Appendix I

Data Sources

This report consolidates the most current data on the health of the population of the United States, the availability and use of health resources, and health care expenditures. The information was obtained from the data files and/or published reports of many Federal Government and private and global agencies and organizations. In each case, the sponsoring agency or organization collected data using its own methods and procedures. Therefore, the data in this report 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, users should be aware of the general strengths and weaknesses of the different data collection systems. For example, population-based surveys obtain socioeconomic data, data on family characteristics, and information on the impact of an illness, such as days lost from work or limitation of activity. These data are limited by the amount of information a respondent remembers or is willing to report. A respondent may not know detailed medical information, such as precise diagnoses or the types of operations performed, and therefore cannot report it. Health care providers, such as physicians and hospitals, usually have good diagnostic information but little or no information about the socioeconomic characteristics of individuals or the impact of illnesses on individuals.

The populations covered by different data collection systems may not be the same, and understanding the differences is critical to interpreting the data. Data on vital statistics and national expenditures cover the entire population. Most data on morbidity and utilization of health resources cover only the civilian noninstitutionalized population. Such statistics do not include data for military personnel who are usually young, for institutionalized people who may be any age, or for nursing home residents who are usually old.

All data collection systems are subject to error, and records may be incomplete or contain inaccurate information. People 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 impact on the data. Where possible, table notes describe the universe and method of data collection, to enable the user to place his or her own evaluation on the 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 Household Survey on Drug Abuse, the Monitoring the Future Survey, and the Youth Risk Behavior Survey using slightly different questions of persons of differing ages, interviewed in different settings (at school versus at home), so estimates will differ.

Overall estimates generally have relatively small sampling errors, but estimates for certain population subgroups may be based on small numbers and have relatively large sampling errors. Numbers of births and deaths from the vital statistics system represent complete counts (except for births in those States where data are based on a 50-percent sample for certain years). Therefore, they 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, considerable caution must be observed in interpreting the conditions described by the figures. Estimates that are unreliable because of large sampling errors or small numbers of events are noted with asterisks in selected tables. The criteria used to designate unreliable estimates are indicated in notes to the applicable tables.

Descriptive summaries of the data sets that follow provide a general overview of study design, methods of data collection, and reliability and validity of the data. The agency or organization that sponsored the data collection is specified. More complete and detailed discussions are in the publications and Web sites listed at the end of each summary. The entries are listed alphabetically by dataset name.

Government Sources

Abortion Surveillance

Centers for Disease Control and Prevention

National Center for Chronic Disease Prevention and Health Promotion

In 1969 CDC's National Center for Chronic Disease Prevention and Health Promotion (NCCDPHP) began abortion surveillance to document the number and characteristics of women obtaining legal induced abortions, monitor unintended pregnancy, and assist efforts to identify and reduce preventable causes of morbidity and mortality associated with abortions. For each year from 1973–97 abortion data from central health agencies have been available from 52 reporting areas: 50 States, the District of Columbia, and New York City. Beginning in 1998, abortion data are available from only 46 States (excluding Alaska, California, New Hampshire, and Oklahoma), the District of Columbia, and New York City. While the total number of legal induced abortions is available for those 48 reporting areas, not all areas collect information on the characteristics of women who obtain abortions. Furthermore the number of areas reporting each characteristic and the number of areas with complete data for each characteristic varies from year to year. For example, in 1999, the number of areas reporting different characteristics ranged from 26 areas reporting Hispanic ethnicity and 37 areas reporting race and marital status to 47 areas reporting age. Reporting area data with more than 15 percent unknown for a given characteristic are excluded from the analysis of that characteristic.

For 48 reporting areas, data concerning the number and characteristics of women who obtain legal induced abortions are provided by central health agencies such as State health departments and the health departments of New York City and the District of Columbia. In general the procedures are reported by the State in which the procedure is performed. However, two reporting areas (the District of Columbia and Wisconsin) report characteristics of abortions only for area/State residents; characteristics for out-of-area/State residents are unavailable.

Between 1989 and 1997, the total number of abortions reported to CDC was about 10 percent less than the total estimated independently by the Alan Guttmacher Institute (AGI), a not-for-profit organization for reproductive health research, policy analysis, and public education. Beginning in 1998, the total number of abortions reported to CDC was about 33 percent less than the total estimated by AGI. The four reporting areas (the largest of which was California) that did not report abortions to CDC in 1998 accounted for 18 percent of all abortions tallied by AGI's 1995–96 survey. See *Alan Guttmacher Institute Abortion Survey*.

For more information, see Centers for Disease Control and Prevention, CDC Surveillance Summaries, November 2002. *Morbidity and Mortality Weekly Report* 2002;51 (NoSS-9), Abortion Surveillance—United States, 1999; or contact: Director, Division of Reproductive Health, NCCDPHP, CDC, Atlanta, GA 30341; or visit the NCCDPHP surveillance and research Web site at www.cdc.gov/nccdphp/drh/surveil.htm.

Aerometric Information Retrieval System (AIRS)

Environmental Protection Agency

The Environmental Protection Agency's Aerometric Information Retrieval System (AIRS) compiles data on ambient air levels of particulate matter smaller than 10 microns (PM-10), lead, carbon monoxide, sulphur dioxide, nitrogen dioxide, and tropospheric ozone. These pollutants were identified in the Clean Air Act of 1970 and in its 1977 and 1990 amendments because they pose significant threats to public health. The National Ambient Air Quality Standards (NAAQS) define for each pollutant the maximum concentration level (micrograms per cubic meter) that cannot be exceeded during specific time intervals. Data shown in this publication reflect percent of the population living in areas that exceed the NAAQS for a pollutant in a calendar year (such areas are called nonattainment areas) and population data from the U.S. Bureau of the Census. For 1996 and later years, estimates of the population in the year 2000 are used for this calculation.

Nonattainment areas may include single counties, multiple counties, parts of counties, municipalities, or combinations of the preceding jurisdictions. When an area is designated as "nonattainment," it retains this status for 3 years, regardless of annual changes in air quality. Nonattainment areas may also include jurisdictions in which the source of the pollutants is located, even if that jurisdiction meets all NAAQS. The areas monitored may change over time to reflect changes in air quality or the pollutants being monitored.

The EPA's ambient air quality monitoring program is carried out by State and local agencies and consists of three major categories of monitoring stations, State and Local Air Monitoring Stations (SLAMS), National Air Monitoring Stations (NAMS), and Special Purpose Monitoring Stations (SPMS), that measure the criteria pollutants. Additionally, a fourth category of a monitoring station, the Photochemical Assessment Monitoring Stations (PAMS), which measures ozone precursors (approximately 60 volatile hydrocarbons and carbonyl), has been required by the 1990 Amendments to the Clean Air Act.

The SLAMS consist of a network of about 4,000 monitoring stations whose size and distribution is largely determined by the needs of State and local air pollution control agencies to meet their respective State implementation plan requirements. The NAMS (1,080 stations) are a subset of the SLAMS network with emphasis being given to urban and multi-source areas. In effect they are key sites under SLAMS, with emphasis on areas of maximum concentrations and high population density. The SPMS provide for special studies needed by the State and local agencies to support State implementation plans and other air program activities. The SPMS are not permanently established and can be adjusted easily to accommodate changing needs and priorities. The SPMS are used to supplement the fixed monitoring network as circumstances require and resources permit.

For more information, write: Office of Air Quality Planning and Standards, Environmental Protection Agency, Research Triangle Park, NC 27711; or visit the EPA Office of Air Quality Planning and Standards Web site at www.epa.gov/oar/oaqps.

AIDS Surveillance

Centers for Disease Control and Prevention

National Center for HIV, STD, and TB Prevention

Acquired immunodeficiency syndrome (AIDS) surveillance is conducted by health departments in each State, territory, and the District of Columbia. Although surveillance activities range from passive to active, most areas employ multifaceted active surveillance programs, which include four major reporting sources of AIDS information: hospitals and hospital-based physicians, physicians in nonhospital practice, public and private clinics, and medical record systems (death certificates, tumor registries, hospital discharge abstracts, and communicable disease reports). Using a standard confidential case report form, the health departments collect information that is then transmitted electronically to CDC without personal identifiers.

AIDS surveillance data are used to detect epidemiologic trends, to identify unusual cases requiring followup, and for semiannual publication in the *HIV/AIDS Surveillance Report*. Studies to determine the completeness of reporting of AIDS cases meeting the national surveillance definition suggest reporting at greater than or equal to 90 percent.

Decreases in AIDS incidence and in the number of AIDS deaths, first noted in 1996, have been ascribed to the effect of new treatments, which prevent or delay the onset of AIDS and premature death among HIV-infected persons, and result in an increase in the number of persons living with HIV and AIDS. A growing number of States require confidential reporting of persons with HIV infection and participate in CDC's integrated HIV/AIDS surveillance system that compiles information on the population of persons newly diagnosed and living with HIV infection.

For more information on AIDS surveillance, see: Centers for Disease Control and Prevention. *HIV/AIDS Surveillance Report*, published semiannually; or contact: Chief, Surveillance Branch, Division of HIV/AIDS Prevention—Surveillance and Epidemiology, National Center for HIV, STD, and TB Prevention (NCHSTP), Centers for Disease Control and Prevention, Atlanta, GA 30333; or visit the NCHSTP Web site at www.cdc.gov/nchstp/od/nchstp.html.

Census of Fatal Occupational Injuries (CFOI)

Bureau of Labor Statistics

The Census of Fatal Occupational Injuries (CFOI), administered by the Bureau of Labor Statistics (BLS) in conjunction with participating State agencies, has compiled comprehensive and timely information on fatal work injuries occurring in the 50 States and the District of Columbia since 1992. To compile counts that are as complete as possible, the BLS census uses diverse sources to identify, verify, and profile fatal work injuries. Key information about each workplace fatality (occupation and other worker characteristics, equipment or machinery involved, and circumstances of the event) is obtained by cross-referencing the source records. For a fatality to be included in the census, the decedent must have been employed (that is,

working for pay, compensation, or profit) at the time of the event, engaged in a legal work activity, or present at the site of the incident as a requirement of his or her job. These criteria are generally broader than those used by Federal and State agencies administering specific laws and regulations. Fatalities that occur during a person's commute to or from work are excluded from the census counts.

Data for the CFOI are compiled from various Federal, State, and local administrative sources—including death certificates, workers' compensation reports and claims, reports to various regulatory agencies, medical examiner reports, and police reports—as well as news reports. Diverse sources are used because studies have shown that no single source captures all job-related fatalities. Source documents are matched so that each fatality is counted only once. To ensure that a fatality occurred while the decedent was at work, information is verified from two or more independent source documents or from a source document and a followup questionnaire.

The number of occupational fatalities and fatality rates shown in this report are revised, except for the most recent year, and may differ from original data published by CFOI. States have up to 1 year to update their initial published State counts. States may identify additional fatal work injuries after data collection closeout for a reference year. In addition, other fatalities excluded from the published count because of insufficient information to determine work relationship may subsequently be verified as work related. Increases in the published counts based on additional information have averaged less than 100 fatalities per year, or less than 1.5 percent of the total.

For more information, see: Bureau of Labor Statistics, *National Census of Fatal Occupational Injuries, 2001*. Washington, DC. U.S. Department of Labor. August 2002; or visit the CFOI Web site at www.bls.gov/iif/oshcfoi1.htm.

Consumer Price Index (CPI)

Bureau of Labor Statistics

The Consumer Price Index (CPI) is a monthly measure of the average change in the prices paid by urban consumers for a fixed market basket of goods and services. The all-urban index (CPI-U) introduced in 1978 covers residents of metropolitan areas as well as residents of urban parts of nonmetropolitan areas (about 87 percent of the U.S. population in 2000).

In calculating the index, price changes for the various items in each location were averaged together with weights that represent their importance in the spending of all urban consumers. Local data were then combined to obtain a U.S. city average.

The index measures price changes from a designated reference date, 1982–84, which equals 100. An increase of 22 percent, for example, is shown as 122. Change can also be expressed in dollars as follows: the price of a base period "market basket" of goods and services bought by all urban consumers has risen from $100 in 1982–84 to $179.9 in 2002.

The current revision of the CPI, completed in 2000, reflects spending patterns based on the Survey of Consumer Expenditures from 1993 to 1995, the 1990 Census of Population, and the ongoing Point-of-Purchase Survey. Using an improved sample design, prices for the goods and services required to calculate the index are collected in urban areas throughout the country and from retail and service establishments. Data on rents are collected from tenants of rented housing and residents of owner-occupied housing units. Food, fuels, and other goods and services are priced monthly in urban locations. Price information is obtained through visits or calls by trained BLS field representatives using computer-assisted telephone interviews.

The earlier 1987 revision changed the treatment of health insurance in the cost-weight definitions for medical care items. This change has no effect on the final index result but provides a clearer picture of the role of health insurance in the CPI. As part of the revision, three new indexes have been created by separating previously combined items, for example, eye care from other professional services and inpatient and outpatient treatment from other hospital and medical care services.

Effective January 1997 the hospital index was restructured by combining the three categories—room, inpatient services, and outpatient services—into one category, hospital services. Differentiation between inpatient and outpatient and among service types are all combined under this broad category. In addition new procedures for hospital data collection identify a payor, diagnosis, and the payor's reimbursement arrangement from selected hospital bills.

A new geographic sample and item structure were introduced in January 1998 and expenditure weights were updated to 1993–95. Pricing of a new housing sample using computer-

assisted data collection started in June 1998. In January 1999 the index was rebased from the 1982–84 time period to 1993–95.

For more information, see: Bureau of Labor Statistics, *Handbook of Methods*, BLS Bulletin 2490, U.S. Department of Labor, Washington, DC. April 1997; Revising the Consumer Price Index, *Monthly Labor Review*, Dec 1996. U.S. Department of Labor, Bureau of Labor Statistics, Washington, DC; IK Ford and D Ginsburg, Medical Care and the Consumer Price Index, National Bureau of Economic Research, Research Studies in Income and Wealth vol. 62; or visit the BLS/CPI Web site at www.bls.gov/cpi/home.htm.

Current Population Survey (CPS)

Bureau of the Census

Bureau of Labor Statistics

The Current Population Survey (CPS) is a household sample survey of the civilian noninstitutionalized population conducted monthly by the U.S. Bureau of the Census for the Department of Labor, Bureau of Labor Statistics (BLS). CPS provides estimates of employment, unemployment, and other characteristics of the general labor force, the population as a whole, and various other population subgroups. Estimates of health insurance coverage are derived from the Annual Demographic Supplement (ADS), which includes a series of questions asked each March in addition to regular CPS questions. The ADS is also known as the "March Supplement."

The CPS sample is located in 754 sample areas, with coverage in every State and the District of Columbia. Beginning with 2001 estimates are based on interviews of an expanded sample of 60,000 households per month. Prior to 2001 estimates were based on 50,000 households per month. Also starting in 2001 the State Children's Health Insurance Program (SCHIP) sample expansion was introduced, which increased the March sample to approximately 72,000 households in order to produce statistically reliable State data on the number of low income children who do not have health insurance coverage. The expanded sample for the March Supplement also improved reliability of other national estimates. In an average month the nonresponse rate is about 6–7 percent. In 1994 major changes were introduced, which included a complete redesign of the questionnaire including new health insurance questions, and the introduction of computer-assisted interviewing for the entire survey. In addition, there were revisions to some of the labor force concepts and definitions. Prior to the redesign, CPS data were primarily collected using a paper-and-pencil form.

The estimation procedure used involves inflation by the reciprocal of the probability of selection, adjustment for nonresponse, and ratio adjustment. Beginning in 1994 new population controls that were based on the 1990 census adjusted for the estimated population undercount were used. Starting with *Health, United States, 2003*, poverty estimates for 2000 were recalculated based on the expanded SCHIP sample, and beginning with 2000 data census 2000-based population controls were implemented.

For more information, visit the CPS Web site at www.bls.census.gov/cps/cpsmain.htm.

Department of Veterans Affairs Databases

Department of Veterans Affairs

The Department of Veterans Affairs (VA) maintains the *National Patient Care Database* (NPCD) and the *National Enrollment Database* (NED). Data are collected locally at each VA medical center and are transmitted electronically to the VA Austin Automation Center for use in providing nationwide statistics, reports, and comparisons.

The NPCD is a nationwide system that contains a statistical record for each episode of care provided under VA auspices in VA and non-VA hospitals, nursing homes, and domiciliaries, and in VA outpatient clinics. Three major extracts from the NPCD are the patient treatment file (PTF), the patient census file, and the outpatient clinic file (OPC).

The *Patient Treatment File (PTF)* collects data at the time of the patient's discharge on each episode of inpatient care provided to patients at VA hospitals, VA nursing homes, VA domiciliaries, community nursing homes, and other non-VA facilities. The PTF record contains the scrambled social security number, dates of inpatient treatment, date of birth, State and county of residence, type of disposition, place of disposition after discharge, as well as the ICD-9-CM diagnostic and procedure or operative codes for each episode of care.

The *Patient Census File* collects data on each patient remaining in a VA medical facility at midnight at the end of each quarter of the fiscal year. The census record includes information similar to that reported in the patient treatment file record.

The *Outpatient Clinic File (OPC)* collects data on each instance of medical treatment provided to a veteran in an outpatient setting. The OPC record includes the age, scrambled social security number, State and county of residence, VA eligibility code, clinic(s) visited, purpose of visit, and the date of visit for each episode of care.

The VA also maintains the *National Enrollment Database* (NED) as the official repository of enrollment information for each veteran enrolled in the VA health care system. In addition, an extract containing selected information from the NPCD, the NED, and the cost distribution system is also produced by the Austin Automation Center.

For more information, write: Department of Veterans Affairs, Office of Policy, Planning, and Preparedness, Policy Analysis Service, 810 Vermont Ave., NW, Washington, DC 20420; or visit the National Patient Care Database at www.virec.research.med.va.gov/DATABASES/NPCD/NPCD.HTM.

Drug Abuse Warning Network (DAWN)

Substance Abuse and Mental Health Services Administration

The Drug Abuse Warning Network (DAWN) is a large-scale, ongoing drug abuse data collection system based on information from hospital emergency departments (EDs) and medical examiner jurisdictions. The major objectives of the DAWN data system include monitoring of drug-abuse patterns and trends, identification of substances associated with drug-abuse episodes, and assessment of drug-related consequences and other health hazards. Estimates reported in this publication are from the hospital ED component of DAWN.

Hospitals eligible for DAWN are non-Federal, short-stay general hospitals that have a 24-hour emergency department. Since 1988 the DAWN emergency department data have been collected from a representative sample of eligible hospitals located throughout the coterminous United States, including 21 oversampled metropolitan areas. Within each facility, a designated DAWN reporter is responsible for identifying eligible drug-abuse episodes by reviewing emergency department records and abstracting and submitting data on each reportable case. To be included in DAWN, the patient presenting to the ED must meet all of the following four criteria: (a) patient was between ages 6 and 97 years and was treated in the hospital's ED; (b) patient's presenting problem(s) for the ED visit was induced by or related to drug use, regardless of when drug use occurred; (c) episode involved use of an illegal drug or use of a legal drug or other chemical substance contrary to directions; (d) patient's reason for using the substance(s) was dependence, suicide attempt or gesture, and/or psychic effect.

The data from the DAWN sample are used to generate estimates of the total number of emergency department drug-abuse episodes and drug mentions in all eligible hospitals in the coterminous United States and in the 21 metropolitan areas. Overall, a response rate of 81 percent of sample hospitals was obtained in the 2001 survey.

For further information, see Substance Abuse and Mental Health Services Administration, Office of Applied Studies. Emergency Department Trends from the Drug Abuse Warning Network, Preliminary Estimates January–June 2001 with Revised Estimates 1994 to 2001, DAWN Series D-21, DHHS Publication No. (SMA) 02-3635, Rockville, MD. 2002; or write: Office of Applied Studies, Substance Abuse and Mental Health Services Administration, Room 16-105, 5600 Fishers Lane, Rockville, MD 20857; or visit the SAMHSA Web site at www.drugabusestatistics.samhsa.gov.

Employee Benefits Survey—See *National Compensation Survey*.

Inventory of Mental Health Organizations (IMHO)

Substance Abuse and Mental Health Services Administration (SAMHSA)

The Survey and Analysis Branch of SAMHSA's Center for Mental Health Services conducted a biennial Inventory of Mental Health Organizations and General Hospital Mental Health Services (IMHO/GHMHS) from 1986 until 1994. The core questionnaire included versions designed for specialty mental health organizations and another for non-Federal general hospitals with separate psychiatric services.

IMHO/GHMHS has been the primary source for Center for Mental Health Services data included in *Health, United States*. The data system was based on questionnaires mailed every other year to mental health organizations in the United States, including psychiatric hospitals, non-Federal general hospitals with psychiatric services, Department of Veterans Affairs psychiatric services, residential treatment centers for emotionally disturbed children, freestanding outpatient psychiatric clinics, partial care organizations, freestanding day-night organizations, and multiservice mental health organizations, not elsewhere classified.

IMHO/GHMHS was a redesign of three previous inventory systems with more complicated data collection procedures. In 1998 the IMHO/GHMHS was replaced by the Survey of Mental Health Organizations, General Hospital Mental Health Services, and Managed Behavioral Health Care Organizations (SMHO). A brief 100 percent inventory of organizations was conducted by postcard and used to provide basic information on all organizations and as a sampling frame from which to draw a sample for a more in-depth sample survey. The sample survey questionnaire differed from the previous core questionnaires mainly by inclusion of questions relating to managed behavioral health care organizations.

For more information, write: Survey and Analysis Branch, Division of State and Community Systems Development, Center for Mental Health Services, Room 15C-04, 5600 Fishers Lane, Rockville, MD 20857. For further information on mental health and data from the 1997 Client/Patient Sample Survey, see: Center for Mental Health Services, *Mental Health, United States, 2000*. Manderscheid R, Henderson MJ, eds. DHHS Pub. No. (SMA) 01-3537. Washington, DC; or visit the Center for Mental Health Services Web site at www.samhsa.gov/centers/cmhs/cmhs.html.

Medicaid Data System

Centers for Medicare & Medicaid Services

The primary data sources for Medicaid statistical data are the Medicaid Statistical Information System (MSIS) and the CMS-64 reports.

MSIS is the basic source of State-reported eligibility and claims data on the Medicaid population, their characteristics, utilization, and payments. Beginning in FY1999, as a result of legislation enacted from the Balanced Budget Act of 1997, States are required to submit individual eligibility and claims data tapes to CMS quarterly through the Medicaid Statistical Information System (MSIS). Prior to FY1999, States were required to submit an annual HCFA-2082 report, designed to collect aggregated statistical data on eligibles, recipients, services, and expenditures during a Federal fiscal year (October 1 through September 30). The data reported for each year represented people on the Medicaid rolls, recipients of Medicaid services and payments for claims adjudicated during the year. The data reflected bills adjudicated or processed during the year, rather than services used during the year. States summarized and reported the data processed through their own Medicaid claims processing and payment operations, unless they opted to participate in MSIS, in which case the 2082 report was produced by HCFA (Health Care Financing Administration, the predecessor to CMS).

The CMS-64 is a product of the financial budget and grant system. The CMS-64 is a statement of expenditures for the Medicaid program that States submit to CMS 30 days after each quarter. The report is an accounting statement of actual expenditures made by the States for which they are entitled to receive Federal reimbursement under title XIX for that quarter. The amount claimed on the CMS-64 is a summary of expenditures derived from source documents such as invoices, cost reports and eligibility records.

The CMS-64 shows the disposition of Medicaid grant funds for the quarter being reported and previous years, the recoupments made or refunds received, and income earned on grant funds. The data on the CMS-64 are used to reconcile the monetary advance made on the basis of States' funding estimates filed prior to the beginning of the quarter on the CMS-37. As such, the CMS-64 is the primary source for making adjustments for any identified overpayments and underpayments to the States. Also incorporated into this process are disallowance actions forwarded from other Federal financial adjustments. Finally, the CMS-64 provides information that forms the basis for a series of Medicaid financial reports and budget analyses. Additionally included are third party liability (TPL) collections tables. Third party liability refers to the legal obligation of certain health care sources to pay the medical claims of Medicaid recipients before Medicaid pays these claims. Medicaid pays only after the TPL sources have met their legal obligation to pay.

Users of Medicaid data may note apparent inconsistencies that are primarily due to the difference in the information captured in MSIS versus CMS-64 reports. The most

substantive difference is due to payments made to "disproportionate share hospitals." Payments to disproportionate share hospitals do not appear in MSIS since States directly reimburse these hospitals and there is no fee-for-service billing. Other less significant differences between MSIS and the CMS-64 occur because adjudicated claims data are used in MSIS versus the reporting of actual payments reflected in the CMS-64. Differences also may occur because of internal State practices for capturing and reporting these data through two separate systems. Finally, national totals for the CMS-64 are different because they include other jurisdictions, such as the Northern Mariana Islands and American Samoa.

For further information on Medicaid data, visit the CMS Web site at www.cms.gov/medicaid/datasources.asp or the Research Data Center (ResDAC) Web site at www.resdac.umn.edu/medicaid/data_available.asp. Also see Appendix II, *Medicaid*.

Medical Expenditure Panel Survey (MEPS)

Agency for Healthcare Research and Quality

The Medical Expenditure Panel Survey (MEPS) is a national probability survey conducted on an annual basis since 1996. The survey is designed to produce nationally representative estimates of healthcare use, expenditures, sources of payment, insurance coverage, and quality of care for the U.S. civilian noninstitutionalized population. The panel design of the survey features several rounds of interviewing covering 2 full calendar years. The MEPS consists of three components.

The Household Component (HC) is a nationally representative survey of the civilian noninstitutionalized population drawn from a subsample of households that participated in the prior year's National Health Interview Survey conducted by the National Center for Health Statistics. The sample sizes for the HC are approximately 10,000 families in 1996 and 1998–2000, 13,500 families in 1997 and 2001, and 15,000 families annually beginning in 2002. The full-year response rate has generally been about 66 percent. Missing expenditure data were imputed using data collected in the Medical Provider Component whenever possible.

Data are collected in the Medical Provider Component (MPC) to improve the accuracy of expenditure estimates derived solely from the Household Component (HC). The MPC is particularly useful in obtaining expenditure information for persons enrolled in managed care plans and Medicaid recipients. The MPC collects data from hospitals, physicians, and home health providers that were reported in the HC as providing care to MEPS sample persons. Sample sizes for the MPC vary from year to year depending on the HC sample size and the MPC sampling rates for providers.

The Insurance Component (IC) consists of two subcomponent samples, a household sample and list sample. The household sample collects detailed information from employers on the health insurance held by and offered to respondents to the MEPS HC. The list sample collects data on the types and costs of workplace health insurance from a total of about 40,000 business establishments and governments each year.

The Medical Expenditure Panel Surveys (MEPS) update the 1987 National Medical Expenditure Survey (NMES). The Household Survey (HS) and the Medical Provider Survey (MPS) components of the 1987 NMES were designed to provide nationally representative estimates of the health status, health insurance coverage, and health care use and expenditures for the U.S. civilian noninstitutionalized population for the calendar year 1987. The HS consisted of four rounds of household interviews. Income was collected in a special supplement administered early in 1988. Events under the scope of the MPS included medical services provided by or under the direction of a physician, all hospital events, and home health care. The sample of events included in the MPS was all events for persons covered by Medicaid and for a 25 percent sample of HS respondents. For the first core household interview, 17,500 households were selected. The 12-month joint core questionnaire/health questionnaire/access supplement response rate for the household component of the NMES was 80 percent. Missing expenditure data were imputed.

For further information about the National Medical Expenditure Survey, see: Hahn B and Lefkowitz D. Annual expenses and sources of payment for health care services (AHRQ Pub. No. 93-0007). National Medical Expenditure Survey Research Findings 14, Agency for Healthcare Research and Quality. Rockville, MD. Public Health Service. Nov. 1992. For further information on the MEPS, visit the MEPS Web site at www.meps.ahrq.gov.

Medicare Administrative Data

Centers for Medicare & Medicaid Services

CMS collects and synthesizes a broad range of quantitative information on its programs, from estimates of future Medicare spending to enrollment, spending, and claims data. The Claims and Utilization Data files contain extensive utilization information at various levels of summarization for a variety of providers and services. There are many types and levels of these files, including but not limited to the National Claims History (NCH) files, the Standard Analytic Files (SAF), Stay Records files, Part B Medicare files, and various other files.

The National Claims History 100 Percent Nearline File contains all institutional and noninstitutional claims, and provides records of every Medicare claim submitted, including adjustment claims. The Standard Analytical Files (SAFs) contain final action claims data in which all adjustments have been resolved. These files contain information collected by Medicare to pay for health care services provided to a Medicare beneficiary. SAFs are available for each institutional (inpatient, outpatient, skilled nursing facility, hospice, or home health agency) and noninstitutional (physician and durable medical equipment providers) claim type. The record unit of SAFs is the claim (some episodes of care may have more than one claim). SAF files include the Inpatient SAF, the Skilled Nursing Facility SAF, the Outpatient SAF, the Home Health Agency SAF, the Hospice SAF, the Clinical Laboratory SAF, the Durable Medical Equipment SAF, and a 5-Percent Beneficiary File SAF.

Medicare Provider and Analysis Review (MedPAR) files contain inpatient hospital and skilled nursing facility (SNF) final action stay records. Each MedPAR record represents a stay in an inpatient hospital or SNF. An inpatient "stay" record summarizes all services rendered to a beneficiary from the time of admission to a facility through discharge. Each MedPAR record may represent one claim or multiple claims, depending on the length of a beneficiary's stay and the amount of inpatient services used throughout the stay.

The Denominator File contains demographic and enrollment information about each beneficiary enrolled in Medicare during a calendar year. The information in the Denominator File is 'frozen' in March of the following calendar year. Some of the information contained in this file includes the beneficiary unique identifier, State and county codes, ZIP Code, date of birth, date of death, sex, race, age, monthly entitlement indicators (for Medicare Part A, Medicare Part B, or Part A and Part B), reasons for entitlement, State buy-in indicators, and monthly managed care indicators (yes/no). The Denominator File is used to determine beneficiary demographic characteristics, entitlement, and beneficiary participation in Medicare Managed Care Organizations.

The Vital Status File contains demographic information about each beneficiary ever entitled to Medicare. Some of the information contained in this file includes the beneficiary unique identifier, State and county codes, ZIP Code, date of birth, date of death, sex, race, and age. Often the Vital Status File is used to obtain recent death information for a cohort of Medicare beneficiaries.

The Group Health Plan (GHP) Master File contains data on beneficiaries who are currently enrolled or have ever been enrolled in a Managed Care Organization (MCO) under contract with CMS. Each record represents one beneficiary and each beneficiary has one record. Some of the information contained in this file includes the Beneficiary Unique Identifier number, date of birth, date of death, State and county, and managed care enrollment information such as dates of membership and MCO contract number. The GHP Master File is used to identify the exact MCO in which beneficiaries were enrolled.

Medicare claims are linked to survey-reported events to produce the Cost and Use file that provides complete expenditure and source of payment data on all health care services, including those not covered by Medicare. Data are also combined to produce estimates of expenditures, per-beneficiary utilization, and other statistics.

For more information about Medicare data files, see CMS' Research Data Center (ResDAC) Web site at www.resdac.umn.edu/medicare/data_available.asp or the CMS Web site at http://cms.hhs.gov/data/default.asp. Also see Appendix II, *Medicare*.

Medicare Current Beneficiary Survey (MCBS)

Centers for Medicare & Medicaid Services

The Medicare Current Beneficiary Survey (MCBS) is a continuous survey of a nationally representative sample of about 18,000 aged and disabled Medicare beneficiaries enrolled in Medicare Part A (hospital insurance), or Part B (medical insurance), or both, and residing in households or long-term care facilities. The survey provides comprehensive time-series data on utilization of health services, health and functional status, health care expenditures, and health insurance and beneficiary information (such as income, assets, living arrangement, family assistance, and quality of life). The longitudinal design of the survey allows each sample person to be interviewed three times a year for 4 years, whether he or she resides in the community or a facility or moves between the two settings, using the version of the questionnaire appropriate to the setting. Sample persons in the community are interviewed using computer-assisted personal interviewing (CAPI) survey instruments. Because long-term care facility residents often are in poor health, information about institutionalized patients is collected from proxy respondents such as nurses and other primary care givers affiliated with the facility. The sample is selected from the Medicare enrollment files with oversampling among disabled persons under age 65 and among persons 80 years of age and over.

Medicare claims are linked to survey-reported events to produce the Cost and Use file that provides complete expenditure and source of payment data on all health care services, including those not covered by Medicare. The Access to Care file contains information on beneficiaries' access to health care, satisfaction with care, and usual source of care. The sample for this file represents the "always enrolled" population, those who participated in the Medicare program for the entire year. In contrast, the Cost and Use file represents the "ever enrolled" population, including the experience of those who enter Medicare during the year and those who died.

For more information about the MCBS, see: A profile of the Medicare Current Beneficiary Survey, by GS Adler. Health Care Financing Review, vol 15 no 4. Health Care Financing Administration. Washington, DC. Public Health Service. 1994. For further information on the MCBS, visit the MCBS Web site at www.cms.hhs.gov/mcbs/default.asp.

Monitoring the Future Study (MTF)

National Institute on Drug Abuse

Monitoring the Future Study (MTF) is a large-scale epidemiological survey of drug use and related attitudes. It has been conducted annually since 1975 under a series of investigator-initiated research grants from the National Institute on Drug Abuse to the University of Michigan's Institute for Social Research. MTF is composed of three substudies: (a) annual survey of high school seniors initiated in 1975; (b) ongoing panel studies of representative samples from each graduating class that have been conducted by mail since 1976; and (c) annual surveys of 8th and 10th graders initiated in 1991.

The survey design is a multistage random sample with stage one being selection of particular geographic areas, stage two selection of one or more schools in each area, and stage three selection of students within each school. Data are collected using self-administered questionnaires conducted in the classroom by representatives of the Institute for Social Research. Dropouts and students who are absent on the day of the survey are excluded (about 18 percent of high school seniors, about 12 percent of 10th graders, and about 10 percent of 8th graders in 2001). Recognizing that the dropout population is at higher risk for drug use, this survey was expanded to include similar nationally representative samples of 8th and 10th graders in 1991. Statistics that are published in the *Dropout Rates in the United States: 2000* (published by the National Center for Educational Statistics, Pub. No. NCES 2002-114) stated that among persons 15–16 years of age, 2.9 percent have dropped out of school while the dropout rate increases to 3.5 percent for persons 17 years of age, 6.1 percent for persons 18 years of age, and 9.6 percent for persons 19 years of age. Therefore, surveying eighth graders (where dropout rates are much lower than for high school seniors) should be effective for picking up students at higher risk for drug use. Although the prevalence of drug use is slightly underestimated due to the exclusion of dropouts and absentees, the methodology is consistent over time and trend estimates are little affected.

Approximately 44,300 8th, 10th, and 12th graders in 394 schools were surveyed in 2002. In 2002 the annual senior samples comprised roughly 13,500 seniors in 120 public and private high schools nationwide, selected to be representative of all seniors in the continental United States. The 10th-grade

samples involved about 14,700 students in 133 schools in 2002, and the 2002 8th-grade samples had approximately 15,500 students in 141 schools. Response rates of 83 percent, 85 percent, and 91 percent for 12th, 10th, and 8th graders in 2002 have been relatively constant across time. Absentees constitute virtually all of the nonrespondents.

Estimates of substance use for youth based on the National Household Survey on Drug Abuse (NHSDA) are generally lower than estimates based on the MTF and the Youth Risk Behavior Surveillance System (YRBSS). In addition to the fact that the MTF excludes dropouts and absentees, rates are not directly comparable across these surveys due to differences in populations covered, sample design, questionnaires, interview setting, and statistical approaches to make the survey estimates generalizable to the entire population. The NHSDA survey collects data in homes, whereas the MTF and YRBSS collect data in school classrooms. The NHSDA estimates are tabulated by age, while the MTF and YRBSS estimates are tabulated by grade, representing different ages as well as different populations. See Cowan CD. Coverage, Sample Design, and Weighting in Three Federal Surveys. Journal of Drug Issues 31(3), 595–614, 2001.

For further information on Monitoring the Future Study, see: National Institute on Drug Abuse, Monitoring the Future National Survey Results on Drug Use, 1975–2001. Volume I, Secondary School Students, NIH Pub. No. 02-5106. Bethesda, MD: Public Health Service, August 2002; or visit the NIDA Web site at www.nida.nih.gov or the Monitoring the Future Web site at www.monitoringthefuture.org/.

National Ambulatory Medical Care Survey (NAMCS)

Centers for Disease Control and Prevention

National Center for Health Statistics

The National Ambulatory Medical Care Survey (NAMCS), initiated in 1973, is a continuing national probability sample of ambulatory medical encounters. The scope of the survey covers patient encounters in the offices of non-Federally employed physicians classified by the American Medical Association or American Osteopathic Association as "office-based, patient care" physicians. Patient encounters with physicians engaged in prepaid practices—health maintenance organizations (HMOs), independent practice organizations (IPAs), and other prepaid practices—are included in NAMCS. Excluded are visits to hospital-based physicians, visits to specialists in anesthesiology, pathology, and radiology, and visits to physicians who are principally engaged in teaching, research, or administration. Telephone contacts and nonoffice visits are also excluded.

A multistage probability design is employed. The first-stage sample consists of 84 primary sampling units (PSUs) in 1985 and 112 PSUs in 1992 selected from about 1,900 such units into which the United States has been divided. In each sample PSU, a sample of practicing non-Federal office-based physicians is selected from master files maintained by the American Medical Association and the American Osteopathic Association. The final stage involves systematic random samples of office visits during randomly assigned 7-day reporting periods. In 1985 the survey excluded Alaska and Hawaii. Starting in 1989 the survey included all 50 States.

In 1999 a sample of 2,499 physicians was selected, 1,728 were in scope, and 1,087 participated in the survey for a response rate of 63 percent. Data were provided on 20,760 records. In the 2000 survey a sample of 3,000 physicians was selected, 2,049 were in scope, and 1,388 participated for a response rate of 68 percent. Data were provided on 27,369 records. Data are collected on providers seen; reason for the visit; diagnoses; waiting time; drugs ordered, provided, or continued; and selected procedures and tests performed during the visit.

The estimation procedure used in NAMCS has three basic components: inflation by the reciprocal of the probability of selection, adjustment for nonresponse, and ratio adjustment to fixed totals.

For more detailed information on NAMCS, see: Cherry DK. National Ambulatory Medical Care Survey: 2000 summary. Advance data from vital and health statistics; no. 328. Hyattsville, MD: National Center for Health Statistics. 2002; or visit the NHCS section of the NCHS Web site at www.cdc.gov/nchs/nhcs.htm.

National Compensation Survey

Bureau of Labor Statistics

The National Compensation Survey (NCS) is conducted quarterly by The Bureau of Labor Statistics' Office of Compensation and Working Conditions and provides comprehensive measures of occupational earnings, compensation cost trends, benefit incidence, and detailed plan provisions. Detailed occupational earnings are available for metropolitan and nonmetropolitan areas, broad geographic regions, and on a national basis. The Employment Cost Index (ECI) and Employer Costs for Employee Compensation (ECEC) are compensation measures derived from the National Compensation Survey (NCS). ECI measures changes in labor costs. Average hourly employer cost for employee compensation is presented in the ECEC. Data from the March survey are presented in *Health, United States*.

In separate surveys the National Compensation Survey covers the incidence and detailed provisions of selected employee benefit plans in small private establishments (in even years), medium and large private establishments (in odd years), and State and local governments (in even years). National benefits data are presented for three broad occupational groupings: professional, technical, and related; clerical and sales; and blue-collar and service employees. Broad incidence data were also available by goods- and service-producing, union affiliation, and full- and part-time status.

The Employment Cost Index (ECI) is a quarterly measure of changes in labor costs. It is one of the principal economic indicators used by the Federal Reserve Bank. ECI data show changes in wages, salaries, benefit costs, and total compensation for all workers and separately for private industry and State and local government workers; report compensation changes by industry, occupational group, union and nonunion status, region, and metropolitan/nonmetropolitan status; provide seasonally adjusted and unadjusted data; and present historical data on changes in labor costs.

The Employer Costs for Employee Compensation (ECEC) product is produced quarterly and shows the employers' average hourly cost for total compensation and its components. The key features of ECEC include:

- Compensation costs for wages and salaries and benefits
- Cost data in dollar amounts and as percentages of compensation
- Data on Civilian workers and State and local government workers
- Compensation costs by major occupation, industry, region, union and nonunion status, establishment size, and full- or part-time status
- Reflects today's labor force composition

The sample for the NCS is selected using a three-stage design. The first stage involves the selection of areas. The NCS sample consists of 154 metropolitan and nonmetropolitan areas that represent the Nation's 326 metropolitan statistical areas and the remaining portions of the 50 States. In the second stage, establishments are systematically selected with probability of selection proportionate to their relative employment size within the industry. Use of this technique means that the larger an establishment's employment, the greater its chance of selection.

The third stage of sampling is a probability sample of occupations within a sampled establishment. This step is performed by the BLS field economist during an interview with the respondent establishment in which selection of an occupation is based on probability of selection proportionate to employment in the establishment. Each occupation is classified under its corresponding major occupational group using the Occupational Classification System Manual (OCSM) and the Census Occupation Index, which are based on the 1990 U.S. Census.

Data collection is conducted by BLS field economists. Data are gathered from each establishment on the primary business activity of the establishment, types of occupations, number of employees, wages and salaries and benefits, hours of work, and duties and responsibilities. Wage data obtained by occupation and work level allows NCS to publish occupational wage statistics for localities, census divisions, and the Nation.

The methodology and procedures used to make estimates vary by product line. For the wage series, individual wage rates are weighted by number of workers; sample weight, adjusted for nonresponding establishments and other factors; and the occupation work schedule (hourly, weekly, or annual). The benefit series has three weight-adjustment factors applied

to the data to account for establishment nonresponse, occupational nonresponse, and to adjust the estimated employment totals to actual counts of employment by industry for the survey reference date.

To measure compensation costs free from the influence of employment shifts among occupations and industries, the ECI is calculated with fixed employment weights unlike the method with which wage series and benefit series are calculated. Since December 1994, 1990 employment counts from the Bureau's Occupational Employment Survey have been used. The ECI is a standard Laspeyres fixed-weighted index.

The ECEC estimates are based on data collected for the ECI. Unlike the ECI, ECEC estimates are weighted by the most recently available industry and occupational employment mix derived from data produced by the BLS Current Employment Statistics (CES) program.

For more information, see: U.S. Department of Labor, Bureau of Labor Statistics, *Employment Cost Indexes 1975–99*, Bulletin 2532, Oct. 2000; and visit the BLS Web site at www.bls.gov/ncs/home.htm.

National Health Accounts

Centers for Medicare & Medicaid Services

Estimates of expenditures for health based on National Health Accounts are compiled annually by type of expenditure and source of funds by the Office of the Actuary. The American Hospital Association (AHA) data on hospital finances are the primary source for estimates relating to hospital care. The salaries of physicians and dentists on the staffs of hospitals, hospital outpatient clinics, hospital-based home health agencies, and nursing home care provided in the hospital setting are considered to be components of hospital care. Expenditures for home health care and for services of health professionals (for example, doctors, chiropractors, private duty nurses, therapists, and podiatrists) are estimated primarily using a combination of data from the U.S. Bureau of the Census Services Annual Survey and the quinquennial Census of Service Industries.

The estimates of retail spending for prescription drugs are based on household and industry data on prescription drug transactions. Expenditures for other medical nondurables and vision products and other medical durables purchased in retail outlets are based on estimates of personal consumption expenditures prepared by the U.S. Department of Commerce's Bureau of Economic Analysis, U.S. Bureau of Labor Statistics/Consumer Expenditure Survey; the 1987 National Medical Expenditure Survey and the 1996 Medical Expenditure Panel Survey conducted by the Agency for Healthcare Research and Quality; and spending by Medicare and Medicaid. Those durable and nondurable products provided to inpatients in hospitals or nursing homes, and those provided by licensed professionals or through home health agencies are excluded here, but are included with the expenditure estimates of the provider service category.

Nursing home expenditures cover care rendered in establishments providing inpatient nursing and health-related personal care through active treatment programs for medical and health-related conditions. These establishments cover skilled nursing and intermediate care facilities, including those for the mentally retarded. Spending estimates are primarily based upon data from the U.S. Bureau of the Census Services Annual Survey and the quinquennial Census of Service Industries.

Expenditures for construction include those spent on the erection or renovation of hospitals, nursing homes, medical clinics, and medical research facilities, but not for private office buildings providing office space for private practitioners. Expenditures for noncommercial research (the cost of commercial research by drug companies is assumed to be imbedded in the price charged for the product; to include this item again would result in double counting) are developed from information gathered by the National Institutes of Health and the National Science Foundation.

Source of funding estimates likewise come from a multiplicity of sources. Data on the Federal health programs are taken from administrative records maintained by the servicing agencies. Among the sources used to estimate State and local government spending for health are the U.S. Bureau of the Census' Government Finances, and the National Academy of Social Insurance reports on State-operated Workers' Compensation programs. Federal and State-local expenditures for education and training of medical personnel are excluded from these measures where they are separable. For the private financing of health care, data on the financial experience of health insurance organizations come from special Centers for Medicare & Medicaid Services analyses of private health insurers, and from the Bureau of Labor Statistics' survey on the cost of employer-sponsored health insurance and on consumer expenditures. Information on

out-of-pocket spending from the U.S. Bureau of the Census Services Annual Survey; U.S. Bureau of Labor Statistics Consumer Expenditure Survey; the 1987 National Medical Expenditure Survey and the Medical Expenditure Panel Surveys conducted by the Agency for Healthcare Research and Quality; and from private surveys conducted by the American Hospital Association, American Medical Association, American Dental Association, and IMS Health, an organization that collects data from the pharmaceutical industry, is used to develop estimates of direct spending by customers.

For more specific information on definitions, sources, and methods used in the National Health Accounts contact: Office of the Actuary, Centers for Medicare & Medicaid Services, 7500 Security Blvd., Baltimore, MD 21244-1850; or visit the Centers for Medicare & Medicaid Services National Health Accounts Web site at http://cms.hhs.gov/statistics/nhe.

State Health Expenditures

Estimates of personal health care spending by State are created using the same definitions of health care sectors used in producing the National Health Expenditures (NHE). The same data sources used in creating NHE are also used to create State estimates whenever possible. Additional sources are employed when surveys used to create valid national estimates lack sufficient sample size to create valid State-level estimates. State-level data are used to estimate the State-by-State distribution of health spending, and the NHE national totals for the specific type of service or source of funds are used to control the level of State-by-State distributions. This procedure implicitly assumes that national spending estimates can be created more accurately than State-specific expenditures.

The NHE data that were used as national totals for these State estimates were published in *Health, United States, 2001*, and differ from the sum of State estimates because national totals included expenditures for persons living in U.S. territories and for military and Federal civilian employees and their families stationed overseas. The sum of the State-level expenditures exclude health spending for those groups. Starting with *Health, United States, 2002,* NHE reflect new data and benchmark revisions incorporated after completion of the State estimates, and incorporate a conceptual revision to exclude spending for persons living in U.S. territories and military and Federal civilian employees and their families living overseas.

Starting in *Health, United States, 2002*, State estimates are based on the location of the beneficiary's residence. This differs from previous estimates published in *Health, United States*, which presented spending based on the health care provider's location. State estimates were first constructed based on the provider's location because data available to estimate spending by State primarily comes from providers and represents the State-of-provider location. However, the most useful unit for analyzing spending trends and differences are per capita units, which are based on spending estimates for the State in which people reside. Therefore, State-of-provider-based expenditures are adjusted to a State-of-residence basis using interstate border-crossing flow patterns that represent travel patterns across State borders for health care.

Data for the interstate border-crossing flow patterns are based on Medicare claims. Medicare is the only comprehensive source upon which to base interstate flows of spending between State-of-provider and State-of-beneficiary residence. Data for non-Medicare payers (excluding Medicaid) are also based on Medicare flow patterns, but are further adjusted for age-specific service mix variation in hospital and physician services. Medicaid services are not adjusted because it is assumed that care provided to eligible State residents is most often provided by in-State providers and further assumed that spending by Medicaid is identical on a residence and provider basis.

In addition to differences noted earlier, national totals for residence-based State health expenditures may differ slightly from national totals for provider-based expenditures due to inflows and outflows of health care spending to the U.S. territories. Because flow patterns are based on Medicare data, we are able to adjust for services that Medicare beneficiaries receive outside of the United States, and for services received by Medicare beneficiaries in the United States who either live in the U.S. territories or in other countries. Similar adjustments for the non-Medicare, non-Medicaid population are not possible.

For more information contact: Office of the Actuary, Centers for Medicare & Medicaid Services, 7500 Security Blvd., Baltimore, MD 21244-1850; or visit the Centers for Medicare & Medicaid Services National Health Expenditures Web site at http://cms.hhs.gov/statistics/nhe/#state.

National Health Care Survey (NHCS)

Centers for Disease Control and Prevention

National Center for Health Statistics

The National Health Care Survey is a family of surveys that collect data from health care providers and establishments about the utilization of health services and characteristics of providers and their patients. The components of the NHCS represent the major sectors of the U.S. health care system providing data on ambulatory, inpatient, and long-term care settings. Each survey in the family is based on a multistage sampling design that includes the health care facilities or providers and their records. Data are collected through abstraction of medical records, completion of encounter forms, compilation of data from State and professional associations, purchase of data from commercial abstraction services, and surveys of providers. Data from all survey components are collected from the establishment, and in no case is information received directly from the person receiving care. This family of surveys includes the following components:

- National Ambulatory Medical Care Survey (NAMCS)
- National Hospital Ambulatory Medical Care Survey (NHAMCS)
- National Hospital Discharge Survey (NHDS)
- National Survey of Ambulatory Surgery (NSAS)
- National Home and Hospice Care Survey (NHHCS)
- National Nursing Home Survey (NHHS)

National Health and Nutrition Examination Survey (NHANES)

Centers for Disease Control and Prevention

National Center for Health Statistics

The NHANES program of the National Center for Health Statistics includes a series of cross-sectional nationally representative health examination surveys beginning in 1960. Each cross-sectional survey provides a national estimate for the U.S. population at the time of the survey, enabling examination of trends over time in the U.S. population. In each survey a nationally representative sample of the U.S. population was selected using a complex, stratified, multistage probability cluster sampling design.

For the first program or cycle of the National Health Examination Survey (NHES I), 1960–62, data were collected on the total prevalence of certain chronic diseases as well as the distributions of various physical and physiological measures, including blood pressure and serum cholesterol levels. For that program, a highly stratified, multistage probability sample of 7,710 adults, of whom 86.5 percent were examined, was selected to represent the 111 million civilian noninstitutionalized adults 18–79 years of age in the United States at that time. The sample areas consisted of 42 primary sampling units (PSUs) from the 1,900 geographic units.

NHES II (1963–65) and NHES III (1966–70) examined probability samples of the Nation's noninstitutionalized children ages 6–11 years (NHES II) and 12–17 years (NHES III) focusing on factors related to growth and development. Both cycles were multistage, stratified probability samples of clusters of households in land-based segments and used the same 40 PSUs. NHES II sampled 7,417 children with a response rate of 96 percent. NHES III sampled 7,514 youth with a response rate of 90 percent.

For more information on NHES I, see: Gordon T, Miller HW. Cycle I of the Health Examination Survey: Sample and response, United States, 1960–62. National Center for Health Statistics. Vital Health Stat 11(1). 1974. For more information on NHES II, see: Plan, operation, and response results of a program of children's examinations. National Center for Health Statistics. Vital Health Stat 1(5). 1967. For more information on NHES III, see: Schaible WL. Quality control in a National Health Examination Survey. National Center for Health Statistics. Vital Health Stat 2(44). 1972.

In 1971 a nutrition surveillance component was added and the survey name was changed to the National Health and Nutrition Examination Survey (NHANES). In NHANES I, conducted from 1971 to 1974, a major purpose was to measure and monitor indicators of the nutrition and health status of the American people through dietary intake data, biochemical tests, physical measurements, and clinical assessments for evidence of nutritional deficiency. Detailed examinations were given by dentists, ophthalmologists, and dermatologists with an assessment of need for treatment. In addition, data were obtained for a subsample of adults on overall health care needs and behavior, and more detailed examination data were collected on cardiovascular, respiratory, arthritic, and hearing conditions.

The NHANES I target population was the civilian noninstitutionalized population 1–74 years of age residing in the coterminous United States, except for people residing on any of the reservation lands set aside for the use of American Indians. The sample design was a multistage, stratified probability sample of clusters of persons in land-based segments. The sample areas consisted of 65 PSUs selected from the 1,900 PSUs in the coterminous United States. A subsample of persons 25–74 years of age was selected to receive the more detailed health examination. Groups at high risk of malnutrition were oversampled at known rates throughout the process. Household interviews were completed for more than 96 percent of the 28,043 persons selected for the NHANES I sample, and about 75 percent (20,749) were examined.

For NHANES II, conducted from 1976 to 1980, the nutrition component was expanded. In the medical area primary emphasis was placed on diabetes, kidney and liver functions, allergy, and speech pathology. The NHANES II target population was the civilian noninstitutionalized population 6 months–74 years of age residing in the United States, including Alaska and Hawaii.

NHANES II used a multistage probability design that involved selection of PSUs, segments (clusters of households) within PSUs, households, eligible persons, and finally, sample persons. The sample design provided for oversampling among persons 6 months–5 years of age, 60–74 years of age, and those living in poverty areas. A sample of 27,801 persons was selected for NHANES II. Of this sample 20,322 (73.1 percent) were examined. Race information for NHANES I and NHANES II was determined primarily by interviewer observation.

The estimation procedure used to produce national statistics for NHANES I and NHANES II 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.

For more information on NHANES I, see: Miller HW. Plan and operation of the Health and Nutrition Examination Survey, United States, 1971–73. National Center for Health Statistics. Vital Health Stat 1(10a) and 1(10b). 1977 and 1978; and Engel A, Murphy RS, Maurer K, Collins E. Plan and operation of the NHANES I Augmentation Survey of Adults 25–74 years, United States, 1974–75. National Center for Health Statistics. Vital Health Stat 1(14). 1978.

For more information on NHANES II, see: McDowell A, Engel A, Massey JT, Maurer K. Plan and operation of the second National Health and Nutrition Examination Survey, 1976–80. National Center for Health Statistics. Vital Health Stat 1(15). 1981. For information on nutritional applications of these surveys, see: Yetley E, Johnson C. Nutritional applications of the Health and Nutrition Examination Surveys (HANES). Ann Rev Nutr 7:441–63. 1987.

The Hispanic Health and Nutrition Examination Survey (HHANES), conducted during 1982–84, was similar in content and design to the previous National Health and Nutrition Examination Surveys. The major difference between HHANES and the previous national surveys is that HHANES used a probability sample of three special subgroups of the population living in selected areas of the United States rather than a national probability sample. The three HHANES universes included approximately 84, 57, and 59 percent of the respective 1980 Mexican-, Cuban-, and Puerto Rican-origin populations in the continental United States. Hispanic ethnicity of these populations was determined by self-report.

In the HHANES three geographically and ethnically distinct populations were studied: Mexican Americans living in Texas, New Mexico, Arizona, Colorado, and California; Cuban Americans living in Dade County, Florida; and Puerto Ricans living in parts of New York, New Jersey, and Connecticut. In the Southwest 9,894 persons were selected (75 percent or 7,462 were examined), in Dade County 2,244 persons were selected (60 percent or 1,357 were examined), and in the Northeast 3,786 persons were selected (75 percent or 2,834 were examined).

For more information on HHANES, see: Maurer KR. Plan and operation of the Hispanic Health and Nutrition Examination Survey, 1982–84. National Center for Health Statistics. Vital Health Stat 1(19). 1985.

The third National Health and Nutrition Examination Survey (NHANES III) was a 6-year survey covering the years 1988–94. Over the 6-year period, 39,695 persons were selected for the survey of which 30,818 (77.6 percent) were examined in the mobile examination center. The NHANES III target population was the civilian noninstitutionalized population 2 months of age and over. The sample design provided for oversampling among children 2–35 months of

age, persons 70 years of age and over, black Americans, and Mexican Americans. Race was reported for the household by the respondent.

For more information on NHANES III, see: Ezzati TM, Massey JT, Waksberg J, et al. Sample design: Third National Health and Nutrition Examination Survey. National Center for Health Statistics. Vital Health Stat 2(113). 1992; 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; or visit the NCHS Web site at www.cdc.gov/nchs/nhanes.htm.

Beginning in 1999, NHANES became a continuous, annual survey that can be linked to related Federal Government surveys of the general U.S. population, specifically, the National Health Interview Survey (NHIS) and, in the future, the U.S. Department of Agriculture's (USDA) Continuing Survey of Food Intakes by Individuals (CSFII). The new design also allows increased flexibility in survey content. Since April 1999, NHANES collects data every year from a representative sample of the U.S. population, newborns and older, by in-home personal interviews and physical examinations in the mobile examination center.

The major objectives of continuous NHANES are to:

- estimate the national prevalence of selected diseases and risk factors
- monitor trends in the prevalence, awareness, treatment, and control of selected diseases
- monitor trends in risk behaviors and environmental exposures
- analyze risk factors for selected diseases
- study the relationship between diet, nutrition, and health
- explore emerging public health issues and new technologies
- establish a national probability sample of genetic material for future genetic testing

The sample frame for continuous NHANES is the list of PSUs selected for the current design of the National Health Interview Survey (NHIS). For the current NHIS design, there are 358 PSUs in the annual sample, divided into four panels with each of the four panels comprising a nationally representative sample. Two of the four panels are available for use by the NHANES. Of the approximately 200 PSUs available in the two national panels for the first stage-sampling frame for the NHANES, 120 NHIS PSUs were selected to comprise six annual national samples, and 20 PSUs were randomly assigned to each year in 1999–2004. For each year, a subset of 15 PSUs was selected with the remaining five PSUs held in reserve. With 15 PSUs per year, approximately 5,000 sample persons can be examined.

For 1999, due to a delay in the start of data collection, there were only 12 distinct PSUs. For the purpose of variance estimation, the 1999–2000 survey is considered to have 26 PSUs. In the sample selection for NHANES 1999–2000, there were 22,839 households screened. Of these, 6,005 households had at least one eligible sample person identified for interviewing. There were a total of 12,160 eligible sample persons identified. Of these 9,965 were interviewed and 9,282 were examined. The overall response rate for those interviewed was 81.9 percent (9,965 out of 12,160) and the response rate for those examined was 76.3 percent (9,282 out of 12,160).

With only 2 years of data in NHANES 1999–2000, instead of the 6 years for NHANES III, sample size is smaller and number of geographic units in the sample is more limited. Due to smaller sample sizes, standard errors for a variable in NHANES 1999–2000 will be approximately 70 percent greater than for the corresponding variable in NHANES III.

NHANES 1999–2000 includes oversampling of low-income persons, adolescents 12–19 years, persons 60 years of age and over, African Americans, and Mexican Americans. The sample is not specifically designed to give a nationally representative sample for the total population of Hispanics residing in the United States.

For more information on NHANES 1999–2000, visit the NHANES Web site at www.cdc.gov/nchs/about/major/nhanes/nhanes99–02.htm.

National Health Interview Survey (NHIS)

Centers for Disease Control and Prevention

National Center for Health Statistics

The National Health Interview Survey (NHIS), initiated in 1957, is a continuing nationwide sample survey of the civilian noninstitutionalized population. Data are collected through household interviews. Information is obtained on personal and demographic characteristics including race and ethnicity by

self-report or as reported by an informant. Information is also obtained on illnesses, injuries, impairments, chronic conditions, utilization of health resources, and other health topics.

The sample design plan of NHIS follows a multistage probability design that permits a continuous sampling of the civilian noninstitutionalized population residing in the United States. The survey is designed in such a way that the sample scheduled for each week is representative of the target population, and the weekly samples are additive over time.

In 1985 NHIS adopted several new sample design features although, conceptually, the sampling plan remained the same as the previous design. Two major changes included reducing the number of primary sampling locations from 376 to 198 for sampling efficiency and oversampling the black population to improve the precision of the statistics. The sample was designed so that a typical NHIS sample for the data collection years 1985–94 consisted of approximately 7,500 segments containing about 59,000 assigned households. Of these households, an expected 10,000 were vacant, demolished, or occupied by persons not in the target population of the survey. The expected sample of 49,000 occupied households yielded a probability sample of about 127,000 persons. In 1994 the sample numbered 116,179 persons.

In 1995 the NHIS sample was redesigned again. Major design changes included increasing the number of primary sampling units from 198 to 358 and oversampling the black and Hispanic populations to improve the precision of the statistics. The sample was designed so that a typical NHIS sample for the data collection years 1995–2004 would consist of approximately 7,000 segments. The expected sample of 44,000 occupied respondent households will yield a probability sample of about 106,000 persons. In 1997 the sample numbered 103,477 persons; 98,785 persons in 1998, 97,059 persons in 1999, 100,618 persons in 2000, and 100,760 persons in 2001.

The NHIS questionnaire fielded from 1982 to 1996 consisted of two parts: a set of basic health and demographic items known as the Core questionnaire and one or more sets of questions on current health topics (supplements). Information was collected from responsible family members residing in the household. Proxy responses were acceptable for Core and Supplement questionnaires when family members were not present at the time of interview. Data for children were collected from proxy respondents.

In 1997 the NHIS questionnaire was redesigned and from 1997 through the present consists of three parts: a basic module, a periodic module, and a topical module. The basic module functions as the new Core questionnaire and comprises three components (Family Core, Sample Adult Core, Sample Child Core). For the Family Core, information is obtained about all members of the family by interviewing any adult members of the household who are present and who may respond for themselves and as proxies for other members of the family. Information is obtained by asking respondents or proxy respondents a series of questions in an unfolding family style. For example, questions on activity limitation are asked as follows: “Are you/any family members limited in activities?” If so, “Who is this?” For the Sample Adult Core, one adult in the household is randomly selected to participate; proxy respondents are not used in this component. For families with children under 18 years of age, one child in the household is randomly selected for participation in the Sample Child Core. Data for this component are collected from a knowledgeable adult in the household. Starting with 1998 periodic and topical modules are incorporated into selected years of the NHIS.

In the 1997 NHIS questionnaire redesign the measurement of some basic concepts was changed and some concepts were measured in different ways. While some questions remain the same over time, they may be preceded by different questions or topics. For some questions, there was a change in the reference period for reporting an event or condition.

Also in 1997 the collection methodology changed from paper and pencil questionnaires to computer-assisted personal interviewing (CAPI). Because of the extensive redesign of the questionnaire in 1997 and introduction of the CAPI method of data collection, data from 1997 and later years may not be comparable with data from earlier years.

The household response rate for the ongoing portion of the survey (core) has been between 94 and 98 percent over the years. In recent years the total household response rate was 92 percent in 1997, 90 percent in 1998, 88 percent in 1999, and 89 percent in 2000 and 2001. Response rates for special health topics (supplements) have generally been lower. For example, the response rate was 80 percent for the 1994 Year 2000 Supplement, which included questions about cigarette smoking and use of such preventive services as mammography. In 1997 the final response rate for the sample adult supplement was 80 percent, 74 percent in 1998, 70 percent in 1999, 72 percent in 2000, and 74 percent in

2001. In 1997 the final response rate for the sample child supplement was 84 percent, 82 percent in 1998, 78 percent in 1999, 79 percent in 2000, and 81 percent in 2001.

For more information about the survey design, methods used in estimation, and general qualifications of the data obtained from the survey, see: Botman SL, Moore TF, Moriarity CL, and Parsons VL. Design and estimation for the National Health Interview Survey, 1995–2004. National Center for Health Statistics. Vital Health Stat 2(130). 2000; Massey JT, Moore TF, Parsons VL, Tadros W. Design and estimation for the National Health Interview Survey, 1985–94. National Center for Health Statistics. Vital Health Stat 2(110). 1989; Kovar MG, Poe GS. The National Health Interview Survey design, 1973–84, and procedures, 1975–83. National Center for Health Statistics. Vital Health Stat 1(18). 1985; Blackwell DL, Tonthat L. Summary Health Statistics for U.S. Children: National Health Interview Survey, 1998. National Center for Health Statistics. Vital Health Stat 10(208). 2002; Blackwell DL, Tonthat L. Summary Health Statistics for the U.S. Population: National Health Interview Survey, 1998. National Center for Health Statistics. Vital Health Stat 10(207). 2002; Pleis JR, Coles R. Summary Health Statistics for U.S. Adults: National Health Interview Survey, 1998. National Center for Health Statistics. Vital Health Stat 10(209). 2002; or visit the NHIS section of the NCHS Web site at www.cdc.gov/nchs/nhis.htm.

National Health Provider Inventory (NHPI)

Centers for Disease Control and Prevention

National Center for Health Statistics

The National Master Facility Inventories (NMFIs), forerunners of the National Health Provider Inventory (NHPI), were a series of inventories of inpatient health facilities in the United States conducted by NCHS. The inventories included hospitals, nursing and related-care homes, and other custodial care facilities. The last NMFI was conducted in 1982. In 1986 the inventory was changed to the Inventory of Long-Term Care Places (ILTCP) and included nursing and related-care homes and facilities for the mentally retarded. In 1991 the inventory was again changed to NHPI and included nursing homes, board and care homes, home health agencies, and hospices. The NHPI has not been repeated since 1991. The NMFI, ILTCP, and NHPI served as sampling frames for the NCHS National Nursing Home Survey and National Home and Hospice Care Survey.

National Home and Hospice Care Survey (NHHCS)

Centers for Disease Control and Prevention

National Center for Health Statistics

The National Home and Hospice Care Survey (NHHCS) is a sample survey of health agencies and hospices. Initiated in 1992, it was also conducted in 1993, 1994, 1996, 1998, and 2000. The original sampling frame consisted of all home health care agencies and hospices identified in the 1991 National Health Provider Inventory (NHPI). The 1992 sample contained 1,500 agencies. These agencies were revisited during the 1993 survey (excluding agencies that had been found to be out of scope for the survey). In 1994 in-scope agencies identified in the 1993 survey were revisited, along with 100 newly identified agencies added to the sample. In 1996 the universe was again updated and a new sample of 1,200 agencies was drawn. In 1998 a sample of 1,350 agencies was selected. In 2000, 1,800 agencies were sampled and the response rate was 96.4 percent.

The sample design for the 1992–94 NHHCS was a stratified three-stage probability design. Primary sampling units were selected at the first stage, agencies were selected at the second stage, and current patients and discharges were selected at the third stage. The sample design for the 1996, 1998, and 2000 NHHCS was a two-stage probability design, in which agencies were selected at the first stage and current patients and discharges were selected at the second stage. Current patients were those on the rolls of the agency as of midnight the day before the survey. Discharges were selected to estimate the number of discharges from the agency during the year before the survey. After the samples were selected, a patient questionnaire was completed for each current patient and discharge by interviewing the staff member most familiar with the care provided to the patient. The respondent was requested to refer to the medical records for each patient.

For additional information see: Haupt BJ. Development of the National Home and Hospice Care Survey. National Center for

Health Statistics. Vital Health Stat 1(33). 1994; and visit the National Health Care Survey (NHCS) Web site at www.cdc.gov/nchs/nhcs.htm.

National Hospital Ambulatory Medical Care Survey (NHAMCS)

Centers for Disease Control and Prevention

National Center for Health Statistics

The National Hospital Ambulatory Medical Care Survey (NHAMCS), initiated in 1992, is a continuing annual national probability sample of visits by patients to emergency departments (EDs) and outpatient departments (OPDs) of non-Federal, short-stay, or general hospitals. Telephone contacts are excluded.

A four-stage probability sample design is used in NHAMCS, involving samples of primary sampling units (PSUs), hospitals with EDs and/or OPDs within PSUs, EDs within hospitals and/or clinics within OPDs, and patient visits within EDs and/or clinics. In 1999 the hospital response rate for NHAMCS was 93 percent for EDs and 86 percent for OPDs. In 2000 the hospital response rate was 94 percent for EDs and 88 percent for OPDs. Hospital staff were asked to complete Patient Record Forms (PRF) for a systematic random sample of patient visits occurring during a randomly assigned 4-week reporting period. On the PRF, up to three physicians' diagnoses were collected and coded by NCHS to the *International Classification of Diseases, Clinical Modification* (ICD-9-CM). Additionally, if the cause-of-injury check box was marked on the PRF, up to three external causes of injury were coded by NCHS to the ICD-9-CM Supplementary Classification of External Causes of Injury and Poisoning. In 1999 the number of PRFs completed for EDs was 21,103 and for OPDs 29,487. In 2000 the number of PRFs completed for EDs was 25,622 and for OPDs 27,510.

For more detailed information on NHAMCS, see: McCaig LF, McLemore T. Plan and operation of the National Hospital Ambulatory Medical Care Survey. National Center for Health Statistics. Vital Health Stat 1(34). 1994; and visit the National Health Care Survey (NHCS) section of the NCHS Web site at www.cdc.gov/nchs/nhcs.htm.

National Hospital Discharge Survey (NHDS)

Centers for Disease Control and Prevention

National Center for Health Statistics

The National Hospital Discharge Survey (NHDS), which has been conducted annually since 1965, is a national probability survey designed to meet the need for information on characteristics of inpatients discharged from non-Federal short-stay hospitals in the United States. The survey is conducted in all 50 States and the District of Columbia. Only hospitals with an average length of stay of fewer than 30 days for all patients, general hospitals, or children's general hospitals are included in the survey. Federal, military, and Department of Veterans Affairs hospitals, as well as hospital units of institutions (such as prison hospitals), and hospitals with fewer than six beds staffed for patient use, are excluded. All discharged patients from in-scope hospitals are included in the survey; however, discharged newborn infants are not included in *Health, United States*.

The original sample was selected in 1964 from a frame of short-stay hospitals listed in the National Master Facility Inventory. A two-stage stratified sample design was used, with hospitals stratified according to bed size and geographic region. Sample hospitals were selected with probabilities ranging from certainty for the largest hospitals to 1 in 40 for the smallest hospitals. Within each sample hospital, a systematic random sample of discharges was selected from the daily listing sheet. Initially, the within-hospital sampling rates for selecting discharges varied inversely with the probability of hospital selection, so that the overall probability of selecting a discharge was approximately the same across the sample. Those rates were adjusted for individual hospitals in subsequent years to control the reporting burden of those hospitals.

In 1985, for the first time, two data-collection procedures were used for the survey. The first was the traditional manual system of sample selection and data abstraction. In the manual system, sample selection and transcription of information from the hospital records to abstract forms were performed by either the hospital staff or representatives of NCHS or both. The second was an automated method, used in approximately 17 percent of the sample hospitals in 1985, involving the purchase of data tapes from commercial

abstracting services. These tapes were then subjected to the NCHS sampling, editing, and weighting procedures.

In 1988 NHDS was redesigned. The hospitals with the most beds and/or discharges annually were selected with certainty, but the remaining sample was selected using a three-stage stratified design. The first stage is a sample of PSUs used by the National Health Interview Survey. Within PSUs, hospitals were stratified or arrayed by abstracting status (whether subscribing to a commercial abstracting service) and within abstracting status arrayed by type of service and bed size. Within these strata and arrays, a systematic sampling scheme with probability proportional to the annual number of discharges was used to select hospitals. The rates for systematic sampling of discharges within hospitals varied inversely with probability of hospital selection within the PSU. Discharge records from hospitals submitting data via commercial abstracting services and selected State data systems (approximately 40 percent of sample hospitals) were arrayed by primary diagnoses, patient sex and age group, and date of discharge before sampling. Otherwise, the procedures for sampling discharges within hospitals were the same as those used in the prior design.

In 2000 the hospital sample was updated by continuing the sampling process among hospitals that were NHDS-eligible for the sampling frame in 2000 but not in 1997. The additional hospitals were added at the end of the list for the strata where they belonged, and the systematic sampling was continued as if the additional hospitals had been present during the initial sample selection. Hospitals that were no longer NHDS-eligible were deleted. A similar updating process occurred in 1991, 1994, and 1997.

The basic unit of estimation for NHDS is the sample patient abstract. The estimation procedure involves inflation by the reciprocal of the probability of selection, adjustment for nonresponding hospitals and missing abstracts, and ratio adjustments to fixed totals. In 2000, 509 hospitals were selected, 481 were within scope, 434 participated (90 percent), and 313,259 medical records were abstracted. In 2001, the sample consisted of 504 hospitals, of which 477 were within scope and 448 of those participated, providing data for approximately 330,000 discharges.

Hospital utilization rates per 1,000 population were computed using estimates of the civilian population of the United States as of July 1 of each year. Rates for 1990 through 1999 use postcensal estimates of the civilian population based on the 1990 census adjusted for net underenumeration using the 1990 National Population Adjustment Matrix from the U.S. Bureau of the Census. These estimates will differ from estimates that calculate discharge rates for 1990–1999 based on estimates of the civilian population that incorporate information from the census 2000 (intercensal estimates—not currently available) thereby adjusting for the "error of closure." The estimates for 2000 that appeared in *Health, United States, 2002* were computed using postcensal civilian population estimates based on the 1990 Census adjusted for net underenumeration. The estimates for 2000 and 2001 that appear in *Health, United States, 2003* were calculated using estimates of the civilian population based on census 2000, and therefore are not directly comparable with rates calculated for the 1990s. See related *Population Census and Population Estimates*.

For more detailed information on the design of NHDS and the magnitude of sampling errors associated with NHDS estimates, see: Hall MJ, DeFrances CJ. 2001 National Hospital Discharge Summary Advance data from vital and health statistics; no 332. Hyattsville, MD: National Center for Health Statistics. 2003; Dennison C, Pokras R. Design and operation of the National Hospital Discharge Survey: 1988 redesign. National Center for Health Statistics. Vital Health Stat 1(39). 2000; and visit the National Health Care Survey Web site at www.cdc.gov/nchs/nhcs.htm.

National Household Survey on Drug Abuse (NHSDA)

Substance Abuse and Mental Health Services Administration

The National Household Survey on Drug Abuse (NHSDA), sponsored by the Substance Abuse and Mental Health Services Administration (SAMHSA), collects data on use of tobacco, alcohol, and illicit drugs among persons 12 years of age and over in the civilian noninstitutionalized population in the United States. This includes civilians living on military bases and persons living in noninstitutionalized group quarters, such as college dormitories, rooming houses, and shelters. Persons excluded from the survey include homeless people who do not use shelters, active military personnel, and residents of institutional group quarters, such as jails and hospitals.

The NHSDA survey has been conducted since 1971. In 1999 the NHSDA underwent a major redesign affecting the method of data collection, sample design, sample size, and oversampling. Because of the differences in methodology and impact of the new design on data collection, comparisons should not be made between data from the redesigned surveys (1999 onward) and data obtained from surveys prior to 1999. Beginning in 1999 the survey used a combination of computer-assisted personal interview (CAPI) conducted by the interviewer and a computer-assisted self-interview (ACASI). Use of ACASI is designed to provide the respondent with a highly private and confidential means of responding to questions and to increase the level of honest reporting of illicit drug use and other sensitive behaviors.

A 5-year sample design provides State estimates for years 1999 through 2003. The sample employs a 50-State design with an independent, multistage area probability sample for each of the 50 States and the District of Columbia. The eight States with the largest population (which together account for 48 percent of the total U.S. population age 12 years and over) were designated as large sample States (California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas). For these States the design provided a sample large enough to support direct State estimates. For the remaining 42 States and the District of Columbia, smaller, but adequate, samples were selected to support State estimates using small-area estimation techniques. The design also oversamples youths and young adults so that each State's sample is approximately equally distributed among three major age groups: 12–17 years, 18–25 years, and 26 years and over.

Each State was stratified into regions (48 regions in each of 8 large States, 12 regions in each of 42 small States and the District of Columbia). At the first stage of sampling, 8 area segments were selected in each region, for a total of 7,200 sample units nationally. In these segments, 171,519 addresses were screened and 89,745 persons were interviewed within the screened addresses in 2001. Weighted response rates for household screening and for interviewing were 91.9 percent and 73.3 percent, respectively, for an overall weighted response rate of 67.3 percent. A description of the methodology can be found in Summary of Findings from the 2001 National Household Survey on Drug Abuse, available from SAMHSA's Web site.

Direct survey estimates considered to be unreliable due to unacceptably large sampling errors are not shown in table 62 in this report, and are noted by asterisks (*). The criterion used for suppressing all direct survey estimates was based on the relative standard error (*RSE*), which is defined as the ratio of the standard error (*se*) over the estimate. Proportion estimates (p) within the range [$0<p<1$], rates, and corresponding estimated number of users were suppressed if:

$$[se(p) / p] / [-ln(p)] > 0.175 \text{ when } p<0.5$$

or

$$[se(p) / (1-p)] / [-ln(1-p)] > 0.175 \text{ when } p \geq 0.5$$

The separate formulae for $p<0.5$ and $p \leq 0.5$ produces a symmetric suppression rule; that is, if p is suppressed, then so will $1-p$. This is an ad hoc rule that requires an effective sample size in excess of 50. When $0.05<p<0.95$, the symmetric properties of the rule produce a local maximum effective sample size of 68 at $p=0.5$. Thus, estimates with these values of p along with effective sample sizes falling below 68 are suppressed. A local minimum effective sample size of 50 occurs at $p=0.2$ and again at $p=0.8$ within this same interval; so, estimates are suppressed for values of p with effective sample sizes below 50. A minimum effective sample size of 68 was added to the suppression criteria in the 2000 NHSDA. As p approaches 0.00 or 1.00 outside the interval (0.05, 0.95), the suppression criteria will still require increasingly larger effective sample sizes. Also new to the 2000 survey is a minimum nominal sample size suppression criteria ($n=100$) that protects against unreliable estimates caused by small design effects and small nominal sample sizes. Prevalence estimates are also suppressed if they are close to zero or 100 percent (i.e., if $p<.00005$ or if $p>.99995$).

Estimates of substance use for youth based on the NHSDA are generally lower than estimates based on Monitoring the Future (MTF) and Youth Risk Behavior Surveillance System (YRBSS). In addition to the fact that the MTF excludes dropouts and absentees, rates are not directly comparable across these surveys, due to differences in populations covered, sample design, questionnaires, interview setting, and statistical approaches to make the survey estimates generalizable to the entire population. The NHSDA survey collects data in homes, whereas the MTF and YRBSS collect data in school classrooms. The NHSDA estimates are tabulated by age, while the MTF and YRBSS estimates are tabulated by grade, representing different ages as well as different populations. See Cowan CD. Coverage, Sample

Design, and Weighting in Three Federal Surveys. Journal of Drug Issues 31(3), 595–614, 2001.

For more information on the National Household Survey on Drug Abuse (NHSDA), see: NHSDA Series: H-13 Summary of Findings from the 2000 National Household Survey on Drug Abuse, DHHS Pub No (SMA) 01-3549; or write: Office of Applied Studies, Substance Abuse and Mental Health Services Administration, Room 16C-06, 5600 Fishers Lane, Rockville, MD 20857; for the 2001 NHSDA Summary of National Findings, visit the SAMHSA Web site at www.drugabusestatistics.samhsa.gov.

National Immunization Survey (NIS)

Centers for Disease Control and Prevention

National Center for Health Statistics and National Immunization Program

The National Immunization Survey (NIS) is a continuing nationwide telephone sample survey to gather data on children 15–35 months of age. Estimates of vaccine-specific coverage are available for national, State, and 28 urban areas considered to be high risk for undervaccination.

NIS uses a two-phase sample design. First, a random-digit-dialing (RDD) sample of telephone numbers is drawn. When households with age-eligible children are contacted, the interviewer collects information on the vaccinations received by all age-eligible children. In 2001 the overall response rate was 70 percent, yielding data for 33,437 children aged 15–35 months. The interviewer also collects information on the vaccination providers. In the second phase, all vaccination providers are contacted by mail. The vaccination information from providers was obtained for 72 percent of all children who were eligible for provider followup in 2001. Providers' responses are combined with information obtained from the households to provide a more accurate estimate of vaccination coverage levels. Final estimates are adjusted for households without telephones.

For more information about the survey design and methods used in estimation, see: Zell ER, Ezzati-Rice TM, Battaglia PM, Wright RA. National Immunization Survey: The Methodology of a Vaccination Surveillance System. Public Health Reports 115:65–77. 2000; or visit the NCHS Web site at www.cdc.gov/nis.

National Medical Expenditure Survey (NMES)—See *Medical Expenditure Panel Survey*.

National Notifiable Diseases Surveillance System (NNDSS)

Centers for Disease Control and Prevention

Epidemiology Program Office

The Epidemiology Program Office (EPO) of CDC, in partnership with the Council of State and Territorial Epidemiologists (CSTE), operates the National Notifiable Diseases Surveillance System. The primary purpose of this system is to provide weekly provisional information on the occurrence of diseases defined as notifiable by CSTE. The system also provides annual summaries of the data. State epidemiologists report cases of notifiable diseases to EPO who tabulates and publishes these data in the *Morbidity and Mortality Weekly Report (MMWR)* and the *Summary of Notifiable Diseases, United States* (entitled *Annual Summary* before 1985). Notifiable disease surveillance is conducted by public health practitioners at local, State, and national levels to support disease prevention and control activities.

Notifiable disease reports are received from health departments in the 50 States, 5 territories, New York City, and the District of Columbia. Policies for reporting notifiable disease cases can vary by disease or reporting jurisdiction, depending on case status classification (i.e., confirmed, probable, or suspect). CSTE and CDC annually review the status of national infectious 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, Q fever and tularemia became nationally notifiable in 2000. However, reporting nationally notifiable diseases to CDC is voluntary. Reporting is currently mandated by law or regulation only at the local and State level. Therefore, the list of diseases that are considered notifiable varies slightly by State. For example, reporting of cyclosporiasis to CDC is not done by some States in which this disease is not notifiable to local or State authorities. More information regarding notifiable diseases, including case definitions for these conditions, is available on the Web at www.cdc.gov/epo/dphsi/phs.htm.

Notifiable disease data are useful for analyzing disease trends and determining relative disease burdens. However, these

data must be interpreted in light of reporting practices. Some diseases that cause severe clinical illness (for example, plague and rabies) are most likely reported accurately if diagnosed by a clinician. However, persons who have diseases that are clinically mild and infrequently associated with serious consequences (for example, salmonellosis) might 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 also is 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.

For more information, see: Centers for Disease Control and Prevention, Summary of Notifiable Diseases, United States, 2001 *Morbidity and Mortality Weekly Report* 50(53) Public Health Service, DHHS, Atlanta, GA, 2003; or write: Chief, Surveillance Systems Branch, Division of Public Health Surveillance and Informatics. Epidemiology Program Office, Centers for Disease Control and Prevention, 4770 Buford Highway, MS K74, Atlanta, GA 30341-3717; or visit the EPO Web site at www.cdc.gov/epo/.

National Nursing Home Survey (NNHS)

Centers for Disease Control and Prevention

National Center for Health Statistics

NCHS conducted six National Nursing Home Surveys (NNHS), the first survey from August 1973–April 1974; the second from May–December 1977; the third from August 1985–January 1986; the fourth from July–December 1995; the fifth from July–December 1997; and the sixth from July–December 1999. The next NNHS, which has undergone a major redesign, is scheduled to be conducted during calendar year 2004.

For the initial NNHS conducted in 1973–74, the universe included nursing homes that provided some level of nursing care and excluded homes providing only personal or domiciliary care. The sample of 2,118 homes was selected from the 17,685 homes listed in the 1971 National Master Facility Inventory (NMFI) or those that opened for business in 1972. Data were obtained from about 20,600 staff and 19,000 residents. Response rates were 97 percent for facilities, 88 percent for expenses, 82 percent for staff, and 98 percent for residents.

The 1977 NNHS encompassed all types of nursing homes, including personal care and domiciliary care homes. The sample of about 1,700 facilities was selected from 23,105 nursing homes in the sampling frame, which consisted of all homes listed in the 1973 NMFI and those opening for business between 1973 and December 1976. Data were obtained from about 13,600 staff, 7,000 residents, and 5,100 discharged residents. Response rates were 95 percent for facilities, 85 percent for expenses, 81 percent for staff, 99 percent for residents, and 97 percent for discharges.

The 1985 NNHS was similar to the 1973–74 survey in that it excluded personal or domiciliary care homes. The sample of 1,220 homes was selected from a sampling frame of 20,479 nursing and related-care homes. The frame consisted of all homes in the 1982 NMFI; homes identified in the 1982 Complement Survey of NMFI "missing" from the 1982 NMFI; facilities that opened for business between 1982 and June 1984; and hospital-based nursing homes obtained from the Centers for Medicare & Medicaid Services. Information on the facility was collected through a personal interview with the administrator. Accountants were asked to complete a questionnaire on expenses or provide a financial statement. Resident data were provided by a nurse familiar with the care provided to the resident. The nurse relied on the medical record and personal knowledge of the resident. In addition to employee data that were collected during the interview with the administrator, a sample of registered nurses completed a self-administered questionnaire. Discharge data were based on information recorded in the medical record. Additional data about current and discharged residents were obtained in telephone interviews with next of kin. Data were obtained from 1,079 facilities, 2,763 registered nurses, 5,243 current residents, and 6,023 discharges. Response rates were 93 percent for facilities, 68 percent for expenses, 80 percent for registered nurses, 97 percent for residents, 95 percent for discharges, and 90 percent for next of kin.

The 1995, 1997, and 1999 NNHS also included only nursing homes that provided some level of nursing care, and excluded homes providing only personal or domiciliary care,

similar to the 1985 and 1973–74 surveys. The 1995 sample of 1,500 homes was selected from a sampling frame of 17,500 nursing homes. The frame consisted of an updated version of the 1991 National Health Provider Inventory (NHPI). Data were obtained from about 1,400 nursing homes and 8,000 current residents. Data on current residents were provided by a staff member familiar with the care received by residents and from information contained in residents' medical records.

The 1997 sample of 1,488 nursing homes was the same basic sample used in 1995. Excluded were out-of-scope and out-of-business places identified in the 1995 survey. Included were a small number of additions to the sample from a supplemental frame of places not in the 1995 frame. The 1997 NNHS included the discharge component not available in the 1995 survey.

The 1999 sample of 1,423 nursing homes was selected from a sampling frame of 18,419 possible facilities from the most current National Health Provider Inventory. A supplemental frame was used to add facilities not in the 1997 frame. Like the 1995 and 1997 surveys, the 1999 survey excluded out-of-scope and out-of-business nursing homes identified in 1997. The 1999 NNHS also included a discharge resident component.

Statistics for the National Nursing Home Surveys are derived by a multistage estimation procedure that provides essentially unbiased national estimates and has three major components: (a) inflation by the reciprocals of the probabilities of sample selection, (b) adjustment for nonresponse, and (c) ratio adjustment to fixed totals. The surveys are adjusted for three types of nonresponse: (1) when an eligible nursing facility did not respond; (2) when the facility failed to complete the sampling lists; and (3) when the facility did not complete the facility questionnaire but did complete the questionnaire for residents in the facility.

For more information on the 1973–74 NNHS, see: Meiners MR. Selected operating and financial characteristics of nursing homes, United States, 1973–74 National Nursing Home Survey. National Center for Health Statistics. Vital Health Stat 13(22). 1975. For more information on the 1977 NNHS, see: Van Nostrand JF, Zappolo A, Hing E, et al. The National Nursing Home Survey, 1977 summary for the United States. National Center for Health Statistics. Vital Health Stat 13(43). 1979. For more information on the 1985 NNHS, see: Hing E, Sekscenski E, Strahan G. The National Nursing Home Survey: 1985 summary for the United States. National Center for Health Statistics. Vital Health Stat 13(97). 1989. For more information on the 1995 NNHS, see: Strahan G. An overview of nursing homes and their current residents: Data from the 1995 National Nursing Home Survey. Advance data from vital and health statistics; no 280. Hyattsville, MD: National Center for Health Statistics. 1997. For more information on the 1997 NNHS, see: The National Nursing Home Survey: 1997 summary. National Center for Health Statistics. Vital Health Stat 13(147). 2000. For more information on the 1999 NNHS, see: Jones A. The National Nursing Home Survey: 1999 summary. National Center for Health Statistics. Vital Health Stat 13(152). 2002. Information about the 1995, 1997, 1999, and 2001 NNHS is also available at the National Health Care Survey Web site at www.cdc.gov/nchs/nhcs.htm.

National Survey of Ambulatory Surgery (NSAS)

Centers for Disease Control and Prevention

National Center for Health Statistics

The National Survey of Ambulatory Surgery (NSAS) is a nationwide sample survey of ambulatory surgery patient discharges from short-stay non-Federal hospitals and freestanding surgery centers. NSAS was conducted in 1994, 1995, and 1996. The sample consisted of eligible hospitals listed in the 1993 SMG Hospital Market Database and the 1993 SMG Freestanding Outpatient Surgery Center Database or Medicare Provider-of-Service files. Facilities specializing in dentistry, podiatry, abortion, family planning, or birthing were excluded.

A three-stage stratified cluster design was used, and facilities were stratified according to primary sampling unit (PSU). The second stage consisted of the selection of facilities from sample PSUs, and the third stage consisted of a systematic random sample of cases from all locations within a facility where ambulatory surgery was performed. Locations within hospitals dedicated exclusively to dentistry, podiatry, pain block, abortion, or small procedures (sometimes referred to as "lump and bump" rooms) were not included. In 1996, of the 751 hospitals and freestanding ambulatory surgery centers selected for the survey, 601 were in-scope and 488 responded for an overall response rate of 81 percent. These facilities provided information for approximately 125,000 ambulatory surgery discharges. Up to six procedures were

coded to the *International Classification of Diseases, 9th Revision, Clinical Modification*. Estimates were derived using a multistage estimation procedure: inflation by reciprocals of the probabilities of selection; adjustment for nonresponse; and population weighting ratio adjustments.

For more detailed information on the design of NSAS, see: McLemore T, Lawrence L. Plan and operation of the National Survey of Ambulatory Surgery. National Center for Health Statistics. Vital Health Stat 1(37). 1997; and visit the National Health Care Survey Web site at www.cdc.gov/nchs/nhcs.htm.

National Survey of Family Growth (NSFG)

Centers for Disease Control and Prevention

National Center for Health Statistics

Data from the National Survey of Family Growth (NSFG) are based on samples of women ages 15–44 years in the civilian noninstitutionalized population of the United States. The first and second cycles, conducted in 1973 and 1976, excluded most women who had never been married. The third, fourth, and fifth cycles, conducted in 1982, 1988, and 1995, included all women ages 15–44 years.

The purpose of the survey is to provide national data on factors affecting birth and pregnancy rates, adoption, and maternal and infant health. These factors include sexual activity, marriage, divorce and remarriage, unmarried cohabitation, contraception and sterilization, infertility, breastfeeding, pregnancy loss, low birthweight, and use of medical care for family planning and infertility.

Interviews are conducted in person by professional female interviewers using a standardized questionnaire. In 1973–88 the average interview length was about 1 hour. In 1995 the average interview lasted about 1 hour and 45 minutes. In all cycles black women were sampled at higher rates than white women, so that detailed statistics for black women could be produced.

Interviewing for Cycle 1 of NSFG was conducted from June 1973 to February 1974. Counties and independent cities of the United States were sampled to form a frame of primary sampling units (PSUs), and 101 PSUs were selected. From these 101 PSUs, 10,879 women 15–44 years of age were selected, 9,797 of these were interviewed. Most never-married women were excluded from the 1973 NSFG.

Interviewing for Cycle 2 of NSFG was conducted from January to September 1976. From 79 PSUs, 10,202 eligible women were identified; of these, 8,611 were interviewed. Again, most never-married women were excluded from the sample for the 1976 NSFG.

Interviewing for Cycle 3 of NSFG was conducted from August 1982 to February 1983. The sample design was similar to that in Cycle 2: 31,027 households were selected in 79 PSUs. Household screener interviews were completed in 29,511 households (95.1 percent). Of the 9,964 eligible women identified, 7,969 were interviewed. For the first time in NSFG, Cycle 3 included women of all marital statuses.

Interviewing for Cycle 4 was conducted between January and August 1988. The sample was obtained from households that had been interviewed in the National Health Interview Survey in the 18 months between October 1, 1985 and March 31, 1987. For the first time, women living in Alaska and Hawaii were included so that the survey covered women from the noninstitutionalized population of the entire United States. The sample was drawn from 156 PSUs; 10,566 eligible women ages 15–44 years were sampled. Interviews were completed with 8,450 women.

Between July and November 1990, 5,686 women were interviewed by telephone in the first NSFG telephone reinterview. The average length of interview in 1990 was 20 minutes. The response rate for the 1990 telephone reinterview was 68 percent of those responding to the 1988 survey and still eligible for the 1990 survey.

Interviewing for Cycle 5 of NSFG was conducted between January and October 1995. The sample was obtained from households that had been interviewed in 198 PSUs in the National Health Interview Survey in 1993. Of the 13,795 eligible women in the sample, 10,847 were interviewed. For the first time, Hispanic as well as black women were sampled at a higher rate than other women.

In order to make national estimates from the sample for the millions of women ages 15–44 years in the United States, data for the interviewed sample women were (a) inflated by the reciprocal of the probability of selection at each stage of sampling (for example, if there was a 1 in 5,000 chance that a woman would be selected for the sample, her sampling weight was 5,000), (b) adjusted for nonresponse, and (c) forced to agree with benchmark population values based on data from the Current Population Survey of the U.S. Bureau of the Census (this last step is called "poststratification").

Quality control procedures for selecting and training interviewers, and coding, editing, and processing data were built into NSFG to minimize nonsampling error.

More information on the methodology of NSFG is available in the following reports: French DK. National Survey of Family Growth, Cycle I: Sample design, estimation procedures, and variance estimation. National Center for Health Statistics. Vital Health Stat 2(76). 1978; Grady WR. National Survey of Family Growth, Cycle II: Sample design, estimation procedures, and variance estimation. National Center for Health Statistics. Vital Health Stat 2(87). 198l; Bachrach CA, Horn MC, Mosher WD, Shimizu I. National Survey of Family Growth, Cycle III: Sample design, weighting, and variance estimation. National Center for Health Statistics. Vital Health Stat 2(98). 1985; Judkins DR, Mosher WD, Botman SL. National Survey of Family Growth: Design, estimation, and inference. National Center for Health Statistics. Vital Health Stat 2(109). 1991; Goksel H, Judkins DR, Mosher WD. Nonresponse adjustments for a telephone followup to a National In-Person Survey. Journal of Official Statistics 8(4):417–32. 1992; Kelly JE, Mosher WD, Duffer AP, Kinsey SH. Plan and operation of the 1995 National Survey of Family Growth. Vital Health Stat 1(36). 1997; Potter FJ, Iannacchione VG, Mosher WD, Mason RE, Kavee JD. Sampling weights, imputation, and variance estimation in the 1995 National Survey of Family Growth. Vital Health Stat 2(124). 1998; or visit the NCHS Web site at www.cdc.gov/nchs/nsfg.htm.

National Survey of Substance Abuse Treatment Services (N-SSATS)

Substance Abuse and Mental Health Services Administration

The National Survey of Substance Abuse Treatment Services (N-SSATS) is part of the Drug and Alcohol Services Information System (DASIS) maintained by the Substance Abuse and Mental Health Services Administration (SAMHSA). N-SSATS is a census of all known substance abuse treatment facilities. It seeks information from all specialized facilities that treat substance abuse. These include facilities that treat only substance abuse, as well as specialty substance abuse units operating within larger mental health facilities (for example, community mental health centers), general health (for example, hospitals), social service (for example, family assistance centers), and criminal justice (for example, probation departments) agencies. N-SSATS solicits data concerning facility and client characteristics for a specific reference day (October 1 in 1998 and 2000 and March 29 in 2002) including number of individuals in treatment, substance of abuse (alcohol, drugs, or both), and types of services. Public and private facilities are included.

Treatment facilities contacted through N-SSATS are identified from the Inventory of Substance Abuse Treatment Services (I-SATS) that lists all known substance abuse treatment facilities. Response rates to the surveys were 91, 94, and 95 percent in 1998, 2000, and 2002, respectively. The full survey was not conducted in 1999 or 2001.

For further information on N-SSATS, contact:
Office of Applied Studies, Substance Abuse and Mental Health Services Administration, Room 16-105,
5600 Fishers Lane, Rockville, MD 20857; or visit the OAS statistical information section of the SAMHSA Web site at www.drugabusestatistics.samhsa.gov.

National Vital Statistics System

Centers for Disease Control and Prevention

National Center for Health Statistics

Through the National Vital Statistics System, the National Center for Health Statistics (NCHS) collects and publishes data on births, deaths, marriages, and divorces in the United States. Fetal deaths are classified and tabulated separately from other deaths. The Division of Vital Statistics obtains information on births and deaths from the registration offices of all States, New York City, the District of Columbia, Puerto Rico, the U.S. Virgin Islands, and Guam. Geographic coverage for births and deaths has been complete since 1933. Trend tables in this book show data for the aggregate of 50 States, New York City, and the District of Columbia, as well as for each individual State and the District of Columbia.

Until 1972 microfilm copies of all death certificates and a 50-percent 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 percent of their death and birth records and sent the entire data file to NCHS on computer tapes. Currently, data are sent to NCHS

through the Vital Statistics Cooperative Program (VSCP), following the same procedures as CHSS. The number of participating States grew from 6 in 1972 to 46 in 1984. Starting in 1985 all 50 States and the District of Columbia participated in VSCP.

U.S. Standard Certificates—U.S. Standard Live Birth and Death Certificates and Fetal Death Reports are revised periodically, allowing careful evaluation of each item and addition, modification, and deletion of items. Beginning with 1989 revised standard certificates replaced the 1978 versions. The 1989 revision of the birth certificate includes items to identify the Hispanic parentage of newborns and to expand information about maternal and infant health characteristics. The 1989 revision of the death certificate includes 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.

Birth File

The birth file is comprised of demographic and medical information from birth certificates. Demographic information, such as race and ethnicity, is provided by the mother at the time of birth. Medical and health information is based on hospital records. The number of States reporting information on maternal education, Hispanic origin, marital status, and tobacco use during pregnancy has increased over the years spanned by this report. Interpretation of trend data should take into consideration expansion of reporting areas and immigration. See Appendix II for methodologic and reporting area changes for the following birth certificate items: *Age* (maternal age); *Education* (maternal education); *Hispanic origin; Marital status; Prenatal care; Race; Tobacco use*.

For more information, see: National Center for Health Statistics, *Vital Statistics of the United States*, Vol. I Natality, Technical Appendix, available at the NCHS Web site at www.cdc.gov/nchs/births.htm.

Mortality File

The mortality data file is comprised of demographic and medical information from death certificates. Demographic information is provided by the funeral director based on information supplied by an informant. Medical certification of cause of death is provided by a physician, medical examiner, or coroner. The mortality data file is a fundamental source of cause-of-death information by demographic characteristics and for geographic areas, such as States. The mortality file is one of the few sources of comparable health-related data for smaller geographic areas in the United States and over a long time period. Mortality data can be used not only to present the characteristics of those dying in the United States, but also to determine life expectancy and to compare mortality trends with other countries. Data for the entire United States refer to events occurring within the United States; data for geographic areas are by place of residence. See Appendix II for methodologic and reporting area changes for the following death certificate items: *Education; Hispanic origin; Race*.

For more information, see: Grove RD, Hetzel AM. *Vital statistics rates in the United States, 1940–60*. Washington: U.S. Government Printing Office, 1968; and National Center for Health Statistics, *Vital Statistics of the United States*, Vol. II Mortality Part A, Technical Appendix, available at the NCHS Web site at www.cdc.gov/nchs/datawh/statab/pubd/ta.htm.

Multiple Cause of Death File

The National Center for Health Statistics (NCHS) is responsible for compiling and publishing annual national statistics on causes of death. In carrying out this responsibility, NCHS adheres to the World Health Organization Nomenclature Regulations. These Regulations require that (1) cause of death be coded in accordance with the applicable revision of the *International Classification of Diseases* (ICD) (see Appendix II, table IV and ICD); and (2) 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. National single-cause mortality statistics go back to the year 1900.

Starting with data year 1968, electronic files exist with multiple cause of death information. These files contain codes for all diagnostic terms and related codable information recorded on the death certificate. These codes comprise the entity axis, and are the input for a software program called TRANSAX. The TRANSAX program eliminates redundant entity axis codes and combines other entity axis codes to create the best set of ICD codes for a record. The output of

the TRANSAX program is the record axis. Record axis data are generally used for research and analysis of multiple or nonunderlying cause of death. Because the function of the TRANSAX program is not to select a single underlying cause of death, record axis data may or may not include the underlying cause. Tabulations of underlying and nonunderlying cause of death in table 48 (selected occupational diseases) are compiled by searching both underlying cause of death and record axis data.

For more information, see www.cdc.gov/nchs/products/elec_prods/subject/mortmcd.htm.

Linked Birth/Infant Death Data Set

National linked files of live births and infant deaths are data sets for research on infant mortality. To create these data sets, death certificates are linked with corresponding birth certificates for infants who die in the United States before their first birthday. Linked data files include all variables on the national natality file, including the more accurate racial and ethnic information, as well as variables on the national mortality file, including cause of death and age at death. The linkage makes available for the analysis of infant mortality extensive information from the birth certificate about the pregnancy, maternal risk factors, and infant characteristics and health items at birth. Each year 97–98 percent of infant death records are linked to their corresponding birth records.

National linked files of live births and infant deaths were first produced for the 1983 birth cohort. Birth cohort linked file data are available for 1983–91 and period linked file data for 1995–2000. Period linked file data starting with 1995 are not strictly comparable with birth cohort data for 1983–91. While birth cohort linked files have methodological advantages, their production incurs substantial delays in data availability, since it is necessary to wait until the close of a second data year to include all infant deaths to the birth cohort.

Starting with data year 1995, more timely linked file data are produced in a period data format preceding the release of the corresponding birth cohort format. Other changes to the data set starting with 1995 data include addition of record weights to correct for the 2.2–2.5 percent of records that could not be linked and addition of an imputation for not stated birthweight. The 1995–2000 weighted mortality rates are less than 1 percent to 4.1 percent higher than unweighted rates for the same period. The 1995–2000 weighted mortality rates with imputed birthweight are less than 1 percent to 6.3 percent higher than unweighted rates with imputed birthweight for the same period.

For more information, see: Mathews TJ, Menacker F, MacDorman MF. Infant mortality statistics from the 2000 period linked birth/infant death data set. National vital statistics reports; vol 50 no 12. Hyattsville, MD: National Center for Health Statistics. 2002; or visit the NCHS Web site at www.cdc.gov/nchs/linked.htm.

Compressed Mortality File

The Compressed Mortality File (CMF) used to compute death rates by urbanization level is a county-level national mortality and population database. The mortality database of CMF is derived from the detailed mortality files of the National Vital Statistics System starting with 1968. The population database of CMF is derived from intercensal and postcensal population estimates and census counts of the resident population of each U.S. county by age, race, and sex. Counties are categorized according to level of urbanization based on an NCHS-modified version of the 1993 rural-urban continuum codes for metropolitan and nonmetropolitan counties developed by the Economic Research Service, U.S. Department of Agriculture. See Appendix II, *Urbanization*.

For more information about the CMF, contact: D. Ingram, Office of Analysis, Epidemiology, and Health Promotion, National Center for Health Statistics, 3311 Toledo Road, Mailstop 6226, Hyattsville, MD 20782.

Nurse Supply Estimates

Health Resources and Services Administration (HRSA)

Nurse supply estimates in this report are based on a model developed by HRSA's Bureau of Health Professions to meet the requirements of Section 951, P.L. 94-63. The model estimates for each State (a) population of nurses currently licensed to practice; (b) supply of full- and part-time practicing nurses (or available to practice); and (c) full-time equivalent supply of nurses practicing full time plus one-half of those practicing part time (or available on that basis). The three estimates are divided into three levels of highest educational preparation—associate degree or diploma, baccalaureate, and master's and doctorate. Among the factors considered are new graduates, changes in educational status, nursing employment rates, age, migration patterns, death rates, and

licensure phenomena. The base data for the model are derived from the National Sample Surveys of Registered Nurses, conducted by the Division of Nursing, Bureau of Health Professions, HRSA. Other data sources include National League for Nursing for data on nursing education and National Council of State Boards of Nursing for data on licensure. For further information, visit HRSA's Division of Nursing Web site at www.bhpr.hrsa.gov/nursing/.

Online Survey Certification and Reporting Database (OSCAR)

Centers for Medicare & Medicaid Services

The Online Survey Certification and Reporting (OSCAR) database has been maintained by the Centers for Medicare & Medicaid Services (CMS), formerly the Health Care Financing Administration (HCFA), since 1992. OSCAR is an updated version of the Medicare and Medicaid Automated Certification System that has been in existence since 1972. OSCAR is an administrative database containing detailed information on all Medicare and Medicaid health care providers in addition to all currently certified Medicare and Medicaid nursing home facilities in the United States and Territories. (Data for the Territories are not shown in this report.) The purpose of the nursing home facility survey certification process is to ensure that nursing facilities meet the current CMS long-term care requirements and thus can participate in serving Medicare and Medicaid beneficiaries. Included in the OSCAR database are all certified nursing facilities, certified hospital-based nursing homes, and certified units for other types of nursing home facilities (for example, life-care communities or board and care homes). Facilities not included in OSCAR are all noncertified facilities (that is, facilities that are only licensed by the State and are limited to private payment sources) and nursing homes that are part of the Department of Veterans Affairs. Also excluded are nursing homes that are intermediate care facilities for the mentally retarded.

Information on the number of beds, residents, and resident characteristics is collected during an inspection of all certified facilities. The information in OSCAR is based on each facility's own administrative record system in addition to interviews with key administrative staff members.

All certified nursing homes are inspected by representatives of the State survey agency (generally the department of health) at least once every 15 months. Therefore a complete census must be based on a 15-month reporting cycle rather than a 12-month cycle. Some nursing homes are inspected twice or more often during any given reporting cycle. In order to avoid overcounting, the data must be edited and duplicates removed. Data editing and compilation were performed by Cowles Research Group and published in the group's *Nursing Home Statistical Yearbook* series

For more information, see: Cowles CM, 1995; 1996; 1997 Nursing Home Statistical Yearbook. Anacortes, WA: Cowles Research Group (CRG), 1995; 1997; 1998; Cowles CM, 1998; 1999; 2000; 2001 Nursing Home Statistical Yearbook. Washington, DC: American Association of Homes and Services for the Aging (AAHSA), 1999; 2000; 2001; 2002; HCFA: OSCAR Data Users Reference Guide, 1995, available from CMS, Health Standards and Quality Bureau, HCFA/HSQB S2 11-07, 7500 Security Boulevard, Baltimore, MD 21244; and visit the CMS Web site at www.cms.gov or the CRG Web site at www.longtermcareinfo.com/crg or the AAHSA Web site at www.aahsa.org.

Population Census and Population Estimates

Bureau of the Census

Decennial Census

The census of population (decennial census) has been held in the United States every 10 years since 1790. The decennial census has enumerated the resident population as of April 1 of the census year ever since 1930. Data on sex, race, age, and marital status are collected from 100 percent of the enumerated population. More detailed information such as income, education, housing, occupation, and industry are collected from a representative sample of the 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 Statistical Policy Directive 15, Race and Ethnicity Standards for Federal Statistics and Administrative Reporting." This document specified rules for the collection, tabulation, and reporting of race and ethnicity 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 of the Standards for the Classification of Federal Data on Race and Ethnicity" (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 from four to five the 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. 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, Hispanics may be of any race.

Modified Decennial Census Files

For several decades the Census Bureau has produced modified decennial census files. These modified files incorporate adjustments to the 100 percent April 1 count data for (1) errors in the census data discovered subsequent to publication, (2) misreported age data, and (3) nonspecified race.

For the 1990 census, the 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 on 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 persons did not specify their race (over 95 percent 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 census, the Census Bureau modified the race data on the census and produced the Modified Race Data Summary File. For this file, persons who reported "Some other race" as part of their race response were assigned to one of the 31 race groups, which are the single- and multiple-race combinations of the five race categories specified in the 1997 race and ethnicity standards. Persons who did not specify their race were assigned to one of the 31 race groups using imputation. Of the 18.5 million persons who reported "Some other race" as part of their race response or who did not specify their race, 16.8 million (90.4 percent) were of Hispanic origin.

Bridged-Race Population Estimates for Census 2000

Race data on the 2000 census are not comparable with race data on other data systems that are continuing to collect data using the 1977 standards on race and ethnicity during the transition to full implementation of the 1997 standards. For example, most of the States in the Vital Statistics Cooperative Program will revise their birth and death certificates to conform to the 1997 standards after 2000. Thus, population estimates for 2000 and beyond with race categories comparable to the 1977 categories are needed so that race-specific birth and death rates can be calculated. To meet this need, NCHS, in collaboration with the U.S. Census Bureau, developed methodology to bridge the 31 race groups in census 2000 to the four single-race categories specified under the 1977 standards.

The bridging methodology was developed using information from the 1997–2000 National Health Interview Survey (NHIS) (Ingram DD, Weed JA, Parker JD, et al. U.S. census 2000 population with bridged race categories. Vital Health Stat 2. Forthcoming, 2003.) The NHIS is an annual survey sponsored by NCHS and conducted by the Census Bureau (see *National Health Interview Survey*). The NHIS provides a unique opportunity to investigate multiple-race groups because since 1982, the NHIS 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. Bridging proportions derived from these models were applied by the U.S. Census Bureau to the Census 2000 Modified Race Data Summary File. This application resulted in bridged counts of the April 1, 2000 resident single-race populations for four racial groups, American Indian or Alaska Native, Asian or Pacific Islander, black, and white.

For more information about bridged-race population estimates, see www.cdc.gov/nchs/about/major/dvs/popbridge/popbridge.htm.

Postcensal Population Estimates

Postcensal population estimates are estimates made for the years following a census, before the next census has been taken. National postcensal population estimates are derived by updating the resident population enumerated in the decennial census using a components of population change approach. The following formula is used to update the decennial census counts:

(1) decennial census enumerated resident population

(2) + births to U.S. resident women,

(3) – deaths to U.S. residents,

(4) + net international migration,

(5) + net movement of U.S. Armed Forces and civilian citizens of the U.S.

State postcensal estimates are based on similar data and a variety of other data series, including school statistics from State departments of education and parochial school systems. The postcensal estimates are consistent with official decennial census figures and do not reflect estimated decennial census underenumeration.

The Census Bureau has produced a postcensal series of estimates of the July 1 resident population of the United States based on census 2000 by applying the components of change methodology to the Modified Race Data Summary File. These postcensal estimates have race data for 31 race groups, in accordance with the 1997 race and ethnicity standards. So that the race data for the 2000-based postcensal estimates would be comparable with race data on vital records, the Census Bureau applied the NHIS bridging methodology to the 31-race group postcensal population estimates to obtain postcensal estimates for the four single-race categories (American Indian or Alaska Native, Asian or Pacific Islander, black, and white). Bridged-race postcensal population estimates are available at www.cdc.gov/nchs/about/major/dvs/popbridge/popbridge.htm.

Note that before the bridged-race April 1, 2000 population counts and the bridged-race 2000-based postcensal estimates were available, the Census Bureau extended their postcensal series of estimates based on the 1990 census (with the four single-race categories needed to compute vital rates) to July 1, 2001. NCHS initially calculated vital rates for 2000 using 1990-based July 1, 2000 postcensal population estimates and vital rates for 2001 using 1990-based July 1, 2001 postcensal estimates. Vital rates for 2000 have been revised using the bridged-race April 1, 2000 population counts and vital rates for 2001 have been revised using the 2000-based bridged-race July 1, 2001 postcensal population estimates.

Intercensal Population Estimates

The further from the census year on which the postcensal estimates are based, the less accurate are the postcensal estimates. With the completion of the decennial census at the end of the decade, intercensal estimates for the preceding decade were prepared to replace the less accurate postcensal estimates. Intercensal population estimates take into account the census of population at the beginning and end of the decade. Thus intercensal estimates are more accurate than postcensal estimates as they correct for the "error of closure" or difference between the estimated population at the end of the decade and the census count for that date. The "error of closure" at the national level was quite small for the 1960s (379,000). However, for the 1970s it amounted to almost 5 million, for the 1980s, 1.5 million, and for the 1990s, about 6 million. The error of closure differentially affects age, race, sex, and Hispanic origin subgroup populations as well as the rates based on these populations. Vital rates that were calculated using postcensal population estimates are routinely revised when intercensal estimates become available because the intercensal estimates correct for the error of closure.

Intercensal estimates for the 1990s with race data comparable to the 1977 standards have been derived so that vital rates for the 1990s could be revised to reflect census 2000. Calculation of the intercensal population estimates for the 1990s was complicated by the incomparability of the race data on the 1990 and 2000 censuses. The Census Bureau, in collaboration with National Cancer Institute and NCHS, derived race-specific intercensal population estimates for the 1990s using the 1990 MARS file as the beginning population base and the bridged-race population estimates for April 1, 2000 as the ending population base. Bridged-race intercensal population estimates are available at www.cdc.gov/nchs/about/major/dvs/popbridge/popbridge.htm.

Revised bridged-race population estimates for women aged 15–17 and 18–19 years based on the 2000 census for computing teenage birth rates in table 3 were not available from the U.S. Census Bureau when this report was prepared. The 1991–99 population estimates for these teenage subgroups were prepared by the Division of Vital Statistics, National Center for Health Statistics (NCHS). The NCHS population estimates were prepared by applying proportions derived from the 1990-based population estimates (according to data year, race, and Hispanic origin for the teenage population) to the 2000-based population of women aged 15–19 years within each race/Hispanic origin group, and adjusting the sum of the population estimates to be consistent with the total population of women aged 15–19 years for each race/Hispanic origin group (2000 based). Rates based on these population estimates are intended as interim measures and caution should be used in interpreting the rates and trends. When the necessary intercensal population estimates based on the 2000 census become available from the U.S. Bureau of the Census, the rates for women aged 15–17 and 18–19 years in table 3 will be revised on the Web site at www.cdc.gov/nchs/hus.htm.

Special Population Estimates

Special population estimates are prepared for the education reporting area for mortality statistics because educational attainment of decedent is not reported by all 50 States. The Housing and Household Economics Statistics Division of the U.S. Bureau of the Census currently produces unpublished estimates of populations by age, race, sex, and educational attainment for NCHS. These population estimates are based on the Current Population Survey, adjusted to resident population controls. The control totals used for July 1, 1994–96 are 1990-based population estimates for 45 reporting States and the District of Columbia (DC); for July 1, 1997–2000, 1990-based population estimates for 46 reporting States and DC; and for July 1, 2001, 2000-based population estimates for 47 reporting States and DC. See Appendix II, *Education*.

For more information about the population census and population estimates, visit the U.S. Bureau of the Census Web site at www.census.gov/.

Sexually Transmitted Disease (STD) Surveillance

Centers for Disease Control and Prevention

National Center of HIV, STD, and TB Prevention

The Division of STD Prevention (DSTD) of the National Center of HIV, STD, and TB Prevention (NCHSTP), Centers for Disease Control and Prevention (CDC) compiles sexually transmitted disease (STD) surveillance information from the following sources of data: (1) case reports from STD project areas; (2) prevalence data from the Regional Infertility Prevention Program, the National Job Training Program (formerly the Job Corps), the Jail STD Prevalence Monitoring Projects, the adolescent Women Reproductive Health Monitoring Project, the Men Who Have Sex With Men (MSM) Prevalence Monitoring Project, and the Indian Health Service; (3) sentinel surveillance of gonococcal antimicrobial resistance from the Gonococcal Isolate Surveillance Project (GISP); and (4) national sample surveys implemented by federal and private organizations.

Case reports of STDs are reported to CDC by STD surveillance systems operated by State and local STD control programs and health departments in 50 States, the District of Columbia, selected cities, 3,139 U.S. counties, and outlying areas comprised of U.S. dependencies and possessions, and independent nations in free association with the United States. Case report data is the source of statistical data in table 52. Because of incomplete diagnosis and reporting, the number of STD cases reported to CDC is less than the actual number of cases occurring among the United States population. Data from outlying areas are not included in table 52.

STD data are submitted to CDC on a variety of hardcopy summary reporting forms (monthly, quarterly, and annually) and electronic summary or individual case-specific (line-listed) formats via the National Electronic Telecommunications System for Surveillance (NETSS). Reports and corrections sent to CDC on hardcopy forms and for NETSS electronic data through May 3, 2002, are included in table 52.

Crude incidence rates (new cases/population) were calculated on an annual basis per 100,000 civilian population. The 2001 rates for the United States were calculated by dividing the number of cases reported in 2001 by the post-1990 estimated 2000 population.

For more information, see: Centers for Disease Control and Prevention. *Sexually Transmitted Disease Surveillance, 2001*. Atlanta, GA: U.S. Department of Health and Human Services, September 2002; or visit the STD Prevention Web site at: www.cdc.gov/std/stats/.

Surveillance, Epidemiology, and End Results Program (SEER)

National Cancer Institute

In the Surveillance, Epidemiology, and End Results (SEER) Program, the National Cancer Institute (NCI) contracts with population-based registries throughout the United States to provide data on all residents diagnosed with cancer during the year and to provide current followup information on all previously diagnosed patients.

Analysis of cancer survival rates in this report covers residents at the time of the initial diagnosis of cancer in the following SEER 9 registries: Atlanta, Georgia; Connecticut; Detroit, Michigan; Hawaii; Iowa; New Mexico; San Francisco-Oakland; Seattle-Puget Sound; and Utah. Analysis of cancer incidence covers residents in the following SEER 12 registries: the SEER 9 registries plus Los Angeles and San Jose-Monterey, California and the Alaska Native Tumor Registry.

Population estimates (1990-based postcensal estimates) used to calculate incidence rates are obtained from the U.S. Bureau of the Census. NCI uses estimation procedures as needed to obtain estimates for years and races not included in data provided by the U.S. Bureau of the Census. Rates presented in this report may differ somewhat from previous reports due to revised population estimates and the addition and deletion of small numbers of incidence cases.

Life tables used to determine normal life expectancy when calculating relative survival rates were obtained from NCHS and in-house calculations. Separate life tables are used for each race-sex-specific group included in the SEER Program.

For further information, see: Ries LAG, Eisner MP, Kosary CL, et al. (eds). *SEER Cancer Statistics Review 1973–99*. National Cancer Institute. Bethesda, MD. 2002; or visit the SEER Web site at www.seer.cancer.gov.

Survey of Occupational Injuries and Illnesses (SOII)

Bureau of Labor Statistics

Since 1971 the Bureau of Labor Statistics (BLS) has conducted an annual survey of establishments in the private sector to collect statistics on occupational injuries and illnesses. The Survey of Occupational Injuries and Illnesses is a Federal/State program in which employer reports are collected from about 169,000 private industry establishments and processed by State agencies cooperating with BLS. Data for the mining industry and for railroad activities are provided by Department of Labor's Mine Safety and Health Administration and Department of Transportation's Federal Railroad Administration. Excluded from the survey are self-employed individuals; farmers with fewer than 11 employees; private households; Federal Government agencies; and employees in State and local government agencies. Establishments are classified in industry categories based on the 1987 Standard Industrial Classification (SIC) Manual, as defined by the Office of Management and Budget.

Survey estimates of occupational injuries and illnesses are based on a scientifically selected probability sample, rather than a census of the entire population. An independent sample is selected for each State and the District of Columbia that represents industries in that jurisdiction. BLS includes all the State samples in the national sample.

Establishments included in the survey are instructed in a mailed questionnaire to provide summary totals of all entries for the previous calendar year to its Log and Summary of Occupational Injuries and Illnesses (OSHA No. 200 form). Additionally, from the selected establishments, approximately 550,000 injuries and illnesses with days away from work are sampled to obtain demographic and detailed case characteristic information. An occupational injury is any injury such as a cut, fracture, sprain, or amputation that results from a work-related event or from a single instantaneous exposure in the work environment. An occupational illness is any abnormal condition or disorder other than one resulting from an occupational injury, caused by exposure to factors associated with employment. It includes acute and chronic illnesses or diseases that may be caused by inhalation, absorption, ingestion, or direct contact. Lost workday cases involve days away from work, days of restricted work activity, or both. The response rate is about 92 percent.

The number of injuries and illnesses reported in any given year can be influenced by the level of economic activity, working conditions and work practices, worker experience and training, and the number of hours worked. Long-term latent illnesses caused by exposure to carcinogens are believed to be understated in the survey's illness measures. In contrast, new illnesses such as contact dermatitis and carpal tunnel syndrome are easier to relate directly to workplace activity.

For more information, see: Bureau of Labor Statistics, Workplace Injuries and Illnesses in 2001, Washington, DC. U.S. Department of Labor, December 2002; or visit the BLS occupational safety and health Web site at www.bls.gov/iif/home.htm.

Youth Risk Behavior Survey (YRBS)

Centers for Disease Control and Prevention

National Center for Chronic Disease Prevention and Health Promotion

The national Youth Risk Behavior Survey (YRBS) is conducted by the Centers for Disease Control and Prevention's National Center for Chronic Disease Prevention and Health Promotion to monitor the prevalence of priority health risk behaviors among high school students in grades 5–12 that contribute to morbidity and mortality in both adolescence and adulthood.

The national YRBS of high school students was conducted in 1990, 1991, 1993, 1995, 1997, 1999, and 2001. The national YRBS school-based surveys employ a three-stage cluster sample design to produce a nationally representative sample of students in grades 5–12 attending public and private high schools. The first-stage sampling frame contains primary sampling units (PSUs) consisting of large counties or groups of smaller, adjacent counties. The PSUs are then stratified based on degree of urbanization and relative percent of black and Hispanic students in the PSU. The PSUs are selected from these strata with probability proportional to school enrollment size. At the second sampling stage, schools are selected with probability proportional to school enrollment size. To enable separate analysis of data for black and Hispanic students, schools with substantial numbers of black and Hispanic students are sampled at higher rates than all other schools. The third stage of sampling consists of randomly selecting one or two intact classes of a required subject from grades 5–12 at each chosen school. All students in the selected classes are eligible to participate in the survey. A weighting factor is applied to each student record to adjust for nonresponse and for the varying probabilities of selection, including those resulting from the oversampling of black and Hispanic students. The sample size for the 2001 YRBS was 13,601. The school response rate was 75 percent, and the student response rate was 83 percent, for an overall response rate of 63 percent.

National YRBS data are subject to at least two limitations. First, these data apply only to adolescents who attend regular high school. These students may not be representative of all persons in this age group because those who have dropped out of high school or attend an alternative high school are not surveyed. Second, the extent of underreporting or overreporting cannot be determined, although the survey questions demonstrate good test-retest reliability.

Estimates of substance use for youth based on the YRBS differ from the National Household Survey on Drug Abuse (NHSDA) and Monitoring the Future (MTF). Rates are not directly comparable across these surveys due to differences in populations covered, sample design, questionnaires, interview setting, and statistical approaches to make the survey estimates generalizable to the entire population. The NHSDA survey collects data in homes, whereas the MTF and YRBS collect data in school classrooms. The NHSDA estimates are tabulated by age, while the MTF and YRBS estimates are tabulated by grade, representing different ages as well as different populations. See Cowan CD. Coverage, Sample Design, and Weighting in Three Federal Surveys. Journal of Drug Issues 31(3), 595–614, 2001.

For further information on the YRBS, see: CDC. Youth risk behavior surveillance—United States, 1999. CDC surveillance summaries, June 9, 2000. MMWR 2000:49(SS-05); CDC. Youth risk behavior surveillance—United States, 2001. CDC surveillance summaries, June 21, 2002. MMWR 2002:51(SS-04); or write: Director, Division of Adolescent and School Health, National Center for Chronic Disease Prevention and Health Promotion, Centers for Disease Control and Prevention, 4770 Buford Highway NE, Mail Stop K-32, Atlanta, GA 30341-3717; or visit the Division of Adolescent and School Health Web site at www.cdc.gov/nccdphp/dash/.

Private and Global Sources

Alan Guttmacher Institute Abortion Survey

The Alan Guttmacher Institute (AGI) conducts periodic surveys of abortion providers. Data are collected from clinics, physicians, and hospitals identified as potential providers of abortion services. For 1999 and 2000, 2,442 facilities were surveyed. In addition, State health statistics agencies were contacted, requesting all available data reported by providers to each State health agency on the number of abortions performed in 1999 and 2000. For States that provide data to AGI, the health agency figures were used for providers who did not respond to the survey. Of the 2,442 potential providers, 1,931 performed abortions between January 1999 and June 2001. Of the abortions reported for 2000, 77 percent were reported by the providers, 10 percent came from health department data, 11 percent were estimated by knowledgeable sources, and 2 percent were projections or other estimates.

The number of abortions estimated by AGI through the mid- to late-1980s was about 20 percent higher than the number reported to the Centers for Disease Control and Prevention (CDC). Between 1989 and 1997 the AGI estimates were about 12 percent higher than those reported by CDC. Beginning in 1998, health departments of four States did not report abortion data to CDC. The four reporting areas (the largest of which was California) that did not report abortions to CDC in 1998 accounted for 18 percent of all abortions tallied by AGI's 1995–96 survey.

For more information, write: The Alan Guttmacher Institute, 120 Wall Street, New York, NY 10005; or visit AGI's Web site at www.agi-usa.org.

American Association of Colleges of Osteopathic Medicine

The American Association of Colleges of Osteopathic Medicine (AACOM) compiles data on various aspects of osteopathic medical education for distribution to the profession, the government, and the public. Questionnaires are sent annually to schools of osteopathic medicine requesting information on characteristics of applicants and students, curricula, faculty, grants, contracts, revenues, and expenditures. The response rate is 100 percent.

For more information, see: *2001 Annual Report on Osteopathic Medical Education*, American Association of Colleges of Osteopathic Medicine: 5550 Friendship Blvd, Suite 310, Chevy Chase, Maryland 20815; or visit the AACOM Web site at www.aacom.org.

American Association of Colleges of Pharmacy

The American Association of Colleges of Pharmacy (AACP) compiles data on the Colleges of Pharmacy, including information on student enrollment and types of degrees conferred. Data are collected through an annual survey; the response rate is 100 percent.

For further information, see: *Profile of Pharmacy Students*. The American Association of Colleges of Pharmacy, 1426 Prince Street, Alexandria, VA; or visit the AACP Web site at www.aacp.org.

American Association of Colleges of Podiatric Medicine

The American Association of Colleges of Podiatric Medicine (AACPM) compiles data on the Colleges of Podiatric Medicine, including information on the schools and enrollment. Data are collected annually through written questionnaires. The response rate is 100 percent.

For further information, write: The American Association of Colleges of Podiatric Medicine, 1350 Piccard Drive, Suite 322, Rockville, MD 20850-4307; or visit the AACPM Web site at www.aacpm.org.

American Dental Association

The Division of Educational Measurement of the American Dental Association (ADA) conducts annual surveys of predoctoral dental educational institutions. The questionnaire, mailed to all dental schools, collects information on student characteristics, financial management, and curricula.

For more information, see: American Dental Association, *1999–2000 Survey of Predoctoral Dental Educational Institutions*. Chicago, IL. 2001; or visit the ADA Web site at www.ada.org.

American Hospital Association Annual Survey of Hospitals

Data from the American Hospital Association (AHA) annual survey are based on questionnaires sent to all AHA-registered and nonregistered hospitals in the United States and its associated areas. 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 percent. 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, and facilities. 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.

For more information on the AHA Annual Survey of Hospitals, see: Health Forum, LLC, an American Hospital Association Company. Hospital Statistics, 2002. Chicago, IL. 2002; or visit the AHA Web site at www.aha.org.

American Medical Association Physician Masterfile

A masterfile of physicians has been maintained by the American Medical Association (AMA) since 1906. The Physician Masterfile contains data on almost every physician in the United States, members and nonmembers of the AMA, and on those graduates of American medical schools temporarily practicing overseas. The file also includes graduates of international medical schools who are in the United States and meet education standards for primary recognition as physicians.

A file is initiated on each individual upon entry into medical school or, in the case of international graduates, upon entry into the United States. Between 1965–85 a mail questionnaire survey was conducted every 4 years to update the file information on professional activities, self-designated area of specialization, and present employment status. Since 1985 approximately one-third of all physicians are surveyed each year.

For more information on the AMA Physician Masterfile, see: Division of Survey and Data Resources, American Medical Association, *Physician Characteristics and Distribution in the U.S., 2002–2003* ed. Chicago, IL. 2002; or visit the AMA Web site at www.ama-assn.org.

Association of American Medical Colleges

The Association of American Medical Colleges (AAMC) collects information on student enrollment in medical schools through the annual Liaison Committee on Medical Education questionnaire, the fall enrollment questionnaire, and the American Medical College Application Service (AMCAS) data system. Other data sources are the institutional profile system, the premedical students questionnaire, the minority student opportunities in medicine questionnaire, the faculty roster system, data from the Medical College Admission Test, and one-time surveys developed for special projects.

For more information, see: Association of American Medical Colleges, *Statistical Information Related to Medical Education*, Washington, DC. 2001; or visit the AAMC Web site at www.aamc.org.

Association of Schools and Colleges of Optometry

The Association of Schools and Colleges of Optometry (ASCO) compiles data on various aspects of optometric education including data on schools and enrollment. Questionnaires are sent annually to all schools and colleges of optometry. The response rate is 100 percent.

For further information, write: Annual Survey of Optometric Educational Institutions, Association of Schools and Colleges of Optometry, 6110 Executive Blvd., Suite 510, Rockville, MD 20852; or visit the ASCO Web site at www.opted.org.

Association of Schools of Public Health

The Association of Schools of Public Health (ASPH) compiles data on schools of public health in the United States and Puerto Rico. Questionnaires are sent annually to all member schools. The response rate is 100 percent.

Unlike health professional schools that emphasize specific clinical occupations, schools of public health offer study in specialty areas such as biostatistics, epidemiology, environmental health, occupational health, health administration, health planning, nutrition, maternal and child health, social and behavioral sciences, and other population-based sciences.

For further information, write: Association of Schools of Public Health, 1101 15th Street, NW, Suite 910, Washington, DC 20005; or visit the ASPH Web site at www.asph.org.

European Health for All Database

World Health Organization Regional Office for Europe

The WHO Regional Office for Europe (WHO/Europe) provides country-specific and topic-specific health information via the Internet for people who influence health policy in the WHO European Region and the media.

WHO/Europe collects statistics on health and makes them widely available through:

- The European health for all database (HFA-DB) that contains data on about 600 health indicators collected from national counterparts in 51 European countries, and data from other WHO technical programs and some international organizations.
- Highlights on health in countries in the WHO European Region that give an overview of the health situation in each country in comparison with other countries. Highlights complement the public health reports produced by a number of member States in the region.
- Health status overview for countries of central and eastern Europe that are candidates for accession to the European Union (Bulgaria, the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia and Slovenia).

WHO/Europe helps countries strengthen their national health information systems, particularly by supporting:

- the development of national health indicator databases
- the exchange of experience on national public health reports between countries; a database of public health reports is maintained and available for consultation and networking
- implementation of international classifications and definitions in countries
- regional networks of health information professionals

For more information, visit the European health for all database at http://hfadb.who.dk/hfa/.

InterStudy National Health Maintenance Organization Census

From 1976 to 1980 the Office of Health Maintenance Organizations conducted a census of health maintenance organizations (HMOs). Since 1981 InterStudy has conducted the census. A questionnaire is sent to all HMOs in the United States asking for updated enrollment, profit status, and Federal qualification status. New HMOs are also asked to provide information on model type. When necessary, information is obtained, supplemented, or clarified by telephone. For nonresponding HMOs State-supplied information or the most current available data are used.

In 1985 a large increase in the number of HMOs and enrollment was partly attributable to a change in the categories of HMOs included in the census: Medicaid-only and Medicare-only HMOs have been added. Also component HMOs, which have their own discrete management, can be listed separately, whereas, previously the oldest HMO reported for all of its component or expansion sites, even when the components had different operational dates or were different model types.

For further information, see: *The InterStudy Competitive Edge*. InterStudy Publications, St. Paul, MN. 2002; or visit the InterStudy Web site at www.hmodata.com.

National League for Nursing

The division of research of the National League for Nursing (NLN) conducts The Annual Survey of Schools of Nursing in October of each year. Questionnaires are sent to all graduate nursing programs (master's and doctoral), baccalaureate programs designed exclusively for registered nurses, basic registered nursing programs (baccalaureate, associate degree, and diploma), and licensed practical nursing programs. Data on enrollments, first-time admissions, and graduates are completed for all nursing education programs. Response rates of approximately 80 percent are achieved for other areas of inquiry.

For more information, see: National League for Nursing, *Nursing Data Review* 1997, New York, NY. 1997; or visit the NLN Web site at www.nln.org.

Organization for Economic Cooperation and Development Health Data

The Organization for Economic Cooperation and Development (OECD) provides annual data on statistical indicators on health and economic policies collected from 30 member countries since the 1960s. The international comparability of health expenditure estimates depends on the quality of national health accounts in OECD member countries. In recent years the OECD health accounts have become an informal standard for reporting on health care systems. Additional limitations in international comparisons include differing boundaries between health care and other social care particularly for the disabled and elderly, and underestimation of private expenditures on health.

The OECD was established in 1961 with a mandate to promote policies to achieve the highest sustainable economic growth and a rising standard of living among member countries. The Organization now comprises 30 member countries: Australia, Austria, Belgium, Canada, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Spain, Sweden, Switzerland, Turkey, the United Kingdom, and the United States.

As part of its mission, the OECD has developed a number of activities in relation to health and health care systems. The main aim of OECD work on health policy is to conduct cross-national studies of the performance of OECD health systems and to facilitate exchanges between member countries of their experiences of financing, delivering, and managing health services. To support this work, each year the OECD compiles cross-country data in OECD Health Data, one of the most comprehensive sources of comparable health-related statistics. OECD Health Data is an essential tool to carry out comparative analyses and draw lessons from international comparisons of diverse health care systems. This international database now incorporates the first results arising from the implementation of the OECD manual, A System of Health Accounts (2000), which provide a standard framework for producing a set of comprehensive, consistent, and internationally comparable data on health spending. The OECD collaborates with other international organizations such as the WHO.

For further information, see www.oecd.org/health.

United Nations Demographic Yearbook

The Statistical Office of the United Nations prepares the *Demographic Yearbook*, a comprehensive collection of international demographic statistics.

Questionnaires are sent annually and monthly to more than 220 national statistical services and other appropriate government offices. Data forwarded on these questionnaires are supplemented, to the extent possible, by data taken from official national publications and by correspondence with the national statistical services. To ensure comparability, rates, ratios, and percents have been calculated in the statistical office of the United Nations.

Lack of international comparability among estimates arises from differences in concepts, definitions, and time of data collection. The comparability of population data is affected by several factors, including (a) definitions of the total population, (b) definitions used to classify the population into its urban and rural components, (c) difficulties relating to age reporting, (d) extent of over- or underenumeration, and (e) quality of population estimates. The completeness and accuracy of vital statistics data also vary from one country to another. Differences in statistical definitions of vital events may also influence comparability.

International demographic trend data are available on a CD-ROM entitled United Nations, 2000. Demographic Yearbook—Historical Supplement 1948–97. CD-ROM Special Issue. United Nations publication sales number E/F.99.XIII.12.

For more information, see: United Nations, *Demographic Yearbook 2000*, United Nations, New York, 2002; or visit the United Nations Web site at www.un.org or their Web site locator at www.unsystem.org.

World Health Statistics Annual

World Health Organization

The World Health Organization (WHO) prepares the *World Health Statistics Annual*, an annual volume of information on vital statistics and causes of death designed for use by the medical and public health professions. Each volume is the

result of a joint effort by the national health and statistical administrations of many countries, the United Nations, and WHO. United Nations estimates of vital rates and population size and composition, where available, are reprinted directly in the *Statistics Annual*. For those countries for which the United Nations does not prepare demographic estimates, primarily smaller populations, the latest available data reported to the United Nations and based on reasonably complete coverage of events are used.

Information published on infant mortality is based entirely on official national data either reported directly or made available to WHO.

Selected life table functions are calculated from the application of a uniform methodology to national mortality data provided to WHO, in order to enhance their value for international comparisons. The life table procedure used by WHO may often lead to discrepancies with national figures published by countries, due to differences in methodology or degree of age detail maintained in calculations.

The international comparability of estimates published in the *World Health Statistics Annual* is affected by the same problems as is the United Nations *Demographic Yearbook*. Cross-national differences in statistical definitions of vital events, in the completeness and accuracy of vital statistics data, and in the comparability of population data are the primary factors affecting comparability.

For more information, see: World Health Organization, *World Health Statistics Annual 2000*, World Health Organization, Geneva, 2002; World Health Statistics 1997–99 at www.who.int/whosis; or visit the WHO Web site at www.who.int.