Limitations of the Data

Introduction The data presented in this *Statistical Abstract* came from many sources. The sources include not only Federal statistical bureaus and other organizations that collect and issue statistics as their principal activity, but also governmental administrative and regulatory agencies, private research bodies, trade associations, insurance companies, health associations, and private organizations such as the National Education Association and philanthropic foundations. Consequently, the data vary considerably as to reference periods, definitions of terms and, for ongoing series, the number and frequency of time periods for which data are available.

The statistics presented were obtained and tabulated by various means. Some statistics are based on complete enumerations or censuses while others are based on samples. Some information is extracted from records kept for administrative or regulatory purposes (school enrollment, hospital records, securities registration, financial accounts, social security records, income tax returns, etc.), while other information is obtained explicitly for statistical purposes through interviews or by mail. The estimation procedures used vary from highly sophisticated scientific techniques, to crude "informed guesses."

Each set of data relates to a group of individuals or units of interest referred to as the *target universe* or *target population*, or simply as the *universe* or *population*. Prior to data collection the target universe should be clearly defined. For example, if data are to be collected for the universe of households in the United States, it is necessary to define a "household." The target universe may not be completely tractable. Cost and other considerations may restrict data collection to a *survey universe* based on some available list, such list may be inaccurate or out of date. This list is called a *survey frame* or *sampling frame*.

The data in many tables are based on data obtained for all population units, *a census,* or on data obtained for only a portion, or *sample,* of the population units. When the data presented are based on a sample, the sample is usually a scientifically selected *probability sample.* This is a sample selected from a list or sampling frame in such a way that every possible sample has a known chance of selection and usually each unit selected can be assigned a number, greater than zero and less than or equal to one, representing its likelihood or probability of selection.

For large-scale sample surveys, the probability sample of units is often selected as a multistage sample. The first stage of a multistage sample is the selection of a probability sample of large groups of population members, referred to as primary sampling units (PSUs). For example, in a national multistage household sample, PSUs are often counties or groups of counties. The second stage of a multistage sample is the selection, within each PSU selected at the first stage, of smaller groups of population units, referred to as secondary sampling units. In subsequent stages of selection, smaller and smaller nested groups are chosen until the ultimate sample of population units is obtained. To qualify a multistage sample as a probability sample, all stages of sampling must be carried out using probability sampling methods.

Prior to selection at each stage of a multistage (or a single-stage) sample, a list of the sampling units or sampling frame for that stage must be obtained. For example, for the first stage of selection of a national household sample, a list of the counties and county groups that form the PSUs must be obtained. For the final stage of selection, lists of households, and sometimes persons within the households, have to be compiled in the field. For surveys of economic entities and for the economic censuses the Census Bureau

generally uses a frame constructed from the Census Bureau's Business Register. The Business Register contains all establishments with payroll in the United States including small single establishment firms as well as large multiestablishment firms.

Wherever the quantities in a table refer to an entire universe, but are constructed from data collected in a sample survey, the table quantities are referred to as *sample estimates.* In constructing a sample estimate, an attempt is made to come as close as is feasible to the corresponding universe quantity that would be obtained from a complete census of the universe. Estimates based on a sample will, however, generally differ from the hypothetical census figures. Two classifications of errors are associated with estimates based on sample surveys: (1) *sampling error*—the error arising from the use of a sample, rather than a census, to estimate population quantities and (2) *nonsampling error*—those errors arising from nonsampling sources. As discussed below, the magnitude of the sampling error for an estimate can usually be estimated from the sample data. However, the magnitude of the nonsampling error for an estimate can rarely be estimated. Consequently, actual error in an estimate exceeds the error that can be estimated.

The particular sample used in a survey is only one of a large number of possible samples of the same size which could have been selected using the same sampling procedure. Estimates derived from the different samples would, in general, differ from each other. The *standard error* (SE) is a measure of the variation among the estimates derived from all possible samples. The standard error is the most commonly used measure of the sampling error of an estimate. Valid estimates of the standard errors of survey estimates can usually be calculated from the data collected in a probability sample. For convenience, the standard error is sometimes expressed as a percent of the estimate and is called the relative standard error or *coefficient of variation* (CV). For example, an estimate of 200 units with an estimated standard error of 10 units has an estimated CV of 5 percent.

A sample estimate and an estimate of its standard error or CV can be used to construct interval estimates that have a prescribed confidence that the interval includes the average of the estimates derived from all possible samples with a known probability. To illustrate, if all possible samples were selected under essentially the same general conditions, and using the same sample design, and if an estimate and its estimated standard error were calculated from each sample, then: 1) approximately 68 percent of the intervals from one standard error below the estimate to one standard error above the estimate would include the average estimate derived from all possible samples; 2) approximately 90 percent of the intervals from 1.6 standard errors below the estimate to 1.6 standard errors above the estimate would include the average estimate derived from all possible samples; and 3) approximately 95 percent of the intervals from two standard errors below the estimate to two standard errors above the estimate would include the average estimate derived from all possible samples.

Thus, for a particular sample, one can say with the appropriate level of confidence (e.g., 90 percent or 95 percent) that the average of all possible samples is included in the constructed interval. Example of a confidence interval: An estimate is 200 units with a standard error of 10 units. An approximately 90 percent confidence interval (plus or minus 1.6 standard errors) is from 184 to 216.

All surveys and censuses are subject to nonsampling errors. Nonsampling errors are of two kinds—*random* and *nonrandom.* Random nonsampling errors arise because of the varying interpretation of questions (by respondents or interviewers) and varying actions of coders, keyers, and other processors. Some randomness is also introduced when respondents must estimate. Nonrandom nonsampling errors result from total nonresponse (no usable data obtained for a sampled unit), partial or item nonresponse (only a portion of a response may be usable), inability or unwillingness on the part of respondents to provide correct information, difficulty interpreting questions, mistakes

in recording or keying data, errors of collection or processing, and coverage problems (overcoverage and undercoverage of the target universe). Random nonresponse errors usually, but not always, result in an understatement of sampling errors and thus an overstatement of the precision of survey estimates. Estimating the magnitude of nonsampling errors would require special experiments or access to independent data and, consequently, the magnitudes are seldom available.

Nearly all types of nonsampling errors that affect surveys also occur in complete censuses. Since surveys can be conducted on a smaller scale than censuses, non-sampling errors can presumably be controlled more tightly. Relatively more funds and effort can perhaps be expended toward eliciting responses, detecting and correcting response error, and reducing processing errors. As a result, survey results can sometimes be more accurate than census results.

To compensate for suspected nonrandom errors, adjustments of the sample estimates are often made. For example, adjustments are frequently made for nonresponse, both total and partial. Adjustments made for either type of nonresponse are often referred to as *imputations.* Imputation for total nonresponse is usually made by substituting for the questionnaire responses of the nonrespondents the "average" questionnaire responses of the respondents. These imputations usually are made separately within various groups of sample members, formed by attempting to place respondents and nonrespondents together that have "similar" design or ancillary characteristics. Imputation for item nonresponse is usually made by substituting for a missing item the response to that item of a respondent having characteristics that are "similar" to those of the non-respondent.

For an estimate calculated from a sample survey, the *total error* in the estimate is composed of the sampling error, which can usually be estimated from the sample, and the nonsampling error, which usually cannot be estimated from the sample. The total error present in a population quantity obtained from a complete census is composed of only nonsampling errors. Ideally, estimates of the total error associated with data given in the *Statistical Abstract* tables should be given. However, due to the unavailability of estimates of nonsampling errors, only estimates of the levels of sampling errors, in terms of estimated standard errors or coefficients of variation, are available. To obtain estimates of the estimated standard errors from the sample of interest, obtain a copy of the referenced report which appears at the end of each table.

Source of Additional Material: The Federal Committee on Statistical Methodology (FCSM) is an interagency committee dedicated to improving the quality of federal statistics <http://fcsm.ssd.census.gov>.

Principal databases—Beginning below are brief descriptions of 35 of the sample surveys and censuses that provide a substantial portion of the data contained in this *Abstract.*

U.S. DEPARTMENT OF AGRICULTURE, National Agriculture Statistics Service

Basic Area Frame Sample

Universe, Frequency, and Types of Data: June agricultural survey collects data on planted acreage and livestock inventories. The survey also serves to measure list incompleteness and is subsampled for multiple frame surveys.

Type of Data Collection Operation: Stratified probability sample of about 11,000 land area units of about 1 sq. mile (range from 0.1 sq. mile in cities to several sq. miles in open grazing areas). Sample includes 42,000 parcels of agricultural land. About 20 percent of the sample replaced annually.

Data Collection and Imputation Procedures: Data collection is by personal enumeration. Imputation is based on enumerator observation or data reported by respondents having similar agricultural characteristics.

Estimates of Sampling Error: Estimated CVs range from 1 percent to 2 percent for regional estimates to 3 percent to 6 percent for state estimates of major crop acres and livestock inventories.

Appendix III   943

Other (nonsampling) Errors: Minimized through rigid quality controls on the collection process and careful review of all reported data.

Sources of Additional Material: U.S. Department of Agriculture, National Agricultural Statistics Service, USDA's National Agricultural Statistics Service: The Fact Finders of Agriculture, September 1994.

Census of Agriculture

Universe, Frequency, and Types of Data: Complete count of U.S. farms and ranches conducted once every 5 years with data at the national, state, and county level. Data published on farm numbers and related items/characteristics.

Type of Data Collection Operation: Complete census for number of farms; land in farms; agriculture products sold; total cropland; irrigated land; farm operator characteristics; livestock and poultry inventory and sales; and selected crops harvested. Market value of land and buildings, total farm production expenses, machinery and equipment, fertilizer and chemicals, and farm labor are estimated from a sample of farms.

Data Collection and Imputation Procedures: Data collection takes place by mailing questionnaires to all farmers and ranchers. Nonrespondents are contacted by telephone and correspondence follow-ups. Imputations were made for all nonresponse items/characteristics. Coverage adjustments were made to account for missed farms and ranches.

Estimates of Sampling Error: Variability in the estimates is due to the sample selection and estimation for items collected by sample and census nonresponse and coverage estimation procedures. The CVs for national and state estimates are generally very small. The response rate is approximately 81 percent.

Other (nonsampling) Errors: Nonsampling errors are due to incompleteness of the census mailing list, duplications on the list, respondent reporting errors, errors in editing reported data, and in imputation for missing data. Evaluation studies are conducted to measure certain nonsampling errors such as list coverage and classification error. Results from the evaluation program for the 2002 census indicate the net under-coverage amounted to about 18 percent of the nation's total farms.

Sources of Additional Material: U.S. Department of Agriculture (NASS), *2002 Census of Agriculture, Volume 1, Subject Series C Part 1, Agriculture Atlas of the U.S.;* Part 2, *Coverage Evaluation;* Part 3, *Rankings of States and Counties;* Part 4, *History;* Part 5, *ZIP Code Tabulation of Selected Items;* and Volume 3 *Special Studies,* Part 1, *Farm and Ranch Irrigation Survey;* Part 2, *Census of Horticultural Specialties;* Part 3, *Census of Aquaculture.*

Multiple Frame Surveys

Universe, Frequency, and Types of Data: Surveys of U.S. farm operators are taken to obtain data on major livestock inventories, selected crop acreage and production, grain stocks, and farm labor characteristics; farm economic data and chemical use data.

Type of Data Collection Operation: Primary frame is obtained from general or special purpose lists, supplemented by a probability sample of land areas used to estimate for list incompleteness.

Data Collection and Imputation Procedures: Mail, telephone, or personal interviews used for initial data collection. Mail nonrespondent follow-up by phone and personal interviews. Imputation based on average of respondents.

Estimates of Sampling Error: Estimated CV for number of hired farm workers is about 3 percent. Estimated CVs range from 1 percent to 2 percent for regional estimates to 3 percent to 6 percent for state estimates of livestock inventories and crop acreage.

Other (nonsampling) Errors: In addition to above, replicated sampling procedures used to monitor effects of changes in survey procedures.

Sources of Additional Material: U.S. Department of Agriculture, National Agricultural Statistics Service), USDA's National Agricultural Statistics Service: The Fact Finders of Agriculture, September 1994.

Objective Yield Surveys

Universe, Frequency, and Types of Data: Surveys for data on corn, cotton, potatoes, soybeans, and wheat, to forecast and estimate yields.

Type of Data Collection Operation: Random location of plots in probability sample. Corn, cotton, soybeans, spring wheat, and durum wheat selected in June from Basic Area Frame Sample (see above). Winter wheat and potatoes selected from March and June multiple frame surveys, respectively.

Data Collection and Imputation Procedures: Enumerators count and measure plant characteristics in sample fields. Production measured from plots at harvest. Harvest loss measured from post harvest gleanings.

Estimates of Sampling Error: CVs for national estimates of production are about 2–3 percent.

Other (nonsampling) Errors: In addition to above, replicated sampling procedures used to monitor effects of changes in survey procedures.

Sources of Additional Material: U.S. Department of Agriculture, National Agricultural Statistics Service), USDA's National Agricultural Statistics Service: The Fact Finders of Agriculture, September 1994.

U.S. BUREAU OF JUSTICE STATISTICS (BJS)

National Crime Victimization Survey

Universe, Frequency, and Types of Data: Monthly survey of individuals and households in the United States to obtain data on criminal victimization of those units for compilation of annual estimates.

Type of Data Collection Operation: National probability sample survey of about 42,000 interviewed households in 203 PSUs selected from a list of addresses from the 1990 census, supplemented by new construction permits and an area sample where permits are not required.

Data Collection and Imputation Procedures: Interviews are conducted every 6 months for 3 years for each household in the sample; 7,000 households are interviewed monthly. Personal interviews are used in the first interview; the intervening interviews are conducted by telephone whenever possible.

Estimates of Sampling Error: CVs for 2003 estimates are: 3.6 percent for personal crimes (includes all crimes of violence plus purse-snatching crimes); 3.6 percent for crimes of violence; 14.0 percent for estimate of rape/sexual assault counts; 8.7 percent for robbery counts; 3.9 percent for assault counts; 14.5 percent for purse snatching (it refers to purse snatching and pocket picking); 1.9 percent for property crimes; 3.8 percent for burglary counts; 2.1 percent for theft (of property); and 5.9 percent for motor vehicle theft counts.

Other (nonsampling) Errors: Respondent recall errors which may include reporting incidents for other than the reference period; interviewer coding and processing errors; and possible mistaken reporting or classifying of events. Adjustment is made for a household noninterview rate of about 8 percent and for a within-household noninterview rate of 14 percent.

Sources of Additional Material: U.S. Bureau of Justice Statistics, *Criminal Victimization in the United States,* annual.

U.S. Bureau of Labor Statistics

Consumer Expenditure Survey (CE)

Universe, Frequency and Types of Data: Consists of two continuous components: a quarterly Interview survey and a weekly diary or record-keeping survey. They are nationwide surveys that collect data on consumer expenditures, income, characteristics, and assets and liabilities. Samples are national probability samples of households that are representative of the civilian noninstitutionalized population. The surveys have been ongoing since 1980.

Type of Data Collection Operation: The Interview Survey is a panel rotation survey. Each panel is interviewed for five quarters and then dropped from the survey. About 7,500 consumer units are

interviewed each quarter. The Diary Survey sample is new each year and consists of about 7,500 consumer units. Data have been collected on an ongoing basis in 105 PSUs since 1996.

Data Collection and Imputation Procedures: For the Interview Survey, data are collected by personal interview with each consumer unit interviewed once per quarter for five consecutive quarters. Designed to collect information that respondents can recall for 3 months or longer, such as large or recurring expenditures. For the Diary Survey, respondents record all their expenditures in a self-reporting diary for two consecutive one-week periods. Designed to pick up items difficult to recall over a long period, such as detailed food expenditures. Missing or invalid attributes or expenditures are imputed. Income, assets, and liabilities are not imputed. The U.S. Census Bureau collects the data for the Bureau of Labor Statistics.

Estimates of Sampling Error: Standard error tables have been available since 2000.

Other (nonsampling) Errors: Includes incorrect information given by respondents, data processing errors, interviewer errors, and so on. They occur regardless of whether data are collected from a sample or from the entire population.

Sources of Additional Material: Bureau of Labor Statistics, Internet site <http://www.bls.gov/cex>.

Consumer Price Index (CPI)

Universe, Frequency, and Types of Data: Monthly survey of price changes of all types of consumer goods and services purchased by urban wage earners and clerical workers prior to 1978, and urban consumers thereafter. Both indexes continue to be published.

Type of Data Collection Operation: Prior to 1978, and since 1998, sample of various consumer items in 87 urban areas; from 1978–1997, in 85 PSUs, except from January 1987 through March 1988, when 91 areas were sampled.

Data Collection and Imputation Procedures: Prices of consumer items are obtained from about 50,000 housing units, and 23,000 other reporters in 87 areas. Prices of food, fuel, and a few other items are obtained monthly; prices of most other commodities and services are collected every month in the three largest geographic areas and every other month in others.

Estimates of Sampling Error: Estimates of standard errors are available.

Other (nonsampling) Errors: Errors result from inaccurate reporting, difficulties in defining concepts and their operational implementation, and introduction of product quality changes and new products.

Sources of Additional Material: U.S. Bureau of Labor Statistics, Internet site <http://www.stats.bls.gov/cpi/home.htm> and *BLS Handbook of Methods,* Chapter 17, Bulletin 2490.

Current Employment Statistics (CES) Program

Universe, Frequency, and Types of Data: Monthly survey drawn from a sampling frame of over 8 million unemployment insurance tax accounts in order obtain data by industry on employment, hours, and earnings.

Type of Data Collection Operation: In 2004, the CES sample included about 160,000 businesses and government agencies, which represent approximately 400,000 individual worksites.

Data Collection and Imputation Procedures: Each month, the state agencies cooperating with BLS, as well as BLS Data Collection Centers, collect data through various automated collection modes and mail. BLS-Washington staff prepares national estimates of employment, hours, and earnings while states use the data to develop state and area estimates.

Estimates of Sampling Errors: The relative standard error for total nonfarm employment is 0.1 percent.

Other (nonsampling) Errors: Estimates of employment adjusted annually to reflect complete universe. Average adjustment is 0.3 percent over the last decade, with an absolute range from less than 0.05 percent to 0.5 percent.

Sources of Additional Material: U.S. Bureau of Labor Statistics, Employment and Earnings, monthly, *Explanatory Notes and Estimates of Errors,* Tables 2-A through 2-F.

National Compensation Survey

Universe, Frequency, and Types of Data: Nationwide sample survey of establishments of all employment size classes, stratified by geographic area, in private industry and state and local government. Data collected include wages and salaries, and employer costs for employee compensation, employment cost index, and employee benefits. Data produced include percent changes in the cost of employment cited in the Employment Cost Index (ECI) and costs per hour worked for individual benefits cited in the Employer Costs for Employee Compensation (ECEC). The survey provides data by ownership (private industry and state and local government), industry sector, major industry divisions, major occupational groups, bargaining status, metropolitan area status, and census region. ECEC also provides data by establishment size class.

Type of Data Collection Operation: Probability proportionate to size sample of establishments. The sample is replaced on a continual basis. Establishments are in the survey for approximately 5 years.

Data Collection and Imputation Procedures: For the initial visit, data are primarily collected in a personal visit to the establishment. Quarterly updates are obtained primarily by mail, fax, and telephone. Imputation is done for individual benefits.

Estimates of Sampling Error: Because standard errors vary from quarter to quarter, the ECI uses a 5-year moving average of standard errors to evaluate published series. These standard errors are available at <http://www.bls.gov/ncs/ect/home.htm>.

Other (nonsampling) Errors: Nonsampling errors have a number of potential sources. The primary sources are (1) survey nonresponse and (2) data collection and processing errors. Nonsampling errors are not measured.

Procedures have been implemented for reducing nonsampling errors, primarily through quality assurance programs. These programs include the use of data collection reinterviews, observed interviews, computer edits of the data, and systematic professional review of the reports on which the data are recorded. The programs also serve as a training device to provide feedback to the field economists, or data collectors, on errors. They also provide information on the sources of error which can be remedied by improved collection instructions or computer processing edits. Extensive training of field economists is also conducted to maintain high standards in data collection.

Sources of Additional Material: Bureau of Labor Statistics, *BLS Handbook of Methods,* Chapter 8 (Bulletin 2490) and <http://www.bls.gov/ncs>.

Producer Price Index (PPI)

Universe, Frequency, and Types of Data: Monthly survey of producing companies to determine price changes of all commodities and services produced in the United States for sale in commercial transactions. Data on agriculture, forestry, fishing, manufacturing, mining, gas, electricity, public utilities, wholesale trade, retail trade, transportation, healthcare, and other services.

Type of Data Collection Operation: Probability sample of approximately 30,000 establishments that result in about 100,000 price quotations per month.

Data Collection and Imputation Procedures: Data are collected by mail and facsimile. If transaction prices are not supplied, list prices are used. Some prices are obtained from trade publications, organized exchanges, and government agencies. To calculate index, price changes are multiplied by their relative weights taken from the Census Bureau's 1997 shipment values from their Census of Industries.

Estimates of Sampling Error: Not applicable.

Other (nonsampling) Errors: Not available at present.

Sources of Additional Material: U.S. Bureau of Labor Statistics, *BLS Handbook of Methods,* Chapter 14, Bulletin 2490. U.S. Bureau of Labor Statistics Internet sites <http://stats.bls.gov/ppi>.

BOARD OF GOVERNORS OF THE FEDERAL RESERVE SYSTEM

Survey of Consumer Finances

Universe, Frequency, and Types of Data: Periodic sample survey of families. In this survey a given household is divided into a primary economic unit and other economic units. The primary economic unit, which may be a single individual, is generally chosen as the unit that contains the person who either holds the title to the home or is the first person listed on the lease. The primary unit is used as the reference family. The survey collects detailed data on the composition of family balance sheets, the terms of loans, and relationships with financial institutions. It also gathered information on the employment history and pension rights of the survey respondent and the spouse or partner of the respondent.

Type of Data Collection Operation: The survey employs a two-part strategy for sampling families. Some families were selected by standard multistage area probability sampling methods applied to all 50 states. The remaining families in the survey were selected using statistical records derived from tax returns, under the strict rules governing confidentiality and the rights of potential respondents to refuse participation.

Data Collection and Imputation Procedures: NORC at the University of Chicago has collected data for the survey since 1992. Since 1995, the survey has used computer-assisted personal interviewing. Adjustments for nonresponse are made through multiple imputation of unanswered questions and through weighting adjustments based on data used in the sample design for families that refused participation.

Estimates of Sampling Error: Because of the complex design of the survey, the estimation of potential sampling errors is not straightforward. A replicate-based procedure is available.

Other (nonsampling) Errors: The survey aims to complete 4,500 interviews, with about two thirds of that number deriving from the area-probability sample. The response rate is typically about 70 percent for the area-probability sample and about 35 percent over all strata in the tax-data sample. Proper training and monitoring of interviewers, careful design of questionnaires, and systematic editing of the resulting data were used to control inaccurate survey responses.

Sources of Additional Material: Board of Governors of the Federal Reserve System, "Recent Changes in U.S. Family Finances: Evidence from the 1998 and 2001 Survey of Consumer Finances," *Federal Reserve Bulletin,* January 2003.

U.S. CENSUS BUREAU

2002 Economic Census (Industry Series, Geographic Area Series and Subject Series Reports) (for NAICS sectors 22, 42, 44-45, 48-49, and 51-81)

Universe, Frequency, and Types of Data: Conducted every 5 years to obtain data on number of establishments, number of employees, total payroll size, total sales/receipts/revenue, and other industry-specific statistics. In 2002, the universe was all employer and nonemployer establishments primarily engaged in wholesale, retail, utilities, finance & insurance, real estate, transportation & warehousing, information, education, health care, and other service industries.

Type of Data Collection Operation: All large employer firms were surveyed (i.e., all employer firms above payroll size cutoffs established to separate large from small employers) plus a 5 percent to 25 percent sample of the small employer firms. Firms with no employees were not required to file a census return.

Data Collection and Imputation Procedures: Mail questionnaires were used with both mail and telephone follow-ups for nonrespondents. Data for nonrespondents and for small employer firms not mailed a questionnaire were obtained from administrative records of

948   Appendix III

other federal agencies or imputed. Non-employer data were obtained exclusively from IRS 2002 income tax returns.

Estimates of Sampling Error: Not applicable for basic data such as sales, revenue, receipts, payroll, etc.

Other (nonsampling) Errors: Establishment response rates by NAICS sector in 2002 ranged from 80 percent to 89 percent. Item response rates generally ranged from 50 percent to 90 percent with lower rates for the more detailed questions. Nonsampling errors may occur during the collection, reporting, and keying of data, and due to industry misclassification.

Sources of Additional Material: U.S. Census Bureau, *2002 Economic Census: Industry Series, Geographic Area Series* and *Subject Series Reports* (by NAICS sector), Appendix C and <http://www.census.gov/econ/census02/guide/index.html>.

American Community Survey

Universe, Frequency, and Types of Data: Nationwide survey to obtain data about demographic, social, economic, and housing characteristics of people, households and housing units. Covers household population and excludes the population living in institutions, college dormitories, and other group quarters.

Type of Data Collection Operation: Two-stage stratified annual sample of approximately 829,000 housing units. The ACS samples housing units from the Master Address File (MAF). The first stage of sampling involves dividing the United States into primary sampling units (PSUs) most of which comprise a metropolitan area, a large county, or a group of smaller counties. Every PSU falls within the boundary of a state. The PSUs are then grouped into strata on the basis of independent information, that is, information obtained from the decennial census or other sources. The strata are constructed so that they are as homogeneous as possible with respect to social and economic characteristics that are considered important by ACS data users. A pair of PSUs were selected from each stratum. The probability of selection for each PSU in the stratum is proportional to its estimated 1996 population. In the second stage of sampling, a sample of housing units within the sample PSUs is drawn. Ultimate sampling units (USUs) are housing units. The USUs sampled in the second stage consist of housing units which are systematically drawn from sorted lists of addresses of housing units from the MAF.

Data Collection and Imputation Procedures: The American Community Survey (ACS) is conducted every month on independent samples. Each housing unit in the independent monthly samples is mailed a prenotice letter announcing the selection of the address to participate, a survey questionnaire package, and a reminder postcard. These sample units receive a second (replacement) questionnaire package if the initial questionnaire has not been returned by a scheduled date. In the mail-out/mail-back sites, sample units for which a questionnaire is not returned in the mail and for which a telephone number is available are defined as the telephone nonresponse follow-up universe. Interviewers attempt to contact and interview these mail nonresponse cases. Sample units from all sites that are still unresponsive two months after the mailing of the survey questionnaires and directly after the completion of the telephone followup operation are sub-sampled at a rate of 1 in 3. The selected nonresponse units are assigned to Field Representatives, who visit the units, verify their existence or declare them nonexistent, determine their occupancy status, and conduct interviews. After data collection is completed, any remaining incomplete or inconsistent information was imputed during the final automated edit of the collected data.

Estimates of Sampling Error: The data in the ACS products are estimates of the actual figures that would have been obtained by interviewing the entire population using the same methodology. The estimates from the chosen sample also differ from other samples of housing units and persons within those housing units.

Other (nonsampling) Errors: Nonsampling Error — In addition to sampling error,

data users should realize that other types of errors may be introduced during any of the various complex operations used to collect and process survey data. An important goal of the ACS is to minimize the amount of nonsampling error introduced through nonresponse for sample housing units. One way of accomplishing this is by following up on mail nonrespondents.

Sources of Additional Material: U.S. Census Bureau, American Community Survey Web site available on the Internet, <http://www.census.gov/acs/www/index.html>. U.S. Census Bureau, American Community Survey, Accuracy of the Data documents available on the Internet, <http://www.census.gov/acs/www/UseData/Accuracy/Accuracy1.htm>.

American Housing Survey

Universe, Frequency, and Types of Data: Conducted nationally in the fall in odd numbered years to obtain data on the approximately 121 million occupied or vacant housing units in the United States (group quarters are excluded). Data include characteristics of occupied housing units, vacant units, new housing and mobile home units, financial characteristics, recent mover households, housing and neighborhood quality indicators, and energy characteristics.

Type of Data Collection Operation: The national sample was a multistage probability sample with about 61,000 units eligible for interview in 2003. Sample units, selected within 394 PSUs, were surveyed over a 4-month period.

Data Collection and Imputation Procedures: For 2003, the survey was conducted by personal interviews. The interviewers obtained the information from the occupants or, if the unit was vacant, from informed persons such as landlords, rental agents, or knowledgeable neighbors.

Estimates of Sampling Error: For the national sample, illustrations of the S.E. of the estimates are provided in Appendix D of the 2003 report. As an example, the estimated CV is about 0.2 percent for the estimated percentage of owner-occupied units with two persons.

Other (nonsampling) Errors: Response rate was about 92 percent. Nonsampling errors may result from incorrect or incomplete responses, errors in coding and recording, and processing errors. For the 2003 national sample, approximately 2.2 percent of the total housing inventory was not adequately represented by the AHS sample.

Sources of Additional Material: U.S. Census Bureau, *Current Housing Reports,* Series H-150 and H-170, *American Housing Survey.* <http://www.census.gov/hhes/www/ahs.html>.

Annual Survey of Manufactures

Universe, Frequency, and Types of Data: The Annual Survey of Manufactures (ASM) is conducted annually, except for years ending in 2 and 7 for all manufacturing establishments having one or more paid employees. The purpose of the ASM is to provide key intercensal measures of manufacturing activity, products, and location for the public and private sectors. The ASM provides statistics on employment, payroll, worker hours, payroll supplements, cost of materials, value added by manufacturing, capital expenditures, inventories, and energy consumption. It also provides estimates of value of shipments for 1,800 classes of manufactured products.

Type of Data Collection Operation: The ASM includes approximately 57,000 establishments selected from the census universe of 366,000 manufacturing establishments. Some 27,000 large establishments are selected with certainty, and some 30,000 other establishments are selected with probability proportional to a composite measure of establishment size. The survey is updated from two sources: Internal Revenue Service administrative records are used to include new single-unit manufacturers and the Company Organization Survey identifies new establishments of multiunit forms.

Data Collection and Imputation Procedures: Survey is conducted by mail with phone and mail follow-ups of nonrespondents. Imputation (for all nonresponse items) is based on previous year reports, or for new establishments in survey, on industry averages.

Estimates of Sampling Error: Estimated standard errors for number of employees, new expenditures, and for value added totals are given in annual publications. For U.S. level industry statistics, most estimated standard errors are 2 percent or less, but vary considerably for detailed characteristics.

Other (nonsampling) Errors: Response rate is about 85 percent. Nonsampling errors include those due to collection, reporting, and transcription errors, many of which are corrected through computer and clerical checks.

Sources of Additional Material: U.S. Census Bureau, *Annual Survey of Manufactures,* and Technical Paper 24. <http://www.census.gov/econ/www/mancen.html>

Annual Surveys of State and Local Government

Universe, Frequency, and Types of Data: Sample survey conducted annually to obtain data on revenue, expenditure, debt, and employment of state and local governments. Universe is all governmental units in the United States (about 87,500).

Type of Data Collection Operation: Sample survey includes all state governments, county governments with 100,000+ population, municipalities with 75,000+ population, townships with 50,000+ population, all independent school districts with 10,000+ enrollment in March 2002, all school districts providing college level (postsecondary) education, and other governments meeting certain criteria; probability sample for remaining units.

Data Collection and Imputation Procedures: Field and office compilation of data from official records and reports for states and large local governments; central collection of local governmental financial data through cooperative agreements with a number of state governments; mail canvass of other units with mail and telephone follow-ups of nonrespondents. Data for nonresponses are imputed from previous year data or obtained from secondary sources, if available.

Estimates of Sampling Error: State and local government totals are generally subject to sampling variability of less than 3 percent.

Other (nonsampling) Errors: Nonresponse rate is less than 10 percent for local governments. Other possible errors may result from undetected inaccuracies in classification, response, and processing.

Sources of Additional Material: Publications: <http://www.census.gov/prod/www/abs/govern.html>:U.S. Census Bureau, *Public Employment in 1992,* GE 92, No. 1, *Governmental Finances in 1991–1992,* GF 92, No. 5, and *Census of Governments, 1997 and 2002,* various reports. Web site references: Census of Governments <http://www.census.gov/govs/www/cog2002.html>, <http://www.census.gov/govs/www/cog.html>. Employment—state and local site: <http://www.census.gov/govs/www/apes.html>. Finance—state and local site: <http://www.census.gov/govs/www/estimate.html>.

Census of Population

Universe, Frequency, and Types of Data: Complete count of U.S. population conducted every 10 years since 1790. Data obtained on number and characteristics of people in the U.S.

Type of Data Collection Operation: In 1980, 1990, and 2000, complete census for some items: age, date of birth, sex, race, and relationship to householder. In 1980, approximately 19 percent of the housing units were included in the sample; in 1990 and 2000, approximately 17 percent.

Data Collection and Imputation Procedures: In 1980, 1990, and 2000, mail questionnaires were used extensively with personal interviews in the remainder. Extensive telephone and personal follow-up for nonrespondents was done in the censuses. Imputations were made for missing characteristics.

Estimates of Sampling Error: Sampling errors for data are estimated for all items collected by sample and vary by characteristic and geographic area. The coefficients of variation (CVs) for national and state estimates are generally very small.

Other (nonsampling) Errors: Since 1950, evaluation programs have been conducted to provide information on the magnitude of some sources of nonsampling errors such as response bias and undercoverage in each census. Results from the evaluation program for the 1990 census indicated that the estimated net undercoverage amounted to about 1.5 percent of the total resident population. For Census 2000, the evaluation program indicates a net overcount of 0.5 percent of the resident population.

Sources of Additional Material: U.S. Census Bureau, The Coverage of Population in the 1980 Census, PHC80-E4; *Content Reinterview Study: Accuracy of Data for Selected Population and Housing Characteristics as Measured by Reinterview,* PHC80-E2; *1980 Census of Population,* Vol. 1, (PC80-1), Appendixes B, C, and D. 1990 Census of Population and Housing, Content Reinterview Survey: Accuracy of Data for Selected Population and Housing Characteristics as measured by Reinterview, CPH-E-1; 1990 Census of Population and Housing, Effectiveness of Quality Assurance, CPH-E-2; Programs to Improve Coverage in the 1990 Census, CPH-E-3. For Census 2000, see <http://www.census.gov/pred/www>.

County Business Patterns

Universe, Frequency, and Types of Data: County Business Patterns is an annual tabulation of basic data items extracted from the Business Register, a file of all known single- and multi-location companies maintained and updated by the Census Bureau. Data include number of establishments, number of employees, first quarter and annual payrolls, and number of establishments by employment size class. Data are excluded for self-employed persons, domestic service workers, railroad employees, agricultural production workers, and most government employees.

Type of Data Collection Operation: The annual Company Organization Survey provides individual establishment data for multi-location companies. Data for single establishment companies are obtained from various Census Bureau programs, such as the Annual Survey of Manufactures and Current Business Surveys, as well as from administrative records of the Internal Revenue Service and the Social Security Administration.

Estimates of Sampling Error: Not applicable.

Other (nonsampling) Error: The data are subject to non-sampling errors, such as industry classification errors, as well as errors of response, keying, and non-reporting.

Sources of Additional Materials: U.S. Census Bureau, *General Explanation of County Business Patterns.*

Current Population Survey (CPS)

Universe, Frequency, and Types of Data: Nationwide monthly sample survey of civilian noninstitutionalized population, 15 years old or over, to obtain data on employment, unemployment, and a number of other characteristics.

Type of Data Collection Operation: Multi-stage probability sample of about 50,000 households in 754 PSUs in 1996 expanded to about 60,000 households in July 2001. Over-sampling in some states and the largest MSAs to improve reliability for those areas of employment data on annual average basis. A continual sample rotation system is used. Households are in sample 4 months, out for 8 months, and in for 4 more. Month-to-month overlap is 75 percent; year-to-year overlap is 50 percent.

Data Collection and Imputation Procedures: For first and fifth months that a household is in sample, personal interviews; other months, approximately 85 percent of the data collected by phone. Imputation is done for both item and total nonresponse. Adjustment for total nonresponse is done by a predefined cluster of units, by MSA size and residence; for item nonresponse imputation varies by subject matter.

Estimates of Sampling Error: Estimated CVs on national annual averages for labor force, total employment, and nonagricultural employment, 0.2 percent; for total unemployment and agricultural employment, 1.0 percent to 2.5 percent. The estimated CVs for family income and poverty rate for all persons

in 1986 are 0.5 percent and 1.5 percent, respectively. CVs for subnational areas, such as states, would be larger and would vary by area.

Other (nonsampling) Errors: Estimates of response bias on unemployment are not available, but estimates of unemployment are usually 5 percent to 9 percent lower than estimates from reinterviews. Six to 7.0 percent of sample households unavailable for interviews.

Sources of Additional Material: U.S. Census Bureau and Bureau of Labor Statistics, *Current Population Survey; Design and Methodology,* (Tech. Paper 63), available on Internet <http://www .census.gov/prod/2002pubs/tp63rv.pdf> and Bureau of Labor Statistics, *Employment and Earnings,* monthly, Explanatory Notes and Estimates of Error, Household Data and *BLS Handbook of Methods,* Chapter 1, available on the Internet <http://www.bls.gov/opub /hom/homch1a.htm>.

Foreign Trade—Import Statistics

Universe, Frequency, and Types of Data: The import entry documents collected by U.S. Bureau of Customs and Border Protection are processed each month to obtain data on the movement of merchandise imported into the United States. Data obtained include value, quantity, and shipping weight by commodity, country of origin, district of entry, and mode of transportation.

Type of Data Collection Operation: Import entry documents, either paper or electronic, are required to be filed for the importation of goods into the United States valued over $2,000, or for articles which must be reported on formal entries. U.S. Bureau of Customs and Border Protection officials collect and transmit statistical copies of the documents to the Census Bureau on a flow basis for data compilation. Estimates for shipments valued under $2,001 and not reported on formal entries are based on estimated established percentages for individual country totals.

Data Collection and Imputation Procedures: Statistical copies of import entry documents, received on a daily basis from ports of entry throughout the country, are subjected to a monthly processing cycle. They are fully processed to the extent they reflect items valued at $2,001 and over or items which must be reported on formal entries.

Estimates of Sampling Error: Not applicable.

Other (nonsampling) Errors: The goods data are a complete enumeration of documents collected by the U.S. Bureau of Customs and Border Protection and are not subject to sampling errors; but they are subject to several types of nonsampling errors. Quality assurance procedures are performed at every stage of collection, processing and tabulation; however the data are still subject to several types of nonsampling errors. The most significant of these include reporting errors, undocumented shipments, timeliness, data capture errors, and errors in the estimation of low-valued transactions.

Sources of Additional Material: U.S. Census Bureau, FT 900 U.S. International Trade in Goods and Services, FT 925 (discounted after 1996) U.S. Merchandise Trade, FT 895 U.S. Trade with Puerto Rico and U.S. Possessions, FT920 U.S. Merchandise Trade: selected highlights, and Information Section on Goods and Services at <http://www.census.gov/ft900>.

Foreign Trade—Export Statistics

Universe, Frequency, and Types of Data: The export declarations collected by U.S. Bureau of Customs and Border Protection are processed each month to obtain data on the movement of U.S. merchandise exports to foreign countries. Data obtained include value, quantity, and shipping weight of exports by commodity, country of destination, district of exportation, and mode of transportation.

Type of Data Collection Operation: Shipper's Export Declarations (paper and electronic) are generally required to be filed for the exportation of merchandise valued over $2,500. U.S. Bureau of Customs and Border Protection officials collect and transmit the documents to the Census Bureau on a flow basis for data

compilation. Data for shipments valued under $2,501 are estimated, based on established percentages of individual country totals.

Data Collection and Imputation Procedures: Statistical copies of Shipper's Export Declarations are received on a daily basis from ports throughout the country and subject to a monthly processing cycle. They are fully processed to the extent they reflect items valued over $2,500. Estimates for shipments valued at $2,500 or less are made, based on established percentages of individual country totals.

Estimates of Sampling Error: Not applicable.

Other (nonsampling) Errors: The goods data are a complete enumeration of documents collected by the U.S. Bureau of Customs and Border Protection and are not subject to sampling errors; but they are subject to several types of nonsampling errors. Quality assurance procedures are performed at every stage of collection, processing and tabulation; however the data are still subject to several types of nonsampling errors. The most significant of these include reporting errors, undocumented shipments, timeliness, data capture errors, and errors in the estimation of low-valued transactions.

Sources of Additional Material: U.S. Census Bureau, FT 900 U.S. International Trade in Goods and Services, FT 925 (discounted after 1996) U.S. Merchandise Trade, FT 895 U.S. Trade with Puerto Rico and U.S. Possessions, FT 920 U.S. Merchandise trade: selected highlights, and Information Section on Goods and Services at <http://census.gov/ft900>.

Monthly Retail Trade and Food Service Survey

Universe, Frequency, and Types of Data: Provides monthly estimates of retail and food service sales by kind of business and end-of-month inventories of retail stores.

Type of Data Collection Operation: Probability sample of all firms from a list frame. The list frame is the Census Bureau's Business Register updated quarterly for recent birth Employer Identification (EI) Numbers issued by the Internal Revenue Service and assigned a kind-of-business code by the Social Security Administration. The largest firms are included monthly; a sample of others is included every month also.

Data Collection and Imputation Procedures: Data are collected by mail questionnaire with telephone follow-ups and fax reminders for nonrespondents. Imputation is made for each nonresponse item and each item failing edit checks.

Estimates of Sampling Error: For the 2004 monthly surveys, CVs are about 0.5 percent for estimated total retail sales and 1.2 percent for estimated total retail inventories. Sampling errors are shown in monthly publications.

Other (nonsampling) Errors: Imputation rates are about 20 percent for monthly retail and food service sales, and 28 percent for monthly retail inventories.

Sources of Additional Material: U.S. Census Bureau, Current Business Reports, Annual Benchmark Report for Retail Trade and Food Services.

Monthly Survey of Construction

Universe, Frequency, and Types of Data: Survey conducted monthly of newly constructed housing units (excluding mobile homes). Data are collected on the start, completion, and sale of housing. (Annual figures are aggregates of monthly estimates.)

Type of Data Collection Operation: For permit issuing places probability sample of 850 housing units obtained from 19,000 permit issuing places. For non-permit places, multistage probability sample of new housing units selected in 169 PSUs. In those areas, all roads are canvassed in selected enumeration districts.

Data Collection and Imputation Procedures: Data are obtained by telephone inquiry and field visit.

Estimates of Sampling Error: Estimated CV of 3 percent to 4 percent for estimates of national totals, but may be

higher than 20 percent for estimated totals of more detailed characteristics, such as housing units in multiunit structures.

Other (nonsampling) Errors: Response rate is over 90 percent for most items. Nonsampling errors are attributed to definitional problems, differences in interpretation of questions, incorrect reporting, inability to obtain information about all cases in the sample, and processing errors.

Sources of Additional Material: All data are available on the Internet at <http://www.census.gov/const/www/newsresconstindex.html>. Further documentation of the survey is also available at that site.

Nonemployer Statistics

Universe, Frequency, and Types of Data: Nonemployer statistics are an annual tabulation of economic data by industry for active businesses without paid employees that are subject to federal income tax. Data showing the number of establishments and receipts by industry are available for the U.S., states, counties, and metropolitan areas. Most types of businesses covered by the Census Bureau's economic statistics programs are included in the nonemployer statistics. Tax-exempt and agricultural-production businesses are excluded from nonemployer statistics.

Type of Data Collection Operation: The universe of nonemployer establishments is created annually as a by-product of the Census Bureau's Business Register processing for employer establishments. If a business is active but without paid employees, then it becomes part of the potential nonemployer universe. Industry classification and receipts are available for each potential nonemployer business. These data are obtained primarily from the annual business income tax returns of the Internal Revenue Service (IRS). The potential nonemployer universe undergoes a series of complex processing, editing, and analytical review procedures at the Census Bureau to distinguish nonemployers from employers, and to correct and complete data items used in creating the data tables.

Estimates of Sampling Error: Not applicable.

Other (nonsampling) Errors: The data are subject to nonsampling errors, such as errors of self-classification by industry on tax forms, as well as errors of response, keying, nonreporting, and coverage.

Sources of Additional Material: U. S. Census Bureau, Nonemployer Statistics: 2002 (Introduction; Coverage and Methodology). See also <http://www.census.gov/epcd/nonemployer/view/cov&meth.htm>

Service Annual Survey

Universe, Frequency, and Types of Data: The U.S. Census Bureau conducts the Service Annual Survey to provide national estimates of revenues, expenses, and e-commerce revenues for taxable and tax-exempt firms classified in selected service industries. Estimates are summarized by industry classification based on the *1997 North American Industry Classification System (NAICS).* Industries covered by the Service Annual Survey include all or part of the following NAICS sectors: Transportation and Warehousing (NAICS 48–49); Information (NAICS 51); Finance and Insurance (NAICS 52); Real Estate and Rental and Leasing (NAICS 53); Professional, Scientific, and Technical Services (NAICS 54); Administrative and Support and Waste Management and Remediation Services (NAICS 56); Health Care and Social Assistance (NAICS 62); Arts, Entertainment, and Recreation (NAICS 71); and Other Services, except Public Administration (NAICS 81). Data items collected include total revenue, revenue from e-commerce transactions; and for selected industries, revenue from detailed service products, total expenses, and expenses by major type, revenue from exported services, and inventories. Questionnaires are mailed in January and request annual data for the prior year. Estimates are published approximately 12 months after the initial survey mailing.

Type of Data Collection Operation: The *Service Annual Survey* estimates are developed using data from a probability sample and administrative records. *Service Annual Survey* questionnaires are mailed to a probability sample that is periodically reselected from a universe

of firms located in the United States and having paid employees. The sample includes firms of all sizes and covers both taxable firms and firms exempt from Federal income taxes. Updates to the sample are made on a quarterly basis to account for new businesses. Firms without paid employees, or non-employers, are included in the estimates through imputation and/or administrative records data provided by other Federal agencies. Links to additional information about confidentiality protection, sampling error, nonsampling error, sample design, definitions, and copies of the questionnaires may be found on the Internet at <http://www.census.gov/econ/www/servmenu.html>.

Estimates of Sampling Error: Coefficients of variation for the 2003 *Service Annual Survey* estimates range from 0.6% to 2.2% for total revenue estimates computed at the NAICS sector (2-digit NAICS code) level. Sampling errors for more detailed industries are shown in the corresponding publications. The full 2003 *Service Annual Survey* results, including coefficients of variations, can be found at <http://www.census.gov/econ/www/servmenu.html>. Links to additional information regarding sampling error may be found at: <http://www.census.gov/svsd/www/cv.html>.

Other (Nonsampling) Errors: Data are imputed for unit nonresponse, item nonresponse, and for reported data that fails edits. The percent of imputed data for total revenue for the 2003 *Service Annual Survey* is approximately 14%.

Sources of Additional Material: U.S. Census Bureau, Current Business Reports, *Service Annual Survey,* Census Bureau Web site: <http://www.census.gov/econ/www/servmenu.html>.

U.S. DEPARTMENT OF EDUCATION

National Center for Education Statistics Higher Education General Information Survey (HEGIS), Degrees and Other Formal Awards Conferred. Beginning 1986, Integrated Postsecondary Education Data Survey (IPEDS), Completions

Universe, Frequency, and Types of Data: Annual survey of all institutions and branches listed in the *Education Directory, Colleges and Universities* to obtain data on earned degrees and other formal awards, conferred by field of study, level of degree, sex, and by racial/ethnic characteristics (every other year prior to 1989, then annually).

Type of Data Collection Operation: Complete census.

Data Collection and Imputation Procedures: Data are collected through a Web-based survey in the fall of every year. Missing data are imputed by using data of similar institutions.

Estimates of Sampling Error: Not applicable.

Other (nonsampling) Errors: For 2002–03, approximately 100.0 percent response rate for degree-granting institutions.

Sources of Additional Material: U.S. Department of Education, National Center for Education Statistics, *Postsecondary Institutions in the United States: Fall 2003 and Degrees and Other Awards Conferred: 2002–03.* <http://www.nces.ed.gov/ipeds/>.

National Household Education Surveys Program

Universe, Frequency, and Types of Data: The National Household Education Surveys Program (NHES) is a system of telephone surveys of the noninstitutionalized civilian population of the United States. Surveys in NHES have varying universes of interest depending on the particular survey. Specific topics covered by each survey are at the NHES Web site <http://nces.ed.gov/nhes>. A list of the surveys fielded as part of NHES, each universe, and the years they were fielded is provided below. 1. Adult Education Interviews were conducted with a representative sample of civilian, noninstitutionalized persons age 16 and older who were not enrolled in grade 12 or below (1991, 1995, 1999, 2001, 2003). 2. Before-and After-School Programs and Activities Interviews were conducted with parents of a representative sample of students in grades K through 8 (1999, 2001). 3. Civic Involvement Interviews were conducted with a representative sample of parents,

youth, and adults(1996, 1999). 4. Early Childhood Program Participation Interviews were conducted with parents of a representative sample of children from birth through grade 3, with the specific age groups varying by survey year (1991, 1995, 1999, 2001). 5. Household and Library Use Interviews were conducted with a representative sample of U.S. households (1996). 6. Parent and Family Involvement in Education Interviews were conducted with parents of a representative sample of children age 3 through grade 12 or in grades K through 12 depending on the survey year (1996, 1999, 2003). 7. School Readiness Interviews were conducted with parents of a representative sample of 3- to 7-year-old children (1993, 1999). 8. School Safety and Discipline Interviews were conducted with a representative sample of students in grades 6–12, their parents, and the parents of a representative sample of students in grades 3–5 (1993).

Type of Data Collection Operation: NHES uses telephone interviews to collect data.

Data Collection and Imputation Procedures: Telephone numbers are selected using random digit dialing techniques. Approximately 45,000 to 64,000 households are contacted in order to identify persons eligible for the surveys. Data are collected using computer-assisted telephone interviewing (CATI) procedures. Missing data are imputed using hot-deck imputation procedures.

Estimates of Sampling Error: Unweighted sample sizes range between 2,500 and 21,000. The average root design effects of the surveys in NHES range from 1.1 to 4.5.

Other (nonsampling) Errors: Because of unit nonresponse and because the samples are drawn from households with telephone instead of all households, nonresponse and/or coverage bias may exist for some estimates. However, both sources of potential bias are adjusted for in the weighting process. Analyses of both potential sources of bias in the NHES collections have been studied and no significant bias has been detected.

Sources of Additional Material: Please see the NHES Web site at <http://nces.ed.gov/nhes>

U.S. FEDERAL BUREAU OF INVESTIGATION

Uniform Crime Reporting (UCR) Program

Universe, Frequency, and Types of Data: Monthly reports on the number of criminal offenses that become known to law enforcement agencies. Data are collected on crimes cleared by arrest; by age, sex, and race of arrestees and for victims and offenders for homicides, on fatal and nonfatal assaults against law enforcement officers, and on hate crimes reported.

Type of Data Collection Operation: Crime statistics are based on reports of crime data submitted either directly to the FBI by contributing law enforcement agencies or through cooperating state UCR programs.

Data Collection and Imputation Procedures: States with UCR programs collect data directly from individual law enforcement agencies and forward reports, prepared in accordance with UCR standards, to FBI. Accuracy and consistency edits are performed by FBI.

Estimates of Sampling Error: Not applicable.

Other (nonsampling) Errors: Coverage of 93 percent of the population (95 percent in MSAs, 85 percent in "cities outside of metropolitan areas," and 83 percent in nonmetropolitan counties) by UCR Program, through varying number of agencies reporting.

Sources of Additional Material: U.S. Federal Bureau of Investigation, *Crime in the United States,* annual, *Hate Crime Statistics,* annual, *Law Enforcement Officers Killed & Assaulted,* annual, <http://www.fbi.gov/ucr.htm>.

U.S. INTERNAL REVENUE SERVICE

Corporation Income Tax Returns

Universe, Frequency, and Types of Data: Annual study of unaudited corporation income tax returns, Forms 1120,

1120-A, 1120-F, 1120-L, 1120-PC, 1120-REIT, 1120-RIC, and 1120S, filed by corporations or businesses legally defined as corporations. Data provided on various financial characteristics by industry and size of total assets, and business receipts.

Type of Data Collection Operation: Stratified probability sample of approximately 147,000 returns for Tax Year 2002, allocated to sample classes which are based on type of return, size of total assets size of net income or deficit, and selected business activity. Sampling rates for sample classes varied from .25 percent to 100 percent.

Data Collection and Imputation Procedures: Computer selection of sample of tax return records. Data adjusted during editing for incorrect, missing, or inconsistent entries to ensure consistency with other entries on return and to comply with statistical definitions.

Estimates of Sampling Error: Estimated CVs for Tax Year 2002: Returns with assets over $250 million are self-representing. For other returns grouped by assets, CVs ranged from 0.01 percent to 2.16 percent; for amount of net income CV is 0.19 percent.

Other (nonsampling) Errors: Nonsampling errors include coverage errors, processing errors, and response errors.

Sources of Additional Material: U.S. Internal Revenue Service, *Statistics of Income, Corporation Income Tax Returns,* annual.

Partnership Income Tax Returns

Universe, Frequency, and Types of Data: Annual study of unaudited income tax returns of partnerships, Form 1065. Data provided on various financial characteristics by industry.

Type of Data Collection Operation: Stratified probability sample of approximately 34,800 partnership returns from a population of 2.4 million filed during calendar year 2003. The sample is classified based on combinations of gross receipts, net income or loss, and total assets, and on industry. Sampling rates vary from 0.12 percent to 100 percent.

Data Collection and Imputation Procedures: Computer selection of sample of tax return records. Data are adjusted during editing for incorrect, missing, or inconsistent entries to ensure consistency with other entries on return. Data not available due to regulations are not imputed.

Estimates of Sampling Error: Estimated CVs for tax year 2002 (latest available): For number of partnerships, 0.29 percent; business receipts, 0.34 percent; net income, 0.80 percent; net loss, 1.72 percent.

Other (nonsampling) Errors: Processing errors and errors arising from the use of tolerance checks for the data.

Sources of Additional Material: U.S. Internal Revenue Service, *Statistics of Income, Partnership Returns* and *Statistics of Income Bulletin,* Vol. 24, No. 2 (fall 2004).

Sole Proprietorship Income Tax Returns

Universe, Frequency, and Types of Data: Annual study of unaudited income tax returns of nonfarm sole proprietorships, form 1040 with business schedules. Data provided on various financial characteristics by industry.

Type of Data Collection Operation: Stratified probability sample of approximately 50,000 sole proprietorships for tax year 2002. The sample is classified based on presence or absence of certain business schedules; the larger of total income or loss; and size of business plus farm receipts. Sampling rates vary from 0.05 percent to 100 percent.

Data Collection and Imputation Procedures: Computer selection of sample of tax return records. Data adjusted during editing for incorrect, missing, or inconsistent entries to ensure consistency with other entries on return.

Estimates of Sampling Error: Estimated CVs for tax year 2002 are available. For sole proprietorships, business receipts, 0.69 percent; depreciation 1.41 percent.

Other (nonsampling) Errors: Processing errors and errors arising from the use of tolerance checks for the data.

Individual Income Tax Returns

Universe, Frequency, and Types of Data: Annual study of unaudited individual income tax returns, forms 1040, 1040A, and 1040EZ, filed by U.S. citizens and residents. Data provided on various financial characteristics by size of adjusted gross income, marital status, and by taxable and nontaxable returns. Data by state, based on 100 percent file, also include returns from 1040NR, filed by nonresident aliens plus certain self-employment tax returns.

Type of Data Collection Operation: Annual 2002 stratified probability sample of approximately 176,000 returns broken into sample strata based on the larger of total income or total loss amounts as well as the size of business plus farm receipts. Sampling rates for sample strata varied from 0.05 percent to 100 percent.

Data Collection and Imputation Procedures: Computer selection of sample of tax return records. Data adjusted during editing for incorrect, missing, or inconsistent entries to ensure consistency with other entries on return.

Estimates of Sampling Error: Estimated CVs for tax year 2002: Adjusted gross income less deficit 0.12 percent; salaries and wages 0.21 percent; and tax exempt interest received 1.78 percent. (State data not subject to sampling error.)

Other (nonsampling) Errors: Processing errors and errors arising from the use of tolerance checks for the data.

U.S. NATIONAL CENTER FOR HEALTH STATISTICS (NCHS)

National Health Interview Survey (NHIS)

Universe, Frequency, and Types of Data: Continuous data collection covering the civilian noninstitutionalized population to obtain information on demographic characteristics, conditions, injuries, impairments, use of health services, health behaviors, and other health topics.

Type of Data Collection Operation: Multistage probability sample of 49,000 households (in 198 PSUs) from 1985 to 1994; 36–40,000 households (358 design PSUs) from 1995 on.

Data Collection and Imputation Procedures: Some missing data items (e.g., race, ethnicity) are imputed using a hot deck imputation value. Unit nonresponse is compensated for by an adjustment to the survey weights.

Estimates of Sampling Error: Estimates of Standard Error (SE): For 2003 medically attended injury episodes rates in the past 12 months by falling for: females 29.72 (1.94), and males 26.15 (1.82) per 1,000 population; for 2003 injury episodes rates during the past 12 months inside the home–20.72 (1.18) per 1,000 population.

Other (nonsampling) Errors: The response rate was 93.8 percent in 1996; in 2003, the total household response rate was 89.2 percent, with the final family response rate of 87.9 percent, and the final sample adult response rate of 74.2 percent. (Note: The NHIS sample redesign was conducted in 1995, and the NHIS questionnaire was redesigned in 1997.)

Estimation for the National Health Interview Survey, 1995–2004, Vital and Health Statistics, Series 2 #130.

National Vital Statistics System

Universe, Frequency, and Types of Data: Annual data on births and deaths in the United States.

Type of Data Collection Operation: Mortality data based on complete file of death records, except 1972, based on 50 percent sample. Natality statistics 1951–71, based on 50 percent sample of birth certificates, except a 20 percent to 50 percent sample in 1967, received by NCHS. Beginning 1972, data from some states received through Vital Statistics Cooperative Program (VSCP) and complete file used; data from other states based on 50 percent sample. Beginning 1986, all reporting areas participated in the VSCP.

Data Collection and Imputation Procedures: Reports based on records from registration offices of all states, District of Columbia, New York City, Puerto Rico, Virgin Islands, Guam, American Samoa, and Northern Marianas.

Estimates of Sampling Error: For recent years, there is no sampling for these files; the files are based on 100 percent of events registered.

Other (nonsampling) Errors: Data on births and deaths believed to be at least 99 percent complete.

Sources of Additional Material: U.S. National Center for Health Statistics, *Vital Statistics of the United States,* Vol. I and Vol. II, annual, and *National Vital Statistics Reports.* NCHS Web site at <http://www.cdc.gov/nchs/nvss.htm>.

National Highway Traffic Safety Administration (NHTSA)

Fatality Analysis Reporting System (FARS)

Universe, Frequency, and Types of Data: FARS is a census of all fatal motor vehicle traffic crashes that occur throughout the United States including the District of Columbia and Puerto Rico on roadways customarily open to the public. The crash must be reported to the state/jurisdiction and at least one directly related fatality must occur within thirty days of the crash.

Type of Data Collection Operation: One or more analysts, in each state, extract data from the official documents and enter the data into a standardized electronic database.

Data Collection and Imputation Procedures: Detailed data describing the characteristics of the fatal crash, the vehicles and persons involved are obtained from police crash reports, driver and vehicle registration records, autopsy reports, highway department, etc. Computerized edit checks monitor that accuracy and completeness of the data. The FARS incorporates a sophisticated mathematical multiple imputation procedure to develop a probability distribution of missing blood alcohol concentration (BAC) levels in the database for drivers, pedestrians and cyclists.

Estimates of Sampling Error: Since this is census data, there are no sampling errors.

Other (nonsampling) Errors: Fatal motor vehicle traffic crash data are more than 97 percent complete. However, these data are highly dependent on the accuracy of the police accident reports. Errors or omissions within police accident reports may not be detected.

Sources of Additional Material: The FARS Coding and Validation Manual, ANSI D16.1 Manual on Classification of Motor Vehicle Traffic Accidents (Sixth Edition).