccinated. Prior to 2014, the adequate provider data definition had more criteria, and it was based on a comparison of provider report of vaccination history with parental report of vaccination history, either by shot card report or recall.

To assess the effect of the change in the adequate provider definition criteria on vaccination coverage estimates, NIS recomputed estimates from the 2006–2013 survey. In general, 2013 NIS–Teen vaccination coverage estimates using the revised adequate provider data definition were different, and generally lower, than original 2013 NIS–Teen estimates. Differences between revised

and original 2013 national vaccination estimates ranged from –0.1 percentage point to –2.2 percentage points. For more information on the revised adequate provider data criteria, see: <https://www.cdc.gov/vaccines/imz-managers/coverage/nis/teen/apd-report.html>, and for revised 2013 estimates based on the 2014 criteria, see: Reagan-Steiner S, Yankey D, Jeyarajah J, Elam-Evans LD, Singleton JA, Curtis CR, et al. National, regional, state, and selected local area vaccination coverage among adolescents aged 13–17 years—United States, 2014. *MMWR Morb Mortal Wkly Rep* 64(29):784–92. 2015. Available from: <https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6429a3.htm>. Because of the revision in the adequate provider definition, NIS–Teen vaccination coverage estimates for 2013 and beyond cannot be directly compared with previously published 2006–2013 NIS–Teen survey vaccination coverage estimates based on the previous adequate provider definition.

### References

Hill HA, Elam-Evans LD, Yankey D, Singleton JA, Kang Y. Vaccination coverage among children aged 19–35 months—United States, 2017. *MMWR Morb Mortal Wkly Rep* 67(40):1123–8. 2017. Available from: <https://www.cdc.gov/mmwr/volumes/67/wr/pdfs/mm6740a4-H.pdf>.

Walker TY, Elam-Evans LD, Yankey D, Markowitz LE, Williams CL, Mbaeyi SA, et al. National, regional, state, and selected local area vaccination coverage among adolescents aged 13–17 years—United States, 2017. *MMWR Morb Mortal Wkly Rep* 67(33): 909–917. 2017. Available from: <https://www.cdc.gov/mmwr/volumes/67/wr/pdfs/mm6733a1-H.pdf>.

## National Income and Product Accounts (NIPA)

### Bureau of Economic Analysis (BEA)

*Overview.* NIPA are a set of economic accounts that provide detailed measures of the value and composition of national output and the incomes generated in the production of that output. Essentially, NIPA provides a detailed snapshot of the myriad transactions that make up the economy, such as buying and selling goods and services, hiring of labor, investing, renting property, and paying taxes. NIPA estimates show U.S. production, distribution, consumption, investment, and saving.

The best-known NIPA measure is the gross domestic product (GDP), which is defined as the market value of the goods, services, and structures produced by the economy in a given period. NIPA calculates GDP as the sum of the final expenditure components: personal consumption expenditures, private fixed investment, change in private inventories, net exports of goods and services, government spending, and government investment. However, GDP is just one of many economic measures presented in NIPA. Another key NIPA indicator presented in *Health, United States* is the

implicit price deflator for GDP.

The conceptual framework of NIPA is illustrated by seven summary accounts: the domestic income and product account, the private enterprise income account, the personal income and outlay account, the government receipts and expenditures account, the foreign transactions current account, the domestic capital account, and the foreign transactions capital account. These summary accounts record a use (or expenditure) in one account for one sector and a corresponding source (or receipt) in an account of another sector or of the same sector. This integrated system provides a comprehensive measure of economic activity in a consistently defined framework without double counting.

*Coverage.* Source data for NIPA domestic estimates cover all 50 states and the District of Columbia. The U.S. national income and product statistics were first presented as part of a complete and consistent double-entry accounting system in the summer of 1947.

*Methodology.* NIPA estimates are revised quarterly, annually, and quinquennially. For GDP and most other NIPA series, a set of three current quarterly estimates is released each year. Quarterly estimates provide the first look at the path of U.S. economic activity. Annual revisions of NIPA are usually carried out each summer. These revisions incorporate source data that are based on more extensive annual surveys, on annual data from other sources, and on later revisions to the monthly and quarterly source data, and they generally cover the 3 previous calendar years. Comprehensive revisions are carried out at about 5-year intervals and may result in revisions that extend back many years. These estimates incorporate all of the best available source data, such as data from the quinquennial U.S. Economic Census.

NIPA measures are built up from a wide range of source data using a variety of estimating methods. To ensure consistency and accuracy, NIPA uses various adjustment and estimation techniques to estimate data. Three general types of adjustments are made to the source data that are incorporated into the NIPA estimates. The first consists of adjustments that are needed so that the data conform to appropriate NIPA concepts and definitions. The second type of adjustment involves filling gaps in coverage. The third type of adjustment involves time of recording and valuation. Source data must be adjusted occasionally to account for special circumstances that affect the accuracy of the data. For example, quarterly and monthly NIPA estimates are adjusted seasonally at the detailed-series level when the series demonstrate statistically significant seasonal patterns. Source data may also be used as indicators to extrapolate annual estimates. For more information, see “An introduction to the National Income and Product Accounts. Methodology papers: U.S. National Income and Product Accounts,” available from: [https://www.bea.gov/scb/pdf/national/nipa/methpap/mpi1\\_0907.pdf](https://www.bea.gov/scb/pdf/national/nipa/methpap/mpi1_0907.pdf); and “Concepts and methods of the U.S. National Income and Product Accounts,” available from: <https://www.bea.gov/sites/default/files/methodologies/nipa-handbook-all-chapters.pdf>.

*Issues Affecting Interpretation.* NIPA source data and estimates are revised frequently. Data are released at different times, estimates are updated as they become available, new concepts and definitions are incorporated, and source data may change due to improvements in collection and new methodologies. As a result, major estimates, such as GDP and its major components, undergo frequent revision, and historical data are changed. For more information, see the BEA (NIPA) website at: [https://apps.bea.gov/iTable/index\\_nipa.cfm](https://apps.bea.gov/iTable/index_nipa.cfm).

### Reference

U.S. Bureau of Economic Analysis. A guide to the National Income and Product Accounts of the United States. 2017. Available from: <https://www.bea.gov/national/pdf/nipaguid.pdf>.

For more information, see the BEA website at: <https://www.bea.gov/national/index.htm>.

**National Medical Expenditure Survey (NMES)**—See [Appendix I](#), Medical Expenditure Panel Survey (MEPS).

## National Notifiable Diseases Surveillance System (NNDSS)

### Centers for Disease Control and Prevention (CDC)

*Overview.* NNDSS is a nationwide collaboration that enables all levels of public health (local, state, territorial, federal, and international) to share health information to monitor, control, and prevent the occurrence and spread of state-reportable and nationally notifiable infectious and some noninfectious diseases and conditions. NNDSS is a multifaceted program that includes the surveillance system for collection, analysis, and sharing of health data, resources, and information about policies and standards, at the local, state, and national levels. NNDSS provides weekly provisional and annual finalized information on the occurrence of diseases defined as notifiable by the Council of State and Territorial Epidemiologists (CSTE). Data include incidence of nationally notifiable reportable diseases, which are reported using uniform surveillance case definitions.

*Coverage.* Notifiable disease reports are received from health departments in the 50 states, 5 territories, the District of Columbia, and New York City. Policies for reporting notifiable disease cases can vary by disease or reporting jurisdiction, depending on case status classification (i.e., confirmed, probable, or suspect).

*Methodology.* CDC, in partnership with CSTE, administers NNDSS. Reportable disease surveillance is conducted by public health practitioners at local, state, and national levels to support disease prevention and control.

Data on a subset of reportable conditions that have been designated nationally notifiable are then submitted to CDC. The system also provides annual summaries of the finalized data. CSTE and CDC annually review the status of national notifiable disease surveillance and recommend additions or deletions to the list of nationally notifiable diseases, based on the need to respond to emerging priorities. For example, Zika virus disease and Zika virus infection became nationally notifiable in 2016. However, reporting nationally notifiable diseases to CDC is voluntary. Because reporting is currently mandated by law or regulation only at the local and state levels, the list of diseases that are considered reportable varies by state. For example, reporting of coccidioidomycosis to CDC is not done by some states in which this disease is not reportable to local or state authorities.

State epidemiologists report cases of nationally notifiable diseases to CDC, which tabulates and publishes these data in *Morbidity and Mortality Weekly Reports* (MMWR) and in *Summary of Notifiable Diseases, United States* (titled *Annual Summary* before 1985). Beginning in 2016, national notifiable disease data are released via the NNDSS website, available from: <https://www.cdc.gov/nndss/infectious-tables.html>.

*Issues Affecting Interpretation.* NNDSS data must be interpreted in light of reporting practices. Some diseases that cause severe clinical illness (e.g., meningococcal disease, plague, and rabies) are likely reported accurately if diagnosed by a clinician. However, persons who have diseases that are clinically mild and infrequently associated with serious consequences (e.g., salmonellosis) may not seek medical care from a health care provider. Even if these less-severe diseases are diagnosed, they are less likely to be reported.

The degree of completeness of data reporting is also influenced by the diagnostic facilities available, the control measures in effect, public awareness of a specific disease, and the interests, resources, and priorities of state and local officials responsible for disease control and public health surveillance. Finally, factors such as changes in case definitions for public health surveillance, introduction of new diagnostic tests, or discovery of new disease entities can cause changes in disease reporting that are independent of the true incidence of disease.

### References

Centers for Disease Control and Prevention. MMWR: Summary of notifiable infectious diseases. Available from: [https://www.cdc.gov/mmwr/mmwr\\_nd/index.html](https://www.cdc.gov/mmwr/mmwr_nd/index.html).

Centers for Disease Control and Prevention. National Notifiable Diseases Surveillance System. 2016 annual tables of infectious disease data. 2017. Available from: [https://wonder.cdc.gov/nndss/nndss\\_annual\\_tables\\_menu.asp](https://wonder.cdc.gov/nndss/nndss_annual_tables_menu.asp).

For more information, see the NNDSS website at: <https://wwwn.cdc.gov/nndss/>.



# National Study of Long-Term Care Providers (NSLTCP)

## National Center for Health Statistics (NCHS)

*Overview.* NSLTCP is a biennial study to monitor the major sectors of paid, regulated long-term care services. NSLTCP uses administrative data from the Centers for Medicare & Medicaid Services (CMS) about the home health, nursing home, and hospice sectors and collects survey data on the residential care community and adult day services sectors. Information includes the supply, organizational characteristics, staffing, and services offered by providers of long-term care services and the demographic, health, and functional status of users of these services. NSLTCP replaces NCHS' periodic National Nursing Home Survey and National Home and Hospice Care Survey, and the one-time National Survey of Residential Care Facilities.

*Coverage.* Initiated in 2012, the NSLTCP included providers that were licensed, registered, listed, certified, or otherwise regulated by the federal or state governments.

*Methodology.* Data on adult day services centers and residential care communities were obtained through surveys. Information on nursing homes, home health agencies, and hospices was obtained from CMS administrative records.

Survey data were collected through three modes: self-administered, hard copy mail questionnaires; self-administered web questionnaires; and computer-assisted telephone interviews (CATI). To the extent possible, the questionnaires included topics comparable across all five LTC sectors, as well as selected topics specific to a particular sector.

An adult day services center frame was obtained from the National Adult Day Services Association (NADSA) containing 5,349 self-identified adult day services centers; duplicates were removed from the frame, leaving 5,348 centers in the sampling frame for the 2016 wave. A census of all adult day services centers on the sampling frame were included in the 2016 NSLTCP. In addition to being included in the NADSA database, centers were eligible if they: 1) were licensed or certified by the state specifically to provide adult day services, or accredited by the Commission on Accreditation of Rehabilitation Facilities (CARF); or authorized or otherwise set up to participate in Medicaid (Medicaid state plan, Medicaid waiver, or Medicaid managed care) or part of a Program of All-Inclusive Center for the Elderly (PACE); 2) had average daily attendance of at least one participant based on a typical week; and 3) had at least one participant enrolled at the center at the time of the survey.

The residential care community sampling frame was constructed from lists of licensed residential care communities obtained from the state licensing agencies in each of the 50 states and the District of Columbia. The state lists were checked for duplicate RCCs and concatenated, resulting in a sampling frame of 42,149 residential care communities for the 2016 wave. The residential care

community component included a mix of sampled communities from states that had enough residential care communities to produce reliable state estimates and a census of residential care communities in states that did not have enough communities to produce reliable state estimates. A sample of 11,688 residential care communities were selected for the 2016 wave. Among states where a sample was selected, the primary sampling strata were defined by state and community bed size. To be eligible for the survey, residential communities had to be state-licensed with four or more licensed beds; provide room and board of at least two meals a day, around-the-clock supervision, and offer assistance with personal care (like dressing) or health-related services (such as medication management); have at least one resident; and serve primarily an adult population. Residential care communities licensed to exclusively serve the mentally ill or intellectually disabled/developmentally disabled populations were excluded.

Every nursing home, home health agency, or hospice in the United States that was certified to provide services under Medicare, Medicaid, or both, and had user data, was included in the data. Facility data was obtained from the CMS' administrative records in Certification and Survey Provider Enhanced Reporting ([CASPER], formerly known as Online Survey Certification and Reporting); the third quarter file of the data year was used. User data were obtained from Minimum Data Set Active Resident Episode Table (MARET) for nursing home sector, Medicare Provider Analysis and Review (MedPAR) for nursing home sector, Outcome-Based Quality Improvement (OBQI) Case Mix Roll Up from the Outcome and Assessment Information Set (OASIS) for home health sector, and Institutional Provider and Beneficiary Summary (IPBS) for home health and hospice sectors.

*Sample Size and Response Rates.* Every certified nursing home, home health agency, and hospice with user information, and all users during the data time frame, was included. The 2015–2016 NSLTCP file included administrative data on 12,208 active home health agencies, 4,348 hospices, and 15,638 nursing homes. Of the 5,348 adult day services centers on the sampling frame, 4,586 were found to be in-scope or were presumed in-scope; 2,836 of them completed the questionnaire, for a response rate of 61.8%. From 42,149 residential communities in the sampling frame, 11,688 residential care communities were sampled; 5,485 communities could not be contacted by the end of data collection and 4,578 completed the questionnaire. This yielded a weighted response rate of 50.7%.

*Issues Affecting Interpretation.* The estimates for adult day services center participants, nursing home residents, and residential care community residents are for current service users on any given day, rather than all users in a year. The estimate for home health patients includes only those who ended care in the prior year (discharges). The same person may be included in this sum more than once, if a person received care in more than one sector in a similar time period (e.g., a residential care resident receiving care from a home health agency). While every effort was made to match question wording in the NSLTCP surveys to the administrative

data available through CMS, some differences remained and may affect comparisons between these two data sources. For example, because not all LTC providers are residential, information on capacity is not comparable across provider types. In addition, different data sources used different reference periods. For instance, user-level data used for home health agencies and hospices were from patients who received home health or hospice care services at any time in the calendar year prior to the survey. In contrast, survey data on residential care community residents and adult day services center participants, and CMS data on nursing home residents, were from current users on any given day or active residents on the last day of the third quarter of the data year.

### References

Harris-Kojetin L, Sengupta M, Lendon JP, Rome V, Valverde R, Caffrey C. Long-term care providers and services users in the United States, 2015–2016. National Center for Health Statistics. *Vital Health Stat* 3(43). 2019. Available from: [https://www.cdc.gov/nchs/data/series/sr\\_03/sr03\\_43-508.pdf](https://www.cdc.gov/nchs/data/series/sr_03/sr03_43-508.pdf).

Harris-Kojetin L, Sengupta M, Park-Lee E, Valverde R, Caffrey C, Rome V, Lendon J. Long-term care providers and services users in the United States: Data from the National Study of Long-Term Care Providers, 2013–2014. National Center for Health Statistics. *Vital Health Stat* 3(38). 2016. Available from: [https://www.cdc.gov/nchs/data/series/sr\\_03/sr03\\_038.pdf](https://www.cdc.gov/nchs/data/series/sr_03/sr03_038.pdf).

Harris-Kojetin L, Sengupta M, Park-Lee E, Valverde R. Long-term care services in the United States: 2013 overview. National Center for Health Statistics. *Vital Health Stat* 3(37). 2013. Available from: [https://www.cdc.gov/nchs/data/series/sr\\_03/sr03\\_037.pdf](https://www.cdc.gov/nchs/data/series/sr_03/sr03_037.pdf).

For more information, see the NSLTCP website at: <https://www.cdc.gov/nchs/nsltcp/index.htm>.

## National Survey on Drug Use and Health (NSDUH)

### Substance Abuse and Mental Health Services Administration (SAMHSA)

*Overview.* NSDUH reports on the prevalence, incidence, and patterns of illicit drug use and alcohol use among the U.S. civilian noninstitutionalized population aged 12 and over. NSDUH also reports on substance use disorders, substance use treatment, mental health problems, and mental health care.

*Coverage.* NSDUH is representative of persons aged 12 and over in the civilian noninstitutionalized population of the United States, and in each state and the District of Columbia (D.C.).

The survey covers residents of households (including those living in houses, townhouses, apartments, and condominiums), persons in noninstitutional group quarters

(including those in shelters, boarding houses, college dormitories, migratory work camps, and halfway houses), and civilians living on military bases. Persons excluded from the survey include people experiencing homelessness who do not use shelters, active military personnel, and residents of institutional group quarters such as jails, nursing homes, mental institutions, and long-term care hospitals.

*Methodology.* Data are collected via in-person interviews conducted with a sample of individuals at their place of residence. Computer-assisted interviewing methods, including audio computer-assisted self-interviewing, are used to provide a private and confidential setting to complete the interview.

NSDUH uses a 50-state (and D.C.) sample design that is revised periodically. In 2014, NSDUH introduced an independent multistage area probability sample within each state and D.C. States are the first level of stratification, and each state was then stratified into approximately equally populated state sampling regions (SSRs). Census tracts within each SSR were then selected, followed by census block groups within census tracts and area segments (i.e., a collection of census blocks) within census block groups. Finally, dwelling units (DUs) were selected within segments, and within each selected DU, up to two residents who were at least 12 years old were selected for the interview.

Also starting in 2014, changes were made in the sample sizes allocated to each state and to different age groups, in order to increase the precision of national estimates, many state estimates, and estimates for older adults. States with sample increases will have more precise estimates than in previous years, whereas states with smaller sample sizes will have some reductions in precision. However, all states will still have reasonable levels of precision. This allocation of sample to states is also thought to be more cost efficient. Starting in 2014, the sample size was redistributed by age group so that 25% of the sample is allocated to those aged 12–17, 25% to those aged 18–25, and 50% to those aged 26 and over. Although the sample sizes for age groups 12–17 and 18–25 were reduced, these two groups are still considered to be oversampled since they represent approximately 10% and 13% of the total population, respectively.

*Sample Size and Response Rate.* In 2017, screening was completed at 138,061 addresses, and 68,032 completed interviews were obtained, including 17,033 interviews from adolescents aged 12–17 and 50,999 interviews from adults aged 18 and over. Weighted response rates for household screening and for interviewing were 75.1% and 67.1%, respectively, for an overall response rate of 50.4% for people aged 12 and over.

*Issues Affecting Interpretation.* Several improvements to NSDUH were implemented in 2002. The data collected in 2002 represent a new baseline for tracking trends in substance use and other measures. Special questions on methamphetamine were added in 2005 and 2006. Data for years prior to 2007 were adjusted for comparability. Starting with 2011 data, 2010 Census-based control totals were used in the weighting process. Analysis weights in the 2002 through 2010 NSDUHs were derived from the 2000 Census

data. This reweighting to the 2010 Census data could affect comparisons between estimates for 2011 and subsequent years and those from prior years. However, an analysis of the impact of reweighting showed that the percentages of substance users were largely unaffected. For more information, see: <https://archive.samhsa.gov/data/NSDUH/NSDUHCensusEffects/Index.aspx>.

The NSDUH questionnaire underwent a partial redesign in 2015 to improve the quality of data and to address the changing needs of policymakers and researchers with regard to substance use and mental health issues. Due to the changes, only 2015, 2016, and 2017 data are presented for certain estimates until comparability with prior years can be established. Trends continue to be presented for estimates that are assumed to have remained comparable with those in earlier years. For more information, see: <https://www.samhsa.gov/data/sites/default/files/NSDUH-TrendBreak-2015.pdf>.

Estimates of substance use for youth based on NSDUH are not directly comparable with estimates based on the Monitoring the Future (MTF) Study and the Youth Risk Behavior Surveillance System (YRBSS). In addition to the fact that MTF excludes dropouts and absentees, rates are not directly comparable across these surveys because of differences in the populations covered, sample design, questionnaires, and interview setting. NSDUH collects data in residences, whereas MTF and YRBSS collect data in school classrooms. Furthermore, NSDUH estimates are tabulated by age, whereas MTF and YRBSS estimates are tabulated by grade, representing different ages as well as different populations.

### References

Bose J, Hedden SL, Lipari RN, Park-Lee E. Key substance use and mental health indicators in the United States: Results from the 2017 National Survey on Drug Use and Health. 2018. Available from: <https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHFFR2017/NSDUHFFR2017.pdf>.

Substance Abuse and Mental Health Services Administration. Results from the 2017 National Survey on Drug Use and Health: Detailed tables. 2018. Available from: <https://www.samhsa.gov/data/sites/default/files/cbhsq-reports/NSDUHDetailedTabs2017/NSDUHDetailedTabs2017.pdf>

For more information, see the NSDUH website at: <https://www.samhsa.gov/data/data-we-collect/nsduh-national-survey-drug-use-and-health>, and the Center for Behavioral Health Statistics and Quality (the data collection agency) website at: <https://www.samhsa.gov/about-us/who-we-are/offices-centers/cbhsq>.

## National Vital Statistics System (NVSS)

### National Center for Health Statistics (NCHS)

*Overview.* NVSS collects and publishes official national statistics on births, deaths, fetal deaths, and, prior to 1996, marriages and divorces occurring in the United States, based on U.S. Standard Certificates. Fetal deaths are classified and tabulated separately from other deaths. The vital statistics files—Birth, Fetal Death, Mortality Multiple Cause-of-Death, Linked Birth/Infant Death, and Compressed Mortality—are described in detail below.

*Coverage.* NVSS collects and presents U.S. resident data for the aggregate of 50 states, New York City, and the District of Columbia (D.C.), as well as for each individual state, D.C., and the U.S.-dependent areas of Puerto Rico, Virgin Islands, Guam, American Samoa, and Northern Marianas. Vital events occurring in the United States to non-U.S. residents and vital events occurring abroad to U.S. residents are excluded. Starting with *Health, United States, 2013*, information on vital events for Puerto Rico, Virgin Islands, Guam, American Samoa, and Northern Marianas is shown in selected state tables but is not included in U.S. totals.

*Methodology.* NCHS' Division of Vital Statistics obtains information on births and deaths from the registration offices of each of the 50 states, New York City, D.C., Puerto Rico, Virgin Islands, Guam, American Samoa, and Northern Marianas. Until 1972, microfilm copies of all death certificates and a 50% sample of birth certificates were received from all registration areas and processed by NCHS. In 1972, some states began sending their data to NCHS through the Cooperative Health Statistics System (CHSS). States that participated in the CHSS program processed 100% of their death and birth records and sent the entire data file to NCHS on computer tapes. Currently, data are sent to NCHS following procedures similar to those under CHSS. The number of participating states grew from 6 in 1972 to 46 in 1984. Starting in 1985, all 50 states and D.C. participated in the Vital Statistics Cooperative Program.

*U.S. Standard Certificates.* U.S. Standard Certificates of Live Birth and Death and Fetal Death Reports are revised periodically, allowing evaluation and addition, modification, and deletion of items. Beginning with 1989, revised Standard Certificates replaced the 1978 versions. The 1989 revision of the death certificate included items on educational attainment and Hispanic origin of decedents, as well as changes to improve the medical certification of cause of death. Standard Certificates recommended by NCHS are modified in each registration area to serve the area's needs. However, most certificates conform closely in content and arrangement to the Standard Certificate, and all certificates contain a minimum data set specified by NCHS. The 2003 revision of vital records went into effect in some states and territories beginning in 2003; full implementation in all states, D.C., and territories (other than American Samoa) will be achieved with 2018 data. The 2003 revision of the death certificate included changes in the determination of multiple

racess, education level, prenatal care, tobacco use, and maternal mortality.

## **Birth File**

*Overview.* Vital statistics natality data are a fundamental source of demographic, geographic, and medical and health information on all births occurring in the United States. This is one of the few sources of comparable health-related data for small geographic areas over an extended time period. The data are used to present the characteristics of babies and their mothers, track trends such as birth rates for teenagers, and compare natality trends with those in other countries.

The Birth file includes characteristics of the baby, such as sex, birthweight, and weeks of gestation; demographic information about the parents, such as age, race, Hispanic origin, parity, educational attainment, marital status, and state of residence; medical and health information, such as prenatal care, based on hospital records; and behavioral risk factors for the birth, such as mother's tobacco use during pregnancy.

*Coverage.* Birth data presented in *Health, United States* are based on reporting from all 50 states and D.C. Data for Alaska have been included starting in 1959, while data for Hawaii have been included starting in 1960 after each received statehood. Data for Puerto Rico, Virgin Islands, Guam, American Samoa, and Northern Marianas are shown in selected state tables but are not included in U.S. totals. Beginning with 1970, births to nonresidents of the United States are excluded.

*Methodology.* In the United States, state laws require birth certificates to be completed for all births. The registration of births is the responsibility of the professional attendant at birth, generally a physician or midwife. The birth certificate must be filed with the local registrar of the district in which the birth occurs. Each birth must be reported promptly; the reporting requirements vary from state to state, ranging from 24 hours to as much as 10 days after the birth.

Federal law mandates national collection and publication of birth and other vital statistics data. NVSS is the result of cooperation between NCHS and the states to provide access to statistical information from birth certificates. Standard forms for the collection of the data, and model procedures for the uniform registration of the events, are developed and recommended for state use through cooperative activities of the states and NCHS. NCHS shares the costs incurred by the states in providing vital statistics data for national use.

*Issues Affecting Interpretation.* The 2003 revision of the birth certificate was phased in from 2003 to 2017. As of January 1, 2016, all states, territories (except American Samoa), and reporting areas had adopted the 2003 revision of the U.S. Standard Certificate of Live Birth. The 2003 certificate uses revised race and ethnicity sections conforming to the 1997 *Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity*. However, to provide uniformity and comparability of data for trend

comparison, bridged race categories are still presented in *Health, United States*. Interpretation of trend data should take into consideration changes to reporting areas. For methodological and reporting area changes for the following birth certificate items, see [Appendix II](#), Age; Hispanic origin; Marital status; Race.

## **Reference**

Martin JA, Hamilton BE, Osterman MJK, Driscoll AK, Drake P. Births: Final data for 2017. National Vital Statistics Reports; vol 67 no 8. Hyattsville, MD: National Center for Health Statistics. 2018. Available from: [https://www.cdc.gov/nchs/data/nvsr/nvsr67/nvsr67\\_08-508.pdf](https://www.cdc.gov/nchs/data/nvsr/nvsr67/nvsr67_08-508.pdf).

For more information, see the NVSS Birth Data website at: <https://www.cdc.gov/nchs/nvss/births.htm>, and Vitalstats at: [https://www.cdc.gov/nchs/data\\_access/Vitalstatsonline.htm](https://www.cdc.gov/nchs/data_access/Vitalstatsonline.htm).

## **Fetal Death Data Set**

*Overview.* Fetal mortality refers to the intrauterine death of a fetus at any gestational age. In *Health, United States*, data are presented for fetal deaths at 20 weeks or more. The Fetal Death data set includes characteristics of the fetus, such as sex, birthweight, and weeks of gestation; demographic information about the mother, such as age, race, Hispanic origin, and live-birth order; and medical and health information, such as maternal diabetes and hypertension.

*Coverage.* Data presented in *Health, United States* are based on reporting from all 50 states and D.C. Data for Puerto Rico, Virgin Islands, Guam, American Samoa, and Northern Marianas are not included in U.S. totals but are included in the fetal death user's guides, available from the NCHS website at: [https://www.cdc.gov/nchs/data\\_access/VitalStatsOnline.htm](https://www.cdc.gov/nchs/data_access/VitalStatsOnline.htm), and in periodic reports.

*Methodology.* Fetal death means the death of a fetus prior to delivery from the mother, irrespective of the duration of pregnancy. Fetal deaths do not include induced terminations of pregnancy. This definition of fetal death, adopted by NCHS as the nationally recommended standard, is based on the definition published by the World Health Organization (WHO) in 1950 and revised in 1988. The term fetal death encompasses other commonly used terms, including stillbirth, spontaneous abortion, and miscarriage. All U.S. states and registration areas have definitions similar to the standard definition, except for Puerto Rico and Wisconsin, which have no formal definition.

State laws require the reporting of fetal deaths, and federal law mandates national collection and publication of fetal death data. States and reporting areas submit fetal mortality data to NCHS as part of a cooperative agreement. Standard forms and procedures for the collection of the data are developed and recommended for state use through cooperative activities of the states and NCHS. NCHS shares the costs incurred by the states in providing vital statistics data for national use.

In addition to fetal mortality rates, perinatal mortality rates are also presented in *Health, United States*. Perinatal mortality includes both late fetal deaths (of at least 28 weeks of gestation) and early infant (neonatal) deaths (within 7 days of birth). Data on early infant deaths come from the Linked Birth/Infant Death data set.

**Issues Affecting Interpretation.** Reporting requirements for fetal deaths vary by state, and these differences have important implications for comparisons of fetal mortality rates by state. The majority of states require reporting of fetal deaths at 20 weeks of gestation or more, or a minimum of 350 grams birthweight (roughly equivalent to 20 weeks), or some combination of the two. In 2017, six states required reporting of fetal deaths at all periods of gestation, two states required reporting beginning at 12 weeks of gestation, and one required reporting beginning at 16 weeks of gestation. Further, one state required the reporting of fetal deaths with birthweights of 500 grams or more (roughly equivalent to 22 weeks of gestation).

Starting with 2014 data, the obstetric estimate of gestation at delivery (OE) is used to determine gestational age, instead of the last normal menses (LMP), which was used for earlier years. The adoption of OE for gestational age had no or negligible impact on total fetal mortality rates. However, late fetal mortality rates based on the OE were lower than those based on the LMP. For more information, see User Guide to the 2017 Fetal Death Public Use File at: [ftp://ftp.cdc.gov/pub/Health\\_Statistics/NCHS/Dataset\\_Documentation/DVS/fetaldeath/2017FetalUserGuide.pdf](ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/DVS/fetaldeath/2017FetalUserGuide.pdf).

There is substantial evidence that not all fetal deaths for which reporting is required are, in fact, reported. Underreporting of fetal deaths is most likely to occur in the earlier part of the required reporting period for each state. For example, in 2013, for those states requiring reporting of fetal deaths at all periods of gestation, 56.4% of fetal deaths at 20 weeks of gestation or more were at 20–27 weeks, whereas for states requiring reporting of fetal deaths of 500 grams or more, only 33.8% were at 20–27 weeks, thus indicating substantial underreporting of early fetal deaths in some states.

### References

MacDorman MF, Gregory ECW. Fetal and perinatal mortality: United States, 2013. National Vital Statistics Reports; vol 64 no 8. Hyattsville, MD: National Center for Health Statistics. 2015. Available from: [https://www.cdc.gov/nchs/data/nvsr/nvsr64/nvsr64\\_08.pdf](https://www.cdc.gov/nchs/data/nvsr/nvsr64/nvsr64_08.pdf).

Gregory ECW, MacDorman MF, Martin JA. Trends in fetal and perinatal mortality in the United States, 2006–2012. NCHS Data Brief, no 169. Hyattsville, MD: National Center for Health Statistics. 2014. Available from: <https://www.cdc.gov/nchs/data/databriefs/db169.pdf>.

For more information, see the NCHS Fetal Deaths data website at: [https://www.cdc.gov/nchs/fetal\\_death.htm](https://www.cdc.gov/nchs/fetal_death.htm).

## Mortality Multiple Cause-of-Death File

**Overview.** Vital statistics mortality data are a fundamental source of demographic, geographic, and underlying and multiple cause-of-death information. Multiple cause-of-death data reflect all medical information reported on death certificates and complement traditional underlying cause-of-death data. Multiple-cause data give information on diseases that are a factor in death, whether or not they are the underlying cause of death; on associations among diseases; and on injuries leading to death.

The Mortality multiple cause-of-death file includes demographic information on age, sex, race, Hispanic origin, state of residence, and educational attainment, as well as medical information on causes of death. This data set is one of the few sources of comparable health-related data for small geographic areas over an extended time period. The data are used to present the characteristics of those dying in the United States, to determine life expectancy, and to compare mortality trends with those in other countries.

**Coverage.** Mortality data presented in *Health, United States* are based on reporting from all 50 states and D.C. Data for Alaska have been included starting in 1959, while data for Hawaii have been included starting in 1960 after each received statehood. Data for Puerto Rico, Virgin Islands, Guam, American Samoa, and Northern Marianas are shown in selected state tables, but are not included in U.S. totals. Beginning with 1970, mortality statistics for the United States exclude deaths of U.S. nonresidents. Mortality statistics for Puerto Rico, Virgin Islands, American Samoa, and Northern Marianas excluded deaths of nonresidents for each area. For Guam, mortality statistics exclude deaths that occurred to a resident of any place other than Guam or the United States (50 states and D.C.).

**Methodology.** By law, the registration of deaths is the responsibility of the funeral director. The funeral director obtains demographic data for the death certificate from an informant. The physician in attendance at the death is required to certify the cause of death. Where cause of death is from other than natural causes, a coroner or medical examiner may be required to examine the body and certify the cause of death. For more information on the role of the medical examiner and coroner systems for data quality, see Miniño et al.

NCHS is responsible for compiling and publishing annual national statistics on causes of death. In carrying out this responsibility, NCHS adheres to WHO Nomenclature Regulations. These regulations require (a) that cause of death be coded in accordance with the applicable revision of the *International Classification of Diseases* (ICD) (see [Appendix II, International Classification of Diseases \[ICD\]; Table III](#)); and (b) that underlying cause of death be selected in accordance with international rules. Traditionally, national mortality statistics have been based on a count of deaths, with one underlying cause assigned for each death.

Prior to 1968, mortality medical data were based on manual coding of an underlying cause of death for each certificate in accordance with WHO rules. Starting with 1968,

NCHS converted to computerized coding of the underlying cause and manual coding of all causes (multiple causes) on the death certificate. In this system, called Automated Classification of Medical Entities (ACME), multiple-cause codes serve as inputs to the computer software, which employs WHO rules to select the underlying cause. ACME is used to select the underlying cause of death for all death certificates in the United States, and cause-of-death data in *Health, United States* are coded using ACME.

In addition, NCHS has developed two computer systems as inputs to ACME. Beginning with 1990 data, the Mortality Medical Indexing, Classification, and Retrieval system (MICAR) was introduced to automate coding of multiple causes of death. MICAR provides more detailed information on the conditions reported on death certificates than is available through the ICD code structure. Then, beginning with data year 1993, SuperMICAR, an enhancement of MICAR, was introduced. SuperMICAR allows for literal entry of the multiple cause-of-death text as reported by the certifier. This information is then processed automatically by the MICAR and ACME computer systems. Records that cannot be processed automatically by MICAR or SuperMICAR are multiple-cause-coded manually and then further processed through ACME. Starting in 2003, SuperMICAR was used to process all of the country's death records.

Data for the entire United States refer to events occurring within the 50 states and D.C.; data for geographic areas are by place of residence. For methodological and reporting area changes for the following death certificate items, see [Appendix II](#), Hispanic origin; Race.

*Issues Affecting Interpretation.* ICD, by which cause of death is coded and classified, is revised approximately every 10 to 20 years. Because revisions of ICD may cause discontinuities in trend data by cause of death, comparison of death rates by cause of death across ICD revisions should be done with caution and with reference to the comparability ratio (see [Appendix II](#), Comparability ratio). Prior to 1999, modifications to ICD were made only when a new revision of ICD was implemented. A process for updating ICD was introduced with the 10th revision (ICD–10) that allows for midrevision changes. These changes, however, may affect comparability of data between years for select causes of death. Minor changes may be implemented every year, whereas major changes may be implemented every 3 years (e.g., 2003 data year). In data year 2006, major changes were implemented, including the addition and deletion of several ICD codes. For more information, see Heron et al.

Multiple-cause data were obtained from all certificates for 1968–1971, 1973–1980, and 1983–present. Data were obtained from a 50% sample of certificates for 1972. Multiple-cause data for 1981 and 1982 were obtained from a 50% sample of certificates from 19 registration areas. For the other states, data were obtained from all certificates.

The death certificate has been revised periodically. A revised U.S. Standard Certificate of Death was recommended for state use beginning January 1, 1989. Among the changes were the addition of a new item on educational attainment and Hispanic origin of the decedent and changes to improve

the medical certification of cause of death. The U.S. Standard Certificate of Death was revised again in 2003; states are adopting this new certificate on a rolling basis.

The 2003 revision permits reporting of more than one race (multiple races). This change was implemented to reflect the increasing diversity of the U.S. population and to be consistent with the decennial census. Some states, however, are still using the 1989 revision of the U.S. Standard Certificate of Death, which allows only a single race to be reported. Until all states adopt the new death certificate, the race data reported using the 2003 revision are “bridged” for those for whom more than one race was reported (multiple race) to one single race, to provide comparability with race data reported on the 1989 revision. For more information on the impact of the 2003 certificate revisions on mortality data presented in *Health, United States*, see [Appendix II](#), Race.

### References

Kochanek KD, Murphy SL, Xu JQ, Arias E. Deaths: Final data for 2017. *Natl Vital Stat Rep*; vol 68 no 9. Hyattsville, MD: National Center for Health Statistics. 2019. Available from: [https://www.cdc.gov/nchs/data/nvsr/nvsr68/nvsr68\\_09-508.pdf](https://www.cdc.gov/nchs/data/nvsr/nvsr68/nvsr68_09-508.pdf).

Heron MP, Hoyert DL, Murphy SL, Xu JQ, Kochanek KD, Tejada-Vera B. Deaths: Final data for 2006. *Natl Vital Stat Rep*; vol 57 no 14. Hyattsville, MD: National Center for Health Statistics. 2009. Available from: [https://www.cdc.gov/nchs/data/nvsr/nvsr57/nvsr57\\_14.pdf](https://www.cdc.gov/nchs/data/nvsr/nvsr57/nvsr57_14.pdf).

Miniño AM, Anderson RN, Fingerhut LA, Boudreault MA, Warner M. Deaths: Injuries, 2002. *Natl Vital Stat Rep*; vol 54 no 10. Hyattsville, MD: National Center for Health Statistics. 2006. Available from: [https://www.cdc.gov/nchs/data/nvsr/nvsr54/nvsr54\\_10.pdf](https://www.cdc.gov/nchs/data/nvsr/nvsr54/nvsr54_10.pdf).

National Center for Health Statistics. Multiple causes of death in the United States. *Monthly Vital Stat Report*; vol 32 no 10 suppl 2. Hyattsville, MD: National Center for Health Statistics. 1984. Available from: [https://www.cdc.gov/nchs/data/mvsr/supp/mv32\\_10s2.pdf](https://www.cdc.gov/nchs/data/mvsr/supp/mv32_10s2.pdf).

For more information, see the NCHS Mortality Data website at: <https://www.cdc.gov/nchs/deaths.htm>.

### Linked Birth/Infant Death Data Set

*Overview.* National linked files of live births and infant deaths are used for research on infant mortality. The Linked Birth/Infant Death data set links information from the birth certificate to information from the death certificate for each infant death in the United States. The purpose of the linkage is to use the many additional variables from the birth certificate, including the more accurate race and ethnicity data, for more detailed analyses of infant mortality patterns. The Linked Birth/Infant Death data set includes all variables on the natality (Birth) file, including racial and ethnic information, birthweight, and maternal smoking, as well as variables on the Mortality file, including cause of death and age at death.

**Coverage.** To be included in the U.S.-linked file, both the birth and death must have occurred in the 50 states, D.C., Puerto Rico, Virgin Islands, or Guam. Data for Puerto Rico, Virgin Islands, and Guam are shown in selected state tables but are not included in U.S. totals. Linked birth/infant death data are not available for American Samoa and Northern Marianas.

**Methodology.** Infant deaths are defined as a death before the infant's first birthday. About 98%–99% of infant death records can be linked to their corresponding birth certificates. The linkage makes available extensive information from the birth certificate about the pregnancy, maternal risk factors, infant characteristics, and health items at birth that can be used for more detailed analyses of infant mortality. The linked file is used for calculating infant mortality rates by race and ethnicity, which are more accurately measured from the birth certificate.

Starting with 1995 data, linked birth/infant death data files are available in two different formats: period data and birth cohort data. The numerator for the period linked file consists of all infant deaths occurring in a given data year linked to their corresponding birth certificates, whether the birth occurred in that year or the previous year. The numerator for the birth cohort linked file consists of deaths to infants born in a given year. In both cases, the denominator is all births occurring in the year. For example, the 2013 period linked file contains a numerator file that consists of all infant deaths occurring in 2013 that have been linked to their corresponding birth certificates, whether the birth occurred in 2012 or 2013. In contrast, the 2013 birth cohort linked file will contain a numerator file that consists of all infant deaths to babies born in 2013, whether the death occurred in 2013 or 2014. Although the birth cohort format has methodological advantages, it creates substantial delays in data availability because it is necessary to wait until the close of the following data year to include all infant deaths in the birth cohort. Starting with 1995 data, period linked files are used for infant mortality rate tables in *Health, United States*.

Other changes to the data set starting with 1995 include the addition of record weights to compensate for the 1%–2% of infant death records that could not be linked to their corresponding birth records. In addition, not-stated birthweight was imputed if the period of gestation was known. This imputation was done to improve the accuracy of birthweight-specific infant mortality rates because the percentage of records with not-stated birthweight is generally higher for infant deaths (4.09% in 2017) than for live births (0.07% in 2017). In 2017, not-stated birthweight was imputed for 0.06% of births.

**Issues Affecting Interpretation.** Period linked file data starting with 1995 are not strictly comparable with birth cohort data for 1983–1991. A new revision of the birth certificate was introduced in 2003, and as of 2016 data, all states and reporting areas (except for American Samoa) had adopted the 2003 version of the birth certificate.

## References

Mathews TJ, Driscoll AK. Trends in infant mortality in the United States, 2005–2014. NCHS Data Brief, no 279. Hyattsville, MD: National Center for Health Statistics. 2017. Available from: <https://www.cdc.gov/nchs/data/databriefs/db279.pdf>.

Mathews TJ, MacDorman MF, Thoma ME. Infant mortality statistics from the 2013 period linked birth/infant death data set. National Vital Statistics Reports; vol 64 no 9. Hyattsville, MD: National Center for Health Statistics. 2015. Available from: [https://www.cdc.gov/nchs/data/nvsr/nvsr64/nvsr64\\_09.pdf](https://www.cdc.gov/nchs/data/nvsr/nvsr64/nvsr64_09.pdf).

For more information, see the NCHS Linked Birth and Infant Death Data website at: <https://www.cdc.gov/nchs/linked.htm>.

## National Youth Tobacco Survey (NYTS)

### Centers for Disease Control and Prevention (CDC), Office on Smoking and Health (OSH) and U.S. Food and Drug Administration (FDA), Center for Tobacco Products (CTP)

**Overview.** NYTS is an annual school-based survey of U.S. middle and high school students that collects data on tobacco use and tobacco-related attitudes, beliefs, and influences. Data collected include use of cigarettes, cigars, smokeless tobacco, e-cigarettes, hookahs, pipe tobacco, and bidis within the past 30 days.

**Coverage.** Data are nationally representative of 6th through 12th graders in public and private middle and high schools in the United States.

**Methodology.** Prior to 2011, the survey was administered on an approximately biennial basis. From 2011–2018, the survey has been conducted annually. The survey uses a three-stage cluster sampling design to generate a nationally representative sample of U.S. students attending public and private schools in grades 6 through 12. Data are collected using a voluntary, school-based, self-administered, pencil-and-paper questionnaire. Make-up surveys are pursued for eligible students who are absent on the scheduled date of data collection. Ineligible students, including those who had moved and were no longer attending the school, dropouts, those with extended absences, and students who could not complete the survey independently, were excluded from survey participation. Data were weighted to account for the complex survey design and adjusted for nonresponse.

**Sample Size and Response Rate.** In 2018, a total of 20,189 students from 238 public and private schools in the United States participated, with an overall response rate of 68%. The 20,189 participants were broken down by grade as follows: 2,568 12th graders, 2,824 11th graders, 2,664 10th graders, 2,935 9th graders, 3,012 8th graders, 3,140 7th graders, and 2,903 6th graders (143 students were missing data on self-reported grade level). During 2011–2018, sample

sizes ranged from a low of 17,711 (in 2015) to a high of 24,658 (in 2012), and response rates ranged from a low of 63% (2015) to a high of 74% (2012).

*Issues Affecting Interpretation.* Estimates of substance use among youth based on NYTS are not directly comparable with estimates based on the National Survey on Drug Use and Health (NSDUH), the Monitoring the Future (MTF) Study, and the Youth Risk Behavior Surveillance System (YRBSS). This is because of differences in populations covered, sample design, questionnaires, interview settings, and data cleaning procedures. NSDUH collects data in residences, whereas NYTS, MTF, and YRBSS collect data in school classrooms. In addition, NSDUH estimates are tabulated by age, whereas NYTS, MTF, and YRBSS estimates are tabulated by school level or grade, representing different ages as well as different populations.

### References

Gentzke AS, Creamer M, Cullen KA, Ambrose BK, Willis G, Jamal A, King BA. Vital signs: Tobacco product use among middle and high school students—United States, 2011–2018. *MMWR Morb Mortal Wkly Rep* 68(6):157–64. 2019. Available from: <https://www.cdc.gov/mmwr/volumes/68/wr/mm6806e1.htm>.

Centers for Disease Control and Prevention. 2018 National Youth Tobacco Survey: Methodology report. Available from: [https://www.cdc.gov/tobacco/data\\_statistics/surveys/nyts/data/index.html](https://www.cdc.gov/tobacco/data_statistics/surveys/nyts/data/index.html).

For more information, see the NYTS website at: [https://www.cdc.gov/tobacco/data\\_statistics/surveys/nyts/index.htm](https://www.cdc.gov/tobacco/data_statistics/surveys/nyts/index.htm).

## Population Census and Population Estimates

### U.S. Census Bureau

#### Decennial Census

The census of population (decennial census) has been held in the United States every 10 years since 1790. Since 1930, it has enumerated the resident population as of April 1 of the census year. Data on sex, race, Hispanic origin, age, and marital status are collected from 100% of the enumerated population.

#### Race Data on the 1990 Census

The question on race on the 1990 Census was based on the Office of Management and Budget's (OMB) 1977 *Race and Ethnic Standards for Federal Statistics and Administrative Reporting* (Statistical Policy Directive 15). This document specified rules for the collection, tabulation, and reporting of racial and ethnic data within the federal statistical system. The 1977 Standards required federal agencies to report

race-specific tabulations using four single-race categories: American Indian or Alaska Native, Asian or Pacific Islander, black, and white. Under the 1977 Standards, race and ethnicity were considered to be two separate and distinct concepts. Thus, persons of Hispanic origin may be of any race.

#### Race Data on the 2000 Census

The question on race on the 2000 Census was based on OMB's 1997 *Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity* (see: Federal Register 62(210):58782–90. 1997.). (Also see [Appendix II](#), Race.) The 1997 Standards incorporated two major changes in the collection, tabulation, and presentation of race data. First, the 1997 Standards increased the minimum set of categories to be used by federal agencies for identification of race from four to five: American Indian or Alaska Native, Asian, black or African American, Native Hawaiian or Other Pacific Islander, and white. Second, the 1997 Standards included the requirement that federal data collection programs allow respondents to select one or more race categories when responding to a query on their racial identity. This provision means that there are potentially 31 race groups, depending on whether an individual selects one, two, three, four, or all five of the race categories. The 1997 Standards continue to call for use, when possible, of a separate question on Hispanic or Latino ethnicity and specify that the ethnicity question should appear before the question on race. Thus, under the 1997 Standards, as under the 1977 Standards, persons of Hispanic origin may be of any race.

#### Race Data on the 2010 Census

Similar to race data on the 2000 Census, the question on race on the 2010 Census was based on OMB's 1997 *Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity* (see: Federal Register 62(210):58782–90. 1997.). (Also see [Appendix II](#), Race.) The 1997 Standards required a minimum set of categories to be used by federal agencies for identification of race: American Indian or Alaska Native, Asian, black or African American, Native Hawaiian or Other Pacific Islander, and white, and require that federal data collection programs allow respondents to select one or more race categories when responding to a query on their racial identity. The 1997 Standards continue to call for use, when possible, of a separate question on Hispanic or Latino ethnicity and specify that the ethnicity question should appear before the question on race. Thus, under the 1997 Standards, as under the 1977 Standards, persons of Hispanic origin may be of any race.

#### Modified Decennial Census Files

For several decades, the U.S. Census Bureau has produced Modified Decennial Census files. These modified files incorporate adjustments to the 100% April 1 count data for (a) errors in the census data discovered subsequent to



publication, (b) misreported age data, and (c) nonspecified race.

For the 1990 Census, the U.S. Census Bureau modified the age, race, and sex data on the census and produced the Modified Age-Race-Sex (MARS) file. The differences between the population counts in the original census file and the MARS file are primarily due to modification of the race data. Of the 248.7 million persons enumerated in 1990, 9.8 million did not specify their race (more than 95% were of Hispanic origin). For the 1990 MARS file, these persons were assigned the race reported by a nearby person with an identical response to the Hispanic-origin question.

For the 2000 and 2010 Censuses, the U.S. Census Bureau modified the race data and produced the Modified Race Data Summary files. For these files, persons who did not report a race (reported only the category “Some other race”) as part of their race response were assigned by imputation to one of the 31 race groups, which are the single- and multiple-race combinations of the five race categories specified in the 1997 OMB race and ethnicity standards. For the 2000 Census, 97% of the 15.4 million persons who did not report a race were of Hispanic origin. Because a large proportion of those identifying their race as “Some other race” are Hispanic, for the 2010 Census, a new instruction was added that, for the census, Hispanic origins are not races. For the 2010 Census, 97% of the 19.1 million persons who did not report a race (reported only the category “Some other race”) were of Hispanic origin.

### ***Postcensal Population Estimates***

Postcensal population estimates are estimates made for the years following a census, before the next census is taken. Postcensal population estimates are derived annually by updating the resident population enumerated in the decennial census using a components-of-population-change approach. Each annual series includes estimates for the current data year and revised estimates for the earlier years in the decade. The following formula is used to derive national estimates for a given year from those for the previous year, starting with the decennial census enumerated resident population as the base:

Resident population estimate  
+ births to U.S. resident women  
– deaths to U.S. residents  
+ net international migration

The postcensal estimates are consistent with official decennial census figures and do not reflect estimated decennial census under-enumeration.

Estimates for the earlier years in a given series are revised to reflect changes in the components-of-change data sets (for example, births to U.S. resident women from a preliminary natality file are replaced with counts from a final natality file). To help users keep track of which postcensal estimate is being used, each annual series is referred to as “vintage,” and the last year in the series is used to name the

series. For example, both the vintage 2011 and the vintage 2012 postcensal series have revised estimates for July 1, 2011, but the estimates for July 1, 2011, from the vintage 2011 and vintage 2012 postcensal series differ.

The U.S. Census Bureau also produces postcensal estimates of the resident population of states and counties, using the components-of-population-change method. An additional component of population change—net internal migration—is involved.

### ***Intercensal Population Estimates***

Intercensal population estimates are estimates made for the years between two decennial censuses and are produced once the census at the end of the decade has been completed. They replace the postcensal estimates produced prior to the completion of the census at the end of the decade. Intercensal estimates are more accurate than postcensal estimates because they are based on both the census at the beginning and the census at the end of the decade. They are derived by adjusting the final postcensal estimates for the decade to correct for the error of closure (the difference between the estimated population at the end of the decade and the census count for that date). The patterns of population change observed over the decade are preserved. The intercensal estimates for the 1990s were produced using the same methodology used to generate the intercensal estimates for the 1980s. The revised intercensal population estimates for 2000–2009 were produced using a modified version of the methodology used previously. Vital rates calculated using postcensal population estimates are routinely revised when intercensal estimates become available.

### ***Bridged-race Population Estimates***

During the transition to full implementation of the 1997 OMB standards on race and ethnicity, race data on the 2000 and 2010 Censuses were not comparable with race data on other data systems that were continuing to collect data using the 1977 OMB Standards on race and ethnicity. For example, states implemented the revised birth and death certificates—which have race and ethnicity items that are compliant with the 1997 OMB Standards—at different times and some states still used the 1989 death certificates that collect race and ethnicity data in accordance with the 1977 OMB Standards. Thus, population estimates for 1990 and beyond with race categories comparable with the 1977 OMB categories were needed so that race-specific birth and death rates can be calculated. To meet this need, the National Center for Health Statistics (NCHS), in collaboration with the U.S. Census Bureau, developed methodology to bridge the 31 race groups in Census 2000 and Census 2010 to the four single-race categories specified under the 1977 OMB Standards. As of January 1, 2016, all states and D.C., in addition to Puerto Rico, the U.S. Virgin Islands, Guam, and Northern Marianas, use the 2003 revision of the U.S. Standard Certificate of Live Birth and report race according

to the 1997 revised OMB standards. However, to provide uniformity and comparability of data for trend comparison, bridged single-race categories are still presented in *Health, United States*.

The bridging methodology was developed using information from the 1997–2000 National Health Interview Survey (NHIS). NHIS provides a unique opportunity to investigate multiple-race groups because, since 1982, it has allowed respondents to choose more than one race but has also asked respondents reporting multiple races to choose a primary race. The bridging methodology developed by NCHS involved the application of regression models relating person-level and county-level covariates to the selection of a particular primary race by the multiple-race respondents. The bridging proportions derived from these models have been applied by the U.S. Census Bureau to various unbridged resident population files. These applications have resulted in bridged-race population estimates for each of the four single-race categories: American Indian or Alaska Native, Asian or Pacific Islander, black, and white.

In *Health, United States*, vital rates for 1991–1999 were calculated using the July 1, 1991–July 1, 1999 bridged-race intercensal estimates. Vital rates for 2000 were calculated using the bridged-race April 1, 2000, census counts, and those for 2010 were calculated using the bridged-race April 1, 2010, census counts. Starting with *Health, United States*, 2012, vital rates for 2001–2009 have been recalculated using the July 1, 2001–July 1, 2009, revised intercensal bridged-race population estimates. Vital rates for 2011 and beyond are calculated using bridged-race estimates of the July 1 population from the corresponding postcensal vintage.

### Reference

Ingram DD, Parker JD, Schenker N, Weed JA, Hamilton B, Arias E, Madans JH. United States Census 2000 population with bridged race categories. National Center for Health Statistics. Vital Health Stat 2(135). 2003. Available from: [https://www.cdc.gov/nchs/data/series/sr\\_02/sr02\\_135.pdf](https://www.cdc.gov/nchs/data/series/sr_02/sr02_135.pdf).

For more information, see the U.S. Census Bureau website at: <https://www.census.gov>, and the NCHS website for U.S. census populations with bridged-race categories at: [https://www.cdc.gov/nchs/nvss/bridged\\_race.htm](https://www.cdc.gov/nchs/nvss/bridged_race.htm).

## Sexually Transmitted Disease (STD) Surveillance

### Centers for Disease Control and Prevention (CDC), National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention (NCHHSTP)

*Overview.* Surveillance information on the incidence and prevalence of STDs is used to inform public and private health efforts to control these diseases. Case reporting data are available for nationally notifiable diseases, including chancroid, chlamydia, gonorrhea, and syphilis. Enhanced

surveillance of these conditions and surveillance of other STDs or STD sequelae, such as genital herpes simplex virus, genital warts, and trichomoniasis use data collected from other sources, including data from sentinel surveillance and national surveys.

*Coverage.* Case reports of STDs are reported to CDC by STD surveillance information systems operated by state and local STD control programs and health departments in 50 states, the District of Columbia, selected cities, all U.S. counties, and outlying areas consisting of U.S. dependencies, possessions, and independent countries in free association with the United States. Data from outlying areas are not included in *Health, United States*.

*Methodology.* Information is obtained from the following data sources: (a) case reporting of nationally notifiable STDs from state and local STD programs through the National Notifiable Disease Surveillance System (NNDSS); (b) projects that monitor STD positivity and prevalence in various settings, including the National Job Training Program, the STD Surveillance Network (SSuN), and the Gonococcal Isolate Surveillance Project (GISP); and (c) national sample surveys implemented by federal and private organizations.

*Issues Affecting Interpretation.* Most STDs are asymptomatic. Because of incomplete diagnosis and reporting, the number of STD cases reported to CDC undercounts the actual number of infections occurring among the U.S. population.

### Reference

Centers for Disease Control and Prevention. Sexually transmitted disease surveillance 2017. 2018. Available from: <https://www.cdc.gov/std/stats17/toc.htm>.

For more information, see the CDC website on STD data and statistics at: <https://www.cdc.gov/std/stats>, and the CDC website on STD diseases and related conditions: <https://www.cdc.gov/std/general/default.htm>.

## Private and Global Sources

### American Dental Association (ADA)

The ADA Masterfile contains the most up-to-date information on dentists in the United States. The Masterfile is a database of all dentists, both practicing and nonpracticing, in the United States. It is updated through a variety of methods, including reconciliation with state licensure databases, death records, and various surveys and censuses of dentists carried out by ADA.

ADA's Health Policy Institute conducts annual surveys of predoctoral dental educational institutions. A questionnaire, mailed to all dental schools, collects information on academic programs, admissions, enrollment, attrition, graduates, educational expenses and financial assistance, patient care, advanced dental education, and faculty positions.

## **References**

American Dental Association, Health Policy Institute. Tables 1 and 3. In: Supply of dentists in the U.S.: 2001–2017. Available from: <https://www.ada.org/en/science-research/health-policy-institute/data-center/supply-and-profile-of-dentists>.

American Dental Association, Health Policy Institute. Survey of Dental Education series. Report 1: Academic programs, enrollment and graduates. 2015–2016 report 1. 2016. Available from: [https://www.ada.org/~media/ADA/Science%20and%20Research/HPI/Files/2015-16\\_SDE1\\_final.xlsx?la=en](https://www.ada.org/~media/ADA/Science%20and%20Research/HPI/Files/2015-16_SDE1_final.xlsx?la=en).

For more information, see the ADA website at: <https://www.ada.org>.

## **American Hospital Association (AHA) Annual Survey of Hospitals**

Data from AHA's annual survey are based on questionnaires sent to all AHA-registered and nonregistered hospitals in the United States and its associated areas: American Samoa, Guam, the Marshall Islands, Puerto Rico, and the Virgin Islands. U.S. government hospitals located outside the United States are excluded. Overall, the average response rate over the past 5 years has been approximately 83%. For nonreporting hospitals and for the survey questionnaires of reporting hospitals on which some information was missing, estimates are made for all data except those on beds, bassinets, facilities, and services. Data for beds and bassinets of nonreporting hospitals are based on the most recent information available from those hospitals. Data for facilities and services are based only on reporting hospitals. Estimates of other types of missing data are based on data reported the previous year, if available. When unavailable, estimates are based on data furnished by reporting hospitals similar in size, control, major service provided, length of stay, and geographic and demographic characteristics.

### **Reference**

American Hospital Association. AHA hospital statistics, 2018 edition. Chicago, IL: AHA. 2018.

For more information, see the AHA website at: <https://www.aha.org>.